

Reconhecimento de Emoções em Cenários Competitivos de Jogos Usando Rede Neural Convolucional e Raspberry Pi

Eduardo Zaparoli¹, Rafael Rieder¹

¹Instituto de Tecnologia (ITec) – Curso de Ciência da Computação
Universidade de Passo Fundo (UPF) – Passo Fundo, RS, Brasil

{183132, rieder}@upf.br

Abstract. *Emotion recognition from facial expressions has gained prominence in competitive sports, where emotional monitoring can provide insights into players' performance and mental state. This study aims to embed a convolutional neural network on a Raspberry Pi device to perform real-time emotion recognition in players' facial expressions, offering insights into their emotional state without interfering with game performance. The research focuses on analyzing the performance of the approach, which considers the adaptation of a predefined model, trained with the FER-2013 dataset, and running on low-cost hardware. The work explores, as a case study, the feasibility of applying emotion recognition to a first-person shooter game. The results show that the emotion recognition system predominantly detected neutral expressions, with variations for happiness, anger, and sadness. Despite this, users reported feeling neutral or happy before and during the game rounds and happy and sad after the match - indicating that the system detected emotions not perceived by the players. Regarding the embedded approach, the application performed efficiently on the Raspberry Pi, performing real-time emotion processing and recognition despite the hardware processing limitations, demonstrating the feasibility of using intelligent models in low-cost devices.*

Keywords: *convolutional neural network, emotion recognition, game, Raspberry Pi.*

Resumo. *O reconhecimento de emoções nas expressões faciais tem ganhado destaque em áreas como esportes competitivos, onde o monitoramento emocional pode fornecer percepções sobre o desempenho e estado mental dos jogadores. Este estudo tem como objetivo embarcar uma rede neural convolucional em um dispositivo Raspberry Pi para realizar o reconhecimento de emoções em expressões faciais oferecendo insights sobre o estado emocional dos jogadores, em tempo real, sem interferir no desempenho do jogo. A pesquisa foca na análise do desempenho da abordagem, que considera a adaptação de um modelo pré-definido, treinado com o dataset FER-2013, e executando em hardware de baixo custo. O trabalho explora como estudo de caso a viabilidade da aplicação de reconhecimento emocional em um jogo de tiro em primeira pessoa. Os resultados mostram que o sistema de reconhecimento de emoções detectou predominantemente expressões neutras, com variações para felicidade, raiva e tristeza. Apesar disso, os usuários relataram sentir-se neutros ou alegres antes e durante as rodadas do jogo, e*

alegres e tristes após a partida - o que indica que o sistema detectou emoções não percebidas pelos jogadores. Em relação à abordagem embarcada, a aplicação se comportou de forma eficiente no Raspberry Pi, realizando o processamento e reconhecimento de emoções em tempo real, apesar das limitações de processamento do hardware, mostrando viabilidade de usar modelos inteligentes em dispositivos de baixo custo.

Palavras-chave: *jogo, reconhecimento de emoções, Raspberry Pi, rede neural convolucional.*

1. Introdução

As descobertas científicas indicam que as emoções desempenham um papel essencial na tomada de decisão racional, percepção, aprendizado e uma variedade de outras funções cognitivas. As emoções não se limitam a arte, entretenimento e interação social, elas influenciam os próprios mecanismos do pensamento racional. As pessoas em geral sabem, por experiência, que as emoções podem prejudicar a tomada de decisão [Picard 1997].

O reconhecimento de emoções nas expressões faciais desempenha um papel crucial em vários campos, como segurança, interação humano-computador e avaliação de saúde mental. Recentemente, redes neurais convolucionais (CNNs) têm sido aplicadas para o reconhecimento automatizado de emoções em expressões faciais [Minaee et al. 2021]. Emoção poderia ser definida como uma condição complexa e momentânea que surge em experiências de caráter afetivo, provocando alterações em várias áreas do funcionamento psicológico e fisiológico, preparando o indivíduo para a ação [Atkinson et al. 2002, Davis & Lang 2003, Frijda 2008].

No entanto, a maior parte das aplicações ocorre em ambientes de alta capacidade computacional, o que dificulta a implementação em dispositivos embarcados e de baixo custo, como o Raspberry Pi. A portabilidade de sistemas de reconhecimento de emoções para esses dispositivos pode abrir novas possibilidades em áreas como esportes eletrônicos, onde é necessário um monitoramento contínuo e em tempo real. A capacidade de executar CNNs em hardware limitado apresenta desafios significativos, tanto em termos de otimização de desempenho quanto de eficiência energética [Capra et al. 2020].

O uso de um Raspberry Pi para monitorar indivíduos em jogos, ao invés de notebooks com câmera integrada, pode apresentar vantagens significativas. Primeiramente, em ambientes competitivos, a carga de processamento gráfico é elevada, pois os jogos modernos têm consumo alto de GPU [Kattakinda et al. 2024]. A utilização de um dispositivo externo, como o Raspberry Pi, para realizar o reconhecimento de emoções pode aliviar essa carga, evitando interferências no desempenho do jogo. Além disso, a separação dos sistemas permite que a câmera do jogador continue disponível para o uso em plataformas de streaming ou comunicação com a equipe. Essa abordagem também possibilita que a análise emocional seja transmitida em tempo real para o público ou utilizada pela equipe técnica do jogador para identificar momentos de estresse e dificuldades durante a partida, contribuindo para melhorar estratégias de treinamento [Beres et al. 2023].

As reações a determinadas situações variam conforme o contexto social em que o indivíduo está envolto. Porém, algumas expressões faciais podem ser consideradas como sendo universais [Ekman 2003]. Apesar das variações culturais, existem algumas

emoções básicas que são amplamente reconhecidas e interpretadas de forma semelhante em diferentes sociedades. A partir dessas “emoções básicas”, pode-se inferir estados emocionais como frustração, engajamento ou desinteresse, o que é particularmente útil para entender o comportamento dos usuários durante o uso de aplicações computacionais. Esse monitoramento pode ajudar na adaptação dinâmica da experiência, no aprimoramento de interfaces e na personalização de conteúdo, contribuindo para a usabilidade e a satisfação do usuário [Alipour et al. 2023].

Em jogos eletrônicos, a análise das expressões faciais dos jogadores pode fornecer *insights* sobre seu nível de estresse ou engajamento, permitindo intervenções para otimizar a experiência de jogo [Mandryk & Atkins 2007]. Fora do contexto competitivo, o reconhecimento emocional é utilizado em ambientes educacionais para avaliar o engajamento dos alunos e adaptar métodos de ensino [Picard 1997].

Com isso em mente, o objetivo deste trabalho é embarcar uma CNN em um dispositivo Raspberry Pi para realizar o reconhecimento de emoções em expressões faciais oferecendo *insights* sobre o estado emocional dos jogadores, em tempo real, sem interferir no desempenho do jogo. A pesquisa foca na análise do desempenho da abordagem, que considera a adaptação de um modelo pré-definido, treinado com o dataset FER-2013, e executando em hardware de baixo custo. O estudo faz uma avaliação preliminar mediante um estudo de caso para reconhecimento emocional durante a interação em um game.

