

UMA ESTRUTURA DE APRENDIZADO PROFUNDO COM MÚLTIPLOS FILTROS E TRANSFERÊNCIA PARA DETECÇÃO DE TRANSTORNO DO ESPECTRO AUTISTA BASEADA EM IMAGENS

Victor José Souza Bernardino; Profa. Dra. Luciene Cavalcanti Rodrigues

e-mail:

victor.j.s.bernardino@gmail.com; luciene.rodrigues@fatec.sp.gov.br.

Resumo: O Transtorno do Espectro Autista (TEA) afeta aproximadamente 1% da população mundial e é caracterizado por dificuldades na comunicação social e comportamentos repetitivos ou obsessivos. A detecção precoce do autismo é crucial, pois permite que intervenções terapêuticas sejam iniciadas mais cedo, aumentando significativamente a eficácia dos tratamentos. No entanto, diagnosticar o TEA continua sendo um desafio, pois é tradicionalmente realizado por meio de métodos que são frequentemente subjetivos e baseados em entrevistas e observações clínicas. Com o avanço das técnicas de visão computacional e reconhecimento de padrões, novas possibilidades estão surgindo para automatizar e aprimorar a detecção de características associadas ao TEA, particularmente na análise de traços faciais. Neste estudo, foi proposto um novo *framework* que integra aumento de dados, rotinas de multifiltragem, equalização de histograma e um processo de redução de dimensionalidade em duas etapas para enriquecer a representação em modelos de rede neural profunda pré-treinados e congelados aplicados ao reconhecimento de padrões de imagem. Experimentos foram conduzidos usando oito modelos pré-treinados em um conjunto de dados faciais bem estabelecido na literatura, que compreende amostras de indivíduos autistas e não autistas. Os resultados mostraram que o *framework* proposto melhorou a precisão dos modelos em até 9% em termos absolutos. Esses achados sugerem que o *framework* é uma ferramenta promissora para a detecção automatizada de autismo, oferecendo um aprimoramento eficiente em abordagens tradicionais baseadas em aprendizado profundo para auxiliar no diagnóstico precoce e mais preciso.

Palavras-chave: Detecção automatizada do autismo. Análise de traços faciais com redes neurais. Aprendizado profundo com múltiplos filtros. Transferência de aprendizado em imagens. Framework híbrido de visão computacional.

Abstract: Autism Spectrum Disorder (ASD) affects approximately 1% of the global population and is characterized by difficulties in social communication and repetitive or obsessive behaviors. Early detection of autism is crucial, as it allows therapeutic interventions to be initiated earlier, significantly increasing the effectiveness of treatments. However, diagnosing ASD remains a challenge, as it is traditionally carried out through methods that are often subjective and based on interviews and clinical observations. With the advancement of computer vision and pattern recognition techniques, new possibilities are emerging to automate and enhance the detection of characteristics associated with ASD, particularly in the analysis of facial features. In this study, we propose a novel framework that integrates data augmentation, multi-filtering routines, histogram equalization, and a two-stage dimensionality reduction process to enrich the representation in pre-trained and frozen deep learning neural network models applied to image pattern recognition. Experiments were conducted using eight pre-trained models on a well-established facial dataset in the literature, which comprises samples of both autistic and non-autistic individuals. The results showed that the proposed framework improved the accuracy of the models by up to 9% in absolute terms. These findings suggest that the framework is a promising tool for the automated detection of autism, offering an efficient improvement in traditional deep learning-based approaches to assist in early and more accurate diagnosis.

Keywords: Automated autism detection. Facial trait analysis with neural networks. Deep learning with multiple filters. Transfer learning in imaging. Hybrid computer vision framework.

1 Introdução

Pesquisa sobre a definição e classificação do Transtorno do Espectro Autista (TEA) Lord et al. (2018), também conhecido como autismo, tem atraído considerável atenção de especialistas e recebido significativos investimentos privados e governamentais ao longo das últimas cinco décadas. No entanto, o interesse da humanidade por essa neurodivergência remonta a mais de 500 anos (DONVAN; ZUCKER, 2016). O TEA descreve um amplo grupo de indivíduos que apresentam dificuldades na comunicação social, juntamente com comportamentos atípicos, repetitivos ou obsessivos (LORD et al., 2020). Estima-se que aproximadamente 1% da população global seja afetada por essa condição (BAXTER et al., 2015). A detecção precoce do TEA e o início imediato de um suporte profissional adequado são cruciais para maximizar a eficácia das intervenções e melhorar os resultados a longo prazo (WEITLAUF et al., 2014). No entanto, diagnosticar o Transtorno do Espectro Autista (TEA) não é uma tarefa simples (HARM; HOPE; HOUSEHOLD, 2013). Tradicionalmente, o diagnóstico é realizado por meio de entrevistas detalhadas conduzidas por especialistas, baseadas em protocolos clínicos estabelecidos (JANVIER et al., 2022). Um exemplo amplamente utilizado é a Escala de Avaliação de Autismo na Infância (CARS) (SCHOPLER et al., 1980), que consiste em um conjunto de 15 observações clínicas e comportamentais para avaliar se um indivíduo é autista. Esta escala atribui uma pontuação que varia de 15 a 60, com valores acima de 35 indicando a presença de TEA, e pontuações mais altas refletindo maior gravidade da condição. Várias outras escalas para avaliar a gravidade do autismo também são amplamente reconhecidas na literatura. A Entrevista Diagnóstica de Autismo Revisada (ADI-R) (RUTTER et al., 2003) destaca 183 questões relacionadas à história de desenvolvimento e antecedentes familiares; a Escala de Avaliação de Autismo de Gillian (LECAVALIER, 2005), que avalia 56 itens agrupados em quatro áreas comportamentais: comportamentos estereotipados, comunicação, interação social e distúrbios de desenvolvimento; e a Entrevista Diagnóstica da Síndrome de Asperger (ASDI) (GILLBERG et al., 2001), uma entrevista de 20 minutos focada especificamente na Síndrome de Asperger. O processo de pontuação de qualquer uma dessas escalas depende fortemente da interação humana, seja com profissionais

especializados ou com os cuidadores do indivíduo potencialmente autista, constituindo assim uma forma de classificação manual.

2 Justificativa

Para auxiliar na determinação de um diagnóstico de TEA de maneira mais precisa, menos subjetiva e mais rápida, técnicas baseadas em inteligência artificial e processamento e análise de sinais estão sendo cada vez mais consideradas por profissionais de saúde. Exemplos incluem sinais sonoros (CHO et al., 2019), sinais de eletroencefalografia (EEG) (TAWHID et al., 2021), sinais de ressonância magnética (MRI) (WADHERA; MAHMUD; BROWN, 2023; MARTINO et al., 2017), sinais de vídeo com rastreamento ocular (ATYABI et al., 2023) e outras características (PANDIAN et al., 2022; ALI et al., 2022). Entre todos esses sinais, os baseados em imagens faciais (ALKAHTANI; ALDHYANI; ALZAHRANI, 2023) estão entre os mais considerados devido à facilidade de amostragem, pois a coleta de uma fotografia é rápida e minimamente invasiva para o paciente. Além disso, é bem conhecido que o TEA está potencialmente associado a características faciais (ALDRIDGE et al., 2011; HAMMOND et al., 2008). Vale destacar que a classificação automática dos sinais mencionados é realizada por meio de técnicas de aprendizado de máquina, especialmente aquelas envolvendo aprendizado profundo (SQUIRES et al., 2023), que têm demonstrado um desempenho notável em tarefas de determinação de diagnósticos (CHAN; HADJIISKI; SAMALA, 2020). No entanto, o treinamento de um modelo de aprendizado profundo é computacionalmente caro e geralmente requer conjuntos de dados altamente representativos e, conseqüentemente, grandes, o que pode ser problemático na detecção de autismo através de imagens, dada a escassez de exemplos disponíveis na literatura. Para superar esse desafio, neste trabalho, propomos o uso do conceito de aprendizado profundo por transferência (KORA et al., 2022) para adaptação de domínio e reconhecimento de autismo por meio de imagens faciais.

3 Objetivo

Esses modelos, que foram previamente treinados em grandes conjuntos de dados de imagens, foram ajustados para identificar padrões específicos relacionados ao autismo, aproveitando a capacidade dos modelos pré-treinados de generalizar para novos domínios com menos dados rotulados. Especificamente, propomos uma nova estrutura baseada em aumento de dados, multi-filtragem, equalização de histograma e etapas de redução de dimensionalidade para aprimorar a representação de imagens faciais por esses modelos pré-treinados, aumentando assim a precisão no diagnóstico de TEA por meio de classificadores de aprendizado de máquina, como Máquinas de Vetores de Suporte (SVM) (NOBLE, 2006). Assim, as principais contribuições deste trabalho incluem:

- Um novo *framework* para aprimorar a representação de imagens faciais em modelos pré-treinados a fim de melhorar a detecção de TEA;
- Experimentações envolvendo o aprimoramento de oito modelos de aprendizado profundo pré-treinados para detecção de padrões em imagens e o respectivo desempenho na tarefa de detectar autismo por características faciais.

4 Fundamentação teórica

Numerosos estudos utilizam estruturas pré-treinadas para estabelecer um modelo para a detecção automática do autismo por meio da análise de imagens. Intuitivamente, o conceito de aprendizado por transferência é modelado na capacidade humana de aproveitar o conhecimento adquirido em uma categoria de problemas para resolver outra. Matematicamente, Pan e Yang (2009) definem esse modelamento como a utilização de uma função de classificação f originalmente ajustada em uma amostra X do espaço de características X com uma distribuição de probabilidade $P(X)$ ou seja, ajustada sobre o domínio fonte $D_S = \{X, P(X)\}$ cuja saída reside dentro do conjunto de rótulos Y e constitui a tarefa de classificação $T_S = \{Y, f\}$. Isso é aplicado para resolver outra tarefa alvo T_T sobre um domínio alvo D_T onde $T_S \neq T_T$ ou $D_S \neq D_T$. O objetivo do aprendizado por transferência é construir uma função de classificação f_T para um novo domínio D_T com base

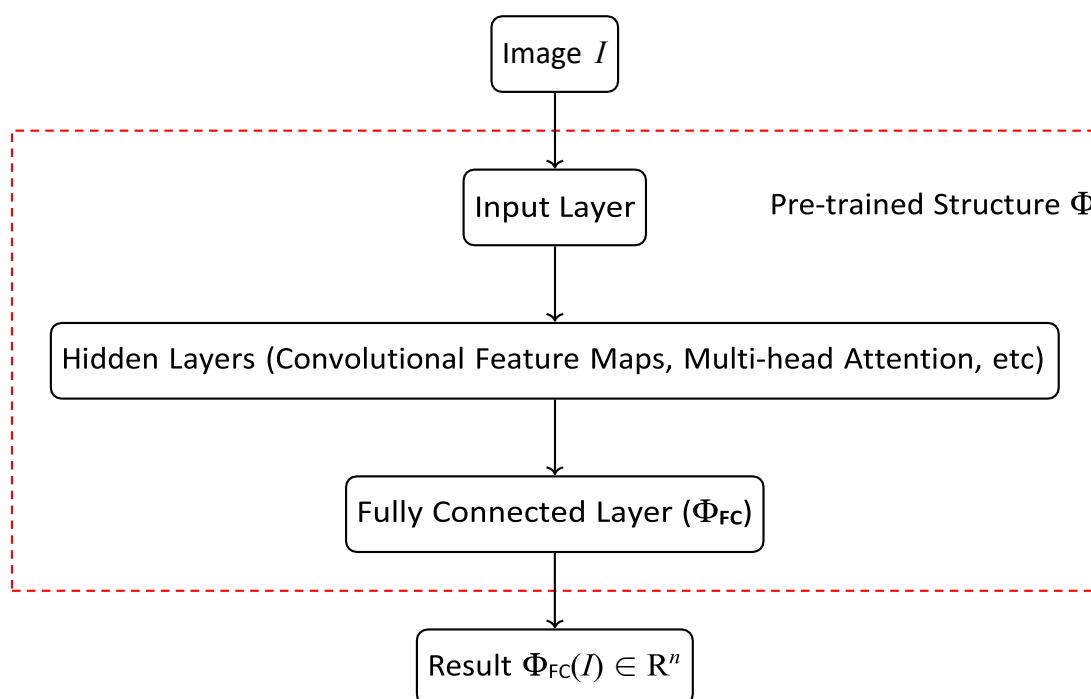
em f . Na prática, esse tipo de modelagem envolve o pré-treinamento de uma rede neural em um conjunto de dados específico para abordar um problema e, em seguida, a utilização de seus pesos para definir outro modelo usando um novo conjunto de dados, que geralmente é menor e menos generalizado do que o original. Essa abordagem é comum em problemas de classificação envolvendo imagens clínicas (KIM et al., 2022). Geralmente, uma rede pré-treinada em conjuntos de dados de imagens consiste em três conjuntos de camadas (KORA et al., 2022): uma camada de entrada dedicada a receber a amostra; um conjunto de camadas de extração de características, que pode ser representado por mapas de características convolucionais em camadas de CNN ou atenção multi-head em redes transformadoras; e, finalmente, uma camada totalmente conectada (FC) correspondente ao número de classes na tarefa de classificação.

Para adaptar a rede pré-treinada a um novo domínio ou tarefa, o processo de ajuste fino (TOO et al., 2019) pode ser empregado. Isso envolve redefinir e re-treinar a camada totalmente conectada (FC) final para ajustar a rede ao novo problema, mantendo as outras camadas inalteradas. Além disso, novas camadas podem ser adicionadas ou re-treinadas dentro da rede original durante esse processo. Alternativamente, a saída da camada FC pode ser usada como uma representação da amostra para treinar outros tipos de classificadores, como uma Máquina de Vetores de Suporte (SVM) (DAWUD; YURTKAN; OZTOPRAK, 2019). Nesse caso, o modelo de transferência de aprendizado funciona como um extrator de características para o conjunto de dados analisado, que é como essa tecnologia será aplicada neste estudo. Assim, matematicamente, consideramos D_T como um conjunto de dados de imagens de indivíduos com TEA e TD, e definimos a função f_Φ da seguinte forma:

$$\begin{aligned} f_\Phi: D_T &\rightarrow n\Phi \\ I &\mapsto f_\Phi(I) = \\ &\Phi_{FC}(I), \end{aligned} \tag{1}$$

onde $\Phi_{FC}(I)$ representa a saída da rede pré-treinada Φ para a imagem $I \in D_T$, e n_Φ denota o número de classes para as quais a rede original foi treinada. A Figura 1 ilustra um pipeline para o uso de redes neurais pré-treinadas como extratores de características para uma imagem I .

Figura 1 Representação do uso de uma rede pré-treinada Φ como extrator de características de imagem.



Fonte: Elaboração do autor.

5 Trabalhos similares

O autismo Lord et al. (2020), que é o foco central deste estudo, é um transtorno do neurodesenvolvimento caracterizado por dificuldades na comunicação social, padrões de comportamento repetitivos e interesses restritos. O diagnóstico dessa condição é geralmente realizado por meio de avaliações clínicas Bishop e Lord (2023), como observações comportamentais e entrevistas estruturadas, utilizando ferramentas especializadas como a consagrada CARS Schopler et al. (1980) e ADI-R (RUTTER et al., 2003). No entanto, esses

métodos dependem fortemente da interação direta entre o profissional e o paciente, sendo tanto subjetivos quanto demorados. Como alternativa aos procedimentos clínicos tradicionais, o reconhecimento automático de padrões a partir de sinais tem sido cada vez mais empregado por pesquisadores da área, oferecendo diagnósticos de TEA mais rápidos e precisos. Por exemplo, modelos de aprendizado de máquina Farooq et al. (2023) podem ser treinados para identificar características faciais específicas associadas ao transtorno, como pequenas diferenças na simetria facial ou no contato visual, que são difíceis de detectar clinicamente. A análise dessas características por meio de observação direta depende intrinsecamente da habilidade e experiência do profissional envolvido, o que limita tanto a escalabilidade quanto a precisão do diagnóstico, tornando a análise baseada em humanos impraticável. Para aprimorar a precisão e a objetividade na detecção do autismo, técnicas baseadas na extração de características de sinais, como imagens faciais do paciente, e sua subsequente classificação por algoritmos de aprendizado de máquina, especialmente aqueles baseados em aprendizado profundo de características, têm se tornado cada vez mais comuns nos últimos anos, como Uddin et al. (2024) destaca em sua revisão da literatura especializada. Esses métodos automatizados demonstraram o potencial de reduzir a subjetividade e melhorar a precisão diagnóstica, auxiliando os profissionais de saúde na identificação mais rápida e eficaz do autismo. O leitor interessado em mais detalhes e comparações sobre trabalhos nesse segmento pode analisar os levantamentos sobre detecção automática de autismo de Hyde et al. (2019) e Parlett-Pelleriti et al. (2023), que apresentam, respectivamente, um resumo das técnicas de aprendizado supervisionado e não supervisionado utilizadas nesse problema. A seguir, discutimos alguns dos principais estudos nesta área de pesquisa.

