

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/334150667>

# Reconhecimento de emoções em imagens utilizando técnicas de construção e otimização em métodos ensembles baseados em árvores de decisão

Article · January 2019

CITATIONS

0

READS

16

4 authors, including:



**Hermino Barbosa de Freitas Júnior**

Universidade Federal de Roraima (UFRR)

1 PUBLICATION 0 CITATIONS

SEE PROFILE



**Filipe Dwan Pereira**

Universidade Federal de Roraima (UFRR)

6 PUBLICATIONS 5 CITATIONS

SEE PROFILE



**Andre Luiz da Silva Pereira**

Universidade Federal de Roraima (UFRR)

1 PUBLICATION 0 CITATIONS

SEE PROFILE



## Reconhecimento de emoções em imagens utilizando técnicas de construção e otimização em métodos ensembles baseados em árvores de decisão.

Hermínio B. de F. Junior, Filipe Dwan Pereira, Andre Luiz da S. Pereira, Luciano F. Silva

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Roraima (UFRR)  
Av. Cap. Ene Garcês, 2413 - Aeroporto, Boa Vista - RR, 69310-000

{heminojunior, andr234ibatera}@gmail.com, {filipe.dwan, luciano.silva}@ufrr.br

**Abstract.** *This paper shows a comparative study of machine learning techniques to recognize emotions in images through the characterization of facial expressions. A well reputed research shows that the human beings emotions could be characterized using a Facial Action Coding System (FACS), which is a taxonomy to classify the human being facial expressions. The same study presented a discrete set of universal emotions that a human being could feel, which are: anger, contempt, disgust, fear, happy, sadness and surprise. We use these FACS to recognize the universal emotions using ensemble methods in images from the Cohn Kanade dataset and lastly in a variation of the same dataset (augmented data set), in which we added some rotation and transformations on the images in order to make them less laboratorial. To conduct the experiments, we used two scenarios. At the first one, only the images of happiness and sadness were submitted to the methods. At the second one, we used all the universal emotions. As a result, we reach up to 93% of accuracy with the original data set and 98% with the augmented data set. At the second scenario the average accuracy was 75% with the original data set and 77% with the augmented data set. The best results were achieved using the ensemble method Gradient Tree Boosting with a random search hyperparameter optimization.*

**Resumo.** *Este artigo realiza um estudo comparativo de técnicas de aprendizado de máquina, especificamente métodos ensembles, para reconhecer emoções em imagens através da caracterização de expressões faciais. Emoções humanas podem ser caracterizadas usando um sistema de codificação de movimentos faciais (FACS), que é uma taxonomia para classificar as expressões faciais humanas. Além disso, existe um conjunto discreto de emoções universais que um ser humano pode sentir, que são: raiva, desprezo, nojo, medo, alegria, tristeza e surpresa. Usamos o FACS para reconhecer as emoções universais usando métodos conjuntos em imagens da base de dados Cohn Kanade e, por fim, em uma variação da mesma base de dados (base de dados aumentada), no qual adicionamos algumas rotações e transformações nas imagens para torná-las menos laboratoriais. Para conduzir os experimentos, usamos dois cenários. No primeiro, apenas as imagens de alegria e tristeza foram submetidas aos métodos. No segundo, usamos todas as emoções universais. No final, o melhor resultado foi obtido com o método ensemble Gradient Tree Boosting, alcançando até 93% de acurácia com a base de dados original e 98% com a base de dados aumentada. No segundo cenário, a acurácia média foi de 75% com a base de*

*dados original e 77% com a base de dados aumentada. O otimização foi realizada empregando a técnica de random search para executar a otimização de hiperparâmetros dos pipelines de aprendizado de máquina.*

## 1. Introdução

Expressões faciais são movimentos voluntários ou não, realizados pelos músculos da face a fim de representar um estado emocional (Tomkins 1980). São identificadas seis expressões faciais universais, sendo elas representações através das seguintes emoções: alegria; tristeza; raiva; medo; surpresa e nojo (Ekman 1989). Suas existências foram notadas em pessoas de diferentes culturas modernas e em tribos primitivas. Reconhecer automaticamente tais emoções universais pode ser útil em diversas áreas para a compreensão do que está acontecendo no: processo educacional de uma pessoa; ambiente de entretenimento; circuito de câmeras de segurança; no *e-commerce*, através da recomendação de produtos que mais agrada o usuário; na psicologia, com objetivo de entender o comportamento dos indivíduos; etc.

