

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

**ESTUDO E ANÁLISE DE TÉCNICAS DE VISÃO
COMPUTACIONAL PARA RECONHECIMENTO DE
EXPRESSÕES FACIAIS NO MONITORAMENTO INDIVIDUAL
VISANDO A PREVENÇÃO DE COLAPSOS AUTISTAS**

**STUDY AND ANALYSIS OF COMPUTATIONAL VISION
TECHNIQUES FOR FACIAL EXPRESSIONS RECOGNITION IN
INDIVIDUAL MONITORING AIMING TO PREVENT MELTDOWNS**

Jacqueline Lopes Silva¹, Italo Oliveira², Zabelita Hinselmann Topolniak³, Ana Beatriz Alvarez⁴

E-mail: jacqueline.silva@ifac.edu.br¹, italo.oliveira@sou.ufac.br²,
zabelitahinselmann@gmail.com³, anabe@ufac.br⁴

Instituição: Universidade Federal do Acre^{1,2,4}, Claretiano Centro Universitário³

Artigo submetido em 05/2021 e aceito em 10/2021

Resumo

Este artigo apresenta o estudo e análise de algoritmos de visão computacional para reconhecimento e identificação de expressões faciais. O processo de análise inicia com a realização de testes para medir a acurácia alcançada utilizando a base de dados FER2013, base que contém expressões padrão já conhecidas: Raiva, Nojo, Medo, Felicidade, Tristeza, Surpresa e Neutro. Os resultados alcançados mostram que a identificação de expressões faciais relacionadas aos colapsos autistas é promissora.

Palavras-chave: Reconhecimento de Expressões Faciais, Redes Neurais Convolucionais, Visão Computacional.

Abstract

This article presents the study and analysis of computer vision algorithms for recognition and identification of facial expressions. The analysis process starts with the performance of tests to measure the accuracy achieved using the FER2013 database, which contains standard expressions already known: Anger, Disgust, Fear, Happiness, Sadness, Surprise and Neutral. The results obtained show that the identification of facial expressions related to autistic meltdown is promising.

Keywords: Facial Expressions Recognition, Convolutional Neural Networks, Computer Vision.

1 INTRODUÇÃO

As emoções humanas possuem um papel fundamental na socialização humana. As expressões faciais, assim como outras formas de expressão de estados emocionais, são importantes vias para a comunicação (ANDRADE et

al., 2013). A expressão das emoções é de extrema relevância na interação social, uma vez que ao manifestar emoções, uma pessoa transmite e comunica a outra o que sente, seja de forma intencional ou não (ADRIANO; ARRIAGA, 2016).

Atualmente existe uma grande preocupação na inserção social dos indivíduos que apresentam o Transtorno do Espectro Autista (TEA). Define-se então, que o TEA caracteriza-se por uma condição no neurodesenvolvimento que causa uma deterioração significativa estando presente desde a infância do indivíduo, persistente do domínio da comunicação social e comportamental (atividades ou interesses repetitivos) (SOARES; CAVALCANTE NETO, 2015). Pessoas que apresentam esse distúrbio tem uma visão diferenciada, se privando do contato com a sociedade, apresentando alterações para se conectar com o meio social prejudicando significativamente o seu desenvolvimento.

Segundo Henriques (HENRIQUES, 2018), indivíduos com TEA apresentam dificuldades em diversas áreas de desenvolvimento, entre elas, a interação social, comunicação e processamento emocional. O indivíduo autista quando entra em um *Meltdown* (crise nervosa), rapidamente passa de modo passivo para agressivo por não ter meios de controlar suas emoções, por isso é necessário ter um acompanhamento adequado com táticas para estabelecer suporte ao indivíduo autista.

Por outro lado, a Visão Computacional é uma área de pesquisa que foca no desenvolvimento de técnicas e metodologias para emular a visão humana de maneira que o computador possa enxergar o mundo e realizar toma de decisões a partir de imagens capturadas do ambiente, usando para isso abordagens de Inteligência Computacional (SHAPIRO; STOCKMAN, 2001).

Neste artigo se apresenta um estudo inicial detalhado para análise de algoritmos de visão computacional que possibilitem a detecção da face, o reconhecimento facial e finalmente o reconhecimento de expressões de emoções faciais. Tudo isto visando a realização do monitoramento individual automático de pessoas com TEA. O monitoramento, quando realizado de maneira robusta e eficaz, possibilitará um acompanhamento não invasivo e a identificação de expressões faciais relacionadas com os possíveis *Meltdowns* que poderá auxiliar profissionais da área de saúde e/ou educação que atuam diretamente com indivíduos autistas.

Os estudos realizados e apresentados neste artigo fazem parte do trabalho de dissertação (SILVA, 2021) da primeira autora deste artigo e foram primordiais para a proposta de uma abordagem eficaz para a identificação automática das expressões faciais descrito detalhadamente em (SILVA et al., 2021). Também, a partir destes estudos e para testar a eficácia dos algoritmos em tempo real, foi idealizado e desenvolvido o módulo EmotiTEA, ferramenta descrita detalhadamente em (OLIVEIRA et al., 2021).

1.1 Trabalhos Relacionados

Dentro dos trabalhos antecedentes que utilizam técnicas inteligentes para o reconhecimento de padrões, alguns deles contribuem de forma significativa para o desenvolvimento realizado, dentro deles o trabalho de Jaques (DE OLIVEIRA; JAQUES, 2013) que apresenta duas abordagens para inferência de emoções por expressões faciais, sendo elas: o uso de classificadores (*Haar-like features*, Máquinas de Vetores de Suporte – SVM e redes neurais) e modelos psicológicos de classificação facial, como por exemplo o *Facial Action Coding System* (FACS). Já no trabalho do grupo de pesquisa de Balasubramanian (BALASUBRAMANIAN et al., 2019) é apresentada uma revisão cobrindo conjuntos de dados e algoritmos que são utilizados para reconhecimento de emoções faciais (FER), que vai desde SVM simples até complexas Redes Neurais Convolucionais (CNN). Os autores explicam os algoritmos através de artigos com aplicações em FER. O estudo conduzido por Lim (LIM et al., 2018) examina como a emoção humana pode ser reconhecida usando visão computacional. Algoritmos e técnicas são revisados pelos autores, entre eles o uso de SVM combinado com *Adaboost* ou com *Kernel* da Função Base Radial (*Kernel RBF*), algoritmos genéticos, Redes Neurais (RNN, CNN, *Multilayer Perceptron* (MPL) e k-NN), Redes Bayesianas (*Naive-Bayes* e *Gaussian-TAN*), *SURF*, entre outros.