A detecção automática de TEA é baseada na análise computacional de dados associados ao paciente, que podem ser obtidos a partir da observação comportamental do indivíduo utilizando, por exemplo, mapeamentos computacionais de escalas bem conhecidas. Bone et al. (2016), por exemplo, representaram computacionalmente as respostas a testes de duas escalas, a ADI-R Rutter et al. (2003) e a Escala de Responsividade Social (SRS) Constantino (2021), associadas a cada paciente analisado, e avaliaram o desempenho dos classificadores SVM e *Random Forest* (RF) Parmar, Katariya e Patel (2019) por meio de validação cruzada em um banco de dados com mais de 1700 amostras, alcançando sensibilidades acima de 86% em seus resultados. Além de considerar várias escalas como o

escore de gravidade da *Autism Diagnostic Observation Schedule* (ADOS) Lord (1999), CARS e os escores globais da *Echelle d'évaluation des Comportements Autistiques* (ECA-R) Lelord e Barthélemy (1995) para representar um paciente, Silleresi et al. (2020) também utilizaram medidas extraídas da análise observacional da estrutura da linguagem, envolvendo repetição de frases e palavras não verbais, e habilidades não verbais. A representação final foi reduzida por meio da análise de componentes principais (PCA) Greenacre et al. (2022) para permitir a construção e análise visual de cinco clusters determinados pela técnica k-means (AHMED; SERAJ; ISLAM, 2020). De maneira similar, Zheng, Grove e Eapen (2019) desenvolveram um modelo baseado em agrupamento hierárquico utilizando 9 componentes principais determinados por PCA a partir de 188 crianças em idade pré-escolar. Augé et al. (2024) analisaram clusters determinados pela Análise de Perfil Latente (LPA) para examinar a relação entre características sensoriais e dificuldades executivas, representadas pelo Inventário de Avaliação Comportamental das Funções Executivas (BRIEF) Gioia et al. (2002), e dificuldades atencionais, representadas pela Escala de Classificação do Transtorno de Déficit de Atenção e Hiperatividade (ADHD-RS) DuPaul et al. (2016), em indivíduos com TEA, detectando três perfis principais considerando os valores brutos e dois perfis principais considerando os valores normalizados. No entanto, vale destacar que a maior e mais diversa parte dos estudos de detecção automática de TEA utilizando aprendizado de máquina foca na análise de características relacionadas aos aspectos fisiológicos dos indivíduos.

Utilizando Imagens por Ressonância Magnética Funcional (fMRI), Bhandage et al. (2023) propuseram uma abordagem baseada na otimização de uma Rede Neural de Crença Profunda (DBN) Hua, Guo e Zhao (2015) por meio da meta-heurística *Adam War Strategy Optimization* (AWSO) (AYYARAO et al., 2022; KINGMA, 2014) para detectar a presença de autismo em regiões pivotais de interesse. Park e Cho (2023) introduziram uma Rede Residual Convolutiva Gráfica que considera mudanças temporais nas conexões entre regiões em imagens cerebrais de FMRI, diagnosticando TEA ao identificar padrões localizados no Sulco Temporal Superior (STS). De forma semelhante, analisando dados de FMRI, Easson, Fatima e McIntosh (2019) utilizaram a técnica de agrupamento k-means para distinguir de forma otimizada dois subtipos distintos de padrões de conectividade funcional em participantes com autismo e em sujeitos de controle. Duffy e Als (2019) empregaram 40 características calculadas a partir de sinais de eletroencefalograma (EEG), mapeando fatores de coerência por todo o

cérebro, e utilizaram tanto agrupamentos simples quanto hierárquicos para visualizar a separabilidade entre indivíduos de controle, aqueles com TEA e aqueles com Síndrome de Asperger. Além disso, Bekele et al. (2013) examinaram clusters derivados da mistura Gaussiana e da análise k-means em componentes principais de EEG e outros sinais fisiológicos, demonstrando que indivíduos de controle e com TEA reagem de maneira diferente a emoções coletadas durante interações com um sistema de realidade virtual.

Os padrões de movimento ocular, ou o foco do olhar, em pacientes com TEA podem apresentar características atípicas (Pantelis e Kennedy, 2017), que podem ser mapeadas computacionalmente e utilizadas para a classificação automática do autismo. Tao e Shyu (2019) propuseram uma combinação de Redes Neurais Convolucionais (CNNs) e Redes de Memória de Longo e Curto Prazo (LSTM) para detectar autismo em um conjunto de dados de movimentos oculares de 300 indivíduos (DUAN et al., 2019). De maneira similar, Liu, Li e Yi (2016) desenvolveram uma arquitetura baseada em aprendizado de máquina onde crianças realizaram tarefas de reconhecimento facial, e seus movimentos oculares foram utilizados para treinar uma Máquina de Vetores de Suporte (SVM), resultando na construção de um modelo automatizado de diagnóstico. Atyabi et al. (2023) integraram dados de movimentos oculares, combinando informações espaciais como o ponto para onde a pessoa está olhando com dados temporais, como a velocidade com que o olhar se desloca, para alimentar uma CNN na detecção de TEA. Outra característica fisiológica que pode ser analisada temporalmente é o “esqueleto” do paciente, inferido como um grafo de Árvore Geradora Mínima (MST) do corpo do indivíduo. Por exemplo, Kojovic et al. (2021) extraíram pontos esqueléticos chave de pacientes utilizando a tecnologia OpenPose (Cao et al., 2017) e definiram um modelo baseado na integração de redes CNN e LSTM. Em uma abordagem similar, S et al. (2022) modelaram comportamentos de “*stimming*” (movimentos repetitivos) utilizando vídeos brutos e características extraídas de pontos chave e mapas de calor do esqueleto inferido de crianças, para treinar uma Rede Neural Profunda RGBPose-SlowFast (Duan et al., 2022) para a segregação automática de indivíduos com TEA e sujeitos de controle. Para calcular a frequência e a intensidade dos movimentos repetitivos de “bater os braços” em crianças com TEA, Dundi et al. (2023) empregaram técnicas de visão computacional e a estrutura MediaPipe (Lugaresi et al., 2019).

As expressões faciais em crianças com autismo frequentemente diferem das produzidas por indivíduos com desenvolvimento típico (TD) (Marchetti et al., 2014; Aldridge

et al., 2011; Hammond et al., 2008). Isso ocorre devido à dificuldade que indivíduos com autismo enfrentam tanto na produção quanto no processamento de emoções e expressões faciais, como demonstrado computacional e experimentalmente por (Guha et al., 2016). Consequentemente, técnicas de análise baseadas em imagens faciais, consideradas um dos métodos de coleta de sinais menos invasivos, têm ganhado destaque na literatura. Por exemplo, Shukla et al. (2017) treinaram uma CNN AlexNet otimizada (Krizhevsky, Sutskever e Hinton, 2012), que processa tanto a imagem facial completa quanto quatro sub-regiões para extrair representações reduzidas por PCA. Essas foram então utilizadas para definir um modelo de classificação baseado em SVM. A classificação de emoções a partir de um pequeno número de quadros de imagens foi realizada por Han et al. (2018), utilizando a renomada Rede Neural Convolutiva Profunda VGG16, pré-treinada pelo Visual Geometry Group da Universidade de Oxford (Simonyan e Zisserman, 2014), e representações esparsas por transferência de espaço de características. Leo et al. (2018b) e Leo et al. (2018a) utilizaram sequências de imagens para extrair características calculadas manualmente via uma CNN, a fim de quantificar a capacidade de crianças com TEA de produzir expressões faciais. Leo et al. (2019) generalizaram esse processo usando o Convolutional Experts Constrained Local Model (CECLM) (Zadeh et al., 2017) para detecção facial e realizaram experimentos adicionais para demonstrar a eficácia da abordagem proposta. De maneira semelhante, Rani (2019) empregaram o conhecido Padrão Binário Local (LBP) (Ojala, Pietikainen e Maenpaa, 2002) para treinar uma SVM e uma Rede Neural Artificial (ANN) na detecção de quatro emoções em crianças com autismo. Tamilarasi e Shanmugam (2020) utilizaram a Rede Neural Residual Profunda pré-treinada com 50 camadas, ResNet-50 (He et al., 2016a), para classificar TEA em crianças utilizando imagens faciais térmicas. De forma semelhante, Banire et al. (2021) classificaram os níveis de atenção em crianças com TEA por meio da análise de imagens faciais, avaliando duas representações computacionais: uma representação vetorial de características geométricas espaciais para ajustar uma SVM, e uma representação matricial das coordenadas dos marcos faciais coletadas ao longo de diferentes quadros para treinar uma CNN.

Akter et al. (2021) propuseram uma estrutura que consiste em modelos de transferência de aprendizado profundo aprimorados com base em imagens e classificadores de aprendizado de máquina clássicos, com representações analisadas usando uma etapa de agrupamento k-means para detectar TEA a partir de imagens faciais estáticas. Rahman e Subashini (2022)

avaliaram cinco CNNs pré-treinadas: MobileNet (Howard, 2017), Xception (Chollet, 2017), EfficientNetB0 (Tan, 2019), EfficientNetB1 e EfficientNetB2 como extratores de características e propuseram uma Rede Neural Profunda como classificador para diferenciar indivíduos com TEA de indivíduos com desenvolvimento típico (TD) com base em uma imagem facial. De maneira semelhante, Alam et al. (2022) avaliaram a otimização de hiperparâmetros de quatro modelos de CNN pré-treinados — VGG19 (Simonyan e Zisserman, 2014), Xception (Chollet, 2017), ResNet50V2 (He et al., 2016b), MobileNetV2 (Sandler et al., 2018) e EfficientNetB0 (Tan, 2019) — cada um conectado a uma camada totalmente conectada com 512 neurônios para detectar autismo a partir de imagens faciais. Jahanara e Padmanabhan (2021) também ajustaram a rede VGG19 em um conjunto de dados de imagens faciais de crianças com TEA e TD. No mesmo conjunto de dados, além de re-treinar as redes VGG19 e Xception, Alsaade e Alzahrani (2022) realizaram experimentos na rede NASNETMobile (Zoph et al., 2018). Arumugam et al. (2021) re-treinaram a rede VGG16 e Rabbi et al. (2021) propuseram um novo modelo de CNN para esse mesmo problema e conjunto de dados. Alkahtani, Aldhyani e Alzahrani (2023) aprimoraram o MobileNet-V1 e propuseram uma estrutura de extração de características utilizando modelos de transferência de aprendizado profundo, avaliando o método com vários classificadores clássicos de aprendizado de máquina. Por fim, Shahzad et al. (2024) concatenaram previsões de dois modelos pré-treinados ajustados, ResNet101 (He et al., 2016b) e EfficientNetB3 (Tan, 2019), com um modelo baseado em atenção para detectar autismo a partir de imagens estáticas.

Embora muitos esforços tenham sido dedicados à detecção do TEA usando técnicas de aprendizado de máquina e modelos baseados em redes neurais profundas, o problema permanece sem solução. Nenhum método atual alcança precisão total em todas as situações, destacando a necessidade de novas abordagens para melhorar a generalização e a robustez dessas técnicas. Além disso, métodos de pré-processamento, como filtragem e redução de dimensionalidade, que poderiam melhorar significativamente os resultados, têm sido pouco explorados nesse contexto. Portanto, este estudo propõe uma nova estrutura que combina várias técnicas de aprendizado profundo e processamento de imagens, incluindo aumento de dados, multifiltragem e equalização de histograma. A estrutura integra oito modelos pré-treinados e aplica uma abordagem em múltiplas fases para aprimorar a identificação de características faciais associadas ao TEA. A inovação reside na combinação dessas técnicas para fornecer uma

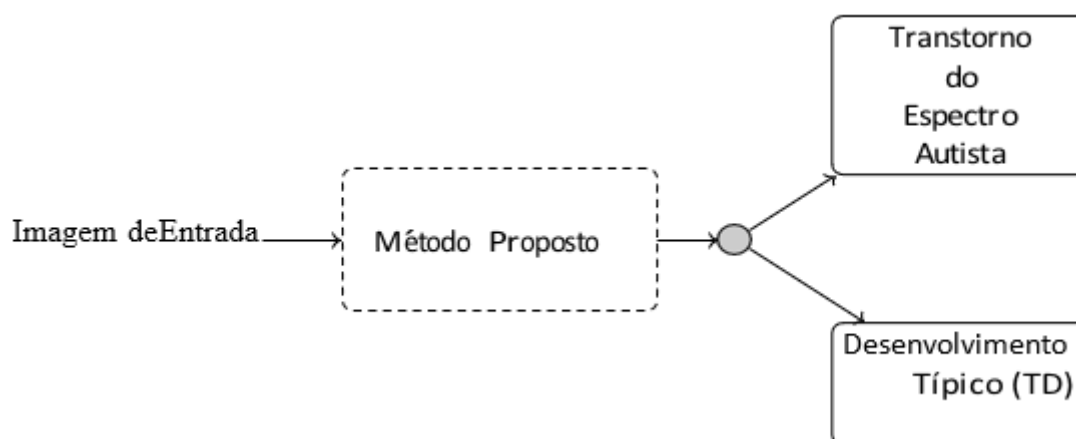
solução mais precisa para a detecção automatizada do transtorno. As seções a seguir detalham os componentes e etapas desta estrutura.

6 Metodologia

Soluções automatizadas para a detecção de autismo podem operar utilizando várias metodologias. Duas das abordagens mais comuns envolvem a definição do escopo do sistema: uma abordagem abrange não apenas a detecção de TEA, mas também etapas de reconhecimento, propostas de interação e até recomendações de programas de tratamento e acompanhamento médico. A outra, mais focada, limita-se à detecção do autismo a partir de uma imagem facial. Neste estudo, optamos por adotar a última abordagem.

Especificamente, a estrutura proposta foi projetada exclusivamente para determinar se uma imagem facial corresponde a um indivíduo com TEA ou a um indivíduo com desenvolvimento típico (TD). Uma representação esquemática do funcionamento da estrutura proposta está ilustrada na Figura 2.

Figura 2 Visão geral do funcionamento da estrutura proposta.



Fonte: Elaboração do autor.