Para tanto, o trabalho está organizado como segue: a Seção 2 mostra trabalhos relacionados a respeito da temática do estudo; a Seção 3 apresenta os materiais e métodos utilizados para concepção da abordagem, bem como detalha o desenho do estudo; a Seção 4 divulga os resultados obtidos com uma avaliação inicial, e discute vantagens e limitações do estudo; e a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

O estudo de [Garcia 2021] investiga o uso de CNNs para reconhecimento de emoções faciais (FER) em dispositivos com recursos limitados, utilizando um Raspberry Pi. Os autores comparam seis arquiteturas de código aberto, avaliando desempenho com e sem Data Augmentation, e implementa a arquitetura mais eficiente com técnicas de poda e quantização, atingindo 68,5% de acerto com 398.000 parâmetros e latência de 168 ms. O trabalho demonstra a viabilidade de sistemas FER em hardware embarcado para aplicações como monitoramento de motoristas e interatividade em jogos.

[Aranha 2022] propõem uma abordagem inovadora para o desenvolvimento de jogos sérios adaptativos, considerando as emoções e traços de personalidade dos jogadores. A pesquisa apresenta um framework que adapta automaticamente os elementos do jogo em tempo real, utilizando o reconhecimento de emoções a partir de expressões faciais e traços de personalidade identificados por meio de um inventário preenchido pelos jogadores. Experimentos foram realizados para validar a eficácia da abordagem na melhoria da experiência dos jogadores, especialmente para aqueles com traços de personalidade específicos. Os resultados indicam que a adaptação automática pode oferecer uma experiência mais personalizada e eficaz, destacando-se como uma contribuição significativa para o campo de jogos sérios e computação afetiva.

[Jofilsan et al. 2021] apresentam um modelo para a mineração de dados das expressões faciais de jogadores durante o uso de jogos digitais, identificando padrões emo-

cionais e agrupando os jogadores com base nesses padrões. Por meio de recursos de visão computacional, o estudo coletou e analisou dados de 232 jogadores, categorizando as expressões faciais em seis emoções básicas: felicidade, tristeza, medo, surpresa, raiva e nojo. A partir desses dados, foram formados clusters que permitiram a identificação de diferentes perfis de jogadores e suas reações emocionais durante momentos de tédio e estresse no jogo. O sistema proposto demonstrou ser eficiente na detecção de micro-expressões de forma não intrusiva, contribuindo para uma melhor compreensão das experiências emocionais dos jogadores e sugerindo aplicações futuras no desenvolvimento de jogos personalizados.

O trabalho de [Ab Wahab et al. 2021] explora a implementação de um sistema de reconhecimento de expressões faciais em tempo real utilizando um Raspberry Pi 4. O modelo híbrido combina a extração de características de uma CNN com um classificador KNN para melhorar a precisão do reconhecimento de emoções. O modelo foi treinado com o dataset FER-2013 e demonstrou uma precisão de 75.26. Comparável a modelos de ponta, sendo otimizado para dispositivos embarcados com processamento limitado e uso em tempo real.

Já o trabalho de [Guglielmo et al. 2022] explora o uso de unidades de ação facial (AUs) para distinguir os níveis de expertise de jogadores em competições do jogo Hearthstone. Com base em dados de expressões faciais durante o torneio, os autores utilizaram um classificador Random Forest para identificar padrões em três AUs principais, AU17 (elevação do queixo), AU25 (lábios afastados) e AU26 (mandíbulas queda) que podiam variar conforme a expertise do jogador. O estudo demonstrou que essas AUs podem prever com precisão a proficiência dos jogadores, com o classificador atingindo precisão de 0,75 na validação cruzada, após balanceamento dos pesos da classe e 0,85 após ter aplicado a técnica de sobreamostragem minoritária sintética (SMOTE).

Com base nessas abordagens, pode-se destacar que o diferencial de nossa proposta é o uso de uma CNN para o reconhecimento de emoções em expressões faciais de jogadores em tempo real, embarcada em um Raspberry Pi 4. Embora estudos anteriores tenham demonstrado a viabilidade do uso de CNNs em dispositivos de baixo custo, nosso trabalho foca especificamente no cenário de jogos, combinando uma análise emocional com condições reais dos jogos e avaliando o impacto que as emoções dos jogadores tem no desempenho das partidas através de métricas e gráficos, além da otimização do modelo para a execução eficiente em um hardware limitado.

3. Materiais e Métodos

Neste trabalho foi utilizado um conjunto de tecnologias essenciais para o desenvolvimento e execução do modelo de reconhecimento de emoções baseado em CNN, embarcada em um dispositivo Raspberry Pi. A seguir são detalhadas as tecnologias, métodos e seus papéis no projeto.

3.1. Raspberry Pi

Em relação ao embarcado, optou-se por utilizar um Raspberry Pi 4 Model B¹, um dispositivo de placa única equipado com um processador quad-core Cortex-A72 de 64 bits,

¹O modelo utilizado pode ser visualizado em <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/>.

1,5 GHz, com 4 GB de RAM. O Raspberry é um dispositivo popular na computação embarcada devido ao seu baixo custo, sendo amplamente utilizado em sistemas de visão computacional e reconhecimento de imagens [de Oliveira 2017].

A configuração do Raspberry incluiu o uso de um sistema operacional baseado em Linux. Para este estudo, procedeu-se com a instalação da distribuição Raspberry Pi OS (64-bit), por meio da ferramenta Raspberry Pi Imager.

3.2. CNN

Para a geração do modelo inteligente de detecção de emoções, a partir da análise de faces, optou-se por utilizar uma CNN pré-definida [Lx 2019], desenvolvida na linguagem Python - amplamente empregada em soluções de aprendizado de máquina. A rede proposta é composta por diferentes camadas convolucionais, como forma de detectar características faciais específicas. A Figura 1 apresenta a arquitetura da rede utilizada.

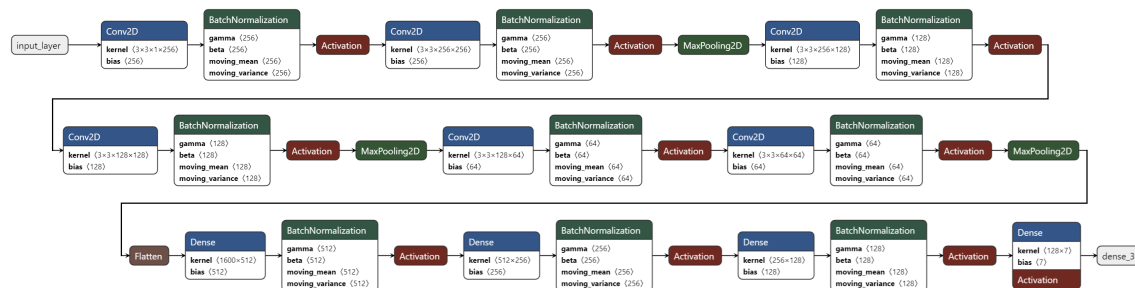


Figura 1. Estrutura da CNN utilizada para gerar o modelo inteligente de detecção de emoções.

O modelo de CNN utiliza recursos da biblioteca Keras, que fornece uma interface de alto nível para construção e treinamento de redes neurais. O modelo foi composto por várias camadas convolucionais, camadas de *pooling*, normalização em lotes e camadas densamente conectadas. Foi usada a função de ativação ReLU em todas as camadas intermediárias, para introduzir não-linearidade ao modelo. A função de ativação Softmax foi aplicada na camada de saída para classificação das emoções em sete categorias (conforme *dataset* utilizado). O modelo foi treinado considerando Entropia Cruzada Categórica como função de perda, e otimizador Adam com taxa de aprendizagem de 0,001.

Para prevenir sobreajuste (*overfitting*) do modelo, foi empregada a técnica de parada antecipada (*early stopping*) para interromper o treinamento quando o desempenho do modelo não melhorava após várias épocas, garantindo que o modelo só treinasse o necessário. Para tanto, definiu-se o número de 50 épocas de treinamento, com 10 épocas de paciência.

