UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS ARARANGUÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

Felipe Zago Canal

Método para Reconhecimento em Tempo Real de Expressões Faciais em Grupos utilizando Redes Neurais Convolucionais

Araranguá
2024

Felipe Zago Canal

Método para Reconhecimento em Tempo Real de Expressões Faciais em Grupos utilizando Redes Neurais Convolucionais

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina para a obtenção do título de mestre em Tecnologias da Informação e Comunicação.
Orientadora: Profª. Eliane Pozzebon, Drª.
Coorientador: Prof. Antonio Carlos Sobieranski, Dr.

Araranguá
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Canal, Felipe Zago
Método para Reconhecimento em Tempo Real de Expressões Faciais em Grupos utilizando Redes Neurais Convolucionais / Felipe Zago Canal ; orientadora, Eliane Pozzebon, coorientador, Antonio Carlos Sobieranski, 2024.
78 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Campus Araranguá, Programa de Pós-Graduação em Tecnologias da Informação e Comunicação, Araranguá, 2024.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2. Reconhecimento de Expressões Faciais. 3. Inteligência Artificial. 4. Análise Afetiva em Grupo. I. Pozzebon, Eliane. II. Sobieranski, Antonio Carlos. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Tecnologias da Informação e Comunicação. IV. Título.

Felipe Zago Canal

Método para Reconhecimento em Tempo Real de Expressões Faciais em Grupos utilizando Redes Neurais Convolucionais

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Roderval Marcelino, Dr.
Universidade Federal de Santa Catarina

Profª. Liane Margarida Rockenbach Tarouco, Drª.
Universidade Federal do Rio Grande do Sul

Prof. Ig Ibert Bittencourt Santana Pinto, Dr.
Universidade Federal de Alagoas

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Tecnologias da Informação e Comunicação.

ICP-Edu
Documento assinado digitalmente
Giovani Mendonca Lunardi
Data: 11/03/2024 10:49:32-0300
CPF: ***.394.559-**
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de Pós-Graduação

GOV.BR
Documento assinado digitalmente
ELIANE POZZEBON
Data: 10/03/2024 09:13:26-0300
CPF: ***.213.139-**
Verifique as assinaturas em https://v.ufsc.br

Profª. Eliane Pozzebon, Drª.
Orientadora
Universidade Federal de Santa Catarina

ICP-Edu
Documento assinado digitalmente
Antonio Carlos Sobieranski
Data: 10/03/2024 19:39:11-0300
CPF: ***.305.809-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Antonio Carlos Sobieranski, Dr.
Coorientador
Universidade Federal de Santa Catarina

Araranguá, 2024.

AGRADECIMENTOS

A elaboração desta dissertação de mestrado constituiu-se em uma jornada de intenso aprendizado e desenvolvimento pessoal e acadêmico, a qual não teria sido possível sem o apoio inestimável de diversas pessoas a quem sou profundamente grato.

Inicialmente, expresso minha mais sincera gratidão aos meus pais, pilares fundamentais da minha vida, cujo amor, sacrifício e incentivo incondicional formaram a base sobre a qual pude construir meus sonhos e aspirações. Seu exemplo de dedicação e resiliência inspira-me continuamente a perseguir meus objetivos com determinação e coragem.

Ao meu irmão, agradeço pelo companheirismo e pela amizade que representa uma fonte constante de apoio e motivação. Sua presença e incentivo foram essenciais para superar os momentos de desafio e incerteza ao longo desta jornada.

À minha namorada, expresso minha profunda gratidão pelo amor, compreensão e paciência demonstrados em cada etapa deste processo. Sua presença e apoio inabalável foram fundamentais para manter meu equilíbrio emocional e foco, permitindo-me avançar mesmo diante das adversidades.

Devo um especial agradecimento à minha orientadora, cuja expertise, dedicação e orientação acadêmica foram cruciais para o desenvolvimento e conclusão desta pesquisa. Sua capacidade de me guiar com sabedoria, promovendo um ambiente de aprendizado estimulante e desafiador, permitiu-me explorar o potencial completo do meu trabalho.

Ao meu coorientador, agradeço pela valiosa contribuição intelectual e pelo suporte técnico oferecidos durante o desenvolvimento deste estudo. Sua perspectiva única e conhecimento especializado enriqueceram significativamente minha pesquisa, proporcionando elementos essenciais para a articulação e refinamento da minha abordagem científica.

Expresso minha gratidão à CAPES pelo suporte e pela oportunidade de desenvolver este trabalho dentro de um ambiente acadêmico estimulante com os recursos necessários. O auxílio da CAPES foi fundamental não apenas para a realização desta pesquisa, mas para toda a minha formação acadêmica, permitindo-me dedicar ao estudo e à investigação científica na área de Tecnologias da Informação e Comunicação.

Estendo igualmente meus sinceros agradecimentos à Universidade Federal de Santa Catarina (UFSC) e ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação. A UFSC, com seu corpo docente qualificado e diversificado, proporcionou-me uma formação robusta e abrangente, enquanto o Programa de Pós-Graduação em Tecnologias da Informação e Comunicação ofereceu um ambiente multidisciplinar rico, onde pude desenvolver minha pesquisa com o apoio e a orienta-

ção de professores e colegas excepcionais. Essas instituições foram pilares na minha trajetória, permitindo-me explorar novas fronteiras do conhecimento e contribuir para o avanço da ciência na minha área de estudo.

Por fim, estendo meus agradecimentos a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, desde colegas de academia até amigos que ofereceram palavras de incentivo e conforto nos momentos necessários. A todos vocês, minha eterna gratidão.

Este trabalho é o resultado do apoio, fé e amor de cada um de vocês.

RESUMO

Com a evolução das Novas Tecnologias da Informação e Comunicação (NTICs), o setor educacional, assim como diversos outro setores da sociedade, se beneficiaram com aplicações e soluções nas mais diversas áreas. A adoção de novas ferramentas com Inteligência Artificial (IA) para auxiliar professores e estudantes no processo de ensino-aprendizado é essencial e frequente neste contexto, especialmente devido às mudanças das dinâmicas sociais dos últimos anos, desencadeadas pela pandemia da COVID-19. Este trabalho aborda a aplicação de (NTICs) no reconhecimento de expressões faciais em grupos, uma área de crescente interesse devido ao seu potencial nestes ambientes e outros contextos sociais. A pesquisa centrou-se no desenvolvimento e aplicação de um modelo de IA baseado em Redes Neurais Convolucionais (CNNs), utilizando a técnica de transferência de aprendizagem sobre o modelo MobileNetV2, ajustado para classificar sete expressões faciais básicas. A metodologia envolveu a seleção do dataset FER-2013 para treinar o modelo, seguida por um processo de treinamento e validação que utilizou métricas como precisão, *recall*, e *F1-score*. A acurácia obtida no conjunto de teste foi de 68.57%em média para todas as expressões classificadas. Um comparativo com trabalhos relacionados revelou que o modelo apresenta resultados promissores, superando a maioria das outras abordagens baseadas no mesmo dataset. A análise destacou a importância de datasets representativos e a necessidade de considerar a variabilidade e sutileza das expressões humanas. Uma aplicação prática do software foi realizada em um grupo de alunos, onde o modelo operou em tempo real e demonstrou eficiência na captura de expressões faciais, alcançando uma média de 12 *frames* por segundo (FPS). O experimento realçou tanto a eficácia quanto os desafios do modelo em capturar a dinâmica das expressões em um ambiente coletivo e reativo. As conclusões do experimento apontam para a importância de continuar refinando a tecnologia de reconhecimento facial, levando em consideração a diversidade das expressões humanas e as nuances das interações grupais. O trabalho demonstrou que, apesar do potencial significativo, a aplicação efetiva de reconhecimento facial em grupos requer um entendimento profundo do comportamento humano e adaptações para contextos específicos. Este trabalho contribui para o avanço da pesquisa, fornecendo percepções valiosos para futuros desenvolvimentos e aplicações.

Palavras-chave: Reconhecimento de Expressões Faciais. Análise Afetiva em Grupo. Inteligência Artificial.

ABSTRACT

With the evolution of New Information and Communication Technologies (NTICs), the educational sector, along with various other sectors of the society, has benefited from applications and solutions in the most diverse areas. The adoption of new tools with Artificial Intelligence (IA) to assist teachers and students in the teaching-learning process is essential and frequent in this context, especially due to changes in social dynamics in recent years, triggered by the COVID-19 pandemic. This work addresses the application of NTICs in the recognition of facial expressions in groups, an area of growing interest due to its potential in these environments and other social contexts. The research focused on the development and application of an IA model based on Convolutional Neural Networks (CNNs), using the technique of transfer learning on the MobileNetV2 model, adjusted to classify seven basic facial expressions. The methodology involved the selection of the FER-2013 dataset to train the model, followed by a training and validation processes that used metrics such as precision, recall, and F1-score. The accuracy obtained in the test set was 68.57% on average for all classified expressions. A comparison with related works revealed that the model presents promising results, outperforming most other approaches based on the same dataset. The analysis highlighted the importance of representative datasets and the need to consider the variability and subtlety of human expressions. A practical application of the software was carried out in a group of students, where the model operated in real-time and demonstrated efficiency in capturing facial expressions, achieving an average of 12 frames per second (FPS). The experiment highlighted both the efficacy and the challenges of the model in capturing the dynamics of expressions in a collective and reactive environment. The conclusions of the experiment point to the importance of continuing to refine facial recognition technology, taking into account the diversity of human expressions and the nuances of group interactions. The work demonstrated that, despite significant potential, the effective application of facial recognition in groups requires a deep understanding of human behavior and adaptations for specific contexts. This work contributes to the advancement of research, providing valuable insights for future developments and applications.

Keywords: Facial Expression Recognition. Group Affective Analysis. Artificial Intelligence.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AM	Attention Mechanisms
AVA	Ambiente Virtual de Aprendizagem
BFN	Rede para Faces Grandes (Big Face Network)
CNN	Rede Neural Convolucional (Convolutional Neural Network)
DBN	Rede Bayesiana Dinâmica (Dynamic Bayesian Network)
FACS	Facial Action Coding System
FER	Reconhecimento de Expressões Faciais (Facial Expression Recognition)
FPS	*Frames* por Segundo
GF	Caracteristicas Geométricas (Geometric Features)
GPU	Unidade de Processamento Gráfico
IA	Inteligência Artificial
IoT	Internet das Coisas (Internet of Things)
LSTM	Memória de Longo e Curto Prazo (Long Short-Term Memory)
MFM	Multiscale Feature Map
ML	Aprendizado de Máquina (Machine Learning)
MTCNN	Multi-Task Cascaded Convolutional Neural Network
NLP	Processamento de Linguagem Natural (Natural Language Processing)
NTIC	Novas Tecnologias da Informação e Comunicação
NVPF	Non-Volume Preserving Fusion
PPGTIC	Programa de Pós-Graduação em Tecnologias da Informação e Comunicação
QLZM	Quantised Local Zernike Moments
RF	Random Forest
RNA	Rede Neural Artificial
RNN	Rede Neural Recorrente (Recurrent Neural Network)
SFN	Rede para Faces Pequenas (Small Face Network)
STI	Sistema Tutor Inteligente
SVM	Máquina de Vetor de Suporte (Support Vector Machine)
SVR	Support Vector Regression
TNVPF	Network Temporal Non-volume Preserving Fusion
TSM	Temporal Shift Module

SUMÁRIO

1 INTRODUÇÃO

As Novas Tecnologias da Informação e Comunicação (NTICs) mudaram a forma humana de executar tarefas do cotidiano na última década (VĂIDEAN; ACHIM, 2022). Diversas áreas foram capazes de evoluir rapidamente com a aplicação de NTICs. No setor da saúde, por exemplo, a Internet da Coisas (IoT), Aprendizado de Maquina (ML), Análise de Big Data e Inteligencia Artificial (IA) vem auxiliando profissionais a detectar e tratar doenças diversas (AHMAD *et al.*, 2022). A indústria 4.0, por outro lado, têm se beneficiado da IA, 5G/6G e computação quântica nos últimos dez anos (SIGOV *et al.*, 2022).

Com a evolução destas NTICs, o setor educacional tem a necessidade de seguir tais tendências e aplicar NTICs no auxílio de estudantes e professores no processo de ensino-aprendizado. Nos últimos anos, principalmente em decorrência da pandemia COVID-19, a educação precisou ser adaptada, e o uso de NTICs foi inevitável.

A utilização de Sistemas Tutores Inteligentes (STIs) e Ambientes Virtuais de Aprendizagem (AVAs) ganhou muito espaço na implementação de aplicações de ciências cognitivas, ciências educacionais e IA (HWANG, 2003; GRAESSER; CONLEY; OLNEY, 2012) e é apresentada como tendência para os próximos anos (VICARI, 2018). Estes sistemas foram capazes de prover suporte para a aplicação de educação remota e foram decisivos para a manutenção dos ambientes educacionais durante os períodos mais críticos da pandemia. Contudo, estes sistemas não são perfeitos. Um dos pontos negativos em suas aplicações é a falta da afetividade envolvida no processo de aprendizado tradicional. Emoções como ansiedade, raiva, confusão e tédio podem influenciar negativamente no aprendizado (REBOLLEDO-MENDEZ *et al.*, 2022).

A importância das emoções nos processos educacionais é ressaltada por Molero Jurado *et al.* (2021). Os autores demonstram em seu trabalho que as emoções podem influenciar positivamente (e.g., motivação) ou negativamente (e.g., ansiedade) no aprendizado. Uma vez que os estudantes estão cada vez mais envolvidos com a tecnologia, a inclusão de aplicações de inteligência emocional nos STIs é uma boa estratégia para melhorar o aprendizado.

A computação afetiva diz respeito à reprodução da capacidade humana de observação, interpretação e representação de características afetivas por meio de computadores (TAO; TAN, 2005). Desta forma, a computação afetiva e a análise de sentimentos podem enaltecer as relações com clientes em sistemas de recomendação, bem como impulsionar processos de tutoria e aperfeiçoar sistemas de entretenimento (CAMBRIA *et al.*, 2017). Além disso, a análise de sentimento está presente em várias aplicações relacionadas ao Big Data, como análise de redes sociais, análise de opiniões, análise de mercado financeiro, entre outras (LIU, B. *et al.*, 2010).

Para contornar este problema enfrentado pelas ferramentas digitais de educa-

ção, STIs e AVAs podem implementar a detecção e reconhecimento automáticos de expressão facial. O conceito de reconhecimento de expressões faciais é comumente baseado na identificação de movimentos em certos músculos faciais que são considerados relativos a determinadas expressões. A análise automática de expressões faciais vem sendo estudada extensivamente na literatura (LI, S.; DENG, W., 2020; BETTADAPURA, 2012; REVINA; EMMANUEL, 2021).

O reconhecimento de expressões faciais tem o potencial de transformar o ambiente educacional ao possibilitar uma compreensão mais profunda das emoções e engajamento dos alunos. A identificação precisa das reações emocionais dos estudantes durante o processo de ensino-aprendizagem pode oferecer entendimentos valiosos para educadores, permitindo-lhes ajustar metodologias e abordagens pedagógicas em tempo real para melhor atender às necessidades dos alunos.

A maioria do sistemas de reconhecimento de expressões faciais é baseado no *Facial Action Codign System* (FACS) (EKMAN; FRIESEN, 1978). FACS é uma taxonomia de expressões faciais, baseado em movimentos de músculos específicos que determinam as expressões executadas. O principal problema ao usar o FACS para classificar emoções é a falta de uma definição unificada para as emoções, dificultando encontrar um conjunto consistente de expressões faciais que sejam universalmente reconhecidas. Alguns estudos tentaram resolver este problema usando o framework Ekman (EKMAN; FRIESEN, 1978). O framework proposto por Ekman se tornou um padrão para determinação de expressões faciais neste ramo, de tal modo que alguns datasets de grande influência na literatura são baseados neste sistema (KANADE; COHN; TIAN, 2000; LUCEY *et al.*, 2010a).

Apesar do reconhecimento de expressões faciais ter sido explorado na literatura, ainda há algumas limitações que devem ser superadas. Neste tipo de problema, os sistemas requerem um alto nível de precisão e poder computacional, sendo que fazer esta leitura de forma automática é um problema complexo. Além disso, diferentes fatores como iluminação, cuidado, idade, raça e gênero podem influenciar nas predições, tornando a resolução do problema complexa. Por se tratar de um problema que usualmente requer alto poder computacional na sua aplicação, este tipo de sistema foi beneficiado pelos recentes avanços computacionais nos quesitos de arquitetura e desempenho de hardware (CHENG *et al.*, 2013).

Conforme constatado em Canal *et al.* (2022), a IA tem se mostrado uma grande aliada para realizar reconhecimento de expressões faciais, principalmente por conta destes avanços computacionais alcançados nos últimos anos. Acredita-se que este problema pode ter uma aplicação prática em diversos ambientes como auditórios e salas de aulas, proporcionando informação sobre a atenção e os sentimentos dos alunos aos professores e/ou responsáveis pela atividade de ensino. Desta forma, esta pesquisa tem como objetivo principal a detecção e reconhecimento de expressões

faciais em grupos e em tempo real para promover apoio a professores em atividades de ensino.

1.1 CONTEXTUALIZAÇÃO DO PROBLEMA E JUSTIFICATIVA DA PESQUISA

A comunicação humana é formada por aproximadamente dois terços de componentes não verbais (MEHRABIAN, 2017). Em geral, as pessoas deduzem o estado emocional dos seus semelhantes, como alegria, tristeza ou raiva, por meio das expressões faciais e tom vocal. Sendo assim, as expressões faciais são fatores muito importantes na comunicação humana pois auxiliam no entendimento das intenções dos outros (KO, 2018).

A detecção e reconhecimento da expressão facial podem ser aplicadas em diversas áreas como, por exemplo, na educação e no turismo. Na educação, tal instrumento pode ser usado pelo professor como uma ferramenta de diagnóstico e como medida da eficiência do ensino (RICHARDSON, 2005) ou no turismo, como ferramenta para entendimento da satisfação dos consumidores (GONZÁLEZ-RODRÍGUEZ; DÍAZ-FERNÁNDEZ; GÓMEZ, 2020).

O reconhecimento de expressões faciais é um campo que vêm sendo estudado na literatura a certo tempo, contudo, a construção de uma solução para aplicar este conceito em grupos de pessoas em tempo real, pode agregar valor a diversos processos essenciais da sociedade, como em processos de aprendizado, por exemplo, onde o professor/tutor pode fazer uso da informação afetiva dos alunos para aumentar o engajamento e, consequentemente, a qualidade do aprendizado.

Nos últimos anos, como os avanços computacionais, principalmente no âmbito de GPU muitos algoritmos sofreram adaptações para processamento paralelo, levando seu tempo de execução a patamares próximos a tempo real (TURABZADEH *et al.*, 2017). Especificamente para o campo de reconhecimento de expressões faciais, esses avanços proporcionaram a evolução de uma série de ferramentas que podem ser aplicadas em diferentes abordagens para resolução do problema. Com o intuito de aproveitar-se destes avanços para prover uma solução aos problemas apresentados, a pergunta a ser respondida neste trabalho é: **"Como detectar expressões de emoção facial em um grupo de pessoas em tempo real?"**

Para cumprir este propósito, buscou-se analisar algoritmos de IA específicos para o campo de visão computacional para compreender as limitações computacionais desta tecnologia. O desenvolvimento e avaliação de algoritmos de detecção de expressões faciais para grupos de pessoas em tempo real foi avaliada nesta dissertação, como resposta à pergunta central da pesquisa.

1.2 OBJETIVOS

Os objetivos gerais e os específicos para a obtenção do resultado do trabalho serão apresentados nesta sessão.

1.2.1 Objetivo Geral

Desenvolver método com IA para ser utilizado na identificação de expressões faciais em um grupo de pessoas em tempo real.

1.2.2 Objetivos Específicos

Para facilitar o alcance do objetivo geral, os seguintes objetivos específicos foram definidos:

- Identificar o estado da arte na literatura para os algoritmos de detecção facial;
- Analisar as métricas e resultados dos principais algoritmos de detecção facial;
- Identificar o estado da arte na literatura para os algoritmos de reconhecimento de expressões faciais;
- Analisar as métricas e resultados dos principais algoritmos de reconhecimento de expressões faciais;
- Realizar procedimentos necessários para aquisição de imagens para treinamento de modelos de detecção e reconhecimento de expressões faciais;
- Analisar os resultados da junção dos algoritmos de detecção e reconhecimento, obtendo as expressões faciais em tempo real;
- Elaborar um estudo de caso para avaliar a eficiência do método desenvolvido.

1.3 ADERÊNCIA E INTERDISCIPLINARIDADE

Segundo o a própria definição do PPGTIC no que diz respeito a linha Computacional do programa, a qual é contemplada neste trabalho:

> "O objetivo da linha é desenvolver modelos, técnicas e ferramentas computacionais auxiliando na resolução de problemas de natureza interdisciplinar. Especificamente, esta linha de pesquisa procura desenvolver novas tecnologias computacionais para aplicação nas áreas de educação e gestão." (PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO, 2022).

Desta forma, a proposta deste projeto está de acordo com o escopo da linha computacional do PPGTIC, visto que compreende algoritmos diversos e inteligência artificial para identificar as expressões faciais de forma automática e em tempo real.

Sobre a linha Educacional, ainda de acordo com o PPGTIC:

> "A linha de pesquisa envolve o estudo, a concepção, o desenvolvimento e a construção de materiais de apoio ao ensino e à aprendizagem (hardware e software) no contexto educacional, nos diferentes níveis de educação. O objetivo é auxiliar a fomentar o desenvolvimento de habilidades e competências para uso de tecnologias como apoio a inovações educacionais." (PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO, 2022).

Neste sentido, a linha educacional também é compreendida, uma vez que os resultados das expressões tem grande valia para o processo de ensino-aprendizagem, podendo proporcionar um instrumento de analise de engajamento e foco dos estudantes a partir da aplicação do software desenvolvido e podendo ser empregada em diferentes níveis de educação.

Desta forma, a interdisciplinaridade desta pesquisa se apresenta na integração de algoritmos de inteligência artificial, juntamente com todo seu ecossistema de desenvolvimento e execução, com ambientes de ensino, composto por um grupo de estudantes e educador(es) que, por sua vez, se beneficiarão da utilização da ferramenta como forma de apoio ao aprendizado.

1.4 METODOLOGIA

As metodologias aplicadas nos processos constituintes deste estudo, são baseadas de acordo com as definições de finalidades e objetivos propostos por Gil *et al.* (2002). O trabalho apoia-se também nas definições de Yin (2016) por se tratar de uma pesquisa de abordagem qualitativa e Yin (2015), no quesito técnico dos procedimentos.

