

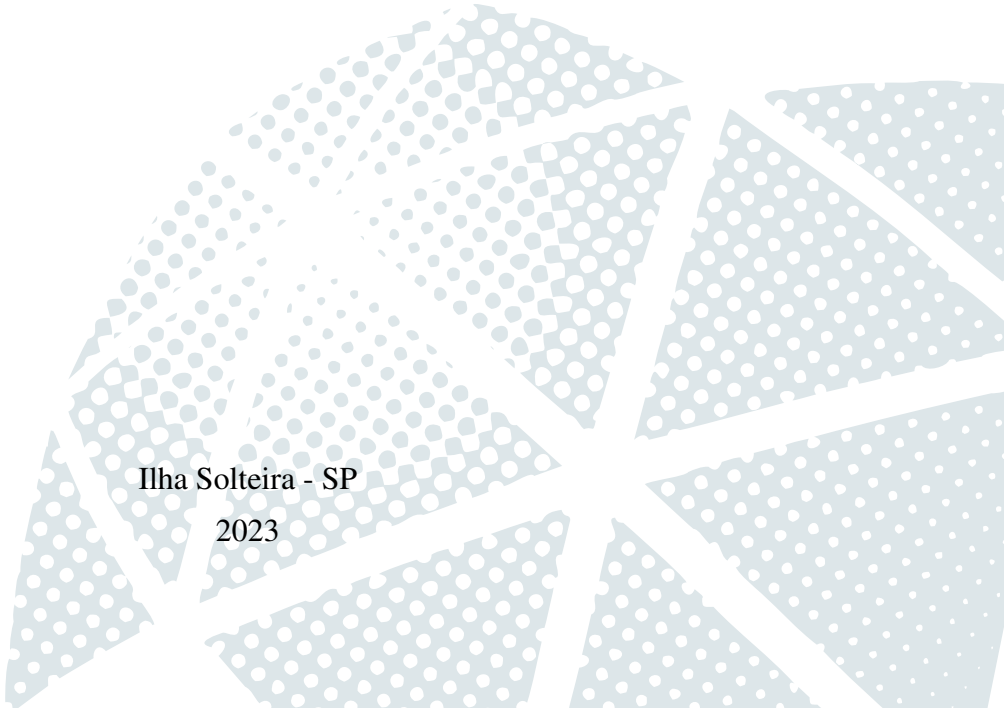
**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA – CAMPUS DE ILHA SOLTEIRA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

REGINALDO JOSÉ DA SILVA

**Diagnóstico da COVID-19 utilizando Redes Neurais
Artificiais baseadas na Teoria da Ressonância Adaptativa**

Ilha Solteira - SP

2023



REGINALDO JOSÉ DA SILVA

Diagnóstico da COVID-19 utilizando Redes Neurais Artificiais baseadas na Teoria da Ressonância Adaptativa

Dissertação apresentada à Faculdade de
Engenharia do Câmpus de Ilha Solteira -
UNESP como parte dos requisitos para ob-
tenção do título de Mestre em Engenharia
Elétrica.

Especialidade: Automação.

Prof.^aDra. Mara Lúcia Martins Lopes
Orientadora

Ilha Solteira - SP
2023

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

S586d Silva, Reginaldo José da.
Diagnóstico da covid-19 utilizando redes neurais artificiais baseadas na teoria da ressonância adaptativa / Reginaldo José da Silva. -- Ilha Solteira: [s.n.], 2023
84 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2023

Orientadora: Mara Lúcia Martins Lopes
Inclui bibliografia

1. ART fuzzy. 2. Biomarcadores sanguíneos. 3. Classificação. 4. Reconhecimento de padrões. 5. SARS-CoV-2. 6. Treinamento continuado.


Raiane da Silva Santos

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Diagnóstico da COVID-19 utilizando Redes Neurais Artificiais baseada na Teoria da Ressonância Adaptativa

AUTOR: REGINALDO JOSÉ DA SILVA

ORIENTADORA: MARA LUCIA MARTINS LOPES

Aprovado como parte das exigências para obtenção do Título de Mestre em Engenharia Elétrica, área: Automação pela Comissão Examinadora:

Prof^a. Dr^a. MARA LUCIA MARTINS LOPES (Participação Presencial)
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. CARLOS ROBERTO MINUSSI (Participação Presencial)
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira - UNESP

Prof. Dr. CARLOS ROBERTO DOS SANTOS JÚNIOR (Participação Presencial)
Campus de Hortolândia / Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - IFSP

Ilha Solteira, 27 de janeiro de 2023

*Dedico este trabalho à minha família, principalmente aos meus pais que sempre me apoiaram,
e também a todos os que sofreram com a pandemia de COVID-19, que as quase 700 mil
vítimas brasileiras jamais sejam esquecidas.*

AGRADECIMENTOS

Agradeço em primeiro lugar à Deus, por me permitir realizar esse sonho e me dar forças diariamente para superar todos os obstáculos.

Aos meus pais, Agnaldo e Delma e aos meus irmãos Rodrigo, Rafaela e Rafael. A todos minha eterna gratidão, respeito e admiração por sempre me incentivarem e torcerem pela minha vitória. Agradeço também a todos os meus familiares que oraram e me apoiaram de alguma forma.

À minha orientadora, Prof.^aDra. Mara Lúcia Martins Lopes pela amizade, apoio e paciência, por acreditar em minha capacidade e por me colocar na direção correta. Não há palavras para agradecê-la.

À Prof.^aDra. Angela Leite Moreno, da Universidade Federal de Alfenas, por me ajudar nos momentos de dúvida, sem seu incentivo jamais teria chegado até aqui. O meu muito obrigado!

Ao Prof. Dr. Carlos Roberto Minussi pelo incentivo e contribuições.

Agradeço ao meu colega André Luiz Caliar Costa da USP de Ribeirão Preto pelas conversas e por me ajudar com as dúvidas da parte biológica.

Agradeço aos meus amigos Bruno, Cibele, Edivan e Gustavo pelos conselhos, incentivos e amizade incondicional. Obrigado por acreditarem em mim quando nem eu mesmo conseguia acreditar.

Aos meus amigos e colegas do laboratório SINTEL, em especial à Paula e Joaquim por me apoiarem e pelos conhecimentos compartilhados.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, pelo apoio financeiro.

Enfim agradeço a todas as pessoas que direta ou indiretamente me incentivaram a chegar até aqui, a todos vocês o meu respeito, carinho e gratidão.

“Esteja seguro, seja inteligente, seja gentil.”

Tedros Adhanom Ghebreyesus

RESUMO

A COVID-19 espalhou-se por todo o mundo levando a óbito cerca de 6,6 milhões de pessoas, diversas estratégias foram utilizadas para evitar a propagação dessas infecções, estratégias essas que influenciaram a vida de diversas pessoas. A forma mais eficaz para controle da doença é a detecção e o isolamento que limita a propagação do vírus, consequentemente, salvando vidas. Dentre os diversos testes existentes para detecção e triagem da COVID-19, o teste RT-PCR é considerado o padrão-ouro. No entanto, o teste possui baixa sensibilidade de diagnóstico, resultando numa taxa elevada de falsos negativos. Diversos estudos clínicos mostraram alterações hematológicas e bioquímicas em pacientes infectados, e diante da necessidade por diagnósticos eficientes e mais precisos em fases iniciais, surgiram estudos que utilizam estas alterações com algoritmos de aprendizagem de máquinas para ajudar a diagnosticar a doença. Desta forma, o presente trabalho tem por objetivo avaliar o desempenho de diferentes redes neurais baseadas na Teoria da Ressonância Adaptativa (ART), bem como desenvolver uma metodologia rápida e eficaz para a classificação da COVID-19 utilizando exames laboratoriais. Diversas redes ART foram testadas, dentre elas autoexpansíveis e com treinamento continuado, vale ressaltar que as redes ART autoexpansíveis realizam o processo de classificação de forma mais rápida que as redes ART tradicionais, enquanto as redes com treinamento continuado permitem a aprendizagem de modo on-line, tornando o treinamento e a análise mais eficazes. Um total de doze conjuntos de dados de cinco países diferentes contendo diversos biomarcadores sanguíneos foram utilizados, a metodologia foi dividida em duas etapas, a primeira responsável por avaliar a melhor rede ART para estes dados e a segunda por avaliar o comportamento desta rede na classificação de pacientes com COVID-19 em dados de diferentes países. A rede ART *Fuzzy* Autoexpansível com Treinamento Continuado obteve o melhor desempenho, superando na maior parte das vezes, modelos tradicionais e de aprendizado profundo frequentemente utilizados pela literatura. Além disso, os atributos selecionados mostraram excelente capacidade preditiva ao classificar com eficiência pacientes de diferentes países. Logo, a metodologia proposta pode servir como um método complementar ao RT-PCR, diminuindo a taxa de falsos positivos ou negativos, ou até mesmo como uma alternativa de baixo custo para a detecção.

Palavras-chave: ART *Fuzzy*. Biomarcadores sanguíneos. Classificação. Reconhecimento de padrões. SARS-CoV-2. Treinamento Continuado.

ABSTRACT

COVID-19 has spread all over the world, killing about 6.6 million people. Several strategies have been used to prevent the spread of these infections, strategies that have influenced the lives of many people. The most effective way to control the disease is the detection and isolation that limits the spread of the virus, consequently saving lives. Among the various existing tests for detection and screening of COVID-19, the RT-PCR test is considered the gold standard. However, the test has low diagnostic sensitivity, resulting in a high false negative rate. Several clinical studies have shown hematological and biochemical alterations in infected patients, and facing the need for efficient and more accurate diagnosis in early stages, studies have emerged that use these alterations with machine learning algorithms to help diagnose the disease. Thus, the present work aims to evaluate the performance of different neural networks based on Adaptive Resonance Theory (ART), as well as to develop a fast and effective methodology for the classification of COVID-19 using laboratory tests. Several ART networks were tested, including self-expanding and continuously trained networks; it is worth noting that self-expanding ART networks perform the classification process faster than traditional ART networks, while continuously trained networks allow learning in an online manner, making training and analysis more effective. A total of twelve datasets from five different countries containing various blood biomarkers were used, the methodology was divided into two stages, the first responsible for evaluating the best ART network for these data and the second for evaluating the behavior of this network in classifying patients with COVID-19 in data from different countries. The Fuzzy Self-expandable ART network with Continuous Training obtained the best performance, outperforming for the most part, traditional and deep learning models often used by the literature. Furthermore, the selected attributes showed excellent predictive ability by efficiently classifying patients from different countries. Therefore, the proposed methodology can serve as a complementary method to RT-PCR, decreasing the false positive or negative rate, or even as a low cost alternative for detection.

Keywords: Fuzzy ART. Blood biomarkers. Classification. Pattern recognition. SARS-CoV-2. Continuous Training.

LISTA DE FIGURAS

Figura 1	Arquitetura ART.	25
Figura 2	Fluxograma de treinamento da rede ART <i>Fuzzy</i> Autoexpansível.	29
Figura 3	Fluxograma de treinamento da rede ART Euclidiana Autoexpansível.	32
Figura 4	Arquitetura ARTMAP Fuzzy.	33
Figura 5	Fluxograma de treinamento da rede ARTMAP <i>Fuzzy</i>	36
Figura 6	Fluxograma de treinamento da rede ART <i>Fuzzy</i> Autoexpansível com Treinamento Continuado.	40
Figura 7	Fluxograma de treinamento da rede ARTMAP <i>Fuzzy</i> Autoexpansível com Treinamento Continuado.	42
Figura 8	Mapa de calor da matriz de correlação.	45
Figura 9	Gráfico de importância por valores SHAP.	49
Figura 10	Histograma e estimativa da função densidade de probabilidade para cada um dos atributos separados por classe.	51
Figura 11	Visão geral da validação.	55
Figura 12	Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para o Dados ₁	57
Figura 13	Curva ROC para o conjunto de Dados ₁	58
Figura 14	Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados ₂	59
Figura 15	Curva ROC para o conjunto de Dados ₂	60
Figura 16	Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados ₃	61
Figura 17	Curva ROC para o conjunto de Dados ₃	62
Figura 18	Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados ₃ Balanceado.	63
Figura 19	Curva ROC para o conjunto de Dados ₃ Balanceado.	64
Figura 20	Curvas ROC obtidas pelas redes ART para os conjuntos Dados _{4T} e Dados _{4TA}	66
Figura 21	Curva ROC e AUC para o conjunto de dados combinados.	70
Figura 22	Curva ROC e AUC obtidos por cada modelo treinado com os dados combinados e validado pelo Dados ₁₂	71

Figura 23	Curva ROC e AUC obtidos por cada modelo treinado com os dados combinados e validado pelo Dados ₁	72
-----------	---	----

LISTA DE TABELAS

Tabela 1	Visão geral dos trabalhos relatados.	23
Tabela 2	Quantidade de dados e atributos.	43
Tabela 3	Atributos contidos em cada conjunto de dados.	44
Tabela 4	Média e desvio padrão dos atributos.	46
Tabela 5	Seleção de atributos por diferentes métodos.	50
Tabela 6	Interpretação dos Valores AUC.	53
Tabela 7	Resultados obtidos por cada modelo para o Dados ₁	56
Tabela 8	Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados ₁	57
Tabela 9	Resultados obtidos por cada modelo para o Dados ₂	58
Tabela 10	Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados ₂	59
Tabela 11	Resultados obtidos por cada modelo para o Dados ₃	60
Tabela 12	Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados ₃	61
Tabela 13	Resultados obtidos por cada modelo para o Dados ₃ Balanceado.	62
Tabela 14	Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados ₃ Balanceado.	63
Tabela 15	Resultados obtidos para os Dados _{4T}	64
Tabela 16	Resultados após remoção de hematócrito e hemoglobina dos Dados _{4T}	65
Tabela 17	Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados _{4T}	65
Tabela 18	Comparação dos resultado para os conjuntos de dados utilizados.	67
Tabela 18	Comparação dos resultado para os conjuntos de dados utilizados.	68
Tabela 19	Tempo computacional (em segundos) para cada conjunto de dados.	69
Tabela 20	Resultados obtidos por cada modelo para o conjunto de dados combinados.	70
Tabela 21	Resultados obtidos por cada modelo treinado com os dados combinados e validado pelo Dados ₁₂	71
Tabela 22	Resultados obtidos por cada modelo treinado com os dados combinados e validado pelo Dados ₁	72

LISTA DE ABREVIACES E SIGLAS

ACC	Acurcia
ADABoost	<i>Adaptive Boosting</i>
AE	Autoexpansvel
ALT	Alanina Aminotransferase
ART	<i>Adaptive Resonance Theory</i>
ARTFAETC	ART Fuzzy Autoexpansvel com Treinamento Continuado
AST	Aspartato Aminotransferase
AUC	<i>Area Under the Curve</i>
BA	Basfilos
BAT	Contagem de Basfilos
CK-MB	Isoenzima Creatino Quinase MB
CNN	Rede Neural Convolutacional
DNN	Redes Neural Profunda
DT	rvore de Deciso
EO	Eosinfilos
EOT	Contagem de Eosinfilos
ET	<i>Extremely Randomized Trees</i>
FN	Falso Negativo
FP	Falso Positivo
GBDT	<i>Gradient Boosting Decision Trees</i>
GSH	Glutaciona Reduzida
HCT	Hematcrito
HGB	Hemoglobina
INR	Razo Normalizada Internacional
IRCCS	<i>Scientific Institute for Research, Hospitalization and Healthcare</i>
KNN	<i>K-Nearest Neighbors</i>
LDH	Lactato Desidrogenase
LR	Regresso Logstica
LSTM	<i>Long Short-Term Memory</i>
LY	Linfcitos
LYT	Contagem de Linfcitos
MCH	<i>Mean Corpuscular Hemoglobin</i>
MCHC	<i>Mean Corpuscular Hemoglobin Concentration</i>
MCV	<i>Mean Corpuscular Volume</i>

MLP	<i>Multilayer Perceptron</i>
MO	Monócitos
MOT	Contagem de Monócitos
NB	<i>Naïve Baye</i>
NE	Neutrófilos
NET	Contagem de Neutrófilos
PCR	Proteína C Reativa
PDW	Índice de Anisocitose Plaquetária
PLT	Plaquetas
RBC	Glóbulos Vermelhos
RDW	Amplitude de Distribuição dos Glóbulos Vermelhos
RF	<i>Random Forest</i>
RNA	Rede Neural Artificial
RNN	Rede Neural Recorrente
ROC	<i>Receiver Operating Characteristic</i>
RT-PCR	Transcrição Reversa seguida de Reação em Cadeia da Polimerase
<i>Se</i>	Sensibilidade
<i>Sp</i>	Especificidade
SGD	<i>Stochastic Gradient Descent</i>
SVM	<i>Support Vector Machine</i>
TC	Treinamento Continuado
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
TWRF	<i>Three-Way Random Forest</i>
WBC	Glóbulos Brancos
XGB	<i>eXtreme Gradient Boosting</i>

SUMÁRIO

1	INTRODUÇÃO	15
2	REVISÃO BIBLIOGRÁFICA	19
3	TEORIA DA RESSONÂNCIA ADAPTATIVA	25
3.1	REDE NEURAL ART <i>FUZZY</i> AUTOEXPANSÍVEL	26
3.2	REDE NEURAL ART EUCLIDIANA AUTOEXPANSÍVEL	30
3.3	REDE NEURAL ARTMAP <i>FUZZY</i>	33
3.4	REDE NEURAL ART AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO	37
3.5	REDE NEURAL ARTMAP AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO	41
4	METODOLOGIA	43
4.1	CONJUNTOS DE DADOS	43
4.2	PRIMEIRA ETAPA	44
4.3	SEGUNDA ETAPA	46
4.4	MÉTRICAS DE AVALIAÇÃO	52
4.5	CLASSIFICADOR	54
4.6	VALIDAÇÃO	54
5	RESULTADOS COMPUTACIONAIS	56
5.1	PRIMEIRA ETAPA	56
5.1.1	CONJUNTO DE DADOS ₁	56
5.1.2	CONJUNTO DE DADOS ₂	58
5.1.3	CONJUNTO DE DADOS ₃	60
5.1.4	CONJUNTO DE DADOS ₃ BALANCEADO	62
5.1.5	CONJUNTO DE DADOS _{4T}	64
5.2	COMPARAÇÃO DOS RESULTADOS OBTIDOS NA PRIMEIRA ETAPA	66
5.3	SEGUNDA ETAPA	70
6	CONCLUSÕES	73
6.1	SUGESTÕES PARA TRABALHOS FUTUROS	74
	REFERÊNCIAS	75

1 INTRODUÇÃO

Um grupo de casos de uma doença incomum semelhante à pneumonia, mas que não respondia bem aos tratamentos, foi relatado pela comissão Municipal de Saúde de Wuhan, China, em 12 de dezembro de 2019 ([Centros de Controle e Prevenção de Doenças, 2022](#)). No dia 31 de dezembro, a organização Mundial da Saúde (OMS) tornou público a declaração da Comissão municipal de Saúde de Wuhan, alertando por uma pneumonia causada por um vírus desconhecido ([Centros de Controle e Prevenção de Doenças, 2022](#)). No dia 10 de janeiro de 2020 a OMS anuncia que o surto é causado pelo novo coronavírus (2019-nCoV), posteriormente chamado de SARS-CoV-2 ([Centros de Controle e Prevenção de Doenças, 2022](#)). Em 11 de fevereiro a OMS anuncia o nome COVID-19, como sendo o nome oficial da doença causada pelo vírus SARS-CoV-2 ([Centros de Controle e Prevenção de Doenças, 2022](#)). Somente em 11 de março de 2020 a OMS declara a COVID-19 uma pandemia ([Centros de Controle e Prevenção de Doenças, 2022](#)), e quatro meses após a identificação dos primeiros casos, havia mais de 2,3 milhões de casos confirmados e um total de 162 mil mortes em todo o mundo ([Organização Mundial da Saúde, 2020a](#)). Atualmente o vírus já infectou mais de 650 milhões de pessoas e causou a morte de 6,6 milhões em todo o mundo ([Organização Mundial da Saúde, 2020b](#); [DONG; DU; GARDNER, 2020](#)). A perda de vidas devido a uma pandemia leva a danos inequívocos a sociedade em geral.

As estratégias necessárias para forçar o achatamento da curva e evitar a propagação dessas infecções que levam a novos casos de COVID-19 ([BLOCK et al., 2020](#)), influenciou severamente a vida cotidiana das pessoas, impactou profundamente a economia global, atingindo diretamente diferentes setores da indústria, instituições de educação e instalações de saúde ([MISHRA et al., 2020](#)). Ao longo da pandemia, a disseminação rápida do SARS-CoV-2 elevou as taxas de infecção, e trouxe desafios sem precedentes para os sistemas de saúde em todo mundo, já que muitos países em desenvolvimento não tem estrutura para comportar todos os pacientes infectados ([TANNE et al., 2020](#)). Em particular, outro ponto muito importante é o risco aumentado de infecção dos profissionais de saúde que atuam na linha de frente, sendo esta a maior vulnerabilidade dos sistemas de saúde, considerando que a maioria dos profissionais de saúde não consegue trabalhar remotamente, ao se infectarem existe um declínio da disponibilidade do profissional de saúde, influenciando no dimensionamento do quantitativo de profissionais aptos à jornada de trabalho ([GODLEE, 2020](#); [MARTINS et al., 2022](#)). Além disso, fatores como, altos custos para manter o sistema de saúde funcionando, escassez de equipamentos de proteção, baixo número de leitos de UTI e ventiladores mecânicos, e por fim, a falta de oxigênio hospitalar para pacientes em estados mais graves, um insumo crucial para manter o

paciente com vida no leito de UTI (SALISBURY, 2020). Estes impedimentos e falhas levaram os sistemas de saúde a serem incapazes de diagnosticar e testar pacientes na taxa exigida. Esta sobrecarga do sistema de saúde exigiu a necessidade de utilizar múltiplas formas de diagnóstico e triagem para atender a demanda, exigindo métodos de diagnósticos mais rápidos e eficazes (BARBOSA, 2020).

Para que a prevenção da COVID-19 ocorra é importante que a população crie hábitos nunca exigidos, e se adapte a esses hábitos para que o número de infecções diminua e o sistema de saúde não entre em colapso, mas também é preciso que haja uma detecção eficaz, isolamento e atendimento médico imediato para indivíduos infectados (WEISSLEDER et al., 2020; COR-MAN et al., 2020; ZHANG et al., 2021; SALATHÉ et al., 2020), evitando que o vírus se espalhe rapidamente, sendo estes, os métodos mais eficientes para se combater a disseminação da COVID-19. Dentre os diversos testes existentes para detecção e triagem da COVID-19, o teste de transcrição reversa seguida de reação em cadeia da polimerase (RT-PCR) que detecta o material genético do vírus é considerado o padrão-ouro para detecção da COVID-19, realizado a partir da coleta de secreções nasofaríngeas e/ou orofaríngeas (REIS; SANTOS, 2020). Embora o RT-PCR possua alta especificidade analítica, ele possui baixa sensibilidade diagnóstica, tendo dificuldade em distinguir os verdadeiros positivos dos verdadeiros negativos em pacientes afetados pela COVID-19 (PRADHAN et al., 2022; AI et al., 2020), uma vez que nos primeiros dias de infecção e apresentação dos sintomas clínicos, a concentração inicial do vírus não é alta o suficiente para um bom teste, e ocorrem falsos positivos facilmente nas amostras testadas (TAHAMTAN; ARDEBILI, 2020). Outro ponto, é que o SARS-CoV-2 é um vírus de RNA com alta propensão de adquirir mutações no seu genoma, e ao longo da atual pandemia, vem ocorrendo a evolução do vírus com o surgimento de cepas com diversas eficiências de replicação e manifestação clinicopatológicas (SHRESTHA et al., 2022), o que pode ser outro motivo responsável pelas diferenças na sensibilidade analítica de ensaios RT-PCR disponíveis e, consequentemente, os resultados de teste divergentes obtidos. Além disso, o teste possui algumas limitações para testagem de pacientes em larga escala, exigindo a necessidade de laboratórios certificados, equipamentos caros e pessoal treinado, além da demora na obtenção dos resultados (LI et al., 2020; KISTENEV et al., 2022). Tais limitações tornam-se ainda mais severas naqueles países com recursos limitados como em países subdesenvolvidos (BRINATI et al., 2020).

Outro meio utilizado para diagnósticos de COVID-19 é a tomografia computadorizada e o raio-X, que também desempenham um papel importante de diagnóstico em muitas outras doenças, sendo um poderoso auxiliar na detecção e julgamento da progressão da doença (RI-QUELME; AKHLOUFI, 2020; CHUNG et al., 2020). Entretanto, este tipo de diagnóstico não possibilita um manejo eficiente dos pacientes e nem do controle da doença para evitar a disseminação do vírus, visto que bons resultados com uma sensibilidade e especificidade mais alta são obtidos apenas quando os pacientes já estão em estágios mais avançados da COVID-19 ou

outras doenças ([KOVÁCS et al., 2021](#); [PAN et al., 2020](#); [RAPTIS et al., 2020](#)).

Em muitos casos a admissão na UTI ocorre entre o oitavo e o décimo dia de infecção, quando os sintomas respiratórios mais graves como dispneia são aparentes e compromete a rotina do paciente, neste ponto a infecção já comprometeu parte do pulmão dos pacientes infectados e as chances de recuperação são diminuídas a cada hora para pacientes em estado avançado ([HUANG et al., 2020](#); [WANG et al., 2020](#)). Desta forma, métodos de diagnósticos precisos podem conseguir detectar diferenças nos padrões de exames laboratoriais clínicos de rotina entre os pacientes infectados e não infectados, podem aumentar em muitas vezes as chances dos pacientes infectados não evoluírem para casos graves, predizer pacientes que não estão infectados o quanto ele pode ser suscetível a causas graves de COVID-19, e por fim, um diagnóstico precoce e um manejo mais eficiente dos paciente já infectados possibilita que medidas de prevenção possam ser tomadas, evitando o contágio excessivo.

Atualmente o aprendizado de máquina tem sido uma ferramenta importante para auxiliar os médicos no diagnóstico de diversas doenças ([AHSAN; LUNA; SIDDIQUE, 2022](#); [MIRZAEI; ADELI, 2022](#)) inclusive a COVID-19 ([WU et al., 2020](#); [SUN et al., 2020b](#)), seja por meio de imagens de raio-x ou tomografia computadorizada ([HUSSAIN et al., 2020](#); [KHUZANI; HEIDARI; SHARIATI, 2021](#); [MARAR; MARAR, 2022](#)). Muitos estudos clínicos mostraram alterações hematológicas em pacientes infectados ([CARELLI et al., 2020](#); [SILVA et al., 2022](#); [LIU et al., 2020](#); [KHARTABIL et al., 2020](#)). Outros estudos focaram em analisar parâmetros hematológicos ao decorrer do curso da doença e a um pior desfecho clínico, em pacientes infectados por SARS-CoV-2 ([PALLADINO, 2021](#); [SUN et al., 2020a](#); [ELDERDERY et al., 2022](#)). Diante da demanda por diagnósticos eficientes e mais precisos em fases iniciais da COVID-19, os atributos relevantes de exames laboratoriais podem ser utilizados para treinar algoritmos de aprendizados de máquina para previsão de risco e evolução da doença, e classificação (diagnósticos) precoce dos pacientes infectados. Além disso, exames laboratoriais como o de sangue tem a vantagem de ser relativamente mais barato que o RT-PCR e exames de imagens como o raio-X e tomografia computadorizada, consequentemente, evitando exposição à radiação. Desta forma, uma avaliação por hemograma pode possibilitar o seu uso em larga escala, com algoritmos especialistas, atendendo a demanda em países subdesenvolvidos quanto nos desenvolvidos, evitando problemas sociais e econômicos.

Sendo assim, este estudo tem como objetivo avaliar o desempenho de diferentes redes neurais baseadas na Teoria da Ressonância Adaptativa (do inglês *Adaptive Resonance Theory* – ART), bem como desenvolver uma metodologia para a classificação da COVID-19 utilizando exames laboratoriais. Para isso, redes neurais ART, autoexpansíveis e com treinamento contínuo foram utilizadas.