Para realizar o reconhecimento automático dessas emoções universais, faz-se necessário que o computador seja capaz de reconhecer o estado emocional de um ser humano (Dantas et al. 2015). Uma alternativa para realizar essa tarefa é fazendo uso de algoritmos de aprendizagem de máquina (AM) (Melo et al. 2014). Com a AM podemos inferir dados das emoções geradas por uma determinada pessoa e diante disso podemos analisar cada característica da expressão facial. Os métodos *ensembles*, são exemplos de abordagens amplamente usadas em AM, pois os mesmos utilizam a manipulação de classificadores individuais, cujas saídas é a combinação ou a fusão dos resultados gerados por esses classificadores (Opitz and Maclin 1999). Ainda nesses métodos podemos realizar os ajustes de hiperparâmetros afim de otimizar a acurácia dos modelos preditivo (Dwan et al. 2017).

Dessa forma, visando contribuir com essa temática, este estudo objetiva realizar comparações de técnicas de construção e otimização de *pipelines*<sup>1</sup> de AM utilizando *ensembles* baseadas em Árvores de Decisão (AD) e *random search* (Bergstra and Bengio 2012) para otimização de hiperparâmetros, com suporte de estudos da teoria *Facial Action Coding System* (FACS) (Ekman and Rosenberg 1997, Ekman et al. 1978). Segundo Ekman and Rosenberg 1997, todos sentem emoções e as expressam em vários níveis, principalmente em termos faciais, o que nos permite perceber, por exemplo, o que outros estão sentindo. Ainda de acordo com Ekman and Rosenberg 1997, as pessoas quando experienciam uma emoção traduzem a mesma através de uma expressão facial e assim sendo, pode-se concluir que as expressões faciais podem ser utilizadas para a identificação de emoções.

Com base nessa premissa, este trabalho realiza uma análise comparativa de algoritmos de AM aplicados sobre uma base de dados chamada Cohn Kanade (CK), fundamentada nas FACS. Além disso, aborda-se um segundo conjunto de dados aumentados com variações na CK como rotações e transformações nas imagens para torná-las menos laboratoriais. Os experimentos para detecção das emoções universais foram conduzidos

<sup>1</sup>Em AM, *pipeline* é o conjunto de etapas para a construção de modelos preditivos. Tais etapas vão desde a obtenção e preparação dos dados a otimização e validação do algoritmo de AM para a construção do modelo preditivo.

em duas frentes: Abordagem 1 (A1), que consiste na análise dos dados das emoções alegria e tristeza; Abordagem 2 (A2), que utiliza os dados de todas as emoções básicas derivadas do estudo de Ekman et al. 1978, porém com a expressão desprezo incluída. Essa divisão foi adotada na prática com o intuito de analisar emoções mais usuais na abordagem A1 e emoções mais complexas na abordagem A2, isto é, raiva, desprezo, nojo, medo, alegria, tristeza e surpresa.

No decorrer da pesquisa a técnica que obteve o melhor resultado nos testes realizados neste estudo foi a *Gradient Tree Boosting*, atingindo, na A1, 93% de acurácia aplicada a base de dados CK e 98% no conjunto de dados aumentados. Já na A2 a técnica alcançou 75% de precisão aplicado à CK e 77% com o conjunto de dados aumentados. Esses resultados são fundamentados e discutidos com maiores detalhes nas próximas seções.