A pesquisa deste projeto é **bibliográfica**, por ter sido realizada uma busca ampla em bases de dados, revistas, livros e artigos científicos que serviu de embasamento teórico para o seu desenvolvimento. Quanto ao objetivo, a pesquisa tem caráter **exploratório** para, segundo Gil *et al.* (2002), proporcionar maior familiaridade com o problema e torná-lo mais explícito. O caráter **descritivo** também está presente neste estudo, pois busca-se avaliar a utilização do sistema a ser desenvolvido no ensino por uma visão específica de educadores.

Em relação aos procedimentos técnicos, a pesquisa é considerada um **estudo de caso**, ao passo que procura investigar empiricamente um fenômeno em profundidade e em seu contexto de mundo real, aqui representados pelo desenvolvimento do sistema de reconhecimento de expressões faciais e sua avaliação de performance da visão dos educadores (YIN, 2015).

Com o objetivo de gerar conhecimento para aplicações práticas e soluções de prolemas, a natureza desta pesquisa é **aplicada** (GIL *et al.*, 2002) e, pela aquisição de dados para verificação da eficiência do sistema no processo de ensino, utiliza-se o método **qualitativo** para representar a opinião e perspectiva dos participantes deste estudo (YIN, 2016). A tabela 1 sintetiza a metodologia aplicada na pesquisa, conforme descrito:

Tabela 1 – Metodologia de pesquisa

Tipo	**Objetivos**	**Abordagem**	**Procedimentos Técnicos**	**Natureza**
Bibliográfica	Exploratório e Descritivo	Qualitativa	Estudo de Caso	Aplicada

Fonte: Elaborado pelo autor

Inicialmente, um levantamento de referencial teórico foi executado pelo estudo de documentos nas seguintes bases de dados: Scopus, IEEExplore, ACM Library, Google Scholar e Periódicos UFSC. As pesquisas foram elaboradas de forma sistemática, seguindo a metodologia proposta por Kitchenham (2004) para análise de trabalhos relacionados à detecção e reconhecimento de expressões faciais em grupos de pessoas e em tempo real. As pesquisas foram restringidas majoritariamente a um período de tempo recente (a partir de 2018) por conta da natureza do problema, restringindo o uso de materiais mais antigos para abordagens nas quais se fez necessário.

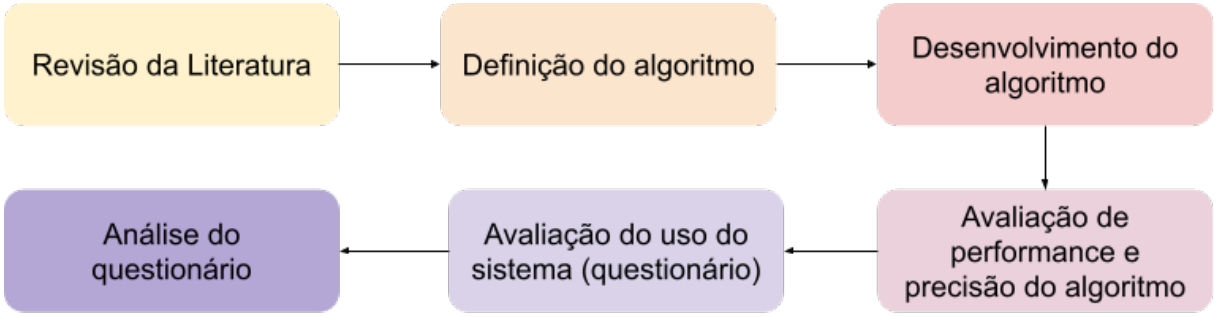
A partir dos conhecimentos adquiridos no levantamento literário, um algoritmo próprio para ser utilizado na detecção e reconhecimento de expressões faciais em grupo foi definido, desenvolvido e avaliado de forma técnica quanto ao seu desempenho e precisão nas classificações. O sistema foi então aplicado em grupos de estudantes em ambiente de ensino acompanhado por seus educadores.

Para validar os resultados gerados pelo sistema de reconhecimento facial em um ambiente de sala de aula realista, um experimento foi feito com alunos do curso PPGTIC da UFSC. Neste experimento, os alunos foram convidados a assistir três vídeos de conteúdos distintos e, após a exibição de cada vídeo, aplicou-se um questionário aos participantes. O questionário foi composto por perguntas relacionadas à experiência dos alunos com o conteúdo e suas percepções emocionais durante os vídeos. As questões buscaram identificar as emoções que os alunos sentiram que mais expressaram, alinhando as percepções subjetivas dos alunos com os dados objetivamente coletados pelo sistema. Este método permitiu uma comparação direta entre a efetividade do sistema de reconhecimento facial e as experiências reais dos usuários, fornecendo uma valiosa análise qualitativa e quantitativa. Este questionário foi desen-

volvido na ferramenta Google Forms e contou com perguntas dissertativas e optativas relevantes para o experimento.

A descrição dos passos realizados nesta pesquisa podem ser visualizados na Figura 1.

Figura 1 – Demonstração dos fluxos das entidades metodológicas do trabalho.



Fonte: Elaborado pelo autor

1.5 ESTRUTURA DO TRABALHO

Para facilitar a leitura e compreensão dos temas apresentados neste trabalho, o mesmo será organizado em 5 capítulos.

No **primeiro capítulo** uma introdução ao tema é apresentada, de forma a ambientar o leitor aos assuntos que serão discorridos nos capítulos subsequentes. Neste mesmo capítulo, a motivação e as justificativas para desenvolvimento deste projeto são expostas, juntamente com os objetivos geral e específicos. Este capítulo compreende ainda a metodologia adotada na execução de todas as etapas do projeto, a estrutura do documento e a apresentação dos argumentos que conferem a aderência às linhas de pesquisa do Programa de Pós-Graduação em Tecnologias da Informação e Comunicação.

O **segundo capítulo** trata da apresentação da metodologia e dos resultados encontrados na revisão da literatura em termos de bases de dados, principais abordagens aplicadas nos trabalhos avaliados, resultados e métodos de destaque. Ainda neste capítulo, é feita uma apresentação dos conceitos fundamentais para a compreensão dos assuntos que serão abordados ao longo deste documento. Inicialmente será apresentado um fluxo geral para treinamento e utilização dos modelos de IA, bem como o funcionamento das principais técnicas e tecnologias empregadas no desenvolvimento do sistema de detecção e reconhecimento de expressões faciais em tempo real.

Em seguida, no **terceiro capítulo**, é apresentada a descrição do modelo a ser implementado, bem como os datasets escolhidos para compor o conjunto de trei-

namento do modelo. Algumas preparações e padronizações aplicadas nos datasets também são apresentadas. Além disso, nesse capítulo serão descritas todas as etapas do processo de criação do sistema como um todo, as ferramentas utilizadas, os tempos de execução e os resultados técnicos obtidos a partir da aplicação do modelo desenvolvido em imagens determinadas previamente.

No **quarto capítulo**, é descrito o experimento realizado com a turma, representando uma situação real de ensino. Neste capítulo são apresentados os resultados obtidos pelo modelo criado neste trabalho em comparação com a percepção dos estudantes quanto às expressões durante o experimento.

O **quinto capítulo** apresenta as discussões e conclusão do trabalho, elencando os pontos positivos e negativos de todo o processo, desde a revisão da literatura até a aplicação em situações reais de práticas de ensino. Neste capítulo, trabalhos futuros e a visão do autor sobre o projeto também são apresentados.

Por fim, nos elementos pós-textuais, são apresentadas todas as referências utilizadas no decorrer do trabalho, seguidas dos apêndices e anexos.

2 INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO

A IA é um campo da ciência da computação que se dedica ao desenvolvimento de algoritmos e sistemas que possam simular a inteligência humana, com a capacidade de tomar decisão e resolver problemas (BELLMAN, 1978). O principal conceito da IA, é que ao aplicar técnicas de Aprendizado de Máquina (ML), o sistema seja capaz de analisar o ambiente em que se encontra e adaptar-se a novas circunstâncias no mesmo (DELLERMANN *et al.*, 2019).

Um dos principais objetivos da IA é desenvolver sistemas capazes de realizar tarefas que exigem, ou até então exigiam, inteligência humana (SILVA, J. A. S. da; MAIRINK, 2019), como reconhecer padrões, resolver problemas e tomar decisões complexas. Além disso, a IA busca criar sistemas que possam aprender e se aperfeiçoar com base em suas próprias experiências, tornando-se mais eficientes ao longo do tempo.

Atualmente, a IA é aplicada em diversas áreas, como robótica, reconhecimento de fala e imagem, processamento de linguagem natural, jogos, sistemas de recomendação, entre outros.

A aplicação da inteligência artificial na educação é um campo em crescimento promissor. Com a evolução da tecnologia, surgem novas formas de facilitar e melhorar o processo de ensino-aprendizagem. A IA pode ser utilizada para personalizar o ensino, adaptando o conteúdo e a metodologia de acordo com as necessidades individuais (SOUZA *et al.*, s.d.).

2.1 RECONHECIMENTO DE EXPRESSÃO FACIAL EM GRUPOS

Reconhecimento computacional de expressões faciais ainda é uma tarefa desafiadora. Apesar de ser uma atividade muito natural no ponto de vista de interações humanas, trata-se de um campo de visão computacional que demanda grande poder computacional e desenvolvimento, sendo, portanto, um problema ainda não completamente resolvido para execução em máquinas (BISWAS; SIL, 2015). No entanto, as aplicações que podem se beneficiar desse tipo de tecnologia são diversas, abrangendo áreas como robótica, marketing digital e educação (LOPES, Andre Teixeira; DE AGUIAR; OLIVEIRA-SANTOS, 2015). Embora essa tecnologia ainda esteja em aberto, na literatura há muitos trabalhos que foram capazes alcançar bom desempenho e resultados em geral. Redes Neurais Artificiais (RNAs) (ALI, H. *et al.*, 2015; LOPES, André Teixeira *et al.*, 2017; JAIN; SHAMSOLMOALI; SEHDEV, 2019) e outros métodos como Máquinas de Vetor de Suporte(SVM) e Lógica Fuzzy (ALI, G.; IQBAL; CHOI, 2016; HAPPY; ROUTRAY, 2014; BISWAS; SIL, 2015; GHASEMI; AHMADY, 2014) são alguns exemplos de técnicas aplicadas a esse tipo de problema.

Mapeamentos anteriores da literatura recente sobre FER foram desenvolvidos

(CANAL *et al.*, 2022; LI, S.; DENG, W., 2020), mas nenhum deles concentrou-se em algoritmos capazes de detectar e reconhecer emoções em grupos de pessoas (FER em nível de grupo). O indivíduo sempre foi o objeto de estudo e, por causa disso, o FER em nível de grupo ainda está em seus estágios iniciais em comparação com o FER individual (QUACH *et al.*, 2022). Com o avanço do FER em nível de grupo, um novo leque de aplicações é possibilitado como sistemas para realizar a seleção automática de fotos para um álbum de fotos, por exemplo. Esse tipo de algoritmo também pode ser empregado para auxiliar cientistas sociais e pesquisadores no campo da educação a analisar as interações entre estudantes no processo de aprendizado colaborativo (HUANG, X. *et al.*, 2019).

Os resultados apresentados em Canal *et al.* (2022) demonstram uma clara tendência na utilização de algoritmos de Redes Neurais Convolucionais (CNNs) em comparação com outros algoritmos clássicos de IA (os detalhes sobre esse tipo de RNA serão explorados nas próximas seções). Os métodos propostos na literatura que utilizam CNN, não são necessariamente os métodos que alcançaram o melhor resultado na classificação de expressões faciais, contudo, vários fatores devem ser levados em consideração nessa análise. Os principais fatores para este estudo são: o dataset (seção 2.4.1) utilizado nos experimentos e; o tempo de execução do algoritmo.

Os trabalhos mencionados possuem foco nos métodos computacionais relacionados ao FER. Todo o processo envolvido nesse contexto será explorado na sequencia deste trabalho, seguido dos resultados de uma aplicação no contexto educacional. Desta forma, é necessário compreender as etapas para desenvolvimento de um modelo computacional capaz de atuar no reconhecimento de expressões facial a nível de grupo.

2.2 PIPELINE GERAL PARA RECONHECIMENTO DE EXPRESSÕES FACIAIS

Em revisão à literatura existente sobre FER a nível de grupo, pôde-se observar aspectos interessantes que indicam um fluxo computacional geral, aplicável para resolver o desafio em questão a partir vídeos. Esse fluxo pode ser ilustrado de maneira geral conforme a Figura 2, da seguinte forma: (i) etapa de aquisição de frame (imagem): imagens de entrada obtida do vídeo em tempo real do publico alvo de detecção das expressões faciais. Este frame pode ser submetido a alguns algoritmos de pre processamento como redimensionamento ou alteração de cores (colorido para preto e branco, por exemplo), conteúdo, trata-se apenas de operações simples a serem aplicadas nessa etapa, tornando o frame adequado para seguir o fluxo de processamento; (ii) etapa de detecção de faces: algumas abordagens têm a capacidade de aplicar algoritmos de FER na imagem integral do grupo (PETROVA; VAUFREYDAZ; DESSUS, 2020; SUN, M. *et al.*, 2020; SRIVASTAVA *et al.*, 2020). Nesse contexto, a execução desta etapa poderia ser considerada dispensável. No entanto, existem dois motivos

fundamentais pelos quais essa etapa é adotada na maioria dos métodos encontrados na literatura:

- A ausência desta etapa resultou, em geral, em desempenhos inferiores quando comparada àqueles métodos que a incorporaram;
- Conjuntos de dados de treinamento que contemplam expressões faciais individuais são substancialmente mais prevalentes e difundidos na literatura;

Esses fatores destacam a relevância da detecção de faces como uma etapa preliminar crucial no processo de reconhecimento de expressões faciais, contribuindo significativamente para a precisão e eficácia dos métodos propostos; (iii) etapa de seleção/extração de características: tem como objetivo extrair características relevantes do sinal de entrada para discretizá-lo e torná-lo único para cada indivíduo - esta etapa funciona como uma assinatura ou impressão digital para cada emoção facial específica; (iv) etapa de classificação: para cada uma das faces detectadas no processo anterior, um algoritmo de classificação é aplicado para identificar a expressão facial do indivíduo em questão; (v) Análise do conjunto de expressões: após a identificação das expressões faciais de cada indivíduo previamente detectado, um algoritmo é empregado para calcular a expressão facial média do quadro. Esta abordagem se justifica pelo contexto de análise do grupo como um todo, em detrimento do foco individual.

Os passos mencionados acima, ilustrados pela Figura 2, podem ser parcial ou totalmente identificados em uma ampla variedade de abordagens atualmente presentes na literatura. Para cada bloco de etapas, os algoritmos utilizados podem variar de acordo com a particularidade do problema abordado, bem como as características do conjunto de dados utilizado para treinamento.

Desta forma, nas próximas seções, serão explorados os algoritmos de detecção de face, bem como o reconhecimento individual de expressões faciais que, juntos, compões os dois processos básicos da abordagem proposta por este trabalho.

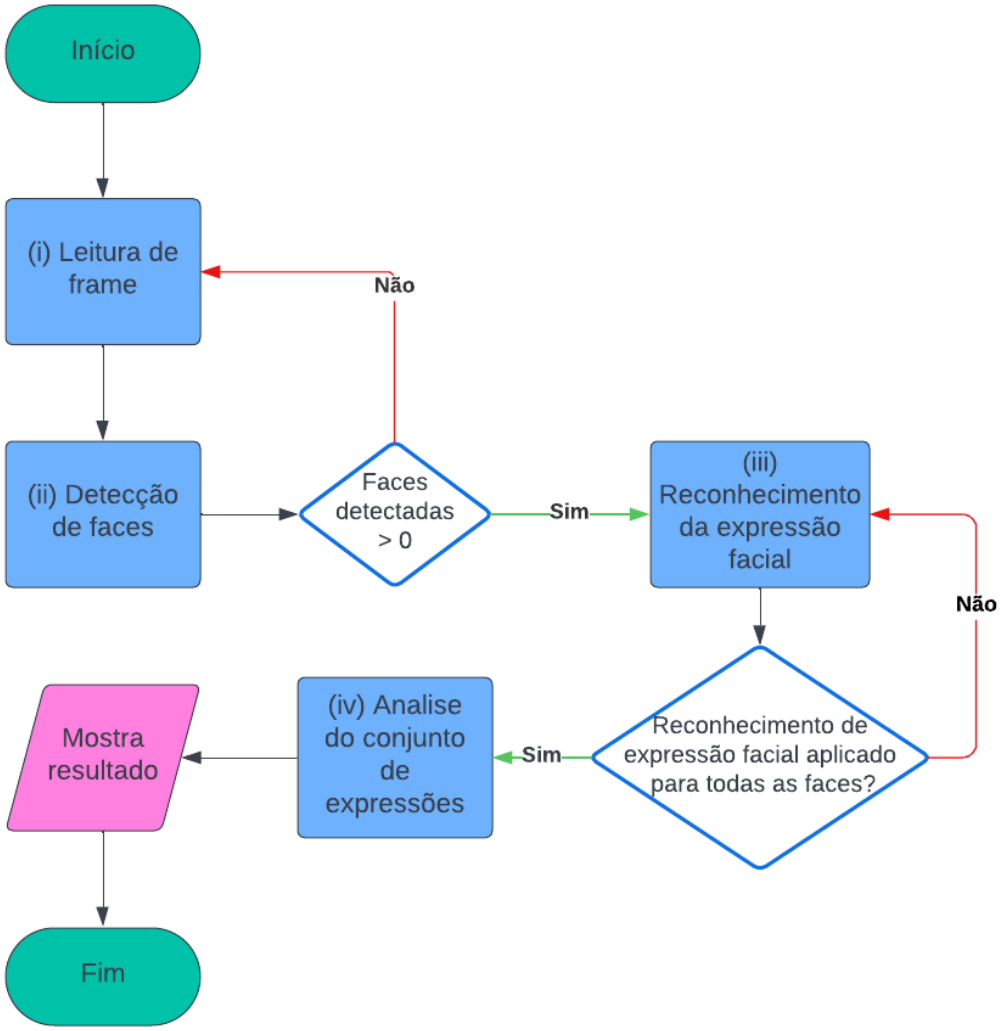
2.3 DETECÇÃO DE FACES

Uma das etapas mais relevantes no reconhecimento de expressões faciais é a seleção inicial da região de interesse na imagem de entrada (GHASEMI; AHMADY, 2014). Existem abordagens algorítmicas eficazes capazes de encontrar a região de interesse (a região facial) mesmo com a presença de perturbações problemáticas (ruído) que podem afetar seu desempenho. Normalmente, esse tipo de algoritmo precisa lidar com variações nas imagens, como pose e iluminação (ANILA; DEVARAJAN, 2012).

As três principais abordagens presentes na literatura para realizar a detecção de faces em imagens são Viola-Jones, Haar Cascade e CNNs.

- **Viola-Jones**: Um dos algoritmos mais conhecidos e amplamente utilizados para detecção de faces. Este algoritmo utiliza um classificador em cascata

Figura 2 – Demonstração do fluxo geral para detecção de expressões faciais em grupo.



Fonte: Elaborado pelo autor, 2024

treinado com o algoritmo AdaBoost para identificar regiões de interesse que podem conter uma face. Esse método é conhecido por sua eficiência e precisão na detecção de faces em tempo real.

- **Haar Cascade**: Outro método popular, o Haar Cascade, também se baseia no uso de classificadores em cascata, mas utiliza recursos de Haar para identificar características em uma imagem. Esse método é amplamente utilizado devido à sua capacidade de detecção robusta e eficiente.
- **CNNs**: As Redes Neurais Convolucionais revolucionaram a área de visão computacional, incluindo a detecção de faces. Elas são capazes de aprender representações hierárquicas de características faciais diretamente das

imagens, o que as torna extremamente eficazes na detecção de faces em uma ampla variedade de condições.

Esses métodos têm sido extensivamente estudados e comparados em termos de desempenho e robustez (DAVID *et al.*, 2022; MENDES *et al.*, 2019; COSTA *et al.*, 2021). A escolha do método mais adequado depende das características específicas da aplicação, tais como restrições de tempo, qualidade das imagens e disponibilidade de recursos computacionais.

Ao longo desta seção, serão abordados os principais conceitos, algoritmos e técnicas utilizados na detecção e segmentação de imagens faciais, bem como suas aplicações no reconhecimento de expressões faciais. Além disso, serão discutidos os desafios e limitações associados a essas abordagens.

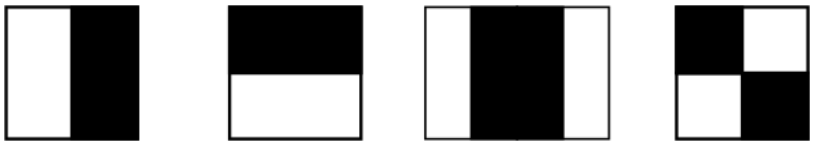
2.3.1 Viola-Jones

O algoritmo desenvolvido por Paul Viola e Michael J. Jones em 2003 teve um impacto significativo na área de detecção facial, pois, além de apresentar grande precisão, possui relativamente baixo custo computacional. Embora ainda seja amplamente utilizado, requer uma etapa de treinamento que pode ser demorada. No entanto, após esse treinamento, a detecção facial ocorre de forma mais eficiente.

Conforme indicado pelos autores (JONES; VIOLA, 2003), essa técnica baseia-se nas características de Haar-like (utilizado para detecção de pedestres), que é determinada pela soma dos pixels dentro da área de retângulos brancos, subtraindo, em seguida, a área dos retângulos pretos (Figura 3), permitindo a detecção de padrões em uma imagem. Embora existam filtros mais avançados, a eficiência oferecida por Haar acaba sendo vantajosa, conforme destacado por (VIOLA; JONES *et al.*, 2001). O detector Viola-Jones realiza a localização facial e tem seu algoritmo construtivo baseado em quatro etapas:

1. Seleção de características retangulares Haar-Like, como demonstrado na Figura 3;
2. Criação da imagem integral que permite cálculos rápidos das características Haar mencionadas anteriormente;
3. Treinamento de um classificador baseado em AdaBoost capaz de selecionar características relevantes;
4. Utilização de classificadores em cascata que descartam regiões onde é improvável a existência de um rosto, concentrando-se em regiões prováveis de conter um rosto.

Figura 3 – Exemplos de características Haar.