No trabalho de Lasri (LASRI; SOLH; EL BELKACEMI, 2019) foi proposto um sistema de reconhecimento de emoções, a partir do rosto de alunos. O sistema consiste em três fases, sendo elas a detecção de rosto utilizando *Haar Cascade*, normalização e o reconhecimento de emoção através da utilização de uma Rede Neural Convolucional (CNN). Os autores utilizaram o banco de dados FER2013 com sete tipos de emoções. Os resultados obtidos mostraram que o

reconhecimento de emoções é viável na educação, e podem ajudar os professores a modificarem suas apresentações de acordo com as expressões demonstradas pelos alunos.

Considerando os trabalhos que focam na aplicação, o trabalho de Patnam (PATNAM et al., 2017) apresenta a utilização de técnicas de aprendizagem profunda para o reconhecimento de graves colapsos autistas (*Meltdowns*), por meio de gestos não verbais confiáveis. O sistema mostra bons resultados, notificando os pais ou responsáveis para que eles possam controlar a situação em menor tempo. Como técnica, os autores utilizaram um classificador treinado com imagens coletadas de vídeos e fontes confiáveis de internet com a maioria de gestos preditivos, para detecção de colapsos com mais precisão.

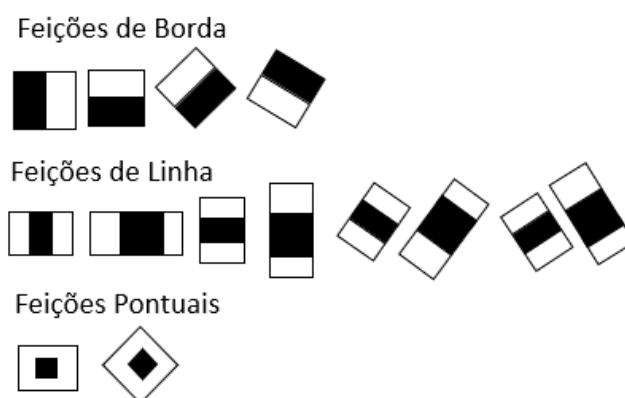
2 MARCO TEÓRICO

2.1 Detecção Facial

A detecção facial é umas das etapas essenciais na aplicação da visão computacional, e consiste em identificar e extrair faces de uma determinada imagem. Estas informações extraídas podem ser de tamanho, postura, posição, expressões, dentre outras, e são utilizadas pelo reconhecimento facial e de expressões de emoções faciais (GONZALEZ; WOODS, 2007).

2.1.1 Algoritmos de Detecção Facial

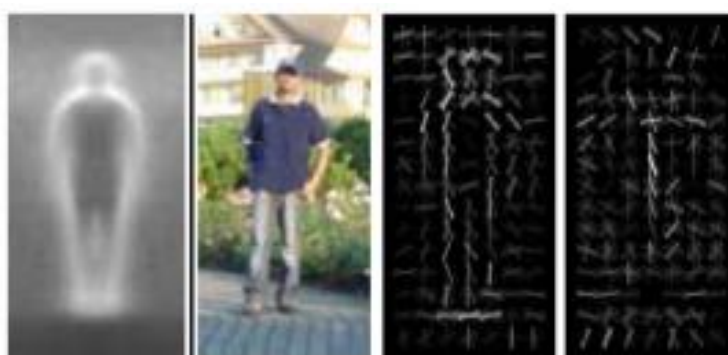
O algoritmo *Haar Cascade* (VIOLA; JONES, 2001), é uma abordagem baseada em *Machine Learning* (ML) que utiliza o algoritmo de aprendizado *Adaboost* (FREUND; SCHAPIRE, 1995). Este algoritmo utiliza filtros, os chamados *Haar Features*, que passam janela por janela da imagem, capturando valores que caracterizam um certo tipo de objeto. Esses filtros podem ser de linha, centrais ou de borda, como mostra a Figura 1. No final, uma pequena quantidade de características importantes é selecionada, e utilizada na etapa de detecção. O cálculo realizado pelo algoritmo *Haar Cascade* quando percorre pelas janelas é a diferença entre a soma dos *pixels* em duas regiões retangulares (VIOLA; JONES, 2004).

Figura 1: Haar Features

Fonte: Adaptado de (LIENHART; KURANOV; PISAREVSKY, 2003).

Histograma de Gradientes Orientados (do inglês, *Histogram of Oriented Gradients* - HOG), é um descritor de recursos utilizado em visão computacional para detecção e reconhecimento de objetos (DALAL; TRIGGS, 2005) que utiliza partes ou porções da imagem, operando em células locais, com invariância a transformações geométricas.

Este método tem como objetivo extrair informações que se referem à orientação das arestas existentes em uma imagem, sendo estas calculadas através de métodos de detecção de bordas (PANCERI et al., 2015), conforme apresenta a Figura 2.

Figura 2: Contornos Obtidos com HOG

Fonte: (DALAL; TRIGGS, 2005).

2.2 Reconhecimento Facial

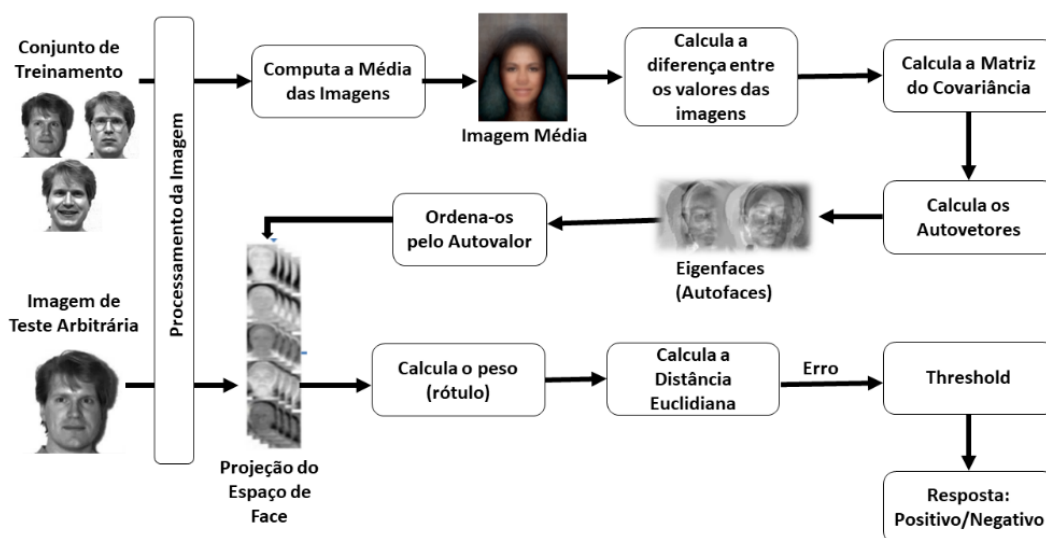
O reconhecimento facial é uma abordagem que funciona comparando características faciais selecionadas de uma determinada imagem com os rostos já existentes em um banco de dados, e tem o objetivo de verificar ou reconhecer

a identidade de uma pessoa com base em suas características. O reconhecimento facial pode ser dividido em 4 etapas: Aquisição de imagens, detecção facial, extração de características, e por último o reconhecimento facial, que se trata da verificação/identificação da face (GONZALEZ; WOODS, 2007).

