

## Treinando Modelos Afetivos de Visão Computacional por Crowdsourcing Soft-Target Rótulos

Peter Washington\*, Departamento de Bioengenharia, Universidade de Stanford, peterwashington@stanford.edu

Haik Kalantarian, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, haik.kalantarian@gmail.com

Jack Kent, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, jackkent@stanford.edu

Arman Husic, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, ahusic@stanford.edu

Aaron Kline, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, akline@stanford.edu

Emilie Leblanc, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, emilie.leblanc@stanford.edu

Cathy Hou, Departamento de Ciência da Computação, Universidade de Stanford, cathyhou@stanford.edu

Cezmi Mutlu, Departamento de Engenharia Elétrica, Universidade de Stanford, cezmi@stanford.edu

Kaitlyn Dunlap, Departamento de Pediatria (Medicina de Sistemas), kaiti.dunlap@stanford.edu

Yordan Penev, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, ypenev@stanford.edu

Nate Stockham, Departamento de Neurociência, Universidade de Stanford, stockham@stanford.edu

Brianna Chrisman, Departamento de Bioengenharia, Universidade de Stanford, briannac@stanford.edu

Kelley Paskov, Departamento de Ciência de Dados Biomédicos, Universidade de Stanford, kpaskov@stanford.edu

Jae-Yoon Jung, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, jaeyjung@stanford.edu

Catalin Voss, Departamento de Ciência da Computação, Universidade de Stanford, catalin@cs.stanford.edu

Nick Haber, Escola de Pós-Graduação em Educação, Universidade de Stanford, nhaber@stanford.edu

Dennis P. Wall\* . Departamentos de Pediatria (Medicina de Sistemas), Ciência de Dados Biomédicos e Psiquiatria e Comportamento  
Ciências, Universidade de Stanford, dpwall@stanford.edu

\*Autores correspondentes: Peter Washington (peterwashington@stanford.edu), Dennis P. Wall (dpwall@stanford.edu)

Conflitos de interesse: Dr. Dennis P. Wall é o fundador científico da Cognoa, uma empresa focada em saúde pediátrica digital; a abordagem e os achados apresentados neste artigo são independentes/não relacionados à Cognoa. Todos os outros autores declararam que não existem interesses concorrentes.

## ABSTRATO

Os classificadores de emoções tradicionalmente predizem emoções discretas. No entanto, as expressões de emoção são muitas vezes subjetivas, exigindo assim um método para lidar com rótulos subjetivos. Exploramos o uso de crowdsourcing para adquirir rótulos *soft-target* confiáveis e avaliamos um classificador de detecção de emoções treinado com esses rótulos. Centramos nosso estudo no conjunto de dados *Child Affective Facial Expression (CAFE)*, uma coleção padrão-ouro de imagens que retratam expressões faciais pediátricas, juntamente com 100 rótulos humanos por imagem. Para testar a viabilidade do crowdsourcing para gerar esses rótulos, usamos *Microworkers para adquirir rótulos para 207 imagens CAFE*. Avaliamos tanto os trabalhadores não filtrados quanto os selecionados por meio de um curto processo de filtragem de multidão. Em seguida, treinamos duas versões de um classificador em rótulos CAFE de destino flexível usando as 100 anotações originais fornecidas com o conjunto de dados: (1) um classificador treinado com rótulos codificados tradicionais e (2) um classificador treinado com rótulos de vetor representando o distribuição de respostas do anotador CAFE. Comparamos as distribuições de saída softmax resultantes dos dois classificadores com um teste t independente de 2 amostras de distâncias L1 entre a distribuição de probabilidade de saída do classificador e a distribuição de rótulos humanos. Embora a concordância com o CAFE seja fraca para trabalhadores de multidão não filtrados, a multidão filtrada concorda com os rótulos CAFE 100% do tempo para muitas emoções. Enquanto o F1-score para um classificador codificado one-hot é muito maior (94,33% vs. 78,68%) em relação aos rótulos CAFE de verdade, o vetor de probabilidade de saída do classificador treinado pela multidão se assemelha mais à distribuição de rótulos humanos ( $t=3,2827$ ,  $p=0,0014$ ). Relatar uma distribuição de probabilidade de emoção que explica a subjetividade da interpretação humana. O crowdsourcing, incluindo um mecanismo de filtragem suficiente, é uma solução viável para a aquisição de rótulos de alvo fácil.

## INTRODUÇÃO

Modelos de aprendizado de máquina que preveem emoções humanas a partir de imagens de expressões faciais são cada vez mais usados em sistemas interativos [5, 33-34, 51, 83] e aplicações como análise multimodal de sentimentos [42, 67-68], saúde [74, 100], e veículos autônomos [23]. O reconhecimento de emoções é tradicionalmente modelado como um problema de classificação, onde o modelo prevê uma categoria de emoção discreta. No entanto, as expressões faciais são muitas vezes ambíguas [15, 104-107], e muitas vezes não é ideal para um modelo de aprendizado de máquina produzir uma única classe para um rótulo subjetivo. Felizmente, a maioria dos métodos de aprendizado supervisionado produz uma distribuição de probabilidade sobre todas as classes possíveis. Algumas vezes, o sistema de computação afetiva irá visualizar essa distribuição para o usuário [33]. Exemplos incluem serviços comerciais de detecção de emoções como Affectiva [55-56] e displays de veículos autônomos [83]. Em um grande número de casos de uso, no entanto, apenas a classe com maior probabilidade é visualizada [34, 51].

Enquanto o paradigma de treinar um modelo com um rótulo codificado discreto e prever uma distribuição de probabilidade é razoável quando os dados de treinamento têm rótulos indiscutíveis, imagens de expressões faciais podem ter rótulos ambíguos ou mesmo vários rótulos corretos simultaneamente, e o rótulo deve idealmente representar essa incerteza inerente. A rotulagem de alvo flexível, em que os rótulos de treinamento representam uma distribuição probabilística em vez de um rótulo codificado one-hot, é uma solução estabelecida para esse problema. O treinamento com rótulos *soft-target* resulta em classificadores que preveem distribuições de probabilidade representativas dos rótulos *soft-target* [3, 25, 101].

Nossa hipótese é que o crowdsourcing pode gerar distribuições que espelham aquelas geradas em um ambiente de laboratório.

Aqui, exploramos o uso de crowdsourcing para adquirir uma distribuição de rótulos para imagens com classes ambíguas ou múltiplas (chamamos essas "rótulos subjetivos"). Primeiro descrevemos a aquisição de rótulos de crowdsourcing para quatro imagens representativas que exibimos ao leitor junto com a distribuição das respostas da multidão a

demonstrar os fenômenos de rótulos subjetivos na computação afetiva. Em seguida, fazemos crowdsourcing da rotulagem de um subconjunto do conjunto de dados Child Affective Facial Expression (CAFE), uma coleção de imagens emotivas de crianças que vem convenientemente com 100 anotações humanas independentes por imagem. A seguir, mostramos que a distribuição de crowdsourcing espelha a distribuição original do CAFE, validando a viabilidade do crowdsourcing para gerar uma distribuição confiável e representativa de rótulos humanos para uma imagem. Finalmente, comparamos o desempenho de duas versões de uma rede neural convolucional (CNN) treinada em CAFE: uma com vetores codificados a quente tradicionais e outra com rótulos soft-target baseados em respostas do anotador CAFE. Descobrimos que o classificador treinado com alvos fáceis resulta em previsões do classificador que espelham muito mais de perto a verdadeira distribuição humana em sujeitos independentes não incluídos no conjunto de treinamento. Esperamos que este trabalho seja útil para designers e desenvolvedores de modelos de aprendizado de máquina para sistemas de computação afetiva que desejam fornecer saídas probabilísticas ao usuário final.

## TRABALHO RELATADO

Embora os rótulos de crowdsourcing e soft-target tenham sido estudados em computação afetiva, somos os primeiros a explorar a viabilidade de usar crowdsourcing para adquirir rótulos de soft-target confiáveis para detecção de emoções por visão computacional. Descrevemos o trabalho relacionado abaixo.

### Detecção de emoções faciais

A detecção de emoções faciais é um desafio fundamental para o aprendizado de máquina. Para que as máquinas inteligentes passem de forma convincente no teste de Turing [80], uma compreensão da emoção humana é crucial. Tem havido um forte corpo de literatura de aprendizado de máquina para detectar o efeito humano de uma variedade de fluxos de dados, incluindo áudio [67-68, 103], texto [5, 48], imagens [75] e vídeo [28, 84-85]. Aqui, nos concentramos na detecção de emoções baseada em imagens a partir de expressões faciais.

Fundamental para uma abordagem de visão computacional bem-sucedida para computação afetiva é a representação de recursos da imagem, e existem várias abordagens para a engenharia de tais recursos. Uma abordagem comum é extrair pontos-chave faciais e usar uma representação de características consistindo nas coordenadas dos pontos-chave [16, 28, 60, 71].

Essa abordagem funciona bem quando o conjunto de dados é pequeno, pois a representação em si é compacta e, portanto, passível de abordagens de aprendizado leves, como regressão logística, máquinas de vetor de suporte e árvores de decisão. Outra abordagem de extração de recursos, CNNs, pode aprender automaticamente mapas de recursos não lineares relevantes.