Fonte: Elaborado pelo autor, 2024

2.3.2 Haar Cascade

O Classificador de Cascata Haar é um método de detecção de objetos que insere recursos Haar em uma série de classificadores para identificar objetos em uma imagem. As características de Haar são baseadas em ondas de Haar, em vez das intensidades usuais de imagem. Segundo (SOO, 2014), esse método foi adotado e desenvolvido por Viola e Jones (VIOLA; JONES *et al.*, 2001), embora tenha sido proposto por (PAPAGEORGIOU; OREN; POGGIO, 1998). O conjunto de recursos alternativos surgiu devido ao custo computacional do cálculo de recursos em uma imagem com pixels RGB.

Uma característica Haar-like considera regiões retangulares adjacentes em uma posição específica na janela de detecção, soma as intensidades de pixel em cada região e calcula a diferença entre essas somas. A diferença calculada pode ser usada para identificar subseções na imagem. Os classificadores Haar devem ser treinados fornecendo muitas imagens (positivas e negativas em termos de conter o objeto a ser detectado). Feito isso, as características podem ser extraídas usando janelas deslizantes de blocos. Esses blocos são semelhantes a matrizes de convolução aplicadas em diferentes partes da imagem para encontrar correspondências parciais.

Posteriormente, mesmo sendo capaz de calcular rapidamente as intensidades usando imagens integrais, muitos recursos extraídos nas etapas anteriores não agregam valor ao resultado. Se uma característica for detectada em uma região que não é de interesse, ela não tem valor real para o classificador. Para isso, uma técnica de reforço é aplicada no classificador Haar Cascade, chamada AdaBoost (FREUND; SCHAPIRE, 1997). Essa técnica reduz significativamente o número de classificadores, sem prejudicar o resultado final.

O Classificador de Haar Cascade é composto pelas seguintes etapas:

1. Uma seção da imagem é selecionada como região de trabalho;
2. A primeira etapa avalia a presença de uma característica conforme detalhado anteriormente;
3. Se o retorno da etapa anterior for positivo, o classificador passa para a próxima etapa, e assim por diante até a última etapa. Por outro lado, se o retorno da etapa anterior for negativo, outra seção da imagem é selecionada,

e o processo é reiniciado;

4. Finalmente, se a imagem for completamente avaliada, o classificador encerra sua execução.

O algoritmo recebe esse nome justamente pela sua aplicação em cascata (cascade) que o ocorre em pequenos segmentos da imagem para identificar o objeto de interesse. Apesar de essa abordagem ser utilizada em diversos métodos de reconhecimento e detecção de objetos, mostrou-se muito eficaz para o problema de localização facial e é altamente difundido na literatura que diz respeito a esse contexto.

2.3.3 Rede Neural Convolucional

As CNNs são, atualmente, abordagens muito corriqueiras no âmbito de reconhecimento e classificação de imagens, detecção de objetos e, evidentemente, localização/reconhecimento facial. As classificações fornecidas pelas CNNs (ALBAWI; ABED MOHAMMED; ALZAWI, 2017) ocorrem quando uma imagem é capturada e classificada da seguinte maneira: o modelo visualiza a imagem de entrada como uma matriz de pixels; esta matriz passará por uma série de filtros de convolução, pooling e funções, como softmax, que classifica um objeto com valores probabilísticos entre 0 e 1; dessa forma, com uma rede treinada, é possível utilizar essas funções para classificar imagens nunca "vistas" pela rede anteriormente (viés zero) com uma precisão consideravelmente boa. A Seção 2.4.2.3 fornecerá detalhes adicionais sobre o funcionamento de uma rede neural convolucional.

Uma arquitetura de CNN que se tornou muito popular devido ao alto desempenho em detecção facial foi a *Multi-Task Cascaded Convolutional Neural Network* (MTCNN) (ZHANG, K. *et al.*, 2016). Experimentos realizados em conjuntos de dados de detecção facial em 2016 demonstraram um desempenho superior a todos os outros algoritmos testados até o momento. Diversas implementações dessa arquitetura estão disponíveis para uso por meio de bibliotecas em Python, popularizando ainda mais o algoritmo.

2.4 RECONHECIMENTO DE EXPRESSÕES FACIAIS (FER)

O reconhecimento de expressões faciais é um campo de estudo de grande relevância no processamento de linguagem natural e na inteligência artificial, pois é essencial para o entendimento das emoções humanas e para a comunicação não-verbal. Essa tarefa envolve a análise de imagens de faces, a detecção de características relevantes e a classificação das emoções faciais com base em algoritmos de aprendizado de máquina e inteligência artificial.

Apesar de todas as operações serem de extrema importância para alcançar bons resultados ao classificar emoções a partir das expressões faciais, geralmente

a etapa mais valorizada em todo o processo é a classificação. Essa fase é também aquela que apresenta a maior diversidade em termos de métodos entre os trabalhos na literatura (CANAL *et al.*, 2022). Nos últimos anos, ocorreu um aumento na utilização de métodos fundamentados em arquiteturas de redes neurais, sobretudo em decorrência do aumento de poder computacional no âmbito de *Hardware*.

A escolha do modelo de inteligência artificial reflete diretamente na qualidade do resultado a ser atingido, contudo, a qualidade do conjunto de dados de treinamento pode ter um impacto significativo no desempenho e na precisão dos sistemas de reconhecimento facial de emoções. Um modelo de rede neural convolucional (ConvNet) bem construído, por exemplo, quando treinado em um conjunto de dados abrangente e diversificado, pode levar a avanços significativos no reconhecimento facial de emoções (DEBNATH *et al.*, 2022). Portanto, é fundamental selecionar cuidadosamente tanto o modelo de inteligência artificial quanto o conjunto de dados de treinamento para garantir o desenvolvimento de sistemas de reconhecimento facial de emoções robustos e precisos. Essa observação sublinha a importância de considerar ambos os fatores no desenvolvimento dos sistemas de reconhecimento facial de emoções.

Na seção seguinte (2.4.1), serão apresentados os principais datasets presentes na literatura, bem como suas características e aplicações específicas.

2.4.1 Datasets

Conjuntos de dados, ou datasets, contendo imagens faciais ou vídeos especificamente desenvolvidos para o reconhecimento de expressões faciais não são uma novidade. Os primeiros estudos são de Sakai, Nagao e Fujibayashi (1969), e desde então, a necessidade de desenvolver abordagens computacionais para inspecionar emoções ou comportamentos em imagens e vídeos faciais tem crescido significativamente. Como consequência, muitos conjuntos de dados de expressões faciais surgiram, variando em seu ambiente de aquisição, número de expressões reconhecíveis, regiões, entre outras características (PANTIC *et al.*, 2005). Por outro lado, devido à maneira como alguns desses conjuntos de dados foram criados, muitos outros pontos também foram incluídos ao longo do tempo, como gênero, etnia, idade, qualidade da imagem (tamanho, cor, conjuntos de dados mais antigos realizaram uma digitalização de fotos físicas) e número de participantes. Todas essas variáveis tendem a influenciar a qualidade do conjunto de dados escolhido, bem como impactar nos resultados que as abordagens computacionais para reconhecimento de emoções produzem.

Uma vez que os conjuntos de dados desempenham um papel fundamental para o reconhecimento baseado em aprendizado orientado a dados e reconhecimento de emoções (Shin, 2016), esta seção visa apresentar uma revisão dos conjuntos de dados mais importantes para o reconhecimento de emoções em imagens ou vídeos, reconhecidos e amplamente utilizados na literatura. Adicionalmente, nesta seção são

destacados detalhes relevantes de implementação usados para a construção desses conjuntos de dados, bem como suas particularidades e possíveis desvantagens.

2.4.1.1 JAFFE

O conjunto de dados Japanese Female Facial Expression (JAFFE)(LYONS *et al.*, 1998) é um dataset gratuito publicado em 1998 por um grupo de pesquisadores japoneses do *ATR Human Information Processing Research Laboratory* e do Departamento de Psicologia da Universidade de Kyushu. O JAFFE apresenta imagens de 10 mulheres japonesas expressando as seguintes emoções: felicidade, tristeza, surpresa, raiva, nojo, medo e neutro. Para cada participante foram capturadas de 3 a 4 fotos referentes a cada emoção, olhando através de uma folha de plástico semi-reflexiva em direção à câmera, totalizando 219 imagens. A câmera do experimento foi posicionada dentro de uma caixa preta para prevenir e mitigar reflexos de luz. O cabelo foi removido da frente do rosto de maneira que as expressões ficassem mais evidentes.

Como o processo fotográfico foi realizado de forma analógica, as fotos foram digitalizadas posteriormente. A resolução digital final de cada imagem é de 256x256 pixels. As imagens foram avaliadas para cada emoção por um grupo de 92 graduandas japonesas em uma escala de 5 pontos. Cada imagem tem um vetor de 5 ou 6 componentes representando as avaliações médias para cada emoção. Apesar de não ser um dataset atual, este conjunto de dados tem sido utilizado em várias abordagens de reconhecimento de emoções ao longo dos anos (MEHTA; JADHAV, 2016; ALI, H. *et al.*, 2015; BISWAS; SIL, 2015).

2.4.1.2 Cohn-Kanade

O Cohn-Kanade AU-Coded Expression Database (CK) (KANADE; COHN; TIAN, 2000), originalmente chamado de CMU-Pittsburgh AU-Coded Facial Expression Image Database, é uma fonte de dados bem estabelecida e amplamente utilizada na literatura desde a sua primeira versão (YEASIN; BULLOT; SHARMA, R., 2006; SUN, J.-M.; PEI; ZHOU, 2008; CHEN *et al.*, 2012; LI, Y. *et al.*, 2013; JAIN; SHAMSOLMOALI; SEHDEV, 2019). Criado pelo Grupo de Análise de Afeto do Laboratório de Pesquisa da Universidade de Pittsburgh (Affect Analysis Group of the Research Lab at the University of Pittsburgh - AGG), o CK foi desenvolvido para superar a escassez de conjuntos grandes de imagens faciais, apoiando estudos de rastreamento e análise de características faciais.

A primeira versão do CK, lançada em 2000, inclui 486 sequências de cerca de 97 diferentes sujeitos, com uma distribuição demográfica de 69% mulheres, 31% homens, 81% euro-americanos, 13% afro-americanos e 6% de outros grupos. As sequências variam de uma expressão neutra a uma expressão de pico, sendo a última totalmente codificada pelo Sistema de Codificação de Ação Facial (FACS) e rotulada

com uma emoção específica. Importante ressaltar que os rótulos de emoção referem-se à expressão solicitada, e não necessariamente à expressão real do participante. As imagens foram capturadas em um ambiente controlado em termos de iluminação, fundo e posição da câmera, e todas as expressões foram realizadas a pedido dos pesquisadores.

A segunda e mais recente versão do banco de dados, chamada de Extended Cohn-Kanade Dataset (CK+) (LUCEY *et al.*, 2010b), foi lançada em 2010. Esta versão incluiu 107 sequências de transições emocionais e 26 sujeitos, com a adição de imagens não posadas e cada imagem recebendo uma etiqueta de emoção nominal baseada na impressão do sujeito sobre cada uma das 7 categorias básicas de emoções: Raiva, Desprezo, Nojo, Medo, Felicidade, Tristeza e Surpresa. As imagens do CK+ oferecem resoluções de 640x490 ou 640x480 e listagem de pixels em escala de cinza de 8 bits ou valores de cor de 24 bits.

Atualmente, os autores estão trabalhando na terceira versão do banco de dados, com a intenção de adicionar imagens sincronizadas de 30 graus a partir do vídeo frontal original, possibilitando análises como reconhecimento facial e emocional em 3D.

2.4.1.3 FER2013

A base de dados FER2013 foi originalmente publicada na Conferência Internacional de Aprendizado de Máquina (ICML em 2013) (GOODFELLOW *et al.*, 2013). Este conjunto de imagens é publicamente acessível e foi criado para um projeto da Pierre-Luc Carrier e Aaron Courville, sendo posteriormente compartilhado publicamente para uso em uma competição de reconhecimento de expressões faciais no Kaggle[1].

Este conjunto de dados consiste em 35.887 imagens de rostos com 48x48 pixels em escala de cinza, todas as imagens são rotuladas para sete expressões faciais e distribuídas da seguinte forma: 4953 imagens de raiva, 547 de nojo, 5121 de medo, 8989 de felicidade, 6077 de tristeza, 4002 de surpresa e 6198 neutras. O conjunto de dados foi desenvolvido usando a API de busca de imagens do Google, na qual foram buscadas imagens de rostos que correspondessem a um conjunto de 184 palavras-chave relacionadas a emoções, como “feliz”, “triste”, entre outras.

Durante a competição do Kaggle, 28709 imagens deste conjunto de dados foram compartilhadas entre os participantes para treinar suas redes neurais, e 3589 imagens foram usadas no conjunto de teste e validação para determinar o algoritmo de reconhecimento vencedor na competição. Após o término da competição, o conjunto de dados foi disponibilizado para o público em geral. Além do grande número de imagens, este conjunto de dados possui uma resolução espacial muito reduzida, visto que seu propósito é ser utilizado como entrada para treinar e testar classificadores computacionais.

[1] https://www.kaggle.com

2.4.1.4 MMI

O conjunto de dados MMI foi constituído a partir da colaboração de 25 indivíduos pertencentes a diversas etnias, incluindo Europeia, Asiática e Sul-Americana. A distribuição por gênero neste conjunto é de 44% feminino e 56% masculino, com idades variando entre 19 e 62 anos. As expressões faciais foram capturadas de maneira espontânea, isto é, os participantes foram incentivados a manifestar expressões faciais autênticas através da visualização de vídeos estimulantes, como conteúdos humorísticos e repulsivos, visando assim assegurar uma maior fidedignidade nas expressões naturais. Este conjunto de dados foi objeto de estudo em diversas publicações acadêmicas, como demonstram os trabalhos de Mohseni, Zarei e Ramazani (2014), Cruz, Bhanu e Thakoor (2014) e Hu *et al.* (2019).

Diferenciando-se de outras bases de dados contemporâneas que se limitavam a analisar apenas seis expressões faciais básicas, o MMI inclui uma ampla gama de expressões não-básicas e indefinidas, as quais foram registradas durante trocas expressivas dinâmicas. As imagens do banco de dados, todas em cores reais, passaram por um processo de digitalização, resultando numa resolução de 720x576 pixels. Os vídeos foram gravados a 24 quadros por segundo, com sequências variando de 40 a 520 quadros, partindo de expressões neutras para expressivas e neutras novamente. Ao todo, o MMI compreende uma hora e trinta e dois minutos de material audiovisual. É importante salientar que os fundos utilizados nas fotografias e vídeos não são uniformes em todas as amostras. O acesso ao conjunto de dados MMI é facilitado pela disponibilidade online, sendo livremente acessível para a comunidade científica.

2.4.2 Algoritmos de classificação

Por muitos anos, as abordagens clássicas para o processamento de imagens, em geral, representaram a melhor tentativa de solucionar alguns problemas com alto custo de processamento. Com o passar do tempo, e o desenvolvimento das capacidades computacionais, aliado ao surgimento de novas arquiteturas, novos métodos surgiram por múltiplas razões, incluindo a complexidade de implementação e custo de execução, possibilitando e a obtenção de resultados expressivos.

Dentre as abordagens presentes na literatura, pode-se elencar os principais algoritmos como:

- SVM
- DBN
- CNN
- Redes Neurais Recorrentes (RNN) e Memória de Longo e Curto Prazo (LSTM)

Dentre estes algoritmos, pode-se observar alguns métodos mais clássicos da área da visão computacional, bem como métodos de aprendizado profundo (CNN e RNN), conforme proposto em Canal *et al.* (2022). Nas próximas seções, serão explorados esses algoritmos de forma geral em termos do seu funcionamento e caracteristicas.

2.4.2.1 Máquinas de Vetores de Suporte

As SVMs (Máquinas de Vetores de Suporte) são métodos de aprendizado supervisionado, isto é, algoritmos capazes de gerar funções de mapeamento entrada-saída a partir de um conjunto específico de dados rotulados com um ou mais vetores de características. As SVMs classificam elementos determinando fronteiras, conhecidas como hiperplanos, que separam as classes entre si. Embora esse hiperplano possa ser orientado de diferentes formas e ainda cumprir seu propósito, o objetivo da SVM é orientá-lo de tal maneira que esteja o mais distante possível do vetor mais próximo (denominado vetor de suporte) de cada classe (HUANG, S. *et al.*, 2018). Inicialmente, essas técnicas de aprendizado de máquina foram desenvolvidas em 1963 para realizar classificação em conjuntos de dados linearmente separáveis (VAPNIK, 1963). No entanto, com o auxílio de funções de *kernel* não lineares, é possível transformar os dados de entrada em um espaço de características de alta dimensão no qual os dados se tornam linearmente separáveis e, portanto, classificáveis pelo algoritmo SVM (WANG, L., 2005).

Dentre os algoritmos mais clássicos, SVM é o método mais utilizado na literatura para classificação de expressões faciais (BAILENSON *et al.*, 2008; CRUZ; BHANU; THAKOOR, 2014; ESKIL; BENLI, 2014; LUO *et al.*, 2017; KARTALI *et al.*, 2018; WANG, F. *et al.*, 2019).

Devido à sua simplicidade em termos de poder de processamento necessário (em comparação com algoritmos de aprendizado profundo), SVM é uma ferramenta excelente para aprendizado de máquina e, com o auxílio de *kernels* não lineares, é capaz de competir com os métodos mais novos e complexos em termos de resultados. Além do reconhecimento de expressões, as SVMs têm sido utilizadas para resolver a classificação em diversas áreas nos últimos anos, incluindo a medicina e muitas outras (ZHANG, Y.-D. *et al.*, 2016).

2.4.2.2 Rede Bayesiana Dinâmica

Uma DBN representa uma extensão formal das Redes Bayesianas, adaptada para modelar séries de dados temporais ou sequências. Redes Bayesianas, de acordo com Silander e Myllymaki (2012), são modelos gráficos probabilísticos que representam um conjunto de variáveis e suas dependências condicionais via um grafo acíclico dirigido. Cada nó no grafo representa uma variável, e as arestas representam dependências probabilísticas entre estas variáveis.

As DBNs estendem as Redes Bayesianas para lidar com dados que variam no tempo. Elas modelam processos estocásticos e são especialmente úteis quando as observações são sequenciais e interdependentes. Sendo assim, em uma DBN, os nós são replicados ao longo do tempo, representando diferentes instâncias temporais das variáveis. Isso permite modelar transições de estado e a evolução de variáveis ao longo do tempo.

Algoritmos de inferência e aprendizado são utilizados para atualizar as variáveis ocultas, dada a evidência observada. Métodos como o filtro de Kalman e algoritmos de suavização e previsão são frequentemente aplicados. Apesar de ser um método que contém um certo nível de complexidade, principalmente no seu desenvolvimento, as DBNs podem ser aplicadas em uma alta gama de processos temporais e são capazes de lidar com dados afetados por ruídos.

2.4.2.3 Rede Neural Convolucional

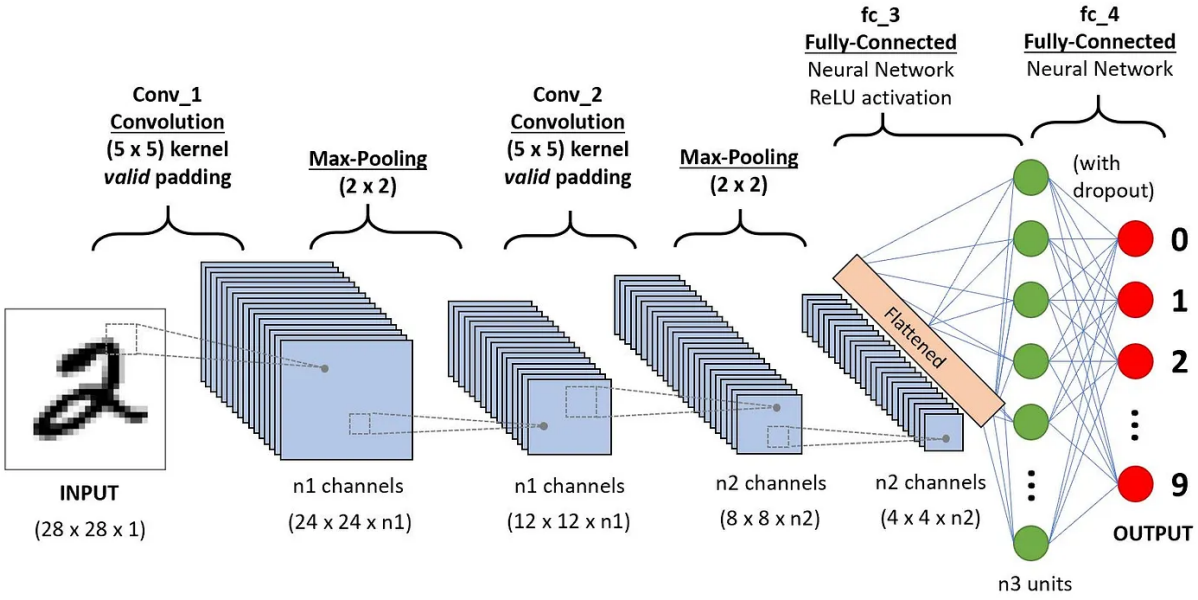
Abordar as Redes Neurais Convolucionais em uma seção sobre algoritmos de classificação pode ser um pouco superficial. A razão para isso é que, embora as CNNs sejam classificadores, elas não se limitam apenas a este papel, mas englobam a combinação da etapa de extração de características com a etapa de classificação. As redes neurais convolucionais surgiram como uma alternativa para a extração de características sem a necessidade do conhecimento humano específico sobre um determinado tópico.

A ideia principal do algoritmo é utilizar a operação de convolução para extrair características da imagem, tornando a informação da posição do pixel em relação ao seu entorno mais atômica para o processo de classificação. Este é o motivo para qual as CNNs são tão amplamente utilizadas em problemas da área de visão computacional, destacando-se pela capacidade de aprender características relevantes de forma autônoma, sem a intervenção ou preconcepções humanas.

Como o nome sugere, as CNNs são redes neurais que utilizam primordialmente camadas convolucionais. Essas camadas recebem uma entrada e aplicam um determinado número de filtros a ela, produzindo uma saída, conhecida como mapa de características (LI, Z. *et al.*, 2021). Os filtros são geralmente distintos para cada camada e são aprimorados durante o treinamento. Inicialmente, um determinado número de filtros aleatórios é aplicado à imagem de entrada e, à medida que o processo de treinamento evolui, a rede pode definir automaticamente quais são os filtros que melhor descrevem a entrada em termos de características relevantes para a classificação. Este processo destaca a capacidade adaptativa das CNNs em aprender e aperfeiçoar continuamente a identificação de características úteis para tarefas específicas, como a classificação, através da sua arquitetura de aprendizado profundo.