3.3. Dataset

A CNN foi treinada, validada e testada com imagens do *dataset* FER-2013², um conjunto de dados fornecido pela Kaggle, apresentado na Conferência Internacional sobre Aprendizado de Máquina em 2013, e comumente utilizado para treinar modelos de reconhecimento de emoções. Este conjunto de dados contém 35.887 imagens em escala de cinza, categorizadas em sete emoções: “Angry”, “Sad”, “Fear”, “Happy”, “Disgust”, “Surprise” e “Neutral”. A Figura 2 ilustra exemplos de cada uma das classes.

²Disponível em <https://www.kaggle.com/datasets/msambare/fer2013>



Figura 2. Classes disponíveis no *dataset* FER-2013.

O *dataset* foi dividido em três partes, na proporção 80/10/10. Adicionalmente, aplicou-se aumento de dados nas imagens utilizadas no treinamento, para melhorar a diversidade de dados e a robustez do modelo. Nesse contexto, utilizou-se transformações de rotação (10° de variação), deslocamento (10% das dimensões da imagem) e inversão horizontal (espelhamento), aplicados ou não, de maneira aleatória.

3.4. Reconhecimento da Face

Para aquisição das imagens faciais em tempo real, utilizou-se uma câmera modelo Webbookers WB 1080p, conectada no Raspberry Pi, e distante aproximadamente 80cm da face do usuário. A captura considerou somente o uso de iluminação artificial em um ambiente convencional de trabalho (sala), com lâmpadas fluorescentes tubulares de luz branca-fria.

O modelo CNN treinado foi exportado e carregado no Raspberry Pi. O código implementado na plataforma embarcada realizava a análise de faces a cada cinco frames, como forma de otimizar a inferência ao modelo de detecção de emoções. Para melhorar mais o desempenho, optou-se pelo uso combinado de um modelo para detecção primária de faces na imagem da câmera, recortando a face e submetendo somente a área de interesse ao modelo de emoções.

Para tanto, foi escolhido um modelo de detecção de faces leve, pré-treinado e de baixo custo computacional. O modelo e pesos de rede neural profunda escolhidos são baseados no método Single Shot-Multibox Detector (SSD), disponibilizados pelo módulo DNN da OpenCV (res10_300x300_ssd_iter_140000.caffemodel). Ele possui a arquitetura ResNet-10 [Gong et al. 2022], consagrada para tarefas de visão computacional.

Após inferência ao modelo de detecção de faces, utilizou-se recursos da biblioteca OpenCV³ para recorte do objeto de interesse na imagem. Em seguida, aplicou-se conversão para escala de cinza, normalização e redimensionamento da imagem resultante (48x48), para a entrada de dados ficar adequada ao modelo de detecção de emoções.

Depois da inferência a este segundo modelo, foram armazenadas em disco apenas as imagens e o momento do tempo que emoções “Angry”, “Sad”, “Fear”, “Happy”, “Disgust” e “Surprise” são detectadas, para fins posteriores de análise dos sentimentos do jogador durante o processo interativo. As capturas de face eram somente confirmadas e salvas quando a detecção apontava para um nível de confiança maior que 0.5.

Para fins de otimização, não foram feitos registros em imagem de emoções do tipo “Neutral”, uma vez que é a expressão mantida na maior parte do tempo. Por outro lado, as mesmas foram apenas contabilizadas, juntamente com as demais, salvas por meio de gráficos de contagem total e de registro temporal de emoções detectadas ao longo do

³Disponível em <https://opencv.org>

processo interativo.

Estas e as demais emoções são desenhadas na imagem a ser exibida na tela, com recursos da OpenCV para destacar o rosto da imagem e escrever a descrição da emoção.

3.5. Desenho do Estudo

Para investigar o reconhecimento das emoções, optou-se em monitorar o processo interativo de usuários durante sua experiência em um jogo de computador. Para tanto, este estudo selecionou o jogo Valorant⁴, um jogo de tiro em primeira pessoa desenvolvido pela Riot Games, caracterizado pela sua dinâmica tática e intensiva.

A escolha do jogo Valorant para este estudo foi por conta de ser um jogo altamente competitivo e pela sua ampla popularidade no cenário de jogos eletrônicos. Como um jogo de tiro em primeira pessoa que combina elementos táticos intensivos e necessidade de trabalho em equipe, Valorant exige dos jogadores rápidas tomadas de decisão e estratégias, proporcionando um ambiente que os jogadores podem experimentar uma variedade de emoções.

O jogo foi instalado em um computador desktop com sistema operacional Windows 10, processador Intel i7-7700k 3.6GHz Quad-core, 16 GB de memória RAM DDR3, e placa de vídeo NVIDIA GTX 1070 com 8GB GDDR5 de memória.

Uma amostragem por conveniência, composta por cinco alunos voluntários dos cursos de Computação da UPF, foi adotada para validação inicial da ferramenta. Os testes foram feitos de forma individual, no Laboratório de Realidade Virtual e Visão Computacional da UPF, entre os dias 29 de outubro e 1º de novembro de 2024. Os voluntários participavam do experimento em posição sentada, em virtude da câmera de monitoramento ter sido fixada na parte superior do monitor onde o jogo era exibido, e porque um jogo competitivo é normalmente jogado desta forma.

O procedimento no dia dos testes foi o seguinte: cada participante primeiramente era instruído sobre os objetivos do projeto e os procedimentos do experimento, informando que o mesmo teria suas emoções monitoradas durante o jogo. Em seguida, os jogadores voluntários eram convidados a jogar uma partida no modo frenético do jogo, o qual propicia um ambiente intenso para a manifestação de diversas emoções (cada partida poderia ter nove rodadas de embate). Ao final da partida, o participante respondia a um questionário para avaliar sua experiência durante o jogo, do ponto de vista emocional.

Para complementar a coleta de dados, foi implementado um questionário pós-experimento aos participantes. Inicialmente, foram coletados dados de categorização da amostra: idade, gênero, nível de formação e frequência com que interagia com jogos de tiro tático (mais que duas horas diariamente, até duas horas diariamente, mais que duas horas semanalmente, até duas horas semanalmente, ou num tempo inferior aos citados).

Em relação às perguntas sobre a experiência emocional no jogo, os participantes registravam:

- Q1) Seu estado emocional predominante durante o experimento com o jogo Valorant, modo frenético (marcando uma das sete macro emoções):
 - Antes de iniciar a partida;

⁴Disponível em <https://playvalorant.com>.

- Durante o jogo, em geral;
- Ao final da partida.
- Q2) Momentos em que tinham certeza que seu estado emocional mudou ao longo do tempo, em relação ao início da partida. Para tanto, selecionavam uma ou mais emoções que sentiram, caso lembrassem, por rodada do jogo;
- Q3) Se haviam vencido ou perdido a partida;
- Q4) Se lembravam quantos jogadores abateram, e quantas vezes foram abatidos;
- Q5) Sugestões, críticas e aspectos positivos e negativos do monitoramento de emoções durante a experiência.

A Figura 3 mostra a configuração do experimento. O tempo do experimento durou, em média, 25 minutos para cada participante, considerando a passagem de instruções, a atuação no jogo e o preenchimento do questionário proposto.

4. Resultados e Discussão

De modo geral, pode-se dizer que o modelo de CNN se comportou dentro do esperado na detecção de faces e categorização de emoções, executando em um Raspberry Pi. Também não observou-se atrasos perceptíveis entre a detecção da face e seu salvamento em disco. Porém, ficou evidente um pequeno atraso para a exibição da detecção rotulada na interface de monitoramento do avaliador. A Figura 4 apresenta exemplos de capturas realizadas corretamente pelo modelo durante o processo interativo.