2.2.1 Algoritmos de Reconhecimento Facial

Eigenfaces é um algoritmo de reconhecimento que define um espaço de recursos onde há a redução da dimensionalidade dos dados originais, tornando com isso o processo mais eficaz e menos custoso computacionalmente (BHELE; MANKAR, 2012). O fluxo básico de reconhecimento facial através do *Eigenfaces* pode ser visto através da Figura 3.

Figura 3: Fluxo básico do reconhecimento facial através do Algoritmo



Fonte: Adaptado de (CORREIA et al., 2014).

O método PCA (do inglês, *Principal Component Analysis*) - um dos métodos populares para seleção de recursos, e seu principal objetivo é encontrar os vetores que melhor explicam a distribuição das imagens de rosto em todo o espaço da imagem e produzir direções de projeção que maximizam a dispersão total em todas as classes, ou seja, em todas as imagens de todas as faces (BELHUMEUR; HESPANHA; KRIEGMAN, 1997).

A abordagem *Eigenface* transforma as faces em pequenos vetores de características, chamadas "*Eigenfaces*", e a partir delas pode-se reconstruir a

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

imagem original, através de uma soma ponderada entre elas onde cada uma possuirá um peso, conforme mostra a Figura 4. Este peso especifica em que medida a característica específica está presente na imagem original.

Figura 4: Vetores de Características *Eigenfaces*

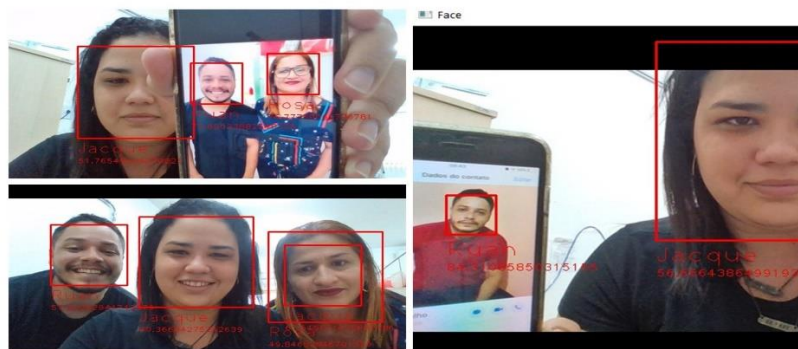


Fonte: (CHAO, 2007).

Fisherfaces é um algoritmo de reconhecimento facial que usa a matriz de transformação específica da classe, e não considera a iluminação uma característica tão importante quanto o método *Eigenfaces*, já que faz a remoção dos três primeiros componentes principais responsáveis pelas alterações na intensidade da luz. O algoritmo utiliza tanto o método PCA, quanto o método LDA (do inglês, *Linear Discriminant Analysis*) para redução de dimensionalidade (BELHUMEUR; HESPANHA; KRIEGMAN, 1997).

Como vantagem do método *Fisherfaces* pode-se citar a capacidade de tirar proveito das informações dentro das classes, com minimização da taxa de dispersão enquanto que maximiza a taxa entre as classes (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), ou seja, as distâncias entre as classes serão maiores, enquanto as distâncias entre as amostras da mesma classe são menores. Um exemplo de funcionamento do algoritmo *fisherfaces* já implementado conforme mostra Figura 5.

Figura 5: Reconhecimento Facial utilizando *Fisherfaces*



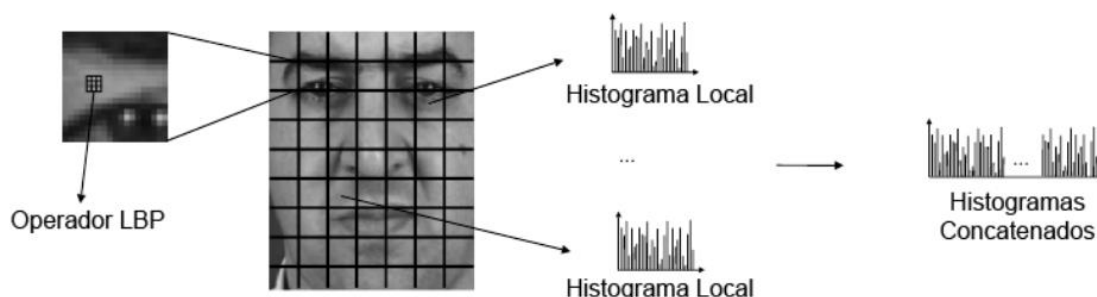
Fonte: Elaborado pelos autores (2021).

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

O LBPH (do inglês, *Local Binary Patterns Histograms*) é um operador de textura simples e eficiente que resume estruturas locais de imagens com eficiência, comparando cada *pixel* com seus pixels vizinhos e considera o resultado como um número binário (OMAR; TAWFIQ, 2019) e tornou-se uma boa técnica em termos de classificação de textura.

O LBP pode ser combinado com histogramas, representando as imagens como um vetor de dados simples. O algoritmo utiliza um operador LBP aprimorado, e divide uma imagem em várias sub-regiões e extrai o LBP de cada *pixel* da sub-região, estabelecendo um histograma estatístico do espectro característico do LBP (Figura 6), de modo que cada sub-região possa usar um histograma estatístico para descrever a imagem inteira por vários componentes estatísticos do histograma (TURK; PENTLAND, 1991).

Figura 6: LBPH



Fonte: Elaborado pelos autores (2021).

