Aplicação de DeepFace e OpenFace para Identificação de Sentimentos Básicos em Vídeos de Teleconsulta

Aurea Ferreira Nascimento[1], Fábio Castro Araújo[1]

[1]Departamento de Computação
Universidade Luterana do Brasil – Palmas – TO

aureaf11@rede.ulbra.br, fabio.araujo@ulbra.br

Resumo. *A teleconsulta dificulta a percepção de sinais não verbais fundamentais para o diagnóstico psicológico. Para superar essa limitação, este artigo propõe a integração do OpenFace e do DeepFace na análise facial automatizada em tempo real. O sistema extrai Unidades de Ação (AUs) e realiza inferência emocional, estruturando os dados em dashboards analíticos. A principal contribuição consiste em uma arquitetura que converte expressões sutis em indicadores visuais de apoio ao terapeuta. Como resultado, a solução permite o monitoramento objetivo de padrões afetivos, atuando como um suporte tecnológico que complementa, sem substituir, o julgamento clínico.*

Abstract. *Teleconsultations hinder the perception of nonverbal cues that are fundamental to psychological diagnosis. To address this limitation, this article proposes the integration of OpenFace and DeepFace for automated real-time facial analysis. The system extracts Action Units (AUs) and performs emotional inference, organizing the data into analytical dashboards. The main contribution is an architecture that transforms subtle facial expressions into visual indicators to support therapists. As a result, the proposed solution enables objective monitoring of affective patterns, serving as a technological aid that complements, rather than replaces, clinical judgment.*

1. Introdução

O desenvolvimento da visão computacional tem sido marcado por avanços progressivos no reconhecimento facial e na análise automática de expressões. Um marco histórico nesse processo foi a introdução das redes neurais convolucionais (CNNs), inicialmente propostas por LeCun et al. (1989) para reconhecimento de padrões visuais. Desde então, a consolidação de arquiteturas de aprendizado profundo, aliada ao aumento da capacidade computacional e à disponibilidade de grandes bases de dados, ampliou significativamente a precisão e a robustez de sistemas de análise visual (Goodfellow, Bengio e Courville, 2016).

Esses avanços permitiram a transição de aplicações restritas a ambientes controlados para contextos operacionais mais complexos. Atualmente, técnicas de reconhecimento facial e análise de expressões são empregadas em áreas como segurança, interação humano-computador, interfaces inteligentes, educação e análise comportamental (Cohn e De la Torre, 2015; Li e Deng, 2020). Em particular, o reconhecimento automático de emoções em imagens e vídeos tem despertado crescente interesse, sendo aplicado, por exemplo, no monitoramento de engajamento, na detecção de estados afetivos e no suporte a decisões em sistemas sensíveis ao contexto emocional.

No campo da computação afetiva, diferentes abordagens têm sido propostas para inferir estados emocionais a partir de sinais visuais. Métodos baseados em aprendizado profundo utilizam redes neurais convolucionais para extrair representações discriminativas diretamente das imagens faciais, enquanto abordagens baseadas no Facial Action Coding System (FACS) modelam explicitamente os movimentos musculares da face por meio de Action Units (AUs), permitindo uma análise mais interpretável das expressões (Ekman e Friesen, 1978). A combinação dessas estratégias tem se mostrado promissora para a identificação de emoções básicas em fotos e vídeos.

Nesse contexto, ferramentas como *DeepFace* e *OpenFace* tornaram-se amplamente utilizadas. O *DeepFace*, introduzido por Taigman et al. (2014), emprega modelos de deep learning capazes de representar faces por meio de *embeddings* robustos, viabilizando tarefas como verificação, identificação e análise emocional (Schroff, Kalenichenko e Philbin, 2015; Deng et al., 2019). Já o *OpenFace* é uma ferramenta *open source* baseada no *FACS*, que fornece uma decomposição detalhada da ativação muscular facial por meio da extração automática de Action Units, possibilitando a análise de microexpressões e padrões sutis de movimento facial (Baltrušaitis et al., 2018).

A identificação de emoções básicas apresenta relevância consolidada em áreas como psicologia, psiquiatria, educação e atendimento ao cliente (Ekman e Friesen, 1971; Ekman, 1999). No entanto, no contexto de teleconsultas psicológicas, a interpretação emocional enfrenta limitações decorrentes de fatores como qualidade variável de vídeo, ausência de profundidade, variações de iluminação e restrições na comunicação não verbal mediada por câmera (Luxton et al., 2011; Shore, Schneck e Mishkind, 2020). Essas limitações podem comprometer a leitura precisa das expressões faciais durante o atendimento clínico.

Diante desse cenário, torna-se relevante investigar soluções computacionais capazes de fornecer indicadores auxiliares à prática clínica, reforçando a interpretação do terapeuta sem substituí-la. Destacam-se, nesse âmbito, (i) modelos baseados no Facial Action Coding System, que descrevem objetivamente os movimentos faciais por meio de Action Units (Ekman e Friesen, 1978), e (ii) modelos de deep learning, especialmente redes neurais convolucionais, amplamente empregadas na extração automática de padrões visuais em imagens faciais (LeCun, Bengio e Hinton, 2015).

Assim, este estudo é orientado pela seguinte questão de pesquisa: como identificar de forma objetiva emoções básicas expressas por pacientes em sessões de teleconsulta psicológica a partir de vídeos? Para responder a essa questão, o presente artigo propõe e descreve a implementação integrada das bibliotecas *DeepFace* e *OpenFace* aplicadas ao processamento de vídeos de teleconsulta. O objetivo é apresentar uma arquitetura operacional para a detecção de emoções básicas, detalhando o pipeline de processamento, os componentes do sistema e sua integração com um dashboard clínico. A proposta visa fornecer sinais analíticos complementares ao profissional de saúde mental, preservando o protagonismo do julgamento clínico.

2. Fundamentação Teórica

A análise computacional das emoções faciais consolidou-se como um dos vetores mais relevantes dentro da inteligência artificial aplicada ao comportamento humano. A maturidade das redes neurais convolucionais (*CNNs*) permitiu a extração automática de

padrões de alto nível, ampliando a capacidade de interpretar expressões em ambientes não controlados e em fluxos de vídeos comuns [Goodfellow, Bengio e Courville, 2016; Li e Deng, 2020]. Esse movimento viabilizou soluções operacionais que hoje permeiam saúde, educação, segurança e serviços digitais [Cohn e De la Torre, 2015].

A presente fundamentação organiza-se em cinco eixos conceituais: (i) emoções básicas, (ii) análise de expressões faciais, (iii) classificação por *deep learning*, (iv) mapeamento por *Action Units* via *OpenFace* e (v) aplicações em telemedicina. Essa segmentação sustenta tecnicamente o *pipeline* proposto. As Figuras 1 a 4, referenciadas ao longo da seção, ilustram os constructos abordados.

2.1 Emoções Básicas

A teoria das emoções básicas, proposta por Ekman e Friesen (1971; 1978), define um conjunto de expressões universais observáveis em todas as culturas humanas. As seis emoções fundamentais (alegria, medo, surpresa, tristeza, nojo e raiva) apresentam padrões musculares específicos e distinguíveis. Esses padrões foram mapeados e codificados no Facial Action Coding System (FACS), desenvolvido por Ekman e Friesen (1978), permitindo a padronização da análise facial para fins científicos e clínicos.