Aparentemente a emoção neutra é mantida na maior parte do tempo, apresentando variações para as emoções “Angry”, “Sad” e “Happy” no decorrer da partida. Essa predominância pode ter relação com o nível de concentração exigido pelo jogo, onde os jogadores precisam manter o foco para ter um bom desempenho. As variações para raiva e tristeza podem estar relacionadas a momentos específicos do jogo, como derrotas em confrontos ou falhas para atingir um objetivo, refletindo uma frustração ou desapontamento do jogador. Os gráficos da Figura 5 comprovam essa situação.

Em relação do questionário todos os participantes responderam. Quanto ao “estado emocional predominante durante o experimento” (Q1), 80% responderam que se sentiram neutros e 20% alegres antes da partida. Durante a partida 60% neutros e 40% alegres. E após a partida, houve uma mudança, onde 60% relatando sentimento de felicidade, 20% tristeza e 20% neutros, mostrando que essa alteração pode estar associada ao resultado da partida e satisfação com o desempenho no jogo.

A partir dos dados coletados pela aplicação, observou-se que todos os jogadores começaram com emoções neutras. Em média, durante as partidas, manteve-se a neutralidade, com emoções de felicidade mais recorrentes. Esses dois momentos corroboram com o que foi relatado nos questionários. Porém, apesar de os participantes terem relatado no questionário que, ao final da partida, a emoção predominante era de felicidade, detectou-se nas imagens capturadas uma predominância em emoções neutras.

Nesse contexto, a emoção neutra foi a que teve mais ocorrências ao longo das partidas, enquanto as emoções de tristeza e raiva apareceram somente em momentos pontuais durante o jogo. Estes dados refletem o estado emocional predominante dos jogadores em cenários de que exigem alta concentração, onde manter uma expressão neutra pode ser uma resposta natural às demandas do jogo.



Figura 3. Usuários interagindo com o jogo Valorant, monitorado pelo protótipo desenvolvido no estudo.

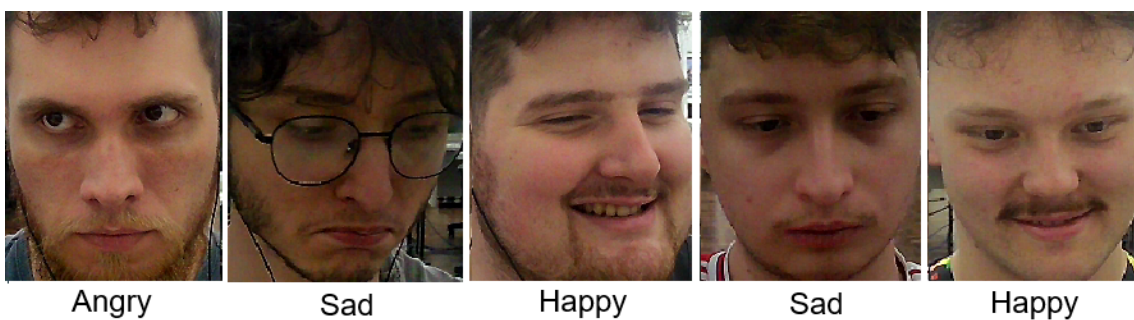


Figura 4. Exemplos de imagens capturadas da face dos participantes.

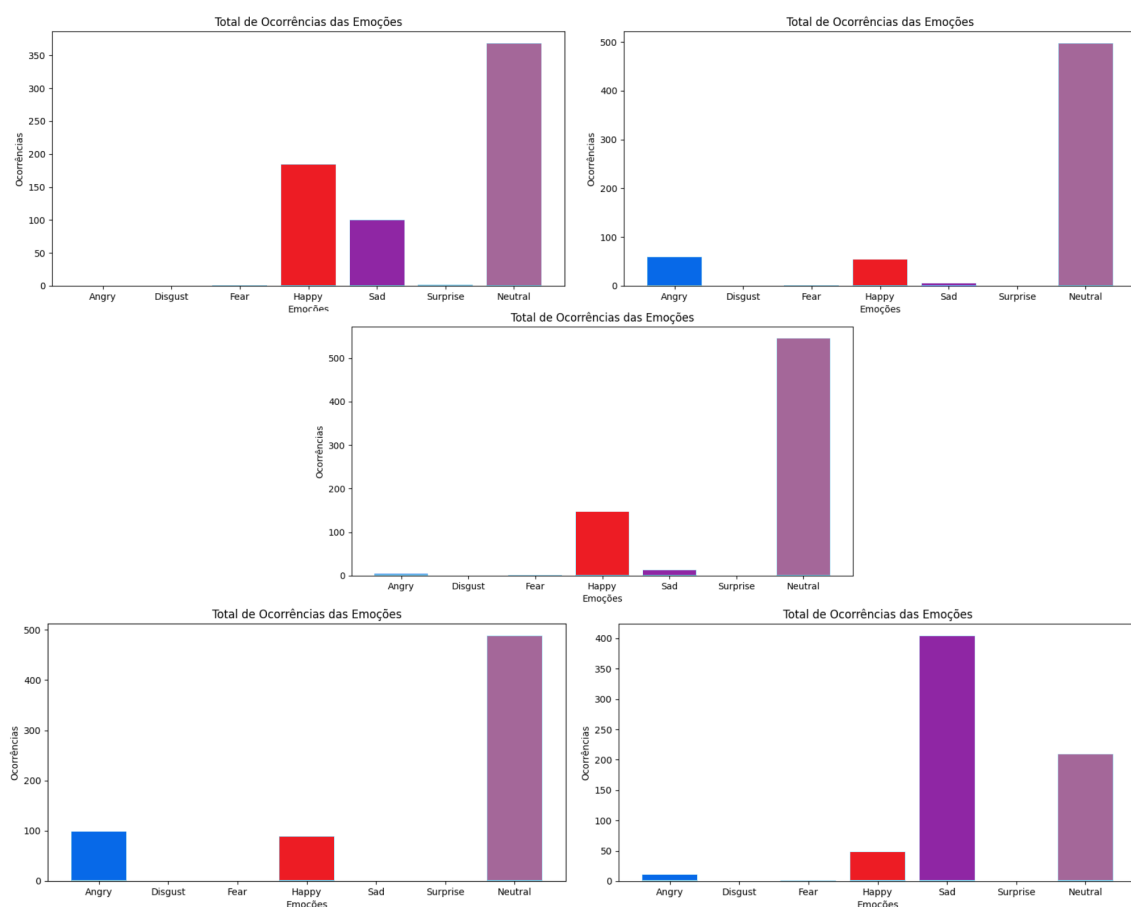


Figura 5. Contagem das emoções detectadas no experimento, por participante.

O cruzamento da emoção percebida pelo usuário e reconhecida pela aplicação mostra que, mesmo os jogadores percebendo seu estado como neutros ou alegres, podem expressar emoções negativas durante a partida sem estarem conscientes disso. Isso mostra a complexidade do reconhecimento de emoções em cenários competitivos e ressalta a importância de combinar outras abordagens para uma compreensão mais abrangente das respostas emocionais dos jogadores.

Para comparar o que foi respondido sobre “momentos em que tinham certeza que seu estado emocional mudou ao longo do tempo” (Q2) e o que a aplicação capturou por imagens, foi considerado nas análises somente o tempo após o começo efetivo de cada rodada que é de, no máximo, 1 minuto e 40 segundos. O tempo de escolha do agente e o período de preparação foram desconsiderados, pois nesses momentos não há ação e as expressões tendem a ser neutras. Só a primeira rodada conta com um tempo de 45 segundos para preparação e compra de itens, ao passo que as rodadas seguintes o tempo é de 30 segundos. Após a inversão de papéis no jogo, a primeira rodada novamente conta com 45 segundos e os demais com 30 segundos.