CNNs muitas vezes resultam em desempenho superior a outros métodos quando o conjunto de dados é suficientemente grande [24, 89].

### Detecção de emoções com rótulos subjetivos

Paul Ekman postulou que existem sete emoções humanas fundamentais que são universais em todas as culturas e fronteiras geográficas: feliz, triste, surpresa, raiva, medo, nojo e desprezo [21-22]. No entanto, essas expressões não são mutuamente exclusivas. Du et al. discutiram a existência de *emoções compostas*, ou combinações de emoções existentes para formar novas [17]. Exemplos de emoções compostas incluem “felizmente surpreso”, “terrivelmente surpreso” e “terrivelmente enojado”. Por meio do sensoriamento do smartphone, Zhang et al. descobriram que os pares de emoções que são frequentemente apresentados simultaneamente incluem (feliz, surpreso), (triste, desgosto) e (triste, medo) [105]. Esta questão tem sido explorada para o discurso emocional [15, 107]. Enquanto algumas emoções podem ser expressas em conjunto, outras podem ser singulares, mas ambíguas. A questão da subjetividade nos rótulos de formação, seja por

rótulos ambíguos ou múltiplos rótulos corretos, foi documentado nas áreas de saúde digital e computação afetiva em particular [54, 57, 69, 82, 104].

O tópico de rótulos subjetivos em conjuntos de dados de computação afetiva contendo dados de fala e áudio foi explorado em trabalhos anteriores. Mower et al. representam rótulos de emoção na granularidade dos enunciados, representando assim um perfil de tempo de como a emoção dominante na fala muda rapidamente ao longo do tempo [58-59]. Fujioka alterna entre atualizar os parâmetros da rede neural e atualizar os parâmetros de importância da amostra em cada iteração de treinamento [26]. Ando et al. utilizam treinamento soft-target, onde os rótulos de emoção são baseados na proporção de anotações humanas em vez da tradicional codificação one-hot [3].

O treinamento soft-target é um método geral de aprendizado de máquina para lidar com rótulos de treinamento subjetivos. Essa abordagem é particularmente desejável quando vários rótulos são adquiridos por imagem. A classificação de rótulos suaves pode ser benéfica porque eles podem levar em conta a subjetividade herdada nos rótulos e são robustos contra ruídos aleatórios [79]. Uma variação deste método é uma função de perda suave, que consiste em subtrair a distância mínima entre as classes da distância máxima dentro das classes [99]. O treinamento de alvo suave e de perda demonstrou superar o treinamento de alvo rígido (codificação one-hot) quando o objetivo do treinamento é produzir uma distribuição de saída como a distribuição de rótulos de anotadores [65], e esse fenômeno foi observado em vários conjuntos de dados e tarefas [81].

A questão dos rótulos subjetivos também foi explorada na análise multimodal de sentimentos, onde o objetivo é prever o sentimento de vários fluxos de dados [42, 67-68], incluindo vídeos enriquecidos com afeto. Chaturvedi et al. criaram um classificador fuzzy para prever o grau em que várias emoções são expressas em uma determinada imagem [6]. Outra abordagem é prever a quantidade de valência e excitação exibida em eixos contínuos (regressão) em vez de prever categorias (classificação) [62, 64, 73, 102, 106].

### **Crowdsourcing com rótulos subjetivos**

Existem vários corpos de trabalho que descrevem abordagens para lidar com rótulos de crowdsourcing. Kairam e Heer levantam a hipótese de que existem diferenças intrínsecas, mas válidas, entre os trabalhadores da multidão ao rotular os pontos de dados e, portanto, categorizam os trabalhadores por seus padrões de rotulagem [40]. Outros exemplos de categorização de trabalhadores incluem medir a informatividade da amostra, estratégias de aprendizado ativo e cooperativo e controlar as métricas de confiabilidade do rotulador [69].

Existem outras técnicas de aprendizado estatístico além da rotulagem de alvo flexível discutida acima, que foram bem-sucedidas com rótulos de crowdsourcing. Rodrigues e Pereira adicionam uma "camada de multidão" extra no final de uma arquitetura CNN tradicional treinada para prever as saídas de cada rotuladora individualmente e, portanto, os vieses dos trabalhadores da multidão [70].

Crowdsourcing tem sido usado para adquirir rótulos de emoção de imagens. Korovina et al. descobriram que os trabalhadores da multidão que rotulam categorias de emoção discretas em uma roda de cores tiveram baixas pontuações de concordância (valor Kappa menor que 0,15) [46], enquanto a consistência entre os trabalhadores ao rotular a valência e a excitação foi muito mais forte [45].

## **MÉTODOS**

### **Adquirindo rótulos de multidão para imagens CAFE**

Usamos o conjunto de dados CAFE [52-53], que é o maior conjunto de dados público de imagens frontais que retratam crianças emocionante. CAFE é usado como referência em várias publicações de computação afetiva [52-53] e é o padrão

conjunto de dados de avaliação para computação afetiva pediátrica. CAFE foi originalmente rotulado por 100 avaliadores humanos não treinados, e a distribuição bruta de 100 rótulos humanos por imagem é fornecida junto com os rótulos de verdade.

Por exemplo, a primeira imagem no conjunto de dados foi rotulada como “irritada” por 62% dos avaliadores e como “desgostosa” por 25% dos avaliadores. Todas as outras emoções receberam 5% ou menos rótulos. Pode-se supor a partir desses números que a imagem parece “principalmente zangada” com “algum desgosto”. A inspeção manual da imagem revela uma expressão facial que pode ser razoavelmente categorizada como “raiva” ou “nojo” dependendo do contexto. (As imagens CAFE são protegidas por direitos autorais e não podem ser republicadas, portanto, nos referimos a esta imagem pelo nome do arquivo no conjunto de dados disponível publicamente: *F-AA-01\_052-Angry.jpg*).

Para validar a capacidade do crowdsourcing de produzir uma distribuição confiável de rótulos de verdade, nós reunimos a tarefa de rotular imagens CAFE e comparamos a distribuição gerada pela multidão resultante com a distribuição relatada no conjunto de dados CAFE original. Todo o crowdsourcing foi realizado em Microworkers.com, uma plataforma de crowdsourcing semelhante ao Amazon Mechanical Turk [63], mas com um pool mais representativo globalmente de trabalhadores [32]. Cada tarefa consistia em rotular uma das sete categorias de emoção (feliz, triste, surpreso, zangado, com medo, nojo e neutro) para um subconjunto de imagens no CAFE. Optamos por limitar os rótulos dos avaliadores a avaliações absolutas (representações one-hot) porque queríamos capturar o peso relativo de cada emoção dentro de uma imagem.

Como os humanos são notoriamente pobres em quantificar com precisão as contribuições relativas de componentes individuais em representações mistas, especialmente no caso de reconhecimento de emoções humanas [8-9, 27], pedimos a cada avaliador que fornecesse apenas a emoção mais saliente de acordo com sua interpretação. Ao adquirir rótulos de 100 trabalhadores independentes da multidão por imagem, cada um dando seu voto para a emoção mais proeminente, criamos uma representação descrevendo a subjetividade entre-sujeitos da emoção expressa na imagem.

Adquirimos rótulos para 131 imagens selecionadas aleatoriamente do CAFE e solicitamos 100 rótulos de multidão por imagem. Verificamos manualmente a exatidão de cada rótulo, e os trabalhadores com rótulos consistentemente de alta qualidade foram recrutados para tarefas de rotulagem adicionais para 76 imagens separadas. Aqui, “alta qualidade” significa que os autores poderiam concordar com o rótulo (por exemplo, um rótulo “feliz” para uma imagem claramente “triste” não seria aceito, mas um rótulo “com medo” para uma expressão “terrivelmente surpreso” seria aceito). Nosso objetivo ao excluir trabalhadores sem rótulos consistentemente de “alta qualidade” era filtrar os trabalhadores de multidão que estavam respondendo aleatoriamente para receber o pagamento, pois esse é um problema comum no crowdsourcing [2, 4, 10, 49]. Analisamos os rótulos de trabalhadores filtrados e não filtrados em diferentes conjuntos de imagens CAFE para medir a possibilidade de que os trabalhadores de filtragem pudessem mascarar a ambiguidade dos rótulos.

Todas as tarefas de crowdsourcing foram aprovadas pelo Institutional Review Board (IRB) da Universidade de Stanford. Todos os trabalhadores foram obrigados a assinar um termo de consentimento eletrônico aprovado pelo IRB antes de participar da tarefa.

### Treinamento e teste com distribuições de probabilidade de multidão

Tradicionalmente, os modelos multiclasse são treinados com perda categórica de entropia cruzada, onde  $\tilde{y}$  é o operador de soma,  $C$  é o número de classes,  $p_i$  é a probabilidade de verdade da classe  $i$  e  $q_i$  é a previsão do classificador para a classe  $i$ :

$$-\sum_{i=1}^C \tilde{y}_i \log(q_i)$$

Quando as classes verdadeiras são indiscutíveis, que é a suposição usual para classificação, então a distribuição de probabilidade de verdade fundamental  $p_i$  é uma codificação one-hot (ou seja, uma probabilidade de 1 para a classe “verdadeira” e uma probabilidade

de 0 para todas as outras classes). No caso de classes subjetivas em que o rótulo verdadeiro pode consistir em uma combinação ponderada de várias classes, como em conjuntos de dados de emoções em que emoções complexas estão presentes, hipotetizamos que fornecer rótulos de alvo flexível em vez de codificações de um único resultado resultará em previsões de classificador para sujeitos humanos separados que se assemelham à distribuição de resposta do anotador humano.