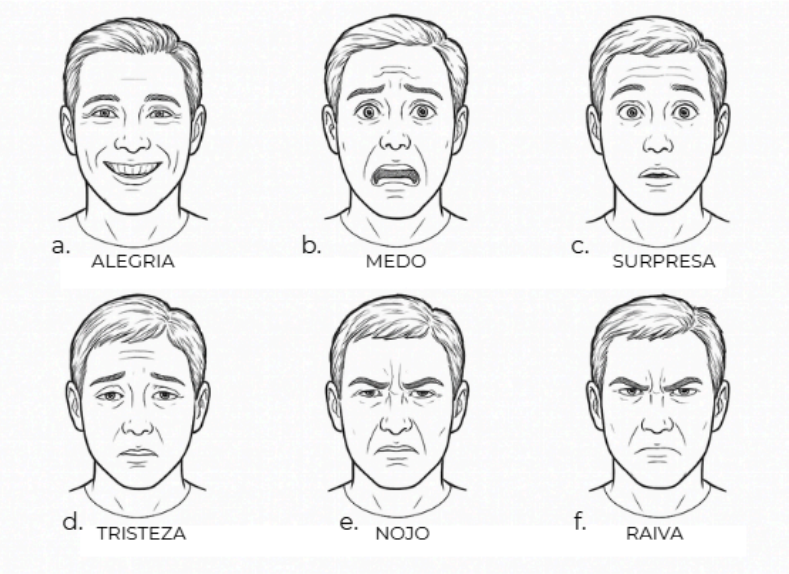


Figura 1 - Expressões faciais típicas de seis emoções básicas

A Figura 1 ilustra exemplos prototípicos dessas expressões básicas, evidenciando as características faciais mais marcantes de cada emoção: o sorriso amplo e a elevação das bochechas na alegria (a), o arqueamento acentuado das sobrancelhas e o alargamento dos olhos no medo (b), o olhar ampliado e a boca aberta na surpresa (c), o abaixamento dos cantos da boca na tristeza (d), a contração nasal típica do nojo (e) e o franzimento das sobrancelhas associado à raiva (f). Esses padrões foram sistematizados no Facial Action Coding System (FACS), que codifica cada expressão em termos de Action Units (AUs) específicas, permitindo a padronização da análise facial tanto em estudos clínicos quanto em aplicações computacionais (EKMAN; FRIESEN, 1978). Essa estrutura conceitual viabilizou o treinamento supervisionado de modelos em bases

como FER-2013 e AffectNet [Mollahosseini et al., 2019], amplamente empregadas em pesquisas contemporâneas de reconhecimento de emoções.

Autores contemporâneos, como Barrett et al. (2019), ressaltam que emoções não são eventos puramente discretos, mas construções influenciadas por fatores culturais e contextuais. Essa crítica recomenda cautela na interpretação algorítmica e reforça a necessidade de combinar sinais computacionais com conhecimento clínico.

Trabalhos introdutórios sobre reconhecimento automático de emoções destacam que é essencial diferenciar emoções espontâneas, expressões moduladas e microexpressões, pois cada tipo envolve intensidades e tempos de ativação distintos, afetando diretamente a sensibilidade dos modelos de IA [Silva, 2023]. A literatura também reforça que as emoções possuem natureza dinâmica: uma expressão não surge de forma instantânea, mas percorre fases de início, pico e recuperação (EKMAN, 2003; COHN; DE LA TORRE, 2015). Considerar essa variação temporal torna a análise mais precisa, permitindo que os sistemas computacionais identifiquem nuances afetivas que se manifestam ao longo do tempo.

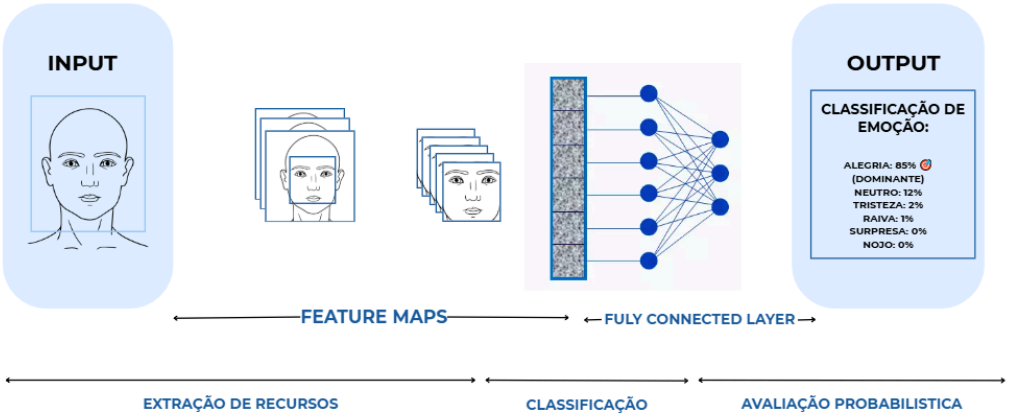
2.2 Análise de Expressões Faciais

Com o amadurecimento da Visão Computacional, a Análise de Expressões Faciais tornou-se um dos campos mais desenvolvidos do reconhecimento de padrões. As metodologias podem ser agrupadas em duas abordagens principais: geométrica, baseada na extração de *landmarks* faciais, e baseada em deep learning, fundamentada em modelos de aprendizado profundo (COHN; DE LA TORRE, 2015; LI; DENG, 2020).

A abordagem geométrica baseia-se na detecção de pontos-chave faciais (*landmarks*), correspondentes a coordenadas anatômicas localizadas em regiões como olhos, sobrancelhas, nariz, boca e contorno facial. A partir desses pontos, são mensuradas distâncias e ângulos, garantindo maior interpretabilidade dos resultados. Contudo, essa abordagem apresenta limitações sob iluminação irregular, oclusões parciais (mãos, cabelo, óculos) ou variações significativas de pose facial (ZENG et al., 2009).

Já a abordagem baseada em *Deep Learning* utiliza redes neurais convolucionais (CNNs) para aprender representações hierárquicas diretamente das imagens, dispensando extração manual de características. As *CNNs* se mostram mais robustas frente a ruído, variação de pose e condições ambientais, sendo o método predominante nas pesquisas atuais [GoodFellow; Bengio; Courville, 2016; Li e Deng, 2020].

Figura 2 - Arquitetura Conceitual de uma CNN. (Fluxo: Entrada → Camadas Convolucionais → Classificação)



A Figura 2 apresenta uma representação conceitual do fluxo interno de uma rede neural convolucional (CNN) aplicada à classificação de emoções faciais. O processo inicia-se com a imagem de entrada (Input), contendo o rosto detectado e alinhado, que é submetida às camadas convolucionais. Estas são responsáveis por extrair padrões visuais como contornos, texturas e a geometria de elementos-chave, como sobrancelhas, boca e olhos, gerando os chamados *feature maps*, que representam diferentes níveis de abstração das características faciais.

Em seguida, essas informações são encaminhadas às camadas totalmente conectadas, que integram os padrões extraídos para a etapa final de decisão. Por fim, o modelo produz um vetor de probabilidades associado a cada categoria emocional no Output, onde expressões como alegria, tristeza e raiva recebem valores percentuais de confiança. Essa sequência ilustra o funcionamento dos modelos utilizados neste estudo (*DeepFace*), demonstrando como a imagem facial é transformada em uma predição emocional interpretável.