É fundamental contextualizar como o desempenho nas partidas influenciou as emoções dos jogadores. A partir dos dados coletados durante o jogo nas respostas da Q2, tentou-se compreender a relação entre as ações dos jogadores, os resultados obtidos em cada rodada e as emoções capturadas. Para tanto, as análises também examinam o

desempenho de cada jogador (J) em termos vitórias ou derrotas (Q3) e abates (Q4) nas rodadas, e como esses fatores podem ter impactado suas emoções durante a partida.

É importante enfatizar que o ser humano nem sempre demonstra aquilo que realmente sente, as emoções exibidas publicamente podem ser moduladas por normas culturais, pressões sociais e objetivos pessoais [Ekman 1972]. Por exemplo, um sorriso pode não refletir genuína felicidade, podendo ser apenas um sinal de cortesia, e a neutralidade pode mascarar raiva ou tristeza [Ekman & Friesen 1969].

O gráfico da Figura 6 apresenta as emoções ao longo do tempo do jogador J1. Com base no histórico do jogo, ele começou a primeira rodada interagindo durante, aproximadamente, 86s (45s de preparação mais 41s de ação). Aos 74s, ele foi abatido e, neste tempo, a aplicação capturou emoções de tristeza. J1 marcou no questionário que a emoção durante a segunda rodada foi neutra. Pelo gráfico, a segunda rodada começa em 116s, e o jogador foi abatido 2s após o início da mesma (118s). Observando o gráfico temporal de J1, pode-se ver que entre 116s e 148s (contemplando partes da primeira e segunda rodada), a expressão neutra (cor rosa) foi capturada, de certa forma corroborando com o indicado no questionário.

Na terceira rodada (entre 178s e 228s), J1 marcou no questionário que sentiu tristeza. Aos 227s, o jogador abateu um adversário e logo na sequência foi abatido (228s). Neste intervalo, pode-se notar que a expressão de felicidade (cor vermelha) e raiva (cor azul) foram capturadas, não condizendo com o que foi relatado no questionário. Nesse caso, é possível que expressões de tristeza estejam associadas a raiva. J1 abateu um inimigo na terceira rodada e um na rodada seguinte.

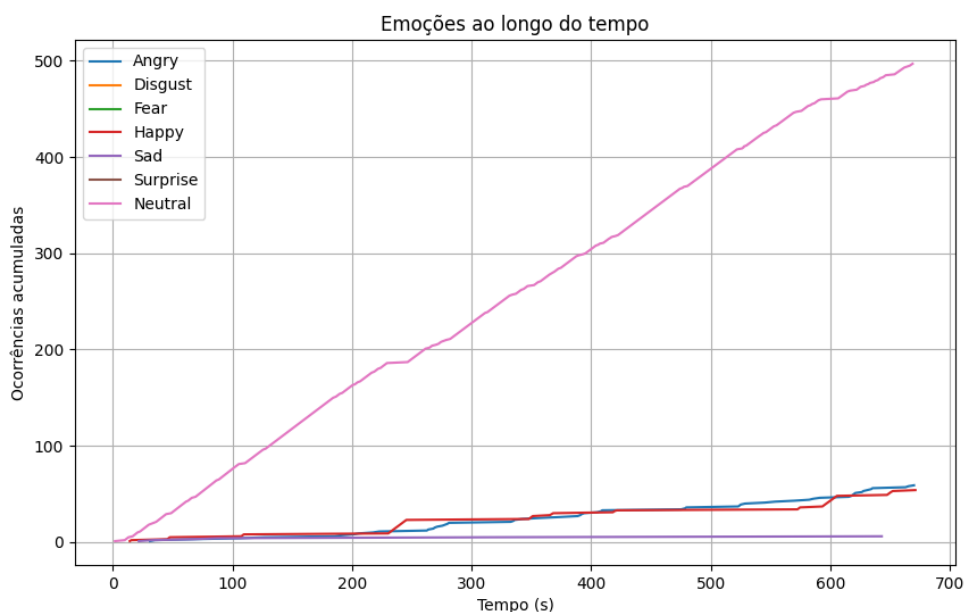


Figura 6. Gráfico temporal de emoções ao longo do tempo do jogador J1.

O jogador J2 marcou no questionário que sua emoção na terceira rodada foi de raiva. A terceira rodada aconteceu entre 202s e 272s. Observando o gráfico da Figura 7,

pode-se confirmar que o modelo detectou corretamente a expressão informada pelo jogador. J2 relatou no questionário que sentiu também emoções neutras durante a quarta e a quinta rodada (intervalo entre 302s e 441s), o que também pode ser confirmado no gráfico. Para este jogador, o modelo também detectou expressões de felicidade nessas mesmas rodadas. Porém, J2 só marcou que isso ocorreu na segunda rodada (entre 129s e 172s). J2 abateu um inimigo na primeira rodada, dois na segunda e dois na terceira.

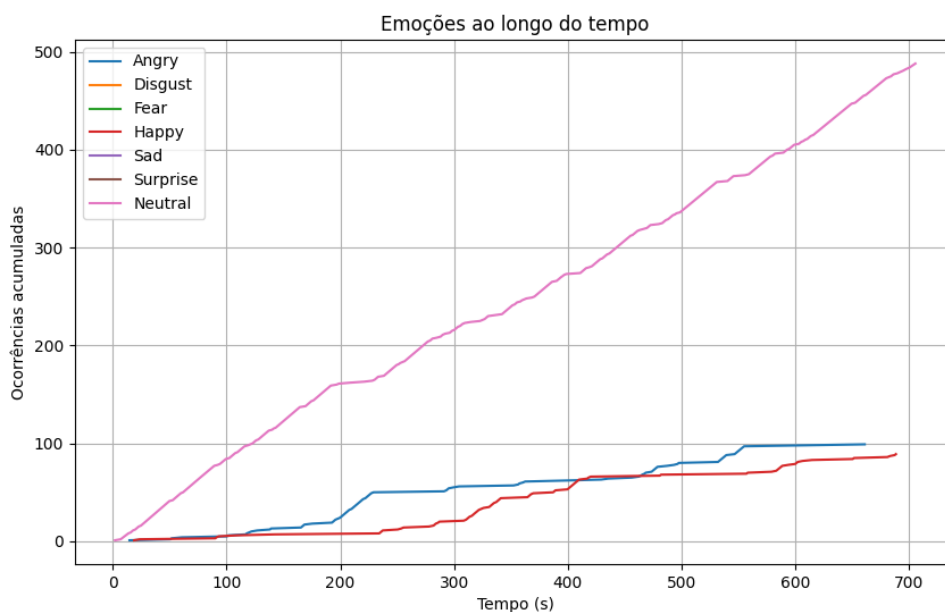


Figura 7. Gráfico temporal de emoções ao longo do tempo do jogador J2.

O jogador J3 teve a emoção de tristeza capturada na maior parte do tempo, como mostra na Figura 8. Este jogador relatou no questionário sentir-se neutro na metade das rodadas, relatou também sentir-se feliz nas três primeiras rodadas, que aconteceram entre 45s e 265s. No gráfico pode-se notar que existem picos de felicidade (cor vermelha) e neutralidade (cor rosa) nesse período. Sobre a emoção de tristeza que o jogador sentiu, pode ter sido um erro do modelo, já que o jogador só relata esta emoção nas duas últimas rodadas que aconteceram entre 414s e 583s. Neste caso, o modelo pode ter errado em virtude do participante ter ficado com a cabeça inclinada para baixo em boa parte do tempo. O jogador abateu um inimigo na quarta rodada e dois na quinta rodada.