2.3 Reconhecimento de Expressões Faciais

Existem diversas formas de uma pessoa manifestar suas emoções, e mais variados ainda são os métodos computacionais para captar e reconhecer emoções transmitidas. A voz, as ações do usuário na interface com o sistema, as expressões faciais e os sinais fisiológicos são considerados os principais modos de reconhecimento de emoções (JAQUES; VICARI, 2005).

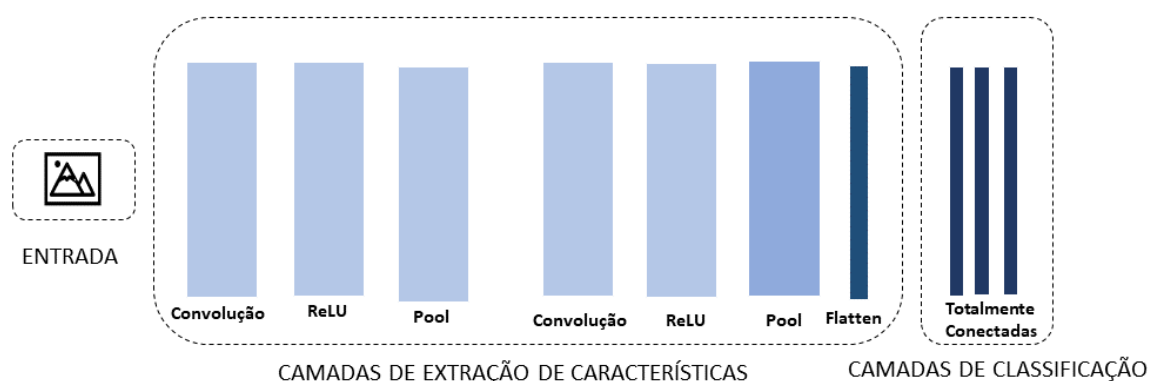
Algoritmos computacionais de reconhecimento de expressões emocionais, em ambientes controlados, podem obter sucesso entre 74% até 98% (SEBE et al., 2005), mas pode-se obter resultados ainda melhores se

combinado com mecanismos de reconhecimento facial e vocal, considerados como principais aspectos utilizados por uma pessoa para expressar e também para reconhecer emoções.

2.3.1 Algoritmos de Reconhecimento de Expressões faciais

As Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks - CNN*) são tipos de redes neurais artificiais que utilizam multicamadas complexas e profundas capazes de reconhecer elementos em imagens (IAN GOODFELLOW, YOSHUA BENGIO, 2015). São compostas de camadas de convolução, *Pooling*, *Flatten* e Camadas Totalmente Conectadas, conforme mostra Figura 7.

Figura 7: Exemplo de Camadas de uma rede Neural Convolucional



Fonte: Elaborado pelos autores (2021).

A principal camada das ConvNets é a camada de Convolução, que é uma transformação linear acompanhada de ativação não-linear. O objetivo desta camada é aplicar filtros nas imagens de entrada, e produzir uma saída (matrizes) que armazenam os pesos das conexões entre os neurônios (PARKHI; VEDALDI; ZISSERMAN, 2015). A camada de *Pooling* é responsável pela redução de dimensionalidade retendo somente as informações mais importantes, criando representações de entrada menores e mais gerenciáveis, e controlando o *overfitting* (ajuste excessivo aos dados de entrada).

Os resultados da camada de Convolução e da camada de *Pooling*, são utilizados para classificar a qual classe pertence a imagem de entrada. Para que esses resultados possam ser recebidos pela última camada (camada densamente conectada), as matrizes precisam passar por uma alteração em sua

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

estrutura, chamada de *Flattening*, onde as informações são colocadas de forma linear (FARIA, 2018).

Finalmente, na última camada 'redes densamente conectadas' estão os classificadores. Normalmente, é aplicada a função de ativação que normaliza as saídas entre 0 e 1, a chamada *Softmax*, que traz como saída as probabilidades do resultado. Estar totalmente conectada implica dizer que todos os neurônios da camada anterior estão conectados a todos os neurônios da camada seguinte. Portanto, na CNN todas as camadas estão interligadas, e fazem o mapeamento da entrada para fazer possíveis predições de acordo com as classes predeterminadas (IAN GOODFELLOW, YOSHUA BENGIO, 2015).

As CNNs são algoritmos que ganharam popularidade na área do aprendizado profundo devido à sua arquitetura, capacidade de fornecer resultados promissores sem a necessidade de extração manual de recursos de dados brutos da imagem (ULJAREVIC; HAMILTON, 2013), além de possuir uma maior precisão se comparado a outros e uma boa adaptabilidade ao adquirir novos conhecimentos.

3 METODOLOGIA

Para estudo e análise foram consideradas abordagens e técnicas de detecção facial, reconhecimento facial e reconhecimento de expressões faciais. Os algoritmos foram codificados utilizando a linguagem *Python* e implementados no ambiente de desenvolvimento *PyCharm Community Edition v2020.3*. Após a análise e respectivas comparações é proposta uma abordagem que implementa os algoritmos escolhidos, cujas etapas são mostradas na Figura 8.

Na etapa de detecção facial, foram utilizados os algoritmos *Haar Cascade*, HOG (do inglês, *Histogram of Oriented Gradients*) e CNN MMOD, que implementa o método *Max-Margin Object Detect*, proporcionando uma detecção mais robusta (KING, 2015).

Na etapa de reconhecimento facial, foram testados os algoritmos *Eigenfaces*, *Fisherfaces*, *Local Binary Patterns Histograms* (LBPH) e CNN *Inception-Resnet-v1*. O *Inception-Resnet-v1*, é um modelo treinado de CNN que é executado sobre as faces para obter os vetores de características; e para classificação pelos nomes foi utilizado o algoritmo KNN (do inglês, *k-Nearest Neighbors Algorithm*), ambos implementados através da DLIB (KING, 2009) que

é uma biblioteca com grande variedade de ferramentas e algoritmos de aprendizagem de máquina disponível para uso gratuito.

Figura 8: Metodologia da abordagem proposta



Fonte: Elaborado pelos autores (2021).

Para a etapa de reconhecimento de expressões de emoções, uma estrutura de CNN foi proposta baseada no trabalho de (LASRI; SOLH; EL BELKACEMI, 2019), aplicando mudanças na quantidade de camadas e nos valores nos hiper parâmetros com o propósito de conseguir bons resultados no reconhecimento das expressões faciais a partir de uma estrutura simples. A arquitetura mais adequada de CNN a ser implementada no Reconhecimento de Expressões, considerou as delimitações das características do problema dentro da aplicação proposta. Foram testados datasets com diferentes iluminações e poses para que pudesse ter uma melhor avaliação dos algoritmos.