Nos sistemas modernos, ambas as técnicas são frequentemente combinadas: os *landmarks* garantem estabilidade e rastreabilidade da face, enquanto as CNNs operam sobre regiões faciais alinhadas, produzindo *embeddings* que alimentam classificadores de emoção. Essa combinação é adotada neste estudo, integrando o rastreamento geométrico via *OpenFace* e a classificação probabilística via *DeepFace*.

2.3 DeepFace: A Classificação por Deep Learning

O *DeepFace*, desenvolvido pelo *Facebook AI Research* [Taigman et al., 2014], foi um marco no reconhecimento facial por aprendizado profundo, alcançando desempenho próximo ao humano na verificação de identidade (*LFW ~97,35%*). Sua implementação em *Python*, mantida por Serengil e Özpınar (2020), expandiu o escopo para reconhecimento, verificação, análise de atributos e classificação emocional.
O fluxo geral de processamento organiza-se em três etapas fundamentais:

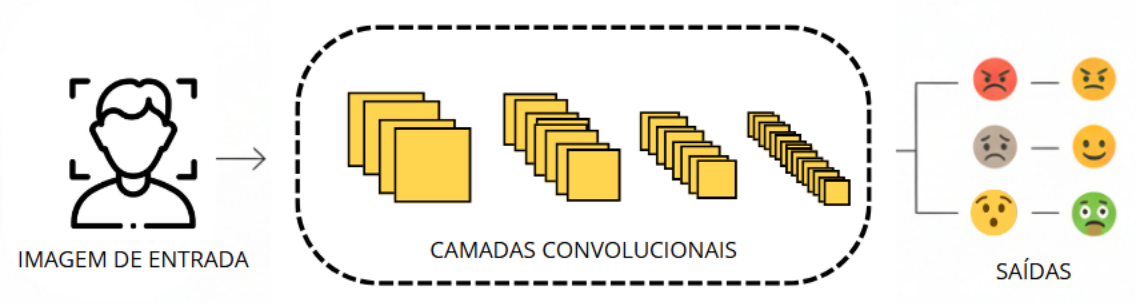


Figura 3 - Fluxograma de Processamento.

A Figura 3 apresenta uma visão simplificada do fluxo de processamento do *DeepFace*. A partir da imagem facial de entrada, o sistema aplica uma série de camadas convolucionais, responsáveis por extrair automaticamente padrões relevantes do rosto, como contornos, texturas e variações musculares. Esses mapas de características são então utilizados para inferir a emoção predominante, resultando em saídas classificadas, como raiva, tristeza, surpresa, alegria ou nojo. O fluxo resume o funcionamento essencial do modelo na análise facial.

O *DeepFace* suporta múltiplos *backbones*, entendidos como arquiteturas centrais de redes neurais profundas responsáveis pela extração das representações faciais de alto nível. Cada *backbone* define a forma como as características visuais são aprendidas e codificadas, influenciando a robustez e a capacidade discriminativa do modelo.
Entre os principais *backbones* suportados estão:

- **VGG-Face** (Parkhi et al., 2015);
- **FaceNet** (Schroff et al., 2015);
- **ArcFace** (Deng et al., 2019);
- **DeepID** (Sun et al., 2014).

Essas ferramentas, originalmente projetadas para verificação de identidade, podem ser adaptadas para análise de emoção mediante bases rotuladas como FER-2013. Em *benchmarks* públicos, a acurácia na classificação de emoções varia amplamente conforme *dataset* e protocolo experimental, geralmente inferior aos resultados de verificação facial [Mollahosseini et al., 2019; LI; Deng, 2020].

Assim, qualquer aplicação clínica deve ser precedida por calibração contextual e validação experimental. A arquitetura modular do *DeepFace* permite combinar diferentes CNNs e ajustar sensibilidade e latência, tornando-o útil em protótipos clínicos exploratórios.

2.4 OpenFace e o Facial Action Coding System (FACS)

O *OpenFace* é um *toolkit* de código aberto para análise de comportamento facial, baseado no *Facial Action Coding System* [Ekman; Friesen, 1978]. Ele realiza detecção de *landmarks*, rastreamento de movimento, extração de *Action Units* (AUs) e estimativa de pose e olhar.

Figura 4 - Ilustração do FACS e Action Units (AUs).

Fonte: Adaptado de Cohn e De la Torre (2015, Fig. 10.2).

A Figura 4 apresenta uma representação esquemática do FACS, destacando os principais grupos musculares faciais utilizados para codificar expressões. Embora a ilustração não exiba explicitamente os pontos de referência facial (*landmarks*), é com base neles que sistemas computacionais rastreiam o movimento das regiões do rosto. Esses *landmarks* servem como marcadores que indicam a direção típica das ativações musculares associadas às *Action Units* (AUs), permitindo compreender como diferentes combinações de movimentos dão origem às expressões analisadas.

Diferentemente do *DeepFace*, o *OpenFace* não classifica emoções diretamente, mas identifica a ativação e a intensidade dos músculos faciais por meio das Action Units (AUs) definidas no Facial Action Coding System (FACS). Cada emoção básica pode ser descrita como uma combinação específica de AUs, o que permite inferir estados emocionais a partir dos padrões de ativação muscular, fornecendo uma representação mais interpretável da expressão facial.

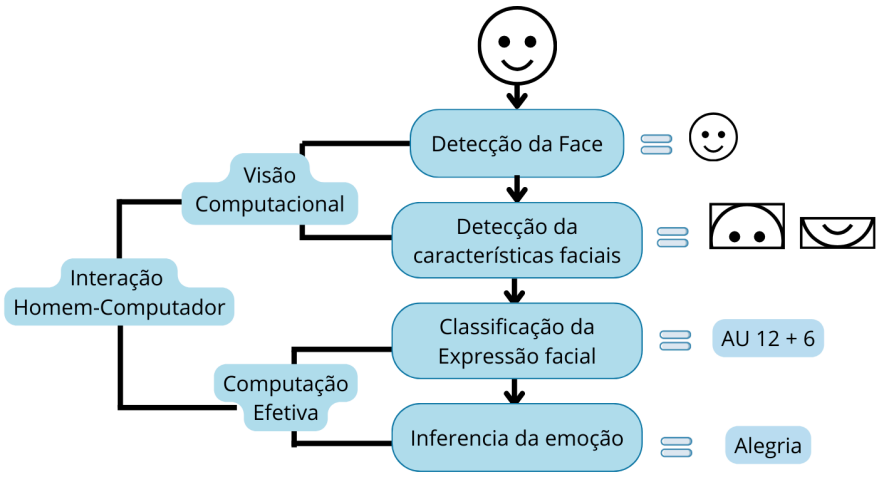


Figura 5 - Metodologia da aplicação do OpenFace.

Na Figura 5, um indivíduo posiciona-se em frente à câmera para a captura, que realiza a detecção de sua face e a separa do restante da imagem. A partir da face detectada, são extraídas e separadas as principais características faciais dos olhos, sobrancelhas, boca, nariz e bochechas de modo que classificadores possam estimar as Unidades de Ação (AUs) associadas ao movimento muscular. Do ponto de vista de Visão Computacional, o fluxo descreve a etapa de processamento automático da imagem, enquanto, sob a perspectiva da Interação Homem-Computador, ilustra como o sistema capta e traduz gestos espontâneos do usuário em sinais interpretáveis pela máquina. Por fim, o somatório e a combinação dessas unidades permitem inferir a emoção aferida, incluindo microexpressões rápidas (≈40–200 ms) que revelam respostas emocionais involuntárias (Ekman, 2003).