O jogador J4 foi o único que ganhou a partida, demonstrando emoção de felicidade mais vezes que os demais. No questionário o jogador não lembrou em quais rodadas teve emoções específicas, mas relatou que no início e no decorrer da partida, sentiu-se neutro e, no final, feliz, mostrando que o resultado positivo pode ter influenciado seu estado emocional. Na segunda rodada, aos 145s, o jogador foi abatido. Observando o gráfico da Figura 9, pode-se perceber que por volta deste tempo foram registradas expressões de raiva e tristeza. Na quarta rodada, aos 314s o jogador abateu um inimigo, tendo um pico na emoção de felicidade registrado pela solução.

O jogador J5 relatou no questionário que sentiu nojo, desgosto (“Disgust”) durante as rodadas 1, 2, 6, 7 e 8. Porém, a aplicação não capturou nenhuma expressão relacionada,

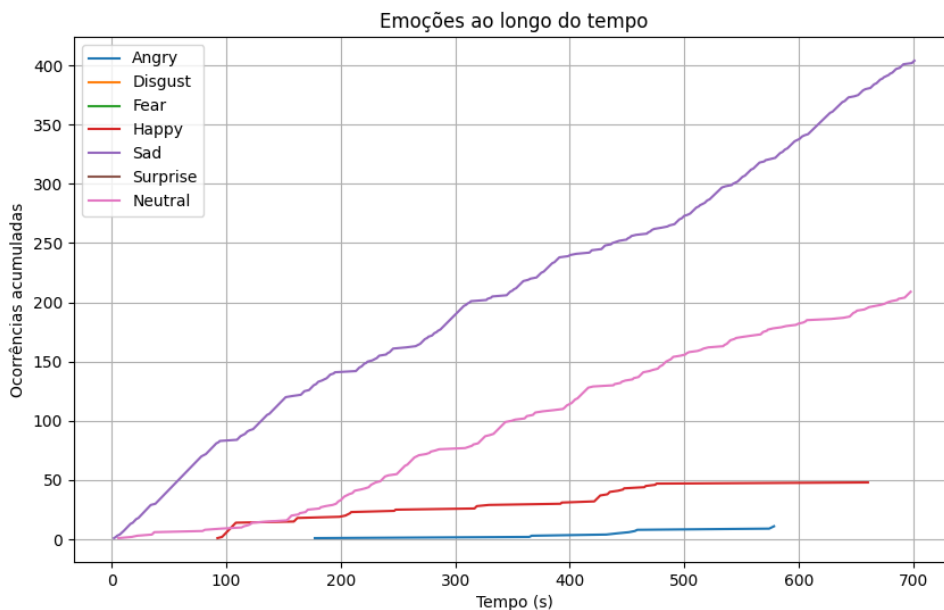


Figura 8. Gráfico temporal de emoções ao longo do tempo do jogador J3.

como mostra o gráfico da Figura 10. J5 também relatou que se sentiu neutro em todas as rodadas. Além destas, o jogador marcou emoções de felicidade e raiva na terceira rodada. Comparando com seu desempenho no jogo, essas sensações podem estar associadas aos momentos em que abateu adversários ou que foi abatido. Observando a rodada citada, ela inicia aos 186s e, aos 217 s, o jogador realiza um abate e logo (222s) é abatido. Pelo gráfico, é possível perceber apenas um pico de felicidade, mas não de raiva.

A Figura 11 resume as respostas obtidas na Q2. Pode-se perceber a frequência da emoção neutra, seguida pela emoção de felicidade.

Observou-se ainda que 80% dos participantes perderam a partida, e 20% venceram. Além disso, todos os jogadores foram abatidos mais vezes do que abateram adversários, o que contribuiu para o registro em questionário de sensações negativas (raiva, tristeza e nojo), como também pode visto na Figura 11.

Em relação a Q5, os participantes relataram que o experimento foi interessante e que os resultados poderiam trazer informações relevantes não só para o cenário competitivo de jogos, mas também para diversas áreas que a análise de emoções se traz necessária. Além disso, foi registrado que seria interessante observar a ferramenta ativa em tempo real nos campeonatos, acrescido de medição de batimentos cardíacos para melhor entendimento das emoções.

Ao longo dos experimentos com os voluntários, pode-se observar que, em determinada distância não mensurada, o modelo de CNN tinha a capacidade de capturar faces de pessoas que estavam circulando próximo ao ambiente de testes. No entanto, a aplicação foi preparada para considerar somente a captura da imagem da pessoa que estava mais próxima à câmera (jogador).

Por outro lado, o modelo de CNN apresentou problemas de acurácia na detecção

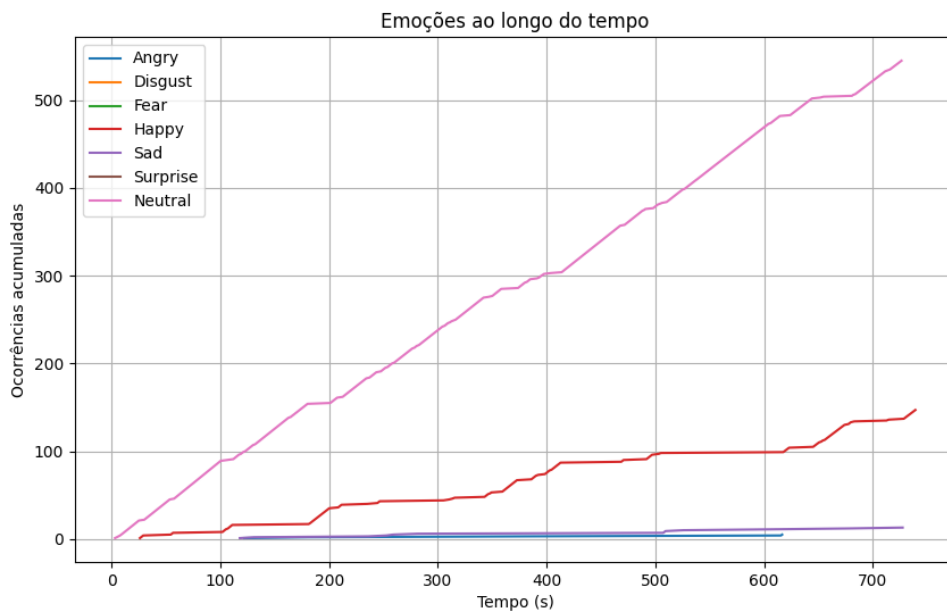


Figura 9. Gráfico temporal de emoções ao longo do tempo do jogador J4.