3.1 Datasets Utilizados

Para os testes referentes aos algoritmos de detecção facial, 20 fotos aleatórias do banco de imagens denominado WIDER (YANG et al., 2016) foram utilizadas. O *dataset* WIDER é um banco de dados que contém 32.203 imagens diversificadas em poses e escalas e focado em detecção de rostos, o que permite realizar diferentes testes, se necessário.

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

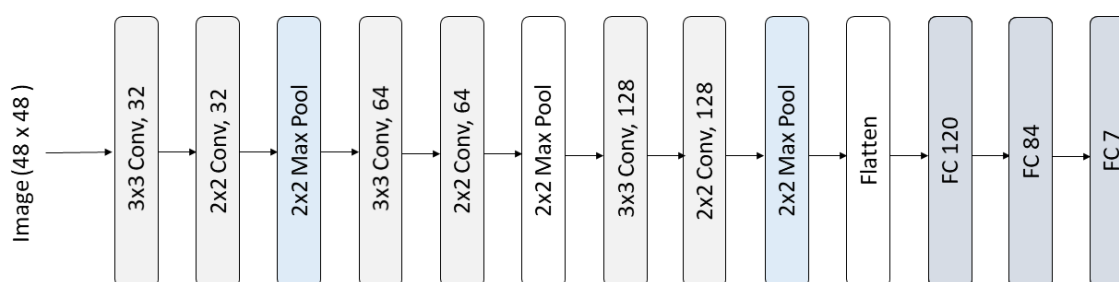
Para os testes de reconhecimento facial foram utilizados 5 bases de dados diversificados, sendo elas: FEI (OLIVEIRA JR; THOMAZ, 2006), Yale Faces (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), Caltech Faces (WEBER, 2006), Georgia Tech (NEFIAN, 1999) e Face Recognition Data (SPACEK, 2007).

Para os testes de reconhecimento de expressões foi utilizado o *dataset* FER2013 (PRAMERDORFER; KAMPEL, 2016), que apresenta 7 expressões com rótulos, contendo o total de 35.887 imagens, geradas utilizando API de pesquisa de imagens do Google, todas possuindo dimensão 48x48 *pixels*.

3.2 Abordagem de CNN proposta para Reconhecimento de expressões faciais.

A abordagem de CNN proposta neste artigo para reconhecimento de emoções se baseia no trabalho de Lasri (LASRI; SOLH; EL BELKACEMI, 2019), visto que se trata de uma arquitetura simples, contando com 6 camadas convolucionais intercaladas com camadas de *Maxpooling*, e recebe entradas de dimensão 48 x 48. As camadas utilizaram filtros de tamanho 32, 64 e 128 com *kernel* 3x3 e 2x2, respectivamente (Figura 9).

Figura 9: Estrutura da Rede Neural Convolucional Proposta



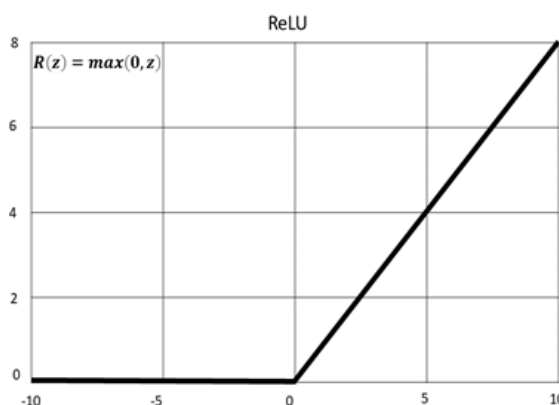
Fonte: Elaborado pelos autores (2021).

A primeira e segunda camada totalmente conectada, ambas no final da arquitetura, utilizam 120 e 84 neurônios, respectivamente, e são acompanhadas com *Dropout*. O objetivo da camada totalmente conectada é nivelar os recursos de alto nível, que são aprendidos por camadas convolucionais, e misturar todos os elementos (SAKIB et al., 2018). A última camada totalmente conectada utiliza a função *softmax* para classificação multiclasse, que traz como saída 7 rótulos de classes, que são as emoções a serem classificadas.

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

Como função de ativação foi aplicada a função ReLU em todas as camadas objetivando generalizar melhor o modelo e induzindo a não linearidade. A operação de ReLU substitui todos os valores de *pixels* negativos no mapa por zero, e mantém os valores positivos sem nenhuma modificação (MASSUCATTO, 2018), sendo muito eficiente computacionalmente quando comparada a outras funções. Conforme mostra a Figura 10, $R(z)$ é zero quando z é menor que zero, e $R(z)$ é igual a z quando z for maior ou igual a zero.

Figura 10: Função ReLU



Fonte: Adaptado de (LASRI; SOLH; EL BELKACEMI, 2019).

Stuart e Peter (RUSSELL; NORVIG, [s.d.]) destacam a importância da divisão do conjunto de dados de treinamento em um subconjunto, que é chamado de conjunto de validação, com o objetivo de ajustar os parâmetros para obter uma estimativa inicial da habilidade do modelo. Já o conjunto de testes, são utilizadas amostras de dados para fornecer uma avaliação imparcial, e previsões do quão bem será o desempenho do modelo em futuras amostras (KUHN; JOHNSON, 2013). Para o treinamento da CNN foi utilizado 80% do banco de imagens para treino, 10% para validação, 10% para os testes.

Com o propósito de gerar novos dados virtuais a partir das imagens existentes para prover mais volume e dar mais suporte ao treinamento, foi aplicada a técnica de *Data Augmentation*, com o objetivo de realizar modificações nas imagens melhorando a capacidade de generalização da rede, conforme Figura 11. Os valores utilizados foram: *rotation range* = 10, *zoom range* = 0.1, *width shift range* = 0.1, *height shift range* = 0.1, *horizontal flip* = True.

Figura 11: Data Augmentation

Fonte: Elaborado pelos autores (2021).

Para a implementação da CNN foi utilizada a biblioteca *Tensorflow* (NELLI; NELLI, 2018) voltada para a prática de Redes Neurais Profundas juntamente com o *Keras* (MOOLAYIL, 2019), uma API (*Application Programming Interface*) de alto nível.