Treinamos um modelo de aprendizado de máquina usando dois conjuntos de rótulos de imagem: (1) os rótulos CAFE originais como um vetor codificado a quente e (2) vetores de alvo flexível representando a distribuição de 100 respostas humanas do conjunto de dados CAFE original. Nós estendemos todas as imagens de 5 crianças selecionadas aleatoriamente do CAFE (F-AA-01, F-EA-39, M-LA-08, M-AA-11 e F-LA-13, correspondendo a uma mulher africana americano, uma mulher europeia americana, um homem latino-americano, um homem afro-americano e uma mulher latino-americana) e usou-as como imagens do conjunto de teste. As demais imagens foram usadas para treinar o classificador. 1.141 imagens (196 irritadas, 180 revoltadas, 135 assustadas, 206 felizes, 222 neutras, 103 tristes e 99 surpresas) foram usadas no trem e 51 imagens (9 irritadas, 11 revoltadas, 5 assustadas, 9 felizes, 8 neutras, 5 tristes e 4 surpresas) foram usados em

o conjunto de teste.

Transferimos o aprendizado em um ResNet-152 [31] CNN pré-treinado em ImageNet [13]. Treinamos cada rede neural usando o framework Keras [7] com um backend TensorFlow [1] para 100 épocas com um tamanho de lote de 16 e uma taxa de aprendizado de 0,0003 usando a otimização Adam [43]. Para aumentar a generalização do processo de treinamento e reduzir o overfitting, aplicamos as seguintes estratégias de aumento de dados: uma faixa de rotação de 7 graus, uma faixa de zoom de 15%, uma faixa de cisalhamento de 5%, uma faixa de brilho de 70% a 130%, e inversão horizontal.

## RESULTADOS

### Demonstração de Emoções Subjetivas

Os métodos descritos aqui não são específicos do CAFE. Nós nos concentramos no CAFE neste artigo como um estudo de caso de um conjunto de dados de computação afetiva popular e como um conjunto de dados que fornece rótulos de verdade para muitos anotadores humanos (100) por imagem. No entanto, as imagens CAFE estão sujeitas a direitos autorais e não podem ser republicadas. Para fornecer ao leitor exemplos visuais de expressões faciais com um grande número de anotações de multidão por imagem, exibimos imagens gratuitas para republicação na Figura 1. Para cada imagem, adquirimos 200 rótulos crowdsourced do Microworkers.com, conforme descrito acima.

A Figura 1A mostra uma imagem que pode ser rotulada como zangada ou enojada, e a Figura 1C mostra uma imagem possivelmente zangada, com medo, surpresa ou alguma combinação dos 3. Mais contexto é necessário para alcançar total confiança sobre as verdadeiras classes. A Figura 1B mostra uma emoção composta, onde o indivíduo parece estar "terrivelmente surpreso". Atribuir apenas uma única categoria à imagem seria enganoso. A Figura 1D retrata uma situação em que não está claro se o rosto neutro do indivíduo parece triste ou se esse indivíduo está fazendo uma cara triste. Em casos como esse, provavelmente seria necessário um modelo de reconhecimento de emoções personalizado.

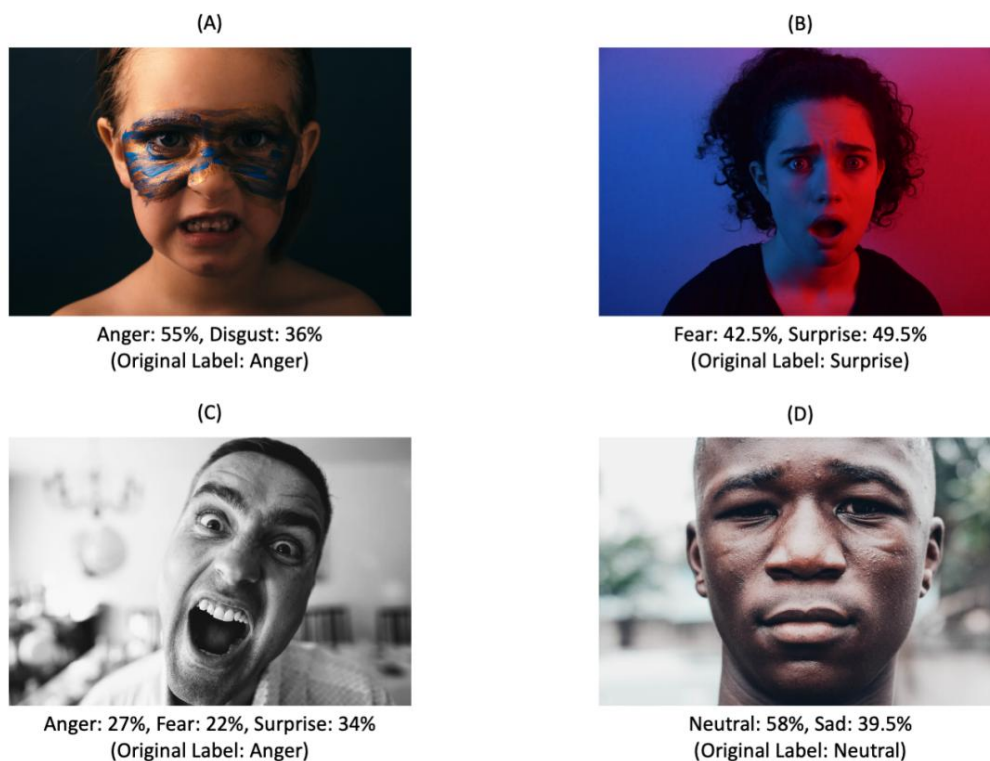


Figura 1: Exemplos de imagens com rótulos de emoção subjetiva. Adquirimos 200 anotações de crowdsourcing para cada imagem acima. As porcentagens de rótulos para cada categoria de emoção são exibidas abaixo da imagem para todas as emoções que receberam pelo menos 10% dos votos. (A) Essa expressão pode ser raiva ou nojo. (B) Esta expressão é uma expressão composta de medo e surpresa. (C) Dependendo do contexto, essa imagem pode ser raiva, medo ou surpresa (ou alguma combinação). (D) Não está claro se esse rosto é neutro ou triste, destacando a necessidade de técnicas personalizadas de reconhecimento de emoções.

Também quantificamos a subjetividade das imagens no CAFE. Medimos o número de imagens com 80% de anotações representadas com os principais  $N$  rótulos mais frequentes para  $N$  variando de 1 a 5 inclusive (histograma na Figura 2). Vemos que, embora muitas emoções não contenham muita subjetividade ( $N = 1$ ), a maioria das imagens é ambígua entre ou composta por 2 ou mais emoções. Quando o ponto de corte é aumentado para 90% (Figura 3), o número de rótulos subjetivos aumenta ainda mais.

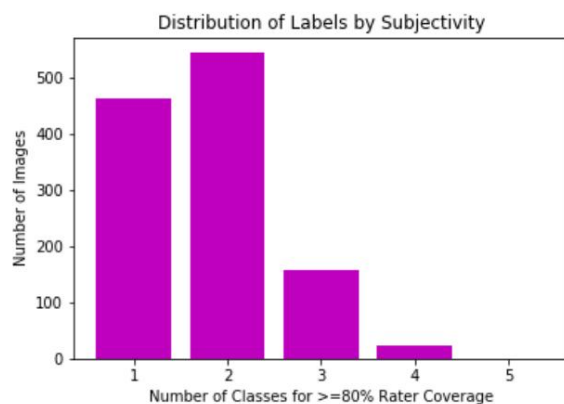


Figura 2: Distribuição dos rótulos por subjetividade. Histograma do número de classes mais votadas necessárias para atingir uma cobertura de avaliador maior ou igual a 80% para cada imagem.

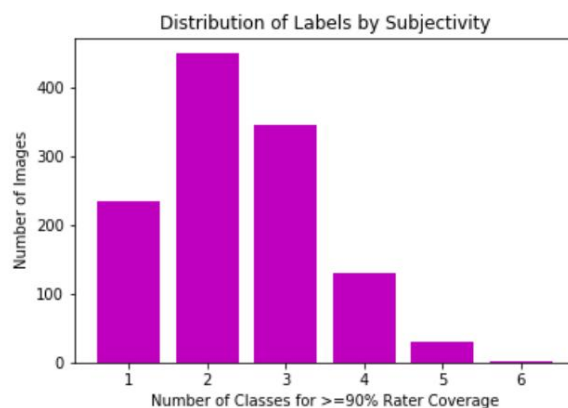


Figura 3: Distribuição dos rótulos por subjetividade. Histograma do número de classes mais votadas necessárias para atingir uma cobertura de avaliador maior ou igual a 90% para cada imagem.