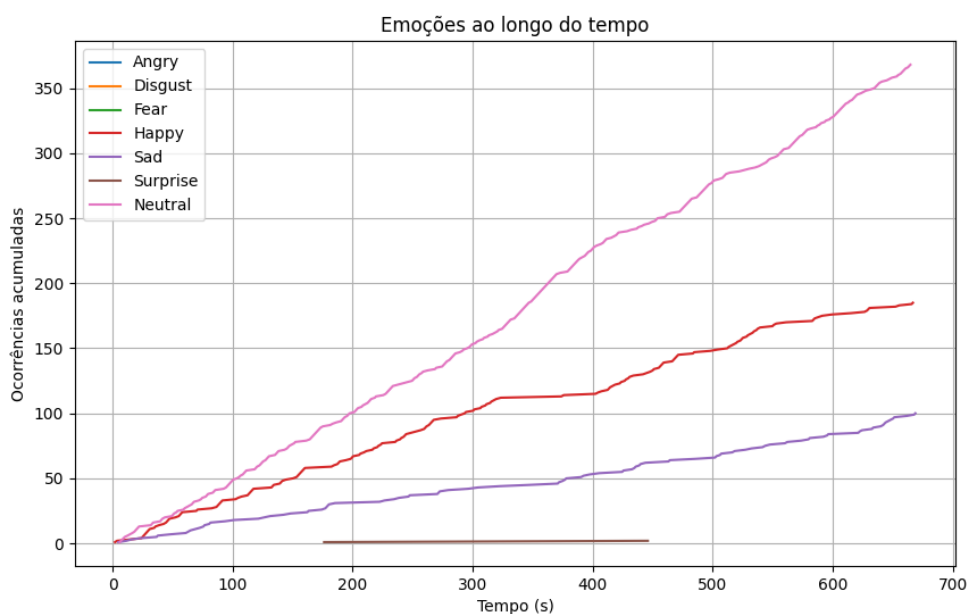


Figura 10. Gráfico temporal de emoções ao longo do tempo do jogador J5.

de emoções quando o usuário utilizava óculos de grau, ou estava com o cabelo posicionado à frente da testa. Este fato foi observado, com evidência, em um dos voluntários, no qual o modelo inferiu expressões como tristeza, raiva ou medo para momentos onde o usuário apresentava uma emoção alegre ou neutra. Esse comportamento pode ter relação com o conjunto de dados disponível para treino, que não apresentava imagens suficientes com este tipo de padrão facial (parte do rosto obstruído ou com algum acessório vestível).

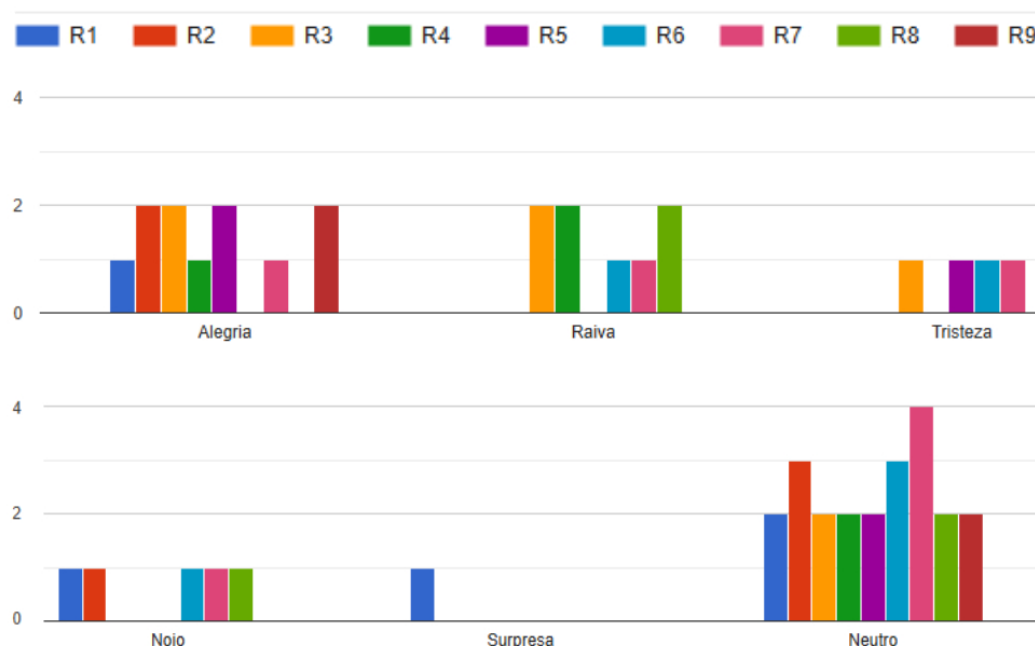


Figura 11. Gráfico de emoções em cada rodada gerado no questionário.

Também foi possível observar que, em situações onde o participante ficava com a cabeça inclinada cerca de 30 a 45 graus para baixo, em relação a câmera, o modelo detectou mais emoções de tristeza - quando, na realidade, o participante demonstrava expressão neutra, apenas concentrado na partida. A Figura 12 exemplifica casos que o modelo errou na predição e captura.

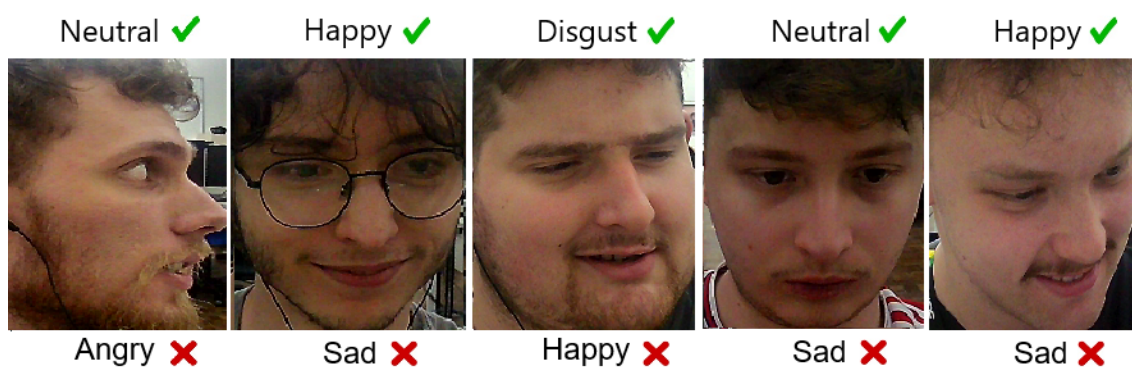


Figura 12. Exemplos de emoções detectadas erroneamente pelo modelo. Na parte superior, destacam-se as emoções esperadas. Na parte inferior, as emoções previstas.

Em relação ao posicionamento da câmera para monitoramento, é importante ressaltar que existem jogadores no cenário competitivo que utilizam o monitor muito perto da face, impossibilitando seu uso de frente ao jogador. Porém, durante os testes, foi possível capturar expressões do jogador quando este ficava com a face de lado para a câmera, levantando a possibilidade de avaliar, futuramente, o monitoramento lateral. Nesse caso, é fundamental considerar um conjunto de dados de treinamento que contenha também

imagens de perfil.

Por fim, pode-se destacar que as principais contribuições desse estudo são: a oferta de uma solução funcional que reconhece emoções primárias, usando modelo inteligente simples e um computador de pequeno porte; e que a detecção de emoções primárias conseguem retratar, em boa parte do tempo, o comportamento humano perante uma determinada atividade (no caso, um jogo competitivo). É uma solução que tem potencial para aplicação em ambientes competitivos e monitoramento de saúde mental em jogos.

A abordagem proposta se mostrou acessível, possibilitando a execução de modelos complexos de reconhecimento de emoções em um dispositivo de baixo custo e com recursos computacionais reduzidos. O sistema conseguiu oferecer *insights* valiosos sobre o estado emocional de jogadores em tempo real, sem interferir no desempenho do jogo, permitindo uma análise emocional contínua. A implementação na Raspberry Pi também pode aliviar a carga de processamento na máquina que executou o jogo, liberando o consumo de recursos para um bom desempenho do próprio jogo.