De acordo com [BALTRUŠAITIS et al., 2016; 2018], o *OpenFace* fornece rastreamento facial em tempo real e detecção de AUs com desempenho dependente de pose, iluminação e base de dados. Sua confiabilidade é reduzida em chamadas de vídeo com iluminação e ângulos variáveis.

Por esse motivo, este estudo combina o *OpenFace* (extração objetiva de AUs) e o *DeepFace* (inferência probabilística de emoções), equilibrando aplicabilidade e abrangência analítica.

2.5 MediaPipe

O *MediaPipe* é um *framework* multiplataforma desenvolvido pelo Google para a construção de pipelines de processamento multimodal em tempo real, amplamente utilizado em aplicações de visão computacional. A ferramenta oferece soluções prontas para tarefas como detecção facial, estimativa de pose, rastreamento de mãos e extração de *landmarks* faciais, combinando eficiência computacional e baixa latência. No contexto da análise de expressões faciais, o *MediaPipe* permite a detecção robusta de pontos-chave do rosto mesmo em cenários com variações de iluminação e pose, sendo especialmente adequado para aplicações em vídeo e ambientes não controlados, como chamadas de teleconsulta [LUGARES et al., 2019; GRISHCHENKO et al., 2020].

2.6 Aplicações de Análise Facial na Telemedicina

A pandemia de COVID-19 acelerou a adoção da telemedicina, consolidando plataformas de videoconferência como parte do atendimento psicológico remoto. Nesse contexto, a comunicação não verbal permanece essencial, mas sofre limitações técnicas que dificultam a leitura emocional.

A integração de sistemas de análise facial oferece suporte promissor, permitindo registrar expressões e visualizar tendências emocionais ao longo das sessões. Ferramentas como *DeepFace* e *OpenFace* possibilitam:

- monitorar variações emocionais durante o atendimento;
- identificar momentos de tensão, ansiedade ou desânimo;
- gerar históricos visuais do comportamento emocional;
- acompanhar respostas expressivas a estímulos terapêuticos.

Somado a esses aspectos, o uso de plataformas digitais levanta questões de segurança, privacidade e proteção de dados sensíveis, exigindo protocolos rigorosos para garantir confidencialidade e conformidade com legislações como a LGPD. Nesse cenário, ferramentas computacionais de análise facial podem oferecer sinais complementares, desde que empregadas com responsabilidade ética e respaldo técnico.

Esses recursos podem enriquecer o planejamento terapêutico, desde que pautados por princípios éticos e legais. O tratamento de dados biométricos requer consentimento explícito, finalidade legítima e segurança da informação, conforme a Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018).

Além disso, é fundamental implementar auditorias de viés, avaliando consistência de desempenho entre grupos populacionais distintos idade, gênero, etnia e variações de iluminação [Rhue, 2019; TERHÖRST et al., 2022].

Observando o campo da teleconsulta a interpretação dos resultados deve ser sempre técnica e humana: os algoritmos indicam probabilidades e padrões, mas a atribuição de significado e contexto cabe exclusivamente ao profissional de saúde.

Ferramentas como *DeepFace* e *OpenFace* podem, assim, monitorar variações emocionais durante o atendimento, identificar momentos de tensão, ansiedade ou desânimo, gerar históricos visuais do comportamento emocional e acompanhar respostas expressivas a estímulos terapêuticos.

3. Trabalhos Correlatos

A literatura apresenta um conjunto de soluções consolidadas para reconhecimento de emoções faciais, com diferentes graus de maturidade tecnológica e escopo operacional. De forma geral, os estudos concentram-se em três linhas: (i) modelos de *deep learning* para classificação direta de emoções, (ii) ferramentas baseadas em *FACS* para análise muscular e microexpressões e (iii) sistemas híbridos voltados ao suporte clínico em contextos remotos.

Silva (2023) discute a importância da interpretação de expressões faciais e outros sinais comportamentais pelo profissional, destacando seu papel no acompanhamento clínico em saúde mental. Partindo dessa premissa, pesquisas em computação afetiva têm buscado desenvolver métodos computacionais capazes de auxiliar a análise desses sinais, oferecendo indicadores objetivos que complementem a avaliação humana, especialmente em contextos de teleatendimento. Neste trabalho,

adota-se a distinção conceitual entre reconhecimento de expressões/emoções (inferência de estados afetivos a partir de sinais observáveis, como padrões faciais) e análise de sentimentos (inferência de polaridade/valência a partir de linguagem e conteúdo textual).

Nesse sentido, além da análise facial em vídeos, há estudos que exploram abordagens multimodais ao combinar sinais visuais com informações linguísticas, extraídas de legendas ou transcrições do áudio do vídeo, para enriquecer a interpretação afetiva (DIAS; SAQUI; MOREIRA, 2024). Assim, evidências visuais e linguísticas podem atuar de forma complementar como suporte analítico em cenários de saúde mental, sem substituir o julgamento clínico. Os modelos baseados exclusivamente em aprendizado profundo, como *VGG-Face, ArcFace e FaceNet*, demonstram desempenho robusto na representação facial e são amplamente empregados como *backbones* para classificação emocional [Parkhi et al., 2015; Schroff et al., 2015; Deng et al., 2019].

Em paralelo, soluções fundamentadas no *Facial Action Coding System* (FACS), como o *OpenFace* [Baltrušaitis et al., 2016; 2018], priorizam a interpretabilidade da análise por meio de *Action Units.* Esses sistemas são recorrentes em pesquisas que demandam rastreabilidade muscular e análise temporal fina, executando detecção de microexpressões e padrões de ativação involuntária. Trabalhos exploratórios presentes na literatura nacional apresentam abordagens iniciais de identificação de emoções utilizando inteligência artificial, ainda que com foco predominantemente teórico e conceitual [Moura; Lopes, 2022]. Tais estudos evidenciam a ampliação do interesse acadêmico pelo tema, embora não empreguem *frameworks* modernos como *DeepFace* ou *OpenFace,* limitando a comparação direta com soluções operacionais.

Entre as abordagens híbridas mais recentes, destaca-se o estudo conduzido por Sugimori e Yamaguchi (2025), na Waseda University, que implementaram um sistema de detecção automática de microexpressões para apoiar avaliações psicológicas de estudantes. O modelo apresentado utiliza análise temporal de movimentos faciais sutis, como indicador complementar de estados emocionais, reforçando a viabilidade de soluções computacionais em contextos sensíveis de saúde mental. A pesquisa evidencia ganhos operacionais derivados da combinação de técnicas de visão computacional e métricas comportamentais, o que converge diretamente com o *pipeline* integrado adotado neste trabalho.

Comparações diretas entre diferentes *frameworks* apontam que sistemas exclusivamente baseados em *CNNs* apresentam maior robustez a ruído e variação de imagem, enquanto soluções fundamentadas em *FACS* entregam maior transparência dos resultados. Os modelos híbridos, portanto, tendem a apresentar melhor equilíbrio entre acurácia prática e interpretabilidade, especialmente em aplicações sensíveis como psicologia e telemedicina.