### Comparação de rótulos CAFE e desempenho da multidão

Ao olhar para o rótulo de consenso da maioria, a multidão filtrada concordou com os rótulos CAFE 100% do tempo para feliz, neutro, triste e surpresa. Houve 90% de concordância para desgosto, 75% de concordância para raiva e 50% de concordância para medo. Ao combinar rótulos comumente confusos em uma classe (“raiva + nojo” e “medo + surpresa”), a multidão filtrada concordou com os rótulos CAFE 100% das vezes para feliz, neutro, triste e “medo + surpresa” e 88,8 % para “raiva + nojo”.



Tabela 1: Distribuição do anotador original do CAFE versus distribuição do trabalhador *filtrado* para o sujeito F-AA-15 no CAFE.

Imagem	Distribuição de rotuladoras CAFE	Distribuição de rotuladora de crowdsourcing
	(contar)	(contar)
Raiva, Nojo, Medo, Feliz, Neutro, Triste, Surpreso		
9990-angry_F-AA-15.jpg	30, 37, 15, 8, 0, 8, 2	7, 3, 0, 4, 0, 0, 0
10108-angryopen_F-AA-15.jpg	29, 6, 35, 1, 1, 23, 5	2, 2, 4, 0, 0, 6, 0
10194-disgust_F-AA-15.jpg	3, 86, 3, 2, 1, 5, 0	2, 10, 0, 0, 0, 2, 0
10288-disgustwithtongue_F-AA-15.jpg	3, 91, 0, 3, 2, 0, 1	1, 6, 1, 5, 0, 0, 1
10383-fearful_F-AA-15.jpg	2, 1, 82, 2, 1, 6, 6	0, 1, 10, 0, 0, 0, 3
10461-fearfulopen_F-AA-15.jpg	2, 3, 58, 2, 3, 1, 31	0, 0, 5, 0, 0, 0, 9
10526-happy_F-AA-15.jpg	1, 0, 0, 96, 2, 1, 0	0, 0, 0, 14, 0, 0, 0
10739-neutral_F-AA-15.jpg	1, 0, 1, 1, 89, 7, 1	0, 0, 0, 0, 14, 0, 0
10867-neutralopen_F-AA-15.jpg	2, 2, 10, 1, 33, 0, 52	0, 0, 0, 0, 7, 0, 7
10967-sad_F-AA-15.jpg	3, 3, 6, 1, 2, 85, 0	2, 0, 0, 0, 0, 12, 0
11027-sadopen_F-AA-15.jpg	0, 5, 22, 0, 0, 72, 1	0, 0, 3, 0, 0, 11, 0
11079-surprise_F-AA-15.jpg	1, 0, 23, 0, 2, 0, 74	0, 0, 1, 0, 0, 0, 13

Por outro lado, os trabalhadores de multidão não filtrados não concordaram tão fortemente com os rótulos CAFE ao analisar o consenso da maioria, destacando a necessidade de medidas de controle de qualidade ao fazer crowdsourcing de anotações de emoção. Houve 100% de concordância para surpresa, 93,3% de concordância para feliz, 83,3% de concordância para triste, 76,9% de concordância para desgostoso, 64,3% de concordância para irritado, 61,5% de concordância para neutro e 30,8% de concordância para medo.

Tabela 2: Distribuição do anotador original do CAFE versus distribuição de classe de crowdsourcing *não filtrada* para o sujeito F-AA-93 no CAFE.

Imagem	Distribuição de rotuladoras CAFE	Distribuição de rotuladora de crowdsourcing
	(contar)	(contar)
Raiva, Nojo, Medo, Feliz, Neutro, Triste, Surpreso		
9979-angry_F-AA-03.jpg	89, 4, 0, 0, 0, 4, 3	78, 31, 2, 8, 3, 2, 3
10100-angryopen_F-AA-03.jpg	16, 0, 36, 5, 1, 2, 40	17, 3, 38, 15, 0, 0, 54
10184-disgust_F-AA-03.jpg	17, 41, 2, 12, 19, 8, 1	12, 75, 1, 10, 25, 3, 1
10280-disgustwithtongue_F-AA-03.jpg	19, 77, 0, 0, 2, 1, 1	19, 85, 2, 13, 6, 0, 2
10375-fearful_F-AA-03.jpg	2, 4, 49, 13, 2, 3, 27	10, 12, 27, 15, 16, 3, 44
10454-fearfulopen_F-AA-03.jpg	0, 1, 27, 4, 1, 0, 67	1, 0, 43, 3, 0, 0, 80
10515-happy_F-AA-03.jpg	1, 0, 0, 98, 1, 0, 0	0, 4, 0, 113, 8, 1, 1
10635-happyopen_F-AA-03.jpg	1, 2, 0, 94, 1, 0, 2	0, 0, 0, 126, 0, 0, 1
10730-neutral_F-AA-03.jpg	3, 2, 0, 2, 73, 19, 1	2, 1, 0, 0, 77, 47, 0
10858-neutralopen_F-AA-03.jpg	4, 4, 4, 2, 17, 3, 66	1, 7, 21, 1, 10, 5, 82
10960-sad_F-AA-03.jpg	2, 1, 2, 1, 2, 92, 0	8, 9, 3, 0, 4, 103, 0
11021-sadopen_F-AA-03.jpg	1, 6, 21, 15, 13, 30, 14	9, 21, 24, 14, 12, 43, 4
11068-surprise_F-AA-03.jpg	2, 1, 13, 20, 1, 0, 63	1, 1, 12, 30, 0, 1, 82

As Tabelas 1 e 2 comparam a distribuição de rótulos dos rotuladores CAFE originais, bem como dos trabalhadores de multidão filtrados e não filtrados (respectivamente) para uma única imagem. Em ambos os casos filtrados e não filtrados, as distribuições se espelham qualitativamente em termos de seus picos. Em todos os casos, os picos que aparecem na distribuição do anotador CAFE também aparecem na distribuição da multidão e vice-versa. No entanto, essas distribuições são barulhentas e a relação entre os picos não pode ser garantida (por exemplo, se “raiva” tiver mais rótulos do que “nojo” para anotadores CAFE, “nojo” pode ter mais classificações para anotadores de multidão). As distribuições de multidão geradas devem, portanto, ser consideradas como uma aproximação ruidosa da verdadeira distribuição de probabilidade, e trabalhos adicionais devem levar em conta esse ruído na representação do rótulo.

### **Treinamento e teste com distribuições de probabilidade de multidão**

Avaliamos os modelos com pontuação F1 em vez de precisão porque CAFE não é um conjunto de dados balanceado. Ao treinar com os rótulos codificados one-hot, a pontuação F1 no conjunto de teste retido é de 94,33%. Ressaltamos que esse alto desempenho é enganoso devido à ambiguidade dos rótulos de verdade. Ao treinar com vetores que representam a distribuição de rótulos humanos, a pontuação F1 no conjunto de teste retido é de 78,68%. Enquanto a pontuação F1 é menor ao treinar com rótulos de distribuição humana, a distribuição das previsões de emoção se assemelha muito mais à distribuição de rótulos humanos para o classificador treinado em distribuição. Para muitas aplicações de computação afetiva, ter uma distribuição de rótulos representativa é mais importante do que a precisão absoluta. A distância média L1 entre a distribuição de rótulo humano para o conjunto de teste e o classificador treinado em distribuição é

0,3727 (DP=0,3000); a distância L1 média entre a distribuição de rótulo humano e o classificador treinado por codificação one-hot é 0,6078 (SD=0,4143). A diferença nas distâncias L1 entre esses dois grupos é estatisticamente significativa de acordo com um teste t independente de 2 amostras ( $t=3,2827$ ,  $p=0,0014$ ). Para visualizar essa diferença,

A Figura 4 compara a distribuição da emoção verdadeira com a distribuição emitida de cada um dos dois classificadores para 3 imagens representativas no conjunto de teste com rótulos subjetivos.