No [Capítulo 2](#) é apresentada uma revisão dos trabalhos realizados pela literatura, no [Capítulo 3](#) apresenta uma visão geral da Teoria da Ressonância Adaptativa, além de descrever as

redes ART que foram utilizadas. Já no [Capítulo 4](#) é apresentada a metodologia, a qual engloba a descrição dos conjuntos de dados a serem utilizados, a seleção dos atributos, as métricas de avaliação e a validação empregada. Dedicada aos resultados, no [Capítulo 5](#) são apresentados os principais resultados obtidos, bem como as comparações e as discussões dos resultados obtidos pelas redes. Finalmente no [Capítulo 6](#) apresentam-se as considerações finais.

2 REVISÃO BIBLIOGRÁFICA

Na literatura há uma grande variedade de estudos utilizando aprendizado de máquina para a detecção da COVID-19 a partir de exames laboratoriais. Um dos primeiros artigos foi realizado por [Wu et al. \(2020\)](#) no qual os autores utilizaram o algoritmo *Random Forest* (RF) e dados coletados de vários hospitais na China, totalizando 253 amostras de 169 pacientes com suspeita de COVID-19. O conjunto de dados utilizado consistia em 24 parâmetros hematológicos e 25 parâmetros bioquímicos, dos quais 11 foram selecionados para as simulações. Atributos como bilirrubina total, glicose, creatinina, lactato desidrogenase (LDH), isoenzima creatinoquinase MB (CK-MB), cálcio, magnésio, índice de anisocitose plaquetária (PDW) e potássio foram identificadas pelo modelo com relevantes para predição de pacientes com COVID-19. O modelo atingiu uma área abaixo da curva (AUC) *receiver operating characteristic* (ROC) de 0,9926, sensibilidade de 100,00% e uma especificidade de 94,44%. Embora tenha alcançado bons resultados durante a validação interna, ao ser testado com amostras de sangue externas, o desempenho do modelo foi inferior.

Utilizando dados que foram coletados do Hospital Israelita Albert Einstein em São Paulo no Brasil entre os dias 17 e 30 de março de 2020 [Batista et al. \(2020\)](#) treinaram cinco algoritmos de aprendizado de máquina, sendo eles: redes neurais artificiais (RNA), *eXtreme Gradient Boosting* (XGB), RF, regressão logística (LR) e *Support Vector Machine* (SVM). Os melhores resultados foram obtidos com o SVM e RF, obtendo uma AUC de 0,847. Os atributos que mais contribuíram foram linfócitos, glóbulos brancos e eosinófilos.

[Sun et al. \(2020b\)](#) utilizaram o algoritmo SVM para prever os pacientes que evoluirão para casos graves ou críticos da COVID-19 utilizando dados de 336 pacientes do Centro Clínico de Saúde Pública de Shanghai na China. O modelo utilizado alcançou uma sensibilidade de 77,5%, especificidade de 78,4% e AUC de 0,9757, atributos como idade, glutathione reduzida (GSH), CD3 e proteína total foram os mais significativas. Apesar do modelo utilizado ter sido robusto e eficaz na previsão dos casos graves/críticos, o tamanho do conjunto de dados utilizado foi relativamente pequeno.

[Tordjman et al. \(2020\)](#) desenvolveram uma estratégia para estimar a probabilidade de um paciente ser diagnosticado com COVID-19 antes do RT-PCR utilizando LR binária. Os dados utilizados são de 605 pacientes dos departamentos de emergência, medicina interna e doenças infecciosas de quatro hospitais na França, coletados entre 10 de março e 30 de abril de 2020. A estratégia proposta obteve AUC de 0,918, sensibilidade e especificidade de 79% e 90%, respectivamente. Linfócitos, eosinófilos, basófilos e neutrófilos foram identificadas como os

mais relevantes para predição.

Yang et al. (2020) desenvolveram um modelo de aprendizado de máquina para prever pacientes com COVID-19 usando exames laboratoriais de rotina. Dados de 3356 pacientes obtidos do centro médico *New York Presbyterian/Weill Cornell* nos Estados Unidos entre 11 de março e 29 de abril de 2020 foram utilizados para os testes. Algoritmos como LR, RF e *Gradient boosting decision tree* (GBDT) foram utilizados, sendo este último o melhor método. Na validação interna o modelo alcançou uma AUC de 0,854, enquanto na validação externa, o modelo alcançou uma AUC de 0,838. Os autores ainda apontam que valores maiores de LDH, proteína C reativa (PCR) e ferritina favorecem a predição correta de pacientes positivos.

Visando prever pacientes negativos para COVID-19 utilizando exames de sangue e sexo, Joshi et al. (2020) desenvolveram uma ferramenta de apoio à decisão usando LR. Os dados utilizados no treinamento foram coletados no *Stanford Health Care* entre 1 de março de 2020 e 20 de março de 2020 e são referentes a 390 pacientes, sendo 33 confirmados com SARS-CoV-2 via RT-PCR positivo. Atributos como contagem de neutrófilos, contagem de linfócitos e hematócrito foram selecionados. Dados de diferentes centros foram utilizados para verificar a generalização da ferramenta, para os dados do *Stanford Health Care* a ferramenta proposta alcançou AUC de 0,78, sensibilidade de 93% e especificidade de 43%. Para os dados de pacientes de Seattle em Washington, AUC de 0,75, com dados de pacientes em Chicago, Illinois AUC de 0,76. Por fim, com os dados de pacientes da Coreia do Sul, a ferramenta alcançou AUC de 0,81.

Goodman-Meza et al. (2020) combinaram as probabilidades de algoritmos como RF, LR, SVM, *multilayer perceptron* (MLP), *stochastic gradient descent* (SGD), XGB e *adaptive boosting* (ADABOOST) para detectar pacientes com COVID-19. Os dados utilizados foram obtidos de pacientes testados para SARS-CoV-2 no pronto-socorro ou internação do Sistema de Saúde da Universidade da Califórnia em Los Angeles entre 1º de março e 24 de maio de 2020, dos 1455 casos contidos no conjunto de dados, 1.273 são casos negativos e apenas 182 casos positivos para COVID-19. O modelo proposto alcançou AUC de 0,91 após a adição ferritina, PCR e LDH.

AlJame et al. (2020) produziram o ERLX, um modelo de aprendizado conjunto que empilha algoritmos como *extremely randomized trees* (ET), RF, LR e XGB para diagnosticar pacientes com COVID-19 a partir de exames de sangue, o modelo é separado em dois níveis, as previsões do primeiro nível, que consiste nos modelos ET, RF e LR, servem como entradas para o classificador do segundo nível (XGB). O conjunto de dados utilizado foi disponibilizado pelo Hospital Israelita Albert Einstein no Brasil e contém informações sobre 5.644 pacientes, dos quais 559 são positivos para COVID-19. O ERLX proposto obteve um bom desempenho, com AUC de 0,9938. Monócitos, plaquetas e glóbulos brancos foram os atributos que mais contribuíram para as previsões precisas do modelo.

Brinati et al. (2020) utilizaram *machine learning* para desenvolver um modelo para prever quais pacientes estavam ou não infectados pela COVID-19. Os dados utilizados foram disponibilizados pelo IRCCS do Hospital San Raffaele contendo informações de 279 pacientes internados e selecionados aleatoriamente entre final de fevereiro de 2020 a meados de março do mesmo ano, 177 pacientes tiveram teste RT-PCR positivo para COVID-19. Diversos algoritmos foram testados, dentre eles: árvore de decisão (DT), ET, *k-nearest neighbors* (KNN), LR, *Naïve Baye* (NB), RF, SVM e *three-way random forest* (TWRf). O melhor desempenho foi obtido pelo RF atingindo AUC de 0,84, sensibilidade e especificidade de 92% e 65%, respectivamente. Os atributos mais relevantes apontados pelo modelo foram aspartato aminotransferase (AST), linfócitos, LDH, PCR, glóbulos brancos, eosinófilos, alanina aminotransferase (ALT), idade e neutrófilos.

Utilizando um conjunto de dados contendo 18 atributos hematológicos e bioquímicos de 600 pacientes do Hospital Israelita Albert Einstein em São Paulo, no Brasil Alakus e Turkoglu (2020) treinaram seis arquiteturas de aprendizado profundo como RNA, rede neural convolucional (CNN), *long short-term memory* (LSTM), rede neural recorrente (RNN) e modelos profundos híbridos CNNLSTM e CNNRNN para detectar pacientes com COVID-19. Na validação cruzada 10-fold, os melhores resultados foram observados com o modelo LSTM, alcançando AUC de 0,625. Resultados melhores foram alcançados na validação *hold-out* com o modelo híbrido CNNLSTM, obtendo AUC de 0,90.

A partir de algoritmos como RF, DNN e XGB, Kukar et al. (2021) criaram o *Smart Blood Analytics* (SBA), um modelo para diagnóstico de COVID-19 utilizando exames de sangue. Os dados utilizados foram coletados de 5333 pacientes internados no Departamento de Doenças Infecciosas do Centro Médico Universitário de Ljubljana na Eslovênia. O melhor resultado foi obtido pelo XGB com AUC de 0,97. Além disso, os autores relataram que exames como concentração de hemoglobina corpuscular média, contagem de eosinófilos, albumina, razão normalizada internacional (INR) e atividade de protrombina foram os mais importantes para o diagnóstico.

Utilizando dados de pacientes extraídos dos prontuários eletrônicos do Hospital Universitário de Kepler em Linz, Áustria, de 1 de março de 2020 a 30 de abril de 2020, Tschoellitsch et al. (2021) empregaram o algoritmo RF para prever o resultado do RT-PCR utilizando exames de sangue. O modelo obteve AUC de 0,74, a contagem de glóbulos brancos foi classificada como o recurso mais importante, seguido da amplitude de distribuição dos glóbulos vermelhos (RDW), hemoglobina e cálcio sérico.

Dois modelos foram treinados por Soltan et al. (2021), um dos modelos foi treinado para identificar pacientes com COVID-19 nos serviços de emergência como positivos, enquanto o outro foi treinado para verificar a necessidade de internação. Os autores utilizaram uma grande base de dados com 115394 pacientes, dos quais 437 testaram positivo para COVID-19, foram

coletados dos serviços de medicina aguda e de emergência dos Hospitais da Universidade de Oxford no Reino Unido, algoritmos como XGB, LR e RF. Para ambos os modelos, o melhor desempenho foi com o XGB, o primeiro modelo obteve AUC de 0,939, tendo como atributos mais relevantes a PCR, eosinófilos, basófilos, oxigênio, frequência respiratória e gases no sangue como cálcio e metemoglobina. O segundo modelo obteve AUC de 0,94, atributos como eosinófilos, basófilos, PCR, glóbulos branco e gases no sangue foram os mais relevantes para prever internação.

[Cabitza et al. \(2021a\)](#) avaliaram algoritmos com RF, NB, LR, SVM e KNN na previsão de pacientes com COVID-19. Foram utilizados dados de pacientes internados entre 19 de fevereiro a 31 de maio de 2020 no Hospital San Raphael, na Itália contendo 72 atributos. A AUC obtida pelos modelos ficou entre 0,74 e 0,90. Na validação interna-externa os modelos obtiveram AUC entre 0,75 a 0,78 e especificidade entre 92% e 96%. Atributos com capacidades preditivas mais significativas identificadas pelos modelos foram LDH, AST, PCR, cálcio, glóbulos brancos, idade, glóbulos brancos e fibrinogênio.

[Alves et al. \(2021\)](#) construiu uma modelo baseado em árvore de decisão para auxiliar no diagnóstico de COVID-19, além disso, o modelo desenvolvido fornece uma explicação que pode ser dada às equipes de saúde. O modelo obteve uma AUC de 0,87, sensibilidade de 67% e especificidade de 91%. Os cinco atributos mais importantes identificados pelo modelo foram glóbulos brancos, plaquetas, eosinófilos, PCR e monócitos.

[Rahman et al. \(2022\)](#) desenvolveu o QCoVSML um modelo de aprendizado de máquina empilhado que usa um sistema de pontuação por nomograma para a detecção de COVID-19 baseado em biomarcadores. Sete conjuntos de dados diferentes, disponíveis publicamente foram utilizados, atributos como contagem de glóbulos brancos, monócitos, linfócitos e idade foram selecionados por serem mais importantes para a previsão da doença COVID-19. O modelo combinado QCoVSML obteve uma AUC de 0,96 e ACC de 91,45%.

A [Tabela 1](#) apresenta uma visão geral dos trabalhos relatados. Analisando-a nota-se que biomarcadores como idade, sexo, febre, cálcio, eosinófilos, basófilos, hemoglobina, glóbulos brancos, plaquetas, linfócitos, creatinina, LDH, PCR, ferritina, CD3, AST, glóbulos brancos, neutrófilos e monócitos apresentaram maiores capacidades preditores para o diagnóstico de COVID-19, tais biomarcadores já se mostraram correlacionados com a triagem para COVID-19 e a previsão de condições graves e mortalidade ([KERMALI et al., 2020](#); [JI et al., 2020](#); [SAMPRATHI; JAYASHREE, 2021](#); [GOROG et al., 2022](#); [KEYKAVOUSI et al., 2022](#)).

Embora os modelos utilizados tenham alcançado bons resultados, eles possuem algumas limitações tais como treinamento *off-line*, esquecimento catastrófico ([GROSSBERG, 2020](#)), aprendizado lento e sensibilidade a *outliers*. Baseando-se nisso, no capítulo seguinte são apresentadas as redes baseadas na Teoria da Ressonância Adaptativa, teoria concebida para solucionar a maioria das limitações existentes nos modelos tradicionais.

Tabela 1 - Visão geral dos trabalhos relatados.

Autor	Algoritmos	Atributos relevantes	Conjunto de Dados	Dados Positivos	Desempenho
Wu et al. (2020)	RF	Bilirrubina total, glicose, creatinina, LDH, CK-MB, potássio, proteína total, cálcio, magnésio, PDW e basófilos	253	105	AUC: 0,9926
Batista et al. (2020)	RNA, XGB, RF, LR e SVM	Linfócitos, glóbulos brancos e eosinófilos	235	102	AUC: 0,85
Sun et al. (2020b)	SVM	Idade, GSH, taxa de CD3 e proteína total	336	-	AUC: 0,9757
Tordjman et al. (2020)	LR Binária	Contagem de linfócitos, eosinófilos, basófilos e neutrófilos	400	258	AUC: 0,889
Yang et al. (2020)	LR, RF, GBDT	LDH, PCR e ferritina	5893	1402	AUC: 0,838
Joshi et al. (2020)	LR	Contagem de neutrófilos, contagem de linfócitos, hematócrito e sexo	2777	368	AUC: 0,78
Goodman-Meza et al. (2020)	RF, LR, SVM, MLP, SGD, XGB e ADA-Boost	PCR, LDH e ferritina	1273	182	AUC: 0,91
AlJame et al. (2020)	ERLX (ET, RF, LR e XGB)	Monócitos, plaquetas e glóbulos brancos	5644	559	AUC: 0,9938
Brinati et al. (2020)	RF, TWRF, DT, ET, KNN, LR, NB e SVM	AST, linfócitos, LDH, PCR, glóbulos brancos, eosinófilos, ALT idade e neutrófilos	279	177	AUC: 0,84

Continua

Tabela 1 - Visão geral dos trabalhos relatados

Autor	Algoritmos	Atributos relevantes	Conjunto de Dados	Dados Positivos	Desempenho
Alakus e Turkoglu (2020)	RNA, CNN, LSTM, RNN, CNNLSTM e CNNRNN	-	600	80	AUC: 0,625
Kukar et al. (2021)	RF, XGB e DNN	Concentração de Hemoglobina Corpuscular Média, contagem de eosinófilos, albumina, INR e atividade de protrombina	5333	160	AUC: 0,97
Tschoellitsch et al. (2021)	RF	Contagem de glóbulos brancos, RDW, hemoglobina e cálcio sérico	1528	65	AUC: 0,74
Soltan et al. (2021)	LR, RF e XGB	PCR, eosinófilos, basófilos, necessidade de oxigênio e frequência respiratória	115394	437	AUC: 0,94
Cabitza et al. (2021a)	RF, NB, LR, SVM e KNN	LDH, AST, PCR, cálcio, glóbulos brancos, glóbulos brancos, idade e fibrinogênio	1624	786	AUC: 0,74 - 0,90
Alves et al. (2021)	DTX	Glóbulos brancos, plaquetas, eosinófilos, PCR e monócitos	608	84	AUC: 0,87
Rahman et al. (2022)	QCoVSML	Glóbulos brancos, monócitos, linfócitos e idade	1624	786	AUC: 0,96

Fonte: Elaborado pelo autor.

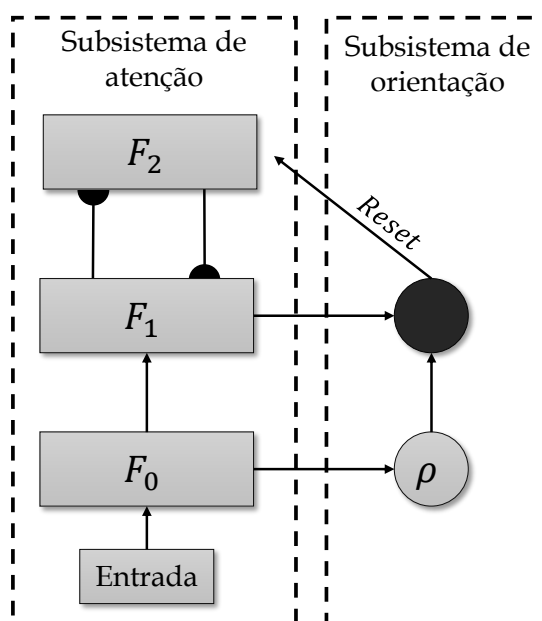
Fim da tabela

3 TEORIA DA RESSONÂNCIA ADAPTATIVA

A Teoria da Ressonância Adaptativa (ART) foi proposta para abordar e superar muitos dos problemas encontrados nas redes neurais mais populares (GROSSBERG, 1987; CARPENTER; GROSSBERG, 1987b). Esta teoria soluciona o dilema da estabilidade-plasticidade (GROSSBERG, 1982), isto é, a capacidade de aprender rapidamente (plasticidade) sem experimentar o esquecimento catastrófico (estabilidade), ou seja, esquecer-se do conhecimento já aprendido (CARPENTER; GROSSBERG, 1987b). A primeira arquitetura utilizando essa teoria foi desenvolvida por Grossberg (1976) e denominada *Grossberg Network* (GN), desde então varias outras foram desenvolvidas com diferentes características (CARPENTER; GROSSBERG, 1987b; CARPENTER; GROSSBERG, 1987a; CARPENTER; GROSSBERG; ROSEN, 1991; CARPENTER; GROSSBERG; REYNOLDS, 1991; CARPENTER et al., 1992).

Grossberg (2020) lista diversas propriedades em que a ART superam os problemas presentes em algoritmos que utilizam o *backpropagation*, inclusive o *Deep Learning*. Dentre tais propriedades está a aprendizagem em tempo real, aprendizagem auto-organizada não supervisionada ou supervisionada, aprendizagem rápida e a capacidade de poder explicar o que a rede aprende. Além disso, a ART concentra a atenção em dados importantes, evitando os *outliers* que levam a um esquecimento catastrófico (GROSSBERG, 2020).

Figura 1 - Arquitetura ART.



Fonte: Adaptado de Carpenter e Grossberg (1987b).

A arquitetura básica de uma rede ART pode ser vista na [Figura 1](#), é composta por dois subsistemas fundamentais, o subsistema de atenção no qual ocorre a aprendizagem e o reconhecimento (camadas F_1 e F_2), e o subsistema de orientação, o qual é responsável por aceitar a categoria obtida no subsistema de atenção ou buscar por uma nova categoria ([GROSSBERG, 2020](#)). De modo geral, os dados de entrada da camada F_0 são replicados para a camada F_1 a qual ativa a busca por uma categoria vencedora na camada F_2 por meio das conexões de baixo para cima, cujo padrão é comparado com a entrada atual em F_1 por meio das conexões de cima para baixo. Caso o nível de correspondência não seja suficiente, há uma busca para encontrar uma nova categoria que possua uma correspondência boa o suficiente entre um padrão de entrada de baixo para cima e uma expectativa de cima para baixo. Quando isso acontece a rede entra no chamado “estado ressonante”, no qual ela aprende o conteúdo atual da camada F_1 , ou cria uma categoria em F_2 com a entrada atual ([CARPENTER et al., 1992](#); [GROSSBERG, 2020](#)). Nas subsecções a seguir são descritas as redes baseadas na Teoria da Ressonância Adaptativa implementadas neste dissertação.

3.1 REDE NEURAL ART FUZZY AUTOEXPANSÍVEL

Desenvolvida por [Moreno \(2010\)](#) a rede ART *Fuzzy* Autoexpansível é baseada na sua antecessora ART *Fuzzy* ([CARPENTER; GROSSBERG; ROSEN, 1991](#)), ambas as redes utilizam o operador *Fuzzy* AND ($\wedge$) da teoria dos conjuntos *Fuzzy* ([ZADEH, 1965](#)) definido pela [Equação 1](#).

$$(\mathbf{x} \wedge \mathbf{y})_i = \min \{x_i, y_i\}. \quad (1)$$

A rede possui um mecanismo de autoexpansão que reduz o tempo de processamento e, consequentemente, o gasto computacional ([MORENO, 2010](#); [MORENO, 2016](#)). O treinamento se inicia com uma única categoria ativa e se expande conforme a necessidade. A verificação de similaridade ocorrerá somente com as categorias ativas, ao contrário da ART *Fuzzy* tradicional, na qual a similaridade é verificada com todas as categorias, seja ela ativa ou inativa. Além disso, ao se inserir um dado incompatível com as categorias existente, a rede inicia uma nova categoria para alocar este novo dado. Vale ressaltar que tais alterações não alteram a estabilidade da rede neural nem comprometem a capacidade de reconhecer novas informações presentes nos dados de treinamento ([MORENO, 2016](#)). O funcionamento da rede e o fluxograma são apresentados na [Figura 2](#).

Algoritmo da Rede Neural ART *Fuzzy* Autoexpansível:

Passo 1: Leitura, normalização e Codificação do padrão de entrada

A [Equação 2](#) representa o vetor contendo os dados de entrada, o vetor é então normalizado

na [Equação 3](#) para impedir a proliferação de categorias ([CARPENTER; GROSSBERG; ROSEN, 1991](#)). O processo de codificação em complemento gera o vetor de entrada I dado pela [Equação 4](#), necessário para preservar a amplitude da informação, deste modo, o novo vetor de entrada será um vetor que terá dimensão $2M$.

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_M], \quad (2)$$

$$\bar{\mathbf{a}} = \frac{\mathbf{a}}{\left| \sum_i^M a_i \right|}, \quad (3)$$

$$\mathbf{I} = [\bar{\mathbf{a}} \ \bar{\mathbf{a}}^c]. \quad (4)$$

Passo 2: Leitura dos parâmetros

Os parâmetros utilizados no processamento da rede ART *Fuzzy* Autoexpansível são:

1. α : Taxa de aprendizagem ($\alpha > 0$), o valor deste parâmetro indica quanto novos padrões vão influenciar a categoria na atualização dos pesos no próximo laço do treinamento;
2. β : Taxa de treinamento ($0 \leq \beta \leq 1$), seu valor baixo indica um treinamento lento, enquanto um valor alto próximo de 1 implica em um treinamento rápido;
3. ρ : Parâmetro de vigilância ($0 \leq \rho \leq 1$), é responsável por determinar o grau de similaridade para que uma nova categoria seja criada.

Passo 3: Inicialização dos Pesos

A matriz de pesos é iniciada contendo apenas o primeiro vetor de entrada indicando haver apenas uma categoria ativa conforme mostra a [Equação 5](#).

$$\mathbf{W}_1 = \mathbf{I}_1. \quad (5)$$

Passo 4: Inicialização do contador de categorias ativas

$$Cluster = 1. \quad (6)$$

Início do Treinamento

Passo 5: Inicialização do contador para criação de uma nova categoria ativa

$$Counter = 1. \quad (7)$$

Passo 6: Cálculo e escolha da categoria vencedora

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{\alpha + |\mathbf{W}_j|}, \quad (8)$$

A categoria escolhida será aquela que possuir a maior similaridade.

$$T_J = \max \{T_j : j = 1, 2, \dots, Cluster\}, \quad (9)$$

Passo 7: Teste de vigilância

$$\frac{|I \wedge W_J|}{|I|} \geq \rho. \quad (10)$$

A ressonância ocorre se o critério de vigilância dado pela [Equação 10](#) for satisfeito, para isso segue-se ao **Passo 10**. Caso contrário, passa-se ao **Passo 8**.

Passo 8: Verificação e Reset

Este passo analisa se todas as categorias já foram verificadas utilizando a [Equação 11](#), caso já tenha verificado todas as categorias, passa-se para o **Passo 9**, caso contrario o contador é atualizado conforme mostra a [Equação 12](#). Além disso, a categoria é eliminada do processo de busca a partir do *Reset* dado pela [Equação 13](#) e volta para o **Passo 6** ([MORENO, 2010](#)).

$$Counter < Cluster, \quad (11)$$

$$Counter^{novo} = Counter^{velho} + 1, \quad (12)$$

$$T_J = 0. \quad (13)$$

Passo 9: Criação de uma nova categoria

Quando o contador é igual ao número de categorias significa que nenhuma categoria existente é similar o suficiente ao dado de entrada para incorporá-lo, sendo assim, é necessária a criação de uma nova categoria para abrigar o padrão. Primeiramente é necessário indicar a criação da nova categoria, por meio da [Equação 14](#), por fim, essa nova categoria incorpora o dado de entrada como mostra a [Equação 15](#). Então retorna-se ao **Passo 5** ([MORENO, 2010](#)).

$$Cluster^{novo} = Cluster^{velho} + 1, \quad (14)$$

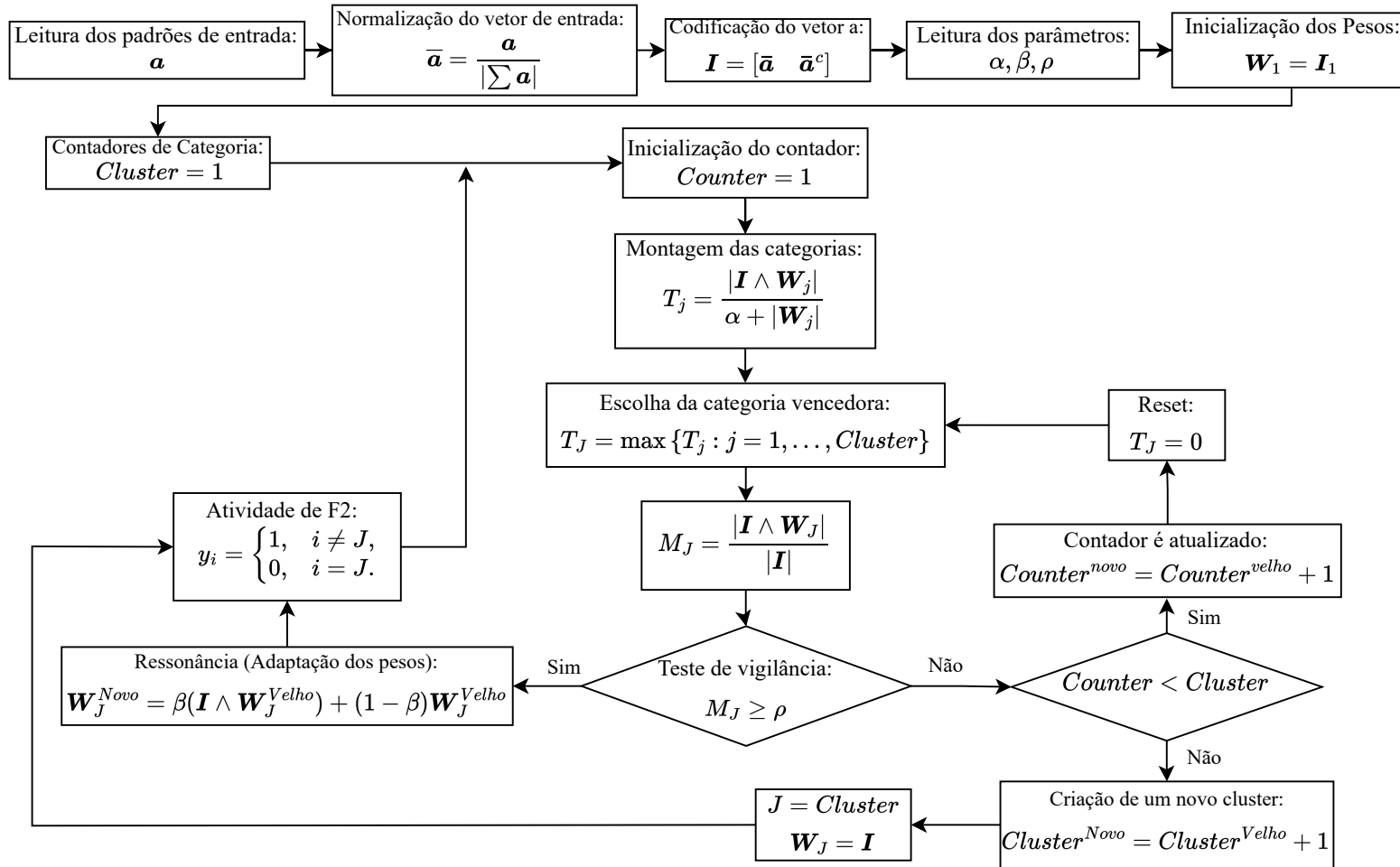
$$W_{Cluster^{novo}} = I. \quad (15)$$

Passo 10: Adaptação dos pesos

Neste passo, os dados da categoria vencedora são atualizados, ocorrendo a adaptação do vetor peso dado pela [Equação 16](#). Então retorna-se ao **Passo 5**.

$$W_J^{novo} = \beta (I \wedge W_J^{velho}) + (1 - \beta) W_J^{velho}. \quad (16)$$

Figura 2 - Fluxograma de treinamento da rede ART Fuzzy Autoexpansível.