5. Conclusão

Este trabalho apresentou uma abordagem para reconhecimento de emoções em cenários de jogos usando CNN e Raspberry Pi. Pode-se concluir que o modelo de reconhecimento de emoções apresentou uma acurácia satisfatória. Apesar de ser uma rede neural profunda que exige mais processamento, o modelo DNN do OpenCV foi bastante útil para otimização da detecção de faces no Raspberry. O sistema detectou predominantemente expressões neutras, com variações para raiva e tristeza. Os usuários relataram sentir-se neutros ou alegres antes e durante o jogo, e alegres e tristes após a partida, o que sugere que os sistema detectou emoções não percebidas conscientemente pelos jogadores, evidenciando o potencial de sistemas de reconhecimento facial em revelar estados emocionais subjacentes.

O trabalho demonstra que um sistema de reconhecimento de emoções pode ser embarcado em um dispositivo de baixo custo, e utilizado em tempo real em contextos competitivos de jogos. Este diferencial amplia as possibilidades de monitoramento emocional dos jogadores de maneira não invasiva e econômica, com potencial para ser aplicado em torneios e treinamentos de equipes. A solução permite que as emoções sejam monitoradas ao longo das partidas, oferecendo uma ferramenta para aprimorar o desempenho de jogadores e entender melhor o impacto emocional em cenários de alta pressão.

Apesar das limitações de processamento do hardware de baixo custo, a aplicação comportou-se de forma eficiente no Raspberry Pi, realizando o processamento e reconhecimento das emoções em tempo real, o que demonstra a viabilidade de implementar a proposta em sistemas embarcados.

Por outro lado, nota-se que é preciso refinar o modelo, agregando mais dados e variabilidade no processo de treinamento, capacitando-o para detectar faces parcialmente ocultas ou com acessórios. Além disso, ainda se manteve um pequeno atraso na captura das emoções, e demora para refletir na tela os movimentos feitos pelo jogador, o que indica que técnicas de otimização para redução de latência podem ser empregadas.

Outro possível trabalho futuro é a implementação de um modelo para detecção de micro expressões faciais, tendo em vista que os jogadores podem demonstrar expressões

muito sutis ao longo do processo interativo. Nesse contexto, também poderia ser agregado ao conjunto de dados que contenham sinais fisiológicos e voz, o que pode proporcionar mais abrangência dos estados emocionais. Por fim, pode-se focar na adoção de uma arquitetura CNN de estado da arte mais robusta, como YOLO, Vision Transformer ou EfficientNet, permitindo que o modelo aprofunde o aprendizado de padrões.

Referências

- Ab Wahab, M. N., Nazir, A., Zhen Ren, A. T., Mohd Noor, M. H., Akbar, M. F., & Mohamed, A. S. A. (2021). Efficientnet-lite and hybrid cnn-knn implementation for facial expression recognition on raspberry pi. *IEEE Access*, 9:134065–134080. DOI: <https://doi.org/10.1109/ACCESS.2021.3113337>.
- Alipour, M., Moghaddam, M. T., Vaidhyanathan, K., & Kjærgaard, M. B. (2023). Emoticontrol: Emotions-based control of user-interfaces adaptations. *Proceedings of the ACM on Human-Computer Interaction*, 7(EICS):1–29. DOI: <https://doi.org/10.1145/3593227>.
- Aranha, R. V. (2022). *Adaptação automática de jogos sérios a partir de emoções e traços de personalidade*. Doutorado em Engenharia Elétrica – Escola Politécnica, Universidade de São Paulo, São Paulo, SP. Disponível em <https://doi.org/10.11606/T.3.2022.tde-31032023-080241>.
- Atkinson, R., Atkinson, R., Smith, E., Bem, D., & Nolen-Hoeksema, S. (2002). *Introducere in psihologie*. Bucuresti: Ed. Tehnica. ISBN: 973-31-1398-0.
- Beres, N. A., Klarkowski, M., & Mandryk, R. L. (2023). Playing with emotions: a systematic review examining emotions and emotion regulation in esports performance. *Proceedings of the ACM on Human-Computer Interaction*, 7(CHI PLAY):558–587. DOI: <https://doi.org/10.1145/3611041>.
- Capra, M., Bussolino, B., Marchisio, A., Masera, G., Martina, M., & Shafique, M. (2020). Hardware and software optimizations for accelerating deep neural networks: Survey of current trends, challenges, and the road ahead. *IEEE Access*, 8:225134–225180. DOI: <https://doi.org/10.1109/ACCESS.2020.3039858>.
- Davis, M. & Lang, P. J. (2003). *Emotion*, chapter 15, pages 405–439. John Wiley & Sons, Ltd. DOI: <https://doi.org/10.1002/0471264385.wei0315>.
- de Oliveira, S. (2017). *Internet das coisas com ESP8266, Arduino e Raspberry PI*. Novatec Editora. ISBN: 978-8575225813.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In Cole, J., editor, *Nebraska Symposium on Motivation*, volume 19, pages 207–283. University of Nebraska Press, Lincoln.
- Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(1):205–221. DOI: <https://doi.org/10.1196/annals.1280.010>.
- Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98. DOI: <https://doi.org/10.1515/semi.1969.1.1.49>.

- Frijda, N. H. (2008). *The psychologists' point of view*, chapter 5, page 68–87. The Guilford Press, 3 edition. Disponível em <https://psycnet.apa.org/record/2008-07784-005>.
- Garcia, L. H. (2021). Detecção de emoções utilizando redes neurais convolucionais em sistemas com recursos limitados de hardware. Graduação em Engenharia Elétrica – Escola de Engenharia, Universidade Federal do Rio Grande do Sul. Disponível em <https://lume.ufrgs.br/handle/10183/234173>.
- Gong, J., Liu, W., Pei, M., Wu, C., & Guo, L. (2022). Resnet10: A lightweight residual network for remote sensing image classification. In *2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pages 975–978. IEEE. DOI: <https://doi.org/10.1109/ICMTMA54903.2022.00197>.
- Guglielmo, G., Blom, P. M., Klinecicz, M., Čule, B., & Spronck, P. (2022). Face in the game: Using facial action units to track expertise in competitive video game play. In *2022 IEEE Conference on Games (CoG)*, pages 112–118. DOI: <https://doi.org/10.1109/CoG51982.2022.9893599>.
- Jofilsan, N. C., de Lima, S. M. L., Alencar, R. C., Silva, S. H. M. T., de Albuquerque, J. V. O., Soares Filho, O. H., Maciel, A. M. A., & dos Santos, W. P. (2021). Classificação de gamers por padrões em suas expressões faciais. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*, pages 796–805. SBC. DOI: https://doi.org/10.5753/sbgames_estendido.2021.21516.
- Kattakinda, J. P., Dhananjaya, S., Mojzis, S., Nadif, A., Hanna, M., Apsan, R., & Malavolta, I. (2024). Towards understanding the energy consumption of virtual reality applications in gaming, education, and entertainment. In *Proceedings of the ACM/IEEE 8th International Workshop on Games and Software Engineering*, pages 50–57. DOI: <https://doi.org/10.1145/3643658.3643922>.
- Lx, Y. (2019). Facial expression recognition using cnn. Disponível em: <https://www.kaggle.com/code/lxyuan0420/facial-expression-recognition-using-cnn/>. Acesso em: 21 de Out. de 2024.
- Mandryk, R. L. & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, 65(4):329–347. DOI: <https://doi.org/10.1016/j.ijhcs.2006.11.011>.
- Minaee, S., Minaei, M., & Abdolrashidi, A. (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, 21(9):3046. DOI: <https://doi.org/10.3390/s21093046>.
- Picard, R. W. (1997). *Affective computing MIT press*. Cambridge: Massachussetts. ISBN: 978-0262161701.