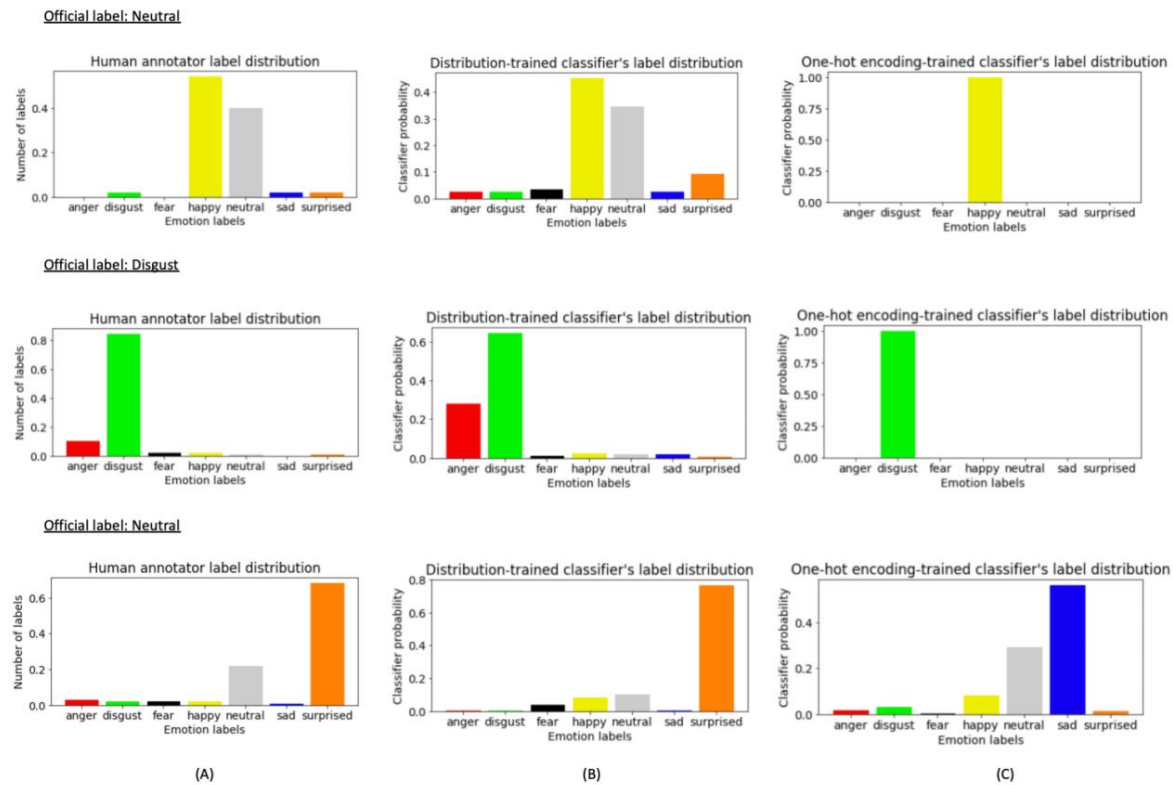


Figura 4: Comparação da distribuição de rótulos humanos (A) e a probabilidade prevista de um classificador treinado com um vetor de probabilidade representando a variedade de rótulos humanos (B) vs. um vetor codificado one-hot para imagens com rótulos de emoção subjetiva (C).

## DISCUSSÃO

Designers de interação e desenvolvedores de sistemas de computação afetiva devem considerar se alvos fáceis ou rígidos são a representação de rótulo mais apropriada para treinar um classificador de visão computacional afetiva para um determinado aplicativo e conjunto de dados. Modelos afetivos de visão computacional que são otimizados para entender a gama potencialmente diversa de interpretações humanas da emoção podem ser usados em várias aplicações de sistemas interativos, como sistemas alimentados por IA que auxiliam indivíduos com autismo e outros atrasos no desenvolvimento [11-12, 14, 29, 35-39, 44, 66, 72, 84-85, 95-96], sistemas de e-learning [41, 78] ou ferramentas de triagem de diagnóstico em casa para condições psiquiátricas [18-20, 30, 47, 50, 76-77, 86-88, 90-94, 97].

Existem várias limitações deste trabalho. Este estudo foi realizado em um único conjunto de dados. Para que esses resultados se generalizem para outros tipos de imagens, inclusive para geração de rótulos soft-target de crowdsourcing em domínios fora do reconhecimento de emoções, outros conjuntos de dados devem ser explorados. Outra limitação é que não registramos ou consideramos possíveis vieses nas etapas de controle de qualidade para filtrar a multidão. Um estudo mais aprofundado sobre como os diferentes mecanismos de qualidade da multidão afetam o resultado seria interessante, pois a qualidade do rótulo de dados pode afetar drasticamente de um

algoritmo de aprendizado de máquina. Uma limitação final é que não tínhamos um método confiável para separar emoções compostas de rótulos ambíguos.

Ao adquirir rótulos de 100 trabalhadores independentes da multidão por imagem, cada um dando seu voto para a emoção mais proeminente, criamos uma representação descrevendo a subjetividade entre-sujeitos da emoção expressa na imagem. Essa representação ofusca notavelmente a subjetividade intra-sujeito, e uma alternativa que deve ser estudada em trabalhos futuros é solicitar que cada avaliador forneça seleções múltiplas por meio de uma escala semântica, como em Korovina et al. [45-46].

## CONCLUSÃO

Para muitas aplicações de computação afetiva, relatar uma distribuição de probabilidade de emoção que explique a subjetividade da interpretação humana pode ser mais importante do que as métricas tradicionais de aprendizado de máquina.

Crowdsourcing é uma solução viável para a aquisição de rótulos soft-target desde um mecanismo de filtragem suficiente para selecionar trabalhadores de multidão confiáveis.

## AGRADECIMENTOS

Este trabalho foi apoiado em parte por fundos para DPW dos Institutos Nacionais de Saúde (1R01EB025025-01, 1R21HD091500-01, 1R01LM013083, 1R01LM013364), National Science Foundation (Prêmio 2014232), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, Weston Havens Foundation e doações de programas do Programa de Inteligência Artificial Centrada no Humano de Stanford, Centro de Diagnóstico Integrado e Saúde de Precisão de Stanford (PHIND), Centro Beckman de Stanford, Centro Bio-X de Stanford, Acelerador de Previsões e Diagnósticos (SPADA) Spectrum, Programa Spark de Stanford em Pesquisa Translacional, Stanford mediaX e Neurociência do Instituto de Neurociências Wu Tsai de Stanford: Programa de tradução. Também agradecemos o apoio generoso de David Orr, Imma Calvo, Bobby DeKesyer e Peter Sullivan. A PW gostaria de agradecer o apoio do Sr. Schroeder e do Stanford Interdisciplinary Graduate Fellowship (SIGF) como o Schroeder Family Goldman Sachs Graduate Fellow.

## CONFORMIDADE COM OS PADRÕES ÉTICOS

Financiamento: Este estudo foi financiado pelo National Institutes of Health (1R01EB025025-01, 1R21HD091500-01, 1R01LM013083, 1R01LM013364), National Science Foundation (Prêmio 2014232), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, a Weston Havens Foundation e doações de programas do Programa de Inteligência Artificial Centrada no Humano de Stanford, Centro de Diagnóstico Integrado e Saúde de Precisão de Stanford (PHIND), Centro Beckman de Stanford, Centro Bio-X de Stanford, Acelerador de Previsões e Diagnósticos (SPADA) Spectrum, Spark de Stanford Programa em Pesquisa Translacional, Stanford mediaX e Neurociência do Instituto de Neurociências Wu Tsai de Stanford: Programa de tradução. Também agradecemos o apoio generoso de David Orr, Imma Calvo, Bobby DeKesyer e Peter Sullivan. A PW gostaria de agradecer o apoio do Sr. Schroeder e do Stanford Interdisciplinary Graduate Fellowship (SIGF) como o Schroeder Family Goldman Sachs Graduate Fellow.

Consentimento informado: O consentimento informado foi obtido de todos os participantes individuais (trabalhadores da multidão em Microworkers.com) incluídos no estudo.

Aprovação ética: Todos os procedimentos realizados em estudos envolvendo participantes humanos estavam de acordo com os padrões éticos da instituição (aprovados pelo Conselho de Revisão Institucional da Universidade de Stanford) e com a declaração de Helsinque de 1964 e suas alterações posteriores ou padrões éticos comparáveis.