No caso da classificação de imagens, como demonstrado na Figura 4, a entrada

Figura 4 – Ilustração geral da estrutura de uma CNN - Neste exemplo, o objetivo da rede neural é identificar qual é o número escrito à mão (INPUT). Pode-se perceber que a entrada da rede neural possui dimensão [28, 28, 1], passando por diversas camadas intermediárias até a saída, onde a profundidade é 10, abrangendo todos os números de 0 a 9.



Fonte: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

geralmente se traduz em um tensor tridimensional, com dimensões para a largura, altura e profundidade (geralmente o número de canais de cor). Por exemplo, a entrada para o conjunto de dados MNIST teria as dimensões [28, 28, 1]. A saída da camada teria a mesma largura e altura que a entrada, mas com uma profundidade diferente. A profundidade da saída é uma função do número de filtros aplicados na camada anterior. Na última camada da rede, a profundidade da saída é igual ao número de classes que a rede está tentando classificar (GU *et al.*, 2018).

Essa estrutura reflete como as CNNs manipulam e transformam os dados de entrada ao longo das suas camadas, adaptando-se progressivamente para identificar as características mais relevantes para a classificação. A mudança na profundidade das camadas ilustra a evolução do processamento da informação desde a entrada bruta até a classificação final.

Além das camadas convolucionais, as camadas de agrupamento (*pooling*) são também comumente aplicadas em CNNs. O propósito dessa camada é reduzir a dimensionalidade dos dados, mantendo apenas as características mais relevantes, otimizando assim o custo computacional para processar a rede e prevenindo que a rede baseie sua classificação em características que, de fato, não são relevantes para o problema. Isso é realizado pela aplicação de um operador de agrupamento (como *max*

pooling, *mean pooling*, entre outros) como demonstrado na Figura 4.

2.4.2.4 Rede Neural Recorrente

As RNNs são um tipo de rede neural artificial onde as conexões não seguem um fluxo sequencial, ou seja, as ligações entre os nós podem criar um ciclo, permitindo que a saída de alguns nós afete a entrada subsequente dos mesmos nós (YU, Y. *et al.*, 2019). Isso resulta em redes neurais com estado, capazes de lembrar informações de entradas anteriores e usar essas informações em entradas futuras. As RNNs são ferramentas poderosas para analisar dados sequenciais, como texto, áudio e vídeo.

Essas redes são frequentemente utilizadas em tarefas como modelagem de linguagem e tradução automática, onde a ordem das palavras é importante. Além disso, podem ser usadas em tarefas como classificação de sentimentos e rotulagem de imagens, onde a ordem dos pontos de dados é relevante. Existem muitos tipos diferentes de RNNs, sendo o tipo mais comum a rede de Memória de Longo Curto Prazo. A LSTM é um tipo de RNN capaz de aprender dependências de longo prazo. Diferentemente das RNNs tradicionais, a LSTM possui uma célula de memória que pode lembrar informações por longos períodos de tempo.

A rede LSTM é composta por uma série de células LSTM. Cada célula possui um portão de entrada, um portão de saída e um portão de esquecimento. Os portões controlam o fluxo de informações para dentro e fora da célula (STAUDEMEYER; MORRIS, 2019). O portão de entrada controla o fluxo de informações da entrada para o estado da célula. O portão de saída controla o fluxo de informações do estado da célula para a saída. O portão de esquecimento controla o fluxo de informações do estado da célula para o estado esquecido.

Durante o treinamento da célula LSTM, os portões são usados para controlar o fluxo de informações para dentro e fora da célula. Um otimizador é aplicado para atualizar os pesos das células LSTM e, assim, minimizar a função de perda.

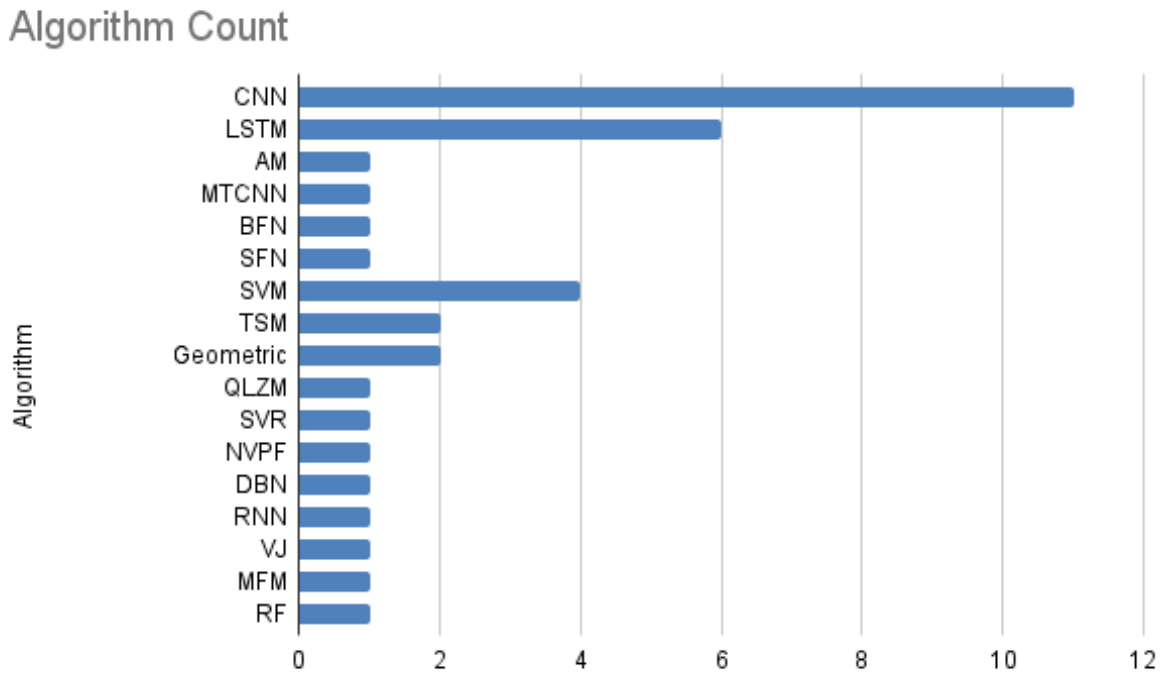
Geralmente, camadas LSTM são aplicadas juntamente com outras camadas, como uma camada totalmente conectada, que é usada para prever a saída, uma camada de *dropout*, que é usada para evitar o sobreajuste, e uma camada recorrente, que é usada para aprender dependências de longo prazo (LI, W. *et al.*, 2021).

As LSTMs podem ser empregadas em uma variedade de tarefas, como Processamento de Linguagem Natural (NLP), tradução automática e, no campo do processamento de imagens, podem ser usadas para classificação e rotulagem de imagens e detecção de objetos. A classificação de imagens pode se beneficiar das redes LSTMs devido à sua capacidade de aprender dependências de longo prazo, o que pode reduzir o risco de desaparecimento do gradiente que as RNNs tradicionais enfrentam.

2.5 TRABALHOS RELACIONADOS

Em pesquisa prévia na literatura, pode-se encontrar algumas abordagens de reconhecimento de expressões faciais, aplicadas a grupos de pessoas. Dentre esses trabalhos, é possível identificar uma grande variedade de algoritmos e métodos aplicados para a solução do problema (CANAL *et al.*, 2023). Contudo, dentre estas abordagens, é clara a distinção de dois principais algoritmos: CNN e LSTM, conforme demonstrado na figura 5.

Figura 5 – Contagem de ocorrência dos algoritmos dentre os trabalhos relacionados. Pode-se perceber a ampla utilização de CNNs e LSTMs.



Fonte: (CANAL *et al.*, 2023)

Como forma de avaliação dos artigos elencados nessa revisão, foi levada em conta a acurácia alcançada pelas propostas de cada autor. É importante destacar que o resultado dos trabalhos não depende somente dos algoritmos aplicados, mas também de diversos outros fatores, como dataset utilizado para treinamento e o própria forma de avaliação dos resultados aplicados por cada autor. Contudo, não seria possível apontar o melhor algoritmo de classificação se a comparação entre os artigos fosse feita de forma a considerar todos estes fatores. Assim sendo, a comparação entre os algoritmos de cada trabalho foi feita levando em conta a acurácia alcançada pelo modelo.

Os dois melhores resultados obtidos pelo uso de métodos CNN, dentre os estudos encontrados, foram os propostos por Guo *et al.* (2018) e Khan *et al.* (2018), com

pontuações de 78.98% e 78.39%, respectivamente. Ambos os métodos foram desenvolvidos e submetidos ao 6º Desafio de Reconhecimento de Emoções no Ambiente Natural (EmotiW 2018) (DHALL *et al.*, 2018).

Guo *et al.* (2018) apresenta uma rede híbrida de aprendizado profundo para resolver o problema de reconhecimento de emoções em grupo. O classificador proposto pelos autores foi desenvolvido para identificar três categorias distintas: emoções positivas, negativas e neutras. A abordagem híbrida proposta é uma fusão de 8 modelos no total, incluindo alguns modelos bem conhecidos como VGG-FACE, Inception-V2, SE-ResNet-50 e até uma implementação de LSTM. Algumas abordagens foram avaliadas separadamente, mas o melhor resultado foi alcançado com a fusão de todos os modelos (78.98%). Os modelos foram treinados e avaliados em uma combinação dos conjuntos de dados FER-2013 (GOODFELLOW *et al.*, 2013) e GENKI-4K (WHITEHILL *et al.*, 2009).

Por outro lado, Khan *et al.* (2018) desenvolveu um algoritmo que funciona basicamente em duas etapas: 1) Detecção de rostos e 2) Detecção das expressões nos rostos individualmente. Para o primeiro passo, os autores aplicaram o MTCNN (ZHANG, K. *et al.*, 2016) para detectar rostos a partir das imagens de entrada. O algoritmo MTCNN é uma rede CNN em cascata de três estágios para detecção conjunta de rostos e localização de marcos faciais. Para o segundo passo, duas Redes diferentes foram aplicadas: *Small Face Network* (SFN) e *Big Face Network* (BFN), dependendo do tamanho dos rostos individuais detectados no passo 1. Ambas SFN e BFN foram baseadas em Redes Residuais (HE *et al.*, 2016). Por fim, a previsão em nível de grupo foi calculada com base nas previsões de cada rosto, de acordo com seu tamanho original, aliviando o efeito de previsões não confiáveis de rostos menores no fundo. Assim como a abordagem anteriormente apresentada, Khan *et al.* (2018) projetou o algoritmo para classificar as imagens nas categorias Positiva, Negativa e Neutra. O conjunto de dados usado para treinar e avaliar o modelo foi obtido do Google Imagens e Flickr com base em pesquisas por palavras-chave.

Esses estudos destacam a eficácia das CNNs no reconhecimento de emoções, um campo desafiador que requer a análise precisa de características visuais complexas. A obtenção de resultados notáveis em um desafio reconhecido internacionalmente como o EmotiW 2018 reforça a relevância das CNNs na pesquisa contemporânea em visão computacional e processamento de imagens.

Além das CNNs, os algoritmos LSTM foram amplamente aplicados nos artigos considerados nesta revisão. Os dois trabalhos que alcançaram os melhores resultados entre aqueles que aplicaram algoritmos LSTM foram Mou, Gunes e Patras (2019) e Guo *et al.* (2018), com 94% e 78.98% de acurácia, respectivamente.

A abordagem proposta por (GUO *et al.*, 2018) já foi explorada acima (o autor aplicou a combinação de CNN com LSTM). Por outro lado, (MOU; GUNES; PATRAS,

2019) apresenta uma abordagem para detecção facial e reconhecimento de emoções que contém componentes LSTM, juntamente com outras técnicas. Inicialmente, os autores adotaram um método multimodal baseado em SVM utilizando características dinâmicas e conduziram experimentos tanto em vídeos individuais quanto em grupo.

O algoritmo foi treinado com imagens individuais e em grupo, utilizando os conjuntos de dados *individualDB* e *groupDB* (MIRANDA-CORREA *et al.*, 2018). Em comparação com outros algoritmos, os autores descobriram que modelos de aprendizado temporal são capazes de superar modelos não temporais em termos de reconhecimento de emoções, tornando o LSTM uma boa técnica para ser aplicada a este tipo de problema. Diferentemente dos outros artigos explorados anteriormente, Mou, Gunes e Patras (2019) optou por usar informações do corpo dos participantes, bem como do rosto, para alcançar os melhores resultados, principalmente para identificar a presença de um grupo de pessoas ou de uma única pessoa na imagem (contexto da imagem).

Esses resultados destacam a eficácia dos algoritmos LSTM em contextos onde a capacidade de retenção de informação de longo prazo é crucial. A habilidade desses algoritmos em evitar o problema do gradiente desaparecendo os torna particularmente adequados para tarefas complexas de sequência temporal, como as encontradas em muitos dos estudos desta revisão.

Apesar de CNN e LSTM terem sido os algoritmos mais aplicados dentre os trabalhos desta pesquisa, o método proposto por Palestra e Pino (2020) é uma abordagem única que implementa um sistema usando características geométricas e um classificador *Random Forest* RF. Este método alcançou uma acurácia de 94,24%, um feito notável apesar do algoritmo tendo sido treinado e testado usando apenas um conjunto de dados, o *Extended Cohn-Kanade* (CK+) (LUCEY *et al.*, 2010a).

Os autores acreditam que a alta acurácia alcançada em seu estudo indica um reconhecimento confiável de expressões faciais, o que pode fornecer informações valiosas para apoiar interações Homem-Robô. A análise revelou que o sistema é capaz de reconhecer expressões faciais em sessões de terapia em grupo assistidas por robôs, mesmo com faces parcialmente ocultas. O método proposto atua como uma ferramenta mediadora e parece promover o engajamento dos participantes. Os resultados da classificação podem oferecer informações robustas ao sistema para basear suas intervenções em uma sessão de terapia em grupo, aprimorando o potencial para o uso de robôs em ambientes de saúde e reabilitação.

Apesar das limitações do conjunto de dados utilizado, os autores demonstraram o potencial de sua abordagem na decodificação automática de expressões faciais e forneceram resultados promissores. Pesquisas futuras poderiam explorar a aplicação do algoritmo desenvolvido por este estudo usando outros conjuntos de dados para avaliar a generalização do método proposto para diferentes populações e contextos mais amplos.

A tabela 2 apresenta todos os artigos analisados, acompanhados da acurácia alcançada e o tipo de algoritmo empregado na resolução do problema.

Trabalho	**Algoritmo**	**Acurácia**
(QUACH *et al.*, 2022)	CNN, NVPF, TNVPF	76,12%
(LIU, C. *et al.*, 2020)	CNN, SVM, LSTM, TSM	76,85%
(PALESTRA; PINO, 2020)	GF, RF	94,24%
(PETROVA; VAUFREYDAZ; DESSUS, 2020)	CNN	59,13%
(SRIVASTAVA *et al.*, 2020)	CNN	35,0%
(SUN, M. *et al.*, 2020)	CNN, TSM	71,93%
(MOU; GUNES; PATRAS, 2019)	GF, QLZM, SVM, SVR, LSTM	94,0%
(SHARMA, A.; MANSOTRA, 2019)	Viola-Jones, CNN, RNN, LSTM, SVM	75,0%
(YU, D. *et al.*, 2019)	MFM, LSTM Profunda Bidirecional	78,0%
(WANG, K. *et al.*, 2018)	CNN	67,48%
(GUO *et al.*, 2018)	CNN, LSTM	78,98%
(GUPTA *et al.*, 2018)	CNN, AM, MTCNN	64,83%
(KHAN *et al.*, 2018)	CNN, SFN, BFN	78,39%

Tabela 2 – Algoritmos aplicados e acurácia alcançada para cada artigo analisado. A coluna referente a acurácia, representa o melhor resultado alcançado pelo trabalho, visto que alguns autores apresentam diferentes algoritmos e datasets em seus artigos.

Fonte: Elaborado pelo autor, 2024.

3 DESENVOLVIMENTO

Neste capítulo, são detalhadas todas as etapas de desenvolvimento, linguagens utilizadas e ferramentas empregadas na construção do algoritmo de reconhecimento de expressões faciais em grupo. Como demonstrado na figura 1, o fluxo de processamento das imagens necessita de duas etapas principais: (i) detecção das faces na imagem e (ii) reconhecimento da expressão facial para cada uma das faces identificadas. Desta forma, este capítulo abordará o processo aplicado na implementação destes dois algoritmos nas sessões a seguir.

3.1 DETECÇÃO DE FACES

Historicamente, a detecção de face começou com métodos simples baseados em características geométricas e rapidamente evoluiu para técnicas mais sofisticadas, incluindo as baseadas em aprendizado de máquina e, mais recentemente, aprendizado profundo. A literatura está repleta de várias abordagens, cada uma buscando melhorar a precisão, a velocidade e a robustez dos sistemas de detecção de face, especialmente em condições desafiadoras que incluem variações de iluminação, pose e expressão.

Os estudos e publicações neste domínio não apenas descrevem os algoritmos e suas melhorias incrementais mas também discutem extensivamente as métricas de avaliação, os conjuntos de dados padrão e as técnicas de validação. Isso demonstra uma busca contínua por padrões de referência, permitindo comparações significativas entre diferentes métodos e, consequentemente, impulsionando o avanço da área.

Portanto, dado o estado avançado da tecnologia de detecção de face, os recursos significativos necessários para desenvolver um algoritmo eficaz e a disponibilidade de soluções robustas, optou-se por adotar uma abordagem existente. Esta abordagem será explorada na próxima seção.

3.1.1 Multi-task Cascaded Convolutional Networks

O *Multi-task Cascaded Convolutional Networks* (MTCNN) é um framework popular para detecção de faces que se destaca pela sua eficiência em várias escalas, o que o torna adequado para imagens de grupo onde os tamanhos das faces podem variar consideravelmente. Quando implementado com MXNet, uma biblioteca de aprendizado de máquina eficiente e escalável, o MTCNN aproveita a computação de alto desempenho e a otimização para processar imagens rapidamente, mantendo a precisão.

O MTCNN executa três tarefas principais simultaneamente: detecção de face, localização de pontos de referência (como olhos, nariz e boca) e calibração de faces. Essas tarefas são realizadas por três redes em cascata: P-Net, R-Net e O-Net, cada

uma focada em diferentes escalas e aspectos da detecção.

O algoritmo desenvolvido por Kaipeng Zhang *et al.* (2016) utiliza a biblioteca MX-Net para a aplicação de uma MTCNN capaz de detectar faces em grupos de pessoas. Este algoritmo encontra-se disponível no github [1].

Este modelo foi escolhido para aplicação neste trabalho por dois motivos: (i) precisão na detecção de faces e (ii) tempo de execução.

O modelo foi comparado com o algoritmo de Haar-Cascade (apresentado na seção 2.3.2), um método mais antigo e amplamente conhecido para detecção de faces. Enquanto o Haar Cascade é eficaz para algumas aplicações e mais fácil de implementar, ele geralmente não se compara ao MTCNN em termos de precisão e robustez, especialmente em ambientes com variação de iluminação, pose e expressão facial.

Além disso, o MTCNN oferece melhores resultados na localização das faces, o que é crucial para aplicações subsequentes, como a análise de expressões, no escopo deste trabalho. Na Figura 6 pode-se avaliar a precisão do algoritmo MTCNN em comparação com o Haar-Cascade.

Outra vantagem significativa do MTCNN sobre o Haar Cascade é sua velocidade de execução quando implementado com *frameworks* otimizados como MXNet. Na figura 6, o algoritmo Haar-Cascade é capaz de produzir um resultado em 89 milissegundos, em média, enquanto o MTCNN leva apenas 82 milissegundos, em média[2]. Isso permite a detecção em tempo real, mesmo em dispositivos com recursos limitados, como *smartphones* e câmeras de segurança. A combinação de eficiência, precisão e velocidade torna o MTCNN uma solução robusta e confiável para detecção e análise de faces em uma variedade de aplicações.

Desta forma, a escolha do MTCNN para este trabalho se justifica pela sua superioridade em termos de precisão de detecção, capacidade de lidar com diferentes escalas e condições de imagem, e eficiência em termos de tempo de execução.

3.2 RECONHECIMENTO DE EXPRESSÕES

Conforme mencionado na seção 2.4, o reconhecimento das expressões faciais é comumente identificado como a etapa mais relevante de todo o fluxo exposto na figura 1 e aplicado neste trabalho. Para obter um resultado satisfatório nesta etapa, dois itens são particularmente importantes: (i) estrutura do modelo e (ii) dataset de treinamento. Desta forma, nas seções seguintes serão apresentados o modelo escolhido para aplicação neste trabalho, bem como o dataset utilizado para treinamento.

[1] https://github.com/YYuanAnyVision/mxnet_mtcnn_face_detection

[2] todos os experimentos descritos neste trabalho foram executados em um computador com processador Ryzen 9 7950x e placa de vídeo RTX 4070ti

Figura 6 – Comparação de precisão entre os algoritmos MTCNN (a) e Haar-Cascade (b).

a)

b)

Fonte: Elaborado pelo autor, 2024

3.2.1 Modelo de classificação de expressões faciais

Para o reconhecimento de expressões faciais, foi escolhido um modelo baseado em Redes Neurais Convolucionais (CNNs), devido à sua capacidade de captar características hierárquicas e detalhes das imagens faciais, que são cruciais para a identificação precisa das expressões. As CNNs têm sido amplamente utilizadas e estudadas no campo da visão computacional, oferecendo um equilíbrio entre precisão e eficiência computacional, o que as torna ideais para a aplicação em tempo real e em

dispositivos com recursos limitados.

A estrutura específica do modelo adotado foi baseada em camadas convolucionais para extração de características, seguidas por camadas de pooling para redução de dimensionalidade e camadas densas para a classificação final das expressões. O modelo também incorporou técnicas modernas de normalização e ativação para melhorar o treinamento e a precisão do modelo. Mais especificamente foi aplicado o conceito de *transfer learning* sobre uma rede MobileNetV2 (SANDLER *et al.*, 2018).

Transfer learning ou aprendizado de transferência é uma técnica que envolve o uso de um modelo pré-treinado em um grande conjunto de dados, geralmente em uma tarefa genérica de visão computacional, como reconhecimento de imagem em grande escala, e então ajustá-lo para uma tarefa específica - neste caso, reconhecimento de expressões faciais. A escolha da MobileNetV2 como base se deu por suas características de eficiência, sendo uma arquitetura otimizada para dispositivos móveis e com recursos limitados, mas ainda assim poderosa o suficiente para tarefas complexas de visão computacional.