Fonte: Adaptado de [Moreno \(2010\)](#).

3.2 REDE NEURAL ART EUCLIDIANA AUTOEXPANSÍVEL

A rede neural ART Euclidiana Autoexpansível (MORENO, 2010) é semelhante à rede apresentada na Seção 3.1. Entretanto, os conectivos lógicos *Fuzzy* são substituídos pela distância Euclidiana, dada pela Equação 17.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (17)$$

Com isso, ao invés de hiper-retângulos, obtém-se hiperesferas cujo centro é a categoria representativa da classe, além disso, ela não necessita da normalização e codificação dos vetores de entrada, reduzindo o custo computacional, visto que a dimensão dos dados inseridos na rede é reduzida pela metade (VUSKOVIC; DU, 2002; KENAYA; CHEOK, 2008). O fluxograma de treinamento pode ser visto na Figura 3.

Algoritmo da Rede Neural ART Euclidiana Autoexpansível:

Passo 1: Leitura do padrão de entrada

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_M]. \quad (18)$$

Passo 2: Leitura dos parâmetros

1. β : Taxa de treinamento ($0 \leq \beta \leq 1$);
2. ρ : Parâmetro de vigilância ($0 \leq \rho \leq 1$).

Passo 3: Inicialização dos Pesos

$$\mathbf{W}_1 = \mathbf{a}_1. \quad (19)$$

Passo 4: Inicialização do contador de categorias ativas

$$Cluster = 1. \quad (20)$$

Início do Treinamento

Passo 5: Inicialização do contador para criação de uma nova categoria ativa

$$Counter = 1. \quad (21)$$

Passo 6: Cálculo e escolha da categoria vencedora

Uma das mudanças no algoritmo está na Equação 22, a qual utiliza a distância Euclidiana

para criar as categorias.

$$T_j = \sqrt{\sum_{i=1}^M (a_i - W_{ji})^2}, \quad (22)$$

Outra mudança está na escolha da categoria, dada pela [Equação 23](#), na rede *Fuzzy* é escolhida a que possui ativação máxima, ao passo que na euclidiana escolhe-se a que possui ativação mínima.

$$T_J = \min \{T_j : j = 1, 2, \dots, Cluster\}, \quad (23)$$

Passo 7: Teste de vigilância

$$\frac{T_J}{\max \left\{ \sum_{i=1}^M (a_i)^2, \sum_{m=1}^M (W_{Jm})^2 \right\}} < \rho. \quad (24)$$

Caso o critério de vigilância dado pela [Equação 24](#) seja satisfeito, segue-se ao **Passo 10**. Caso contrário, passa-se ao **Passo 8**.

Passo 8: Verificação e Reset

Se a [Equação 25](#) for satisfeita, o contador é atualizado conforme mostra a [Equação 26](#), a categoria é então eliminada do processo de busca pela [Equação 27](#) e volta para o **Passo 6**, caso contrario passa-se para o **Passo 9**.

$$Counter < Cluster, \quad (25)$$

$$Counter^{novo} = Counter^{velho} + 1, \quad (26)$$

$$T_J = 10^6. \quad (27)$$

Passo 9: Criação de uma nova categoria

Este passo é idêntico ao descrito na [Seção 3.1](#), a [Equação 28](#) atualiza a quantidade de categorias e na [Equação 29](#), a categoria é incorporada a rede, então retorna-se ao **Passo 5**.

$$Cluster^{novo} = Cluster^{velho} + 1, \quad (28)$$

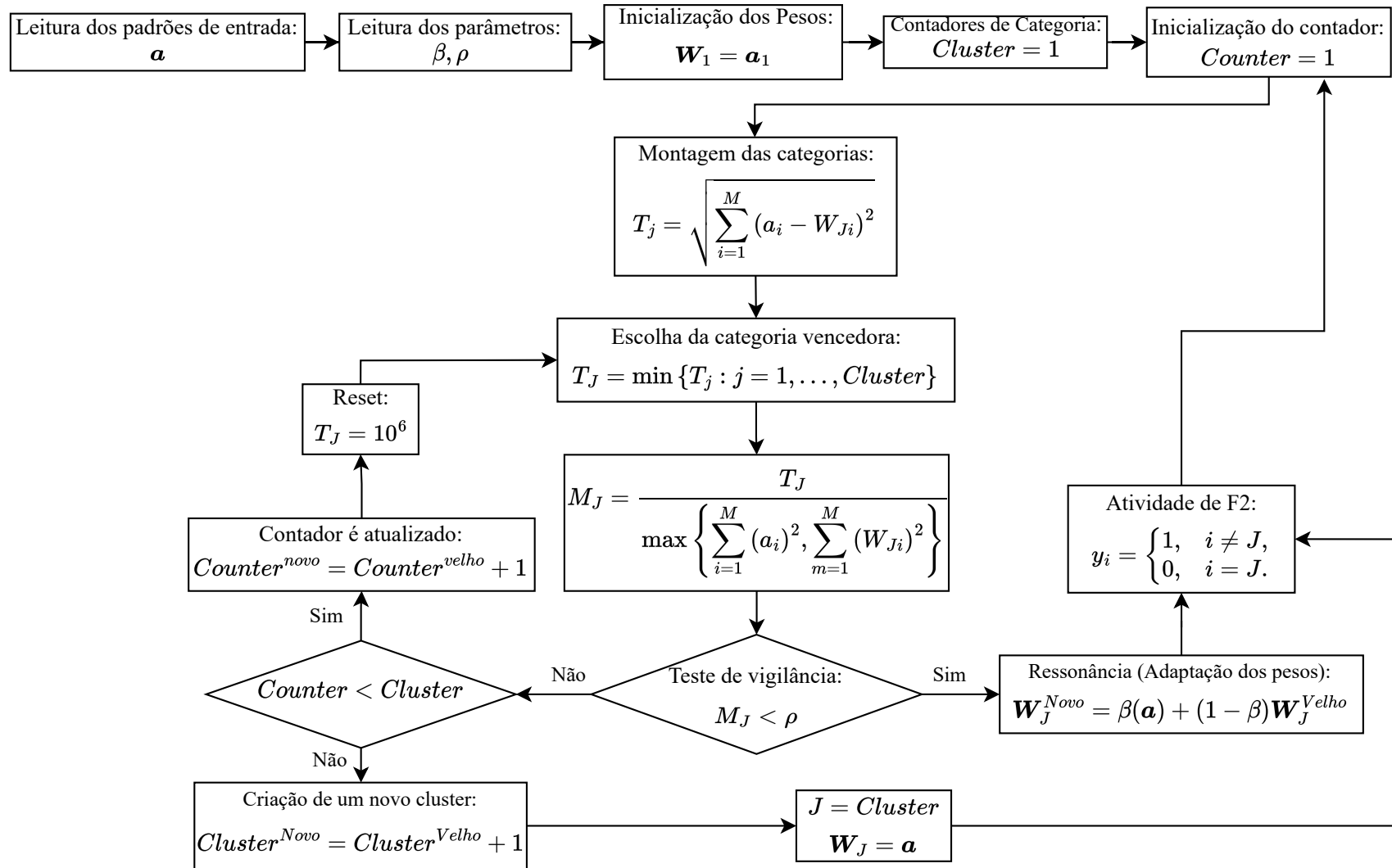
$$W_{Cluster^{novo}} = I. \quad (29)$$

Passo 10: Adaptação dos pesos

Os dados da categoria vencedora são atualizados, ocorrendo a adaptação do vetor peso dado pela [Equação 30](#). Então retorna-se ao **Passo 5**.

$$W_J^{novo} = \beta(a) + (1 - \beta)W_J^{velho}. \quad (30)$$

Figura 3 - Fluxograma de treinamento da rede ART Euclidiana Autoexpansível.

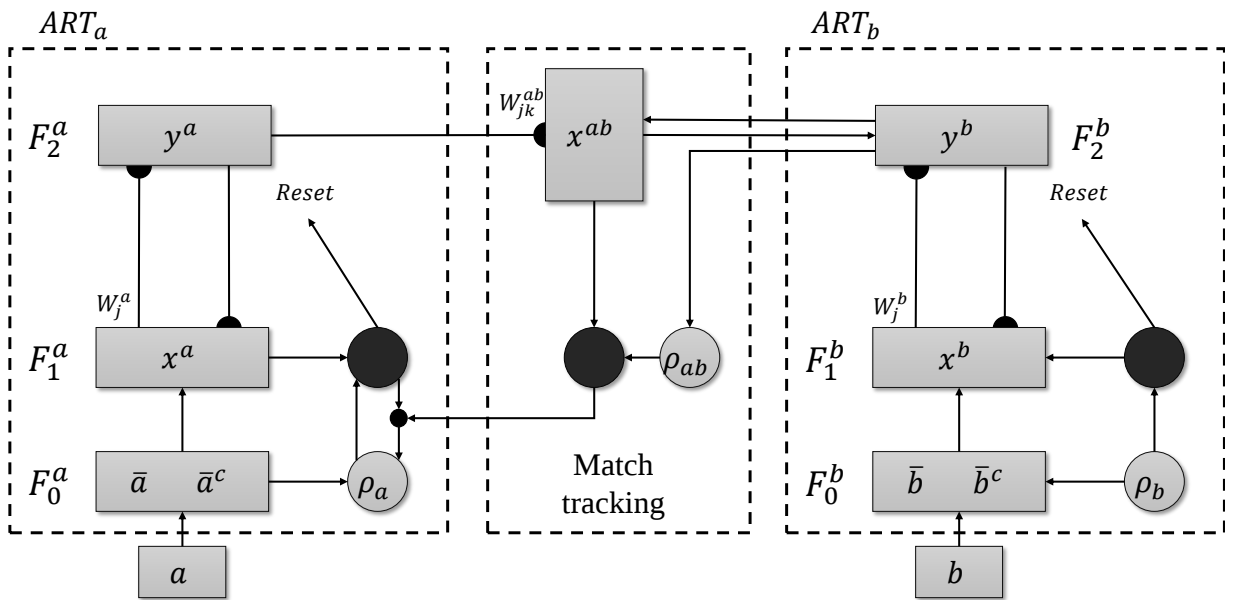


Fonte: Adaptado de Barros (2013).

3.3 REDE NEURAL ARTMAP FUZZY

Possuindo aprendizado supervisionado e incremental, a rede ARTMAP Fuzzy (CARPENTER et al., 1992) contém dois módulos ART conforme mostra a Figura 4, o módulo ART_a responsável por receber os padrões de entrada, enquanto o módulo ART_b recebe a saída desejada, o módulo inter-ART é encarregado de fazer o “casamento” entre o ART_a e ART_b . Além disso, este módulo possui um mecanismo denominado *Match-Traking* que minimiza o erro e maximiza a generalização.

Figura 4 - Arquitetura ARTMAP Fuzzy.



Fonte: Adaptado de Grossberg (2020).

O funcionamento da rede e o fluxograma pode ser visto na Figura 5.

Algoritmo da Rede Neural ARTMAP Fuzzy:

Passo 1: Leitura, normalização e Codificação do padrão de entrada

Os vetores de entrada e saída são definidos pela Equação 31, os vetores são normalizados pela Equação 32 para evitar a proliferação de categorias e a codificação em complemento gera os vetores I^a e I^b dado pela Equação 33.

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_{Ma}] \quad \mathbf{b} = [b_1 \ b_2 \ \dots \ b_{Mb}], \quad (31)$$

$$\bar{\mathbf{a}} = \frac{\mathbf{a}}{\left| \sum_i^{Ma} a_i \right|} \quad \bar{\mathbf{b}} = \frac{\mathbf{b}}{\left| \sum_i^{Mb} b_i \right|}, \quad (32)$$

$$\mathbf{I}_{L \times 2Ma}^a = [\bar{\mathbf{a}} \ \bar{\mathbf{a}}^c] \quad \mathbf{I}_{L \times 2Mb}^b = [\bar{\mathbf{b}} \ \bar{\mathbf{b}}^c]. \quad (33)$$

Passo 2: Leitura dos parâmetros

1. α : Taxa de aprendizagem ($\alpha > 0$);
2. β : Taxa de treinamento ($0 \leq \beta \leq 1$);
3. ρ_a : Parâmetro de vigilância do módulo ART_a ($0 \leq \rho_a \leq 1$);
4. ρ_b : Parâmetro de vigilância do módulo ART_b ($0 \leq \rho_b \leq 1$);
5. ρ_{ab} : Parâmetro de vigilância do módulo ART_{ab} ($0 \leq \rho_{ab} \leq 1$);
6. ε : Incremento do ρ_a .

Passo 3: Inicialização dos Pesos

As matrizes de pesos são iniciadas com o valor 1, indicando que nenhuma categoria está ativa (CARPENTER et al., 1992).

$$\mathbf{W}_{L \times M_a}^a = 1 \quad \mathbf{W}_{L \times M_b}^b = 1 \quad \mathbf{W}_{L \times L}^{ab} = 1. \quad (34)$$

Início do Treinamento**Passo 4: Cálculo e escolha da categoria vencedora nos módulos ART_a e ART_b**

$$T_k^b = \frac{|\mathbf{I}^b \wedge \mathbf{W}_k^b|}{\alpha + |\mathbf{W}_k^b|} \quad T_j^a = \frac{|\mathbf{I}^a \wedge \mathbf{W}_j^a|}{\alpha + |\mathbf{W}_j^a|}, \quad (35)$$

A categoria escolhida será aquela que possuir a maior similaridade (Equação 36) e o vetor atividade é atualizado pela Equação 37.

$$T_K^b = \max \{T_k^b : k = 1, 2, \dots, L\} \quad T_J^a = \max \{T_j^a : j = 1, 2, \dots, L\}, \quad (36)$$

$$Y_k^b = \begin{cases} 0, & k \neq K, \\ 1, & k = K. \end{cases} \quad (37)$$

Passo 5: Teste de vigilância no módulo ART_b

Caso a Equação 38 seja satisfeita, os pesos do módulo ART_b são atualizados por meio da Equação 39 e passa-se ao **Passo 6**, caso contrário ocorre o *reset* e a categoria K é excluída do processo de busca pela Equação 40 e uma nova categoria vencedora é determinada voltando ao **Passo 4**.

$$\frac{|\mathbf{I}^b \wedge \mathbf{W}_K^b|}{|\mathbf{I}^b|} \geq \rho_b. \quad (38)$$

$$\mathbf{W}_K^{b\text{ novo}} = \beta \left(\mathbf{I}^b \wedge \mathbf{W}_K^{b\text{ velho}} \right) + (1 - \beta) \mathbf{W}_K^{b\text{ velho}}, \quad (39)$$

$$T_K^b = 0. \quad (40)$$

Passo 6: Teste de vigilância no módulo ART_a

Se a [Equação 41](#) for satisfeita passa-se para o **Passo 7**, caso contrário ocorre o *reset* e a categoria J é excluída do processo de busca pela [Equação 42](#) e uma nova categoria vencedora é determinada voltando para o **Passo 4**.

$$\frac{|I^a \wedge W_J^a|}{|I^a|} \geq \rho_a. \quad (41)$$

$$T_J^a = 0. \quad (42)$$

Passo 7: Verificação no módulo Inter-ART

A [Equação 43](#) verifica se a categoria ativa em ART_a corresponde à saída desejada apresentada em ART_b, caso ela seja satisfeita, passa-se para o **Passo 8**. Caso contrário, a categoria de ART_a é descartada e o parâmetro ρ_a é incrementado temporariamente pela [Equação 44](#) até que encontre uma categoria que satisfaça a [Equação 43](#) (CARPENTER et al., 1992).

$$\frac{|Y^b \wedge W_J^{ab}|}{|Y^b|} \geq \rho_{ab}, \quad (43)$$

$$\rho_a = \frac{|I \wedge W_J^a|}{|I|} + \varepsilon. \quad (44)$$

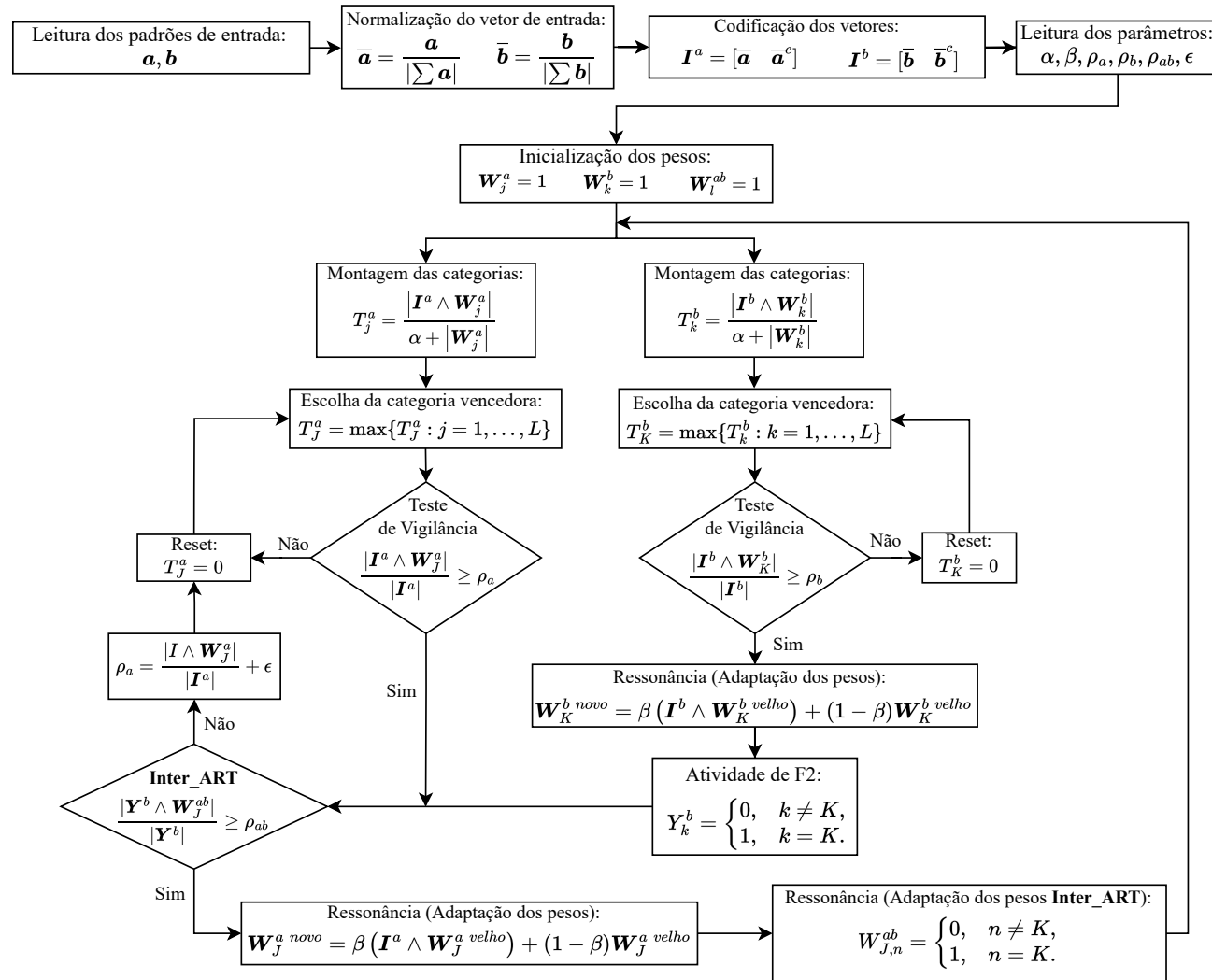
Passo 8: Atualização dos pesos no módulo ART_a e Inter-ART

Os pesos no módulo ART_a e Inter-ART são atualizados, respectivamente, pelas [Equações 45 e 46](#).

$$W_J^{a\text{ novo}} = \beta (I^a \wedge W_J^{a\text{ velho}}) + (1 - \beta) W_J^{a\text{ velho}}, \quad (45)$$

$$W_{J,n}^{ab} = \begin{cases} 0, & n \neq K, \\ 1, & n = K. \end{cases} \quad (46)$$

Figura 5 - Fluxograma de treinamento da rede ARTMAP Fuzzy.

Fonte: Adaptado de [Moreno \(2010\)](#).

3.4 REDE NEURAL ART FUZZY AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO E ART EUCLIDIANA AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO

Moreno (2010) desenvolveu a ART Euclidiana Autoexpansível com Treinamento Contínuo, ela é baseada na rede descrita na Seção 3.2 e conta com a adição do módulo de TC que possibilita a aprendizagem de modo *on-line*, isto significa que a previsão e o treinamento ocorrem simultaneamente, além disso, não há necessidade de reinicializar o treinamento caso novos dados estejam disponíveis, tornando-a mais rápida e eficiente (MORENO, 2010). Vale ressaltar que a inclusão deste módulo à rede só é possível devido à característica de estabilidade e plasticidade que ela possui (MORENO, 2010).

Duas funções de escolha são necessárias para evitar que os dados que aparecem com pouca frequência não sejam incorporados à memória da rede, uma para categorias temporárias (T^T) e outra para categorias definitivas (T^D), a adição do parâmetro permanência ($NMIN$) estabelece a quantidade de padrões necessários para que as categorias temporárias se tornem definitivas, o índice de novidade (η) por outro lado, determina se os pesos da categoria definitiva devem ou não serem atualizados, desta forma, informações redundantes são descartadas (MORENO, 2010), evitando a sobrecarga da memória da rede.

O algoritmo e o fluxograma da rede ART Euclidiana Autoexpansível com Treinamento Contínuo estão disponíveis em Moreno (2010). Já o funcionamento da rede ART Fuzzy Autoexpansível com Treinamento Contínuo bem como seu fluxograma pode ser visto na Figura 6.

Algoritmo da Rede Neural ART Fuzzy Autoexpansível com Treinamento Contínuo:

Passo 1: Leitura, normalização e Codificação do padrão de entrada

$$\mathbf{a} = [a_1 \ a_2 \ \dots \ a_M], \quad (47)$$

$$\bar{\mathbf{a}} = \frac{\mathbf{a}}{\left| \sum_i^M a_i \right|}, \quad (48)$$

$$\mathbf{I} = [\bar{\mathbf{a}} \ \bar{\mathbf{a}}^c]. \quad (49)$$

Passo 2: Leitura dos parâmetros

Os parâmetros utilizados no processamento da rede ART Fuzzy Autoexpansível são:

1. α : Taxa de aprendizagem ($\alpha > 0$);
2. β : Taxa de treinamento ($0 \leq \beta \leq 1$);
3. ρ : Parâmetro de vigilância ($0 \leq \rho \leq 1$);

4. $NMIN$: Parâmetro Permanência ($NMIN > 1$);
5. η : Índice de novidade ($\eta > 0$).