Especificamente, a metodologia proposta consiste em quatro etapas principais descritas a seguir:

- **M_1 Definição do domínio do problema:** Para que a técnica funcione de forma eficaz, é essencial fornecer uma imagem de um rosto humano capturada, por exemplo, por uma câmera, representando o indivíduo para o qual se deseja inferir um diagnóstico de TEA. Assim, o domínio do problema consiste em imagens tipicamente definidas em escala RGB, ou seja, tensores no espaço. $[0, 255]^{3 \times N \times M}$, onde $N \times M$ especifica as dimensões da imagem. No entanto, toda a estrutura desenvolvida pode ser facilmente adaptada para operar com imagens em escala de cinza.
- **M_2 Método proposto:** Conforme mencionado, este estudo foca nos avanços para a detecção de autismo a partir de uma imagem facial estática. Portanto, a principal contribuição deste trabalho reside em uma estrutura que incorpora rotinas de aumento de dados, técnicas de aprimoramento de imagem e o processamento de características extraídas de redes pré-treinadas para melhorar as métricas relacionadas ao número de classificações corretas e incorretas associadas aos modelos padrão de aprendizado por transferência.
- **M_3 Saída do método:** A ferramenta desenvolvida deve ser capaz de determinar se uma dada imagem facial corresponde a um indivíduo com TEA ou TD. Assim, o método opera de acordo com uma rotina de classificação binária, atribuindo um dos seguintes rótulos à imagem de entrada: "TEA" ou "TD".
- **M_4 Validação:** Para validar a eficácia da estrutura proposta, análises serão conduzidas utilizando o benchmark mais amplamente usado na área, que é um conjunto de dados da plataforma de competições Kaggle contendo imagens faciais de crianças com TEA e TD. Em todos os cenários de teste, o sucesso de uma técnica na tarefa de classificação será determinado por métricas de desempenho relacionadas à acurácia do modelo. Especificamente, os seguintes cenários serão avaliados:
 - Análise de desempenho por configuração: Dado que o modelo proposto é uma estrutura geral, permitindo diversas instâncias específicas, são realizadas comparações entre múltiplas configurações propostas. A configuração da estrutura requer a definição de métodos como técnicas de filtragem e

equalização de histograma, entre outros. Além disso, oito redes pré-treinadas são consideradas nos experimentos.

- Comparação com o estado da arte: Além de comparar diferentes configurações da estrutura proposta, é essencial avaliar o desempenho da classificação do método em relação a técnicas existentes na literatura que representam o estado da arte nesta área.

7 Desenvolvimento

Nesta seção, descrevemos os componentes que compõem o método desenvolvido para identificar TEA em indivíduos com base em imagens faciais. Fornecemos uma explicação detalhada sobre o funcionamento de todas as técnicas empregadas por meio de algoritmos e fluxogramas, visando facilitar a compreensão e a replicação do *framework* proposto. Em particular, destacamos as seguintes inovações introduzidas neste trabalho:

- Um *framework* inovador para extração e classificação de características de imagens faciais utilizando redes de aprendizado profundo pré-treinadas, com o objetivo de distinguir amostras em dois grupos distintos: o primeiro grupo consiste em imagens de indivíduos com TEA, enquanto o segundo grupo contém amostras de indivíduos com DT;
- Uma análise experimental de várias configurações do *framework* generalizado proposto é realizada neste estudo.

A ideia de utilizar etapas de aumento de dados e aprimoramento de imagens para aumentar a acurácia de classificadores em bases de dados pequenas não é nova. Neste trabalho, foi proposta uma adaptação do *framework* apresentado por Contreras et al. (2021) e Contreras et al. (2022), originalmente projetado para o problema de detecção de falsificação de impressões digitais por meio da análise de descritores de textura projetados manualmente. Este *framework* foi modificado para abordar o problema de detecção de autismo a partir de imagens faciais, utilizando características extraídas de redes pré-treinadas. A seção seguinte apresenta a adaptação proposta, que consiste em três etapas principais: Aumento de Dados; Processamento

de Imagens de Entrada; e Definição do Modelo de Representação Computacional e Classificação.

7.1 Aumento de Dados

A maioria dos conjuntos de dados de imagens médicas é composta por um número insuficiente de amostras, o que frequentemente é citado como justificativa para a utilização de aprendizado por transferência na formulação de modelos. Para enfrentar esse problema, rotinas de aumento de dados (Mumuni e Mumuni, 2022) podem ser empregadas. De fato, o uso dessas estratégias é relevante na classificação facial com redes neurais profundas (DNNs) (Wang, Wang e Lian, 2020), inclusive no desenvolvimento de modelos baseados em aprendizado por transferência (Uchoa et al., 2020). Assim, o primeiro passo do *framework* é proposto como o aumento sintético do conjunto de amostras de imagens faciais. Matematicamente, seja $\mathcal{A}$ o conjunto de técnicas de aumento de dados consideradas.

$$\mathcal{A} = \{A_1, A_2, \dots, A_{n_{\mathcal{A}}}\}, \quad (2)$$

onde A_i é uma função que mapeia do espaço tensorial de imagens para outro.

Assim, partindo de um conjunto de dados de imagens faciais que compõem o conjunto de amostras de treinamento B_{Train} , o conjunto aumentado $\hat{B}_{\text{Train}}$ é criado, definido como segue:

$$\hat{B}_{\text{Train}} = \bigcup_{i=1}^{n_{\mathcal{A}}} A_i(B_{\text{Train}}). \quad (3)$$

7.2 Processamento de Imagens de Entrada e Multifiltragem

Um dos passos mais comuns em sistemas de análise de imagens é o aprimoramento das imagens de entrada (SHAKER; BAKER; MAHMOOD, 2022). Na maioria dos trabalhos

discutidos, tal processo pode ser encontrado. No *framework* proposto, esta categoria de estratégias é empregada para aprimorar as capacidades de detecção de características de descritores baseados em aprendizado por transferência. Duas sub-rotinas serão consideradas para isso: equalização de histograma adaptativa e multifiltragem. A primeira, representada pela função $HE(\cdot)$, é utilizada para corrigir possíveis anomalias de iluminação nas imagens. A segunda será aplicada para reduzir ruídos e/ou destacar padrões utilizando múltiplas funções de filtragem de diferentes tipos. Matematicamente, o conjunto de multifiltragem é definido como $\mathcal{F}$:

$$\mathcal{F} = \{F_1, F_2, \dots, F_{n_{\mathcal{F}}}\}, \quad (4)$$

onde F_i é uma função de filtro, $\forall i$. Consequentemente, $F_i(I)$ é uma versão filtrada de uma imagem I .

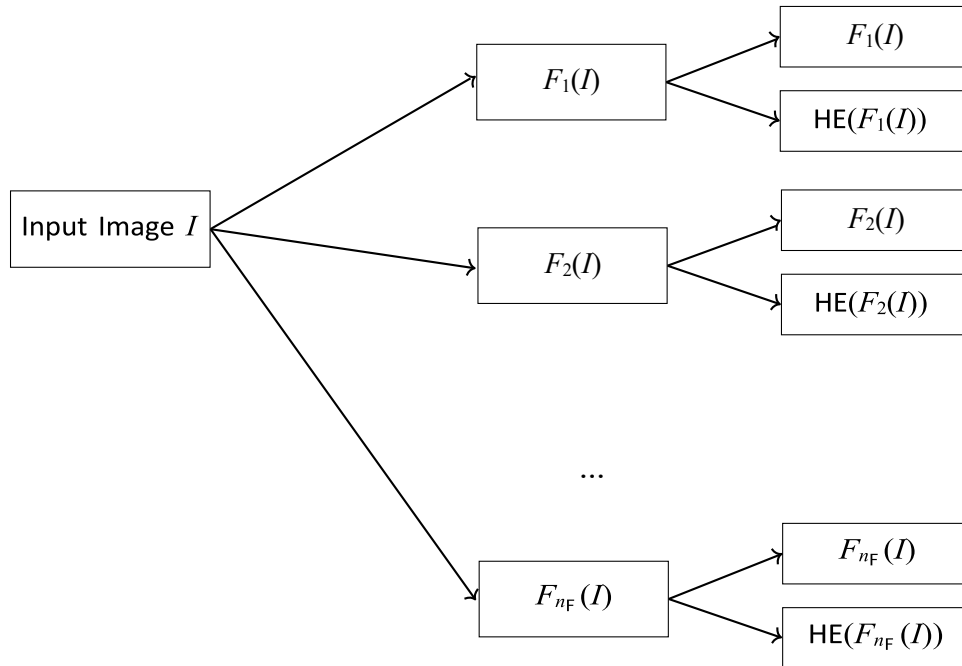
É importante enfatizar que o objetivo de aprimorar a capacidade representativa de uma imagem é garantir que nenhuma de suas versões geradas durante esta etapa seja descartada sem razão, mas que todas sejam consideradas na fase de extração de características. Assim, para cada imagem I , $n_{\mathcal{F}}$ versões filtradas serão geradas, além de um número igual de versões com iluminação corrigida, ou seja, com histogramas equalizados. A Figura 3 apresenta um diagrama ilustrando a geração de versões filtradas e corrigidas em termos de iluminação das imagens de entrada no *framework* proposto.

Ao final desta etapa, dada uma imagem I , será construído um conjunto $I(I)$. Este conjunto compreende a imagem original I , sua versão com problemas de iluminação corrigidos $HE(I)$, e todas as suas versões filtradas com e sem equalização de histograma. Dessa forma, nenhuma característica que tenha sido destacada pelo processo de filtragem ou de equalização de histograma será desconsiderada. Além disso, para garantir que as características associadas à imagem original também sejam computadas e não se percam durante o processo, considera-se que um dos filtros é igual à função identidade ou, em outras palavras, que a imagem original não filtrada e sua versão equalizada são consideradas em $I(I)$. Matematicamente,

$$I(I) = \{I, HE(I), F_1(I), HE(F_1(I)), F_2(I), HE(F_2(I)), \dots, F_{n_{\mathcal{F}}}(I), HE(F_{n_{\mathcal{F}}}(I))\}, \quad (5)$$

onde $I(I)$ é um conjunto contendo $(2 \cdot n_{\mathcal{F}} + 2)$ imagens.

Figura 3 Fluxograma representando o processo de transformação das imagens de entrada por meio das funções $F_i(I)$, seguido pela equalização de histograma $HE(F_i(I))$ para cada função.

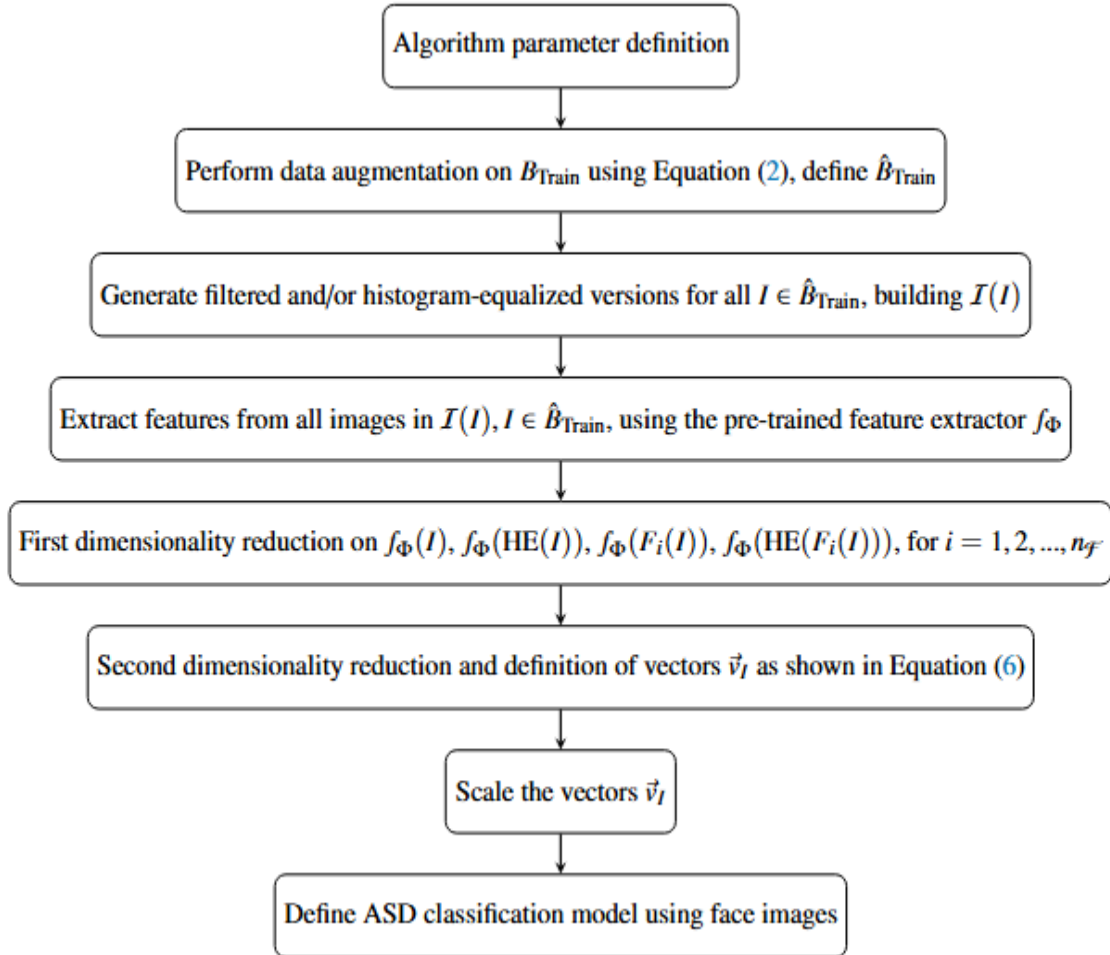


Fonte: Elaboração do autor.

7.3 Algoritmo proposto

O *framework* proposto envolve a execução sequencial de todas as etapas descritas nesta seção. É necessário estabelecer uma configuração prática para todos os parâmetros do algoritmo, uma vez que o *framework* foi generalizado para permitir múltiplas configurações. Em seguida, deve-se realizar a ampliação sintética de dados no conjunto de treinamento. Versões filtradas e/ou imagens com equalização de histograma precisam ser geradas para todas as imagens disponíveis. Para treinar o modelo de detecção de autismo com base na análise de imagens faciais, devem ser obtidas representações computacionais de todas as imagens do conjunto de dados ampliado. Por fim, todas as etapas do *framework* proposto estão descritas no fluxograma apresentado na Figura 4.

Figura 4 Fluxograma representando o processo de transformação das imagens de entrada por meio das funções $F_i(I)$, seguido pela equalização de histograma $HE(F_i(I))$ para cada função.



Fonte: Elaboração do autor.

8 Resultados e discussões

Esta seção realizará os experimentos necessários para avaliar a estrutura proposta. Para isso, será utilizado um benchmark, detalhado posteriormente, amplamente reconhecido na área. Especificamente, este estudo foca na avaliação individual de cada etapa da estrutura. Além disso, os resultados obtidos foram comparados com estudos relevantes que representam o estado da arte na detecção do autismo para validar a eficácia da abordagem proposta.

Para diferenciar o desempenho das diversas versões consideradas e promover a comparação de nossos avanços com trabalhos futuros na mesma área, foram definidas métricas para capturar a acurácia e os erros do método em relação às imagens faciais no conjunto de dados. As seguintes medidas de avaliação foram selecionadas para este problema de classificação binária:

- Falsos Positivos (FP): O número de crianças não autistas classificadas incorretamente como autistas.
- Falsos Negativos (FN): O número de crianças autistas classificadas incorretamente como não autistas. Verdadeiros Positivos (TP): O número de crianças autistas corretamente classificadas como autistas.
- Verdadeiros Negativos (TN): O número de crianças não autistas corretamente classificadas como não autistas.
- Acurácia (ACC):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

que é a proporção de imagens corretamente classificadas, ou seja, crianças com ou sem autismo, sobre o número total de casos.

- Erro Médio de Classificação (ACE):

$$ACE = \frac{FP + FN}{TP + TN + FP + FN}$$

que indica a proporção média de classificações incorretas em todas as previsões.

- Recall ou Sensibilidade:

$$Recall = \frac{TP}{TP + FN}$$

que é a proporção de crianças autistas corretamente identificadas pelo modelo.

- Precisão:

$$Precision = \frac{TP}{TP + FP}$$

que é a proporção de crianças previstas como autistas que realmente são autistas.

- Especificidade:

$$Specificity = \frac{TN}{TN + FP}$$

que é a proporção de crianças não autistas corretamente identificadas pelo modelo.

- F1 Score ($F1$):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

que representa um equilíbrio entre precisão e recall.

- Área Sob a Curva (AUC): A área sob a curva ROC, refletindo a capacidade do modelo de distinguir entre crianças autistas e não autistas com base nas características faciais.
- Taxa de Erro Igual (EER): O ponto na curva ROC onde a taxa de falsos positivos, ou seja, classificar erroneamente uma criança não autista como autista, é igual à taxa de falsos negativos, ou seja, classificar erroneamente uma criança autista como não autista.