4 RESULTADOS E DISCUSSÃO

4.1 Detecção Facial

Na etapa de detecção facial foi realizada uma comparação entre os algoritmos relacionados a eficiência na detecção de rostos nas imagens, onde o algoritmo HOG alcançou uma taxa de acerto de 41% do total de 129 faces com o tempo de 0,69 ms. O algoritmo CNN ficou em segundo lugar com a taxa de acerto de 33% e a pior média de tempo, de aproximadamente 15,00 segundos, seguido do algoritmo HAAR com 20% de acerto e o menor tempo de processamento: 0,08 ms. Os resultados podem ser vistos na Tabela 1.

Tabela 1: Comparativo de Algoritmos de Detecção Facial

Algoritmo	Acurácia (%)	Tempo de Processamento (s)
HAAR	20.0	0,08
HOG	41.0	0,69
CNN MMOD	33.0	15,5

Fonte: Banco de Dados da Pesquisa (2021)

Dos resultados alcançados, o algoritmo de detecção escolhido foi o HOG, pois obteve a melhor taxa de acerto dentre os outros e o tempo de processamento menor que 1 segundo nos testes realizados.

4.2 Reconhecimento Facial

Nos testes com algoritmos de reconhecimento facial, foi realizada uma comparação entre os algoritmos CNN (*Inception-Resnet1*), *Fisherfaces*, LBPH e *Eigenfaces*, os dois últimos obtiveram quase o mesmo padrão de acerto, com o *Fisherfaces* obtendo a menor taxa de acerto com a maioria dos *datasets* e a CNN obtendo o melhor desempenho em todos os *datasets* de imagens, chegando a 100% de acerto no *dataset* "Faces 94" como pode ser visto na Tabela 2.

Tabela 2: Comparativo de Algoritmos de Reconhecimento Facial

Algoritmos	Reconhecimento Facial – Comparativo de Acurácia (%)					
	FEI	YaleFaces	Faces94 – Mulheres	Faces94 - Homens	Georgia Tech	Caltech
Eigenfaces	57.2	60.0	96.3	97.5	59.2	61.1
Fisherfaces	5.5	70.0	90.0	88.5	6.8	42.1
LBPH	65.0	63.3	100.0	96.5	52.8	84.2
CNNInception- Resnet1	82.5	96.7	100.0	100.0	61.6	98.9

Fonte: Banco de Dados da Pesquisa (2021).

Para o reconhecimento facial, o algoritmo que apresentou melhor resultado foi a CNN *Inception-Resnet1*, uma CNN pré-treinada onde a classificação final de imagens reconhecidas foi realizada com o algoritmo *KNN* ambos implementados na linguagem *Python* através da biblioteca *DLIB* – biblioteca multiplataforma para uso gratuito.

4.3 Reconhecimento de Expressões Faciais

Nos testes com reconhecimento de emoções, a avaliação de eficiência da CNN proposta foi realizada através de gráficos de acurácia e função de perda no treino e validação, e a porcentagem de acurácia nos testes. A função de perda (também conhecida como função de custo) quantifica o quão próximo a rede está de acertar (HAYKIN, 2007), portanto quanto menor o valor desta função, melhor o algoritmo está se comportando.

Na tabela 3, são apresentados os resultados de acurácia e perda dos processos realizados, onde como melhor resultado obteve-se uma acurácia de 65,56%.

Tabela 3: Resultados de Acurácia e Perda – Treino, Validação e Teste (%)

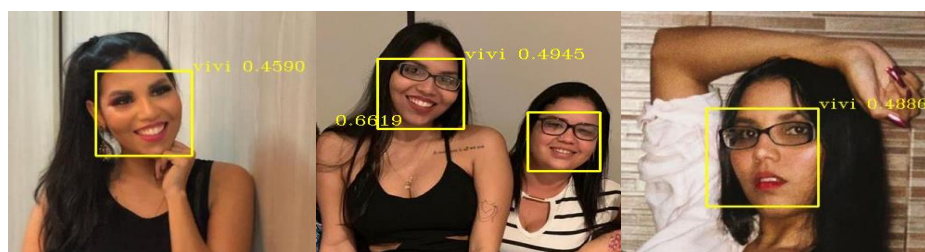
	Acurácia	Perda
Treino	67,02%	0.87
Validação	65,76%	0.91
Teste	65,56%	0.95

Fonte: Banco de Dados da Pesquisa (2021).

Com base nos resultados alcançados com testes anteriores utilizando as bases de dados, são selecionados os algoritmos HOG e CNN para fazer parte da abordagem para as etapas de detecção e reconhecimento facial, respectivamente, e uma CNN customizada pela autora para o reconhecimento de emoções. Assim, nesta última parte de testes são utilizadas imagens reais relacionadas com o problema no qual pretende-se utilizar a CNN customizada.

Considerando que as imagens são capturadas em um ambiente coletivo, o reconhecimento facial é uma etapa fundamental para identificar especificamente o indivíduo que precisa ser monitorado e realizar o reconhecimento de emoções faciais específicas antes de ocorrer o colapso (*Meltdown*). Na Figura 12, pode-se observar o resultado de reconhecimento facial utilizando a abordagem proposta, e na Figura 13 são mostrados alguns exemplos de emoções faciais reconhecidas na pessoa anteriormente identificada.

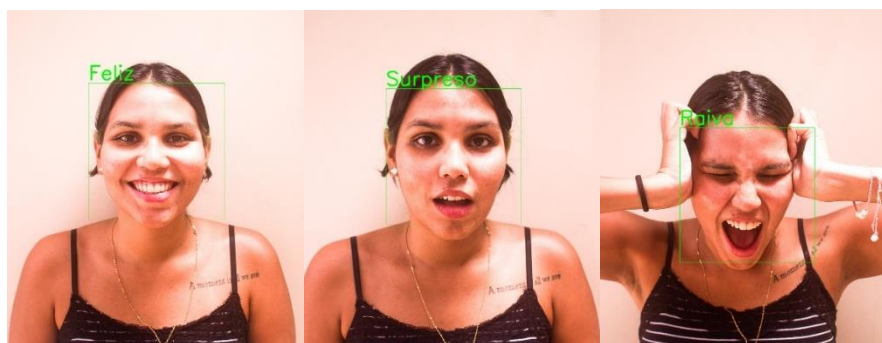
Figura 12: Exemplos de detecção e reconhecimento facial utilizando HOG e CNN



Fonte: Elaborado pelos autores (2021).

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

Figura 13: Exemplo de reconhecimento de algumas emoções faciais: felicidade, surpreso e raiva



Fonte: Elaborado pelos autores (2021).