Passo 3: Inicialização dos Pesos

A matriz de pesos temporários é iniciada com apenas uma categoria ativa (MORENO, 2010), como pode ser visto na Equação 50.

$$\mathbf{W}_1^T = \mathbf{I}_1. \quad (50)$$

Passo 4: Inicialização do contador de categorias temporárias similares

$$ConT_1 = 1. \quad (51)$$

Passo 5: Inicialização do contador de categorias definitivas e temporárias ativas

$$Cluster_{Temp} = 1 \quad Cluster_{Def} = 0 \quad (52)$$

Início do Treinamento

Passo 6: Inicialização do contador para criação de uma nova categoria ativa

$$Counter = 1. \quad (53)$$

Passo 7: Cálculo da categoria

O cálculo e a escolha da categoria vencedora é feito para ambas as categorias temporárias e definitivas (MORENO, 2010).

$$T_j^T = \frac{|\mathbf{I} \wedge \mathbf{W}_j^T|}{\alpha + |\mathbf{W}_j^T|} \quad T_j^D = \frac{|\mathbf{I} \wedge \mathbf{W}_j^D|}{\alpha + |\mathbf{W}_j^D|}, \quad (54)$$

Passo 8: Escolha da categoria vencedora

A categoria escolhida será aquela que possuir a maior similaridade dentre as categorias definitivas e temporárias (MORENO, 2010; MORENO, 2016).

$$T_{JT}^T = \max \{T_j^T : j = 1, 2, \dots, Cluster_{Temp}\}, \quad (55)$$

$$T_{JD}^D = \max \{T_j^D : j = 1, 2, \dots, Cluster_{Def}\}, \quad (56)$$

$$T = \max \{T_{JT}^T, T_{JD}^D\}. \quad (57)$$

Passo 9: Teste de vigilância caso a categoria vencedora seja temporária ou definitiva

$$\frac{|\mathbf{I} \wedge \mathbf{W}_{JT}^T|}{|\mathbf{I}|} \geq \rho \quad \text{ou} \quad \frac{|\mathbf{I} \wedge \mathbf{W}_{JD}^D|}{|\mathbf{I}|} \geq \rho. \quad (58)$$

Se o critério de vigilância dado pela [Equação 58](#) for satisfeito passa-se ao **Passo 11**. Caso contrário, é necessário verificar se todas as categorias temporárias e definitivas foram verificadas utilizando a [Equação 59](#), se ela for satisfeita passa-se para o passo **Passo 10**, caso contrário o contador é atualizado pela [Equação 60](#) e ocorre o *reset* para a categoria temporária ou definitiva conforme a [Equação 61](#) e retorna-se ao **Passo 8** (MORENO, 2010).

$$Counter \geq (Cluster_{Temp} + Cluster_{Def}), \quad (59)$$

$$Counter^{novo} = Counter^{velho} + 1, \quad (60)$$

$$T_{JT}^T = 0 \quad \text{ou} \quad T_{JD}^D = 0. \quad (61)$$

Passo 10: Criação de uma nova categoria temporária

Quando o contador é igual a soma de categorias temporárias e definitivas, significa que nenhuma categoria existente é similar o suficiente ao dado de entrada para incorporá-lo, então cria-se uma categoria temporária para abrigar o padrão, isto é feito pelas Equações 62 e 63, então retorna-se ao **Passo 6** (MORENO, 2010).

$$Cluster_{Temp}^{novo} = Cluster_{Temp}^{velho} + 1, \quad (62)$$

$$\mathbf{W}_{Cluster_{Temp}} = \mathbf{I}. \quad (63)$$

Passo 11: Verificação das categorias

Se a categoria vencedora seja definitiva passa-se ao **Passo 12**, caso seja temporária ocorre a adaptação do vetor peso temporário dado pela [Equação 64](#). Além disso, é necessário registrar que um novo padrão foi incorporado nesta categoria por meio da [Equação 65](#) (MORENO, 2010).

$$\mathbf{W}_{JT}^{T(novo)} = \beta \left(\mathbf{I} \wedge \mathbf{W}_{JT}^{T(velho)} \right) + (1 - \beta) \mathbf{W}_{JT}^{T(velho)}, \quad (64)$$

$$ConT_{JD} = ConT_{JD} + 1 \quad (65)$$

Em seguida, é necessário verificar se este contador atingiu o valor determinado pelo parâmetro *NMIN*, caso não tenha atingido volta ao **Passo 6**, caso contrário esta categoria torna-se definitiva, isto é feito pelas Equações 66–68, então retorna-se ao **Passo 6** (MORENO, 2010).

$$Cluster_{Def}^{novo} = Cluster_{Def}^{velho} + 1 \quad \text{e} \quad \mathbf{W}_{Cluster_{Def}}^D = \mathbf{W}_{JT}^T, \quad (66)$$

$$\mathbf{W}_l^T = \mathbf{W}_{l+1}^T \quad \text{e} \quad ConT_l = ConT_{l+1} \quad l = JT, \dots, Cluster_{Temp} \quad (67)$$

$$Cluster_{Temp}^{novo} = Cluster_{Temp}^{velho} - 1, \quad (68)$$

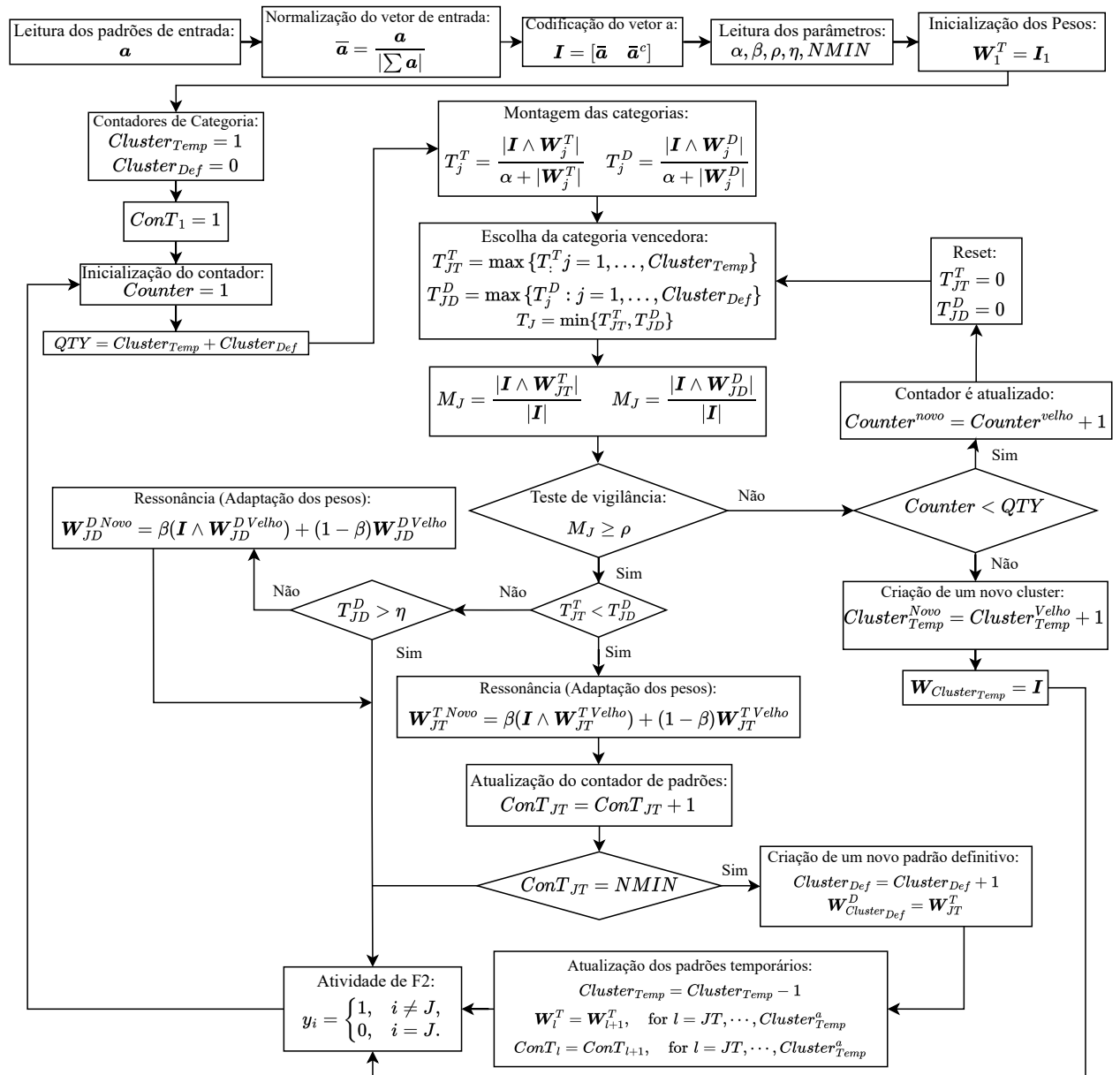
Passo 12: Detecção de Anomalia

Este passo verifica se a categoria definitiva vencedora é uma anomalia, se a [Equação 69](#) for satisfeita não haverá a necessidade da atualização de pesos, caso contrário a matriz de peso é atualizada por meio da [Equação 70](#) (MORENO, 2010).

$$T > \eta \quad (69)$$

$$\mathbf{W}_{JD}^{D(novo)} = \beta \left(\mathbf{I} \wedge \mathbf{W}_{JD}^{D(velho)} \right) + (1 - \beta) \mathbf{W}_{JD}^{D(velho)}, \quad (70)$$

Figura 6 - Fluxograma de treinamento da rede ART *Fuzzy* Autoexpansível com Treinamento Continuo.

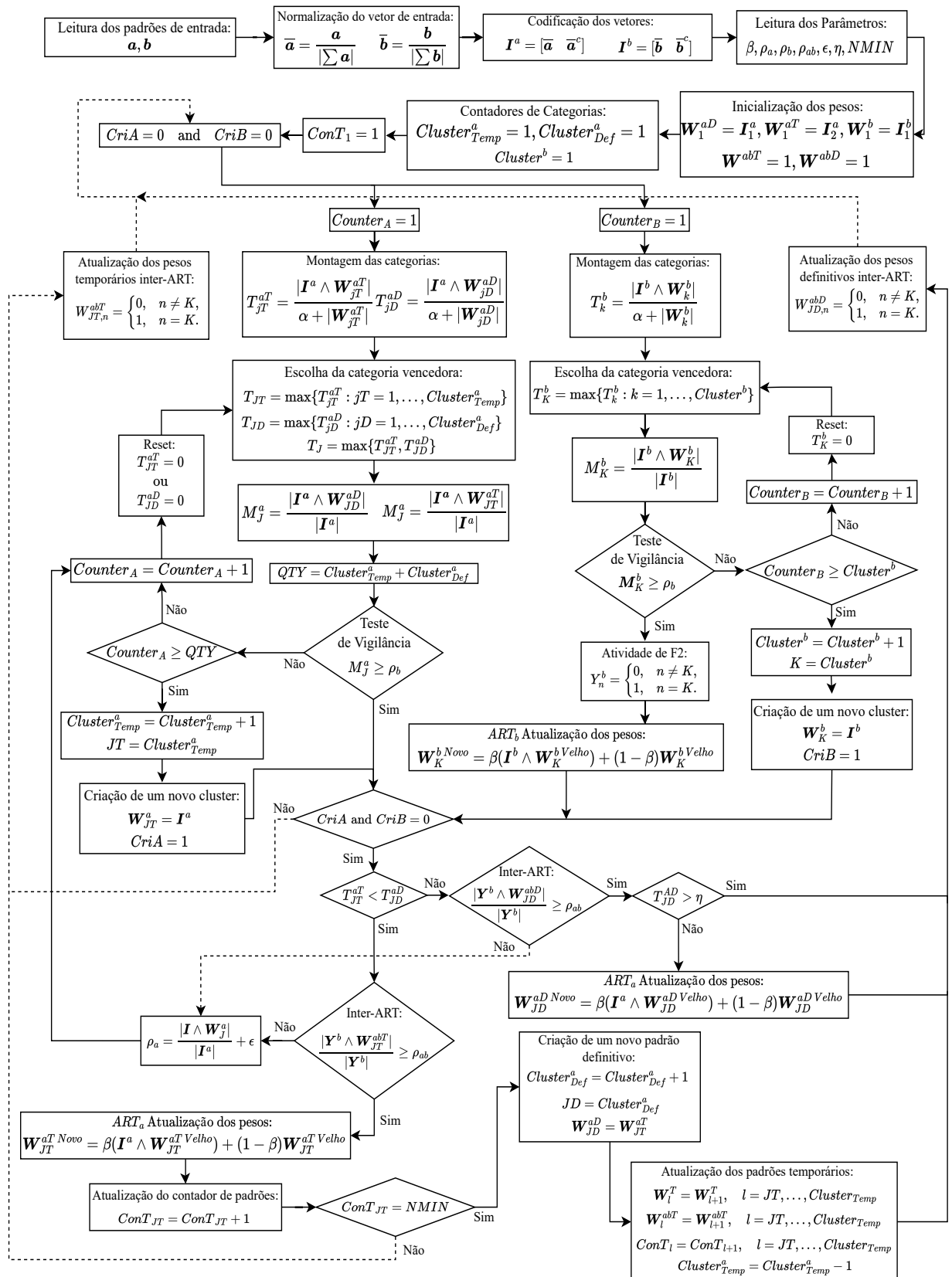


Fonte: Elaborado pelo autor.

3.5 REDE NEURAL ARTMAP FUZZY AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO E ARTMAP EUCLIDIANA AUTOEXPANSÍVEL COM TREINAMENTO CONTINUADO

As redes ARTMAP Euclidiana Autoexpansível com Treinamento Continuoado ([MARCHIORI et al., 2011](#); [BARROS, 2013](#)) e ARTMAP *Fuzzy* Autoexpansível com Treinamento Continuoado combinam as características relatadas anteriormente, tais como ser autoexpansível ([Seção 3.2](#)), possuir aprendizagem supervisionada ([Seção 3.3](#)) e possibilitar o treinamento *on-line* ([Seção 3.4](#)). O algoritmo e o fluxograma da rede baseada na distância euclidiana podem ser encontrados em [Barros \(2013\)](#). Por outro lado, o fluxograma da rede baseada na lógica *Fuzzy* pode ser visto na [Figura 7](#).

Figura 7 - Fluxograma de treinamento da rede ARTMAP *Fuzzy* Autoexpansível com Treinamento Contínuo.



Fonte: Elaborado pelo autor.

4 METODOLOGIA

Neste capítulo serão apresentados em detalhes todos os conjuntos de dados utilizados, os atributos empregados na primeira e na segunda fase das simulações, as métricas utilizadas para avaliar o desempenho das redes, além da validação realizada.

4.1 CONJUNTOS DE DADOS

Um total de 10 conjuntos de dados de acesso aberto foram utilizados, os dados são referentes a biomarcadores clínicos obtidos de diferentes hospitais no Brasil, Espanha, Itália e Etiópia. Uma visão geral dos dados utilizados podem ser observados na [Tabela 2](#).

Tabela 2 - Quantidade de dados e atributos.

Disponível em:	Nome	País	Coleta	Positivos	Negativos
(BRINATI et al., 2020)	Dados ₁	Itália	02/2020-03/2020	177	102
(CABITZA et al., 2021a)	Dados ₂	Itália	02/2020-05/2020	786	838
(ALAKUS; TURKOGLU, 2020)	Dados ₃	Brasil	Início de 2020	80	520
Kaggle ¹	Dados ₄	Brasil	Início de 2020	558	5086
(CABITZA et al., 2021b)	Dados ₅	Itália	03/2020-04/2020	163	174
(CABITZA et al., 2021b)	Dados ₆	Itália	03/2020-04/2020	104	145
(CABITZA et al., 2021b)	Dados ₇	Itália	11/20	118	106
(CABITZA et al., 2021b)	Dados ₈	Brasil	02/2020-06/2020	334	11
(CABITZA et al., 2021b)	Dados ₉	Brasil	02/2020-06/2020	352	949
(CABITZA et al., 2021b)	Dados ₁₀	Brasil	02/2020-06/2020	375	1960
(CABITZA et al., 2021b)	Dados ₁₁	Etiópia	01/2021-03/2021	200	0
(CABITZA et al., 2021b)	Dados ₁₂	Espanha	10/2020	78	42

Fonte: Elaborado pelo autor.

O Dados₁ e Dados₇ foram coletados de pacientes do Hospital IRCCS San Raffaele na Itália, o Dados₂ contém informações de pacientes coletados na admissão nos departamentos de emergência de dois hospitais, o Hospital IRCCS San Raaele e IRCCS Instituto Ortopedico Galeazzi. O Dados₃, Dados₄ e Dados₁₀ foram obtidos do Hospital Israelita Albert Einstein no Brasil, sendo que o Dados₃ é oriundo do conjunto Dados₄, o Dados₅ foi coletado no Hospital Papa Giovanni XXIII e o Dados₆ foi coletado do Hospital Desio, ambos na Itália. O Dados₈ foi coletado no Hospital Sírío-Libanês e Dados₉ nas clínicas particulares Fleury, o Dados₁₁ no Laboratório Nacional de Referência para Química Clínica na Etiópia, por fim, o Dados₁₂ foi

¹O conjunto de dados está disponível em: <https://www.kaggle.com/einsteindata4u/covid19>

coletado no Hospital Universitário Santa Lucía na Espanha. Vale ressaltar que não há instâncias duplicadas para os conjuntos de dados coletados em um mesmo hospital, todos os dados são exclusivos, exceto o Dados₃ que foi obtido por [Alakus e Turkoglu \(2020\)](#) após tratar os dados do conjunto original, ou seja, o Dados₄.

As simulações foram realizadas em duas etapas, a primeira consistiu em avaliar qual rede baseada na Teoria da Ressonância Adaptativa teria o melhor desempenho na predição de pacientes afetados pela COVID-19, enquanto a segunda se baseou em utilizar esta rede como modelo para classificar os pacientes de diferentes Países. Os conjuntos de dados e os atributos utilizados em cada etapa são descritos na [Seção 4.2](#) e [Seção 4.3](#).

4.2 PRIMEIRA ETAPA

Os conjuntos de Dados₁, Dados₂, Dados₃ e Dados₄ foram selecionados para esta etapa por possuírem uma boa diversidade de atributos e por terem sido submetidos a simulações com uma grande variedade de modelos, inclusive modelos de aprendizado profundo. As simulações foram realizadas com os mesmos atributos utilizados por [Brinati et al. \(2020\)](#) no Dados₁, [Cabitza et al. \(2021a\)](#) no Dados₂ e [Alakus e Turkoglu \(2020\)](#) no Dados₃. No conjunto de Dados₄ foi necessário realizar um pré-processamento, devido a grande quantidade de dados ausentes, além de possuir mais de 100 atributos. Vale ressaltar que no conjunto de Dados₄ todas os atributos foram disponibilizados já normalizadas, ou seja, possuem média igual a zero e desvio padrão igual a um. Os atributos presentes em cada conjunto de dados utilizados nesta etapa podem ser vistos na [Tabela 3](#).

Tabela 3 - Atributos contidos em cada conjunto de dados.

Dados	Atributos
Dados ₁	Alanina Aminotransferase, Aspartato Aminotransferase, Basófilos, Classe, Eosinófilos, Gamma GT, Gênero, Idade, Lactato Desidrogenase, Leucócitos, Linfócitos, Monócitos, Neutrófilos, Proteína C-reativa, Plaquetas
Dados ₂	Alanina Aminotransferase, Aspartato Aminotransferase, Basófilos ($10^9 \ell^{-1}$), Basófilos (%), Cálcio, Concentração da Hemoglobina Corpuscular Média, Creatinase, Classe, Creatinina, Eosinófilos ($10^9 \ell^{-1}$), Eosinófilos (%), Fosfatase alcalina, Gamma GT, Gênero, Glicose, Hemoglobina Corpuscular Média, Hemácias, Hematócrito, Hemoglobina, Idade, Lactato Desidrogenase, Leucócitos, Linfócitos ($10^9 \ell^{-1}$), Linfócitos (%), Monócitos ($10^9 \ell^{-1}$), Monócitos (%), Neutrófilos, Neutrófilos (%), Proteína C-reativa, Plaquetas, Potássio, Suspeito COVID-19, Ureia, Volume Corpuscular Médio, Volume Plaquetário médio

Continua

Tabela 3 - Atributos contidos em cada conjunto de dados.

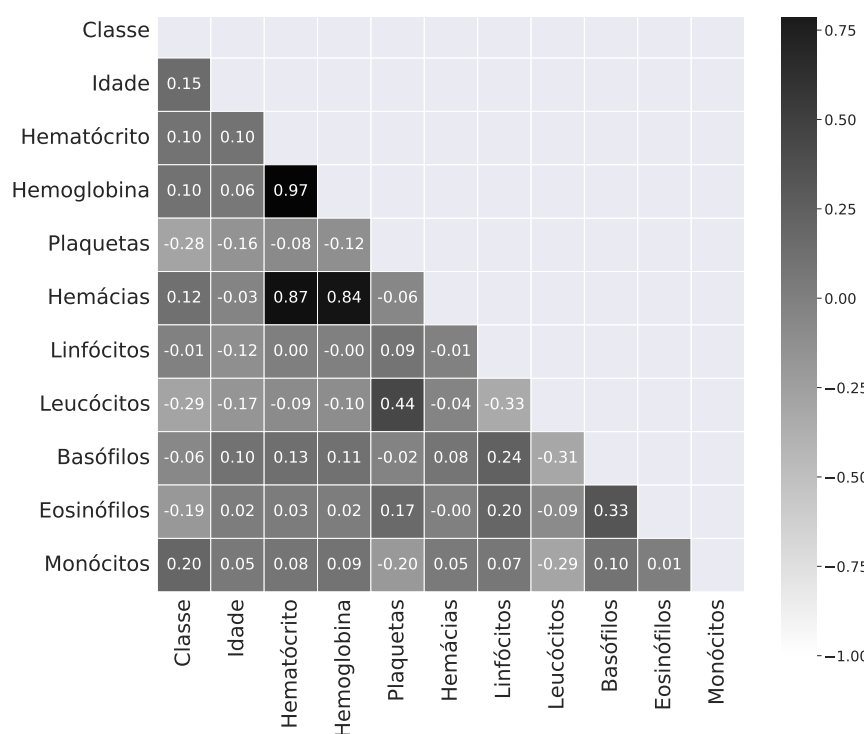
Dados	Atributos
Dados ₃	Alanina Aminotransferase, Aspartato Aminotransferase, Basófilos, Classe, Creatinina, Eosinófilos, glicose Sérica, Hemácias, Hematócrito, Hemoglobina, Idade, Leucócitos, Linfócitos, Monócitos, Neutrófilos, Proteína C-reativa, Plaquetas, Potássio, Sódio, Ureia
Dados _{4T}	Idade, Hematócritos, Hemoglobina, Plaquetas, Hemácias, Linfócitos, Leucócitos, Basófilos, Eosinófilos, Monócitos

Fonte: Elaborado pelo autor.

Fim da tabela

Os atributos do Dados₄ foram selecionados considerando aqueles que possuíam uma quantidade maior de amostras e que fossem relevantes para detectar a COVID-19 (KHARTABIL et al., 2020), os atributos escolhidos foram Idade, Hematócrito, Hemoglobina, Plaquetas, Hemácias, Linfócitos, Leucócitos, Basófilos, Eosinófilos e Monócitos, tais atributos já mostraram boa capacidade preditiva em outros estudos como pode ser visto na Tabela 1. Todas as instâncias que não tivessem informações destes atributos foram excluídas, resultando num conjunto de dados com 601 instâncias. A técnica SMOTE (*Synthetic Minority Oversampling Technique*) que gera dados sintéticos não idênticos foi utilizada para balancear os dados, deixando ambas as classes COVID e não COVID com a mesma quantidade de amostras. Ao final deste processo foi obtido um conjunto de dados com 1000 instâncias e 10 atributos, o conjunto resultante deste pré-processamento será denominado por Dados_{4T}.

Figura 8 - Mapa de calor da matriz de correlação.



Fonte: Elaborado pelo autor.

Analisando a [Figura 8](#) pode-se notar que alguns atributos como hematócrito, hemoglobina e hemácias estão fortemente correlacionados positivamente, ou seja, se um atributo aumenta os outros tendem a aumentar. Sendo assim, as simulações para este conjunto de dados foram realizadas considerando todos os atributos e outra excluindo o hematócrito e hemoglobina. O modelo com o melhor desempenho nesta etapa foi utilizado na etapa seguinte.

4.3 SEGUNDA ETAPA

Para esta etapa os Dados₂, Dados₅, Dados₆, Dados₇, Dados₈, Dados₉, Dados₁₀ e Dados₁₁ foram combinados para formar um novo conjunto de dados com características hematológicas de diferentes países, de modo que fornecesse informações mais gerais para previsão da COVID-19, uma vez que parâmetros hematológicos são afetados por fatores individuais, como idade, sexo e estilo de vida, além de fatores populacionais e ecológicos, como origem étnica, clima, exposição a patógenos e altitude ([ODHIAMBO et al., 2015](#); [ROSENFELD et al., 2019](#); [LUGADA et al., 2004](#)). Desta forma, é importante que a rede seja treinada com informações de múltiplas regiões a fim de assegurar que a rede, seja generalizável e possa reconhecer padrões em qualquer região.

Tabela 4 - Média e desvio padrão dos atributos.

Atributos	Dados ₂	Dados ₅	Dados ₆	Dados ₇	Dados ₈	Dados ₉	Dados ₁₀	Dados ₁₁
Idade	60,55	54,38	66,35	60,53	54,40	47,01	42,87	59,73
	± 19,6	± 25,0	± 18,5	± 20,2	± 15,5	± 17,4	± 17,0	± 13,1
BA (%)	0,34	0,46	0,18	0,32	0,30	0,52	0,48	0,29
	± 0,3	± 0,4	± 0,4	± 0,3	± 0,3	± 0,3	± 0,3	± 0,4
BAT	0,02	0,03	0,02	0,02	0,02	0,03	0,03	0,04
	± 0,04	± 0,04	± 0,05	± 0,04	± 0,03	± 0,02	± 0,02	± 0,1
EO (%)	0,83	1,23	0,74	0,60	1,05	2,30	2,27	0,62
	± 1,6	± 2,1	± 1,6	± 1,2	± 1,6	± 2,4	± 2,5	± 1,1
EOT	0,06	0,09	0,06	0,05	0,06	0,15	0,17	0,08
	± 0,1	± 0,2	± 0,1	± 0,1	± 0,1	± 0,2	± 0,2	± 0,3
HCT	39,41	37,77	38,20	39,67	40,75	41,47	41,18	39,96
	± 5,6	± 7,4	± 6,3	± 6,1	± 5,0	± 4,0	± 4,1	± 7,9
HGB	13,21	12,86	13,21	13,11	13,74	13,82	14,11	13,36
	± 2,0	± 2,7	± 2,3	± 2,2	± 1,9	± 1,5	± 1,5	± 3,2
LY (%)	18,54	21,90	16,56	18,30	23,14	31,10	27,57	7,58
	± 11,1	± 14,5	± 11,6	± 11,9	± 11,4	± 11,2	± 11,0	± 7,1
LYT	1,37	1,84	1,63	1,82	1,31	2,01	2,00	0,77
	± 1,0	± 4,9	± 6,3	± 6,1	± 1,0	± 1,9	± 0,9	± 0,6

Continua

Tabela 4 - Atributos contidos em cada conjunto de dados.

Atributos	Dados ₂	Dados ₅	Dados ₆	Dados ₇	Dados ₈	Dados ₉	Dados ₁₀	Dados ₁₁
MCH	29,20 ± 2,7	30,41 ± 2,9	29,62 ± 3,1	29,51 ± 2,4	29,60 ± 1,9	29,37 ± 1,9	29,59 ± 1,8	29,58 ± 2,8
MCHC	33,47 ± 1,4	33,98 ± 1,4	34,49 ± 1,5	33,00 ± 1,3	33,68 ± 1,1	33,31 ± 1,1	34,25 ± 1,0	33,09 ± 2,0
MCV	87,22 ± 7,1	89,44 ± 7,4	85,72 ± 8,0	89,41 ± 6,5	87,90 ± 5,1	88,19 ± 5,1	86,39 ± 4,7	88,69 ± 8,3
MO (%)	7,76 ± 3,9	8,86 ± 4,7	7,17 ± 3,9	8,13 ± 5,0	9,64 ± 4,4	9,29 ± 3,7	8,64 ± 3,2	5,74 ± 8,2
MOT	0,62 ± 0,5	0,73 ± 0,9	0,64 ± 0,4	0,64 ± 0,3	0,56 ± 0,3	0,59 ± 0,3	0,63 ± 0,3	0,63 ± 0,8
NE (%)	72,50 ± 13,3	67,54 ± 17,2	75,03 ± 14,2	72,48 ± 14,5	65,78 ± 14,1	56,79 ± 12,4	61,02 ± 12,4	85,77 ± 12,7
NET	6,47 ± 4,5	5,62 ± 4,3	7,47 ± 4,9	6,76 ± 4,3	4,35 ± 3,1	3,92 ± 2,1	4,82 ± 2,4	11,60 ± 8,5
PLT	234,49 ± 94,9	204,00 ± 113,7	220,23 ± 90,0	218,00 ± 84,8	199,81 ± 73,4	246,96 ± 70,3	239,92 ± 64,6	268,69 ± 146,8
RBC	4,55 ± 0,7	4,25 ± 0,9	4,49 ± 0,8	4,46 ± 0,8	4,65 ± 0,7	4,72 ± 0,5	4,78 ± 0,5	4,46 ± 0,9
WBC	8,69 ± 4,7	8,31 ± 7,1	9,81 ± 8,0	9,53 ± 7,5	6,30 ± 3,5	6,70 ± 3,0	7,66 ± 2,7	14,18 ± 17,8

BA: Basófilos; BAT: Contagem de Basófilos; EO: Eosinófilos; EOT: Contagem de Eosinófilos; HCT: Hematócrito; HGB: Hemoglobina; LY: Linfócitos; LYT: Contagem de Linfócitos; MCH: *Mean Corpuscular Hemoglobin*; MCHC: *Mean Corpuscular Hemoglobin Concentration*; MCV: *Mean Corpuscular Volume*; MO: Monócitos; MOT: Contagem de Monócitos; NE: Neutrófilos; NET: Contagem de Neutrófilos; PLT: Plaquetas; RBC: Glóbulos Vermelhos

Fonte: Elaborado pelo autor.

Fim da tabela

Na [Tabela 4](#) são apresentadas a média e desvio padrão dos atributos comuns entre todos os conjuntos dados. Os atributos selecionados para as simulações foram obtidos utilizando Análise de Variância (ANOVA), Fator de Inflação da Variância (VIF) e valores SHAP (*SHapley Additive exPlanations*) ([LUNDBERG; LEE, 2017](#)). Além disso, os atributos foram analisados levando-se em consideração os recursos apontados como relevantes pela literatura na [Tabela 1](#).