Para isso o estudo foi dividido em 6 seções, sendo que: a seção 2 apresenta fundamentação teórica da pesquisa; a seção 3 aborda os trabalhos relacionados; a seção 4 explica o método utilizado; a seção 5 demonstra os resultados e discussões; e finalmente a seção 6 apresenta as considerações finais da pesquisa.

## 2. Fundamentação teórica

Esta seção apresenta as principais partes da base conceitual utilizadas para realização deste trabalho. Os conceitos abordados foram: expressões faciais e suas ligações com as emoções e a codificação utilizada para a identificação das mesmas; aprendizagem de máquina, juntamente com aprendizagem supervisionada; e métodos ensemble e otimização de hiperparâmetros.

### 2.1. Facial Action Coding System

Em um experimento envolvendo americanos e japoneses, foi possível notar que expressões faciais são fontes de dados fidedignas para identificação do estado emocional de um indivíduo (Ekman and Oster 1979). Depois da identificação das seis expressões faciais universais e de testar suas validades, Paul Ekman desenvolveu o *Facial Action Coding System* ou FACS, que é um sistema de codificações da face humana, criado com o objetivo de mapear as ações musculares faciais (Ekman et al. 1978). Ações estas chamadas de Action Units (AU), que podem ser frutos da manifestação fisiológica do estado emocional de um indivíduo. No FACS estão descritas cerca de 98 AUs. Quando a face realiza um conjunto específico dessas AUs é possível inferir qual emoção está sendo representada (Ekman and Rosenberg 1997). Os conjuntos necessários para realizar o reconhecimento das seis emoções universais estão dispostos na Tabela 1 e a descrição dos mesmos encontra-se na Tabela 2.

A análise das AUs segue uma linha temporal. Como estado inicial é utilizado um estado neutro como mostrado na Figura 1(a), onde os requisitos presentes nos conjuntos que representam uma emoção não são supridos. Apartir deste ponto as deformidades resultantes das ações musculares na face são observadas, no momento em que um dado conjunto possui seus requisitos supridos a inferência da emoção manifestada pode ser realizada, como exemplificado na Figura 1(h). Ao término do processo a face tende a tornar ao ponto inicial.

<b>Raiva</b>	<b>Tristeza</b>	<b>Medo</b>
{4,5,7,10,22,23,25 ou 26}	{1,4}	{1,2,4}, {5,20}, {5}
{4,5,7,10,23,25 ou 26}	{1,4,11 ou 15}	{1,2,4,5,20,25 ou 26 ou 27}
{4,5,7,17,23 ou 24}	{1,4,15,17}	{1,2,4,5,25 ou 26 ou 27}
{4,5,7,23 ou 24}	{6,15}	{1,2,4,5}
{4,5 ou 7}, {17,24}	{11,17}	{1,2,5,25 ou 26 ou 27}
	{1}	{5,20,25 ou 26 ou 27}
<b>Alegria</b>	<b>Surpresa</b>	<b>Aversão</b>
{12}	{1,2,5,26 ou 27}	{9 ou 10,17}
{6,12}	{1,2,5}	{9 ou 10}
	{1,2,26 ou 27}	{9 ou 10,16,25 ou 26}
	{5,26 ou 27}	

**Tabela 1. Mapeamento das UAs para reconhecimento de emoções.**  
Fonte:(Libralon 2014)

<b>AU</b>	<b>Descrição</b>
1	Levantar da parte interna das sobrancelhas
2	Levantar da parte externa das sobrancelhas
4	Abaixar das Sobrancelhas
5	Levantar das pálpebras superiores
6	Levantar das bochechas
7	Tensionar das pálpebras
9	Franzir do nariz
10	Levantar dos lábios superiores
11	Aprofundar do nasolabial
12	Puxar dos cantos labiais
13	<i>Cheek Puffer</i>
14	<i>Dimpler</i> - Depressão na maçã do rosto
15	Depressão dos cantos labiais
16	Depressão dos lábios inferiores
17	Levantar do queixo
18	Franzir dos lábios
20	Esticar dos lábios
22	Canalizar dos lábios
23	Tensionar os lábios
24	Pressionar os lábios
25	Separação dos lábios
26	Derrubar do queixo
27	Espalhar da boca