A solução proposta neste artigo posiciona-se exatamente nesse espaço híbrido, integrando *OpenFace* e *DeepFace* em um *pipeline* operacional orientado a uso clínico exploratório. Diferencia-se dos trabalhos correlatos ao oferecer arquitetura end-to-end com rastreamento facial, inferência emocional, persistência estruturada em tempo real e *dashboard* móvel, direcionado especificamente ao ambiente de teleconsulta psicológica.

4. Metodologia

Este estudo adota uma abordagem aplicada e exploratória, com foco na implementação e validação operacional de um sistema integrado para análise automática de emoções faciais em vídeos de teleconsulta psicológica. A metodologia prioriza a descrição clara do que foi desenvolvido, da arquitetura proposta e dos componentes que compõem o sistema, sem a realização de avaliações quantitativas formais de desempenho, como métricas de acurácia ou latência.

O modelo proposto foi implementado como uma aplicação móvel com *dashboard* analítico, integrada a um *pipeline* de processamento facial em tempo real. O sistema realiza desde a captura do vídeo até a visualização das emoções inferidas, caracterizando uma solução end-to-end voltada ao apoio à prática clínica.

4.1 Integração com a Plataforma ONTERAPIA

A solução apresentada neste trabalho foi integrada à plataforma ONTERAPIA, um projeto colaborativo voltado ao suporte de teleatendimentos psicológicos, com aplicativo móvel e *dashboard* para acompanhamento da sessão. No ONTERAPIA, este artigo concentra-se especificamente no módulo de análise facial, responsável por capturar frames do vídeo, realizar o alinhamento facial, extrair Action Units (*OpenFace*), inferir emoções básicas (*DeepFace*) e persistir os resultados para visualização. Este artigo descreve exclusivamente o módulo de análise facial da plataforma ONTERAPIA. As demais funcionalidades (por exemplo, transcrição, prontuário e camadas adicionais de segurança) são mencionadas apenas como contexto de integração, não constituindo o foco de implementação e avaliação deste estudo.”

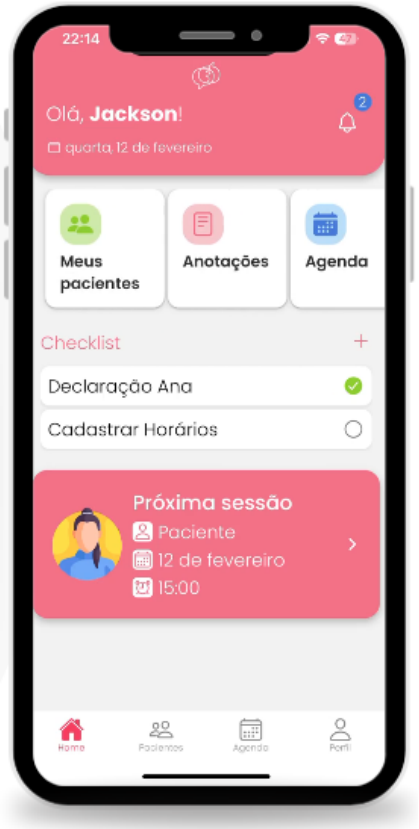


Figura 6 – Visão geral da plataforma ONTERAPIA.

4.2 Materiais

O ambiente experimental foi estruturado com base em ferramentas amplamente utilizadas em visão computacional e análise facial. O conjunto de materiais empregados inclui bibliotecas de *software*, infraestrutura de *hardware* e *frameworks* de visualização e persistência de dados. As Bibliotecas de Visão Computacional, OpenCV (*Open Source Computer Vision Library*), foram utilizadas para captura e manipulação de fluxos de vídeo, pré-processamento de quadros (*frames*) e integração com a *webcam*, e o *MediaPipe*, biblioteca desenvolvida pelo Google, foi responsável pela detecção facial, rastreamento de *landmarks* e alinhamento geométrico do rosto.

Essas ferramentas asseguraram a qualidade da entrada visual, corrigindo variações de iluminação e posicionamento. Para a Análise Facial, o *OpenFace* foi empregado para extração de *Action Units* (AUs) e métricas relacionadas ao *Facial Action Coding System* (FACS), fornecendo dados interpretáveis sobre intensidade e ativação muscular. O *DeepFace* foi utilizado para classificação probabilística das emoções básicas, integrando modelos pré-treinados (VGG-Face, *ArcFace*, *FaceNet* e DeepID) com bases de referência como FER-2013 e *AffectNet*. A Infraestrutura de Armazenamento utilizou o *Supabase*, plataforma *backend as a service* baseada em banco de dados relacional *PostgreSQL*, empregada para armazenar registros de emoções, *timestamps* e intensidades, com acesso seguro via API REST e autenticação JWT.

A interface de visualização foi desenvolvida com *React Native*, *framework* para desenvolvimento de aplicações móveis multiplataforma, implementando o *dashboard* interativo com visualização gráfica (pizza), com o auxílio de SVG e D3.js, bibliotecas auxiliares para renderização vetorial de alta precisão, integradas à camada de visualização do aplicativo. O *Hardware* utilizado consistiu em um computador pessoal com processador Intel Core i5, GPU dedicada NVIDIA GTX 1650, 8 GB de RAM e sistema operacional Windows 11, e uma *Webcam* HD (30 fps) utilizada para captura das expressões faciais durante as sessões simuladas. O conjunto de componentes operou de forma estável e responsiva, entregando processamento conforme os requisitos do experimento.

4.3 Métodos

O método adotado segue uma sequência lógica de etapas encadeadas, compondo um *pipeline* integrado que parte da aquisição de vídeo e termina na visualização analítica dos resultados.

Figura 7 - Metodologia

A figura 7 apresenta a metodologia do estudo, caracterizada como uma pesquisa aplicada, conduzida por meio de um estudo de caso técnico. O núcleo da metodologia é o desenvolvimento de um protótipo *end-to-end*, aplicado em um cenário de teleconsulta simulada, no qual são realizadas a análise de emoções e a validação do funcionamento do sistema. O processo contempla as etapas de desenvolvimento e integração do *pipeline* computacional, seguidas da validação operacional, com foco na viabilidade técnica da solução. Ressalta-se que o estudo prioriza testes funcionais, não contemplando análise estatística formal dos resultados.

O processo metodológico empregado para validar a viabilidade técnica do sistema de análise facial e demonstrar o potencial de integração das bibliotecas *DeepFace* e *OpenFace* iniciou-se com a Aquisição de Dados de Vídeo (Etapa 1). Nessa fase, a captura das imagens foi realizada por meio de webcam em um ambiente controlado, configurado para simular uma teleconsulta psicológica. Os vídeos foram gravados utilizando a configuração padrão da ferramenta de videoconferência, em qualidade suficiente para que os algoritmos de visão computacional identificassem adequadamente a região facial. Durante as gravações, foram solicitadas diferentes variações de expressões faciais, de modo a observar o comportamento do rastreamento e da análise automática em situações diversas.

Em seguida, na Detecção e Alinhamento Facial (Etapa 2), cada frame do vídeo foi processado utilizando *OpenCV* e *MediaPipe* para localizar e alinhar a face. Essa etapa é fundamental para padronizar a região facial, reduzindo ruídos visuais e diferenças de rotação ou inclinação. O *MediaPipe Face Mesh* foi empregado para fornecer até 468 *landmarks*, que serviram como base para um alinhamento geométrico de alta precisão.