- ANOVA é um método estatístico que verifica se existem diferenças significativas entre as médias dos grupos e é calculada pela [Equação 71](#) ([KAUFMANN; SCHERING, 2014](#); [JOHNSON; SYNOVEC, 2002](#)).

$$F = \frac{\frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{X})^2}{k-1}}{\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n-k}} \quad (71)$$

Em que:

- k : número de classes;
- n_j : número de elementos da classe j ;
- $\bar{x}_j$: média da classe j ;
- $\bar{X}$: média global;
- x_{ij} : i -ésima observação na classe j ;
- n : soma do número de elementos de todas as classes.

A significância foi determinada em $p < 0,5$, ou seja, se $p < 0,5$ então rejeita-se H_0 e aceita-se H_1 , em que:

H_0 : Não há diferença significativa entre o atributo e os diferentes grupos.

H_1 : Existe uma diferença significativa entre o atributo e os diferentes grupos.

- O VIF é uma medida estatística usada para avaliar a multicolinearidade entre variáveis e pode ser obtido a partir dos elementos da diagonal principal da matriz V da [Equação 73](#) (MONTGOMERY; PECK; VINING, 2006).

$$\mathbf{R}_{j \times j} = \left(\frac{1}{i-1} \cdot \mathbf{W}'_{j \times i} \cdot \mathbf{W}_{i \times j} \right) \quad (72)$$

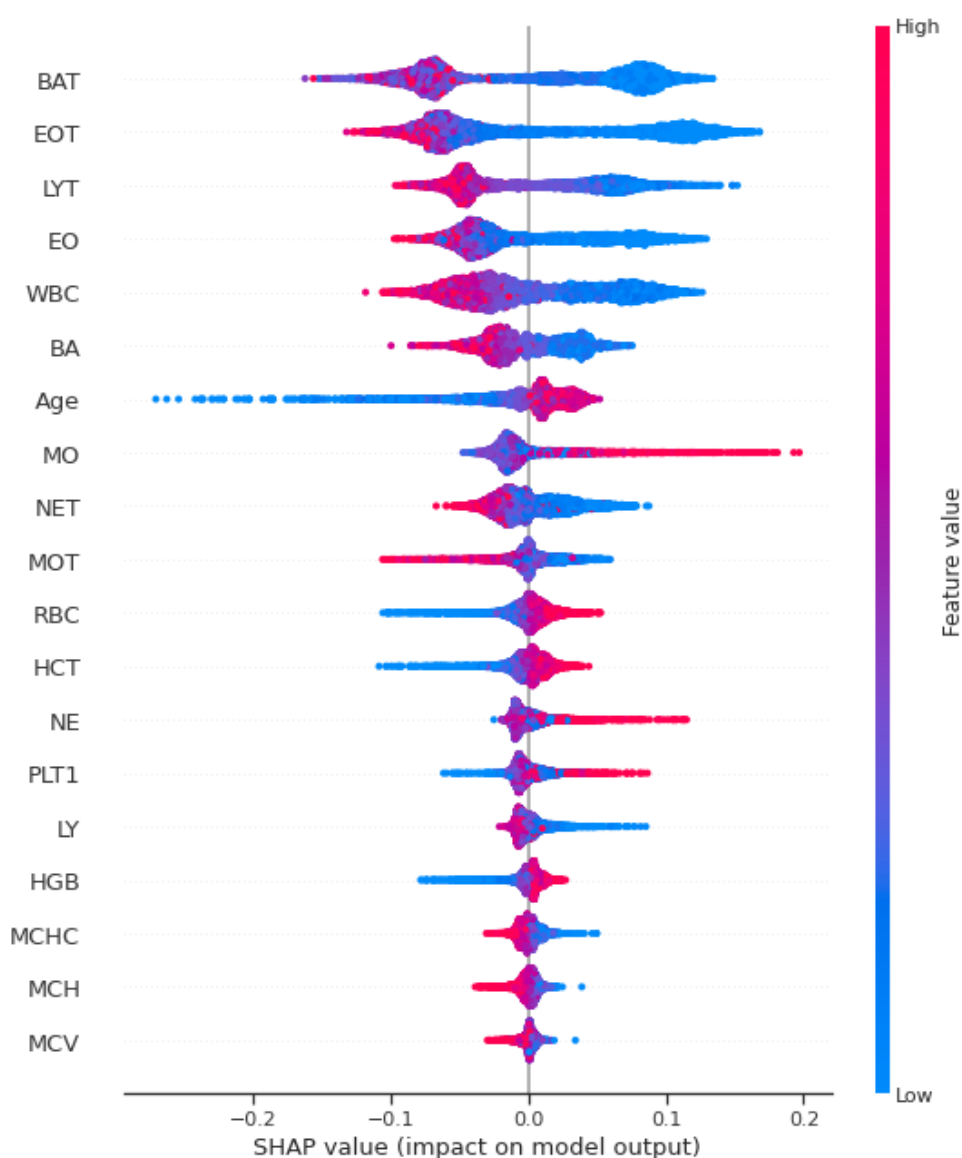
$$\mathbf{V}_{j \times j} = (\mathbf{R}_{j \times j})^{-1} \quad (73)$$

Em que $\mathbf{R}$ é a matriz de correlação e $\mathbf{W}$ é a matriz de dados centrados e reduzidos. Vale ressaltar que a matriz $\mathbf{R}$ será invertível se, e somente se, o posto($\mathbf{W}$) for completo, ou seja, as colunas/linhas de $\mathbf{W}$ forem linearmente independentes (BOLDRINI et al., 1980). A literatura aponta que valores de VIF acima de 5,0 ou 10,0 indicam multicolinearidade (CRANEY; SURLES, 2002), desta forma, atributos com VIF acima de 8,0 não foram selecionados para as simulações.

- O SHAP é uma técnica baseada na teoria dos jogos, usada para explicar a saída dos modelos. O processo de decisão de um modelo pode ser explicado através da utilização

destes valores, que refletem o esforço de cada atributo para a previsão final (LUNDBERG; LEE, 2017). Na Figura 9 é apresentado o gráfico com as contribuições de cada atributo extraídas de um modelo de floresta aleatória.

Figura 9 - Gráfico de importância por valores SHAP.



Fonte: Elaborado pelo autor.

Na Figura 9 também é possível avaliar o comportamento dos valores de cada atributo e sua contribuição, cada ponto do gráfico significa uma amostra e a cor representa o valor, de modo mais resumido, podemos dizer que valores menores de BAT, EOT, LYT, EO e WBC contribuem para as previsões positivas do modelo, ou seja, valores menores destes atributos levam à predição de pacientes positivos para COVID-19. Além disso, valores menores de idade contribuem para a previsão negativa. A ordem em que os atributos aparecem na Figura 9 está relacionado

com o seu grau de importância, sendo BAT, EOT, LYT, EO e WBC os cinco atributos mais relevantes para classificação correta dos pacientes.

A ordem em que cada atributo aparece na coluna importância SHAP da [Tabela 5](#) é baseado na sua contribuição da [Figura 9](#).

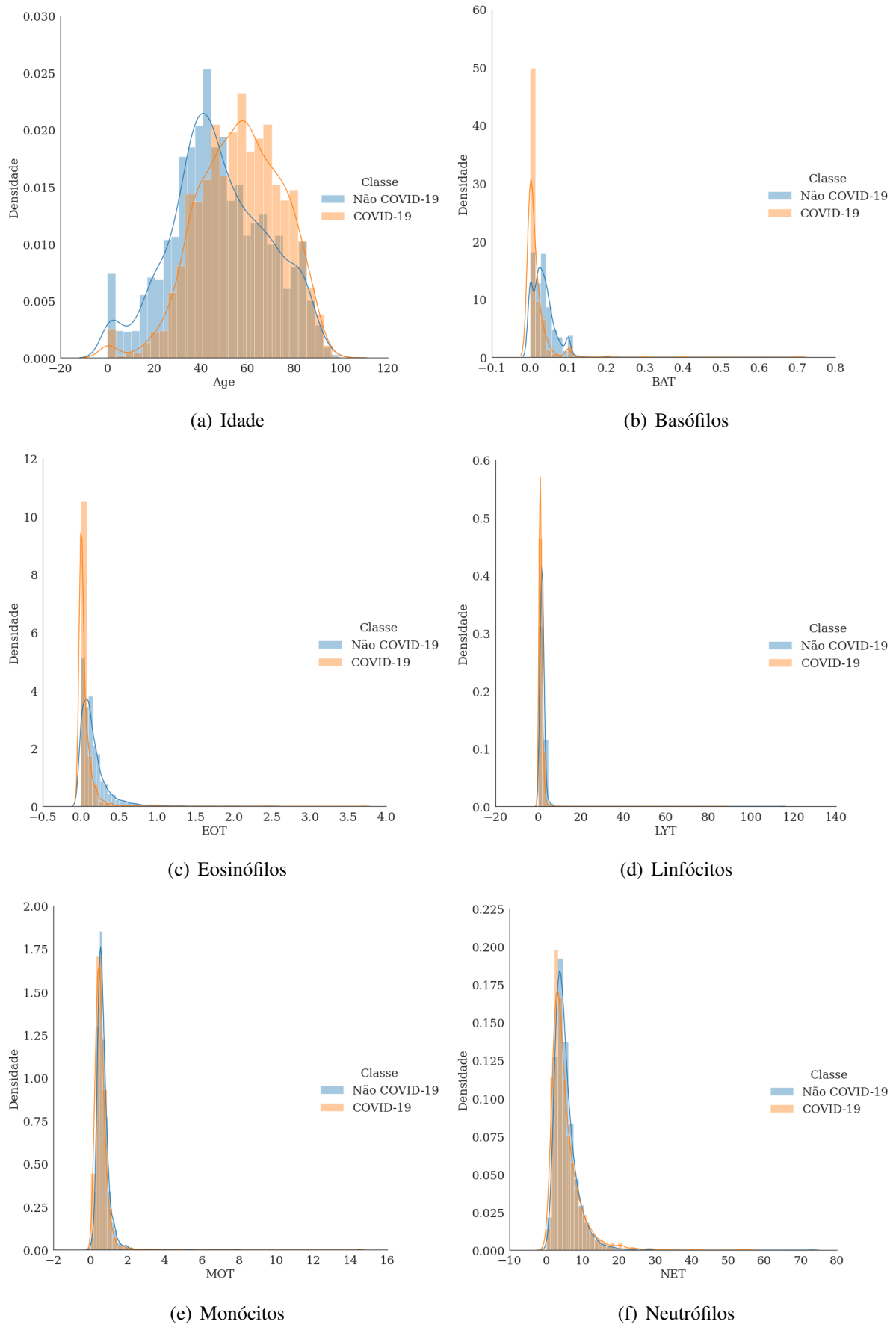
Tabela 5 - Seleção de atributos por diferentes métodos.

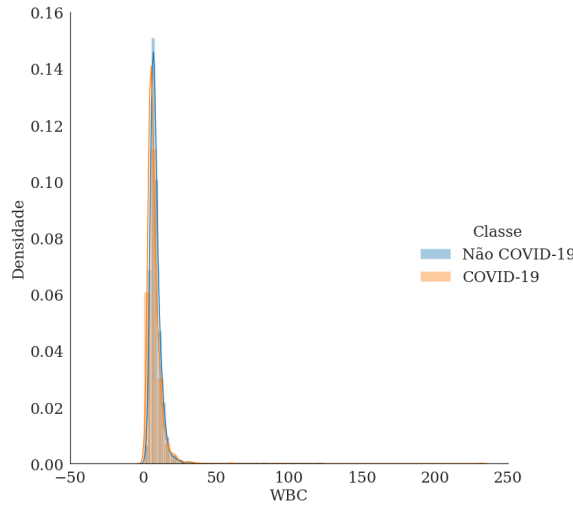
Atributos	<i>p</i> -valor ANOVA	VIF	Importância SHAP
Idade	0,0000	1,4213	7º
BA (%)	0,0000	2,8222	6º
BAT	0,0000	2,6595	1º
EO (%)	0,0000	10,5956	4º
EOT	0,0000	7,2687	2º
HCT	0,0004	400,5673	12º
HGB	0,0236	351,4449	16º
LY (%)	0,0000	100,8982	15º
LYT	0,0000	2,1397	3º
MCH	0,0074	105,2796	18º
MCHC	0,0004	10,5313	17º
MCV	0,2803	81,7446	19º
MO (%)	0,0011	12,1883	8º
MOT	0,0000	2,8672	10º
NE (%)	0,0000	140,5236	13º
NET	0,2321	5,4626	9º
PLT	0,0000	1,2310	14º
RBC	0,0003	59,8770	11º
WBC	0,0000	3,8370	5º

Fonte: Elaborado pelo autor.

Os atributos destacados na [Tabela 5](#) foram selecionados para as simulações. Apenas os atributos com os valores absolutos foram selecionados, uma vez que o valor absoluto tem significado clínico e é mais informativo que os valores relativos (%) ([AGNELLO et al., 2021](#)). Além disso, os atributos selecionados desempenham um papel fundamental no sistema imunológico auxiliando no combate a infecções, vírus e bactérias, analisando-os pode-se constatar algumas informações em relação à pacientes afetados pela COVID-19, como uma diminuição nos basófilos ([Figura 10\(b\)](#)), eosinófilos ([Figura 10\(c\)](#)) e linfócitos ([Figura 10\(d\)](#)), conforme relatado por [Kermali et al. \(2020\)](#), [Samprathi e Jayashree \(2021\)](#) e [Keykavousi et al. \(2022\)](#), tais informações também foram encontradas na [Figura 9](#).

Figura 10 - Histograma e estimativa da função densidade de probabilidade para cada um dos atributos separados por classe.





(g) Leucócitos

Fonte: Elaborado pelo autor.

Após utilizar este conjunto de dados combinado para treinar e testar a rede, ela foi validada com os Dados₁ e Dados₁₂ para avaliar a capacidade de generalização da rede treinada. Vale ressaltar que os mesmos atributos utilizados no treinamento e teste foram também utilizados para validar. O Dados₁ foi utilizado por ser da Itália, um país presente nos dados de treinamento, já o Dados₁₂ foi escolhido por ser da Espanha, um país diferente dos presentes no treinamento.

4.4 MÉTRICAS DE AVALIAÇÃO

As métricas utilizadas para avaliar o desempenho dos modelos foram acurácia, sensibilidade, especificidade, F₁-Score e coeficiente de correlação de Matthews dadas respectivamente pelas Equações 74–78 (PARIKH et al., 2008; HUANG; LING, 2005; BOUGHORBEL; JAR-RAY; EL-ANBARI, 2017; ZHANG; ZHANG, 2009; MATTHEWS, 1975).

- Acurácia (ACC): é a medida que avalia o desempenho geral do modelo.

$$\frac{VP + VN}{VP + VN + FP + FN} \quad (74)$$

- Sensibilidade (Se): indica a capacidade do modelo em detectar os indivíduos verdadeiramente positivos, ou seja, é a probabilidade de um teste dar positivo em quem realmente está doente.

$$\frac{VP}{VP + FN} \quad (75)$$

- Especificidade (Sp): é uma medida complementar à sensibilidade e indica a capacidade do modelo em detectar os verdadeiros negativos, isto é, indica a probabilidade de um teste dar negativo em quem realmente não está doente.

$$\frac{VN}{VN + FP} \quad (76)$$

- F_1 -Score (F_1): é a média harmônica entre precisão e sensibilidade que avalia o equilíbrio entre as duas métricas, recebe valores entre 0 e 1, em que valores próximos de 1 significa melhor desempenho (ZHANG; ZHANG, 2009).

$$\frac{VP}{VP + \frac{1}{2}(FP + FN)} \quad (77)$$

- Coeficiente de Correlação de Matthews (MCC): é uma medida de desempenho usada em classificação binária, e mede a qualidade de uma previsão quando existem desbalanceamentos na distribuição de classes nos dados de treinamento. O MCC varia de -1 a 1, onde valores próximos de 1 indicam um bom desempenho (alta precisão e alta sensibilidade), valores próximos de -1 indicam um desempenho ruim (baixa precisão e baixa sensibilidade) e valores próximos de 0 indicam um desempenho médio. O MCC é uma medida útil quando existe desbalanceamento entre as classes, pois considera todas as combinações de previsões corretas e incorretas, enquanto outras métricas, como a acurácia, podem ser enganosas em tais situações (BOUGHORBEL; JARRAY; EL-ANBARI, 2017).

$$\frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP) \cdot (VP + FN) \cdot (VN + FP) \cdot (VN + FN)}} \quad (78)$$

em que:

- VP : dado positivo classificado corretamente;
- VN : dado negativo classificado corretamente;
- FP : dado negativo classificado incorretamente;
- FN : dado positivo classificado incorretamente.

Além das métricas abordadas, também foi utilizada a curva ROC (*Receiver Operating Characteristic*), um dos estimadores estatísticos mais utilizados para avaliar a qualidade da classificação dos modelos, obtida através da relação entre Sensibilidade e a Taxa de Falso Positivo ($1 - \text{Especificidade}$). Já a AUC resume a curva ROC num único valor que pode ser interpretado conforme a Tabela 6 (MANDREKAR, 2010).

Tabela 6 - Interpretação dos Valores AUC.

Intervalo	Resultado	Intervalo	Resultado
0,5 – 0,6	Ruim	0,7 – 0,8	Aceitável
0,8 – 0,9	Bom	> 0,9	Excelente

Fonte: Elaborado pelo autor.

4.5 CLASSIFICADOR

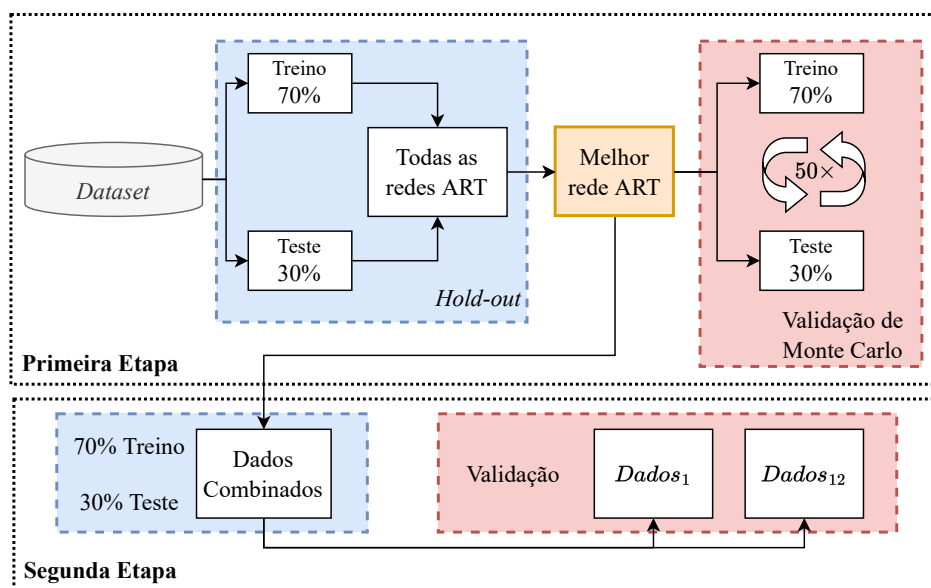
Nas redes com treinamento não supervisionado como a ART Fuzzy AE [Seção 3.1](#), ART Euclidiana AE [Seção 3.2](#), ART Fuzzy AE TC e ART Euclidiana AE TC [Seção 3.4](#) foi necessário utilizar um classificador após o treinamento para rotular todas as categorias. Além de extrair uma saída discreta nesses modelos, também extraiu-se uma probabilidade de cada categoria pertencer a um dos dois rótulos, ou seja, pacientes com “COVID-19” e “não COVID-19”. Tal classificador consistiu na verificação dos padrões dentro de cada categoria, a saída discreta foi determinada por votação majoritária. Desta forma, o rótulo da classe seria definido pela maioria dos padrões dentro da categoria, ou seja, se a maioria dos padrões fossem de pacientes com COVID-19, logo, a classe desta categoria seria definida como COVID-19.

A probabilidade, por outro lado, foi obtida dividindo a quantidade de padrões “COVID-19” pelo número total de padrões presentes dentro da categoria, obtendo desta forma a probabilidade de um padrão nesta categoria ser de um paciente com COVID-19, a probabilidade complementar representa a chance do padrão dentro desta categoria ser de um paciente saudável.

4.6 VALIDAÇÃO

Para a primeira e segunda etapa os dados foram particionados utilizando o método de validação cruzada *hold-out* simples, no qual os dados são divididos aleatoriamente entre treino e previsão de acordo com uma determinada proporção. Além disso, o melhor modelo de cada conjunto de dados na primeira etapa foi submetido ao método de validação cruzada *hold-out* repetido, também conhecido como validação cruzada de Monte Carlo. Este método consiste em divisões aleatórias repetidas do conjunto de dados em duas partes usando uma proporção, uma parte é utilizada como conjunto de treinamento e a outra como conjunto de previsão, este método é útil para avaliar o desempenho do modelo de uma forma mais robusta do que o método *hold-out*, uma vez que são considerados múltiplos subconjuntos de teste em vez de apenas um. Sendo, portanto mais representativo do desempenho do modelo no mundo real ([DUDOIT; LAAN, 2005](#); [BORRA; CIACCIO, 2010](#)). A proporção utilizada para a divisão dos dados em treino e previsão foi 70 – 30, sendo 70% das amostras utilizadas para treinamento e os outros 30% para previsão. Para a validação cruzada de Monte Carlo repetiu-se a validação 50 vezes, vale ressaltar que os dados foram embaralhados e divididos aleatoriamente. Para a segunda etapa foi utilizado o conjunto de dados descrito na [Seção 4.3](#) que combinou exames de diferente países, o método *hold-out* foi aplicado para particionar o conjunto em treino e previsão e utilizado para definir os parâmetros da rede. Após definir os parâmetros, os conjunto de Dados₁ e Dados₁₂, foram utilizados para verificar a capacidade de generalização da rede para novas informações. Na [Figura 11](#) é apresentado o fluxograma do processo de validação utilizado em ambas as etapas.

Figura 11 - Visão geral da validação.



Fonte: Elaborado pelo autor.

5 RESULTADOS COMPUTACIONAIS

Neste capítulo são apresentados os resultados obtidos pelas redes ART para a classificação da COVID-19 nas duas etapas. A implementação das redes foram realizadas utilizando o Sistema Computacional Scilab, versão 6.1.1 (SCILAB, 2012).

5.1 PRIMEIRA ETAPA

Para encontrar os melhores parâmetros de modo que a rede alcançasse os melhores resultados, foi realizada uma busca exaustiva nos parâmetros β e ρ dos modelos para todos os conjuntos de dados, para as redes baseadas na lógica *fuzzy* o parâmetro ρ variou entre $0,001 \leq \rho \leq 0,999$, já a rede baseada na distância euclidiana a busca aconteceu entre $0,000001 \leq \rho \leq 0,2$. Por outro lado, o parâmetro β de ambas as redes variou entre $0,01 \leq \beta \leq 1,0$. Alguns parâmetros foram fixados, como o $\alpha = 0,1$, $\rho_a = 0,6$ (ARTMAP), $\rho_b = 0,99$ e $\rho_{ab} = 0,99$, os outros parâmetros foram definidos por meio de testes.

5.1.1 CONJUNTO DE DADOS₁

A melhor rede para classificar este conjunto de dados foi a ART *Fuzzy* Autoexpansível com treinamento continuado que alcançou acurácia de 0,871 e sensibilidade de 0,926, sendo eficaz em detectar corretamente pacientes com COVID-19. Além disso, quando se observa a Tabela 7 que apresenta o desempenho de todas as redes para o método *hold-out*, fica claro que as redes que usam a lógica *fuzzy* superaram as redes baseadas na distância euclidiana.

Tabela 7 - Resultados obtidos por cada modelo para o Dados₁.

Modelo	ACC	Se	Sp	MCC
ART <i>Fuzzy</i> AE	0,800	0,870	0,670	0,561
ART Euclidiana AE	0,706	0,759	0,613	0,370
ARTMAP <i>Fuzzy</i>	0,835	0,870	0,774	0,645
ART <i>Fuzzy</i> AE TC	0,871	0,926	0,774	0,717
ART Euclidiana AE TC	0,741	0,778	0,677	0,450
ARTMAP <i>Fuzzy</i> AE TC	0,800	0,759	0,871	0,608
ARTMAP Euclidiana AE TC	0,635	0,778	0,387	0,176

Fonte: Elaborado pelo autor.

Na Tabela 8 é apresentado a média das métricas bem como o intervalo com 95% de confiança obtido pela rede ART *Fuzzy* Autoexpansível com Treinamento Continuado na validação

de Monte Carlo, a rede obteve um desempenho reduzido com MCC de 0,6555.

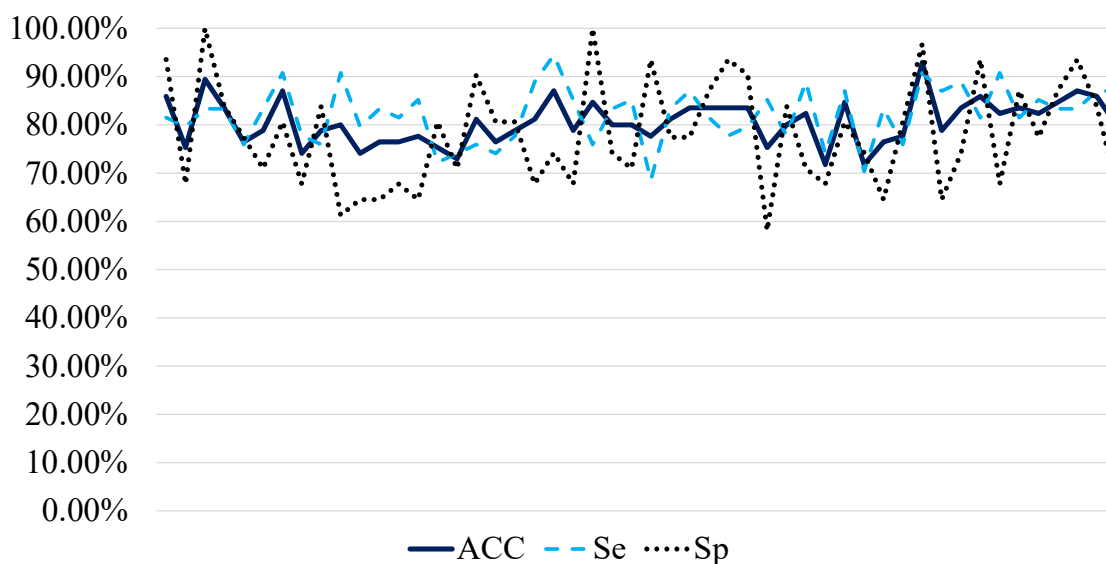
Tabela 8 - Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados₁.

Métricas	Resultado médio (IC 95%)
ACC	0,8061 (0,7932 - 0,8191)
Se	0,8211 (0,8049 - 0,8373)
Sp	0,7800 (0,7495 - 0,8105)
MCC	0,6555 (0,6264 - 0,6846)
F_1	0,8679 (0,8569 - 0,8788)

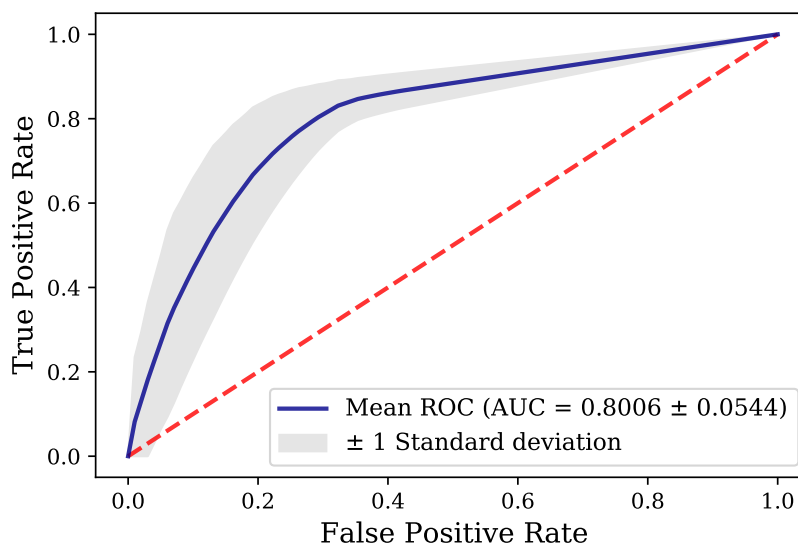
Fonte: Elaborado pelo autor.