As implementações computacionais necessárias para obter os resultados apresentados neste trabalho foram realizadas utilizando a linguagem de programação Python. Além disso, empregamos a biblioteca TensorFlow Developers (2022), PyTorch Paszke et al. (2019) e o repositório Hugging Face (<https://huggingface.com/>, acessado em 30 de setembro de 2024) para a configuração de redes pré-treinadas. Para rotinas de processamento de imagens, foi utilizada a conhecida biblioteca OpenCV Bradski (2000), especificamente sua versão para Python. Por fim, rotinas relacionadas à redução de dimensionalidade e ao treinamento de classificadores foram implementadas utilizando a biblioteca scikit-learn (PEDREGOSA et al., 2011). Todos os desenvolvimentos foram executados em um computador pessoal equipado com 8 GB de RAM e um processador Intel (R) Core (TM) i5-4460 com frequência de 3.20 GHz.

8.1 Benchmark

O conjunto de dados utilizado para as avaliações deste estudo é a coleção de imagens de Piosenka (2021a), originalmente publicada no site de competições Kaggle e atualmente disponível no repositório Google Drive (PIOSENKA, 2021b). O objetivo da construção deste

conjunto de dados é compilar imagens das faces de crianças com Transtorno do Espectro Autista (TEA) e crianças com desenvolvimento típico (TD). Essas imagens foram coletadas automaticamente da internet e recortadas para formar tensores coloridos com dimensões de $224 \times 224 \times 3$. O conjunto de dados é dividido em três subconjuntos: um conjunto de treino contendo 1268 amostras de faces de indivíduos com TEA e a mesma quantidade de amostras de indivíduos TD; um conjunto de validação com 50 amostras de faces de indivíduos com TEA e 50 de indivíduos TD; e um conjunto de teste com 150 imagens de faces de indivíduos com TEA e 150 de indivíduos TD.

8.2 Análise das Etapas da Estrutura

Considerando todas as versões analisadas, a estrutura proposta gerou um total de 1160 configurações. Nesta seção, cada componente da estrutura, juntamente com suas várias configurações, foi avaliado. As Tabelas 1 e 2 apresentam as melhores métricas e configurações para cada arquitetura de rede neural considerada nos conjuntos de Teste e Validação, respectivamente. Essas tabelas detalham a arquitetura do modelo, escala de dados, comprimento do vetor de características, aplicação de aumento de dados (DA) e as principais métricas de avaliação para a melhor configuração, como acurácia, F1 score, AUC, Equal Error Rate (EER), Average Classification Error (ACE), recall, precisão, especificidade e componentes da matriz de confusão (FP, FN, TP, TN). Cada arquitetura de rede foi avaliada utilizando o *framework* (+ FW) e de forma pura no classificador SVM. Os valores em negrito representam os melhores resultados da métrica em cada coluna.

Ao analisar os resultados apresentados, é evidente que o desempenho de todas as arquiteturas de rede melhorou com a aplicação da estrutura proposta na maioria das métricas. Especificamente, no conjunto de testes, a acurácia da representação vetorial de todos os modelos de rede pré-treinados foi aprimorada, com um aumento absoluto de até 3,33%, como observado na rede AffectNet. No conjunto de validação, a rede ResNet50 apresentou uma melhora significativa de 8% na acurácia com o uso da estrutura.

Tabela 1 Resumo das métricas de desempenho para diferentes arquiteturas de aprendizado profundo aplicadas ao conjunto de testes do benchmark.

DL Architecture	Scale	Length	Use DA	ACC	F1	AUC	EER	ACE	Recall	Precision	Specificity	FP	FN	TP	TN
AffectNet	None	1000	False	87.67	88.18	94.16	12.00	12.33	92.00	84.66	83.33	25	12	138	125
AffectNet + FW	None	207	True	91.00	91.26	95.09	10.00	9.00	94.00	88.68	88.00	18	9	141	132
AlexNet	None	1000	False	79.33	80.13	85.56	22.00	20.67	83.33	77.16	75.33	37	25	125	113
AlexNet + FW	Standard	39	True	81.67	82.54	87.66	20.00	18.33	86.67	78.79	76.67	35	20	130	115
ResNet-50	None	1000	False	75.67	76.68	82.45	24.00	24.33	80.00	73.62	71.33	43	30	120	107
ResNet-50 + FW	Standard	30	False	77.00	77.67	79.29	24.67	23.00	80.00	75.47	74.00	39	30	120	111
VGG16	None	1000	False	73.00	73.27	80.27	27.33	27.00	74.00	72.55	72.00	42	39	111	108
VGG16 + FW	None	33	True	74.00	75.32	77.85	27.33	26.00	79.33	71.69	68.67	47	31	119	103
VGG19	None	1000	False	72.00	73.08	78.36	28.67	28.00	76.00	70.37	68.00	48	36	114	102
VGG19 + FW	Robust	40	True	74.00	75.62	79.07	29.33	26.00	80.67	71.18	67.33	49	29	121	101
ViT	None	1000	False	87.67	87.87	94.40	12.00	12.33	89.33	86.45	86.00	21	16	134	129
ViT + FW	Minmax	278	True	90.67	90.91	95.42	10.00	9.33	93.33	88.61	88.00	18	10	140	132
ViTFER	None	1000	False	87.00	87.21	93.44	12.67	13.00	88.67	85.81	85.33	22	17	133	128
ViTFER + FW	Standard	265	True	88.33	88.45	93.33	12.00	11.67	89.33	87.58	87.33	19	16	134	131
ViTSwin	None	1000	False	90.33	90.49	95.35	10.67	9.67	92.00	89.03	88.67	17	12	138	133
ViTSwin + FW	Minmax	163	True	92.67	92.81	95.29	8.67	7.33	94.67	91.03	90.67	14	8	142	136

Fonte: Elaboração do autor.

Tabela 2 Resumo das métricas de desempenho para diferentes arquiteturas de aprendizado profundo aplicadas ao conjunto de validação do benchmark.

DL Architecture	Scale	Length	Use DA	ACC	F1	AUC	EER	ACE	Recall	Precision	Specificity	FP	FN	TP	TN
AffectNet	None	1000	False	82.0	81.63	90.60	18.0	18.0	80.0	83.33	84.0	8	10	40	42
AffectNet + FW	None	167	False	87.0	86.60	90.60	16.0	13.0	84.0	89.36	90.0	5	8	42	45
AlexNet	None	1000	False	82.0	81.25	88.08	22.0	18.0	78.0	84.78	86.0	7	11	39	43
AlexNet + FW	Robust	43	False	82.0	82.69	90.40	18.0	18.0	86.0	79.63	78.0	11	7	43	39
ResNet-50	None	1000	False	72.0	70.83	83.56	24.0	28.0	68.0	73.91	76.0	12	16	34	38
ResNet-50 + FW	Standard	36	False	80.0	80.00	86.76	20.0	20.0	80.0	80.00	80.0	10	10	40	40
VGG16	None	1000	False	70.0	66.67	78.44	28.0	30.0	60.0	75.00	80.0	10	20	30	40
VGG16 + FW	Robust	31	False	77.0	76.77	83.84	22.0	23.0	76.0	77.55	78.0	11	12	38	39
VGG19	None	1000	False	68.0	63.64	76.36	30.0	32.0	56.0	73.68	80.0	10	22	28	40
VGG19 + FW	Standard	38	False	74.0	73.47	79.76	28.0	26.0	72.0	75.00	76.0	12	14	36	38
ViT	None	1000	False	81.0	80.81	92.36	18.0	19.0	80.0	81.63	82.0	9	10	40	41
ViT + FW	None	130	False	85.0	85.15	91.00	16.0	15.0	86.0	84.31	84.0	8	7	43	42
ViTFER	None	1000	False	80.0	79.17	89.68	20.0	20.0	76.0	82.61	84.0	8	12	38	42
ViTFER + FW	Robust	213	False	86.0	84.78	92.64	14.0	14.0	78.0	92.86	94.0	3	11	39	47
ViTSwin	None	1000	False	81.0	80.81	91.68	18.0	19.0	80.0	81.63	82.0	9	10	40	41
ViTSwin + FW	None	186	False	87.0	86.60	91.96	16.0	13.0	84.0	89.36	90.0	5	8	42	45

Fonte: Elaboração do autor.

Para demonstrar a eficácia do uso do aprimoramento de imagens na composição da representação computacional de cada imagem facial, a Tabela 3 apresenta os melhores resultados da estrutura em relação a algumas variações do conjunto de filtros F e ao uso ou não da estratégia de equalização de histograma definida pela função. $HE(\cdot)$. Especificamente, a seguinte nomenclatura foi adotada: “high” e “smooth” indicam, respectivamente, filtragem passa-alta e passa-baixa; “None” indica que a arquitetura não considerou nenhuma etapa da estrutura proposta; “original” indica que a etapa de processamento de imagem, com equalização de histograma e filtragem, foi desconsiderada; “histeq” simboliza o uso da função $HE(\cdot)$; o uso conjunto de equalização de histograma e filtragem é representado por “_”, sendo “histeq_smooth”, por exemplo, o uso conjunto de $HE(\cdot)$ com filtragem passa-baixa; por fim, o símbolo de soma “+” indica que mais de uma estratégia foi utilizada na mesma imagem, compondo a versão completa da estrutura.

Tabela 3 Impacto do desempenho de filtros e equalização de histograma na estrutura proposta em várias arquiteturas de aprendizado profundo (resultados em negrito destacam o melhor valor de acurácia em porcentagem para cada arquitetura no conjunto indicado).

Filters (/) / Structure	AffectNet	AlexNet	ResNet-50	ViTSwin	VGG16	VGG19	ViT	ViTFER	AffectNet	AlexNet	ResNet-50	ViTSwin	VGG16	VGG19	ViT	ViTFER
	Test								Validation							
None	87.67	79.33	75.67	90.33	73.00	72.00	87.67	87.00	82.00	82.00	72.00	81.00	70.00	68.00	81.00	80.00
high	89.33	77.33	69.67	89.33	68.33	62.00	82.00	80.67	83.17	75.00	61.17	81.00	65.00	63.50	80.00	80.17
smooth	90.33	76.33	71.67	88.67	71.67	71.33	88.33	88.33	83.67	82.00	78.00	86.00	74.00	71.00	81.00	86.00
original	91.00	78.67	74.00	91.33	71.67	72.00	89.00	87.33	83.83	82.00	79.00	84.33	72.00	70.00	81.00	82.00
histeq_high	89.33	77.33	69.00	89.00	69.00	63.33	81.67	81.00	81.50	76.00	61.00	81.50	66.00	67.00	79.00	83.00
histeq_smooth	89.33	78.00	73.33	89.33	71.67	72.67	87.33	86.00	82.00	79.00	78.00	82.50	74.00	74.00	83.00	80.17
histeq_original	88.67	79.67	73.67	90.67	73.00	71.00	89.00	88.33	83.33	81.00	71.00	81.50	73.00	71.00	83.00	80.83
high+histeq_high	88.67	74.00	67.00	89.00	67.67	60.33	81.00	80.00	80.50	73.83	60.33	81.17	66.00	66.00	80.00	81.00
original+histeq_high	90.00	79.33	75.33	90.67	73.00	71.00	89.00	86.00	85.83	81.00	71.00	85.17	73.00	70.00	80.67	82.00
original+smooth+high	90.67	78.33	74.33	90.33	74.00	72.00	89.33	87.00	84.33	80.00	74.00	86.17	71.00	72.00	81.17	84.00
smooth+histeq_smooth	89.33	77.33	72.33	90.00	71.33	74.00	87.33	88.33	81.17	78.00	80.00	86.00	77.00	71.67	83.00	84.00
original+histeq_smooth	89.33	78.67	72.67	92.67	70.00	71.33	87.33	87.67	83.00	79.00	76.00	85.50	76.00	71.50	84.00	84.00
original+histeq_original	88.67	79.00	77.00	90.33	72.00	71.00	89.00	86.33	82.17	76.17	74.00	84.67	71.00	69.67	84.00	82.00
histeq_original+smooth+high	89.67	78.33	76.67	91.67	72.67	72.33	90.67	86.33	85.00	81.00	73.00	87.00	72.00	73.00	82.00	83.00
histeq_original+histeq_high	89.00	79.67	75.67	91.33	71.00	69.33	89.00	85.67	86.00	78.00	70.00	83.83	73.00	71.00	83.00	84.00
histeq_original+histeq_smooth	88.33	79.00	74.00	91.67	73.67	70.67	90.00	87.00	82.00	76.00	74.00	83.00	75.00	72.00	85.00	80.50
original+histeq_smooth+histeq_high	90.33	79.33	74.33	91.33	73.00	72.33	89.00	86.67	85.00	80.00	73.00	86.33	71.00	72.00	82.00	81.83
histeq_original+histeq_smooth+histeq_high	89.67	78.67	74.67	92.00	72.00	72.67	89.00	85.00	86.00	81.00	72.00	84.67	75.00	73.00	82.00	85.00
original+smooth+high+histeq_original+histeq_smooth+histeq_high	90.00	81.67	75.67	91.00	73.67	72.33	89.67	87.33	87.00	79.00	74.00	86.50	71.00	73.00	81.00	82.33

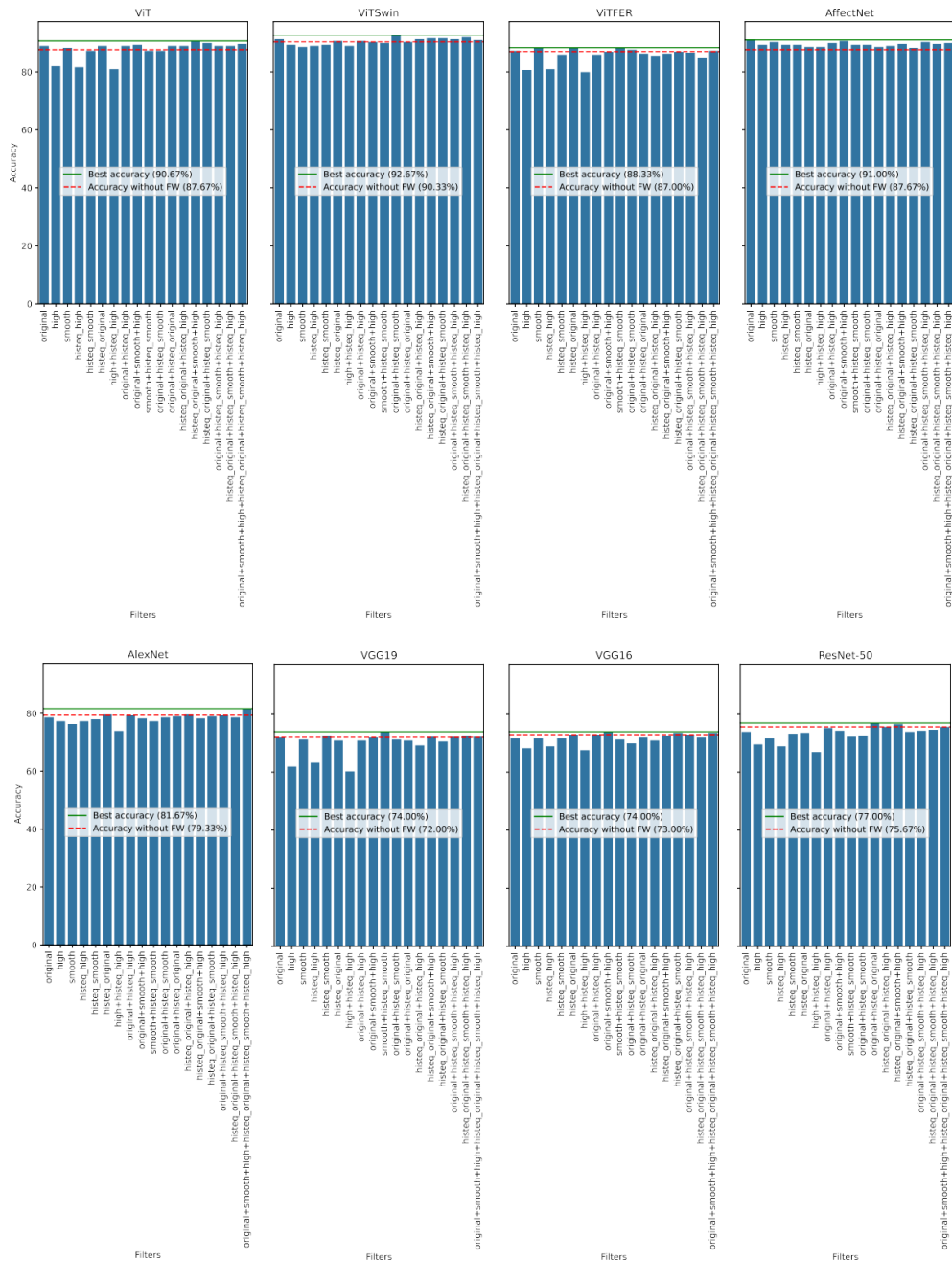
Fonte: Elaboração do autor.