A extração de *features* comportamentais (Etapa 3) ocorreu posteriormente, nessa etapa, as imagens alinhadas foram encaminhadas ao *OpenFace*, responsável por executar a análise muscular por meio das *Action Units* (AUs). Cada AU recebeu valores de intensidade e de ativação binária, possibilitando a identificação de microexpressões

movimentos faciais breves, com duração inferior a 200 milissegundos. Esses dados forneceram um mapa detalhado da atividade muscular facial, relevante para interpretações psicológicas.

Paralelamente, foi executada a classificação de emoções (Etapa 4). Os frames processados foram analisados pelo *DeepFace*, utilizando modelos pré-treinados como VGG-Face e FER-2013. O processamento resultou em um vetor probabilístico das seis emoções básicas, vinculado ao *timestamp* correspondente. Dessa forma, cada instante do vídeo recebeu uma estimativa emocional específica.

Os resultados das análises foram enviados para a persistência de dados em tempo real (Etapa 5), na qual foram armazenados no *Supabase*. Cada registro incluía o *timestamp*, a emoção predita, sua probabilidade, os valores das AUs e um identificador único da sessão (idSessão). A integridade e a segurança dos dados foram garantidas por autenticação via JWT e políticas de controle de acesso.

Por fim, na visualização interativa dos resultados (Etapa 6), os dados armazenados foram consumidos por um aplicativo móvel desenvolvido em *React Native.* A interface apresentou gráficos e elementos visuais que permitiram observar a dinâmica emocional ao longo da sessão, com destaque para um gráfico de pizza que sintetizou a distribuição percentual das emoções detectadas.

A adoção de um método *end-to-end* baseia-se em sua adequação a um contexto clínico exploratório, no qual a prioridade é a funcionalidade coerente e reprodutível, e não a mensuração estatística formal de desempenho.

Ao integrar bibliotecas *open source,* o sistema assegura transparência, baixo custo e replicabilidade, características essenciais para pesquisas acadêmicas em saúde. O uso do *Supabase* como camada de persistência permite futura expansão para análises, correlacionando séries temporais de emoções com contextos de fala ou discurso, fortalecendo o potencial clínico e analítico do modelo.

5. Resultados e Discussão

Esta seção apresenta os resultados obtidos com a execução do *pipeline* proposto, implementado como uma aplicação móvel com *dashboard* destinada ao apoio em teleconsultas psicológicas simuladas. Os resultados discutidos correspondem às informações exibidas na interface do aplicativo, que consolida os dados provenientes da captura de vídeo, da análise facial automática e da persistência estruturada. A análise concentra-se no comportamento técnico do sistema, no desempenho operacional do *pipeline* e na coerência entre as diferentes etapas da arquitetura, considerando desde a aquisição das imagens até a apresentação visual dos indicadores emocionais ao profissional.

5.1 Captura do vídeo

A captura do vídeo foi realizada em um ambiente de teleconsulta simulada utilizando a plataforma *Jitsi Meet* como interface principal entre psicólogo e participante. As sessões ocorreram em salas privadas, acessadas por link exclusivo, reproduzindo a dinâmica real de atendimentos remotos. O vídeo transmitido pelo *Jitsi* foi simultaneamente acessado por um *script* em *Python*, via *OpenCV*, que capturou o fluxo da *webcam* para alimentar o *pipeline* de análise facial.

Nesse fluxo, o participante acessa a sala virtual do Jitsi, momento em que o módulo local de captura é inicializado e passa a coletar o vídeo da webcam. Os frames capturados são encaminhados ao módulo de processamento, onde ocorre a análise facial. Após o processamento, os resultados gerados são então enviados ao backend para armazenamento e posterior utilização nas etapas de pós-processamento e visualização.

Figura 8 - Arquitetura e Comunicação.

A Figura 8 ilustra essa etapa, destacando a comunicação inicial entre o dispositivo do participante e o módulo de captura, bem como o fluxo subsequente para a API de processamento.

5.2 APIs de integração

A comunicação entre os componentes do sistema foi realizada por meio de um conjunto de *APIs REST* implementadas neste trabalho, organizadas sob a forma de um API *Gateway* local. Esse *gateway* foi responsável por orquestrar o fluxo de dados entre a camada de processamento (*Python*), a camada de armazenamento (*Supabase*) e o aplicativo móvel (*React Native)*, utilizando mensagens no formato *JSON* para garantir interoperabilidade entre tecnologias heterogêneas.

No *backend*, foram desenvolvidos endpoints específicos para: (i) receber os registros de análise facial gerados pelo *pipeline* (emoção predita, probabilidades associadas, *Action Units* e *timestamp*); (ii) registrar metadados das sessões de teleconsulta (identificador da sessão, duração e participante); e (iii) disponibilizar consultas agregadas para consumo pelo *dashboard* móvel, como a distribuição percentual das emoções ao longo da sessão. Essa camada intermediária desacopla a lógica de visão computacional das camadas de persistência e visualização, favorecendo modularidade e manutenção do sistema.

Nos experimentos realizados, a arquitetura de integração implementada mostrou-se adequada para operar em tempo quase real. A latência observada entre a geração de um registro emocional no *backend* e sua disponibilização no aplicativo móvel manteve-se na ordem de centenas de milissegundos, o que viabiliza o uso do sistema em contextos clínicos exploratórios nos quais a atualização contínua dos indicadores é desejável, ainda que não estritamente em tempo real.

5.3 Processamento

O módulo de processamento concentra as etapas de visão computacional e inferência emocional. Cada frame recebido da etapa de captura é inicialmente tratado pelo *OpenCV*, que realiza conversão de formato, normalização e ajuste de brilho/contraste quando necessário. Em seguida, o frame é enviado ao *MediaPipe Face Mesh,* responsável por detectar a face e localizar até 468 *landmarks*, permitindo o alinhamento

geométrico da região facial.

As imagens alinhadas são então encaminhadas a dois submódulos complementares. O primeiro, baseado no *OpenFace*, extrai as *Action Units* (AUs) do FACS, produzindo, para cada quadro, um vetor de intensidades e *flags* binárias de ativação muscular. O segundo, baseado no *DeepFace*, aplica um *backbone* de *deep learning* pré-treinado (VGG-Face em combinação com FER-2013), gerando um vetor de probabilidades para as seis emoções básicas. Ambos os resultados são associados ao mesmo *timestamp*, permitindo correlação direta entre camada muscular (AUs) e camada emocional probabilística.

Os resultados foram obtidos a partir de testes experimentais controlados, realizados em ambiente local, com captura contínua de vídeo em tempo real. A avaliação consistiu em medições de desempenho computacional (fps) do *pipeline* de visão computacional e da inferência emocional. A detecção facial e o rastreamento de *landmarks* operaram em média a 30 fps, enquanto a inferência emocional manteve entre 10 e 12 fps, valor compatível com uso em teleconsulta síncrona.

Embora os módulos de detecção facial (*MediaPipe*) e inferência emocional (DeepFace) operem, em testes contínuos, a aproximadamente 30 fps e 10–12 fps, respectivamente, o experimento clínico simulou uma taxa de amostragem de 1 frame por segundo, adotada por decisão de projeto. Essa escolha visou reduzir custo computacional e volume de armazenamento, mantendo rastreabilidade temporal suficiente para análise.

Tabela 1 – Métricas operacionais do pipeline de análise facial em teleconsulta simulada.