Para melhorar a visualização do desempenho em cada iteração da validação de Monte Carlo, na [Figura 12](#) são apresentadas a ACC , Se e Sp , é possível perceber que o desempenho da rede não foi constante, alcançando acurácia máxima de 0,9294 e mínima de 0,7176. Na [Figura 13](#), por outro lado, é apresentada a curva ROC média e a AUC obtida pela rede nesta validação, a rede alcançou pontuação AUC acima de 0,8 podendo ser considerado um bom desempenho conforme a [Tabela 6](#) (MANDREKAR, 2010).

Figura 12 - Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para o Dados₁.



Fonte: Elaborado pelo autor.

Figura 13 - Curva ROC para o conjunto de Dados₁.

Fonte: Elaborado pelo autor.

5.1.2 CONJUNTO DE DADOS₂

Os resultados obtidos por cada modelo para o conjunto de Dados₂ podem ser observados na [Tabela 9](#), a rede ART *Fuzzy* Autoexpansível com Treinamento Continuoado obteve o melhor resultado conseguindo detectar quase que perfeitamente os pacientes positivos e negativos para a COVID-19, as outras redes alcançaram um desempenho aproximado com $MCC \approx 0,5$. Apesar da rede ART Euclidiana Autoexpansível com Treinamento Continuoado ter obtido o melhor resultado em especificidade, ela não foi a melhor, visto que detectou corretamente apenas 54,7% dos paciente com COVID-19, tal diferença de desempenho pode ser observada ao analisar o MCC obtido por cada modelo, tendo a rede ART *Fuzzy* Autoexpansível com Treinamento Continuoado alcançado MCC de 0,967.

Tabela 9 - Resultados obtidos por cada modelo para o Dados₂.

Modelo	ACC	Se	Sp	MCC
ART <i>Fuzzy</i> AE	0,750	0,746	0,754	0,500
ART Euclidiana AE	0,727	0,691	0,762	0,454
ARTMAP <i>Fuzzy</i>	0,783	0,792	0,774	0,566
ART <i>Fuzzy</i> AE TC	0,984	0,970	0,996	0,967
ART Euclidiana AE TC	0,781	0,547	1,000	0,619
ARTMAP <i>Fuzzy</i> AE TC	0,738	0,712	0,762	0,474
ARTMAP Euclidiana AE TC	0,697	0,712	0,683	0,394

Fonte: Elaborado pelo autor.

Por apresentar o melhor desempenho, a rede ART *Fuzzy* Autoexpansível com Treinamento Continuoado foi submetida à validação cruzada de Monte Carlo. Na [Tabela 10](#) são apresentadas

a média das métricas, assim como o intervalo com 95% de confiança obtido pela rede após a validação. A rede alcançou um excelente desempenho com acurácia média de 0,9348 e $MCC = 0,8721$.

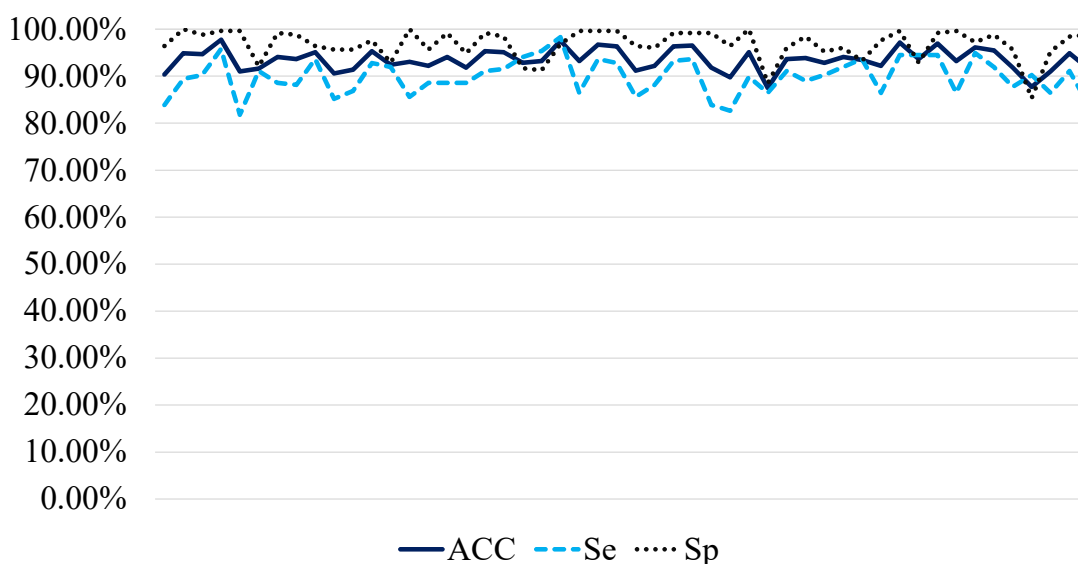
Tabela 10 - Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados₂.

Métricas	Resultado médio (IC 95%)
ACC	0,9348 (0,9282 - 0,9413)
Se	0,8992 (0,8885 - 0,9098)
Sp	0,9682 (0,9594 - 0,9770)
MCC	0,8721 (0,8592 - 0,8850)
F_1	0,9301 (0,9231 - 0,9371)

Fonte: Elaborado pelo autor.

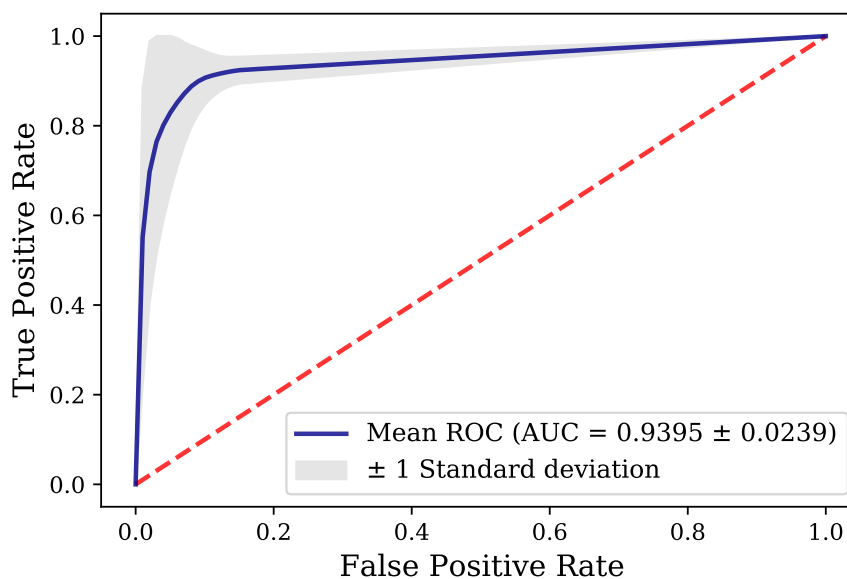
Na [Figura 14](#) é possível observar o excelente desempenho obtido pela rede em cada iteração da validação de Monte Carlo por meio das métricas ACC , Se e Sp , e diferente do que aconteceu para o conjunto de Dados₁, a rede mostrou-se bastante estável mantendo um bom desempenho em praticamente todas as iterações. A acurácia máxima foi 0,9775 e a mínima 0,875, das 50 iterações apenas cinco delas ficaram com acurácia abaixo de 0,91.

Figura 14 - Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados₂.



Fonte: Elaborado pelo autor.

A curva ROC e a AUC média obtida após a validação de Monte Carlo pode ser observada na [Figura 15](#). Nota-se que a AUC obtida foi de 0,9395, indicando uma excelente capacidade de discriminação conforme proposto por [Mandrekar \(2010\)](#).

Figura 15 - Curva ROC para o conjunto de Dados₂.

Fonte: Elaborado pelo autor.

5.1.3 CONJUNTO DE DADOS₃

Com 13,33% dos dados sendo de pacientes com COVID-19 e 86,67% de pacientes não COVID-19, este é o conjunto mais desbalanceado. Como esperado, esse desbalanceamento prejudicou a capacidade de classificação da maioria dos modelos implementados, resultando num baixo desempenho preditivo conforme mostra a [Tabela 11](#).

Tabela 11 - Resultados obtidos por cada modelo para o Dados₃.

Modelo	<i>ACC</i>	<i>Se</i>	<i>Sp</i>	<i>MCC</i>
ART <i>Fuzzy</i> AE	0,889	0,208	0,994	0,382
ART Euclidiana AE	0,878	0,500	0,936	0,452
ARTMAP <i>Fuzzy</i>	0,867	0,375	0,942	0,359
ART <i>Fuzzy</i> AE TC	0,944	0,583	1,000	0,740
ART Euclidiana AE TC	0,972	0,792	1,000	0,876
ARTMAP <i>Fuzzy</i> AE TC	0,889	0,458	0,955	0,468
ARTMAP Euclidiana AE TC	0,872	0,458	0,936	0,417

Fonte: Elaborado pelo autor.

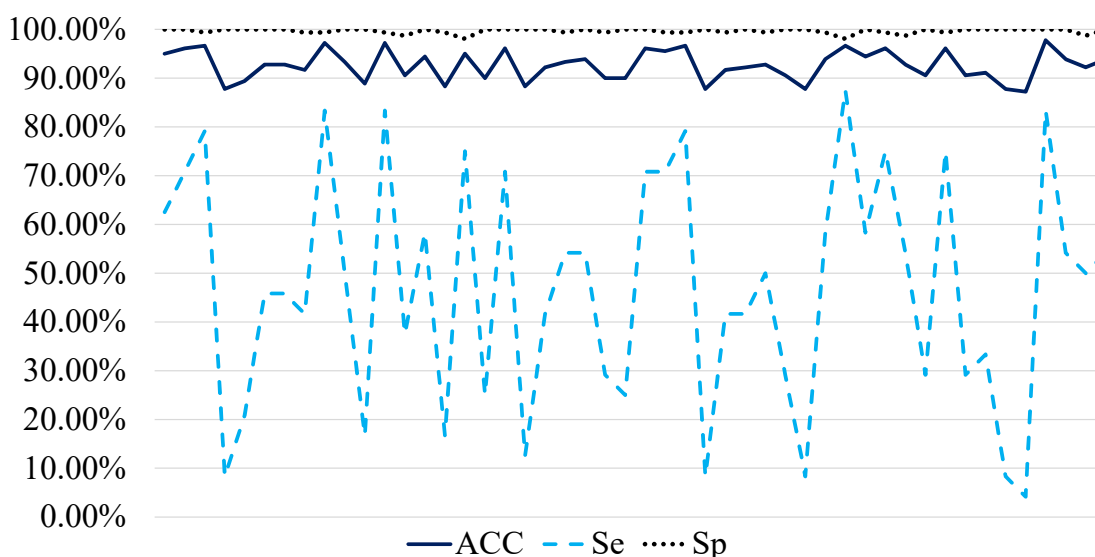
Considerando que este conjunto de dados possui classes desbalanceadas, métricas como a acurácia, sensibilidade e especificidade não são boas indicadas para avaliar o desempenho dos modelos ([HUANG; LING, 2005](#); [PARIKH et al., 2008](#)), sendo necessário utilizar métricas como o *MCC* e *AUC*, visto que elas são pouco influenciadas por dados desbalanceados ([HUANG; LING, 2005](#); [BOUGHORBEL; JARRAY; EL-ANBARI, 2017](#)). A ART Euclidiana Autoexpansível com Treinamento Continuado obteve o melhor desempenho, atingindo *MCC* = 0,876. Já na [Tabela 12](#) são apresentadas a média das métricas e o intervalo com 95% de confiança após a validação de Monte Carlo.

Tabela 12 - Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados₃.

Métricas	Resultado médio (IC 95%)
<i>ACC</i>	0,9265 (0,9180 - 0,9350)
<i>Se</i>	0,4714 (0,4045 - 0,5382)
<i>Sp</i>	0,9965 (0,9951 - 0,9979)
<i>MCC</i>	0,6195 (0,5656 - 0,6734)
<i>F₁</i>	0,5924 (0,5268 - 0,6580)

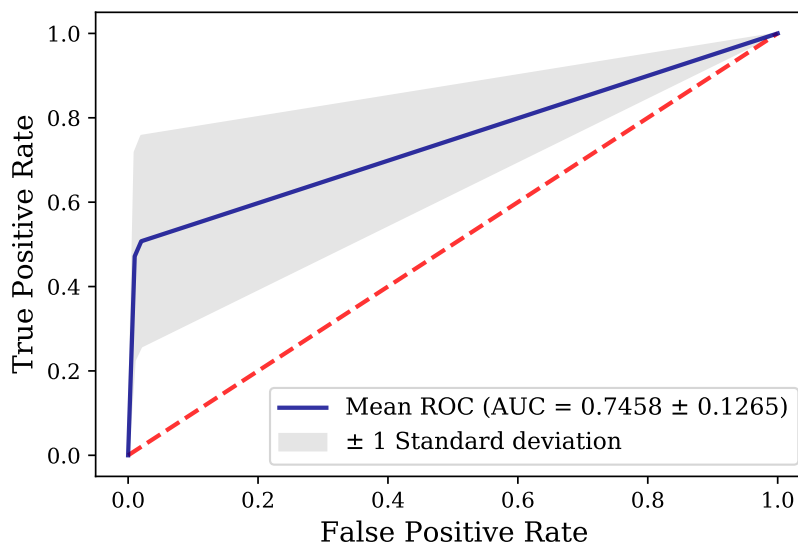
Fonte: Elaborado pelo autor.

Analisando a [Tabela 12](#) pode-se notar que o modelo obteve uma alta especificidade, um resultado já esperado, uma vez que a classe majoritária são de pacientes não infectados pela COVID-19, em contrapartida, a sensibilidade média foi de apenas 0,4714, ao observar o *MCC* nota-se que a rede não obteve bom desempenho, alcançando uma média de 0,6195. Comparando as Tabelas [11](#) e [12](#) observa-se que houve uma redução no desempenho da rede, em algumas iterações da validação de Monte Carlo a rede não foi capaz de detectar corretamente os pacientes com COVID-19, com a sensibilidade ficando abaixo de 30% em algumas iterações conforme ilustra a [Figura 16](#).

Figura 16 - Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados₃.

Fonte: Elaborado pelo autor.

Na [Figura 17](#) esboça-se a curva ROC e a AUC obtida pela rede, analisando-a tem-se que a rede manteve a especificidade estável, entretanto houve uma variabilidade muito grande na sensibilidade, a AUC média alcançada foi de 0,7458 com desvio de 0,1265 indicando uma classificação aceitável conforme [Mandrekar \(2010\)](#).

Figura 17 - Curva ROC para o conjunto de Dados₃.

Fonte: Elaborado pelo autor.

5.1.4 CONJUNTO DE DADOS₃ BALANCEADO

Considerando que o conjunto de Dados₃ é desbalanceado e isso prejudica a qualidade da classificação, foi utilizada a estratégia *Random Undersampling*, com o intuito de reduzir de maneira aleatória os dados da classe majoritária, conforme realizado por Rikan et al. (2022), desta forma ambas as classes COVID-19 e não COVID-19 possuem 80 dados. Na Tabela 13 são apresentados os resultados para cada uma das redes utilizadas e ao analisar a tabela nota-se que a rede ART *Fuzzy* Autoexpansível com Treinamento Continuoado obteve o melhor desempenho dentre as redes testadas, alcançado $MCC = 0,959$.

Tabela 13 - Resultados obtidos por cada modelo para o Dados₃ Balanceado.

Modelo	<i>ACC</i>	<i>Se</i>	<i>Sp</i>	<i>MCC</i>
ART <i>Fuzzy</i> AE	0,812	0,750	0,875	0,630
ART Euclidiana AE	0,729	0,750	0,708	0,459
ARTMAP <i>Fuzzy</i>	0,833	0,833	0,833	0,667
ART <i>Fuzzy</i> AE TC	0,979	0,958	1,000	0,959
ART Euclidiana AE TC	0,896	0,792	1,000	0,809
ARTMAP <i>Fuzzy</i> AE TC	0,812	0,792	0,833	0,625
ARTMAP Euclidiana AE TC	0,750	0,792	0,708	0,502

Fonte: Elaborado pelo autor.

Como a rede ART *Fuzzy* Autoexpansível com Treinamento Continuoado obteve o melhor resultado na validação *hold-out*, a mesma foi selecionada para a validação cruzada de Monte Carlo. O resultado médio alcançado pela rede e o intervalo com 95% de confiança pode ser visto na Tabela 14. Analisando-a pode-se notar que a rede alcançou uma acurácia média de 0,9560 e MCC de 0,9150, indicando um excelente desempenho preditivo.

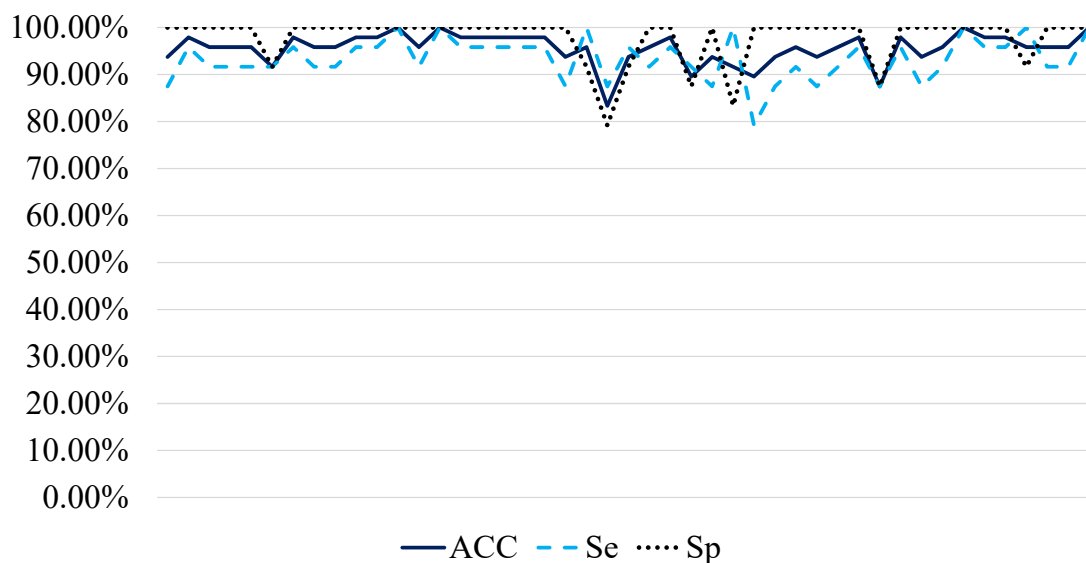
Tabela 14 - Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados₃ Balanceado.

Métricas	Resultado médio (IC 95%)
<i>ACC</i>	0,9560 (0,9462 - 0,9658)
<i>Se</i>	0,9333 (0,9200 - 0,9467)
<i>Sp</i>	0,9787 (0,9641 - 0,9933)
<i>MCC</i>	0,9150 (0,8957 - 0,9342)
<i>F₁</i>	0,9550 (0,9452 - 0,9648)

Fonte: Elaborado pelo autor.

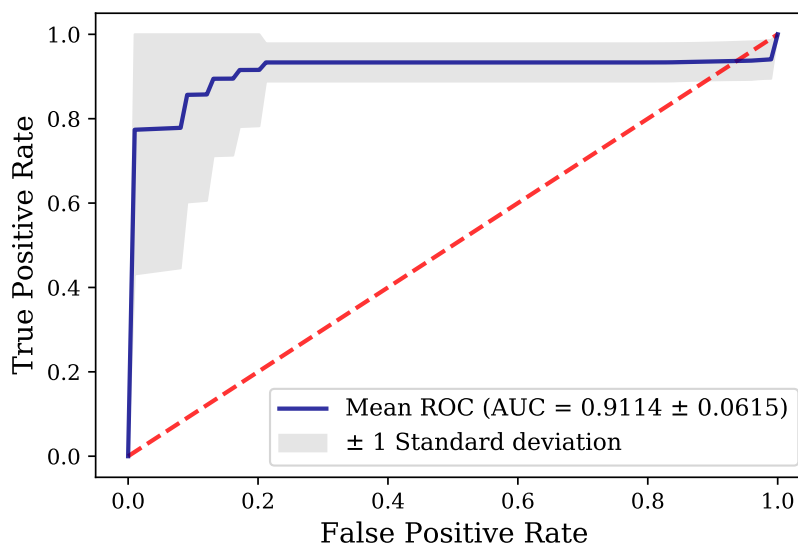
Na [Figura 18](#) é possível visualizar a acurácia, sensibilidade e especificidade obtida em cada umas das 50 iterações da validação de Monte Carlo. A rede manteve seu desempenho acima de 90% na maior parte do tempo e a acurácia mínima registrada foi de 0,8333, ou seja, acertando 40 dos 48 dados destinados para previsão.

Figura 18 - Desempenho para cada uma das 50 iterações da validação cruzada de Monte Carlo para os Dados₃ Balanceado.



Fonte: Elaborado pelo autor.

A AUC alcançada foi de 0,9114, indicando ser um excelente discriminador ([MANDREKAR, 2010](#)), a Curva ROC e a AUC podem ser vistas na [Figura 19](#).

Figura 19 - Curva ROC para o conjunto de Dados₃ Balanceado.

Fonte: Elaborado pelo autor.

5.1.5 CONJUNTO DE DADOS_{4T}

Na Tabela 15 são apresentados os resultados obtidos pelas redes para o conjunto de Dados_{4T}, nela pode-se notar que todas as redes baseadas na distância euclidiana obtiveram bons resultados, tanto em acurácia quanto em sensibilidade ficando acima de 0,9 e 0,933, respectivamente, a rede ART *Fuzzy* Autoexpansível com Treinamento Continuo conseguiu o melhor resultado dentre todas as redes utilizadas atingindo acurácia de 0,95 e $MCC = 0,900$.

Tabela 15 - Resultados obtidos para os Dados_{4T}.

Modelo	<i>ACC</i>	<i>Se</i>	<i>Sp</i>	<i>MCC</i>
ART <i>Fuzzy</i> AE	0,803	0,933	0,673	0,628
ART Euclidiana AE	0,900	0,967	0,833	0,807
ARTMAP <i>Fuzzy</i>	0,850	0,967	0,733	0,720
ART <i>Fuzzy</i> AE TC	0,950	0,967	0,933	0,900
ART Euclidiana AE TC	0,927	0,973	0,880	0,857
ARTMAP <i>Fuzzy</i> AE TC	0,867	0,947	0,787	0,743
ARTMAP Euclidiana AE TC	0,940	0,993	0,887	0,885

Fonte: Elaborado pelo autor.

Já na Tabela 16 é possível ver os resultados obtidos após a remoção dos atributos hematócrito e hemoglobina. Analisando-a nota-se que a remoção destes atributos trouxe um aumento na acurácia das redes mais simples como a ART *Fuzzy* Autoexpansível e ART Euclidiana Autoexpansível, já as redes com aprendizado supervisionado ou treinamento continuado essa diferença pouco aconteceu, por outro lado, ao remover estes atributos algumas redes tiveram um

desempenho maior na classificação correta de pacientes com COVID refletindo na sensibilidade de cada modelo.

Tabela 16 - Resultados após remoção de hematócrito e hemoglobina dos Dados_{4T}.

Modelo	<i>ACC</i>	<i>Se</i>	<i>Sp</i>	<i>MCC</i>
ART <i>Fuzzy</i> AE	0,820	0,900	0,740	0,648
ART Euclidiana AE	0,927	0,987	0,867	0,859
ARTMAP <i>Fuzzy</i>	0,847	0,947	0,747	0,708
ART <i>Fuzzy</i> AE TC	0,950	0,900	1,000	0,904
ART Euclidiana AE TC	0,927	0,980	0,873	0,858
ARTMAP <i>Fuzzy</i> AE TC	0,870	0,940	0,800	0,747
ARTMAP Euclidiana AE TC	0,930	0,973	0,887	0,933

Fonte: Elaborado pelo autor.

Na Tabela 17 são apresentados os resultados obtidos pelas redes ART *Fuzzy* AE TC e ARTMAP Euclidiana AE TC na validação de Monte Carlo, nela pode-se notar que a rede baseada na distância euclidiana obteve um resultado satisfatório para os dois conjuntos de dados testados.

Tabela 17 - Resultado Médio após validação cruzada de Monte Carlo para o conjunto de Dados_{4T}.

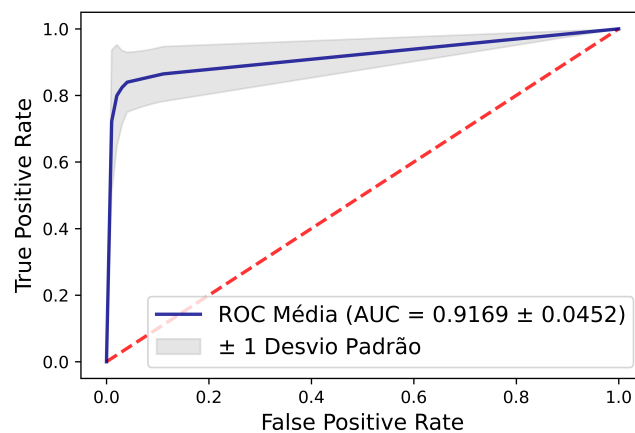
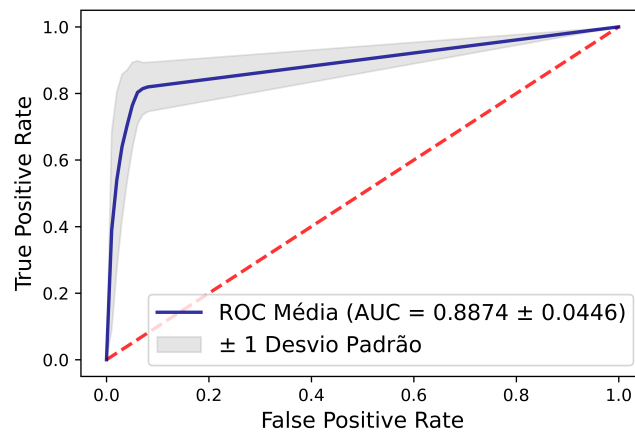
Dados	Métricas	Resultado médio (<i>IC</i> 95%)			
		ART <i>Fuzzy</i> Autoexpansível com Treinamento Continuado		ARTMAP Euclidiana Autoexpansível com Treinamento Continuado	
Dados _{4T}	<i>ACC</i>	0,8814	(0,8662 – 0,8967)	0,9196	(0,9089 – 0,9303)
	<i>Se</i>	0,7709	(0,7422 – 0,7997)	0,9980	(0,9951 – 1,0000)
	<i>Sp</i>	0,9919	(0,9854 – 0,9984)	0,8412	(0,8198 – 0,8226)
	<i>MCC</i>	0,7827	(0,7555 – 0,8098)	0,8163	(0,7933 – 0,8392)
	<i>F₁</i>	0,8661	(0,8426 – 0,8845)	0,9124	(0,8999 – 0,9249)
Dados _{4TA}	<i>ACC</i>	0,8770	(0,8884 – 0,8957)	0,9196	(0,9126 – 0,9266)
	<i>Se</i>	0,7926	(0,7598 – 0,8253)	0,9953	(0,9924 – 0,9981)
	<i>Sp</i>	0,9615	(0,9467 – 0,9763)	0,8439	(0,8308 – 0,8571)
	<i>MCC</i>	0,7659	(0,7305 – 0,8033)	0,8149	(0,7996 – 0,8302)
	<i>F₁</i>	0,8650	(0,8431 – 0,8869)	0,9129	(0,9048 – 0,9209)

Dados_{4T} = Dados **com** os atributos hemoglobina e hematócrito;

Dados_{4TA} = Dados **sem** os atributos hemoglobina e hematócrito.

Fonte: Elaborado pelo autor.