Os resultados apresentados na Tabela 3 mostram o efeito de diferentes estruturas de filtros e técnicas de equalização de histograma na acurácia de diversos modelos de aprendizado profundo utilizados em uma estrutura de detecção de autismo. No conjunto de testes, a combinação de filtros suavizantes e passa-alta com equalização de histograma, aplicada de várias formas, geralmente melhora o desempenho do modelo. Modelos como ViTswin e AffectNet apresentaram melhorias significativas de acurácia utilizando essas técnicas, alcançando até 92,67% e 91,00%, respectivamente. No conjunto de validação, observou-se uma tendência similar, com ViTswin e AffectNet mantendo um desempenho elevado em várias combinações de filtros, com acurácia chegando a 86,50% e 87,00%, respectivamente. Modelos como VGG16 e VGG19 também se beneficiam desses aprimoramentos, embora em menor grau. Notavelmente, a arquitetura ResNet-50 novamente apresenta ganhos inconsistentes, sugerindo que este modelo específico pode não responder de forma tão robusta aos aprimoramentos aplicados quanto outras arquiteturas. Em resumo, enquanto a maioria dos modelos mostra melhorias na acurácia com o uso de filtros e equalização de histograma, o grau de benefício varia entre as arquiteturas. Os resultados sugerem que a estrutura proposta enriquece efetivamente as representações de características para a detecção de autismo, com ViTswin e AffectNet se destacando particularmente sob esses aprimoramentos.

Além disso, ao comparar os casos “None”, que representa a ausência de todas as etapas da estrutura, e “original”, que representa a ausência apenas da etapa de processamento de imagens, observa-se que, na maioria dos casos, já há uma melhora na acurácia. Isso indica que o aumento de dados e a redução de dimensionalidade já são estratégias suficientes para aumentar a robustez da classificação. No entanto, na maioria dos casos, a melhor acurácia dos modelos é obtida quando técnicas de equalização de histograma e multifiltragem são aplicadas. Especificamente, nos Gráficos 1 e 2, são apresentados gráficos de barras que representam a acurácia dos modelos na presença de estratégias de multifiltragem e equalização de histograma nos conjuntos de Teste e Validação, respectivamente. Em grande parte das arquiteturas de redes neurais, observa-se uma tendência de que as barras à direita, associadas ao uso de mais técnicas de processamento de imagem dentro da estrutura, sejam mais altas, enquanto as barras à esquerda tendem a ser mais baixas. Além disso, pode-se notar que todas as estruturas avaliadas no conjunto de teste melhoraram com o uso da estrutura, enquanto, exceto pela arquitetura

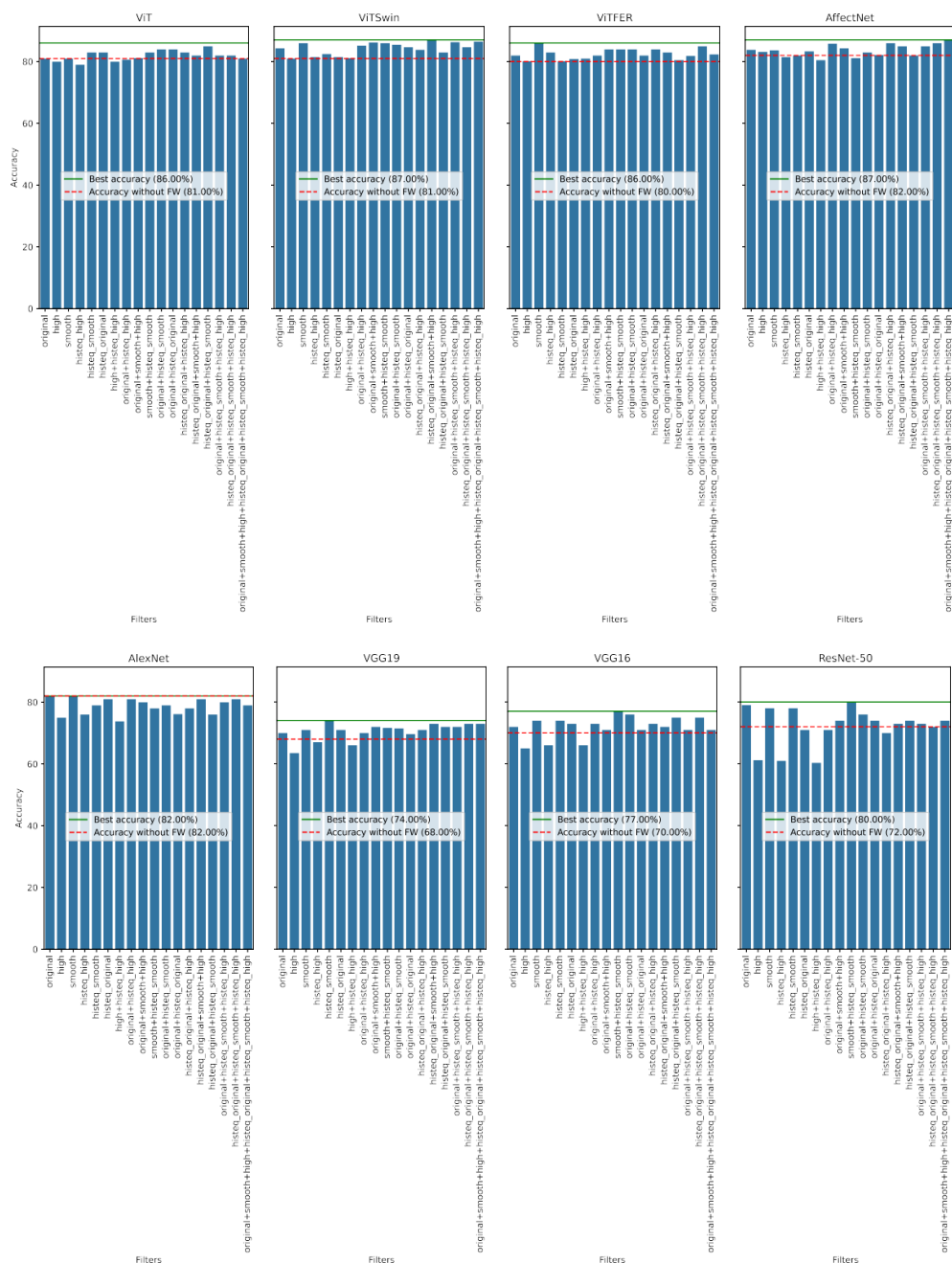
AlexNet, que não mostrou melhora nem declínio, todas as outras redes apresentaram aumento na acurácia no conjunto de validação.

Gráfico 1 Gráfico de barras comparando o desempenho de diferentes estratégias de multifiltragem e equalização de histograma em diferentes arquiteturas de redes neurais no conjunto de Teste.



Fonte: Elaboração do autor

Gráfico 2 Gráfico de barras comparando o desempenho de diferentes estratégias de multifiltragem e equalização de histograma em diferentes arquiteturas de redes neurais no conjunto de Validação.

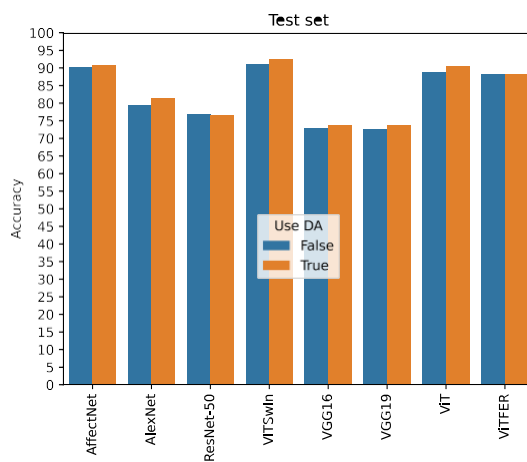


Fonte: Elaboração do autor

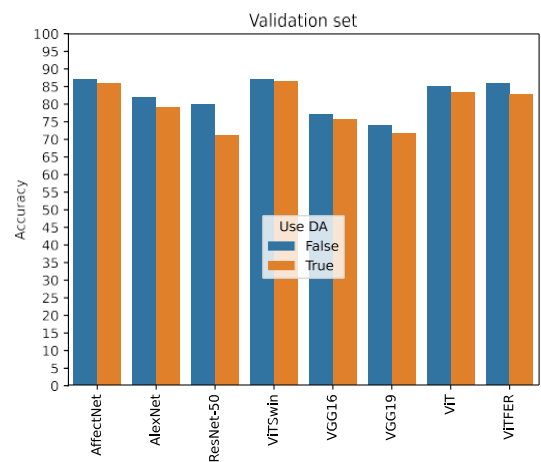
Os resultados relacionados ao uso de aumento de dados também destacam a heterogeneidade entre os conjuntos de teste e validação do benchmark considerado. Especificamente, no Gráfico 3a e 3b, são apresentadas as maiores acurácias alcançadas por cada rede pré-treinada utilizando a estrutura proposta, com barras laranjas representando aquelas que empregaram a etapa de aumento de dados, e barras azuis representando aquelas que não utilizaram, para os conjuntos de teste e validação, respectivamente. Ao analisar os gráficos, é evidente que o uso do aumento de dados aumentou a acurácia de todas as arquiteturas de rede, exceto ResNet-50, avaliadas no conjunto de teste. No entanto, o uso dessa etapa da estrutura não parece ter tido um efeito benéfico nas arquiteturas quando avaliadas no conjunto de validação. Como o conjunto de validação contém apenas 100 amostras, que é exatamente um terço do número de amostras no conjunto de teste, a adição de dados aumentados durante o treinamento pode ter feito com que o modelo se adaptasse bem a certos padrões que podem não estar presentes no conjunto de teste, mas também pode ter causado overfitting no conjunto de validação.

Gráfico 3 Comparação dos melhores valores de acurácia com e sem o uso de aumento de dados para diferentes arquiteturas de redes pré-treinadas, nos conjuntos de teste, em (a), e validação, em

(b).



(a)



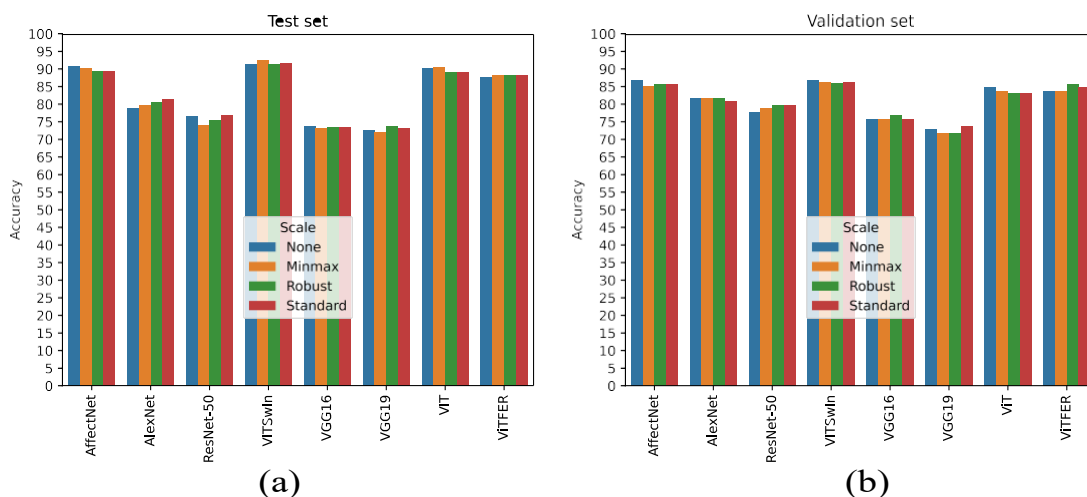
(b)

Fonte: Elaboração do autor

O uso de estratégias de escalonamento com a função $\text{SCALE}(\cdot)$ também foi responsável por melhorar os valores das métricas das arquiteturas de rede ao utilizar a estrutura.

Especificamente, analisando os gráficos de barras no Gráfico 4a e 4b, que apresentam os melhores valores de acurácia para cada rede pré-treinada utilizando a estrutura para os conjuntos de teste e validação, respectivamente, pode-se observar que, na maioria dos casos, 6 de 8 redes no conjunto de teste e 5 de 8 redes no conjunto de validação, alguma estratégia de escalonamento foi associada a um valor de acurácia mais alto em comparação à ausência de escalonamento (“None”). Além disso, também é notável que, no conjunto de validação, apenas os escalonamentos Standard e sua variação, Robust, estão associados a valores de acurácia mais altos na estrutura, superando a ausência de escalonamento. No conjunto de teste, esses mesmos escalonamentos estão associados a metade das redes analisadas, enquanto o escalonamento MinMax está associado a duas redes, nomeadamente ViT e ViTSwin. O maior valor de acurácia obtido, alcançado pela rede ViTSwin utilizando a estrutura proposta, foi atingido quando a função SCALE(·) foi configurada para escalonamento MinMax.

Gráfico 4 Comparação das melhores acurácias obtidas utilizando quatro diferentes estratégias de escalonamento nenhum escalonamento (None), MinMax, Robust e Standard para cada arquitetura de rede pré-treinada nos conjuntos de teste, em (a), e validação, em (b).

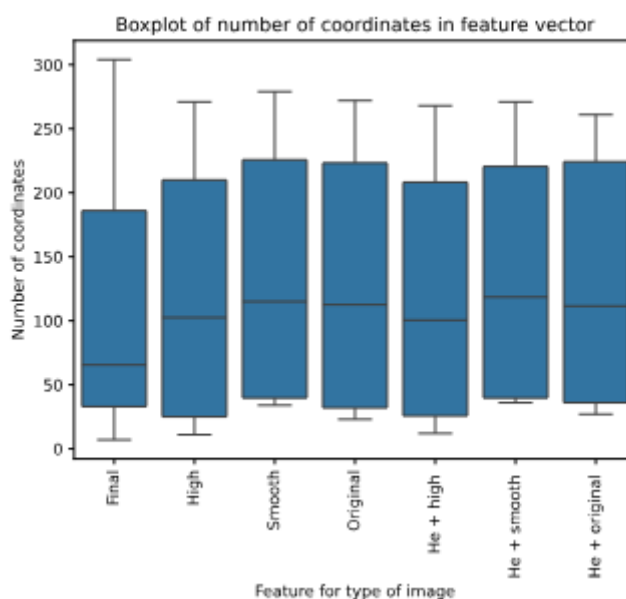


Fonte: Elaboração do autor

Analisando o efeito das duas etapas de redução de dimensionalidade propostas no *framework*, diversos padrões importantes podem ser observados. No Gráfico 4, um boxplot é apresentado para cada tipo de imagem resultante da etapa de aprimoramento de imagens, que

define os conjuntos I para o número de coordenadas que assume após a primeira etapa de redução, estabelecendo $\vec{v}_{\text{first-reducing}}$. Adicionalmente, um boxplot é apresentado para a dimensão do vetor final de características $\vec{v}_I$. Analisando a representação, é evidente que, a partir das 1000 coordenadas iniciais de cada vetor $f\Phi(\cdot)$, o vetor é reduzido na primeira etapa para aproximadamente 10% de seu tamanho original, em média. Essa redução ocorre de maneira consistente para todos os tipos de imagens em I , sejam elas originais, filtradas, equalizadas ou uma combinação dessas melhorias. Também é notável que o vetor final, resultante da segunda etapa de redução de dimensionalidade no método proposto, tende a possuir ainda menos coordenadas do que os vetores intermediários. Assim, vale destacar que o *framework* proposto é também eficaz em gerar um conjunto de características computacionalmente reduzido e representativo, considerando os valores de acurácia mencionados anteriormente.