## REFERÊNCIAS

- [1] Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin et al. "Tensorflow: Um sistema para aprendizado de máquina em larga escala." No *12º simpósio {USENIX} sobre design e implementação de sistemas operacionais ({OSDI} 16)*, pp. 265-283. 2016.
- [2] Allahbakhsh, Mohammad, Boualem Benatallah, Aleksandar Ignjatovic, Hamid Reza Motahari-Nezhad, Elisa Bertino e Shahram Dustdar. "Controle de qualidade em sistemas de crowdsourcing: questões e direções." *IEEE Internet Computing* 17, não. 2 (2013): 76-81.
- [3] Ando, Atsushi, Satoshi Kobashikawa, Hosana Kamiyama, Ryo Masumura, Yusuke Ijima e Yushi Aono. "Treinamento de alvo suave com enunciados emocionais ambíguos para classificação de emoção de fala baseada em DNN." Em *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4964-4968. IEEE, 2018.
- [4] Buchholz, Sabine e Javier Latorre. "Testes de preferência de crowdsourcing e como detectar trapaças." Na *Décima Segunda Conferência Anual da International Speech Communication Association*. 2011.
- [5] Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay e Antonio Feraco. "Computação afetiva e análise de sentimentos." Em *Um guia prático para análise de sentimentos*, pp. 1-10. Springer, Cham, 2017.
- [6] Chaturvedi, Iiti, Ranjan Satapathy, Sandro Cavallari e Erik Cambria. "Raciocínio de senso comum difuso para análise multimodal de sentimentos." *Cartas de reconhecimento de padrão* 125 (2019): 264-270.
- [7] Chollet, François. "Keras: A biblioteca de aprendizado profundo python." *ascl* (2018): ascl-1806.
- [8] Coolican, Jamesie, Gail A. Eskes, Patricia A. McMullen e Erin Lecky. "Vieses perceptuais no processamento de identidade facial e emoção." *Cérebro e Cognição* 66, não. 2 (2008): 176-187.
- [9] Coren, Stanley e James A. Russell. "O domínio relativo de diferentes expressões faciais de emoção sob condições de ambiguidade perceptiva." *Cognição e Emoção* 6, não. 5 (1992): 339-356.
- [10] Daniel, Florian, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah e Mohammad Allahbakhsh. "Controle de qualidade em crowdsourcing: Um levantamento de atributos de qualidade, técnicas de avaliação e ações de garantia." *ACM Computing Surveys (CSUR)* 51, no. 1 (2018): 1-40.
- [11] Daniels, Jena, Jessey N. Schwartz, Catalin Voss, Nick Haber, Azar Fazel, Aaron Kline, Peter Washington, Carl Feinstein, Terry Winograd e Dennis P. Wall. "Estudo exploratório examinando a viabilidade em casa de uma ferramenta vestível para aprendizagem socioafetiva em crianças com autismo." *NPJ medicina digital* 1, nº. 1 (2018): 1-10.
- [12] Daniels, Jena, Nick Haber, Catalin Voss, Jessey Schwartz, Serena Tamura, Azar Fazel, Aaron Kline et al. "Teste de viabilidade de um wearable ajuda comportamental para aprendizagem social em crianças com autismo." *Informática clínica aplicada* 9, nº 1 (2018): 129.
- [13] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li e Li Fei-Fei. "Imagenet: Um banco de dados de imagens hierárquicas em grande escala." Em *2009, conferência IEEE sobre visão computacional e reconhecimento de padrões*, pp. 248-255. Eee, 2009.
- [14] Deriso, David, Joshua Susskind, Lauren Krieger e Marian Bartlett. "Espelho de emoção: uma nova intervenção para o autismo com base no reconhecimento de expressão em tempo real." Em *Conferência Europeia sobre Visão Computacional*, pp. 671-674. Springer, Berlim, Heidelberg, 2012.
- [15] Devillers, Laurence, Laurence Vidrascu e Lori Lamel. "Desafios na anotação de emoções da vida real e aprendizado de máquina baseado detecção." *Neural Networks* 18, nº 4 (2005): 407-422.
- [16] Dinculescu, Adrian, Andra Bȃltoiu, Carmen Strungaru, Livia Petrescu, Cristian Vizitiu, Alexandru Mandu, Nicoarȃ Talpeȃ e Vlad Vȃleanu. "Identificação automática de marcos faciais antropológicos para detecção de emoções." Em *2019 9ª Conferência Internacional sobre Recentes Avanços em Tecnologias Espaciais (RAST)*, pp. 585-590. IEEE, 2019.
- [17] Du, Shichuan, Yong Tao e Aleix M. Martinez. "Expressões faciais compostas de emoção." *Anais da Academia Nacional de Ciências* 111, n. 15 (2014): E1454-E1462.
- [18] Duda, Marlena, Jena Daniels e Dennis P. Wall. "Avaliação clínica de uma avaliação de risco de autismo nova e móvel." *Jornal de autismo e transtornos do desenvolvimento* 46, não. 6 (2016): 1953-1961.
- [19] Duda, M., N. Haber, J. Daniels e DP Wall. "Validação por crowdsourcing de um sistema de classificação de aprendizado de máquina para autismo e TDAH." *Psiquiatria translacional* 7, não. 5 (2017): e1133-e1133.
- [20] Duda, M., R. Ma, N. Haber e DP Wall. "Uso de aprendizado de máquina para distinção comportamental de autismo e TDAH." *Psiquiatria translacional* 6, não. 2 (2016): e732-e732.
- [21] Ekman, Paul. "Existem emoções básicas?" (1992): 550.
- [22] Ekman, Paul. "Emoções básicas." *Manual de cognição e emoção* 98, nº. 45-60 (1999): 16.
- [23] Eyben, Florian, Martin Wöllmer, Tony Poitschke, Björn Schuller, Christoph Blaschke, Berthold Färber e Nhu Nguyen-Thien. "Emoção na estrada - necessidade, aceitação e viabilidade de computação afetiva no carro." *Avanços na interação humano-computador* 2010 (2010).
- [24] Fan, Yingruo, Jacqueline CK Lam e Victor OK Li. "Rede neural convolucional de conjunto multirregional para expressão facial

- reconhecimento." In *International Conference on Artificial Neural Networks*, pp. 84-94. Springer, Cham, 2018.
- [25] Fang, Xi, Jiancheng Yang e Bingbing Ni. "Refinaria de rótulos estocásticos: rumo a uma melhor distribuição de rótulos de destino." Em *2020 25ª Conferência Internacional sobre Reconhecimento de Padrões (ICPR)*, pp. 9115-9121. IEEE, 2021.
- [26] Fujioka, Takuya, Dario Bertero, Takeshi Homma e Kenji Nagamatsu. "Enfrentando a ambiguidade dos rótulos de emoção através do Meta Learning." *pré-impressão do arXiv arXiv:1911.02216* (2019).
- [27] Gray, Katie LH, Wendy J. Adams, Nicholas Hedger, Kristiana E. Newton e Matthew Garner. "Rostos e consciência: nível baixo, não fatores emocionais determinam a dominância perceptiva." *Emotion* 13, no. 3 (2013): 537.
- [28] Haber, Nick, Catalin Voss, Azar Fazel, Terry Winograd e Dennis P. Wall. "Uma abordagem prática para a subtração de recursos neutros em tempo real para reconhecimento de expressões faciais." Em *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1-9. IEEE, 2016.
- [29] Haber, Nick, Catalin Voss e Dennis Wall. "Tornar as emoções transparentes: o Google Glass ajuda crianças autistas a entender expressões através da terapia de realidade aumentada." *IEEE Spectrum* 57, no. 4 (2020): 46-52.
- [30] Halim, Abbas, Garberson Ford, Stuart Liu-Mayo, Eric Glover e Dennis P. Wall. "Abordagem de IA multimodular para simplificar o autismo Diagnóstico em crianças pequenas." *Scientific Reports (Nature Publisher Group)* 10, nº 1 (2020).
- [31] Ele, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun. "Aprendizagem residual profunda para reconhecimento de imagem." Nos *Anais do IEEE conferência sobre visão computacional e reconhecimento de padrões*, pp. 770-778. 2016.
- [32] Hirth, Matthias, Tobias Hofßfeld e Phuoc Tran-Gia. "Anatomia de uma plataforma de crowdsourcing - usando o exemplo de microworkers. com." Em *2011 Quinta conferência internacional sobre serviços móveis e de internet inovadores em computação ubíqua*, pp. 322-329. IEEE, 2011.
- [33] Hupont, Isabelle, Sandra Baldassarri, Eva Cerezo e Rafael Del-Hoyo. "Visualização de Afeto Humano Avançada." Em *2013 IEEE Conferência Internacional sobre Sistemas, Homem e Cibernética*, pp. 2700-2705. IEEE, 2013.
- [34] Jerauld, Roberto. "Sistema de detecção e feedback de emoções vestíveis." Patente dos EUA 9.019.174, emitida em 28 de abril de 2015.
- [35] Kalantarian, Haik, Khaled Jedoui, Peter Washington e Dennis P. Wall. "Um jogo para celular para rotulagem automática de emoções em imagens." *IEEE transações em jogos* 12, não. 2 (2018): 213-218.
- [36] Kalantarian, Haik, Khaled Jedoui, Peter Washington, Qandeel Tariq, Kaiti Dunlap, Jessey Schwartz e Dennis P. Wall. "Marcar imagens com emoção facial e o potencial para a saúde pediátrica." *Inteligência artificial na medicina* 98 (2019): 77-86.
- [37] Kalantarian, Haik, Khaled Jedoui, Kaitlyn Dunlap, Jessey Schwartz, Peter Washington, Arman Husic, Qandeel Tariq, Michael Ning, Aaron Kline e Dennis Paul Wall. "O desempenho de classificadores de emoção para crianças com autismo relatado pelos pais: estudo de viabilidade quantitativa." *JMIR saúde mental* 7, não. 4 (2020): e13174.
- [38] Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber e Dennis P. Wall. "Adivinha?." *Diário de pesquisa em informática em saúde* 3, não. 1 (2019): 43-66.
- [39] Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber e Dennis Wall. "Um sistema móvel gamificado para vídeo de crowdsourcing para pesquisa de autismo." Em *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 350-352. IEEE, 2018.
- [40] Kairam, Sanjay e Jeffrey Heer. "Partindo multidões: caracterizando interpretações divergentes em tarefas de anotação de crowdsourcing." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1637-1648. 2016.
- [41] Kaiser, Robin e Karina Oertel. "Emoções em HCI: um sistema de e-learning afetivo." Nos *Anais do workshop HCSNet sobre Uso de visão na interação humano-computador-Volume 56*, pp. 105-106. Australian Computer Society, Inc., 2006.
- [42] Kaur, Ramandeep e Sandeep Kautish. "Análise de sentimento multimodal: uma pesquisa e comparação." *Jornal Internacional de Serviço Ciência, Gestão, Engenharia e Tecnologia (IJSMET)* 10, nº. 2 (2019): 38-58.
- [43] Kingma, Diederik P. e Jimmy Ba. "Adam: Um método para otimização estocástica." *arXiv pré-impressão arXiv:1412.6980* (2014).
- [44] Kline, Aaron, Catalin Voss, Peter Washington, Nick Haber, Hessey Schwartz, Qandeel Tariq, Terry Winograd, Carl Feinstein e Dennis P. Parede. "Vidro superpoderoso." *GetMobile: Computação e Comunicações Móveis* 23, nº. 2 (2019): 35-38.
- [45] Korovina, Olga, Marcos Baez e Fabio Casati. "Confiabilidade do crowdsourcing como método para coletar rótulos de emoções em fotos." *Notas de pesquisa da BMC* 12, nº 1 (2019): 1-6.
- [46] Korovina, Olga, Fabio Casati, Radoslaw Nielek, Marcos Baez e Olga Berestneva. "Investigando crowdsourcing como um método para coletar rótulos de emoção para imagens." In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1-6. 2018.
- [47] Kosmicki, JA, V. Sochat, M. Duda e DP Wall. "Procurando um conjunto mínimo de comportamentos para detecção de autismo por meio de aprendizado de máquina baseado em seleção de recursos." *Psiquiatria translacional* 5, não. 2 (2015): e514-e514.
- [48] Kratzwald, Bernhard, Suzana Ilić, Mathias Kraus, Stefan Feuerriegel e Helmut Prendinger. "Aprendizado profundo para computação afetiva: reconhecimento de emoção baseado em texto no suporte à decisão." *Sistemas de Apoio à Decisão* 115 (2018): 24-35.
- [49] Arrendamento, Mateus. "Sobre controle de qualidade e aprendizado de máquina em crowdsourcing." *Computação Humana* 11, não. 11 (2011).
- [50] Leblanc, Emilie, Peter Washington, Maya Varma, Kaitlyn Dunlap, Yordan Penev, Aaron Kline e Dennis P. Wall. "Os métodos de substituição de recursos permitem uma análise confiável de vídeo doméstico para detecção de autismo por aprendizado de máquina." *Relatórios científicos* 10, n. 1 (2020): 1-11.
- [51] Liu, Rumpeng, Joseph P. Salisbury, Arshya Vahabzadeh e Ned T. Sahin. "Viabilidade de uma realidade aumentada focada no autismo sistema de smartglasses para comunicação social e coaching comportamental." *Frontiers in Pediatrics* 5 (2017): 145.
- [52] LoBue, Vanessa, Lewis Baker e Cat Thrasher. "Através dos olhos de uma criança: identificação pré-escolar de expressões emocionais do conjunto de expressão facial afetiva da criança (CAFE)". *Cognição e Emoção* 32, nº. 5 (2018): 1122-1130.