**Tabela 2. Parcela da tabela de descrição das AUs e músculos da face humana correspondentes. Adaptada:(Robotics Institute 2002)**

## 2.2. Aprendizagem de Máquina

Aprendizagem de máquina é uma subárea da computação que busca tornar o computador capaz de assimilar informações que não foram explicitamente apresentadas



**Figura 1.** Sequência de imagens disponíveis na *Cohn-Kanade AU-Coded Expression Database* de uma pessoa expressando surpresa. Fonte: (Kanade et al. 2000) e (Lucey et al. 2010)

(Samuel 1959). Para tal, existem duas abordagens que podem ser utilizadas, a aprendizagem supervisionada e a não supervisionada. Esses métodos possuem algumas diferenças, na aprendizagem supervisionada os pares de entrada já são definidos, logo o algoritmo treina com as características já determinadas para uma classe. Já a aprendizagem não supervisionada, não é fornecido ao algoritmo a qual classe o dado pertence, na verdade ele realiza um agrupamento dos dados de entrada que apresentam ser potencialmente úteis e possuem similaridade (Russell and Norvig 2004). Devido ao cunho da pesquisa foram utilizados métodos de aprendizagem supervisionada.

Tais métodos utilizam como base um certo conjunto de treinamento com  $N$  pares de entradas e saídas,  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$  onde cada  $y_j$  é gerado por uma função  $f(x)$  desconhecida. Com base nesses dados, o método tenta descobrir uma função  $h(x)$ , denominada de hipótese, que mais se aproxime dos valores de  $y = f(x)$ , a qual é denominada função original (Russell and Norvig 2004). Como escolher a hipótese que mais se assemelha a função original é a questão central estudada na aprendizagem supervisionada. A fim de calcular a taxa de acerto da hipótese as classes (note que classe nesse contexto é o resultado esperado da função hipótese) dos dados já são conhecidas, por isso é dado o nome de aprendizagem de máquina supervisionada.

### 2.2.1. Métodos Ensemble

Os *ensembles* são um conjunto de técnicas de aprendizagem de máquina, que podem ser definidos como vários algoritmos individuais em paralelo, cujas as saídas são a combinação ou a fusão dos resultados gerados individualmente, afim de encontrar a melhor solução para um determinado problema (Opitz and Maclin 1999). Esses métodos baseiam-se na premissa de que um algoritmo individual não produz resultados ótimos em problemas não triviais do mundo real. Dessa forma, os *ensembles* utilizam a manipulação de vários algoritmos simples de AM, transformando-os em um algoritmo robusto (Daumé III 2012).



As técnicas de AM baseadas em AD são métodos *ensemble* bastante conhecidos, que utilizam um algoritmo determinístico, tendo como entrada um vetor de atributos. Sendo cada nó da árvore representado como uma condição lógica, onde é estabelecida uma pergunta, e as folhas podem ser representadas com as respostas geradas a partir dessas condições (Dwan 2018). Alguns exemplos de métodos *ensemble* estão dispostos a seguir:

- *Random Forest* (RF) é um exemplo de *ensemble* que usa AD. Nesse caso é usado o algoritmo de *Bagging* (Breiman 2001) para minimizar problemas na produção da árvore. O algoritmo de *Bagging* usa uma técnica de *bootstrap resampling* que gera subconjuntos de dados de entrada a partir do conjunto de dados originais, dessa forma, ele realiza a construção de classificadores base (Dwan 2018). O algoritmo manipula o conjunto de treino repetidas vezes para produzir classificadores que combinam as previsões realizadas com base na votação majoritária (Lei et al. 2010).
- *Extremely Randomized Trees* (ERT) (Geurts et al. 2006) é um método parecido ao RF, mas a principal diferença é que em casos extremos o ERT deixa ainda mais aleatória a construção das árvores, isto é, ele produz árvores totalmente diferentes entre si com estruturas independentes. A ERT usa a base de treino inteira para realizar a construção de todas as árvores de decisão, diferentemente do RF que utiliza subamostras do conjunto de treino (Dwan 2018).
- *Gradient Tree Boosting* (GTB) (Friedman 2001) é um *ensemble* baseado em árvores de regressão. Ele utiliza a técnica de *boosting* para melhorar o desempenho do algoritmo, realizando votações, afim de encontrar uma combinação ótima de árvores. O GBT cria uma árvore de regressão simples e depois usa variação de gradiente descendente para melhorar a construção das árvores com uma função de custo. A previsão do GTB é dada pela soma das estimativas das árvores (Dwan 2018).