Além disso, ao remover os atributos hemoglobina e hematócrito houve uma pequena redução na capacidade preditiva dos modelos. Quando analisamos a sensibilidade que avalia a capacidade de diagnosticar corretamente pessoas doentes, ou seja, pacientes infectados com COVID-19, a rede ARTMAP superou ART *Fuzzy* alcançando sensibilidade acima de 0,9950.

Figura 20 - Curvas ROC obtidas pelas redes ART para os conjuntos Dados_{4T} e Dados_{4TA}.(a) ARTFAETC para os Dados_{4T}(b) ARTFAETC para os Dados_{4TA}

Na Figura 20 observa-se a curva ROC e a AUC que a rede ART *Fuzzy* AE TC obteve para cada conjunto de dados testado. Ela apresentou AUC de 0,9169 e 0,8874 para o conjunto de dados_{4T} e dados_{4TA}, respectivamente. Como a curva ROC avalia a sensibilidade e especificidade para diferentes limiares, nenhuma curva ROC e, consequentemente, a AUC foi obtida para a rede ARTMAP. Uma vez que a mesma produziria apenas um ponto no espaço ROC por possuir saída discreta, diferente da rede ART, a qual é possível atribuir uma probabilidade de pertencimento a uma das duas classes, bastando apenas considerar a quantidade de padrões presentes em cada categoria durante o treinamento.

5.2 COMPARAÇÃO DOS RESULTADOS OBTIDOS NA PRIMEIRA ETAPA

A Tabela 18 apresenta as métricas de avaliação obtidas pelas redes utilizadas neste artigo, bem como os resultados encontrados na literatura para os conjuntos utilizados na primeira etapa.

Tabela 18 - Comparação dos resultado para os conjuntos de dados utilizados.

Autor	Modelo	Dados	Partição	ACC	Se	Sp	MCC	F ₁ - Score	AUC
Este estudo	ART <i>Fuzzy</i> AE	Dados ₁	70-30	0,8000	0,8704	0,6774	0,5610	0,8468	0,7739
	ART Euclidiana AE	Dados ₁	70-30	0,7059	0,7593	0,6129	0,3697	0,7664	0,6861
	ARTMAP <i>Fuzzy</i>	Dados ₁	70-30	0,8353	0,8704	0,7742	0,6446	0,8704	-
	ART <i>Fuzzy</i> AE TC	Dados ₁	70-30	0,8706	0,9259	0,7742	0,7170	0,9009	0,8501
	ART Euclidiana AE TC	Dados ₁	70-30	0,7412	0,7778	0,6774	0,4496	0,7925	0,7276
	ARTMAP <i>Fuzzy</i> AE TC	Dados ₁	70-30	0,8000	0,7593	0,8710	0,6078	0,8283	-
	ARTMAP Euclidiana AE TC	Dados ₁	70-30	0,6353	0,7778	0,3871	0,1763	0,7304	-
	ART <i>Fuzzy</i> AE TC	Dados ₁	Monte Carlo	0,8061	0,8211	0,7800	0,6555	0,8679	0,8006
(RIKAN et al., 2022)	ET	Dados ₁	4-Fold	0,8000	0,7500	0,9444	0,6133	0,8478	0,7710
(BRINATI et al., 2020)	DNN	Dados ₁	4-Fold	0,9211	0,9614	0,8456	0,8250	0,9388	0,9220
	RF	Dados ₁	80-20	0,8200	0,9200	0,6500	-	-	0,8400
	TWRF	Dados ₁	70-30	0,8600	0,9500	0,7500	-	-	-
Este estudo	ART <i>Fuzzy</i> AE	Dados ₂	70-30	0,7500	0,7459	0,7540	0,4996	0,7426	0,7499
	ART Euclidiana AE	Dados ₂	70-30	0,7275	0,6907	0,7619	0,4540	0,7102	0,7263
	ARTMAP <i>Fuzzy</i>	Dados ₂	70-30	0,7828	0,7924	0,7738	0,5659	0,7792	-
	ART <i>Fuzzy</i> AE TC	Dados ₂	70-30	0,9836	0,9703	0,9960	0,9674	0,9828	0,9481
	ART Euclidiana AE TC	Dados ₂	70-30	0,7807	0,5466	1,000	0,6194	0,7068	0,7733
	ARTMAP <i>Fuzzy</i> AE TC	Dados ₂	70-30	0,7377	0,7119	0,7619	0,4745	0,7241	-
	ARTMAP Euclidiana AE TC	Dados ₂	70-30	0,6967	0,7119	0,6828	0,3942	0,6942	-
	ART <i>Fuzzy</i> AE TC	Dados ₂	Monte Carlo	0,9348	0,8992	0,9682	0,8721	0,9301	0,9395
(RIKAN et al., 2022)	ET	Dados ₂	4-Fold	0,8498	0,8788	0,8221	0,7012	0,8509	0,8510
(CABITZA et al., 2021a)	DNN	Dados ₂	4-Fold	0,9316	0,9327	0,9302	0,8633	0,9263	0,9320
	KNN	Dados ₂	hold-out	0,8600	0,8000	0,9200	-	-	0,8700

Continua

Tabela 18 - Comparação dos resultado para os conjuntos de dados utilizados.

Autor	Modelo	Dados	Partição	ACC	Se	Sp	MCC	F ₁ - Score	AUC
Este estudo	ART <i>Fuzzy</i> AE	Dados ₃	70-30	0,8889	0,2083	0,9936	0,3824	0,3333	0,6010
	ART Euclidiana AE	Dados ₃	70-30	0,8778	0,5000	0,9359	0,4524	0,5217	0,7179
	ARTMAP <i>Fuzzy</i>	Dados ₃	70-30	0,8667	0,3750	0,9423	0,3595	0,4286	-
	ART <i>Fuzzy</i> AE TC	Dados ₃	70-30	0,9444	0,5833	1,0000	0,7404	0,7368	0,7917
	ART Euclidiana AE TC	Dados ₃	70-30	0,9722	0,7917	1,0000	0,8758	0,8837	0,8962
	ARTMAP <i>Fuzzy</i> AE TC	Dados ₃	70-30	0,8889	0,4583	0,9551	0,4685	0,5238	-
	ARTMAP Euclidiana AE TC	Dados ₃	70-30	0,8722	0,4583	0,9359	0,4175	0,4889	-
	ART Euclidiana AE TC	Dados ₃	Monte Carlo	0,9265	0,4714	0,9965	0,6195	0,5924	0,7458
(RIKAN et al., 2022)	ET	Dados ₃	4-Fold	0,9200	0,7143	0,9301	0,4530	0,4545	0,6590
	DNN	Dados ₃	4-Fold	0,9333	0,7705	0,9547	0,6904	0,7249	0,8597
(ALAKUS; TURKOGLU, 2020)	LSTM	Dados ₃	10-Fold	0,8666	0,9942	-	-	0,9189	0,6250
	CNN-LSTM	Dados ₃	80-20	0,9230	0,9368	-	-	0,9300	0,9000
Este estudo	ART <i>Fuzzy</i> AE	Dados _{3B}	70-30	0,8125	0,8750	0,6299	0,8000	0,8468	0,8125
	ART Euclidiana AE	Dados _{3B}	70-30	0,7292	0,7500	0,7083	0,4587	0,7347	0,7292
	ARTMAP <i>Fuzzy</i>	Dados _{3B}	70-30	0,8333	0,8333	0,8333	0,6667	0,8333	-
	ART <i>Fuzzy</i> AE TC	Dados _{3B}	70-30	0,9792	0,9583	1,0000	0,9592	0,9787	0,9583
	ART Euclidiana AE TC	Dados _{3B}	70-30	0,8958	0,7917	1,0000	0,8094	0,8837	0,8958
	ARTMAP <i>Fuzzy</i> AE TC	Dados _{3B}	70-30	0,8125	0,7917	0,8333	0,6255	0,8085	-
	ARTMAP Euclidiana AE TC	Dados _{3B}	70-30	0,7500	0,7917	0,7083	0,5017	0,7600	-
	ART <i>Fuzzy</i> AE TC	Dados _{3B}	Monte Carlo	0,9560	0,9333	0,9787	0,9150	0,9550	0,9114
(RIKAN et al., 2022)	ET	Dados _{3B}	4-Fold	0,9000	0,9583	0,8125	0,7917	0,9200	0,9070
	DNN	Dados _{3B}	4-Fold	0,9250	1,0000	0,8500	0,8597	0,9302	0,9220

Fonte: Elaborado pelo autor.

Fim da tabela

Para o conjunto de Dados₁ a rede ART *Fuzzy* AE TC obteve uma AUC de 0,8501 com o método *hold-out* 70–30 e de 0,8006 na validação de Monte Carlo, superando modelos tradicionais como RF e ET, indicando ser eficiente na classificação de pacientes com e sem COVID-19. Entretanto, o modelo de aprendizado profundo DNN utilizado por Rikan et al. (2022) obteve o melhor resultado com um valor AUC de 0,9220.

Ao observar os resultados para o conjunto de Dados₂ é possível notar que a ART *Fuzzy* AE TC obteve AUC de 0,9481 na validação *hold-out* e de 0,9395 na validação de Monte Carlo, sendo superior aos modelos de aprendizado de máquina frequentemente utilizadas como ET, KNN e o algoritmo de aprendizado profundo DNN que obteve AUC de 0,9320. A sensibilidade alcançada pela rede ART também foi superior, sendo de 97,03%. Mesmo quando submetida a validação cruzada de Monte Carlo a rede manteve o bom desempenho superando novamente o DNN, apontando a robustez do modelo.

Na Tabela 18 também consta os resultados deste estudo e da literatura para os conjuntos de Dados₃ e Dados_{3B} que é o conjunto com as classes balanceadas, para o conjunto de Dados₃ a rede ART Euclidiana AE TC obteve resultados próximos ao CNN-LSTM utilizado por Alakus e Turkoglu (2020) com AUC de 0,8962 e 0,9000, no entanto, após a validação de Monte Carlo houve uma redução no desempenho da rede devido ao desequilíbrio significativo entre as classes diminuindo a capacidade preditiva dos modelos.

Ao realizar o balanceamento das classes utilizando *Random Undersampling* no conjunto de Dados₃, os resultados obtidos pelas redes ART melhoraram consideravelmente, a rede ART *Fuzzy* AE TC obteve AUC de 0,9583 na validação *hold-out* e de 0,9114 na de Monte Carlo, superando o modelo ET e se aproximando do modelo DNN que obteve AUC de 0,9220.

A Tabela 19 traz uma comparação entre o tempo computacional (em segundos) de alguns dos modelos implementados nesse artigo, com o melhor modelo utilizado por Rikan et al. (2022), ou seja, a Rede Neural Profunda (DNN). Estão destacados em negrito os melhores algoritmos para cada conjunto de dados.

Tabela 19 - Tempo computacional (em segundos) para cada conjunto de dados.

Algoritmo	Dados ₁	Dados ₂	Dados ₃
ART <i>Fuzzy</i> AE TC	1,25	6,34	5,14
ART Euclidiana AE TC	1,22	0,62	1,67
DNN	20,2	40,32	21,01

Fonte: Elaborado pelo autor.

Analisando o tempo computacional gasto por cada modelo, fica evidente que as redes ART são consideravelmente mais rápidas que a DNN. Essa velocidade de convergência é uma característica das redes ART e acentuada pelo fato da rede ser Auto-expansível. Além da vantagem do tempo computacional reduzido, os modelos baseados na teoria da ressonância adaptativa não sofrem do esquecimento catastrófico como aqueles baseados no algoritmo *backpropagation* e o

Deep Learning conforme relata [Grossberg \(2020\)](#).

Nesta primeira etapa o melhor modelo foi a rede ART *Fuzzy* AE TC, uma rede com aprendizagem supervisionada e treinamento *on-line* e será, portanto, utilizada para a segunda etapa.

5.3 SEGUNDA ETAPA

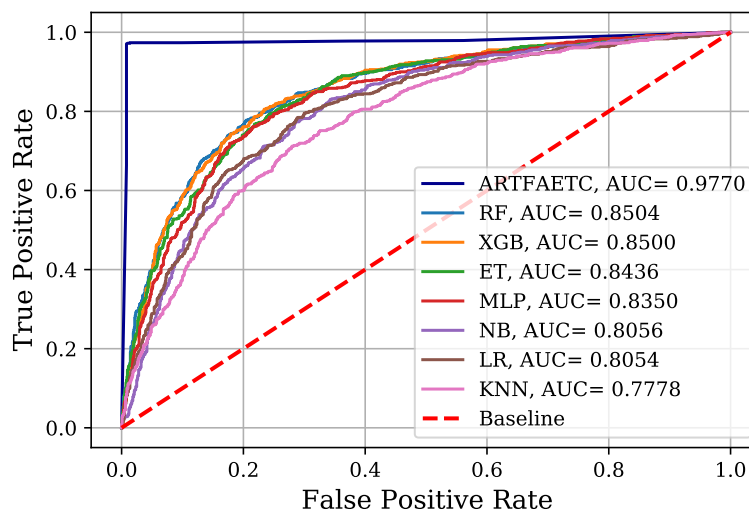
Para que fosse possível realizar uma comparação da rede ART *Fuzzy* AE TC (ARTFAETC), modelos tradicionais empregados pela literatura foram testados com o mesmo conjunto de dados, algoritmos como KNN ([COVER; HART, 1967](#)), RF ([HO, 1995](#)), MLP ([MCCULLOCH; PITTS, 1943](#); [WERBOS, 1974](#)), LR ([WILSON; WORCESTER, 1943](#)), ET ([GEURTS; ERNST; WEHENKEL, 2006](#)), XGB ([CHEN; GUESTRIN, 2016](#)) e NB ([BAYES; PRICE, 1763](#); [ZHANG, 2004](#)) foram utilizados, foram implementados utilizando a linguagem de programação *Python* e a biblioteca *Scikit-Learn* ([PEDREGOSA et al., 2011](#)). Os melhores parâmetros para cada modelo foram definidos pelo método *GridSearchCV*, o qual testa diferentes combinações de hiperparâmetros em um modelo para encontrar a combinação que resulte no melhor desempenho. Os resultados alcançados podem ser observados na [Tabela 20](#), e nela pode-se notar a discrepância entre a rede ART e os modelos tradicionais, superando com folga todos os modelos e obtendo *MCC* de 0,9628.

Tabela 20 - Resultados obtidos por cada modelo para o conjunto de dados combinados.

Modelo	ARTFAETC	KNN	RF	MLP	LR	ET	XGB	NB
<i>ACC</i>	0,9828	0,7277	0,7931	0,7825	0,7617	0,7748	0,7333	0,7429
<i>Se</i>	0,9667	0,5986	0,7014	0,6806	0,6389	0,5806	0,8722	0,6806
<i>Sp</i>	0,9920	0,8019	0,8458	0,8411	0,8323	0,8866	0,6534	0,7788
<i>F₁-Score</i>	0,9762	0,6162	0,7123	0,6955	0,6619	0,6531	0,7048	0,6590
<i>MCC</i>	0,9628	0,4059	0,5510	0,5267	0,4790	0,4988	0,5076	0,4536

Fonte: Elaborado pelo autor.

Figura 21 - Curva ROC e AUC para o conjunto de dados combinados.



Fonte: Elaborado pelo autor.

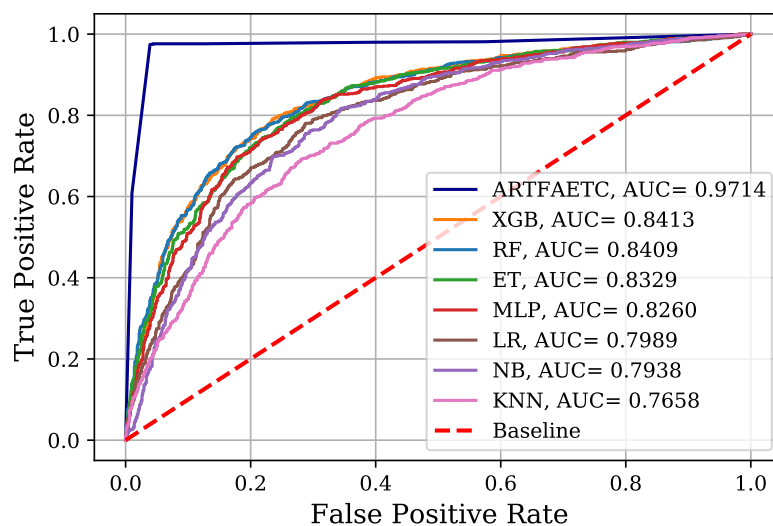
Além disso, na [Figura 21](#) é possível visualizar a curva ROC e a AUC para cada modelo, a rede ART *Fuzzy* AE TC obteve AUC de 0,9770, indicando que a rede possui uma alta precisão na previsão da classe correta. Dentre os modelos tradicionais testados o algoritmo RF obteve o melhor desempenho com AUC e *MCC* de 0,8504 e 0,5510, respectivamente, sendo um resultado bem inferior ao da rede ART. Na [Tabela 21](#) e [Figura 22](#) são apresentados, respectivamente, as métricas e a curva ROC juntamente com a AUC dos modelos treinados com os dados combinados e validado com o Dados₁₂, ou seja, os dados coletados na Espanha.

Tabela 21 - Resultados obtidos por cada modelo treinado com os dados combinados e validado pelo Dados₁₂.

Modelo	ARTFAETC	KNN	RF	MLP	LR	ET	XGB	NB
<i>ACC</i>	0,9646	0,7169	0,7834	0,7724	0,7547	0,7633	0,7308	0,7331
<i>Se</i>	0,9699	0,5834	0,6813	0,6625	0,6286	0,5609	0,8595	0,6637
<i>Sp</i>	0,9614	0,7991	0,8462	0,8400	0,8323	0,8879	0,6515	0,7759
<i>F₁-Score</i>	0,9543	0,6110	0,7057	0,6893	0,6614	0,6436	0,7087	0,6547
<i>MCC</i>	0,9258	0,3903	0,5355	0,5112	0,4714	0,4838	0,5076	0,4374

Fonte: Elaborado pelo autor.

Figura 22 - Curva ROC e AUC obtidos por cada modelo treinado com os dados combinados e validado pelo Dados₁₂.



Fonte: Elaborado pelo autor.

Pode-se perceber que mesmo perdendo um pouco do desempenho quando comparado com [Tabela 20](#) e [Figura 21](#), a rede ART *Fuzzy* AE TC ainda supera os modelos tradicionais, modelos estes que também reduziram o desempenho com os novos dados.

Ao submeter a rede ART *Fuzzy* AE TC treinada utilizando os dados combinados ao conjunto de validação Dados₁, houve uma redução no desempenho da rede, modelos tradicionais

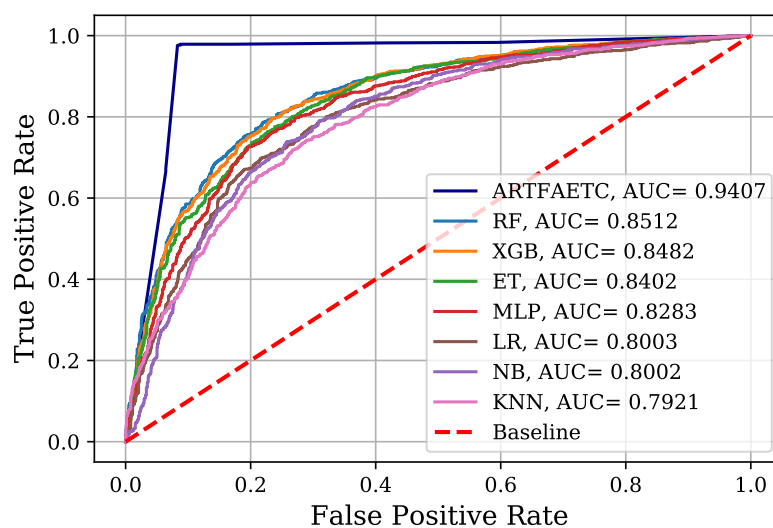
por outro lado, ganharam um aumento no desempenho conforme pode ser visto na [Tabela 22](#) e [Figura 23](#). Embora a rede ART tenha sofrido uma diminuição no desempenho, ela ainda permanece como um modelo mais eficaz em distinguir os pacientes com e sem COVID-19, conforme nota-se pela AUC de 0,9407.

Tabela 22 - Resultados obtidos por cada modelo treinado com os dados combinados e validado pelo Dados₁.

Modelo	ARTFAETC	KNN	RF	MLP	LR	ET	XGB	NB
<i>ACC</i>	0,9387	0,7326	0,7863	0,7699	0,7539	0,7654	0,7326	0,7370
<i>Se</i>	0,9710	0,6388	0,7168	0,6901	0,6589	0,6042	0,8796	0,7101
<i>Sp</i>	0,9173	0,7947	0,8323	0,8227	0,8168	0,8722	0,6352	0,7548
<i>F₁-Score</i>	0,9266	0,6556	0,7278	0,7050	0,6809	0,6725	0,7239	0,6827
<i>MCC</i>	0,8769	0,4377	0,5522	0,5168	0,4817	0,5011	0,5090	0,4597

Fonte: Elaborado pelo autor.

Figura 23 - Curva ROC e AUC obtidos por cada modelo treinado com os dados combinados e validado pelo Dados₁.



Fonte: Elaborado pelo autor.

6 CONCLUSÕES

Em várias partes do mundo, os sistemas de saúde foram desafiados pela COVID-19 e pelas limitações dos testes, inclusive o RT-PCR. Embora este seja o teste padrão ouro, ele possui limitações como baixa sensibilidade, necessidade de laboratórios certificados, equipamentos caros e pessoal treinado. Exames de sangue podem ser uma alternativa auxiliar na detecção correta dos pacientes infectados, visto que alterações laboratoriais podem ser observadas em alguns exames.

Considerando as limitações existentes, este trabalho teve como objetivo avaliar o desempenho de diferentes redes neurais baseadas na Teoria da Ressonância Adaptativa, como também, desenvolver uma metodologia capaz de classificar com precisão a COVID-19 utilizando exames laboratoriais.

Sete redes baseadas na Teoria da Ressonância Adaptativa foram treinadas e validadas com dados de acesso aberto de diferentes países. As redes utilizadas possuem vantagens sobre os modelos tradicionais que sofrem de aprendizagem lenta, sensibilidade a *outliers*, treinamento *off-line* e esquecimento catastrófico. A rede ART *Fuzzy* Autoexpansível com Treinamento Contínuo obteve desempenho superior as outras redes ART implementadas. Além disso, em alguns casos, as redes ART superaram os modelos tradicionais e também algoritmos de aprendizado profundo como DNN e CNN frequentemente utilizadas pela literatura, foram superados tanto em tempo computacional quanto em métricas como acurácia e coeficiente de correlação de Matthews. Mesmo quando submetida a dados externos da Espanha e Itália ela manteve o bom desempenho superando algoritmos como *random forest*, regressão logística e *Multilayer Perceptron*.

Características importantes como a autoexpansão, que torna a rede mais rápida e a capacidade de aprender continuamente sem a necessidade de reinicializar o treinamento tornam esta rede uma ferramenta poderosa para o problema em questão, com convergência rápida e aprendizado em tempo real, além de ser resistente aos problemas listados anteriormente. Desta forma, a metodologia proposta pode ser utilizada como um meio auxiliar para ajudar na identificação de pacientes minimizando a taxa de falsos positivos ou negativos do RT-PCR, bem como uma alternativa de baixo custo para detectar possíveis pacientes com COVID-19 em países subdesenvolvidos, ou naqueles que carecem de reagentes e laboratórios especializados.

6.1 SUGESTÕES PARA TRABALHOS FUTUROS

Para trabalhos futuros algumas sugestões podem ser feitas, como:

- Aplicar a validação cruzada nos dados combinados de diferentes países como forma de validar e avaliar a capacidade de generalização do modelo;
- Verificar a combinação de outros atributos;
- Testar o modelo treinado com mais dados, como forma de verificar o treinamento de modo contínuo;
- Modificar a normalização, ou até mesmo implementar a rede utilizando outras métricas;
- Combinar diferentes redes ART, de modo que cada modelo possa detectar características importantes e contribuir para a classificação final.

REFERÊNCIAS

- AGNELLO, L. et al. The value of a complete blood count (cbc) for sepsis diagnosis and prognosis. **Diagnostics**, Basel, CH, v. 11, n. 10, 2021. DOI:<<https://doi.org/10.3390/diagnostics11101881>>. Citado na página 50.
- AHSAN, M. M.; LUNA, S. A.; SIDDIQUE, Z. Machine-learning-based disease diagnosis: A comprehensive review. **Healthcare**, Basel, CH, v. 10, n. 3, p. 541, 2022. DOI:<<https://doi.org/10.3390/healthcare10030541>>. Citado na página 17.
- AI, T. et al. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. **Radiology**, Oak Brook, US, 2020. DOI:<<https://doi.org/10.1148/radiol.2020200642>>. Citado na página 16.
- ALAKUS, T. B.; TURKOGLU, I. Comparison of deep learning approaches to predict covid-19 infection. **Chaos, Solitons & Fractals**, Oxford, UK, v. 140, p. 110120, 2020. DOI:<<https://doi.org/10.1016/j.chaos.2020.110120>>. Citado 6 vezes nas páginas 21, 24, 43, 44, 68 e 69.
- ALJAME, M.; AHMAD, I.; IMTIAZ, A.; MOHAMMED, A. Ensemble learning model for diagnosing covid-19 from routine blood tests. **Informatics in Medicine Unlocked**, London, UK, v. 21, p. 100449, 2020. DOI:<<https://doi.org/10.1016/j.imu.2020.100449>>. Citado 2 vezes nas páginas 20 e 23.
- ALVES, M. A. et al. Explaining machine learning based diagnosis of covid-19 from routine blood tests with decision trees and criteria graphs. **Computers in Biology and Medicine**, Oxford, UK, v. 132, p. 104335, 2021. DOI:<<https://doi.org/10.1016/j.combiomed.2021.104335>>. Citado 2 vezes nas páginas 22 e 24.
- BARBOSA, L. D. O uso de ventiladores na pandemia do covid-19. **InterAmerican Journal of Medicine and Health**, Campinas, BR, v. 3, 2020. DOI:<<https://doi.org/10.31005/iajmh.v3i0.141>>. Citado na página 16.
- BARROS, A. C. **Detecção e Classificação de Distúrbios de Tensão Usando Redes Neurais ARTMAP Euclidiana Modificada com Treinamento Continuado**. 2013. 109 f. Tese (Doutorado em Engenharia Elétrica) — Departamento de Engenharia Elétrica, Universidade Estadual Paulista, Ilha Solteira, BR, 2013. Citado 2 vezes nas páginas 32 e 41.
- BATISTA, A. F. d. M.; MIRAGLIA, J. L.; DONATO, T. H. R.; FILHO, A. D. P. C. Covid-19 diagnosis prediction in emergency care patients: a machine learning approach. **MedRxiv**, [S.l.], 2020. DOI:<<https://doi.org/10.1101/2020.04.04.20052092>>. Citado 2 vezes nas páginas 19 e 23.
- BAYES, T.; PRICE, n. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s.

Philosophical Transactions of the Royal Society of London, London, UK, v. 53, p. 370–418, 1763. DOI:<<https://doi.org/10.1098/rstl.1763.0053>>. Citado na página 70.

BLOCK, P. et al. Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world. **Nature Human Behaviour**, London, UK, v. 4, n. 6, p. 588–596, 2020. DOI:<<https://doi.org/10.1038/s41562-020-0898-6>>. Citado na página 15.

BOLDRINI, J. L.; COSTA, S. I. R.; FIGUEIREDO, V. L.; WETZLER, H. G. **Álgebra Linear**. 3. ed. São Paulo, BR: Editora Harbra, 1980. Citado na página 48.

BORRA, S.; CIACCIO, A. D. Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods. **Computational statistics & data analysis**, Amsterdam, NL, v. 54, n. 12, p. 2976–2989, 2010. DOI:<<https://doi.org/10.1016/j.csda.2010.03.004>>. Citado na página 54.