O modelo MobileNetV2 é conhecido por suas camadas convolucionais profundas e eficientes, que utilizam uma estrutura chamada de *inverted residuals* e *linear bottlenecks*. Essas estruturas permitem que o modelo seja, ao mesmo tempo, leve e preciso. O processo de transferência de aprendizado começou com a inicialização do modelo com os pesos pré-treinados do conjunto de dados ImageNet (DENG, J. *et al.*, 2009). Em seguida, as últimas camadas do modelo foram adaptadas e treinadas com o dataset específico de expressões faciais FER (GOODFELLOW *et al.*, 2013), já explorado na seção 2.4.1.3, permitindo que o modelo aprendesse as nuances e características específicas relevantes para o processo de reconhecimento de expressões faciais.

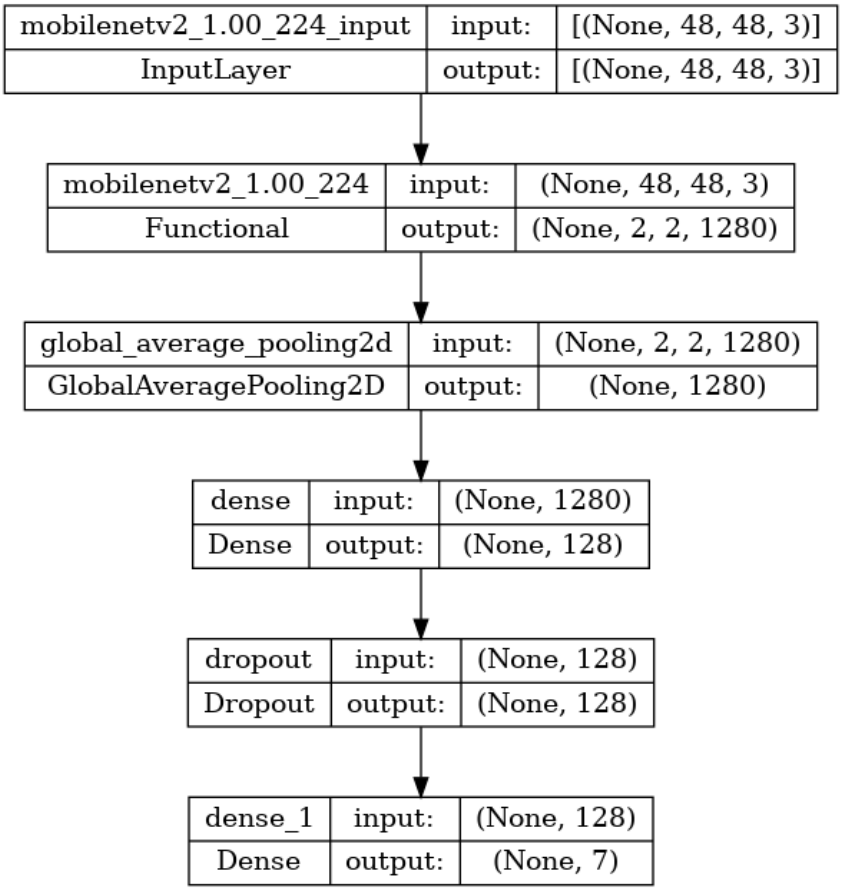
A partir do modelo MobileNetV2, foram adicionadas duas camadas densas de 128 e 64 neurônios respectivamente, juntamente com uma camada de *Global Average Pooling* e uma camada de *Dropout*, confirme demonstrado na figura 7 que representa a estrutura final do modelo.

A camada de *Global Average Pooling* foi utilizada como uma estratégia para reduzir a dimensionalidade das características extraídas pela CNN, condensando cada mapa de características em um único valor médio. Essa técnica não apenas reduz a quantidade de parâmetros e a complexidade computacional do modelo, mas também contribui para diminuir o risco de *overfitting*[3], tornando o modelo mais generalizável e robusto a variações nas imagens de entrada.

A camada de *Dropout*, por sua vez, foi incluída como uma forma de regularização. Durante o treinamento, uma fração dos neurônios é aleatoriamente desativada, o

[3] Quando o modelo se adapta aos dados de treinamento e atinge altos valores de acurácia, mas ao ser aplicado a dados fora deste conjunto, desempenha de forma diferente, ou seja, o modelo se torna ótimo para classificar as imagens do conjunto de treinamento somente.

Figura 7 – Estrutura final do modelo de classificação desenvolvido neste trabalho. Inicialmente, a estrutura do modelo MobileNetV2, sobre o qual foi aplicado o *transfer learning* e por fim, as camadas adicionais inseridas para classificação no contexto de expressões faciais.



Fonte: Elaborado pelo autor, 2024

que força o modelo a aprender representações mais robustas e menos dependentes de qualquer característica específica. Isso ajuda a melhorar a capacidade do modelo de generalizar para novos dados, evitando que ele se ajuste demais às particularidades do conjunto de treinamento, reduzindo ainda mais o risco de *overfitting*.

3.2.2 Dataset de treinamento

Com o modelo definido, a próxima etapa crucial foi a seleção e preparação do dataset de treinamento. Para este fim, foi escolhido o FER-2013 (GOODFELLOW *et al.*, 2013), conforme apresentado na seção 2.4.1.3, um dos datasets mais conhecidos e utilizados na comunidade científica para o reconhecimento de expressões faciais. Este dataset é composto por milhares de imagens de faces humanas, cada uma categorizada em uma das sete expressões básicas. As imagens foram coletadas de maneira a representar uma variedade de demografias, iluminações e condições de fundo, proporcionando uma rica variedade de dados para treinar o modelo.

Todo o processo de treinamento foi executado para que o modelo pudesse ser capaz de classificar as 6 expressões básicas propostas por Ekman e Friesen (1971): raiva, nojo, medo, felicidade, tristeza e surpresa, além da expressão neutra, totalizando 7 expressões. Utilizou-se uma combinação de técnicas para otimização e ajuste fino, como a variação da taxa de aprendizagem e o uso de validação cruzada, para monitorar e melhorar o desempenho do modelo. O objetivo foi alcançar não apenas uma alta precisão geral, mas também garantir que o modelo fosse equilibrado e performático em todas as categorias de expressão, considerando as variações e nuances individuais que cada uma apresenta.

Para enriquecer o conjunto de dados e aumentar a robustez do modelo, foram aplicadas técnicas de aumento de dados, que incluem rotação, inversão, zoom e variações na iluminação das imagens. Essas técnicas ajudaram a simular uma gama mais ampla de condições que o modelo poderia encontrar na prática, preparando-o para reconhecer expressões faciais em diferentes contextos e ambientes.

3.2.3 Treinamento do modelo

O treinamento do modelo foi uma etapa crucial para alcançar o objetivo final de reconhecer expressões faciais com precisão e eficiência. A fase de treinamento começou com a inicialização do modelo utilizando os pesos da MobileNetV2 pré-treinada no ImageNet, seguida pela adaptação das camadas finais conforme demonstrado na figura 7. Este conjunto específico de dados contém uma variedade de expressões faciais obtidas da internet e capturadas em diferentes condições, representando um desafio realista para o sistema.

Durante o treinamento, foram analisadas algumas métricas para garantir a convergência eficiente e eficaz do modelo, como a taxa de aprendizagem, por exemplo. A taxa de aprendizagem foi cuidadosamente ajustada para equilibrar a rapidez da convergência com a precisão do aprendizado. Técnicas como a redução da taxa de aprendizagem em platôs e o uso de momentos ou otimizadores adaptativos. Os otimizadores adaptativos são uma classe de algoritmos de otimização usados no treinamento de modelos de aprendizado de máquina, especialmente em redes neurais profundas. Eles ajustam a taxa de aprendizagem durante o treinamento de maneira adaptativa por parâmetro, com o objetivo de melhorar a convergência e o desempenho do modelo. Diferentemente de métodos de otimização tradicionais, que mantêm uma taxa de aprendizagem constante ou diminuem uniformemente ao longo do tempo para todos os parâmetros, os otimizadores adaptativos ajustam a taxa de aprendizagem para cada parâmetro individualmente, baseando-se em estimativas de momentos de primeira e/ou segunda ordem do gradiente. Isso permite que eles sejam mais eficientes em termos de convergência, especialmente em cenários com gradientes esparsos ou com parâmetros que precisam de atualizações em magnitudes variadas.

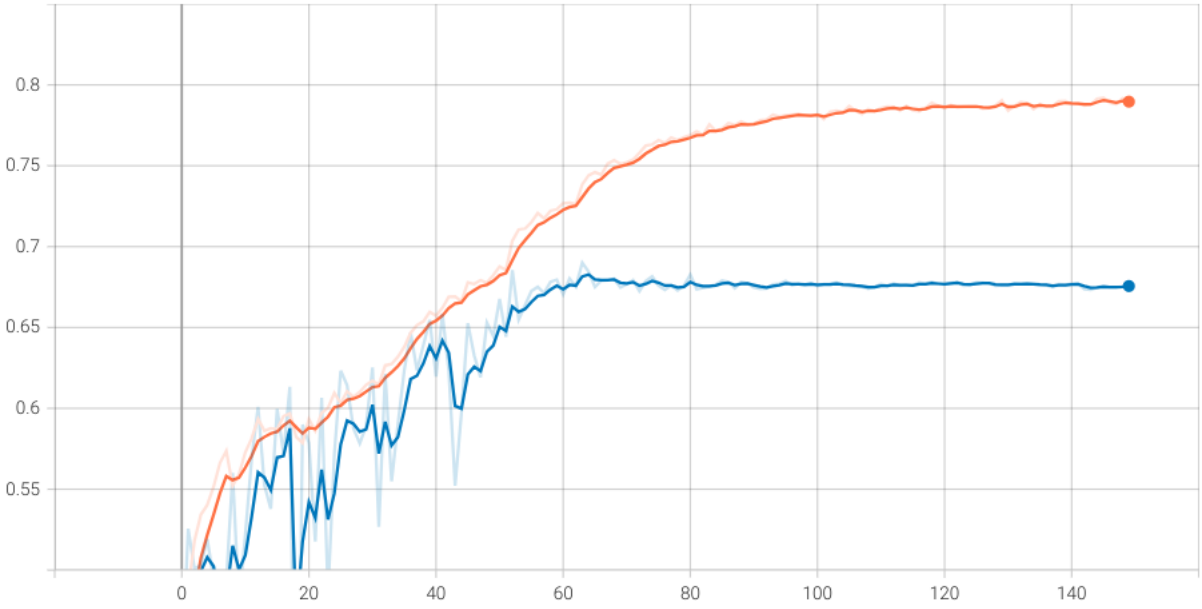
O modelo foi treinado por um período de 150 épocas, cada uma destas envolvendo a passagem de todo o conjunto de dados de treinamento através do modelo e a atualização dos pesos do modelo com base nos erros calculados.

Para garantir que o modelo não apenas aprendesse os dados de treinamento, mas também generalizasse bem para novos dados, o conjunto de dados foi dividido em três partes: treinamento, validação e teste. A divisão típica seguiu uma proporção onde a maior parte dos dados (70%) foi usada para treinamento, uma porção menor para validação durante o treinamento para ajustar hiperparâmetros e evitar *overfitting* (15%), e uma parte final reservada para testar o desempenho do modelo após o treinamento ter sido concluído (15%).

Durante cada época de treinamento, o desempenho do modelo foi avaliado usando o conjunto de validação. Esta avaliação contínua permitiu ajustes em tempo real no processo de treinamento, como atualizações na taxa de aprendizagem e parada antecipada no caso de demonstração de *overfitting*.

O gráfico que apresenta a acurácia do modelo ao longo do processo de treinamento é apresentado na figura 8

Figura 8 – Acurácia do modelo ao longo do treinamento. O a linha laranja representa o resultado no conjunto de treinamento enquanto a linha azul apresenta o desempenho do modelo quando avaliado no conjunto de validação.



Fonte: Elaborado pelo autor, 2024

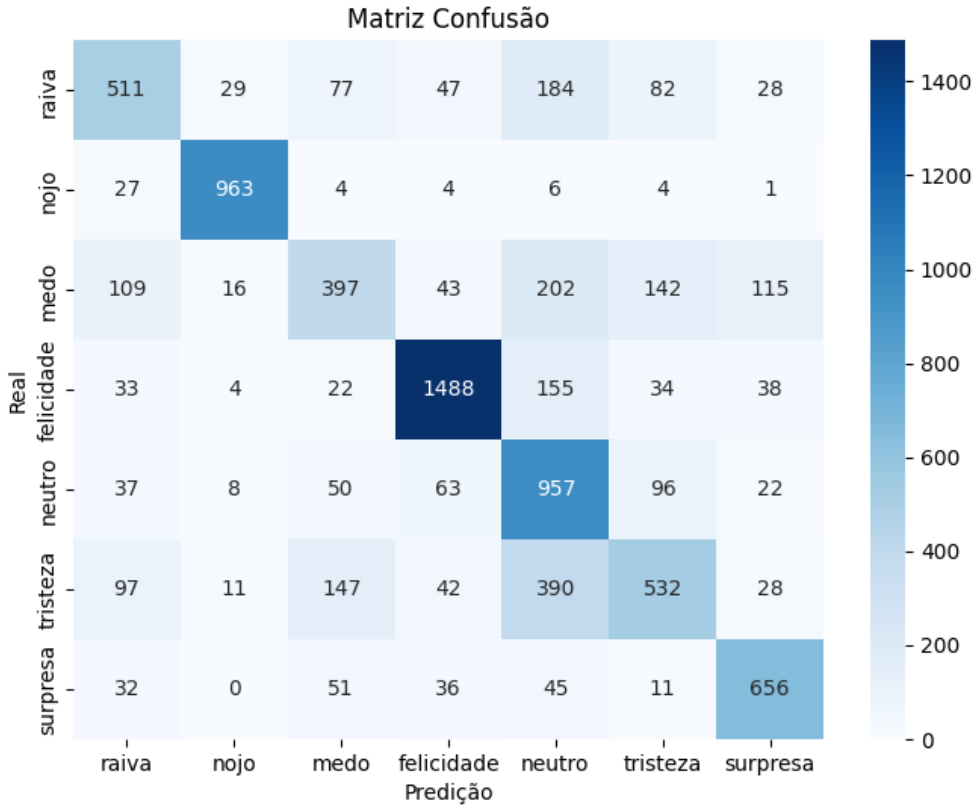
O processo de treinamento foi executado em um computador com processador Ryzen 9 7950x, 32 GB de memória RAM e uma placa gráfica RTX 4070ti e totalizou em torno de 35 minutos de execução. No gráfico, podemos perceber que o conjunto de validação obteve uma acurácia em torno de 67% na ultima época do treinamento,

contudo, o melhor desempenho do modelo em termos de acurácia no conjunto de validação ocorreu na época 63, onde o algoritmo foi capaz de atingir 69,01%. Depois deste período, o iniciou o processo de *overfitting*, contudo, os pesos selecionados para a aplicação do modelo foram os da época 63.

3.3 AVALIAÇÃO TÉCNICA DOS RESULTADOS DO ALGORITMO

A avaliação técnica dos resultados após o treinamento do modelo de reconhecimento de expressões faciais envolveu uma análise detalhada do desempenho do algoritmo em diferentes métricas e cenários. O foco principal foi em avaliar a precisão, a *recall*, a *F1-score* e a matriz de confusão. Estas métricas fornecem uma visão abrangente da performance do modelo, identificando não apenas sua acurácia global, mas também como ele performa especificamente em termos de falsos positivos e falsos negativos.

Figura 9 – Matriz de confusão representativa dos resultados do modelo.



Fonte: Elaborado pelo autor, 2024

A precisão do modelo é crucial, pois indica a proporção de identificações corretas de expressões faciais entre todas as identificações que o modelo fez. Já a *recall*

é importante para entender quantas das expressões faciais reais foram corretamente identificadas pelo modelo. A *F1-score* combina essas duas métricas em uma média harmônica, proporcionando um único score que balanceia tanto a precisão quanto a *recall*. Essas métricas são particularmente úteis quando se lida com um conjunto de dados desequilibrado, que é comum em aplicações do mundo real.

A matriz de confusão (figura 9) oferece uma visão detalhada de como o modelo está confundindo diferentes classes de expressões. Ela é uma tabela que mostra as frequências de classificação para cada classe de expressão facial, comparando os rótulos previstos pelo modelo com os rótulos verdadeiros do conjunto de teste. Isso ajuda a identificar quais expressões são frequentemente confundidas umas com as outras, permitindo ajustes direcionados na fase de pós-processamento ou re-treinamento do modelo.

Por fim, quando avaliado no conjunto de teste, o modelo obteve uma acurácia de 68.57%. O valores de precisão, *recall* e *F1-score* são apresentados na figura 10 para cada uma das expressões. É importante ressaltar que o modelo foi avaliado em termos de precisão e acurácia. Estas métricas, embora pareçam semelhantes, não tem o mesmo significado. A acurácia global reflete a proporção de previsões corretas (tanto verdadeiros positivos quanto verdadeiros negativos) em relação ao total de previsões feitas, enquanto a precisão é uma medida específica que indica a exatidão das previsões positivas do modelo. Desta forma, em média, o modelo obteve uma precisão de 68,71%, recall de 67,29% e F1-score de 67,99%.

3.4 COMPARATIVO COM TRABALHOS RELACIONADOS

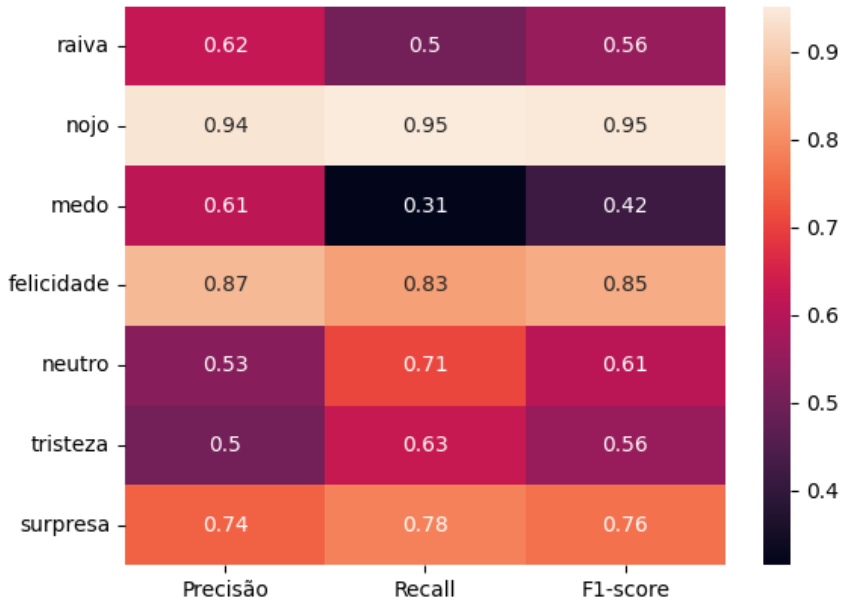
A avaliação técnica do modelo de reconhecimento de expressões faciais foi complementada com um comparativo detalhado com trabalhos relacionados na área. Esta comparação teve como objetivo entender o desempenho do modelo desenvolvido em relação às abordagens existentes, identificar as melhores práticas e tecnologias empregadas e reconhecer áreas de melhoria e oportunidades para futuras pesquisas.

Diversos trabalhos têm abordado o reconhecimento de expressões faciais, aplicando uma variedade de técnicas e metodologias. A análise desses trabalhos foi realizada com foco em dois aspectos principais: a acurácia dos modelos propostos e as peculiaridades das técnicas utilizadas, incluindo os tipos de algoritmos, as características dos conjuntos de dados e as estratégias de treinamento e validação.

A comparação começou com a revisão dos algoritmos mais comuns e bem-sucedidos encontrados na literatura, principalmente CNNs e LSTMs, como mencionado anteriormente. Esses dois tipos de redes são predominantes no campo devido à sua eficácia em processar dados visuais e sequenciais, respectivamente.

O estudo também considerou a diversidade dos conjuntos de dados utilizados pelos diferentes trabalhos, uma vez que a escolha do conjunto de dados pode ter um

Figura 10 – Valores de precisão, *recall* e *F1-score* para cada uma das classes envolvidas no treinamento do modelo.



Fonte: Elaborado pelo autor, 2024

impacto significativo no desempenho do modelo. Foi dada atenção especial àqueles que utilizaram o mesmo dataset utilizado para o treinamento do modelo (FER-2013).

Dentro os trabalhos relacionados avaliados nessa pesquisa, apenas Guo *et al.* (2018) utilizou o dataset em questão no seu treinamento. Contudo além do dataset FER-2013, os autores utilizaram também outro dataset (GENKI-4k) para o treinamento de seu algoritmo. A acurácia alcançada pelo modelo de Guo *et al.* (2018) foi de 78,98%, conforme disposto na tabela 2.

Como o escopo do algoritmo de classificação deste trabalho é uma única face, é valida a comparação com outros trabalhos de reconhecimento de expressões faciais, não necessariamente aplicados a grupos. Desta forma, foi feita uma nova pesquisa de artigos proponentes de algoritmos similares e que utilizaram o dataset FER-2013.

A análise comparativa dos trabalhos relacionados que utilizaram o FER-2013 como dataset para treinamento dos seus modelos revelou uma gama de resultados em termos de acurácia. Estes resultados variam significativamente, refletindo as diferentes abordagens, arquiteturas de rede e técnicas de otimização utilizadas pelos pesquisadores. A tabela 3 apresenta uma síntese dos resultados obtidos, oferecendo uma visão comparativa direta entre o modelo desenvolvido neste trabalho e as outras abordagens existentes.

O trabalho de Guo *et al.* (2018) destaca-se com a maior acurácia, alcançando 78,98%. Esta alta performance pode ser atribuída à combinação de datasets durante o

Trabalho	**Acurácia**
(GUO *et al.*, 2018)	78,98%
(WANG, XuMing *et al.*, 2018)	68,79%
(SALUNKE; PATIL, 2017)	68,0%
(GAN, 2018)	64,46%
(YU, Z.; ZHANG, C., 2015)	61,29%
(NG *et al.*, 2015)	50,50%

Tabela 3 – Trabalhos que utilizaram FER-2013 como dataset para treinamento dos seus modelos e os respectivos resultados obtidos em termos de acurácia.

Fonte: Elaborado pelo autor, 2024.

treinamento, indicando uma robustez significativa e adaptabilidade do modelo. No entanto, é importante considerar que a utilização de datasets adicionais e a combinação de múltiplos modelos podem aumentar a complexidade e os requisitos computacionais do sistema, fatores que devem ser levados em conta ao avaliar a aplicabilidade prática da solução.