É importante ressaltar que, o histórico do indivíduo e a experiência do profissional de saúde que o acompanha cotidianamente serão importantes para definir um limiar no algoritmo para determinar a porcentagem do quanto a expressão facial apresentada pode ser considerada importante para indicar um sinal relacionado com o colapso.

4 CONCLUSÕES

Neste artigo apresenta-se uma análise utilizando algoritmos de visão computacional mediante o reconhecimento de expressões faciais para identificar sinais prévios a um colapso autista.

Os resultados dos testes em *datasets* mostraram na etapa de detecção facial, o HOG obteve a melhor taxa de acerto e um tempo de processamento menor que 1 segundo. Na etapa de reconhecimento facial, a estrutura de CNN *Inception-Resnet1* conseguiu a melhor acurácia com 100% de acerto na validação de alguns *datasets*. Já na etapa de reconhecimento de expressões faciais, a CNN proposta pela autora alcançou uma acurácia acima de 65%, utilizando para treinamento, validação e testes o *dataset* FER2013.

Os resultados alcançados utilizando os melhores algoritmos testados para cada etapa são considerados aceitáveis para a abordagem proposta testada, com uma baixa complexidade na estrutura e possibilitando a sua utilização em tempo real. É importante ressaltar que a abordagem considera que as imagens capturadas possuem boa iluminação.

A identificação prévia dos colapsos poderá ser uma ferramenta importante de apoio para os profissionais da área de saúde e educação possibilitando a tomada de medidas paliativas de maneira eficaz e com isso beneficiar na melhora da qualidade de vida do indivíduo monitorado.

REFERÊNCIAS

ADRIANO, Teresa; ARRIAGA, Patrícia. Exaustão emocional e reconhecimento de emoções na face e voz em médicos. **Psicologia, Saúde & Doenças**, [S. l.], v. 17, n. 1, p. 97–104, 2016.

ANDRADE, Nara Côrtes; ABREU, Neander Silva; DURAN, Victor Riccio; VELOSO, Tainã Jesus; MOREIRA, Narena Alencar. Reconhecimento de expressões faciais de emoções: padronização de imagens do teste de conhecimento emocional. **Psico**, [S. l.], v. 44, n. 3, p. 382–390, 2013.

BALASUBRAMANIAN, Balaji; DIWAN, Pranshu; NADAR, Rajeshwar; BHATIA, Anuradha. Analysis of Facial Emotion Recognition. *In*: 2019 3RD INTERNATIONAL CONFERENCE ON TRENDS IN ELECTRONICS AND INFORMATICS (ICOEI) 2019, **Anais [...]**. [s.l: s.n.] p. 945–949.

BELHUMEUR, Peter N.; HESPANHA, João P.; KRIEGMAN, David J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S. l.], 1997. DOI: 10.1109/34.598228.

BHELE, Sujata G.; MANKAR, V. H. A review paper on face recognition techniques. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, [S. l.], v. 1, n. 8, p. 339–346, 2012.

CHAO, Wei-Lun. Face recognition. **GICE, National Taiwan University**, [S. l.], 2007.

CORREIA, Thaisa; PITERI, Almir Artero; SILVA, Francisco Assis; PEREIRA, Danilo Roberto. Development of an application for security based in face recognition on Android platform. *In*: X WORKSHOP DE VISÃO COMPUTACIONAL 2014, **Anais [...]**. [s.l: s.n.]

DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. *In*: 2005 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05) 2005a, **Anais [...]**. [s.l: s.n.] p. 886–893.

DALAL, Navneet; TRIGGS, Bill. Histograms of Oriented Gradients for Human Detection. *In*: PROCEEDINGS OF THE 2005 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05) - VOLUME 1 - VOLUME 01 2005b, USA. **Anais [...]**. USA: IEEE Computer Society, 2005. p. 886–893. DOI: 10.1109/CVPR.2005.177. Disponível em: <https://doi.org/10.1109/CVPR.2005.177>.

DE OLIVEIRA, Eduardo; JAQUES, Patrícia Augustin. Classificação de emoções básicas através de imagens capturadas por webcam. **Revista**

Revista Conexão na Amazônia, ISSN 2763-7921, n. 3, v. 2, 2021

Brasileira de Computação Aplicada, [S. l.], v. 5, n. 2, p. 40–54, 2013.

FARIA, Elisângela Lopes De. Redes neurais convolucionais e máquinas de aprendizado extremo aplicadas ao mercado financeiro brasileiro. [S. l.], 2018.

FREUND, Yoav; SCHAPIRE, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. *In: EUROPEAN CONFERENCE ON COMPUTATIONAL LEARNING THEORY 1995*, **Anais [...]**. [s.l: s.n.] p. 23–37.

GONZALEZ, Rafael C.; WOODS, Richard E. Digital Image Processing (3rd Edition). **Prentice-Hall, Inc. Upper Saddle River, NJ, USA ©2006**, [S. l.], p. 976, 2007. Disponível em: <http://dl.acm.org/citation.cfm?id=1076432>. Acesso em: 12 out. 2021.

HAYKIN, Simon. **Redes neurais: princípios e prática**. [s.l.] : Bookman Editora, 2007.

HENRIQUES, Tiago. **Autismo e Síndrome de Asperger: O Guia Fácil de Entender para Pais, Educadores e Portadores de Autismo: E se fosse possível realmente entender o autismo?** 1. ed. [s.l.] : Independently Published, 2018.

IAN GOODFELLOW, YOSHUA BENGIO, Aaron Courville. Deep Learning Book. **Deep Learning**, [S. l.], 2015.

JAQUES, Patrícia Augustin; VICARI, Rosa Maria. Estado da arte em ambientes inteligentes de aprendizagem que consideram a afetividade do aluno. **Revista informática na educação: teoria & prática**, [S. l.], v. 8, n. 1, p. 15–38, 2005.

KING, Davis E. Dlib-ml: A Machine Learning Toolkit. **Journal of Machine Learning Research**, [S. l.], v. 10, p. 1755–1758, 2009.

KING, Davis E. Max-Margin Object Detection. [S. l.], 2015. Disponível em: <https://arxiv.org/abs/1502.00046v1>. Acesso em: 9 out. 2021.

KUHN, Max; JOHNSON, Kjell. Applied predictive modeling. **Applied Predictive Modeling**, [S. l.], p. 1–600, 2013. DOI: 10.1007/978-1-4614-6849-3.

LASRI, Imane; SOLH, Anouar Riad; EL BELKACEMI, Mourad. Facial Emotion Recognition of Students using Convolutional Neural Network. *In: 2019 THIRD INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES (ICDS) 2019*, **Anais [...]**. [s.l: s.n.] p. 1–6.