- [53] LoBue, Vanessa e Cat Thrasher. "O conjunto de Expressão Facial Afetiva da Criança (CAFE): Validade e confiabilidade de não treinados adultos." *Frontiers in Psychology* 5 (2015): 1532.
- [54] Lotfian, Reza e Carlos Busso. "Amostragem excessiva de dados de fala emocional com base em avaliações subjetivas fornecidas por múltiplos indivíduos." *IEEE Transactions on Affective Computing* (2019).
- [55] Magdin, Martin e F. Priker. "Reconhecimento de expressões faciais em tempo real usando webcam e SDK afetiva." *IJIMA* 5, não. 1 (2018): 7-15.
- [56] McDuff, Daniel, Rana Kaliouby, Thibaud Senechal, May Amr, Jeffrey Cohn e Rosalind Picard. "Conjunto de dados de expressão facial Affective-mit (am-fed): expressões faciais naturalistas e espontâneas coletadas." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 881-888. 2013.
- [57] McFarland, Dennis J., Muhammad A. Parvaz, William A. Sarnacki, Rita Z. Goldstein e Jonathan R. Wolpaw. "Previsão de subjetividade classificações de imagens emocionais por características de EEG." *Journal of neural engineering* 14, no. 1 (2016): 016009.
- [58] Mower, Emily, Maja J. Matarić e Shrikanth Narayanan. "Uma estrutura para classificação automática de emoções humanas usando profiles." *IEEE Transactions on Audio, Speech, and Language Processing* 19, nº 5 (2010): 1057-1070.
- [59] Mower, Emily, Angeliki Metallinou, Chi-Chun Lee, Abe Kazemzadeh, Carlos Busso, Sungbok Lee e Shrikanth Narayanan. "Interpretando expressões emocionais ambíguas." Em *2009 3ª Conferência Internacional sobre Computação Afetiva e Interação Inteligente e Workshops*, pp. 1-8. IEEE, 2009.
- [60] Nguyen, Binh T., Minh H. Trinh, Tan V. Phan e Hien D. Nguyen. "Uma detecção eficiente de emoções em tempo real usando câmeras e pontos de referência faciais." Em *2017 sétima conferência internacional de ciência e tecnologia da informação (ICIST)*, pp. 251-255. IEEE, 2017.
- [61] Ning, Michael, Jena Daniels, Jessey Schwartz, Kaitlyn Dunlap, Peter Washington, Haik Kalantarian, Michael Du e Dennis P. Wall. "Identificação e quantificação de lacunas no acesso a recursos de autismo nos Estados Unidos: um estudo infodemiológico." *Jornal de pesquisa médica na Internet* 21, não. 7 (2019): e13094.
- [62] Nicolaou, Mihalís A., Hatice Gunes e Maja Pantic. "A previsão contínua de afeto espontâneo a partir de múltiplas pistas e modalidades em espaço de excitação de valência." *IEEE Transactions on Affective Computing* 2, no. 2 (2011): 92-105.
- [63] Paolacci, Gabriele, Jesse Chandler e Panagiotis G. Ipeirotis. "Executando experimentos no Amazon Turk Mecânico." *Julgamento e Tomada de Decisão* 5, n. 5 (2010): 411-419.
- [64] Parthasarathy, Srinivas e Carlos Busso. "Previsão conjunta de excitação, valência e dominância com aprendizagem multitarefa." Em *Interfala*, volume 2017, págs. 1103-1107. 2017.
- [65] Peterson, Joshua C., Ruairidh M. Battleday, Thomas L. Griffiths e Olga Russakovsky. "A incerteza humana torna a classificação mais robusto." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9617-9626. 2019.
- [66] Poggia, Giovanni, Roberta Iglozzi, Marcello Ferro, Arti Ahluwalia, Filippo Muratori e Danilo De Rossi. "Um andróide para melhorar as habilidades sociais e o reconhecimento de emoções em pessoas com autismo." *Transações IEEE em Sistemas Neurais e Engenharia de Reabilitação* 13, no. 4 (2005): 507-515.
- [67] Poria, Soujanya, Erik Cambria e Alexander Gelbukh. "Recursos textuais de rede neural convolucional profunda e aprendizado de vários kernels para análise de sentimento multimodal em nível de expressão." Em *Anais da conferência de 2015 sobre métodos empíricos em processamento de linguagem natural*, pp. 2539-2544. 2015.
- [68] Poria, Soujanya, Erik Cambria, Newton Howard, Guang-Bin Huang e Amir Hussain. "Fusão de pistas de áudio, visuais e textuais para análise de sentimento a partir de conteúdo multimodal." *Neurocomputing* 174 (2016): 50-59.
- [69] Rizos, Georgios e Björn W. Schuller. "Average Jane, Where Art Thou? - Recentes avenidas em aprendizado de máquina eficiente sob incerteza de subjetividade." In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 42-55. Springer, Cham, 2020.
- [70] Rodrigues, Filipe e Francisco Pereira. "Aprendizagem profunda com multidões." Em *Anais da Conferência AAAI sobre Inteligência Artificial*, volume 32, não. 1. 2018.
- [71] Sharma, Mukta, Anand Singh Jalal e Aamir Khan. "Reconhecimento de emoções usando expressão facial fundindo descritor de pontos-chave e recursos de textura." *Ferramentas e aplicativos multimídia* 78, no. 12 (2019): 16195-16219.
- [72] Smitha, Kavallur Gopi e A. Prasad Vinod. "Sistema de reconhecimento de emoções faciais para crianças autistas: um estudo viável baseado em FPGA implementação." *Engenharia médica e biológica e computação* 53, nº 11 (2015): 1221-1229.
- [73] Stappen, Lukas, Alice Baird, Erik Cambria e Björn W. Schuller. "Análise de sentimentos e reconhecimento de tópicos em transcrições de vídeo." *IEEE Sistemas Inteligentes* 36, não. 2 (2021): 88-95.
- [74] Tahir, Madiha, Abdallah Tubaishat, Feras Al-Obeidat, Babar Shah, Zahid Halim e Muhammad Waqas. "Um novo algoritmo genético caótico binário para seleção de recursos e sua utilidade na computação afetiva e na saúde." *Computação Neural e Aplicações* (2020): 1-22.
- [75] Tao, Jianhua e Tieniu Tan. "Computação afetiva: Uma revisão." Em *Conferência Internacional sobre Computação Afetiva e Interação Inteligente*, pp. 981-995. Springer, Berlim, Heidelberg, 2005.
- [76] Tariq, Qandeel, Scott Lanyon Fleming, Jessey Nicole Schwartz, Kaitlyn Dunlap, Conor Corbin, Peter Washington, Haik Kalantarian, Naila Z. Khan, Gary L. Darmstadt e Dennis Paul Wall. "Detecção de atraso no desenvolvimento e autismo por meio de modelos de aprendizado de máquina usando vídeos caseiros de crianças de Bangladesh: estudo de desenvolvimento e validação." *Jornal de pesquisa médica na Internet* 21, não. 4 (2019): e13822.
- [77] Tariq, Qandeel, Jena Daniels, Jessey Nicole Schwartz, Peter Washington, Haik Kalantarian e Dennis Paul Wall. "Detecção móvel de autismo por meio de aprendizado de máquina em vídeo caseiro: um estudo de desenvolvimento e validação prospectiva." *PLoS medicina* 15, não. 11 (2018): e1002705.