Os outros trabalhos listados na tabela variam em acurácia de 50,50% a 68,79%, com diversas abordagens tentando otimizar a precisão na classificação das expressões faciais. As variações nos resultados destacam a natureza desafiadora do reconhecimento de expressões faciais, especialmente ao se utilizar um dataset complexo e diversificado como o FER-2013. Cada abordagem tem seus próprios méritos e limitações, e a escolha de uma sobre a outra pode depender de vários fatores, incluindo a complexidade do modelo, o tempo de treinamento, a capacidade computacional disponível e o contexto específico de aplicação.

A comparação direta entre o modelo desenvolvido neste trabalho e os trabalhos relacionados é um passo crucial para a avaliação técnica. Ela não apenas posiciona o modelo dentro do espectro de soluções existentes, mas também fornece valiosas informações sobre as tendências da pesquisa e as possíveis direções para futuros aprimoramentos.

Em suma, o modelo treinado neste trabalho obteve um resultado satisfatório, superando a maioria dos trabalhos relacionados sobre o mesmo dataset, mesmo utilizando um modelo próprio para execução em dispositivos móveis, com baixo tempo de execução (o tempo de execução será explorado nas próximas seções).

4 APLICAÇÃO DO SOFTWARE

Neste capítulo é apresentada a aplicação da ferramenta desenvolvida com um grupo de alunos do PPGTIC da Universidade Federal de Santa Catarina. Esta aplicação proporcionou dados para avaliação do algoritmo desenvolvido em uma situação real e este processo será descrito e explorado neste capítulo.

4.1 EXPERIMENTO

A avaliação do algoritmo de forma técnica foi descrita na seção 3.3, contudo, uma avaliação do seu desempenho em uma situação real de aplicação é necessária para entender como o modelo se comporta em condições práticas e quais são os resultados tangíveis que ele oferece. A aplicação prática foi realizada com um grupo de alunos do curso de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina, fornecendo uma oportunidade única para observar e avaliar o modelo em um ambiente dinâmico e realista.

4.1.1 Configuração do Teste

A configuração do teste envolveu a utilização da ferramenta desenvolvida em uma sessões com os alunos. Esta sessão foi cuidadosamente planejada para garantir um cenário realista. O espaço foi organizado para simular um ambiente típico de sala de aula, com iluminação adequada e posicionamento de câmera estratégico para capturar as expressões faciais dos participantes.

Antes de iniciar a sessão, foi explicado aos alunos o propósito do teste e como a ferramenta seria usada. Assegurou-se de obter o consentimento informado de todos os participantes, ressaltando a importância da privacidade e do uso ético dos dados coletados. Durante a sessão, os alunos foram expostos a diferentes estímulos e atividades projetadas para evocar uma variedade de expressões faciais, ao passo que, a ferramenta capturou expressões faciais dos alunos em tempo real. Os dados coletados foram armazenados de forma segura e organizada para facilitar a análise posterior.

Os estímulos mencionados se deram por meio de vídeos por algumas razões:

- **Contexto Natural:** Os vídeos representam situações reais em que as pessoas interagem com conteúdo audiovisual, tornando a avaliação mais próxima das experiências cotidianas.
- **Engajamento dos Participantes:** Os vídeos podem ser mais envolventes e cativantes para os participantes, estimulando respostas emocionais genuínas.
- **Desafio de Detecção Variada:** Diferentes vídeos apresentam desafios de detecção diversos, testando a capacidade do algoritmo de reconhecer uma

ampla gama de emoções.

Inicialmente, os alunos foram expostos a um vídeo de comédia sobre inteligência artificial, onde esperava-se detectar a expressão facial de felicidade. Na sequência, foi apresentado aos alunos um vídeo contendo acidentes entre carros autônomos e animais, para o qual era esperada a detecção de expressões de medo e surpresa. Por fim, foi apresentado aos alunos, um vídeo que mostrava um profissional da área médica removendo parasitas da perna de uma pessoa. Para este vídeo, era esperada a detecção das expressões nojo e surpresa.

Antes da apresentação dos dois últimos vídeos, os participantes do experimento foram alertados sobre a presença de conteúdo sensível e lhes foi dada a opção de não visualizar o material caso se sentissem desconfortáveis. Foi reforçada a importância do consentimento e do conforto de todos os participantes, assegurando um ambiente ético e respeitoso durante toda a execução do teste, contudo, nenhum participante optou por retirar-se do ambiente.

Após a execução de cada um dos vídeos, foi solicitado aos estudantes que preenchessem um breve formulário referente à experiencia durante a execução do conteúdo. Este questionário foi elaborado contendo 4 perguntas, foi disponibilizado de forma *online* para facilitação do acesso durante o teste e pode ser visualizado no Apêndice A.

O objetivo da aplicação deste formulário foi adquirir informação necessária para realizar uma comparação dos dados coletados automaticamente pela ferramenta de reconhecimento de expressões faciais com as percepções subjetivas dos alunos sobre suas próprias reações e sentimentos.

4.2 AVALIAÇÃO DOS RESULTADOS

A aplicação do software em um contexto real permitiu uma avaliação abrangente do algoritmo em termos de desempenho e assertividade. Durante as sessões, a ferramenta operou eficientemente, atingindo uma média de 12 *frames* por segundo (FPS) na detecção e classificação de expressões faciais. Esta taxa de processamento é indicativa de que o algoritmo possui a capacidade de ser executado em tempo real, um aspecto crítico para aplicações interativas e dinâmicas. A eficiência do algoritmo em operar em tempo real abre um leque de possibilidades para sua aplicação em diferentes cenários, desde salas de aula até ambientes de trabalho e sociais, onde a leitura instantânea das expressões faciais pode ser valiosa.

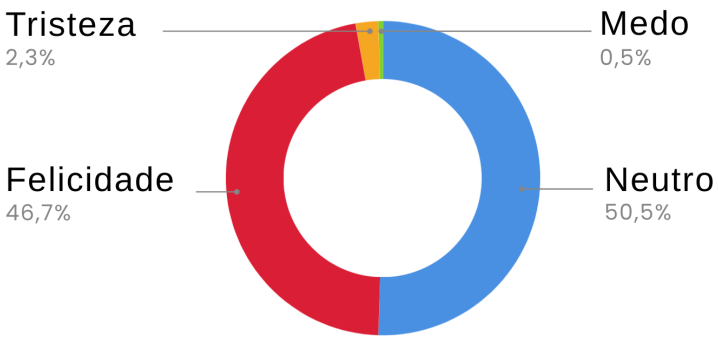
Para cada *frame* capturado, o algoritmo identificou as expressões faciais presentes individualmente em cada rosto detectado na cena. A partir dessas identificações, para determinar a expressão predominante de cada *frame*, o algoritmo selecionou a expressão que apresentou a maior ocorrência entre todos os rostos detectados. Essa abordagem permitiu que a ferramenta fornecesse uma leitura precisa e representativa

da expressão facial dominante em qualquer momento específico do vídeo, capturando assim a dinâmica das reações emocionais do grupo ao longo do tempo. Esse método de agregação das expressões mais frequentes em cada *frame* assegura que a expressão refletida é a mais representativa da reação coletiva dos participantes, permitindo uma análise robusta e detalhada das emoções em resposta aos estímulos apresentados.

4.2.1 Vídeo 1

No primeiro vídeo (comédia), a ferramenta detectou predominantemente as expressões neutra e de felicidade entre os alunos. A figura 11 ilustra a distribuição e a frequência das expressões faciais identificadas durante a exibição do vídeo. Esses resultados são promissores, pois indicam que o algoritmo foi capaz de capturar com precisão as reações emocionais dos participantes a conteúdos humorísticos. Essa capacidade de reconhecer e interpretar corretamente expressões de alegria e neutralidade é essencial para a aplicação da ferramenta em ambientes que visam medir a satisfação, o engajamento ou a reação positiva do público.

Figura 11 – Resultado da aplicação do modelo para o primeiro vídeo do experimento.



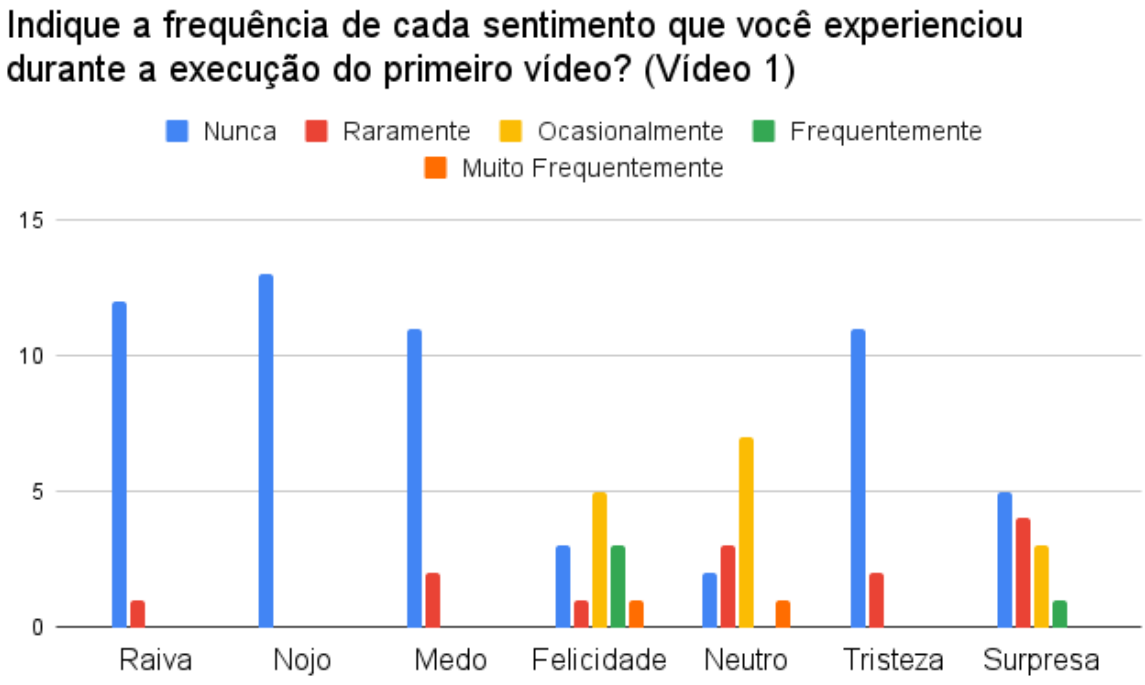
Fonte: Elaborado pelo autor, 2024

O formulário referente ao primeiro vídeo do experimento recebeu 13 respostas no total. Ao comparar os resultados da detecção com as respostas dadas pelos alunos no formulário (figura 12), pode-se perceber uma paridade nas informações. Essa concordância entre a percepção dos usuários e os dados coletados pela ferramenta reforça a validade do algoritmo e a sua aplicabilidade em situações reais.

Quanto à terceira pergunta do formulário (Você acredita ter expressado esse sentimento com expressões faciais durante a execução do vídeo?), foram como “sim” as seguintes quantidades para as respectivas expressões: 1 para tristeza, 2 para medo, 5 para surpresa, 8 para neutro e 11 para felicidade, conforme representado na figura 13.

Interessantemente, as menções a expressões de surpresa, medo e tristeza, embora em menor número, destacam a complexidade das respostas emocionais humanas

Figura 12 – Respostas à segunda pergunta do formulário do primeiro vídeo da aplicação, preenchido pelos alunos logo após a execução do primeiro vídeo.



Fonte: Elaborado pelo autor, 2024

e a diversidade de percepções individuais mesmo em um contexto grupal e unificado. Isso também pode refletir momentos específicos do vídeo que podem ter evocado reações mais variadas ou sutis, ressaltando a importância de considerar a amplitude das respostas emocionais ao avaliar a eficácia de algoritmos de reconhecimento facial.

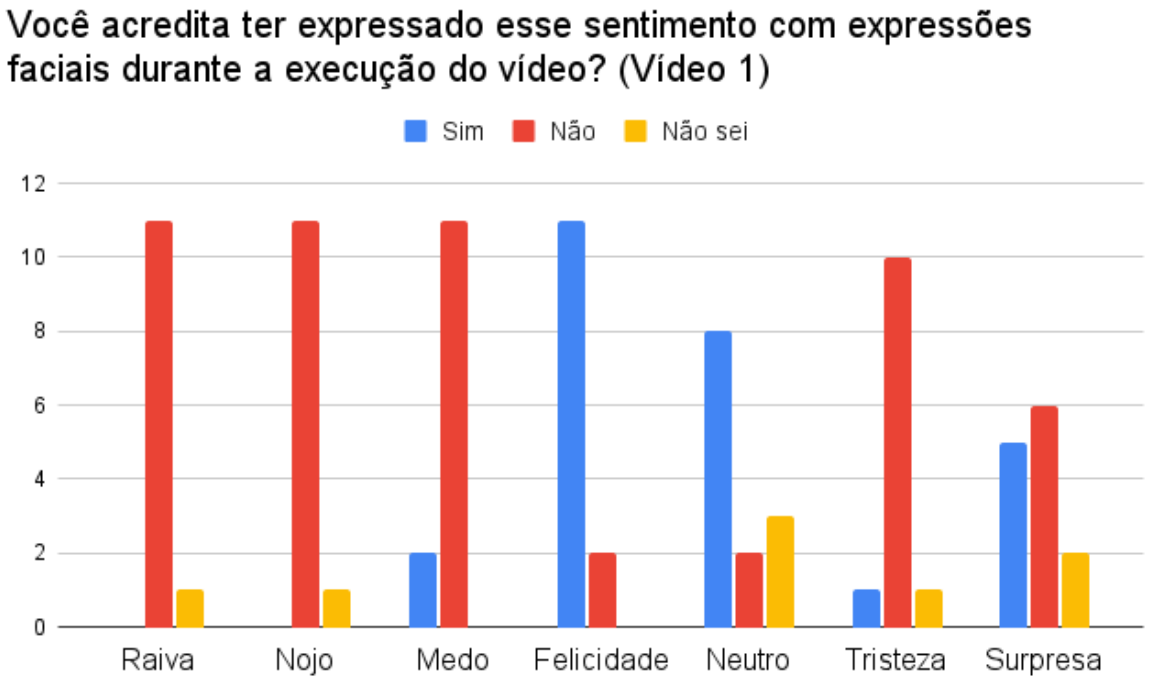
Quanto à pergunta final do formulário, apesar de não ser de resposta obrigatória, foram recebidas as seguintes quatro respostas:

- *“O vídeo era engraçado e despertou sentimentos de alegria”*
- *“Nada”*
- *“Como não é algo que interfere minha realidade profissional, não achei humorístico.”*
- *“O vídeo levou a reflexões bem humoradas.”*

4.2.2 Vídeo 2

Para o segundo vídeo, envolvendo acidentes entre carros autônomos e animais, o objetivo era identificar expressões de tristeza e surpresa. A ferramenta de reconhecimento facial foi capaz de captar uma quantidade significativa de expressões que correspondiam ao conteúdo mais intenso e chocante do vídeo. A figura 14 demonstra

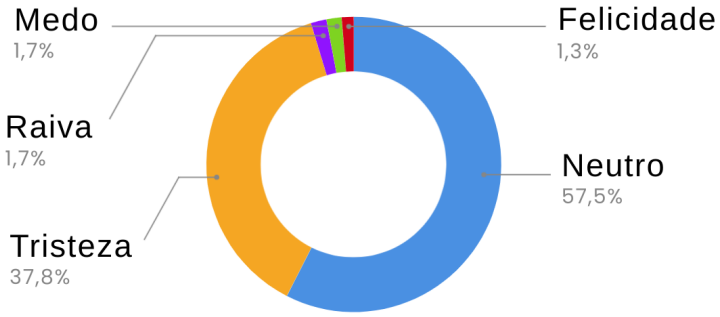
Figura 13 – Respostas à terceira pergunta do formulário do primeiro vídeo da aplicação, preenchido pelos alunos logo após a execução do primeiro vídeo.



Fonte: Elaborado pelo autor, 2024

a distribuição e frequência das expressões faciais identificadas durante a exibição do segundo vídeo. Os resultados mostraram um aumento nas expressões de tristeza e até raiva, refletindo a natureza mais grave e séria do vídeo em comparação com o primeiro.

Figura 14 – Resultado da aplicação do modelo para o segundo vídeo do experimento.

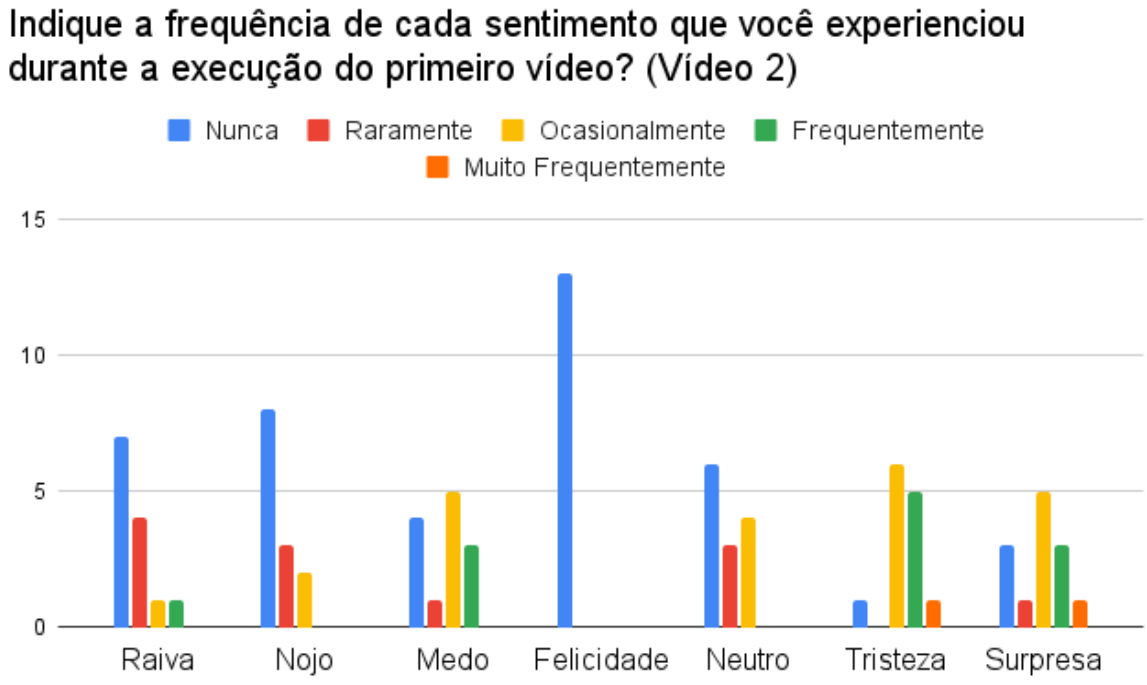


Fonte: Elaborado pelo autor, 2024

O formulário respondido pelos alunos após a visualização do segundo vídeo obteve 15 respostas e ofereceu uma visão complementar dos resultados capturados pela ferramenta. As respostas ao formulário indicaram que os alunos geralmente per-

ceberam uma correspondência entre as expressões detectadas pela ferramenta e suas experiências emocionais durante o vídeo. A figura 15 mostra um resumo das percepções dos alunos sobre a precisão da ferramenta e suas próprias reações emocionais ao conteúdo do segundo vídeo.

Figura 15 – Respostas ao formulário do segundo vídeo da aplicação, preenchido pelos alunos logo após a execução do vídeo.



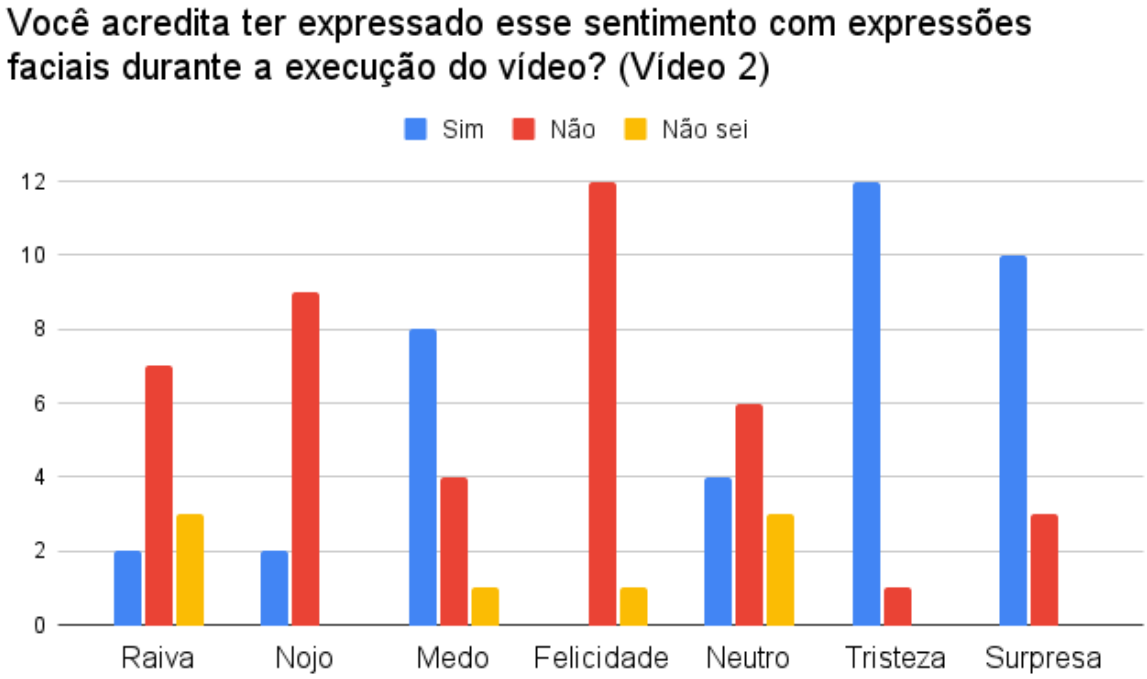
Fonte: Elaborado pelo autor, 2024

Nesta figura, é possível identificar que as expressões felicidade e nojo foram as menos relatadas pelos estudantes, respectivamente. Por outro lado, tristeza foi a expressão mais relatada como “Frequentemente” e, juntamente com surpresa, foi a expressão que recebeu mais respostas de “Muito frequentemente”. Ao passo que tristeza foi uma das expressões mais identificadas ao longo desta etapa do experimento, juntamente com neutro, a expressão de surpresa não foi identificada em nenhum *frame* do vídeo como a expressão dominante dentre todas as faces detectadas. O possível motivo para este efeito será explorado na seção 4.3

A análise das respostas à terceira pergunta do formulário revelou uma tendência dos alunos a reconhecerem em suas expressões faciais os sentimentos de tristeza, surpresa e medo evocados pelo vídeo. A concordância parcial entre as observações subjetivas dos alunos e as detecções objetivas da ferramenta reforça a precisão do algoritmo em capturar as nuances das expressões faciais em resposta a diferentes tipos de estímulo. O gráfico referente às respostas dos usuários à terceira pergunta do

formulário pode ser visualizada na figura 16

Figura 16 – Respostas à terceira pergunta do formulário do segundo vídeo da aplicação, preenchido pelos alunos logo após a execução do conteúdo.