Os *ensemble* baseados em AD, demonstram ter uma maior eficiência em relação a acurácia dos modelos preditivos, possibilitando assim, uma melhor perspectiva dos dados e das relações existentes entre eles.

### 2.3. Otimização de Hiperparâmetros

Os modelos de AM possuem algumas características que podem ser modificadas afim de melhorar o resultado gerado pelo modelo preditivo. Essas características são comumente chamadas de hiperparâmetros (Dwan 2018). Para o classificador AD, podemos definir, por exemplo, a profundidade máxima da árvore com o parâmetro *max depth* ou definir o número mínimo de amostras necessárias para dividir um nó interno com o parâmetro *min samples split*.

No entanto, para realizar esses ajustes é preciso explorar qual o *pipeline* melhor se encaixa para o modelo preditivo, dentre o conjunto de parâmetros dado (Dwan 2018). Para exemplificar, mostraremos um exemplo de conjunto de parâmetros para o classificador RF:

- **max depth** : [3, 4, 5, 6, 7], profundidade máxima da árvore.
- **n estimators** : [100, 200, 300, 400, 500, 600], número de árvores de decisão.

- **min samples split** : [2, 4, 8, 16, 32], número mínimo de amostras necessárias para dividir um nó interno.
- **min samples leaf** : [5, 10, 15, 20, 25], número mínimo de amostras necessárias para estar em um nó de folha.
- **criterion** : ["gini", "entropy"], função para medir a qualidade de uma divisão.

Observando o exemplo acima, vemos que para selecionar um *pipeline* de hiperparâmetros adequados para os modelos preditivos, torna-se uma tarefa inviável para seres humanos. Para isso existem técnicas que realizam seleção de amostras através de pesquisas de otimização de hiperparâmetros. Uma dessas técnicas é a pesquisa aleatória ou *random search* em inglês (Bergstra and Bengio 2012). O *Random Search* envolve pontos de amostragem, com base em alguma técnica de amostragem, para região viável do problema de otimização, avaliando o desempenho da função objetivo nos pontos escolhidos. Posteriormente realiza a atualização da estratégia de amostragem com base no histórico observado, para assim, realizar a próxima iteração (Andradóttir 2006). O número de iterações para a construção do *pipeline* do *random search* é determinado pelo programador, ao final das suas iterações é retornado o melhor *pipeline* encontrado (Dwan 2018).

### 3. Trabalhos Correlatos

No trabalho de Melo et al. 2014 é realizado um estudo comparativo de técnicas de AM combinadas a teoria FACS. Em geral, buscou-se verificar qual técnica de classificação de emoções é mais adequada durante o processo de aprendizagem do aluno. Dessa forma, Melo et al. 2014 realizou uma análise a partir da Base de Dados (BD) Cohn Kanade (CK) (Lucey et al. 2010), onde se encontra informações de pessoas de várias etnias, sexo e idade. As técnicas utilizadas no estudo de Melo et al. 2014 foram as Redes Neurais Artificiais (RNA) do tipo *MultiLayer Perceptron* (MPL), *Rede Basis Networks* (RBFN) e *Redes Bayesianas* (RB) que obtiveram uma precisão total de 86%, 75.72% e 71.76%, respectivamente.