Métrica	Valor	Descrição
Duração do vídeo analisado	17 minutos	Tempo total da sessão de teleconsulta psicológica simulada.
Número total de frames processados	1.020 frames	Total de quadros efetivamente analisados pelo sistema.
Tempo total de processamento da sessão	574,60 segundos (≈ 9,57 minutos)	Tempo computacional necessário para processar todos os frames capturados.
Tempo médio de processamento por frame	≈ 0,56 s/frame	Relação entre o tempo total de processamento e o número de frames (574,6 ÷ 1.020).
Relação processamento / tempo real	≈ 0,56× o tempo real	O pipeline processa o vídeo em pouco mais da metade do tempo da sessão original.

Taxa média de detecção e rastreamento (MediaPipe)	≈ 30 fps	Desempenho do módulo de detecção de face e landmarks em condições experimentais controladas.
Taxa média de inferência emocional (DeepFace)	≈ 10–12 fps	Frequência com que as emoções são inferidas a partir dos frames alinhados.
Latência backend → dashboard móvel	Ordem de centenas de milissegundos	Intervalo entre a geração do registro de emoção e sua disponibilização no aplicativo (quase em tempo real).
Gargalos observados no armazenamento (Supabase)	Não observados	Não foram identificados atrasos relevantes de escrita/leitura para sessões com ≈ 1.000 registros.

Além da análise do comportamento dos módulos de inferência, foram realizadas medições quantitativas operacionais relacionadas ao tempo de processamento do *pipeline*. Considerando uma sessão de teleconsulta simulada com duração de 17 minutos e uma taxa de amostragem definida em 1 frame por segundo, o sistema analisou 1.020 frames em um tempo total de 574,60 segundos (≈ 9,57 minutos). Isso corresponde a um tempo médio de processamento de aproximadamente 0,56 segundos por frame.

Esses valores refletem as condições específicas do experimento realizado, incluindo a taxa de amostragem adotada, o ambiente computacional e o caráter offline da análise. Dessa forma, os resultados permitem indicar a viabilidade operacional do protótipo no cenário experimental avaliado, sem permitir generalizações sobre desempenho em outros contextos ou configurações. No escopo deste estudo, o *pipeline* mostrou-se adequado para uso exploratório e assíncrono como ferramenta complementar em sessões clínicas simuladas.

5.4 Armazenamento

Os dados gerados pelo processamento são persistidos no *Supabase*, que oferece uma camada de banco de dados relacional baseada em *PostgreSQL*, acessível via APIs REST com autenticação JWT. Para cada sessão, são criadas entradas em tabelas lógicas que armazenam: (i) metadados da sessão (idSessão, data, duração, identificador do participante); (ii) registros de emoções (*timestamp*, emoção predominante, vetor de probabilidades); e (iii) registros de AUs (*timestamp*, intensidades e ativações de cada unidade de ação).

Essa estrutura possibilita tanto consultas pontuais, como a recuperação da série temporal completa de emoções de uma sessão específica, quanto análises agregadas, como a proporção de cada emoção ao longo da sessão e a identificação de picos de ativação muscular em intervalos definidos. No experimento realizado, uma sessão de aproximadamente 17 minutos, com taxa de amostragem de 1 frame por segundo, gerou 1.020 registros de análise emocional, além dos registros correspondentes às Action

Units. Esse volume foi persistido sem atrasos perceptíveis de escrita ou leitura durante os testes, indicando adequação do *Supabase* à frequência e ao volume de dados produzidos pelo *pipeline* no cenário avaliado.

Do ponto de vista de segurança e privacidade, o presente trabalho implementa controle de acesso na camada de aplicação, por meio de autenticação baseada em JWT e separação lógica das tabelas. Entretanto, mecanismos de proteção em nível de SGBD, como criptografia em repouso, controle direto de permissões ou auditoria de acesso ao banco, não foram tratados neste protótipo. Essas preocupações são abordadas de forma mais ampla no contexto da plataforma OnTerapia, que contempla estratégias específicas para armazenamento seguro, governança de dados clínicos e conformidade com a LGPD. Assim, no escopo deste estudo, a persistência de dados deve ser compreendida como parte de uma solução experimental, cuja adoção em ambiente real exige a integração com camadas adicionais de segurança e gestão de dados sensíveis.

5.5 Apresentação de Resultados

Na etapa de apresentação de resultados, os dados processados ao longo do *pipeline* são organizados e exibidos ao profissional de saúde mental. É nesse momento que as informações técnicas provenientes da detecção facial, da extração das *Action Units* e da classificação emocional são convertidas em indicadores visuais capazes de apoiar a interpretação clínica.

A Figura 9 representa o fluxo operacional completo do sistema, mostrando, de maneira sequencial, como o vídeo bruto capturado pela webcam passa pelas fases de pré-processamento, extração de características faciais, classificação das emoções e persistência transacional até chegar à visualização final no dispositivo móvel. Cada bloco destacado na figura sintetiza uma etapa funcional da arquitetura: a captura inicial conduzida por *OpenCV* e *MediaPipe*, a extração de *features* pelo *OpenFace*, a classificação emocional pelo *DeepFace* utilizando o modelo VGG-Face/FER-2013, o armazenamento estruturado no *Supabase* e, por fim, a entrega dos resultados ao

Complementando essa visão sistêmica, a Figura 09 apresenta o *dashboard* de visualização, interface destinada ao profissional que acompanha a sessão. Nela, os resultados emocionais aparecem organizados de forma intuitiva, com destaque para o gráfico que exibe a distribuição geral das emoções detectadas ao longo do vídeo, além das informações essenciais da sessão, como nome do paciente, data, arquivo analisado e duração da gravação. A interface permite que o psicólogo observe tendências emocionais, identifique momentos de maior ativação e registre anotações clínicas diretamente na tela. Essa camada final traduz o processamento técnico do *pipeline* em elementos compreensíveis e úteis para o acompanhamento terapêutico, reforçando o caráter complementar da ferramenta no contexto da telepsicologia.

Figura 9 - Dashboard de visualização.

O *dashboard* apresentou funcionamento adequado na apresentação dos dados processados, permitindo ao psicólogo visualizar a distribuição e a variação temporal das emoções ao longo da sessão, conforme ilustrado na Figura 09. Por meio dos gráficos exibidos, foi possível observar flutuações emocionais, identificar momentos de maior intensidade e reconhecer padrões expressivos recorrentes ao longo do atendimento.

A análise foi conduzida a partir da comparação visual entre os picos de probabilidade emocional exibidos no *dashboard* e os valores de ativação das *Action Units (AUs)* extraídas pelo *OpenFace* para os mesmos instantes temporais. Observou-se que, em episódios de maior intensidade emocional, houve correspondência consistente entre a elevação das probabilidades estimadas pelo *DeepFace* e a ativação das AUs esperadas para cada emoção.

Apesar dessas limitações, o *pipeline* como um todo demonstrou fluidez, modularidade e robustez suficiente para apoiar teleconsultas psicológicas, fornecendo indicadores visuais que enriquecem a leitura emocional sem substituir o julgamento clínico do profissional.

6. Considerações Finais

O estudo demonstrou a viabilidade operacional de um *pipeline* integrado para análise automatizada de emoções básicas em vídeos de teleconsulta psicológica. A arquitetura proposta consolidou, em fluxo contínuo, os módulos de captura (*OpenCV*), detecção e

alinhamento facial (*MediaPipe*), extração de *Action Units* (*OpenFace*), classificação probabilística de emoções (*DeepFace*), persistência estruturada (*Supabase*) e visualização analítica (*React Native*).