Fonte: Elaborado pelo autor, 2024

As respostas fornecidas pelos alunos à pergunta final sobre suas impressões gerais do vídeo e quaisquer comentários adicionais refletiram a diversidade de experiências e reações emocionais dentro do grupo. Enquanto um aluno mencionou não ter sentido impacto algum pelo vídeo, sugerindo uma reação mais neutra ou desapegada, outro expressou uma decisão consciente de não se envolver muito com o conteúdo, indicativo de uma possível sensibilidade ou desconforto com o material apresentado. Estes comentários ilustram a complexidade e a variedade de reações emocionais que podem ocorrer em um grupo de indivíduos expostos ao mesmo estímulo, destacando os desafios enfrentados ao capturar e interpretar expressões faciais em ambientes diversificados. Estas respostas subjetivas foram muito importantes para entender a amplitude das experiências emocionais humanas e reforçaram a necessidade de considerar uma gama variada de reações ao desenvolver e aplicar algoritmos de reconhecimento facial.

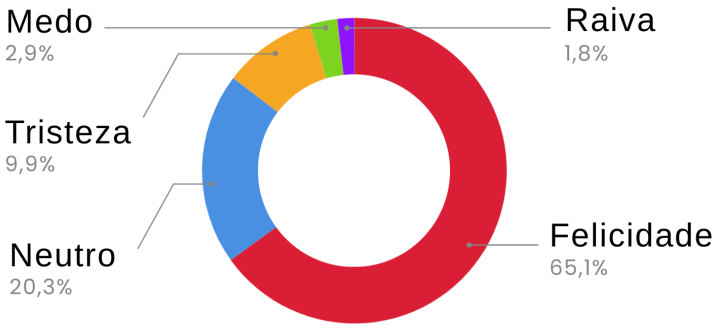
As respostas exatas dadas pelos participantes dos experimentos foram:

- *“Acredito que não senti impacto.”*
- *“Procurei não me envolver muito no vídeo.”*

4.2.3 Vídeo 3

Para o terceiro vídeo, exibindo um procedimento médico gráfico, esperava-se predominantemente detectar expressões de nojo e surpresa entre os participantes. Contudo, curiosamente, durante a visualização do vídeo, uma dinâmica inesperada ocorreu: ao invés de expressões de nojo predominarem, houve momentos de riso coletivo entre os alunos, um fenômeno social que alterou a atmosfera da sala e, consequentemente, as expressões faciais detectadas. A figura 17 ilustra as expressões identificadas durante a exibição do vídeo, mostrando uma predominância inesperada de expressões que não correspondiam inteiramente ao conteúdo perturbador do vídeo.

Figura 17 – Resultado da aplicação do modelo para o terceiro vídeo do experimento.

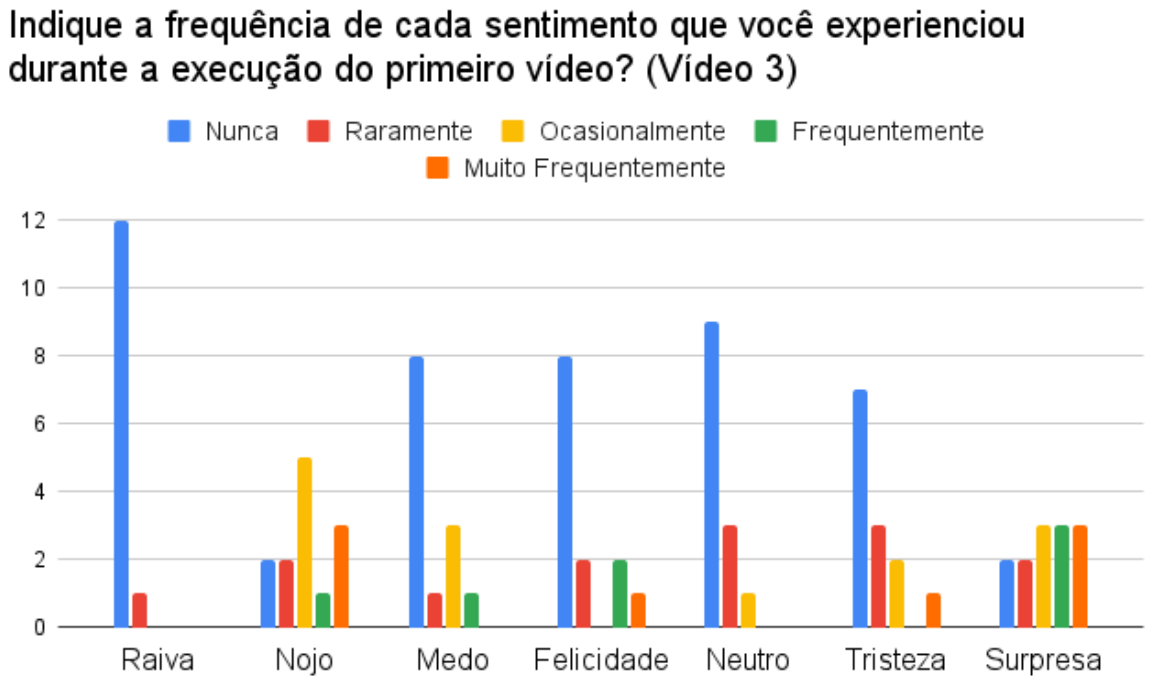


Fonte: Elaborado pelo autor, 2024

Interessantemente, o algoritmo não conseguiu detectar expressões de nojo, uma reação que seria intuitivamente esperada dada a natureza do vídeo. Isso pode ser atribuído a várias razões, incluindo a possível complexidade em identificar e diferenciar nojo de outras expressões faciais em um contexto grupal, ou limitações do modelo em capturar nuances específicas dessa expressão. Além disso, a atmosfera coletiva de riso pode ter influenciado as expressões faciais dos alunos de maneira a prevalecer sobre as reações individuais ao conteúdo, destacando como o comportamento grupal pode afetar as respostas emocionais em situações compartilhadas.

Um total de 15 alunos responderam ao formulário após a visualização do terceiro vídeo, fornecendo dados sobre suas reações pessoais e percepções durante o experimento. Conforme ilustrado na figura 18, enquanto a maioria dos alunos relataram sentir nojo e surpresa, a expressão de felicidade recebeu algumas respostas com para “Frequentemente” e “Muito Frequentemente”. Estas pessoas podem ter desencadeado a presença de risadas, influenciando os demais estudantes e alterando e a atmosfera geral do grupo, fazendo divergir a situação real da aplicação com a reação esperada ao vídeo. Este fenômeno, mais uma vez, destaca a complexidade das dinâmicas grupais e como elas podem influenciar a percepção e a expressão emocional em ambientes coletivos.

Figura 18 – Respostas ao formulário do terceiro vídeo da aplicação, preenchido pelos alunos logo após a execução do vídeo.



Fonte: Elaborado pelo autor, 2024

Esse episódio demonstra que, enquanto ferramentas de reconhecimento facial são poderosas na detecção de emoções, a interpretação desses dados em contextos sociais complexos pode requerer uma compreensão mais profunda das interações humanas e dos fatores ambientais. A influência do comportamento coletivo, a sensibilidade do algoritmo a expressões específicas e a necessidade de adaptar tecnologias de reconhecimento facial para refletir com precisão a riqueza das expressões humanas são aspectos importantes a serem considerados em futuros desenvolvimentos e aplicações. Este experimento com o terceiro vídeo ressalta a importância de continuar refinando a tecnologia de reconhecimento facial, considerando a variedade de contextos humanos e as sutilezas das interações grupais.

Por fim, para a última pergunta do formulário, foram recebidas duas respostas:

- *"Não assisti , pois não gosto deste tipo de vídeo."*
- *"Foi difícil acompanhar o vídeo."*

Estas respostas refletem não apenas as diferentes tolerâncias individuais a conteúdos gráficos e desagradáveis, mas também a importância do consentimento na realização do experimento. Os comentários indicam que, mesmo entre aqueles que escolheram assistir, o conteúdo, foram provocadas reações emocionais intensas e, possivelmente, desconfortáveis. Ambas as respostas sublinham a complexidade

emocional e psicológica envolvida no processamento de estímulos visuais intensos e reiteram a necessidade de sensibilidade ao conduzir experimentos que envolvem reações emocionais, especialmente em grupos.

4.3 CONSIDERAÇÕES SOBRE A APLICAÇÃO

A implementação prática do software em um ambiente real, particularmente com um grupo de alunos de Pós-Graduação em Tecnologias da Informação e Comunicação, proporcionou reflexões valiosas não apenas sobre a capacidade técnica do algoritmo, mas também sobre suas limitações e potencialidades em cenários dinâmicos e complexos que envolvem interações humanas reais. Esta seção aprofunda as observações e conclusões derivadas da aplicação prática.

4.3.1 Análise de Desempenho do Modelo em Condições Reais

A eficiência operacional do software, demonstrada pela capacidade de processar expressões faciais em tempo real, é um indicativo significativo de sua aplicabilidade em diversos cenários. Atingindo uma média de 12 *frames* por segundo na identificação e classificação de expressões, o modelo não apenas provou sua viabilidade técnica, mas também seu potencial para aplicações que exigem respostas imediatas e análises em tempo real, como em ambientes educacionais, de trabalho, ou mesmo em situações sociais que demandam interações ágeis e adaptativas.

4.3.2 Discrepâncias na Detecção de Expressões Específicas e suas Implicações

Apesar de ser bem detectada nos testes de validação (figura 9), a não detecção de expressões de surpresa em um dos vídeos apresentados, sugere uma possível desconexão entre as expressões faciais modeladas no dataset de treinamento e as manifestações reais de surpresa em situações cotidianas. Esta discrepância pode ser atribuída à variação na intensidade das expressões: enquanto os datasets podem apresentar reações altamente expressivas e exageradas, as expressões humanas reais tendem a ser mais sutis e diversificadas. Isso aponta para a necessidade de aprimorar a representatividade e a diversidade dos datasets, incluindo uma gama mais ampla de intensidades emocionais e variações culturais ou individuais nas expressões faciais.

Um aspecto notável identificado durante a aplicação foi a influência significativa da dinâmica grupal nas expressões faciais dos participantes. Como visto no terceiro vídeo, a reação coletiva de riso, apesar do conteúdo gráfico, sugere que o comportamento grupal pode ter um impacto substancial na expressão emocional individual. Isso destaca um desafio importante para o reconhecimento facial em ambientes grupais: a necessidade de distinguir entre expressões emocionais individuais e aquelas influ-

enciadas ou modificadas pela dinâmica do grupo. A capacidade de um algoritmo de reconhecer e interpretar corretamente essas nuances é crucial para aplicações em cenários sociais e educacionais, onde a interação grupal é um componente significativo.

4.3.3 Implicações para o Desenvolvimento Futuro de Algoritmos de Reconhecimento Facial

As observações acima levantam questões críticas para o desenvolvimento futuro de tecnologias de reconhecimento facial, particularmente em ambientes grupais e dinâmicos. A necessidade de algoritmos que possam interpretar com precisão uma ampla variedade de expressões faciais, considerando fatores como intensidade emocional, contextos culturais, e dinâmicas sociais, é clara. A capacidade de ajustar a sensibilidade do modelo para reconhecer variações menos intensas ou expressões atípicas é necessária para a aplicação eficaz dessas tecnologias em situações reais.

Além disso, o experimento sublinha a importância de uma abordagem holística no desenvolvimento de ferramentas de reconhecimento facial. Compreender a psicologia humana, as condições ambientais e as interações sociais é essencial para a criação de modelos que possam ser efetivamente aplicados em cenários da vida real. Isso envolve não apenas a melhoria técnica dos algoritmos, mas também a consideração ética e sensível das diversas experiências e reações humanas.

5 CONCLUSÃO

Neste trabalho, foi desenvolvida uma ferramenta para reconhecimento de expressões faciais em grupos, utilizando tecnologias avançadas de visão computacional e inteligência artificial. O algoritmo foi baseado em um modelo de *transfer learning* com a arquitetura MobileNetV2 e treinado com o dataset FER-2013, resultando em uma acurácia de 68.57% no conjunto de teste.

A aplicação prática do software em um ambiente real com alunos demonstrou a capacidade do modelo de operar em tempo real e capturar expressões faciais em um contexto dinâmico. Os resultados obtidos durante os experimentos com vídeos variados indicaram que o modelo foi capaz de identificar expressões de felicidade, tristeza, raiva, e neutras, entre outras, embora tenham surgido desafios em detectar expressões específicas como nojo e surpresa em certos contextos.

A avaliação dos resultados, em comparação com os trabalhos relacionados, demonstrou que o modelo está em paridade com outras abordagens da literatura, superando muitos em termos de acurácia. A experiência também realçou a importância de datasets representativos e a necessidade de adaptar modelos para reconhecer a diversidade e sutileza das expressões humanas.

A aplicação prática também trouxe à tona a influência das dinâmicas grupais nas expressões emocionais, ressaltando a complexidade de interpretar emoções em contextos coletivos. Os comportamentos coletivos, como o riso em resposta a um conteúdo gráfico, demonstraram a capacidade de influenciar significativamente as expressões faciais dos indivíduos, um fator crítico a ser considerado em futuras melhorias e aplicações da ferramenta.

Este trabalho contribui para o campo do reconhecimento facial, oferecendo percepções valiosos sobre as aplicações práticas e limitações da tecnologia em ambientes reais. As lições aprendidas com os experimentos destacam a necessidade contínua de pesquisa e desenvolvimento na área, especialmente no que diz respeito a aumentar a precisão e a sensibilidade dos modelos, bem como em adaptar as ferramentas para refletir a complexidade das interações humanas e das expressões emocionais.

A realização deste trabalho proporcionou compreensões valiosas sobre a aplicação de tecnologias de reconhecimento facial em grupos, especialmente no contexto educacional. Através do desenvolvimento e aplicação de um algoritmo baseado em Redes Neurais Convolucionais e o conceito de *transfer learning*, foi possível criar uma ferramenta capaz de detectar e reconhecer expressões faciais em tempo real com uma eficácia razoável. O uso do dataset FER-2013 para treinamento e validação mostrou ser uma escolha adequada, fornecendo um bom equilíbrio entre diversidade de expressões e complexidade do dataset.

A aplicação prática do software em um ambiente de sala de aula com estu-

dantes de pós-graduação revelou tanto o potencial quanto as limitações do algoritmo desenvolvido. A ferramenta foi eficaz em capturar as expressões faciais dominantes em situações variadas, refletindo as reações emocionais dos alunos aos diferentes tipos de estímulos visuais apresentados. A experiência demonstrou que, enquanto a ferramenta pode identificar expressões faciais com um nível de precisão considerável, a interpretação desses dados em um contexto grupal dinâmico apresenta desafios significativos.

Os resultados deste trabalho evidenciam a capacidade do algoritmo de reconhecimento de expressões faciais de funcionar efetivamente em tempo real, abrindo possibilidades para sua integração em ambientes educacionais. A utilização dessas tecnologias pode auxiliar professores a identificar rapidamente estados emocionais entre os estudantes. Isso, por sua vez, pode permitir intervenções pedagógicas mais oportunas e personalizadas, promovendo um ambiente de aprendizagem mais responsivo e adaptativo. Além disso, a análise de dados emocionais coletados durante as aulas pode contribuir para o desenvolvimento de estratégias de ensino mais engajadoras e eficazes, melhorando a qualidade da educação e o desempenho dos alunos. Portanto, encoraja-se futuras pesquisas a explorar a integração desta ferramenta em plataformas educacionais digitais e ambientes de sala de aula, avaliando seu impacto no processo de ensino-aprendizagem e nas dinâmicas educacionais.

5.1 TRABALHOS FUTUROS

Este projeto abre várias avenidas para pesquisas e desenvolvimentos futuros. Uma área de melhoria imediata seria a inclusão de um conjunto de dados mais diversificado e representativo durante a fase de treinamento do modelo. Isso poderia ajudar a aumentar a precisão do algoritmo em reconhecer uma gama mais ampla de expressões faciais, especialmente aquelas que são mais sutis ou menos pronunciadas.

Outra direção promissora seria explorar o impacto da dinâmica social e das interações grupais sobre as expressões faciais. O desenvolvimento de modelos que possam considerar esses fatores pode levar a um reconhecimento mais preciso e contextualizado de emoções em grupos. Além disso, aprimorar a capacidade do algoritmo de distinguir entre diferentes tipos de expressões faciais em cenários complexos e diversos poderia expandir significativamente suas aplicações práticas.

A integração da ferramenta com sistemas de apoio à decisão em ambientes educacionais ou corporativos poderia ser outra área de pesquisa valiosa. Isso envolveria não apenas a detecção de emoções, mas também a análise e interpretação desses dados para fornecer informações úteis para educadores, treinadores ou gestores. Por exemplo, o software poderia ser usado para monitorar o engajamento dos alunos em sala de aula ou ambientes virtuais de aprendizagem, fornecendo resposta em tempo real aos educadores sobre a eficácia de suas metodologias de ensino.

Por fim, a aplicação de ferramentas de reconhecimento facial em ambientes educacionais, principalmente, abre um leque de possibilidades para melhorar o processo de ensino e aprendizagem, oferecendo novas formas de entender e responder às necessidades emocionais e cognitivas dos alunos. A continuação da pesquisa e o desenvolvimento nesta área podem proporcionar não apenas avanços tecnológicos, mas também impactos significativos na sociedade, melhorando a comunicação, o bem-estar e a eficácia em vários domínios sociais.

REFERÊNCIAS

AHMAD, Khairol Amali Bin; KHUJAMATOV, Halim; AKHMEDOV, Nurshod; BAJURI, Mohd Yazid; AHMAD, Mohammad Nazir; AHMADIAN, Ali. Emerging trends and evolutions for smart city healthcare systems. **Sustainable Cities and Society**, Elsevier, v. 80, p. 103695, 2022.

ALBAWI, Saad; ABED MOHAMMED, Tareq; ALZAWI, Saad. Understanding of a Convolutional Neural Network. *In*.

ALI, Ghulam; IQBAL, Muhammad Amjad; CHOI, Tae-Sun. Boosted NNE collections for multicultural facial expression recognition. **Pattern Recognition**, Elsevier, v. 55, p. 14–27, 2016.

ALI, Hasimah; HARIHARAN, Muthusamy; YAACOB, Sazali; ADOM, Abdul Hamid. Facial emotion recognition using empirical mode decomposition. **Expert Systems with Applications**, Elsevier, v. 42, n. 3, p. 1261–1277, 2015.

ANILA, Satish; DEVARAJAN, Nanjundappan. Preprocessing technique for face recognition applications under varying illumination conditions. **Global Journal of Computer Science and Technology**, 2012.

BAILENSON, Jeremy N; PONTIKAKIS, Emmanuel D; MAUSS, Iris B; GROSS, James J; JABON, Maria E; HUTCHERSON, Cendri AC; NASS, Clifford; JOHN, Oliver. Real-time classification of evoked emotions using facial feature tracking and physiological responses. **International journal of human-computer studies**, Elsevier, v. 66, n. 5, p. 303–317, 2008.

BELLMAN, Richard Ernest. Artificial intelligence: can computers think? **(No Title)**, 1978.

BETTADAPURA, Vinay. Face expression recognition and analysis: the state of the art. **arXiv preprint arXiv:1203.6722**, 2012.

BISWAS, Suparna; SIL, Jaya. An Efficient Expression Recognition Method using Contourlet Transform. *In*: ACM. PROCEEDINGS of the 2nd International Conference on Perception and Machine Intelligence. [*S.l.*: *s.n.*], 2015. P. 167–174.

CAMBRIA, Erik; DAS, Dipankar; BANDYOPADHYAY, Sivaji; FERACO, Antonio. Affective computing and sentiment analysis. *In*: A practical guide to sentiment analysis. [*S.l.*]: Springer, 2017. P. 1–10.

CANAL, Felipe Zago; LOPEZ, Dennis Paz; POZZEBON, Eliane; SOBIERANSKI, Antonio C. A Systematic Review of Facial Detection and Expression Recognition in Groups of People. **Revista Novas Tecnologias na Educação**, v. 21, n. 2, p. 141–157, 2023.

CANAL, Felipe Zago; MÜLLER, Tobias Rossi; MATIAS, Jhennifer Cristine; SCOTTON, Gustavo Gino; SA JUNIOR, Antonio Reis de; POZZEBON, Eliane; SOBIERANSKI, Antonio Carlos. A survey on facial emotion recognition techniques: A state-of-the-art literature review. **Information Sciences**, Elsevier, v. 582, p. 593–617, 2022.

CHEN, Jingying; CHEN, Dan; GONG, Yujiao; YU, Meng; ZHANG, Kun; WANG, Lizhe. Facial expression recognition using geometric and appearance features. *In*: PROCEEDINGS of the 4th international conference on internet multimedia computing and service. [*S.l.*: *s.n.*], 2012. P. 29–33.

CHENG, Jinkuang; DENG, Yangdong; MENG, Hongying; WANG, Zhihua. A facial expression based continuous emotional state monitoring system with GPU acceleration. *In*: IEEE. 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). [*S.l.*: *s.n.*], 2013. P. 1–6.

COSTA, Lucas José da; SOUSA, Thiago Luz de; SILVA, Francisco Assis da; ALMEIDA, Leandro Luiz; PEREIRA, Danillo Roberto; ARTERO, Almir Olivette; PITERI, Marco Antonio. Análise de Métodos de Detecção e Reconhecimento de Faces Utilizando Visão Computacional e Algoritmos de Aprendizado de Máquina. *In*: 2. COLLOQUIUM Exactarum. ISSN: 2178-8332. [*S.l.*: *s.n.*], 2021. v. 13, p. 1–11.

CRUZ, Albert C; BHANU, Bir; THAKOOR, Ninad S. One shot emotion scores for facial emotion recognition. *In*: IEEE. 2014 IEEE International Conference on Image Processing (ICIP). [*S.l.*: *s.n.*], 2014. P. 1376–1380.

DAVID, Rose Ana Rios *et al.* ANÁLISE COMPARATIVA ENTRE OS PRINCIPAIS ALGORITMOS DE DETECÇÃO FACIAL: HAAR CASCADE, HOG, CNN, YOLO E DEEPFACE. **OPEN SCIENCE RESEARCH V**, Editora Cientıéfica Digital, v. 5, n. 1, p. 439–454, 2022.