LIENHART, Rainer; KURANOV, Alexander; PISAREVSKY, Vadim. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *In: JOINT PATTERN RECOGNITION SYMPOSIUM 2003*, **Anais [...]**. [s.l: s.n.] p. 297–304.

LIM, Andreas Pangestu; KUSUMA, Gede Putra; ZAHRA, Amalia; OTHERS. Facial Emotion Recognition Using Computer Vision. *In: 2018 INDONESIAN ASSOCIATION FOR PATTERN RECOGNITION INTERNATIONAL CONFERENCE (INAPR) 2018*, **Anais [...]**. [s.l: s.n.] p. 46–50.

MASSUCATTO, Jean Daniel Prestes. **Aplicação de conceitos de redes**

- neurais convolucionais na classificação de imagens de folhas.** 2018. Universidade Tecnológica Federal do Paraná, [S. l.], 2018.
- MOOLAYIL, Jojo. **Learn Keras for Deep Neural Networks.** [s.l: s.n.]. DOI: 10.1007/978-1-4842-4240-7.
- NEFIAN, Ara V. **Georgia Tech face database** Georgia Institute of Technology, 1999.
- NELLI, Fabio; NELLI, Fabio. Deep Learning with TensorFlow. *In: Python Data Analytics.* [s.l: s.n.]. DOI: 10.1007/978-1-4842-3913-1_9.
- OLIVEIRA, Italo; SILVA, Jacqueline Lopes; QUISPE, Facundo Palomino; ALVAREZ, Ana Beatriz. **EmotiTEA: A visual monitoring module based on the recognition of facial emotions with CNN.** Rio Branco.
- OLIVEIRA JR, L. L.; THOMAZ, C. E. Captura e alinhamento de imagens: Um banco de faces brasileiro. **Relatório de iniciação científica, Depto. Eng. Elétrica da FEI, São Bernardo do Campo, SP, [S. l.], v. 10, 2006.**
- OMAR, Herman Kh; TAWFIQ, Nada E. Face Recognition Based on Histogram Equalization and LBP Algorithm. **Academic Journal of Nawroz University, [S. l.], v. 8, n. 3, p. 33–38, 2019.**
- PANCERI, João Antonio Campos; PINTO, Luiz Alberto; PEREIRA, Flávio Garcia; CAVALIERI, Daniel Cruz; KOMATI, Karin Satie. RECONHECIMENTO FACIAL BASEADO EM HOG E PCA: UMA COMPARAÇÃO QUANTO À INVARIÂNCIA À ILUMINAÇÃO. **Revista Ifes Ciência-ISSN 2359-4799, [S. l.], v. 1, n. 1, 2015.**
- PARKHI, Omkar M.; VEDALDI, Andrea; ZISSERMAN, Andrew. Deep face recognition. [S. l.], 2015.
- PATNAM, Venkata Sindhoor Preetham et al. Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. **Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017, [S. l.], 2017.** DOI: 10.1109/ICHI.2017.35.
- PRAMERDORFER, Christopher; KAMPEL, Martin. Facial expression recognition using convolutional neural networks: state of the art. **arXiv preprint arXiv:1612.02903, [S. l.], 2016.**
- RUSSELL, Stuart J. (Stuart Jonathan); NORVIG, Peter. Artificial intelligence : a modern approach. [S. l.], p. 1091, [s.d.].
- SAKIB, Shadman; AHMED; JAWAD, Ahmed; KABIR, Jawad; AHMED, Hridon. An Overview of Convolutional Neural Network: Its Architecture and Applications. **ResearchGate, [S. l.], 2018.**
- SEBE, Nicu; COHEN, Ira; GEVERS, Theo; HUANG, Thomas S. Multimodal approaches for emotion recognition: a survey. *In: INTERNET IMAGING VI 2005, Anais [...].* [s.l: s.n.] p. 56–67.
- SHAPIRO, Linda G.; STOCKMAN, George C. Computer vision. [S. l.], p. 580, 2001.

SILVA, Jacqueline Lopes. **Identificação e Reconhecimento de Emoções Faciais utilizando Redes Neurais Convolucionais em Benefício da Tecnologia Assistiva**. 2021. Universidade Federal do Acre - UFAC, [S. l.], 2021.

SILVA, Jacqueline Lopes; OLIVEIRA, Italo; TOPOLNIAK, Zabelita; ALVAREZ, Ana Beatriz. A CNN Approach Implemented to Emotional Facial Expression Recognition for the Prevention of Autistic Meltdowns. [S. l.], p. 1–6, 2021. DOI: 10.1109/SCLA53004.2021.9540183.

SOARES, Angélica Miguel; CAVALCANTE NETO, Jorge Lopes. Evaluation of Motor Behavior in Children with Autism Spectrum Disorder: a Systematic Review. **Revista Brasileira de Educação Especial**, [S. l.], v. 21, n. 3, p. 445–458, 2015.

SPACEK, L. **Collection of facial images: Faces94. Computer vision science and research projects, University of Essex, UK**, 2007. Disponível em: <https://cswww.essex.ac.uk/mv/allfaces/>.

TURK, Matthew; PENTLAND, Alex. Face recognition using eigenfaces. *In*: PROCEEDINGS. 1991 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 1991, **Anais [...]**. [s.l: s.n.] p. 586–587.

ULJAREVIC, Mirko; HAMILTON, Antonia. Recognition of emotions in autism: A formal meta-analysis. **Journal of Autism and Developmental Disorders**, [S. l.], v. 43, n. 7, p. 1517–1526, 2013. DOI: 10.1007/s10803-012-1695-5. Disponível em:

<http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L52284014%0Ahttp://dx.doi.org/10.1007/s10803-012-1695-5>.

VIOLA, Paul; JONES, Michael. Rapid object detection using a boosted cascade of simple features. *In*: PROCEEDINGS OF THE 2001 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. CVPR 2001 2001, **Anais [...]**. [s.l: s.n.] p. I–I.

VIOLA, Paul; JONES, Michael J. Robust real-time face detection. **International journal of computer vision**, [S. l.], v. 57, n. 2, p. 137–154, 2004.

WEBER, Markus. Face database collection of Markus Weber. **Face Database Collection of Markus Weber**, [S. l.], 2006.

YANG, Shuo; LUO, Ping; LOY, Chen Change; TANG, Xiaoou. WIDER FACE: A Face Detection Benchmark. *In*: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR) 2016, **Anais [...]**. [s.l: s.n.]