Os resultados evidenciaram coerência entre os picos emocionais mais intensos e os padrões de AUs associados, indicando complementaridade entre as camadas muscular e probabilística. A solução operou de forma responsiva, mantendo rastreabilidade por *timestamp* e integridade dos registros. O *pipeline* demonstrou alinhamento aos requisitos de um ambiente clínico exploratório, oferecendo insumo visual e analítico que pode apoiar a revisão temporal das sessões.

Persistem limitações relevantes. A ausência de um *ground truth* sincronizado impede mensuração formal de acurácia, F1 ou calibração. O sistema mostrou sensibilidade às condições do ambiente de captura, como iluminação, pose e oclusões, além de suscetibilidade a *dataset shift* em relação às bases de treinamento (FER-2013, AffectNet). Esses fatores reforçam o caráter não conclusivo e não diagnóstico da solução.

Em síntese, o modelo entrega valor como ferramenta complementar ao trabalho do psicólogo, ampliando a visibilidade sobre variações expressivas durante a teleconsulta. A adoção prática requer validação adicional, auditoria de vieses, padronização de protocolos e análise ética contínua. Trabalhos futuros podem incluir calibração individual, comparação sistemática com outros *frameworks* de análise facial e integração com camadas multimodais, como voz e conteúdo verbal.

7. Referências

BALTRUŠAITIS, T.; ROBINSON, P.; MORENCY, L.-P. OpenFace: An Open Source Facial Behavior Analysis Toolkit. In: IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

BALTRUŠAITIS, T.; ZADEH, A.; LIM, Y. C.; MORENCY, L.-P. OpenFace 2.0: Facial Behavior Analysis Toolkit. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2018.

BARRETT, Lisa Feldman et al. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest, v. 20, n. 1, p. 1–68, 2019.Disponível em: https://journals.sagepub.com/doi/10.1177/1529100619832930. Acesso em: 12 out. 2025.

COHN, J. F.; DE LA TORRE, F. Automated Face Analysis for Affective Computing. In: CALVO, R.; D'MELLO, S. (Eds.). The Oxford Handbook of Affective Computing. Oxford: Oxford University Press, 2015.

DENG, J. et al. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

EKMAN, P.; FRIESEN, W. V. Facial Action Coding System (FACS): Manual. Palo Alto, CA: Consulting Psychologists Press, 1978.

EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, [s. l.], v. 17, n. 2, p. 124–129, 1971.

EKMAN, P. Basic Emotions. In: DALGLEISH, T.; POWER, M. (Eds.). Handbook of Cognition and Emotion. Chichester, UK: Wiley, 1999.

EKMAN, P. Emotions Revealed. New York: Times Books, 2003.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge, MA: MIT Press, 2016.

LI, S.; DENG, W. Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, [s. l.], v. 13, n. 3, p. 1195–1215, 2020.

LUXTON, D. D. et al. mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare. Professional Psychology: Research and Practice, [s. l.], v. 42, n. 6, p. 505–512, 2011.

MOLLAHOSSEINI, A.; HASANI, B.; MAHOOR, M. AffectNet: A Database for Facial Expression, Valence, and Arousal in the Wild. IEEE Transactions on Affective Computing, [s. l.], v. 10, n. 4, p. 483–494, 2019.

PARKHI, O.; VEDALDI, A.; ZISSERMAN, A. Deep Face Recognition. In: British Machine Vision Conference (BMVC), 2015.

RHUE, L. Racial Influence on Automated Perceptions of Emotions. Information Systems Research, [s. l.], v. 30, n. 3, p. 696–713, 2019.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

SHORE, J. H.; SCHNECK, C. D.; MISHKIND, M. Telepsychiatry and the Coronavirus Disease 2019 Pandemic—Current and Future Outcomes of the Rapid Virtualization of Psychiatric Care. JAMA Psychiatry, v. 77, n. 12, p. 1211–1212, 2020. Disponível em: https://www.researchgate.net/publication/341926848 .Acesso em: 12 out. 2025.

SUN, Y.; WANG, X.; TANG, X. DeepID: Deep Learning Face Representation by Joint Identification-Verification. In: Neural Information Processing Systems (NIPS), 2014.

TAIGMAN, Y. et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

TERHÖRST, Philipp et al. A comprehensive study on face recognition biases beyond demographics. IEEE Transactions on Technology and Society, v. 3, n. 1, p. 16–32, 2022. Disponível em: https://www.researchgate.net/publication/357317385 .Acesso em: 12 out. 2025.

ZENG, Z. et al. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, [s. l.], v. 31, n. 5, p. 771–807, 2009.

SUGIMORI, K.; YAMAGUCHI, M. AI-based micro-expression detection for mental-health assessment in university students. Relato de pesquisa divulgado por Waseda University News, 2024/2025. Disponível em: https://www.waseda.jp/top/en/news/85707 . Acesso em: 18 nov. 2025.

GOOGLE. MediaPipe Face Mesh – Technical Documentation. MediaPipe Documentation, [s. l.], [2025?]. Disponível em: https://developers.google.com/mediapipe. Acesso em: 12 out. 2025.

INTEL. OpenCV Documentation. OpenCV Documentation, [s. l.], [2025?]. Disponível em: https://docs.opencv.org. Acesso em: 12 out. 2025.

META. React Native – Getting Started. React Native Documentation, [s. l.], [2025?]. Disponível em: https://reactnative.dev/docs/getting-started. Acesso em: 12 out. 2025.

SERENGIL, S.; ÖZPINAR, A. DeepFace – GitHub Repository. GitHub, 2025. Disponível em: https://github.com/serengil/deepface. Acesso em: 12 out. 2025.

SUPABASE. Supabase Docs. Supabase Documentation, [s. l.], [2025?]. Disponível em: https://supabase.com/docs. Acesso em: 12 out. 2025.

PEREIRA, Gabriel Gonçalves; NUNES, Luiz Fernando. Análise comparativa de algoritmos de reconhecimento facial: FaceNet e DeepFace na detecção de emoções em aplicações clínicas. Revista Terra & Cultura: Cadernos de Ensino e Pesquisa, Londrina, v. 41, n. especial, p. 426-450, 2025.

SILVA, Leonardo de Oliveira da. Detecção e reconhecimento de faces para a anonimização de pessoas em vídeos. 2023. 102 f. Trabalho de Conclusão de Curso

(Graduação em Sistemas de Informação) - Universidade Federal de Santa Catarina, Florianópolis, 2023.

MOURA, A.; LOPES, M. Estudo da Identificação de Emoções através da Inteligência Artificial. 2022. Disponível em: https://multivix.edu.br/wp-content/uploads/2018/08/estudo-da-identificacao-de-emocoes-atraves-da-inteligencia-artificial.pdf

DIAS, Jean Ferreira; SAQUI, Diego; MOREIRA, Heber Rocha. Deep Learning para Classificação de Sentimento em Vídeos utilizando legendas. 2024. Disponível:https://josif.ifsuldeminas.edu.br/ojs/index.php/anais/article/view/2349/1881

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: Brasília, DF, 15 ago. 2018.

LUGARESI, C.; TANG, J.; NASH, H.; *et al.* MediaPipe: A Framework for Building Perception Pipelines. *arXiv preprint* arXiv:1906.08172, 2019.