BOUGHORBEL, S.; JARRAY, F.; EL-ANBARI, M. Optimal classifier for imbalanced data using matthews correlation coefficient metric. **PloS one**, San Francisco, US, v. 12, n. 6, p. e0177678, 2017. DOI:<<https://doi.org/10.1371/journal.pone.0177678>>. Citado 3 vezes nas páginas 52, 53 e 60.

BRINATI, D. et al. Detection of covid-19 infection from routine blood exams with machine learning: a feasibility study. **Journal of medical systems**, New York, US, v. 44, n. 8, p. 1–12, 2020. DOI:<<https://doi.org/10.1007/s10916-020-01597-4>>. Citado 6 vezes nas páginas 16, 21, 23, 43, 44 e 67.

CABITZA, F. et al. Development, evaluation, and validation of machine learning models for covid-19 detection based on routine blood tests. **Clinical Chemistry and Laboratory Medicine (CCLM)**, Berlin, DE, v. 59, n. 2, p. 421–431, 2021. DOI:<<https://doi.org/10.1515/cclm-2020-1294>>. Citado 5 vezes nas páginas 22, 24, 43, 44 e 67.

CABITZA, F. et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. **Computer Methods and Programs in Biomedicine**, Shannon, IE, v. 208, p. 106288, 2021. DOI:<<https://doi.org/10.1016/j.cmpb.2021.106288>>. Citado na página 43.

CARELLI, G. Z. et al. Laboratory markers in patients with covid-19. **Research, Society and Development**, Itabira, BR, v. 9, n. 12, p. e30191211115–e30191211115, 2020. DOI:<<https://doi.org/10.33448/rsd-v9i12.11115>>. Citado na página 17.

CARPENTER, G. A.; GROSSBERG, S. Art 2: Self-organization of stable category recognition codes for analog input patterns. **Applied optics**, Washington, US, v. 26, n. 23, p. 4919–4930, 1987. DOI:<<https://doi.org/10.1364/AO.26.004919>>. Citado na página 25.

CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. **Computer vision, graphics, and image processing**, Maryland Heights, US, v. 37, n. 1, p. 54–115, 1987. DOI:<[https://doi.org/10.1016/S0734-189X\(87\)80014-2](https://doi.org/10.1016/S0734-189X(87)80014-2)>. Citado na página 25.

CARPENTER, G. A. et al. Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps. **IEEE Transactions on neural**

networks, Piscataway, US, v. 3, n. 5, p. 698–713, 1992. DOI:<<https://doi.org/10.1109/72.159059>>. Citado 5 vezes nas páginas 25, 26, 33, 34 e 35.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. **Neural Networks**, Oxford, UK, v. 4, n. 5, p. 565–588, 1991. DOI:<[https://doi.org/10.1016/0893-6080\(91\)90012-T](https://doi.org/10.1016/0893-6080(91)90012-T)>. Citado na página 25.

CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. **Neural Networks**, Oxford, UK, v. 4, n. 6, p. 759–771, 1991. DOI:<[https://doi.org/10.1016/0893-6080\(91\)90056-B](https://doi.org/10.1016/0893-6080(91)90056-B)>. Citado 3 vezes nas páginas 25, 26 e 27.

Centros de Controle e Prevenção de Doenças. **CDC Museum COVID-19 Timeline**. 2022. Disponível em: <<https://www.cdc.gov/museum/timeline/covid19.html#print>>. Acesso em: 6 dez. 2022. Citado na página 15.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: KDD '16: THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, San Francisco, US. **Anais eletrônicos...** New York, US: Association for Computing Machinery, 2016. p. 785–794. ISBN 9781450342322. DOI:<<https://doi.org/10.1145/2939672.2939785>>. Citado na página 70.

CHUNG, M. et al. Ct imaging features of 2019 novel coronavirus (2019-ncov). **Radiology**, Oak Brook, US, v. 295, n. 1, p. 202–207, 2020. DOI:<<https://doi.org/10.1148/radiol.2020200230>>. Citado na página 16.

CORMAN, V. et al. Diagnostic detection of wuhan coronavirus 2019 by real-time rt-pcr. **Geneva: World Health Organization**, Berlin, DE, v. 13, 2020. Citado na página 16.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, Piscataway, US, v. 13, n. 1, p. 21–27, 1967. DOI:<<https://doi.org/10.1109/TIT.1967.1053964>>. Citado na página 70.

CRANEY, T. A.; SURLES, J. G. Model-dependent variance inflation factor cutoff values. **Quality Engineering**, New York, US, v. 14, n. 3, p. 391–403, 2002. DOI:<<https://doi.org/10.1081/QEN-120001878>>. Citado na página 48.

DONG, E.; DU, H.; GARDNER, L. An interactive web-based dashboard to track covid-19 in real time. **The Lancet infectious diseases**, London, UK, v. 20, n. 5, p. 533–534, 2020. DOI:<[https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)>. Citado na página 15.

DUDOIT, S.; LAAN, M. J. van der. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. **Statistical methodology**, Amsterdam, NL, v. 2, n. 2, p. 131–154, 2005. DOI:<<https://doi.org/10.1016/j.stamet.2005.02.003>>. Citado na página 54.

ELDERDERY, A. Y. et al. Complete blood count alterations of covid-19 patients in riyadh, kingdom of saudi arabia. **Journal of Nanomaterials**, London, UK, v. 2022, 2022. DOI:<<https://doi.org/10.1155/2022/6529641>>. Citado na página 17.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine learning**,

New York, US, v. 63, p. 3–42, 2006. DOI:<<https://doi.org/10.1007/s10994-006-6226-1>>. Citado na página 70.

GODLEE, F. Protect our healthcare workers. **BMJ**, London, UK, v. 369, 2020. DOI:<<https://doi.org/10.1136/bmj.m1324>>. Citado na página 15.

GOODMAN-MEZA, D. et al. A machine learning algorithm to increase covid-19 inpatient diagnostic capacity. **Plos one**, San Francisco, US, v. 15, n. 9, p. e0239474, 2020. DOI:<<https://doi.org/10.1371/journal.pone.0239474>>. Citado 2 vezes nas páginas 20 e 23.

GOROG, D. A. et al. Current and novel biomarkers of thrombotic risk in covid-19: a consensus statement from the international covid-19 thrombosis biomarkers colloquium. **Nature Reviews Cardiology**, London, UK, p. 1–21, 2022. DOI:<<https://doi.org/10.1038/s41569-021-00665-7>>. Citado na página 22.

GROSSBERG, S. Adaptive pattern classification and universal recoding: Ii. feedback, expectation, olfaction, illusions. **Biological cybernetics**, Heidelberg, DE, v. 23, n. 4, p. 187–202, 1976. DOI:<<https://doi.org/10.1007/BF00340335>>. Citado na página 25.

GROSSBERG, S. How does a brain build a cognitive code? In: _____. **Studies of Mind and Brain: Neural principles of learning, perception, development, cognition, and motor control**. Boston, US: Springer, 1982. cap. 1, p. 1–52. DOI:<https://doi.org/10.1007/978-94-009-7758-7_1>. Citado na página 25.

GROSSBERG, S. Competitive learning: From interactive activation to adaptive resonance. **Cognitive science**, Hoboken, US, v. 11, n. 1, p. 23–63, 1987. DOI:<[https://doi.org/10.1016/S0364-0213\(87\)80025-3](https://doi.org/10.1016/S0364-0213(87)80025-3)>. Citado na página 25.

GROSSBERG, S. A path toward explainable ai and autonomous adaptive intelligence: deep learning, adaptive resonance, and models of perception, emotion, and action. **Front Neurorobot**, Lausanne, CH, v. 14, p. 36, 2020. DOI:<<https://doi.org/10.3389/fnbot.2020.00036>>. Citado 5 vezes nas páginas 22, 25, 26, 33 e 70.

HO, T. K. Random decision forests. In: PROCEEDINGS OF 3RD INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 3., 1995, Montreal, CA. **Anais eletrônicos...** Montreal, CA: IEEE, 1995. v. 1, p. 278–282. DOI:<<https://doi.org/10.1109/ICDAR.1995.598994>>. Citado na página 70.

HUANG, C. et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. **The lancet**, London, UK, v. 395, n. 10223, p. 497–506, 2020. DOI:<[https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)>. Citado na página 17.

HUANG, J.; LING, C. Using auc and accuracy in evaluating learning algorithms. **IEEE Transactions on Knowledge and Data Engineering**, Piscataway, US, v. 17, n. 3, p. 299–310, 2005. DOI:<<https://doi.org/10.1109/TKDE.2005.50>>. Citado 2 vezes nas páginas 52 e 60.

HUSSAIN, L. et al. Machine-learning classification of texture features of portable chest x-ray accurately classifies covid-19 lung infection. **BioMedical Engineering OnLine**, London, UK, v. 19, n. 1, p. 1–18, 2020. DOI:<<https://doi.org/10.1186/s12938-020-00831-x>>. Citado na página 17.

- Jl, D. et al. Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. **Clinical Infectious Diseases**, Cary, US, v. 71, n. 6, p. 1393–1399, 04 2020. DOI:<<https://doi.org/10.1093/cid/ciaa414>>. Citado na página 22.
- JOHNSON, K. J.; SYNOVEC, R. E. Pattern recognition of jet fuels: comprehensive gc×gc with anova-based feature selection and principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, NL, v. 60, n. 1, p. 225–237, 2002. DOI:<[https://doi.org/10.1016/S0169-7439\(01\)00198-8](https://doi.org/10.1016/S0169-7439(01)00198-8)>. Citado na página 47.
- JOSHI, R. P. et al. A predictive tool for identification of sars-cov-2 pcr-negative emergency department patients using routine test results. **Journal of Clinical Virology**, Amsterdam, NL, v. 129, p. 104502, 2020. DOI:<<https://doi.org/10.1016/j.jcv.2020.104502>>. Citado 2 vezes nas páginas 20 e 23.
- KAUFMANN, J.; SCHERING, A. Analysis of variance anova. In: BALAKRISHNAN, N. et al. (Ed.). **Wiley StatsRef: Statistics Reference Online**. Hoboken, US: John Wiley Sons, Ltd, 2014. DOI:<<https://doi.org/10.1002/9781118445112.stat06938>>. Citado na página 47.
- KENAYA, R.; CHEOK, K. C. Euclidean art neural networks. In: WORLD CONGRESS ON ENGINEERING AND COMPUTER SCIENCE 2008, 2., 2008, San Francisco, US. **Anais eletrônicos...** San Francisco, US: IAENG, 2008. p. 963–968. Citado na página 30.
- KERMALI, M. et al. The role of biomarkers in diagnosis of covid-19—a systematic review. **Life sciences**, Philadelphia, US, v. 254, p. 117788, 2020. DOI:<<https://doi.org/10.1016/j.lfs.2020.117788>>. Citado 2 vezes nas páginas 22 e 50.
- KEYKAVOUSHI, K. et al. A review of routine laboratory biomarkers for the detection of severe covid-19 disease. **International journal of analytical chemistry**, London, UK, v. 2022, 2022. DOI:<<https://doi.org/10.1155/2022/9006487>>. Citado 2 vezes nas páginas 22 e 50.
- KHARTABIL, T.; RUSSCHER, H.; VEN, A. van der; RIJKE, Y. D. A summary of the diagnostic and prognostic value of hemocytometry markers in covid-19 patients. **Critical reviews in clinical laboratory sciences**, Oxfordshire, UK, v. 57, n. 6, p. 415–431, 2020. DOI:<<https://doi.org/10.1080/10408363.2020.1774736>>. Citado 2 vezes nas páginas 17 e 45.
- KHUZANI, A. Z.; HEIDARI, M.; SHARIATI, S. A. Covid-classifier: An automated machine learning model to assist in the diagnosis of covid-19 infection in chest x-ray images. **Scientific Reports**, London, UK, v. 11, n. 1, p. 1–6, 2021. DOI:<<https://doi.org/10.1038/s41598-021-88807-2>>. Citado na página 17.
- KISTENEV, Y. V.; VRAZHNOV, D. A.; SHNAIDER, E. E.; ZUHAYRI, H. Predictive models for covid-19 detection using routine blood tests and machine learning. **Heliyon**, London, UK, p. e11185, 2022. DOI:<<https://doi.org/10.1016/j.heliyon.2022.e11185>>. Citado na página 16.
- KOVÁCS, A. et al. The sensitivity and specificity of chest ct in the diagnosis of covid-19. **European radiology**, Heidelberg, DE, v. 31, n. 5, p. 2819–2824, 2021. DOI:<<https://doi.org/10.1007/s00330-020-07347-x>>. Citado na página 17.
- KUKAR, M. et al. Covid-19 diagnosis by routine blood tests using machine learning. **Scientific reports**, London, US, v. 11, n. 1, p. 1–9, 2021. DOI:<<https://doi.org/10.1038/s41598-021-90265-9>>. Citado 2 vezes nas páginas 21 e 24.

- LI, Z. et al. Development and clinical application of a rapid igm-igg combined antibody test for sars-cov-2 infection diagnosis. **Journal of Medical Virology**, Hoboken, US, v. 92, n. 9, p. 1518–1524, 2020. DOI:<<https://doi.org/10.1002/jmv.25727>>. Citado na página 16.
- LIU, Y. et al. Clinical and biochemical indexes from 2019-ncov infected patients linked to viral loads and lung injury. **Science China Life Sciences**, Beijing, CN, v. 63, n. 3, p. 364–374, 2020. DOI:<<https://doi.org/10.1007/s11427-020-1643-8>>. Citado na página 17.
- LUGADA, E. S. et al. Population-based hematologic and immunologic reference values for a healthy ugandan population. **Clinical and Vaccine Immunology**, Washington, US, v. 11, n. 1, p. 29–34, 2004. DOI:<<https://doi.org/10.1128/CDLI.11.1.29-34.2004>>. Citado na página 46.
- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 30., 2017, Long Beach, US. **Anais eletrônicos...** Long Beach, US: Curran Associates, Inc., 2017. p. 4768–4777. Disponível em: <<https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>>. Acesso em: 05 nov. 2022. Citado 2 vezes nas páginas 47 e 49.
- MANDREKAR, J. N. Receiver operating characteristic curve in diagnostic test assessment. **Journal of Thoracic Oncology**, New York, US, v. 5, n. 9, p. 1315–1316, 2010. DOI:<<https://doi.org/10.1097/JTO.0b013e3181ec173d>>. Citado 5 vezes nas páginas 53, 57, 59, 61 e 63.
- MARAR, R. W.; MARAR, H. W. A 15-gbps bicos xnor gate for fast recognition of covid-19 in binarized neural networks. **International Journal of Electrical and Computer Engineering**, Bantul, ID, v. 12, n. 1, p. 997, 2022. DOI:<<https://doi.org/10.11591/ijece.v12i1.pp997-1002>>. Citado na página 17.
- MARCHIORI, S. C. et al. Neural network based on adaptive resonance theory with continuous training for multi-configuration transient stability analysis of electric power systems. **Applied Soft Computing**, Amsterdam, NL, v. 11, n. 1, p. 706–715, 2011. ISSN 1568-4946. DOI:<<https://doi.org/10.1016/j.asoc.2009.12.032>>. Citado na página 41.
- MARTINS, C. et al. Dificuldades e desafios enfrentados pela equipe de enfermagem frente ao cenário da pandemia da covid-19. **Research, Society and Development**, Itabira, BR, v. 11, n. 6, p. e4311627150–e4311627150, 2022. DOI:<<https://doi.org/10.33448/rsd-v11i6.27150>>. Citado na página 15.
- MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. **Biochimica et Biophysica Acta (BBA)-Protein Structure**, Amsterdam, NL, v. 405, n. 2, p. 442–451, 1975. DOI:<[https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)>. Citado na página 52.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, New York, US, v. 5, p. 115–133, 1943. DOI:<<https://doi.org/10.1007/BF02478259>>. Citado na página 70.
- MIRZAEI, G.; ADELI, H. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. **Biomedical Signal**

- Processing and Control**, Amsterdam, NL, v. 72, p. 103293, 2022. DOI:<<https://doi.org/10.1016/j.bspc.2021.103293>>. Citado na página 17.
- MISHRA, N. P. et al. Global impacts of pre-and post-covid-19 pandemic: Focus on socio-economic consequences. **Sensors International**, Dongcheng, CN, v. 1, p. 100042, 2020. DOI:<<https://doi.org/10.1016/j.sintl.2020.100042>>. Citado na página 15.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. New York: John Wiley & Sons, 2006. 612 p. Citado na página 48.
- MORENO, A. L. **Análise da Estabilidade Transitória via Rede Neural ART-ARTMAP Fuzzy Euclidiana Modificada com Treinamento Continuado**. 2010. 144 f. Tese (Doutorado em Engenharia Elétrica) — Departamento de Engenharia Elétrica, Universidade Estadual Paulista, Ilha Solteira, BR, 2010. Citado 9 vezes nas páginas 26, 28, 29, 30, 36, 37, 38, 39 e 40.
- MORENO, A. L. **Redes Neurais ART e ARTMAP com Treinamento Continuado: Uma apresentação detalhada dos principais tipos de redes neurais do tipo ART e ARTMAP**. Saarbrücken, DE: Novas Edicoes Academicas, 2016. Citado 2 vezes nas páginas 26 e 38.
- ODHIAMBO, C. et al. Evaluation of locally established reference intervals for hematology and biochemistry parameters in western kenya. **Plos one**, San Francisco, US, v. 10, n. 4, p. e0123140, 2015. DOI:<<https://doi.org/10.1371/journal.pone.0123140>>. Citado na página 46.
- Organização Mundial da Saúde. **Coronavirus disease 2019 (covid-19) situation report 92**. 2020. Disponível em: <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>>. Acesso em: 05 mai. 2022. Citado na página 15.
- Organização Mundial da Saúde. **Coronavirus disease (covid-19) dashboard**. 2020. Disponível em: <<https://covid19.who.int/>>. Acesso em: 20 nov. 2022. Citado na página 15.
- PALLADINO, M. Complete blood count alterations in covid-19 patients: A narrative review. **Biochemia medica**, Zagreb, HR, v. 31, n. 3, p. 0–0, 2021. DOI:<<https://doi.org/10.11613/BM.2021.030501>>. Citado na página 17.
- PAN, F. et al. Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. **Radiology**, Oak Brook, US, 2020. DOI:<<https://doi.org/10.1148/radiol.2020200370>>. Citado na página 17.
- PARIKH, R. et al. Understanding and using sensitivity, specificity and predictive values. **Indian journal of ophthalmology**, Mumbai, IN, v. 56, n. 1, p. 45, 2008. DOI:<<https://doi.org/10.4103/0301-4738.37595>>. Citado 2 vezes nas páginas 52 e 60.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, Cambridge, US, v. 12, p. 2825–2830, 2011. Citado na página 70.
- PRADHAN, A. et al. Supervised learning models for the preliminary detection of covid-19 in patients using demographic and epidemiological parameters. **Information**, Basel, CH, v. 13, n. 7, p. 330, 2022. DOI:<<https://doi.org/10.3390/info13070330>>. Citado na página 16.
- RAHMAN, T. et al. Qcovsml: A reliable covid-19 detection system using cbc biomarkers by a

stacking machine learning model. **Computers in Biology and Medicine**, Oxford, UK, v. 143, p. 105284, 2022. DOI:<<https://doi.org/10.1016/j.combiomed.2022.105284>>. Citado 2 vezes nas páginas 22 e 24.

RAPTIS, C. A. et al. Chest ct and coronavirus disease (covid-19): a critical review of the literature to date. **AJR Am J Roentgenol**, Leesburg, US, v. 215, n. 4, p. 839–842, 2020. DOI:<<https://doi.org/10.2214/AJR.20.23202>>. Citado na página 17.

REIS, A. A. d. S.; SANTOS, R. d. S. O padrão ouro no diagnóstico molecular na covid19: O que sabemos sobre a soberania destemétodo? **Brazilian Journal of Health Review**, Curitiba, BR, v. 3, n. 3, p. 5986–5992, 2020. DOI:<<https://doi.org/10.34119/bjhrv3n3-157>>. Citado na página 16.

RIKAN, S. B. et al. Covid-19 diagnosis from routine blood tests using artificial intelligence techniques. **Biomedical Signal Processing and Control**, Amsterdam, NL, v. 72, p. 103263, 2022. DOI:<<https://doi.org/10.1016/j.bspc.2021.103263>>. Citado 4 vezes nas páginas 62, 67, 68 e 69.

RIQUELME, D.; AKHLOUFI, M. A. Deep learning for lung cancer nodules detection and classification in ct scans. **Ai**, Basel, CH, v. 1, n. 1, p. 28–67, 2020. DOI:<<https://doi.org/10.3390/ai1010003>>. Citado na página 16.

ROSENFELD, L. G. et al. Valores de referência para exames laboratoriais de hemograma da população adulta brasileira: Pesquisa nacional de saúde. **Revista Brasileira de Epidemiologia**, Rio de Janeiro, BR, v. 22, 2019. DOI:<<https://doi.org/10.1590/1980-549720190003.supl.2>>. Citado na página 46.

SALATHÉ, M. et al. Covid-19 epidemic in switzerland: on the importance of testing, contact tracing and isolation. **Swiss medical weekly**, Muttentz, CH, v. 150, n. 1112, 2020. DOI:<<https://doi.org/10.4414/smw.2020.20225>>. Citado na página 16.

SALISBURY, H. Helen salisbury: Fear in the time of covid. **British Medical Journal Publishing Group**, London, UK, v. 368, 2020. DOI:<<https://doi.org/10.1136/bmj.m1286>>. Citado na página 16.

SAMPRATHI, M.; JAYASHREE, M. Biomarkers in covid-19: an up-to-date review. **Frontiers in pediatrics**, Lausanne, CH, v. 8, p. 607647, 2021. DOI:<<https://doi.org/10.3389/fped.2020.607647>>. Citado 2 vezes nas páginas 22 e 50.

SCILAB, E. **Scilab (version 6.1.1)**. 2012. Disponível em: <<https://www.scilab.org>>. Citado na página 56.

SHRESTHA, L. B. et al. Evolution of the sars-cov-2 omicron variants ba. 1 to ba. 5: Implications for immune escape and transmission. **Reviews in Medical Virology**, Oxford, UK, v. 32, n. 5, p. e2381, 2022. DOI:<<https://doi.org/10.1002/rmv.2381>>. Citado na página 16.

SILVA, G. F. R. et al. Comparison of the biochemical and hematological profiles of patients with covid-19 in their different outcomes: Hospital discharge and death. **Research, Society and Development**, Itabira, BR, v. 11, n. 2, p. e25611225772–e25611225772, 2022. DOI:<<https://doi.org/10.33448/rsd-v11i2.25772>>. Citado na página 17.

- SOLTAN, A. A. et al. Rapid triage for covid-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. **The Lancet Digital Health**, London, UK, v. 3, n. 2, p. e78–e87, 2021. DOI:<[https://doi.org/10.1016/S2589-7500\(20\)30274-0](https://doi.org/10.1016/S2589-7500(20)30274-0)>. Citado 2 vezes nas páginas 21 e 24.
- SUN, D.-w. et al. The underlying changes and predicting role of peripheral blood inflammatory cells in severe covid-19 patients: A sentinel? **Clinica chimica acta**, Amsterdam, NL, v. 508, p. 122–129, 2020. DOI:<<https://doi.org/https://doi.org/10.1016/j.cca.2020.05.027>>. Citado na página 17.
- SUN, L. et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected covid-19. **Journal of Clinical Virology**, Amsterdam, NL, v. 128, p. 104431, 2020. DOI:<<https://doi.org/10.1016/j.jcv.2020.104431>>. Citado 3 vezes nas páginas 17, 19 e 23.
- TAHAMTAN, A.; ARDEBILI, A. Real-time rt-pcr in covid-19 detection: issues affecting the results. **Expert review of molecular diagnostics**, Oxfordshire, UK, v. 20, n. 5, p. 453–454, 2020. DOI:<<https://doi.org/10.1080/14737159.2020.1757437>>. Citado na página 16.
- TANNE, J. H. et al. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. **Bmj**, London, UK, v. 368, 2020. DOI:<<https://doi.org/10.1136/bmj.m1090>>. Citado na página 15.
- TORDJMAN, M. et al. Pre-test probability for sars-cov-2-related infection score: The paris score. **PLoS One**, San Francisco, US, v. 15, n. 12, p. e0243342, 2020. DOI:<<https://doi.org/10.1371/journal.pone.0243342>>. Citado 2 vezes nas páginas 19 e 23.
- TSCHOELLITSCH, T. et al. Machine learning prediction of sars-cov-2 polymerase chain reaction results with routine blood tests. **Laboratory medicine**, Cary, US, v. 52, n. 2, p. 146–149, 2021. DOI:<<https://doi.org/10.1093/labmed/lmaa111>>. Citado 2 vezes nas páginas 21 e 24.
- VUSKOVIC, M.; DU, S. Classification of prehensile emg patterns with simplified fuzzy artmap networks. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 3., 2002, Honolulu, US. **Anais eletrônicos...** Honolulu, US: IEEE, 2002. p. 2539–2544. DOI:<<https://doi.org/10.1109/IJCNN.2002.1007543>>. Citado na página 30.
- WANG, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. **Jama**, Chicago, US, v. 323, n. 11, p. 1061–1069, 2020. DOI:<<https://doi.org/10.1001/jama.2020.1585>>. Citado na página 17.
- WEISSLEDER, R.; LEE, H.; KO, J.; PITTET, M. J. Covid-19 diagnostics in context. **Science translational medicine**, New York, US, v. 12, n. 546, p. eabc1931, 2020. DOI:<<https://doi.org/10.1126/scitranslmed.abc1931>>. Citado na página 16.
- WERBOS, P. J. **Análise da Estabilidade Transitória via Rede Neural ART-ARTMAP Fuzzy Euclidiana Modificada com Treinamento Continuado**. 1974. 906 f. Tese (Doutorado) — Universidade Harvard, Cambridge, US, 1974. Citado na página 70.
- WILSON, E. B.; WORCESTER, J. The determination of l.d.50 and its sampling error in bio-assay. In: PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF

THE UNITED STATES OF AMERICA, 2., 1943, Washington, US. **Anais eletrônicos...** Washington, US: National Academy of Sciences, 1943. v. 29, n. 2, p. 79–85. DOI:<<https://doi.org/10.1073/pnas.29.2.79>>. Citado na página 70.

WU, J. et al. Rapid and accurate identification of covid-19 infection through machine learning based on clinical available blood test results. **MedRxiv**, [S.l.], 2020. DOI:<<https://doi.org/10.1101/2020.04.02.20051136>>. Citado 3 vezes nas páginas 17, 19 e 23.

YANG, H. S. et al. Routine laboratory blood tests predict sars-cov-2 infection using machine learning. **Clinical chemistry**, Berlin, DE, v. 66, n. 11, p. 1396–1404, 2020. DOI:<<https://doi.org/10.1093/clinchem/hvaa200>>. Citado 2 vezes nas páginas 20 e 23.

ZADEH, L. A. Fuzzy sets. **Information and control**, Maryland Heights, US, v. 8, n. 3, p. 338–353, 1965. DOI:<[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)>. Citado na página 26.

ZHANG, E.; ZHANG, Y. F-measure. In: LIU, LING AND ÖZSU, M. TAMER. **Encyclopedia of Database Systems**. Boston, US: Springer, 2009. p. 1147–1147. DOI:<https://doi.org/10.1007/978-0-387-39940-9_483>. Citado 2 vezes nas páginas 52 e 53.

ZHANG, H. The optimality of naive bayes. In: PROCEEDINGS OF THE SEVENTEENTH INTERNATIONAL FLORIDA ARTIFICIAL INTELLIGENCE RESEARCH SOCIETY CONFERENCE, 17., 2004, Miami Beach, US. **Anais eletrônicos...** Palo Alto, US: AAAI Press, 2004. p. 562–567. Citado na página 70.

ZHANG, L. et al. Fast screening and primary diagnosis of covid-19 by atr-ft-ir. **Analytical chemistry**, Washington, US, v. 93, n. 4, p. 2191–2199, 2021. DOI:<<https://doi.org/10.1021/acs.analchem.0c04049>>. Citado na página 16.