Gráfico 5 Número de coordenadas que o vetor de características associado às imagens de 5 possui em cada versão analisada do *framework* após a primeira etapa de redução dimensional do método, compondo o vetor $\vec{v}_{\text{first-reducing}}$. O boxplot para o tamanho do vetor final de características $\vec{v}_I$, representado pelo rótulo “Final”, também é exibido.



Fonte: Elaboração do autor

Para analisar detalhadamente o efeito da redução de dimensionalidade em cada modelo de rede neural pré-treinado considerado, as duas partes do Gráfico 5 apresentam um gráfico de dispersão que associa o tamanho do vetor final de características $\vec{v}_I$ com a acurácia alcançada por diferentes versões de cada rede neural, destacando o uso ou ausência de dados aumentados, referindo-se ao conjunto de teste e ao conjunto de validação separadamente. Os gráficos destacam o uso de aumento de dados, mostrado em vermelho, e sua ausência, mostrada em azul. Os gráficos referentes às acurácias no conjunto de teste são apresentados à esquerda, enquanto os do conjunto de validação são apresentados à direita. Baseando-se nas ilustrações, alguns padrões se destacam e são discutidos a seguir:

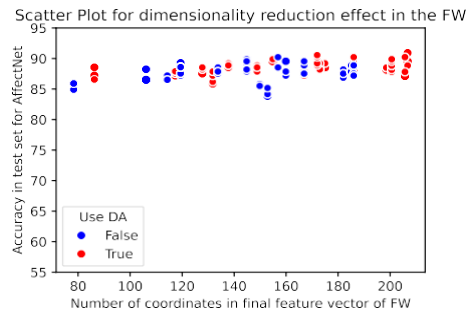
- Entre os modelos baseados em CNN, que geralmente apresentam vetores de características com tamanhos variando de 10 a 50, o AffectNet exibe a representação mais complexa, com vetores variando de 80 a 200 coordenadas. De maneira semelhante, as redes *transformers* também requerem vetores maiores, especificamente entre 100 e 300 coordenadas, para definir mais variância na representação das amostras durante as etapas de redução de dimensionalidade.
- É notável que há similaridade de desempenho entre os subconjuntos em algumas redes, mas não em outras, já que elas apresentam melhor desempenho em um conjunto do que no outro. Por exemplo, analisando o AffectNet, é possível observar que, no caso do conjunto de teste, a acurácia de todas as versões está em torno de 85% a 90%, enquanto no conjunto de validação, a acurácia de todas as versões gira em torno de 80%. Essa queda de desempenho também é observada em *transformers*. Contudo, algumas redes exibiram padrões mais consistentes entre os conjuntos de avaliação. Por exemplo, o AlexNet, apesar de mostrar desempenho inferior em relação aos outros, manteve uma acurácia média próxima de 75% em ambos os conjuntos. Um padrão semelhante pode ser observado nas outras CNNs. A acurácia das versões da ResNet-50 está fortemente associada ao tamanho do vetor de características, já que, quanto mais coordenadas o *framework* utiliza, maior é a acurácia do modelo.
- Na maioria dos modelos analisados, o uso de aumento de dados resultou em mais coordenadas sendo usadas no vetor de características. De fato, exceto pelo AlexNet, pode-se notar um acúmulo de círculos vermelhos no lado esquerdo dos gráficos. No

caso do AlexNet, é evidente que as versões que consideram o aumento de dados formaram clusters concentrados. Isso indica que o aumento de dados adicionou variabilidade às amostras, uma vez que o SVD considera esse fator em sua projeção.

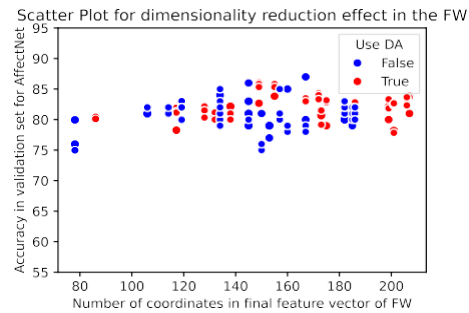
8.3 Comparação com o Estado-da-Arte

Para demonstrar a competitividade do *framework* proposto, os resultados obtidos foram comparados com os valores reportados por trabalhos que definem o estado-da-arte atual para o benchmark das faces de crianças com autismo e crianças típicas (TD) consideradas. Assim, a Tabela 4 apresenta as métricas de desempenho de classificação usadas neste trabalho, as quais devem ser comparadas com as de outros estudos relevantes sobre o tema.

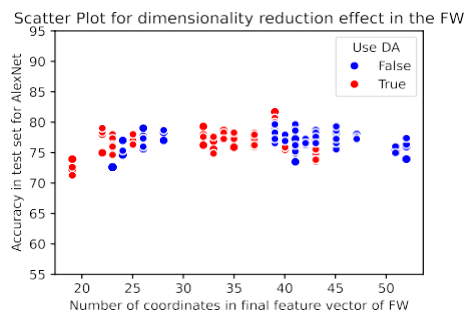
Gráfico 6 Gráfico de dispersão representando o número de coordenadas no vetor de características $\vec{v}_l$ em relação à acurácia alcançada no *framework* para cada rede pré-treinada (Parte 1).



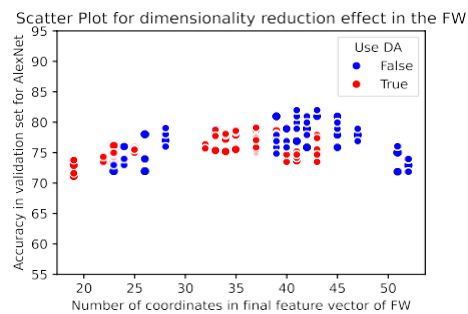
(a) AffectNet (Teste).



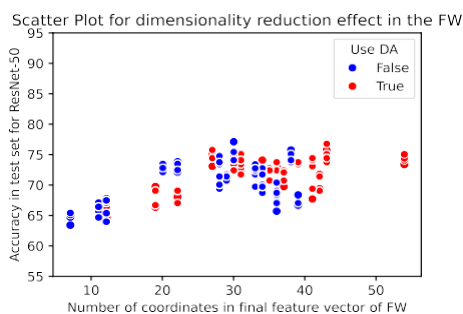
(b) AffectNet (Validação).



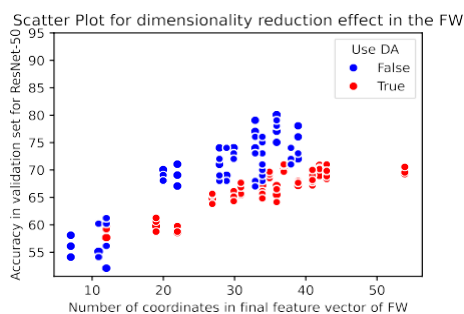
(c) AlexNet (Teste).



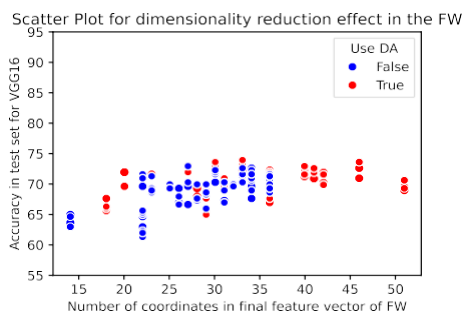
(d) AlexNet (Validação).



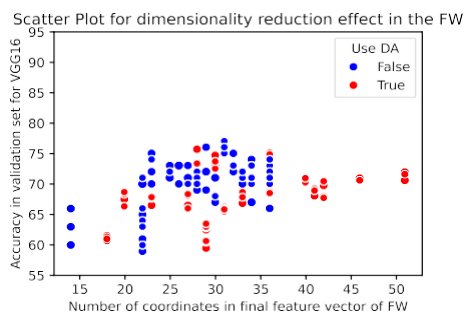
(e) ResNet-50 (Teste).



(f) ResNet-50 (Validação).



(g) VGG16 (Teste).

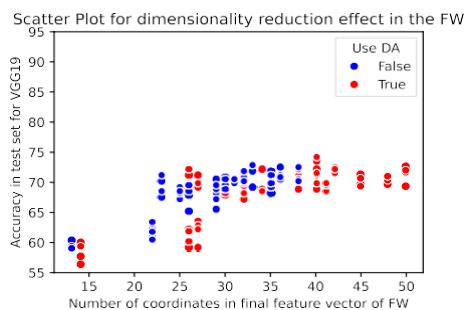


(h) VGG16 (Validação).

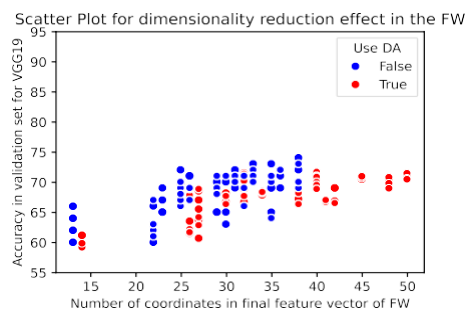
Fonte: Elaboração do autor

Gráfico 6 (continuação) Gráfico de dispersão representando o número de coordenadas no vetor de características

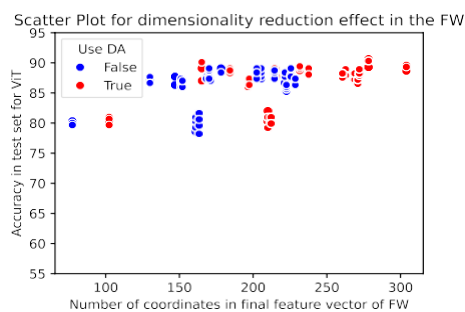
$\vec{v}_l$ em relação à precisão alcançada no *framework* para cada rede pré-treinada.



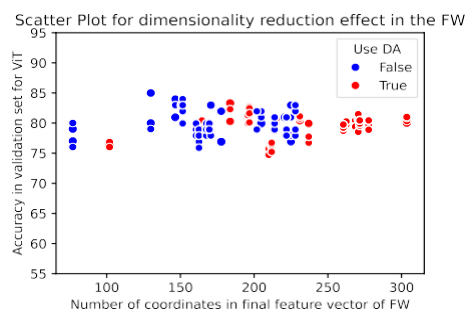
(i) VGG19 (Teste).



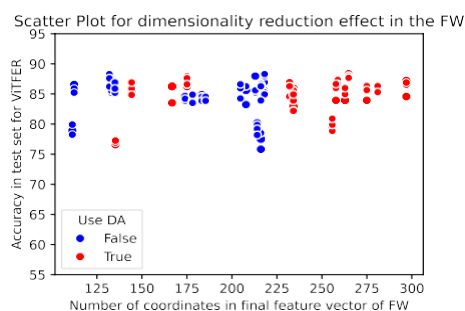
(j) VGG19 (Validação).



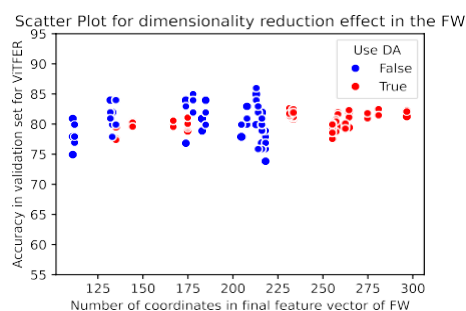
(k) ViT (Teste).



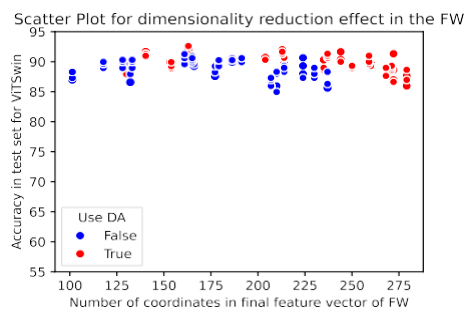
(l) ViT (Validação).



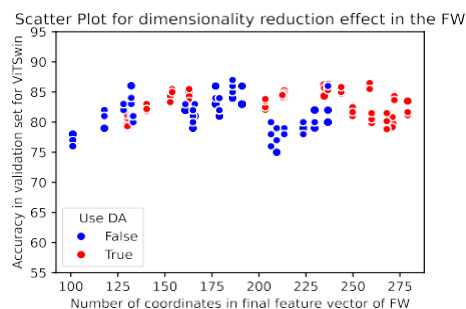
(m) ViTFER (Teste).



(n) ViTFER (Validação).



(o) ViTSwin (Teste).



(p) ViTSwin (Validação).

Fonte: Elaboração do autor

Tabela 4 Comparação das métricas de desempenho entre o método proposto e outras abordagens da literatura. O símbolo “-” indica que o valor não foi reportado no respectivo trabalho. Os valores em negrito destacam o melhor resultado para cada métrica.

Algorithm	ACC	F1	AUC	Recall	Precision	Specificity	FP	FN	TP	TN
MobileNet-V1 [76]	90.67	90.67	90.67	90.67	-	90.67	-	-	-	-
EfficientNetB1/Xception [77]	-	-	96.63	88.46	94.66	94.07	8	18	142	132
Xception [81]	92.01	-	96.25	90.97	90.97	-	12	12	138	138
Xception [85]	91.0	-	-	88.0	-	94.0	35	21	129	115
MobileNet [19]	91.0	92.0	-	92.0	90.47	-	12	12	138	138
VGG [87]	91.0	-	-	-	-	-	-	-	-	-
CNN [88]	92.31	91.54	96.95	93.45	89.72	-	-	-	-	-
Proposed	92.67	92.81	95.29	94.67	91.03	90.67	14	8	142	136

Fonte: Elaboração do autor

Analisando os resultados, o método proposto se destaca em várias métricas. Em termos de ACC, o valor de 92,67% alcançado pelo *framework* desenvolvido é o mais alto entre os métodos considerados, demonstrando a robusta capacidade do *framework* de classificar instâncias corretamente. O valor de F1 do método proposto de 92,81% também é o mais alto, indicando o melhor equilíbrio entre precisão e recall. Além disso, o valor de Recall obtido, 94,67%, supera todos os outros métodos, destacando que o *framework* proposto é o mais eficaz na detecção de verdadeiros positivos. Isso é crucial neste contexto, pois a detecção automática correta dos casos em que o indivíduo analisado possui o Transtorno do Espectro Autista (TEA) é central para o tema.

Embora o método proposto não tenha alcançado o maior AUC, com um valor de 95,29%, métodos de Rahman e Subashini (2022), usando EfficientNetB1/Xception, Alam et al. (2022), usando Xception, e Rabbi et al. (2021), usando CNN, apresentaram valores mais altos, variando de 0,96% a 1,66% acima. No entanto, a acurácia e o valor de F1 para EfficientNetB1/Xception não são conhecidos. Enquanto o valor de Recall do método proposto é 6,21% maior, sua *Precision* é 3,63% menor, resultando em 6 falsos positivos a mais, 10 falsos negativos a menos, o mesmo número de verdadeiros positivos e 4 verdadeiros negativos a mais. Além disso, o Xception de Alam et al. (2022) supera o AUC do método proposto por 0,96%,

mas apresenta menor ACC, com uma diferença de 0,66%, e valores de Precision e Recall mais baixos, com valores competitivos de FP, FN, TP e TN. Por fim, a CNN de Rabbi et al. (2021) alcança o maior AUC entre as técnicas consideradas. No entanto, comparado ao método proposto, apresenta valores mais baixos em todas as outras métricas, o que significa que o *framework* desenvolvido continua competitivo.