DEBNATH, Tanoy; REZA, Md Mahfuz; RAHMAN, Anichur; BEHESHTI, Amin; BAND, Shahab S; ALINEJAD-ROKNY, Hamid. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. **Scientific Reports**, Nature Publishing Group UK London, v. 12, n. 1, p. 6991, 2022.

DELLERMANN, Dominik; EBEL, Philipp; SÖLLNER, Matthias; LEIMEISTER, Jan Marco. Hybrid intelligence. **Business & Information Systems Engineering**, Springer, v. 61, p. 637–643, 2019.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. ImageNet: A Large-Scale Hierarchical Image Database. *In*: CVPR09. [*S.l.*: *s.n.*], 2009.

DHALL, Abhinav; KAUR, Amanjot; GOECKE, Roland; GEDEON, Tom. Emotiw 2018: Audio-video, student engagement and group-level affect prediction. *In*: PROCEEDINGS of the 20th ACM International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2018. P. 653–656.

EKMAN, Paul; FRIESEN, Wallace V. Constants across cultures in the face and emotion. **Journal of personality and social psychology**, American Psychological Association, v. 17, n. 2, p. 124, 1971.

EKMAN, Paul; FRIESEN, Wallace V. Facial action coding system. **Environmental Psychology & Nonverbal Behavior**, 1978.

ESKIL, M Taner; BENLI, Kristin S. Facial expression recognition based on anatomy. **Computer Vision and Image Understanding**, Elsevier, v. 119, p. 1–14, 2014.

FREUND, Yoav; SCHAPIRE, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of computer and system sciences**, Elsevier, v. 55, n. 1, p. 119–139, 1997.

GAN, Yijun. Facial Expression Recognition Using Convolutional Neural Network. *In*: ACM. PROCEEDINGS of the 2nd International Conference on Vision, Image and Signal Processing. [*S.l.*: *s.n.*], 2018. P. 29.

GHASEMI, Raja; AHMADY, Maryam. Facial expression recognition using facial effective areas and Fuzzy logic. *In*: IEEE. 2014 Iranian Conference on Intelligent Systems (ICIS). [*S.l.*: *s.n.*], 2014. P. 1–4.

GIL, Antonio Carlos *et al.* **Como elaborar projetos de pesquisa**. [*S.l.*]: Atlas São Paulo, 2002. v. 4.

GONZÁLEZ-RODRÍGUEZ, M Rosario; DÍAZ-FERNÁNDEZ, M Carmen; GÓMEZ, Carmen Pacheco. Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. **Telematics and Informatics**, Elsevier, v. 51, p. 101404, 2020.

GOODFELLOW, Ian *et al.* Challenges in Representation Learning: A report on three machine learning contests, 2013.

GRAESSER, Arthur C; CONLEY, Mark W; OLNEY, Andrew. Intelligent tutoring systems. **APA educational psychology handbook, Vol 3: Application to learning and teaching.**, American Psychological Association, p. 451–473, 2012.

GU, Jiuxiang *et al.* Recent advances in convolutional neural networks. **Pattern recognition**, Elsevier, v. 77, p. 354–377, 2018.

GUO, Xin; ZHU, Bin; POLANIéA, Luisa F; BONCELET, Charles; BARNER, Kenneth E. Group-level emotion recognition using hybrid deep models based on faces, scenes, skeletons and visual attentions. *In*: PROCEEDINGS of the 20th ACM International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2018. P. 635–639.

GUPTA, Aarush; AGRAWAL, Dakshit; CHAUHAN, Hardik; DOLZ, Jose; PEDERSOLI, Marco. An attention model for group-level emotion recognition. *In*: PROCEEDINGS of the 20th ACM International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2018. P. 611–615.

HAPPY, SL; ROUTRAY, Aurobinda. Automatic facial expression recognition using features of salient facial patches. **IEEE transactions on Affective Computing**, IEEE, v. 6, n. 1, p. 1–12, 2014.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. 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.

HU, Min; WANG, Haowen; WANG, Xiaohua; YANG, Juan; WANG, Ronggui. Video facial emotion recognition based on local enhanced motion history image and

CNN-CTSLSTM networks. **Journal of Visual Communication and Image Representation**, Elsevier, v. 59, p. 176–185, 2019.

HUANG, Shujun; CAI, Nianguang; PACHECO, Pedro Penzuti; NARRANDES, Shavira; WANG, Yang; XU, Wayne. Applications of support vector machine (SVM) learning in cancer genomics. **Cancer Genomics-Proteomics**, International Institute of Anticancer Research, v. 15, n. 1, p. 41–51, 2018.

HUANG, Xiaohua; DHALL, Abhinav; GOECKE, Roland; PIETIKAINEN, Matti K; ZHAO, Guoying. Analyzing group-level emotion with global alignment kernel based approach. **IEEE Transactions on Affective Computing**, IEEE, 2019.

HWANG, Gwo-Jen. A conceptual map model for developing intelligent tutoring systems. **Computers & Education**, Elsevier, v. 40, n. 3, p. 217–235, 2003.

JAIN, Deepak Kumar; SHAMSOLMOALI, Pourya; SEHDEV, Paramjit. Extended deep neural network for facial emotion recognition. **Pattern Recognition Letters**, Elsevier, v. 120, p. 69–74, 2019.

JONES, Michael; VIOLA, Paul. Fast multi-view face detection. **Mitsubishi Electric Research Lab TR-20003-96**, v. 3, n. 14, p. 2, 2003.

KANADE, Takeo; COHN, Jeffrey F; TIAN, Yingli. Comprehensive database for facial expression analysis. *In*: IEEE. PROCEEDINGS Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). [*S.l.*: *s.n.*], 2000. P. 46–53.

KARTALI, Aneta; ROGLIĆ, Miloš; BARJAKTAROVIĆ, Marko; ĐURIĆ-JOVIČIĆ, Milica; JANKOVIĆ, Milica M. Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches. *In*: IEEE. 2018 14th Symposium on Neural Networks and Applications (NEUREL). [*S.l.*: *s.n.*], 2018. P. 1–4.

KHAN, Ahmed Shehab; LI, Zhiyuan; CAI, Jie; MENG, Zibo; O'REILLY, James; TONG, Yan. Group-level emotion recognition using deep models with a four-stream hybrid network. *In*: PROCEEDINGS of the 20th ACM International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2018. P. 623–629.

KITCHENHAM, Barbara. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004.

KO, Byoung Chul. A brief review of facial emotion recognition based on visual information. **sensors**, MDPI, v. 18, n. 2, p. 401, 2018.

LI, Shan; DENG, Weihong. Deep facial expression recognition: A survey. **IEEE transactions on affective computing**, IEEE, 2020.

LI, Weihan; SENGUPTA, Neil; DECHENT, Philipp; HOWEY, David; ANNASWAMY, Anuradha; SAUER, Dirk Uwe. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. **Journal of power sources**, Elsevier, v. 482, p. 228863, 2021.

LI, Yongqiang; WANG, Shangfei; ZHAO, Yongping; JI, Qiang. Simultaneous facial feature tracking and facial expression recognition. **IEEE Transactions on Image Processing**, IEEE, v. 22, n. 7, p. 2559–2573, 2013.

LI, Zewen; LIU, Fan; YANG, Wenjie; PENG, Shouheng; ZHOU, Jun. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, IEEE, 2021.

LIU, Bing *et al.* Sentiment analysis and subjectivity. **Handbook of natural language processing**, Oxfordshire, v. 2, n. 2010, p. 627–666, 2010.

LIU, Chuanhe; JIANG, Wenqiang; WANG, Minghao; TANG, Tianhao. Group level audio-video emotion recognition using hybrid networks. *In*: PROCEEDINGS of the 2020 International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2020. P. 807–812.

LOPES, Andre Teixeira; DE AGUIAR, Edilson; OLIVEIRA-SANTOS, Thiago. A facial expression recognition system using convolutional networks. *In*: IEEE. 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images. [*S.l.*: *s.n.*], 2015. P. 273–280.

LOPES, André Teixeira; AGUIAR, Edilson de; DE SOUZA, Alberto F; OLIVEIRA-SANTOS, Thiago. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. **Pattern Recognition**, Elsevier, v. 61, p. 610–628, 2017.

LUCEY, Patrick; COHN, Jeffrey F; KANADE, Takeo; SARAGIH, Jason; AMBADAR, Zara; MATTHEWS, Iain. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *In*: IEEE. 2010 ieee

computer society conference on computer vision and pattern recognition-workshops. [*S.l.*: *s.n.*], 2010. P. 94–101.

LUCEY, Patrick; COHN, Jeffrey F; KANADE, Takeo; SARAGIH, Jason; AMBADAR, Zara; MATTHEWS, Iain. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *In*: IEEE. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. [*S.l.*: *s.n.*], 2010. P. 94–101.

LUO, Yuan; ZHANG, Ling; CHEN, Yunhua; JIANG, Wenchao. Facial expression recognition algorithm based on reverse co-salient regions (RCSR) features. *In*: IEEE. 2017 4th International Conference on Information Science and Control Engineering (ICISCE). [*S.l.*: *s.n.*], 2017. P. 326–329.

LYONS, Michael; AKAMATSU, Shigeru; KAMACHI, Miyuki; GYOBA, Jiro. Coding facial expressions with gabor wavelets. *In*: IEEE. PROCEEDINGS Third IEEE international conference on automatic face and gesture recognition. [*S.l.*: *s.n.*], 1998. P. 200–205.

MEHRABIAN, Albert. Communication without words. *In*: COMMUNICATION theory. [*S.l.*]: Routledge, 2017. P. 193–200.

MEHTA, Neelum; JADHAV, Sangeeta. Facial emotion recognition using log Gabor filter and PCA. *In*: IEEE. 2016 International Conference on Computing Communication Control and automation (ICCUBEA). [*S.l.*: *s.n.*], 2016. P. 1–5.

MENDES, Kreisler Brenner *et al.* Comparativo de algoritmos clássicos de aprendizado de máquina em um problema de reconhecimento de faces. Universidade Federal de Uberlândia, 2019.

MIRANDA-CORREA, Juan Abdon; ABADI, Mojtaba Khomami; SEBE, Nicu; PATRAS, Ioannis. Amigos: A dataset for affect, personality and mood research on individuals and groups. **IEEE Transactions on Affective Computing**, IEEE, v. 12, n. 2, p. 479–493, 2018.

MOHSENI, Sina; ZAREI, Niloofar; RAMAZANI, Saba. Facial expression recognition using anatomy based facial graph. *In*: IEEE. 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). [*S.l.*: *s.n.*], 2014. P. 3715–3719.

MOLERO JURADO, Maríéa del Mar; PÉREZ-FUENTES, Maríéa del Carmen; MARTOS MARTIéNEZ, África; BARRAGÁN MARTIéN, Ana Belén; SIMÓN MÁRQUEZ, Maríéa del Mar; GÁZQUEZ LINARES, José Jesús. Emotional intelligence as a mediator in the relationship between academic performance and burnout in high school students. **Plos one**, Public Library of Science San Francisco, CA USA, v. 16, n. 6, e0253552, 2021.

MOU, Wenxuan; GUNES, Hatice; PATRAS, Ioannis. Alone versus in-a-group: A multi-modal framework for automatic affect recognition. **ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)**, ACM New York, NY, USA, v. 15, n. 2, p. 1–23, 2019.

NG, Hong-Wei; NGUYEN, Viet Dung; VONIKAKIS, Vassilios; WINKLER, Stefan. Deep learning for emotion recognition on small datasets using transfer learning. *In*: PROCEEDINGS of the 2015 ACM on international conference on multimodal interaction. [*S.l.*: *s.n.*], 2015. P. 443–449.

PALESTRA, Giuseppe; PINO, Olimpia. Detecting emotions during a memory training assisted by a social robot for individuals with Mild Cognitive Impairment (MCI). **Multimedia Tools and Applications**, Springer, v. 79, n. 47, p. 35829–35844, 2020.

PANTIC, Maja; VALSTAR, Michel; RADEMAKER, Ron; MAAT, Ludo. Web-based database for facial expression analysis. *In*: IEEE. 2005 IEEE international conference on multimedia and Expo. [*S.l.*: *s.n.*], 2005. 5–pp.

PAPAGEORGIOU, Constantine P; OREN, Michael; POGGIO, Tomaso. A general framework for object detection. *In*: IEEE. SIXTH International Conference on Computer Vision (IEEE Cat. No. 98CH36271). [*S.l.*: *s.n.*], 1998. P. 555–562.

PETROVA, Anastasia; VAUFREYDAZ, Dominique; DESSUS, Philippe. Group-level emotion recognition using a unimodal privacy-safe non-individual approach. *In*: PROCEEDINGS of the 2020 International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2020. P. 813–820.

PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO, Programa de. **Linhas de Pesquisa**. Edição: PPGTIC. Araranguá: [*s.n.*], 2022. `https://ppgtic.ufsc.br/linhas-de-pesquisa/`. Acesso em: 15 out. 2022.

QUACH, Kha Gia; LE, Ngan; DUONG, Chi Nhan; JALATA, Ibsa; ROY, Kaushik; LUU, Khoa. Non-volume preserving-based fusion to group-level emotion recognition on crowd videos. **Pattern Recognition**, Elsevier, p. 108646, 2022.

REBOLLEDO-MENDEZ, Genaro; HUERTA-PACHECO, N Sofia; BAKER, Ryan S; BOULAY, Benedict du. Meta-affective behaviour within an intelligent tutoring system for mathematics. **International Journal of Artificial Intelligence in Education**, Springer, v. 32, n. 1, p. 174–195, 2022.

REVINA, I Michael; EMMANUEL, WR Sam. A survey on human face expression recognition techniques. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, v. 33, n. 6, p. 619–628, 2021.

RICHARDSON, John TE. Instruments for obtaining student feedback: A review of the literature. **Assessment & evaluation in higher education**, Taylor & Francis, v. 30, n. 4, p. 387–415, 2005.

SAKAI, Toshiyuki; NAGAO, Makoto; FUJIBAYASHI, Shinya. Line extraction and pattern detection in a photograph. **Pattern recognition**, Elsevier, v. 1, n. 3, p. 233–248, 1969.

SALUNKE, Vibha V; PATIL, CG. A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks. *In*: IEEE. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). [*S.l.*: *s.n.*], 2017. P. 1–5.

SANDLER, Mark; HOWARD, Andrew; ZHU, Menglong; ZHMOGINOV, Andrey; CHEN, Liang-Chieh. 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.

SHARMA, Archana; MANSOTRA, Vibhakar. Multimodal decision-Level group sentiment prediction of students in classrooms. **Int. J. Innov. Technol. Explor. Eng**, v. 8, n. 12, p. 4902–4909, 2019.

SIGOV, Alexander; RATKIN, Leonid; IVANOV, Leonid A; XU, Li Da. Emerging enabling technologies for industry 4.0 and beyond. **Information Systems Frontiers**, Springer, p. 1–11, 2022.

SILANDER, Tomi; MYLLYMAKI, Petri. A simple approach for finding the globally optimal Bayesian network structure. **arXiv preprint arXiv:1206.6875**, 2012.

SILVA, Jennifer Amanda Sobral da; MAIRINK, Carlos Henrique Passos. Inteligência artificial. **LIBERTAS: Revista de Ciênciais Sociais Aplicadas**, v. 9, n. 2, p. 64–85, 2019.

SOO, Sander. Object detection using Haar-cascade Classifier. **Institute of Computer Science, University of Tartu**, p. 1–12, 2014.

SOUZA, Lıévia Barbosa Pacheco *et al.* Inteligência Artificial Na Educação: Rumo A Uma Aprendizagem Personalizada.

SRIVASTAVA, Shivam; LAKSHMINARAYAN, Saandeep Aathreya SIdhapur; HINDUJA, Saurabh; JANNAT, Sk Rahatul; ELHAMDADI, Hamza; CANAVAN, Shaun. Recognizing emotion in the wild using multimodal data. *In*: PROCEEDINGS of the 2020 International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2020. P. 849–857.

STAUDEMEYER, Ralf C; MORRIS, Eric Rothstein. Understanding LSTM–a tutorial into long short-term memory recurrent neural networks. **arXiv preprint arXiv:1909.09586**, 2019.

SUN, Jian-Ming; PEI, Xue-Sheng; ZHOU, Shi-Sheng. Facial emotion recognition in modern distant education system using SVM. *In*: IEEE. 2008 International Conference on Machine Learning and Cybernetics. [*S.l.*: *s.n.*], 2008. v. 6, p. 3545–3548.

SUN, Mo; LI, Jian; FENG, Hui; GOU, Wei; SHEN, Haifeng; TANG, Jian; YANG, Yi; YE, Jieping. Multi-Modal Fusion Using Spatio-Temporal and Static Features for Group Emotion Recognition. *In*: PROCEEDINGS of the 2020 International Conference on Multimodal Interaction. [*S.l.*: *s.n.*], 2020. P. 835–840.

TAO, Jianhua; TAN, Tieniu. Affective computing: A review. *In*: SPRINGER. INTERNATIONAL Conference on Affective computing and intelligent interaction. [*S.l.*: *s.n.*], 2005. P. 981–995.

TURABZADEH, Saeed; MENG, Hongying; SWASH, Rafiq M; PLEVA, Matus; JUHAR, Jozef. Real-time emotional state detection from facial expression on embedded devices. *In*: IEEE. 2017 Seventh International Conference on Innovative Computing Technology (INTECH). [*S.l.*: *s.n.*], 2017. P. 46–51.

VĂIDEAN, Viorela Ligia; ACHIM, Monica Violeta. When more is less: Do information and communication technologies (ICTs) improve health outcomes? An empirical investigation in a non-linear framework. **Socio-Economic Planning Sciences**, Elsevier, v. 80, p. 101218, 2022.

VAPNIK, Vladimir. Pattern recognition using generalized portrait method. **Automation and remote control**, v. 24, p. 774–780, 1963.

VICARI, Rosa Maria. Tendências em inteligência artificial na educação no perıéodo de 2017 a 2030: sumário excutivo. Diretoria de Tecnologia e Educação (DIRET). Unidade de Estudos e Prospectiva . . ., 2018.

VIOLA, Paul; JONES, Michael *et al.* Rapid object detection using a boosted cascade of simple features. **CVPR (1)**, v. 1, n. 511-518, p. 3, 2001.

WANG, Fengyuan; LV, Jianhua; YING, Guode; CHEN, Shenghui; ZHANG, Chi. Facial expression recognition from image based on hybrid features understanding. **Journal of Visual Communication and Image Representation**, Elsevier, v. 59, p. 84–88, 2019.

WANG, Kai; ZENG, Xiaoxing; YANG, Jianfei; MENG, Debin; ZHANG, Kaipeng; PENG, Xiaojiang; QIAO, Yu. Cascade attention networks for group emotion recognition with face, body and image cues. *In*: PROCEEDINGS of the 20th ACM international conference on multimodal interaction. [*S.l.*: *s.n.*], 2018. P. 640–645.

WANG, Lipo. **Support vector machines: theory and applications**. [*S.l.*]: Springer Science & Business Media, 2005. v. 177.

WANG, XuMing; HUANG, Jin; ZHU, Jia; YANG, Min; YANG, Fen. Facial expression recognition with deep learning. *In*: ACM. PROCEEDINGS of the 10th International Conference on Internet Multimedia Computing and Service. [*S.l.*: *s.n.*], 2018. P. 10.

WHITEHILL, Jacob; LITTLEWORT, Gwen; FASEL, Ian; BARTLETT, Marian; MOVELLAN, Javier. Toward practical smile detection. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 31, n. 11, p. 2106–2111, 2009.

YEASIN, Mohammed; BULLOT, Baptiste; SHARMA, Rajeev. Recognition of facial expressions and measurement of levels of interest from video. **IEEE Transactions on Multimedia**, IEEE, v. 8, n. 3, p. 500–508, 2006.

YIN, Robert K. **Estudo de Caso-: Planejamento e métodos**. [*S.l.*]: Bookman editora, 2015.

YIN, Robert K. **Pesquisa qualitativa do inıécio ao fim**. [*S.l.*]: Penso Editora, 2016.

YU, Dai; XINGYU, Liu; SHUZHAN, Dong; LEI, Yang. Group emotion recognition based on global and local features. **IEEE Access**, IEEE, v. 7, p. 111617–111624, 2019.

YU, Yong; SI, Xiaosheng; HU, Changhua; ZHANG, Jianxun. A review of recurrent neural networks: LSTM cells and network architectures. **Neural computation**, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . ., v. 31, n. 7, p. 1235–1270, 2019.

YU, Zhiding; ZHANG, Cha. Image based static facial expression recognition with multiple deep network learning. *In*: PROCEEDINGS of the 2015 ACM on international conference on multimodal interaction. [*S.l.*: *s.n.*], 2015. P. 435–442.

ZHANG, Yu-Dong; YANG, Zhang-Jing; LU, Hui-Min; ZHOU, Xing-Xing; PHILLIPS, Preetha; LIU, Qing-Ming; WANG, Shui-Hua. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. **IEEE Access**, IEEE, v. 4, p. 8375–8385, 2016.

ZHANG, Kaipeng; ZHANG, Zhanpeng; LI, Zhifeng; QIAO, Yu. Joint face detection and alignment using multitask cascaded convolutional networks. **IEEE Signal Processing Letters**, IEEE, v. 23, n. 10, p. 1499–1503, 2016.

APÊNDICE A – FORMULÁRIO DE APLICAÇÃO DO ALGORITMO

O questionário aplicado aos estudantes durante o experimento pode ser visualizado no endereço abaixo:

https://forms.gle/RdKbFUhYG5Vy3icm8

Reconhecimento de Expressão Facial para Grupos de Pessoas em Tempo Real

1. E-mail (obrigatório)
2. Indique a frequência de cada sentimento que você experienciou durante a execução do primeiro vídeo? (obrigatório)

	Nunca	Raramente	Ocasionalmente	Frequentemente	Muito Frequentemente
Raiva	○	○	○	○	○
Nojo	○	○	○	○	○
Medo	○	○	○	○	○
Felicidade	○	○	○	○	○
Neutro	○	○	○	○	○
Tristeza	○	○	○	○	○
Surpresa	○	○	○	○	○

3. Você acredita ter expressado esse sentimento com expressões faciais durante a execução do vídeo?

	Não	Sim	Não sei
Raiva	○	○	○
Nojo	○	○	○
Medo	○	○	○
Felicidade	○	○	○
Neutro	○	○	○
Tristeza	○	○	○
Surpresa	○	○	○

4. Caso deseje, deixe seu comentário sobre o vídeo ou sobre as emoções experienciadas durante a atividade.