Em sua pesquisa Dantas et al. 2015 usou o algoritmo de AD para classificação automática de emoções em ambientes virtuais de aprendizagem (AVA). Com essa abordagem propôs o reconhecimento das sete emoções básicas baseadas na teoria FACS. Da mesma maneira que Melo et al. 2014, Dantas et al. 2015 também faz uso da base CK, mas com algumas alterações. Foi utilizado uma técnica de dados aumentados, no qual fez uma compilação das imagens da base CK para criar vídeos. Ao final integrou o modelo preditivo construído à ferramenta *Moodle* e realizou os testes de seu estudo. A acurácia atingida pelo modelo preditivo foi de 86.4%.

Lien et al. 1998 realizaram suas pesquisas afim de averiguar a possibilidade de realizar o reconhecimento de expressões faciais utilizando apenas a parte superior da face. Mais especificamente foram utilizados dados coletados da testa e sobrancelhas dos voluntários e foram analisados tendo como base o FACS. Os dados foram extraídos utilizando 3 técnicas distintas: *Facial Feature Point Tracking*; *Dense Flow Tracking* com *Principal Component Analysis*; e *High Gradient Component Detection*. A classificação dos dados foi realizada utilizando *Hidden Markov Models*. Com estes processos foram obtidos, respectivamente, 85%, 93% e 85% de assertividade.

Diferentemente de Lien et al. 1998, que buscou utilizar um número mínimo de características, Aly et al. 2016 visou potencializar a taxa de sucesso com a junção de dados



2D e 3D obtidos com uma visão frontal e lateral da face. Ambos tipos de dados foram gerados por meio do kinect-v1. Como os dados não possuíam a mesma representação, fez-se necessário realizar a extração e normalização separadamente. Após a normalização, as informações foram classificadas por meio de uma Máquina de Vetor de Suporte gerando assertividade de 80% para imagens frontais e 59% para imagens laterais.

Wei et al. 2016 também utilizaram dados tridimensionais para realizar o reconhecimento de expressões faciais. Em suas pesquisas foram utilizados dados de 17 ângulos distintos, assim pretendiam aumentar o leque de cenários onde a classificação poderia ser utilizada. Foram utilizados 40 pontos faciais para formar vetores em 120 dimensões. Mantendo uma proporção de 2:1, os dados de treinamento foram classificados utilizando *Random Forest*. Ao termino da classificação os resultados variaram de aproximadamente 55% à 80%.

#### 4. Metodologia de Desenvolvimento

Para a elaboração dos experimentos de estudo os dados foram divididos em 70% para treino dos modelos e 30% para teste. E ainda, se utilizou de duas abordagens: A1, os algoritmos foram treinados e testados apenas com dados das expressões faciais alegria e tristeza; A2, os algoritmos treinados e foram testados com dados de seis expressões básicas derivadas do estudo Ekman et al. 1978 (alegria, surpresa, raiva, medo, desgosto, tristeza). Porém com uma diferença, que nesse estudo usará a expressão desgosto contida também no conjunto de dados da CK. Dessa forma, foi feito um agrupamento dessas classes que resultou no seguintes: raiva, desprezo, nojo, medo, alegria, tristeza e surpresa. Ainda nesse estudo, não se utilizou da expressão neutra, pois a mesma apresenta uma grande quantidade de dados em relação as outras expressões o que pode ocasionar me resultados tendenciosos. Vale ressaltar que as bases de dados utilizadas não foram balanceadas, pois o balanceamento faria com que 33% dos dados fossem descartados. o que acarretaria na perda de muito dados importantes para o treino dos modelos. Nos algoritmos de AM, afim de evitar resultado precários, foi realizado 10 iterações sobre a validação cruzada, na qual foi feita 10 partições. Também foi utilizado nos classificadores o ajuste de hiperparâmetros utilizando o a técnica otimização de hiperparâmetros *Random Search*.