Em termos de *Pecision* e *Specifity*, o método proposto alcança valores altos de 91,03% e 90,67%, respectivamente, sendo superado apenas pelo EfficientNetB1/Xception, que tem uma vantagem de 3,63% e 3,4%. No entanto, como observado, o *framework* proposto se destaca quando outras métricas são consideradas. Além disso, o *framework* proposto teve apenas 8 casos em que um indivíduo com TEA foi classificado erroneamente como TD, o menor número de falsos negativos entre todos os métodos comparados. Além disso, o número de indivíduos com TEA corretamente identificados pelo modelo foi 142, o melhor desempenho nesta métrica. Isso reflete a alta sensibilidade do método proposto, que foi capaz de detectar a maioria dos casos positivos.

Também é importante destacar o trabalho de Shahzad et al. (2024), que, ao melhor conhecimento dos autores, apresenta as métricas mais altas do estado da arte em termos de ACC, *Precision*, *Recall* e *F1*, com valores de 96,50%, 96,54%, 96,50% e 96,49%, respectivamente. Esses resultados são entre 1,83% e 5,51% superiores aos alcançados pelo *framework* proposto. No entanto, dois pontos devem ser observados. Primeiro, o benchmark considerado por esses autores não é o mesmo utilizado nas comparações apresentadas aqui, pois o benchmark neste trabalho inclui um conjunto de testes com 300 amostras, enquanto Shahzad et al. (2024) usaram um conjunto de testes com 200 amostras. Além disso, este trabalho considera apenas um modelo de rede pré-treinada por vez no *framework*, enquanto a abordagem de Shahzad et al. (2024) emprega a hibridização e concatenação de múltiplos modelos pré-treinados em um sistema de *Attention Learning*.

8.4 Análise de Tempo e Complexidade

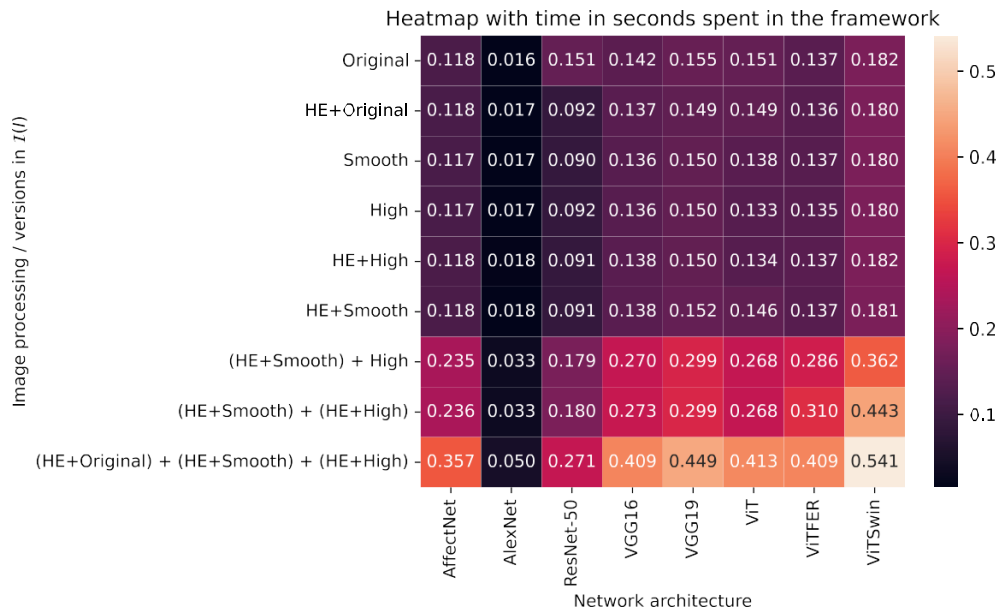
Como o método proposto é definido por uma sequência de etapas de processamento de imagem, extração de características e aprimoramento de características, a complexidade do *framework* é uma função das complexidades das técnicas que definem o conjunto F , a técnica $HE(\cdot)$, o modelo f_Φ , a estratégia de redução de dimensionalidade $DR(\cdot)$, a função de escala $SCALE(\cdot)$ e o modelo de classificação definido. Especificamente, dada uma imagem de teste para o *framework* com todos os parâmetros definidos e com o modelo de classificação já estabelecido, para calcular um diagnóstico, o método precisará filtrar essa imagem através de todos os n_F filtros de F , corrigir problemas de iluminação com $HE(\cdot)$ para as imagens I , $F_1(I)$, ..., $F_{n_F}(I)$, extrair características de todas as imagens geradas $(2n_F + 2)$ usando a rede pré-treinada f_Φ , projetar todas essas características por tipo de imagem usando a primeira estratégia de redução de dimensionalidade definida pelas funções $DR(\cdot)$ pré-configuradas, concatenar e projetar o vetor de características resultante com a segunda estratégia de redução de dimensionalidade $DR(\cdot)$, escalar o vetor usando a função $SCALE(\cdot)$, e finalmente, utilizar o classificador já treinado para estabelecer um diagnóstico. Assim, a ordem de complexidade para o *framework* proposto é dada por O_{Proposed} na Equação (7)

$$O_{\text{Proposed}} = \left(\sum_{i=1}^{n_F} O_{F_i} \right) + (n_F + 1) O_{HE(\cdot)} + (2n_F + 2) [O_{f_\Phi} + O_{DR(\cdot) \text{ first-reducing}}] + O_{DR(\cdot) \text{ second-reducing}} + O_{SCALE(\cdot)} + O_{\text{Classifier}}, \quad (7)$$

onde, O_{F_i} representa a complexidade de cada filtro $F_i \in F$, $O_{HE(\cdot)}$ é a complexidade da técnica de equalização de histograma, O_{f_Φ} corresponde a complexidade da extração de características usando a rede pré-treinada Φ , $O_{DR(\cdot) \text{ first-reducing}}$ é a complexidade das projeções realizadas durante a primeira etapa de redução de dimensionalidade, $O_{DR(\cdot) \text{ second-reducing}}$ representa a complexidade da segunda etapa de redução de dimensionalidade, $O_{SCALE(\cdot)}$ é a complexidade da função de escala, e $O_{\text{Classifier}}$ é a complexidade da predição feita pelo modelo treinado.

Vale ressaltar que, na prática, as estratégias de redução de dimensionalidade escolhidas para os experimentos são baseadas em SVD e, portanto, suas complexidades, $O_{DR()}_{\text{first-reducing}}$ e $O_{DR()}_{\text{second-reducing}}$, são polinomiais e contribuem pouco para a complexidade geral do modelo. O mesmo se aplica à função de escala. Quanto à complexidade do classificador treinado, como foi utilizado um SVM-RBF, ela depende da dimensão do vetor de características, que tende a ser reduzida pelas duas etapas dedicadas a isso no *framework*, e pelo número de vetores de suporte, que depende do tamanho do conjunto de dados de treinamento. Assim, o maior componente que adiciona à complexidade do método será determinado pela função de extração de características e pela etapa de processamento de imagem, ou seja, pelas funções de equalização de histograma e filtragem para visualizar esse efeito, o Gráfico 7 apresenta o tempo médio em segundos para obter uma classificação para uma imagem de rosto, considerando as redes pré-treinadas usadas nos experimentos e as seguintes técnicas de processamento de imagem: nenhuma (Original); equalização de histograma sem filtragem (HE+Original); filtragem passa-baixa (Smooth); filtragem passa-alta (High); filtragem passa-alta com correção de iluminação (HE+High); filtragem passa-baixa com correção de iluminação (HE+Smooth); filtragem passa-baixa com correção de iluminação e filtragem passa-alta ((HE+Smooth)+High); filtragem passa-baixa com correção de iluminação e filtragem passa-alta com correção de iluminação ((HE+Smooth)+(HE+High)); imagem sem filtragem com correção de iluminação (HE+Original); e filtragem passa-baixa com correção de iluminação e filtragem passa-alta com correção de iluminação ((HE+Original)+(HE+Smooth)+(HE+High)).

Gráfico 7 Mapa de calor mostrando o tempo gasto em segundos para criar uma versão da imagem e extrair a característica com uma arquitetura de rede específica.



Fonte: Elaboração do autor

No mapa de calor, é evidente que o tempo computacional é principalmente determinado pelo tempo de extração de características de f_Φ , pois algumas colunas aparecem mais escuras do que outras, como a coluna correspondente ao AlexNet. Além disso, quanto mais funções de filtragem são consideradas, maior o tempo computacional necessário para definir uma classificação, já que as linhas inferiores tendem a ser representadas por células com cores mais claras. De fato, esse aumento de tempo não é apenas devido ao custo introduzido pelas funções de filtragem, mas também pela geração de imagens adicionais que devem ser representadas por f_Φ e projetadas durante a primeira etapa de redução de dimensionalidade.

8.5 Limitações

A versatilidade do método proposto pode levar a altos custos computacionais, como evidenciado na Equação (7). Assim, embora um número maior de funções de filtragem definidas em F possa aprimorar a representação das imagens consideradas, isso também pode aumentar a complexidade do método. Além disso, a definição ampla do *framework* exige uma parametrização cuidadosa para garantir seu desempenho ideal. De fato, como mostrado nesta seção, as métricas de desempenho do método dependem fortemente do modelo pré-treinado considerado e das estratégias de aprimoramento de imagem empregadas.

Por fim, o benchmark utilizado nos experimentos, embora seja um dos mais respeitados no estado da arte possivelmente por ser o único conjunto de dados de imagens estáticas de acesso aberto, como mostrado na Tabela 4, apresenta algumas inconsistências. Entre elas, destacam-se expressões faciais variadas, o uso de edições em Photoshop, sorrisos acompanhados de caretas e até a presença de rostos de crianças com síndromes diferentes do Transtorno do Espectro Autista (TEA). Tais inconsistências decorrem principalmente do fato de que o benchmark é composto por imagens coletadas automaticamente e conta com documentação limitada disponível.

9 Conclusões

Este estudo foi originalmente desenvolvido como parte de um projeto de Iniciação Científica sob orientação do Prof. Dr. Rodrigo Conalgo Contreras, no ano de 2024 e apresenta dois avanços significativos voltados para aprimorar a detecção do Transtorno do Espectro Autista (TEA) por meio da análise automática de imagens faciais estáticas de crianças. O primeiro avanço concentra-se no desenvolvimento de uma estrutura que incorpora características extraídas por modelos pré-treinados a partir de imagens aprimoradas por correção de iluminação e diversas funções de filtragem. Essas características são então reduzidas por meio de técnicas de projeção dupla e ajustadas

utilizando uma estratégia de escalonamento, o que permite uma maior capacidade de representação das imagens e, conseqüentemente, melhora a acurácia do classificador. Essa estrutura abrange diversos processos, como aumento sintético de imagens, enriquecimento e simplificação da representação baseada em vetores, redução de dimensionalidade, normalização dos vetores de características e treinamento do classificador. Cada uma dessas etapas foi analisada na seção de experimentação e demonstrou ser importante para o aprimoramento das diversas métricas avaliadas. Em termos quantitativos, a estrutura conseguiu aumentar a acurácia de modelos pré-treinados em até 8 pontos percentuais, como observado no caso da rede ResNet-50 para o conjunto de validação.

A segunda inovação reside na ampla experimentação conduzida com mais de 1000 configurações práticas da estrutura proposta. Cada componente do *framework* foi cuidadosamente avaliado, com o objetivo de destacar seus pontos fortes e limitações em diversos cenários. Por meio dessa análise, observou-se que a etapa de aumento de dados mostrou-se eficaz quando aplicada ao conjunto de dados final, enquanto a etapa de aprimoramento de imagens apresentou desempenho consistente em todas as configurações testadas. Além disso, a etapa de redução de dimensionalidade foi capaz de comprimir a representação de características do modelo de melhor desempenho, o ViTswin, para um vetor com apenas 163 dimensões. As funções de escalonamento também contribuíram com melhorias sutis nas métricas avaliadas na maioria dos casos. A estrutura proposta foi ainda comparada com estudos consolidados utilizando o mesmo conjunto de dados, demonstrando resultados competitivos, alcançando a maior acurácia e desempenho sólido em outras métricas — mesmo adotando um design de modelo mais simplificado. Por fim, o método proposto obteve os melhores resultados na maioria das métricas em comparação com os trabalhos disponíveis na literatura, atingindo 92,67% de acurácia no conjunto de teste, o que representa uma melhoria absoluta de 0,36%.

Trabalhos futuros buscarão conduzir experimentos utilizando parametrizações mais sofisticadas dentro da estrutura proposta. Por exemplo, estratégias refinadas de fusão de características poderão substituir a concatenação direta de vetores atualmente aplicada após a primeira etapa de redução de dimensionalidade. Além disso, a utilização de múltiplos modelos de redes pré-treinadas poderá aprimorar ainda mais a representação de características,

contribuindo para a melhoria das métricas de acurácia do modelo. Considerando que a estrutura proposta foi projetada com configurações generalizáveis e adequadas para problemas semelhantes, desenvolvimentos futuros também explorarão sua adaptabilidade em outras aplicações.

Referências

AHMED, M.; SERAJ, R.; ISLAM, S. M. S. The k-means algorithm: **A comprehensive survey and performance evaluation**. *Electronics*, MDPI, v. 9, n. 8, p. 1295, 2020.

AKTER, T. et al. **Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage**. *Brain sciences*, MDPI, v. 11, n. 6, p. 734, 2021.

ALAM, M. S. et al. **Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach**. *Bioengineering*, MDPI, v. 9, n. 11, p. 710, 2022.

ALDRIDGE, K. et al. **Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes**. *Molecular autism*, Springer, v. 2, p. 1–12, 2011.

ALI, A. et al. **Video-based behavior understanding of children for objective diagnosis of autism**. In: *VISAPP 2022-17th International Conference on Computer Vision Theory and Applications*. [S.l.: s.n.], 2022.

ALKAHTANI, H.; ALDHYANI, T. H. H.; ALZAHRANI, M. Y. **Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks**. *Applied Sciences*, v. 13, n. 8, 2023. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/13/8/4855>>.

ALSAADE, F. W.; ALZAHRANI, M. S. **[retracted] classification and detection of autism spectrum disorder based on deep learning algorithms.** *Computational Intelligence and Neuroscience*, Wiley Online Library, v. 2022, n. 1, p. 8709145, 2022.

ARUMUGAM, S. R. et al. **A deep convolutional neural network based detection system for autism spectrum disorder in facial images.** In: IEEE. *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. [S.l.], 2021. p. 1255– 1259.

ATYABI, A. et al. **Stratification of children with autism spectrum disorder through fusion of temporal information in eye-gaze scan-paths.** *ACM Transactions on Knowledge Discovery from Data*, ACM New York, NY, v. 17, n. 2, p. 1–20, 2023.

AUGE', P. et al. **Global sensory features are linked to executive and attentional impairments in autism spectrum disorders.** *Journal of Autism and Developmental Disorders*, Springer, p. 1–9, 2024.

AYYARAO, T. S. et al. **War strategy optimization algorithm: a new effective metaheuristic algorithm for global optimization.** *IEEE Access*, IEEE, v. 10, p. 25073–25105, 2022.

BANIRE, B. et al. **Face-based attention recognition model for children with autism spectrum disorder.** *Journal of Healthcare Informatics Research*, Springer, v. 5, n. 4, p. 420–445, 2021.

BAXTER, A. J. et al. **The epidemiology and global burden of autism spectrum disorders.** *Psychological medicine*, Cambridge University Press, v. 45, n. 3, p. 601–613, 2015.

BEKELE, E. et al. **Understanding how adolescents with autism respond to facial expressions in virtual reality environments.** *IEEE transactions on visualization and*

computer graphics, IEEE, v. 19, n. 4, p. 711–720, 2013.