- [78] Thiam, Patrick, Sascha Meudt, Markus Kächele, Günther Palm e Friedhelm Schwenker. "Detecção de eventos emocionais utilizando métodos de vetor de suporte em um cenário de HCI de aprendizado ativo." Em *Anais do workshop de 2014 sobre representação e modelagem de emoções em sistemas de interação humano-computador*, pp. 31-36. 2014.
- [79] Thiel, Christian. "A classificação em rótulos macios é robusta contra o ruído do rótulo." Na *Conferência Internacional sobre Conhecimento e Sistemas Inteligentes de Informação e Engenharia*, pp. 65-73. Springer, Berlim, Heidelberg, 2008.
- [80] Turing, Alan M. "Máquinas de computação e inteligência." Em *Analisando o teste de turing*, pp. 23-65. Springer, Dordrecht, 2009.
- [81] Uma, Alexandra, Tommaso Fornaciari, Dirk Hovy, Silviu Paun, Barbara Plank e Massimo Poesio. "Um caso para funções de perda suave." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 8, não. 1, pp. 173-177. 2020.
- [82] Villon, Olivier e Christine Lisetti. "Para reconhecer a emoção subjetiva do indivíduo a partir de sinais fisiológicos na aplicação prática." In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, pp. 357-362. IEEE, 2007.
- [83] Völkel, Sarah Theres, Julia Graefe, Ramona Schödel, Renate Häuslschmid, Clemens Stachl, Quay Au e Heinrich Hussmann. "Eu Dirijo Meu Carro e Meus Estados Me Conduzem: Visualizando os Estados Emocionais e Físicos do Motorista." In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 198-203. 2018.
- [84] Voss, Catalin, Jessey Schwartz, Jena Daniels, Aaron Kline, Nick Haber, Peter Washington, Qandeel Tariq et al. "Efeito da intervenção digital vestível para melhorar a socialização em crianças com transtorno do espectro do autismo: um ensaio clínico randomizado." *JAMA Pediatrics* 173, no. 5 (2019): 446-454.
- [85] Voss, Catalin, Peter Washington, Nick Haber, Aaron Kline, Jena Daniels, Azar Fazel, Titas De et al. "Vidro superpotente: fornecendo dicas sociais discretas em tempo real em sistemas vestíveis." In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 1218-1226. 2016.
- [86] Wall, Dennis Paul, J. Kosmicki, TF Deluca, E. Harstad e Vincent Alfred Fusaro. "Uso de aprendizado de máquina para encurtar a observação triagem baseada e diagnóstico de autismo." *Translacional psiquiatria* 2, no. 4 (2012): e100-e100.
- [87] Washington, Peter, Emilie Leblanc, Kaitlyn Dunlap, Yordan Penev, Aaron Kline, Kelley Paskov, Min Woo Sun et al. "Telemedicina de precisão por meio de aprendizado de máquina de crowdsourcing: testando a variabilidade de trabalhadores em massa para reconhecimento de recursos de autismo baseado em vídeo." *Revista de medicina personalizada* 10, não. 3 (2020): 86.
- [88] Washington, Peter, Natalie Park, Parishkrita Srivastava, Catalin Voss, Aaron Kline, Maya Varma, Qandeel Tariq et al. "Diagnóstico orientado por dados e o potencial da inteligência artificial móvel para fenotipagem terapêutica digital em psiquiatria computacional". *Psiquiatria Biológica: Neurociência Cognitiva e Neuroimagem* (2019).
- [89] Washington, Peter, Haik Kalantarian, Jack Kent, Arman Husic, Aaron Kline, Emilie Leblanc, Cathy Hou et al. "Treinando um classificador de detecção de emoções usando quadros de um jogo terapêutico móvel para crianças com transtornos do desenvolvimento." *arXiv pré-impressão arXiv:2012.08678* (2020).
- [90] Washington, Peter, Haik Kalantarian, Qandeel Tariq, Jessey Schwartz, Kaitlyn Dunlap, Brianna Chrisman, Maya Varma et al. "Validade da triagem online para autismo: estudo de crowdsourcing comparando tarefas de diagnóstico pagas e não pagas." *Jornal de pesquisa médica na Internet* 21, não. 5 (2019): e13668.
- [91] Washington, Peter, Aaron Kline, Onur Cezmi Mutlu, Emilie Leblanc, Cathy Hou, Nate Stockham, Kelley Paskov, Brianna Chrisman e Dennis P. Wall. "Reconhecimento de atividade com câmeras em movimento e alguns exemplos de treinamento: aplicativos para detecção de headbanging relacionado ao autismo." *arXiv pré-impressão arXiv:2101.03478* (2021).
- [92] Washington, Peter, Emilie Leblanc, Kaitlyn Dunlap, Yordan Penev, Maya Varma, Jae-Yoon Jung, Brianna Chrisman et al. "Seleção de trabalhadores de multidão confiáveis para diagnóstico telemédico de transtorno do espectro do autismo pediátrico." In *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*, pp. 14-25. 2020.
- [93] Washington, Peter, Kelley Marie Paskov, Haik Kalantarian, Nathaniel Stockham, Catalin Voss, Aaron Kline, Ritik Patnaik et al. "Característica seleção e redução de dimensão de dados de autismo social." In *Pac Symp Biocomput*, vol. 25, pp. 707-718. 2020.
- [94] Washington, Peter, Qandeel Tariq, Emilie Leblanc, Brianna Chrisman, Kaitlyn Dunlap, Aaron Kline, Haik Kalantarian et al. "Marcação de recursos preservados por privacidade de crowdsourcing de vídeos caseiros curtos para detecção de ASD de aprendizado de máquina." *Relatórios científicos* 11, n. 1 (2021): 1-11.
- [95] Washington, Peter, Catalin Voss, Nick Haber, Serena Tanaka, Jena Daniels, Carl Feinstein, Terry Winograd e Dennis Wall. "UMA auxílio de interação social vestível para crianças com autismo." In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2348-2354. 2016.
- [96] Washington, Peter, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titas De, Carl Feinstein, Terry Winograd e Dennis Wall. "SuperpowerGlass: uma ajuda vestível para a terapia em casa de crianças com autismo." *Anais do ACM sobre tecnologias interativas, móveis, vestíveis e onipresentes* 1, no. 3 (2017): 1-22.
- [97] Washington, Peter, Serena Yeung, Bethany Percha, Nicholas Tatonetti, Jan Liphardt e Dennis P. Wall. "Conquistando confiança soluções de dados biomédicos." In *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*, pp. 1-13. 2020.
- [98] White, Susan W., Lynn Abbott, Andrea Trubanova Wieckowski, Nicole N. Capriola-Hall, Sherin Aly e Amira Youssef. "Viabilidade de treinamento automatizado para expressão de emoção facial e reconhecimento no autismo." *Terapia comportamental* 49, não. 6 (2018): 881-888.
- [99] Yang, Zhao, Tie Liu, Jiehao Liu, Li Wang e Sai Zhao. "Uma nova função de perda de margem suave para incorporação discriminativa profunda Aprendizagem." *IEEE Access* 8 (2020): 202785-202794.
- [100] Yannakakis, Georgios N. "Aprimorando os cuidados de saúde através da computação afetiva." (2018).
- [101] Yin, Da, Xiao Liu, Xiuyu Wu e Baobao Chang. "Uma estratégia de soft label para classificação de sentimento em nível de destino." Em *Anais do Décimo Workshop sobre Abordagens Computacionais à Subjetividade, Sentimento e Análise de Mídias Sociais*, pp. 6-15. 2019.



- [102] Yu, Liang-Chih, Jin Wang, K. Robert Lai e Xue-jie Zhang. "Prever classificações de excitação de valência de palavras usando um método de gráfico ponderado." In *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 788-793. 2015.
- [103] Zeng, Zhihong, Jilin Tu, Ming Liu, Thomas S. Huang, Brian Pianfetti, Dan Roth e Stephen Levinson. "Efeito audiovisual reconhecimento." *IEEE Transactions on multimedia* 9, no. 2 (2007): 424-428.
- [104] Zhang, Lin, Steffen Walter, Xueyao Ma, Philipp Werner, Ayoub Al-Hamadi, Harald C. Traue e Sascha Gruss. "'BioVid Emo DB': Um banco de dados multimodal para análises de emoções validado por classificações subjetivas." Em *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-6. IEEE, 2016.
- [105] Zhang, Xiao, Wenzhong Li, Xu Chen e Sanglu Lu. "Moodexplorer: Rumo à detecção de emoções compostas através da detecção de smartphones." *Anais do ACM sobre Tecnologias Interativas, Móveis, Vestíveis e Ubíquas* 1, no. 4 (2018): 1-30.
- [106] Zhao, Sicheng, Hongxun Yao e Xiaolei Jiang. "Prevendo distribuição de probabilidade contínua de emoções de imagem no espaço de excitação de valência." In *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 879-882. 2015.
- [107] Zhou, Ying, Xuefeng Liang, Yu Gu, Yifei Yin e Longshan Yao. "Aprendizagem interativa multiclassificador para emoção de fala ambígua Reconhecimento." *pré-impressão do arXiv arXiv:2012.05429* (2020).