Esse trablho, utiliza-se da base de dados CK como principal conjunto de dados. A Cohn Kanade possui a disponibilidade de imagens de pessoas de diferentes etnias, sexos e idades e que dispõe de um conjuntos de AU's (Melo et al. 2014). Inicialmente, a CK foi construída com 486 sequências de imagens de 97 pessoas, sendo que cada sequência varia desde o estado neutro até a expressão de pico. Para a identificação da expressão de pico é aplicado à base CK as métricas do estudo FACS, o que à torna mais confiável. Posteriormente, em uma de suas atualizações a CK recebeu 107 novas sequências de imagens de 26 indivíduos (Lucey et al. 2010), o que lhe confere um total de 593 sequências de 123 pessoas.

Com a intenção de aumentar a validade dos resultados, trouxemos a alternativa de se realizar testes em uma base auxiliar, dessa forma foi construída um base de dados fundamentada na CK, mas com algumas alterações. As alterações realizadas, diz respeito a orientação em que as imagens estão apresentadas e à quantidades de dados na base. Todas as sequências da CK tiveram o seu grau de inclinação alterado, tanto para a esquerda

	CK				BDA			
	A1		A2		A1		A2	
<b>RF</b>	<b>94.33%</b>	<b>5.90%</b>	69.88%	8.03%	97.76%	<b>2.28%</b>	68.32%	<b>3.45%</b>
<b>ERT</b>	93.90%	6.74%	63.61%	<b>6.28%</b>	95.32%	3.90%	62.12%	4.30%
<b>GTB</b>	93.19%	7.18%	<b>74.19%</b>	7.60%	<b>98.43%</b>	2.52%	<b>77.67%</b>	3.67%

**Tabela 3. Resultados gerais dos testes realizado na BDO e BDA**

quanto para a direita, em 15° (quinze graus) em relação ao seu estado original. Isso possibilitou triplicar os dados originais da CK resultando em um total de 1779 sequências de 123 pessoas. Dessa maneira que foi produzida a Base de Dados Aumentados (BDA).

Os modelos preditivos(RF, ERT, GBT) usados neste trabalho derivaram da biblioteca Scikit-learn (Pedregosa et al. 2011), que se trata de uma biblioteca Python que integra vários algoritmos de aprendizado de máquina, supervisionados e não supervisionados, dando a possibilidade de solucionar uma quantidade considerável de problemas do mundo real.

## 5. Resultados e Discussões

Nesta seção são apresentados os resultados obtidos durante os experimentos realizados com A1 e A2, nas bases de dados CK e BDA. A Tabela 3 mostra os resultados gerais alcançados durante os experimentos realizados. Nela, estão apresentados as médias dos classificadores e as margens de erro, respectivamente. Logo abaixo estão algumas ponderações sobre as Matrizes de Confusão (CM) geradas e sobre como as abordagens A1 e A2 foram aplicadas à CK e posteriormente aplicadas à BDA.