BHANDAGE, V. et al. **Autism spectrum disorder classification using adam war strategy optimization enabled deep belief network**. *Biomedical Signal Processing and Control*, v. 86, p. 104914, 2023. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1746809423003476>>.

BISHOP, S. L.; LORD, C. **Commentary: Best practices and processes for assessment of autism spectrum disorder—the intended role of standardized diagnostic instruments**. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 64, n. 5, p. 834–838, 2023.

BONE, D. et al. **Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion**. *Journal of Child Psychology and Psychiatry*, v. 57, n. 8, p. 927–937, 2016. Disponível em: <<https://acamh.onlinelibrary.wiley.com/doi/abs/10.1111/jcpp.12559>>.

BRADSKI, G. *The OpenCV Library*. 2000. Disponível em: <<https://github.com/opencv/opencv-python>>.

CAO, Z. et al. **Realtime multi-person 2d pose estimation using part affinity fields**. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 7291–7299.

CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. **Computer-aided diagnosis in the era of deep learning**. *Medical physics*, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020.

CHO, S. et al. **Automatic detection of autism spectrum disorder in children using acoustic and text features from brief natural conversations**. In: *Interspeech*. [S.l.: s.n.], 2019. p. 2513–2517.

CHOLLET, F. **Xception: Deep learning with depthwise separable convolutions**. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1251–1258.

CONSTANTINO, J. N. **Social responsiveness scale**. In: *Encyclopedia of autism spectrum disorders*. [S.l.]: Springer, 2021. p. 4457–4467.

CONTRERAS, R. C. et al. **A new multi-filter framework with statistical dense sift descriptor for spoofing detection in fingerprint authentication systems**. In: RUTKOWSKI, L. et al. (Ed.). *Artificial Intelligence and Soft Computing*. Cham: Springer International Publishing, 2021. p. 442–455. ISBN 978-3-030-87897-9.

CONTRERAS, R. C. et al. **A new multi-filter framework for texture image representation improvement using set of pattern descriptors to fingerprint liveness detection**. *IEEE Access*, v. 10, p. 117681–117706, 2022.

DAWUD, A. M.; YURTKAN, K.; OZTOPRAK, H. **Application of deep learning in neuroradiology: brain hemorrhage classification using transfer learning**. *Computational Intelligence and Neuroscience*, Wiley Online Library, v. 2019, n. 1, p. 4629859, 2019.

DEVELOPERS, T. Tensorflow. *Zenodo*, 2022.

DONVAN, J.; ZUCKER, C. **In a different key: The story of autism**. [S.l.]: Crown, 2016.

DUAN, H. et al. **A dataset of eye movements for the children with autism spectrum disorder**. In: *Proceedings of the 10th ACM Multimedia Systems Conference*. [S.l.: s.n.], 2019. p. 255–260.

DUAN, H. et al. **Revisiting skeleton-based action recognition**. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.],

2022. p. 2969–2978.

DUFFY, F. H.; ALS, H. **Autism, spectrum or clusters? an eeg coherence study.** *BMC neurology*, Springer, v. 19, p. 1–13, 2019.

DUNDI, U. R. et al. **Computer vision aided machine learning framework for detection and analysis of arm flapping stereotypic behavior exhibited by the autistic child.** In: R, S. C. K. et al. (Ed.). *Computational Intelligence in Data Science*. Cham: Springer Nature Switzerland, 2023. p. 203–217. ISBN 978-3-031-38296-3.

DUPAUL, G. J. et al. **ADHD rating scale? 5 for children and adolescents: checklists, norms, and clinical interpretation.** [S.l.]: Guilford Publications, 2016.

EASSON, A. K.; FATIMA, Z.; MCINTOSH, A. R. **Functional connectivity-based subtypes of individuals with and without autism spectrum disorder.** *Network Neuroscience*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 3, n. 2, p. 344–362, 2019.

FAROOQ, M. S. et al. **Detection of autism spectrum disorder (asd) in children and adults using machine learning.** *scientific reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 9605, 2023.

GILLBERG, C. et al. **The asperger syndrome (and high-functioning autism) diagnostic interview (asdi): a preliminary study of a new structured clinical interview.** *Autism*, Sage Publications Sage CA: Thousand Oaks, CA, v. 5, n. 1, p. 57–66, 2001.

GIOIA, G. A. et al. **Profiles of everyday executive function in acquired and developmental disorders.** *Child neuropsychology*, Taylor & Francis, v. 8, n. 2, p. 121–137, 2002.

GREENACRE, M. et al. **Principal component analysis**. *Nature Reviews Methods Primers*, Nature Publishing Group UK London, v. 2, n. 1, p. 100, 2022.

GUHA, T. et al. **A computational study of expressive facial dynamics in children with autism**. *IEEE transactions on affective computing*, IEEE, v. 9, n. 1, p. 14–20, 2016.

HAMMOND, P. et al. **Face–brain asymmetry in autism spectrum disorders**. *Molecular psychiatry*, Nature Publishing Group, v. 13, n. 6, p. 614–623, 2008.

HAN, J. et al. **Affective computing of children with autism based on feature transfer**. In: IEEE. *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*. [S.l.], 2018. p. 845–849.

HARM, M.; HOPE, M.; HOUSEHOLD, A. American psychiatric association, 2013, diagnostic and statistical manual of mental disorders, 5th edn, washington, dc: American psychiatric association anderson, j, sapey, b, spandler, h (eds), 2012, distress or disability?, lancaster: Centre for disability research, www.lancaster.ac.uk. *Arya*, v. 347, p. 64, 2013.

HE, K. et al. **Deep residual learning for image recognition**. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016.

HE, K. et al. **Deep residual learning for image recognition**. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

HOWARD, A. G. **Mobilenets: Efficient convolutional neural networks for mobile vision applications**. *arXiv preprint arXiv:1704.04861*, 2017.

HUA, Y.; GUO, J.; ZHAO, H. **Deep belief networks and deep learning**. In: IEEE. *Proceedings of 2015 international conference on intelligent computing and internet of things*. [S.l.], 2015. p. 1–4.

HYDE, K. K. et al. **Applications of supervised machine learning in autism spectrum disorder research: a review.** *Review Journal of Autism and Developmental Disorders*, Springer, v. 6, p. 128–146, 2019.

JAHANARA, S.; PADMANABHAN, S. **Detecting autism from facial image.** *International Journal of Advance Research, Ideas and Innovations in Technology*, v. 7, n. 2, p. 219–225, 2021.

JANVIER, D. et al. **Brief report: Examining test-retest reliability of the autism diagnostic observation schedule (ados-2) calibrated severity scores (css).** *Journal of Autism and Developmental Disorders*, Springer, p. 1–7, 2022.

KIM, H. E. et al. **Transfer learning for medical image classification: a literature review.** *BMC medical imaging*, Springer, v. 22, n. 1, p. 69, 2022.

KINGMA, D. P. **Adam: A method for stochastic optimization.** *arXiv preprint arXiv:1412.6980*, 2014.

KOJOVIC, N. et al. **Using 2d video-based pose estimation for automated prediction of autism spectrum disorders in young children.** *Scientific Reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 15069, 2021.

KORA, P. et al. **Transfer learning techniques for medical image analysis: A review.** *Biocybernetics and Biomedical Engineering*, Elsevier, v. 42, n. 1, p. 79–107, 2022.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **Imagenet classification with deep convolutional neural networks.** *Advances in neural information processing systems*, v. 25, 2012.

LECAVALIER, L. **An evaluation of the gilliam autism rating scale.** *Journal of autism*

and developmental disorders, Springer, v. 35, p. 795–805, 2005.

LELORD, G.; BARTHE'LE'MY, C. *Echelle d'evaluation des comportements autistiques*. [S.l.]: Etablissement d'applications psychotechniques, 1995.

LEO, M. et al. **Towards the automatic assessment of abilities to produce facial expressions: The case study of children with asd**. In: *20th Italian National Conference on Photonic Technologies (Fotonica 2018)*. [S.l.: s.n.], 2018. p. 1–4.

LEO, M. et al. **Computational assessment of facial expression production in asd children**. *Sensors*, v. 18, n. 11, 2018. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/18/11/3993>>.

LEO, M. et al. **Computational analysis of deep visual data for quantifying facial expression production**. *Applied Sciences*, v. 9, n. 21, 2019. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/9/21/4542>>.

LIU, W.; LI, M.; YI, L. **Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework**. *Autism Research*, v. 9, n.8, p.888–898, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.1615>>.

LORD, C. **Autism diagnostic observation schedule**. (*No Title*), 1999.

LORD, C. et al. **Autism spectrum disorder**. *Nature reviews Disease primers*, Nature Publishing Group, v. 6, n. 1, p. 1–23, 2020.

LORD, C. et al. **Autism spectrum disorder**. *The lancet*, Elsevier, v. 392, n. 10146, p. 508–520, 2018.

LUGARESI, C. et al. **Mediapipe: A framework for perceiving and processing reality**. In: *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*. [S.l.: s.n.], 2019. v. 2019.

MARCHETTI, A. et al. **Theory of mind in typical and atypical developmental settings: Some considerations from a contextual perspective**. In: *Reflective Thinking in Educational Settings: A Cultural Framework*. [S.l.]: Cambridge University Press, 2014. p. 102–136.

MARTINO, A. D. et al. **Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii**. *Scientific data*, Nature Publishing Group, v. 4, n. 1, p. 1–15, 2017.

MUMUNI, A.; MUMUNI, F. **Data augmentation: A comprehensive survey of modern approaches**. *Array*, Elsevier, v. 16, p. 100258, 2022.

NOBLE, W. S. **What is a support vector machine?** *Nature biotechnology*, Nature Publishing Group UK London, v. 24, n. 12, p. 1565–1567, 2006.

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. **Multiresolution gray-scale and rotation invariant texture classification with local binary patterns**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, 2002.

PAN, S. J.; YANG, Q. **A survey on transfer learning**. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.

PANDIAN, D. et al. **Detecting a child's stimming behaviours for autism spectrum disorder diagnosis using rgbpose-slowfast network**. In: IEEE. *2022 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2022. p. 3356–3360.

PANTELIS, P. C.; KENNEDY, D. P. **Deconstructing atypical eye gaze perception in autism spectrum disorder**. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–10, 2017.

PARK, K.-W.; CHO, S.-B. **A residual graph convolutional network with spatiotemporal features for autism classification from fmri brain images**. *Applied Soft Computing*, v. 142, p. 110363, 2023. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494623003812>>.

PARLETT-PELLERITI, C. M. et al. **Applications of unsupervised machine learning in autism spectrum disorder research: a review**. *Review Journal of Autism and Developmental Disorders*, Springer, v. 10, n. 3, p. 406–421, 2023.

PARMAR, A.; KATARIYA, R.; PATEL, V. **A review on random forest: An ensemble classifier**. In: SPRINGER. *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*. [S.l.], 2019. p. 758–763.

PASZKE, A. et al. **Pytorch: An imperative style, high-performance deep learning library**. *Advances in Neural Information Processing Systems*, v. 32, 2019. Disponível em: <<https://proceedings.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>>.

PEDREGOSA, F. et al. **Scikit-learn: Machine learning in Python**. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PIOSENKA, G. **Detect autism from a facial image**. <https://www.kaggle.com/cihan063/autism-image-data> accessed on 1 de dezembro de 2024, 2021.

PIOSENKA, G. **Detect autism from a facial image**.

<https://drive.google.com/drive/folders/1XQU0pluL0m3TilXqntano12d68peMb8A>, accessed on 1 de dezembro de 2024, 2021.

RABBI, M. F. et al. **A convolutional neural network model for early-stage detection of autism spectrum disorder**. In: *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. [S.l.: s.n.], 2021. p. 110–114.

RAHMAN, K. K. M.; SUBASHINI, M. M. **Identification of autism in children using static facial features and deep neural networks**. *Brain Sciences*, v. 12, n. 1, 2022. ISSN 2076-3425. Disponível em: <<https://www.mdpi.com/2076-3425/12/1/94>>.

RANI, P. **Emotion detection of autistic children using image processing**. In: *2019 Fifth International Conference on Image Information Processing (ICIIP)*. [S.l.: s.n.], 2019. p. 532–535.

RUTTER, M. et al. **Autism diagnostic interview-revised**. *Los Angeles, CA: Western Psychological Services*, v. 29, n. 2003, p. 30, 2003.

S, J. B. et al. **Detecting a child's stimming behaviours for autism spectrum disorder diagnosis using rgbpose-slowfast network**. In: *2022 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2022. p. 3356–3360.

SANDLER, M. et al. **Mobilenetv2: Inverted residuals and linear bottlenecks**. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 4510–4520.

SCHOPLER, E. et al. **Toward objective classification of childhood autism: Childhood autism rating scale (cars)**. *Journal of autism and developmental disorders*, Plenum Publishing Corp., 1980.

SHAHZAD, I. et al. **Enhancing asd classification through hybrid attention-based learning of facial features.** *Signal, Image and Video Processing*, Springer, p. 1–14, 2024.

SHAKER, E.; BAKER, M. R.; MAHMOOD, Z. **The impact of image enhancement and transfer learning techniques on marine habitat mapping.** *Gazi University Journal of Science*, Gazi University, v. 36, n. 2, p. 592–606, 2022.

SHUKLA, P. et al. **A deep learning frame-work for recognizing developmental disorders.** In: IEEE. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.], 2017. p. 705–714.

SILLERESI, S. et al. **Identifying language and cognitive profiles in children with asd via a cluster analysis exploration: Implications for the new icd-11.** *Autism Research*, v. 13, n.7, p. 1155–1167, 2020. Dispon'ível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.2268>>.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition.** *arXiv preprint arXiv:1409.1556*, 2014.

SQUIRES, M. et al. **Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment.** *Brain Informatics*, Springer, v. 10, n. 1, p. 10, 2023.

TAMILARASI, F. C.; SHANMUGAM, J. **Convolutional neural network based autism classification.** In: *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. [S.l.: s.n.], 2020. p. 1208–1212.

TAN, M. **Efficientnet: Rethinking model scaling for convolutional neural networks.** *arXiv preprint arXiv:1905.11946*, 2019.

TAO, Y.; SHYU, M.-L. **Sp-asdnet: Cnn-lstm based asd classification model using observer scanpaths**. In: *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. [S.l.: s.n.], 2019. p. 641–646.

TAWHID, M. N. A. et al. **A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg**. *Plos one*, Public Library of Science San Francisco, CA USA, v. 16, n. 6, p. e0253094, 2021.

TOO, E. C. et al. **A comparative study of fine-tuning deep learning models for plant disease identification**. *Computers and Electronics in Agriculture*, Elsevier, v. 161, p. 272–279, 2019.

UCHO[^]A, V. et al. **Data augmentation for face recognition with cnn transfer learning**. In: IEEE. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.], 2020. p. 143–148.

UDDIN, M. Z. et al. **Deep learning with image-based autism spectrum disorder analysis: A systematic review**. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 127, p. 107185, 2024.

WADHERA, T.; MAHMUD, M.; BROWN, D. J. **A deep concatenated convolutional neural network-based method to classify autism**. In: TANVEER, M. et al. (Ed.). *Neural Information Processing*. Singapore: Springer Nature Singapore, 2023. p. 446–458. ISBN 978-981-99-1648-1.

WANG, X.; WANG, K.; LIAN, S. **A survey on face data augmentation for the training of deep neural networks**. *Neural computing and applications*, Springer, v. 32, n. 19, p. 15503–15531, 2020.

WEITLAUF, A. S. et al. **Therapies for children with autism spectrum disorder: Behavioral interventions update.** 2014.

ZADEH, A. et al. **Convolutional experts constrained local model for 3d facial landmark detection.** In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops.* [S.l.: s.n.], 2017.

ZHENG, L.; GROVE, R.; EAPEN, V. **Spectrum or subtypes? a latent profile analysis of restricted and repetitive behaviours in autism.** *Research in Autism Spectrum Disorders*, Elsevier, v. 57, p. 46–54, 2019.

ZOPH, B. et al. **Learning transferable architectures for scalable image recognition.** In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* [S.l.: s.n.], 2018. p. 8697–8710.