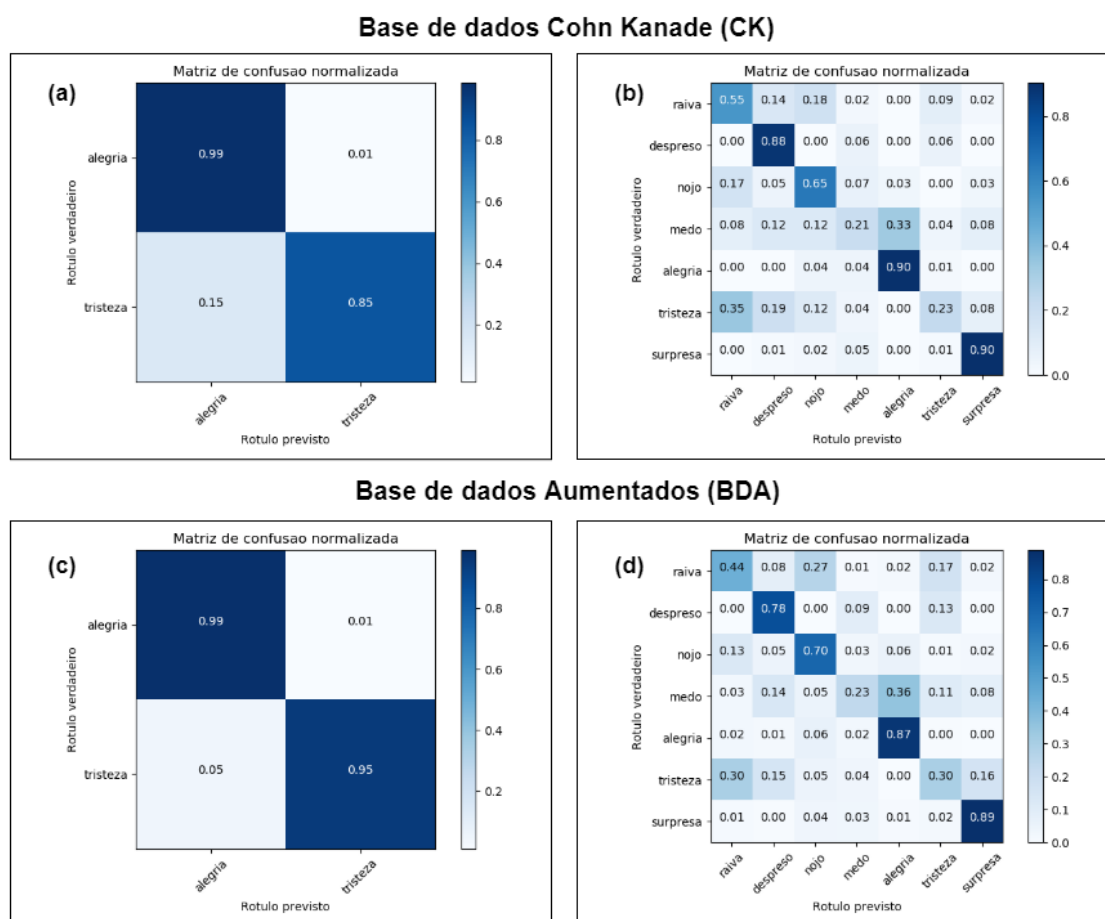
### 5.1. Random Forest

O resultado geral do experimento feito com a CK, no qual se utilizou da A1, pode ser visto na Tabela 3. Vemos que o algoritmo RF consegue selecionar bem as expressões obtendo um resultado de 94.33%, com uma margem de erro de 5.90%. Pode-se observar que na Figura 2(a), o RF obteve uma taxa de erro de apenas 1% na expressão alegria, revelando um bom desempenho. Além disso, o classificador obteve apenas 15% de erro nos dados da expressão tristeza.

Considerando a A2, foi observado que com a CK o RF obteve uma acurácia geral de 69.88%, com margem de erro de 8.03%. A Figura 2(b) demonstra que o classificador conseguiu selecionar bem as expressões desprezo, alegria e surpresa com uma precisão de 88%, 90% e 90%, respectivamente. Quando observamos mais detalhadamente as expressões raiva e nojo, podemos perceber que o RF consegue inferir pouco mais de 50% dos dados corretamente. Enquanto as expressões medo e tristeza obtiveram um baixo nível de reconhecimento, na Figura 2(b) vemos que os dados dessas expressões estão distribuídos entre todas as outras classes. Para o algoritmo RF, 33% dos dados da expressão medo, pertencem a expressão alegria. Já para a expressão tristeza o algoritmo identificou que 35% dos dados é pertencem a raiva.

Isso indica que o algoritmo encontrou ambiguidades nos dados das expressões faciais e às classificou de forma incorreta. Uma forma de contornar esse problema e treinar o algoritmo com um conjunto de dados maior, pois quanto mais dados, mais características

o algoritmo poderá coletar para identificar uma determinada expressão. Vale lembrar que na A1 a emoção tristeza obteve um bom resultado, isso devido ao fato de que o algoritmo estava sendo submetido a apenas duas classes de expressões.



**Figura 2. Matrizes de Confusão do algoritmo *Random Forest* aplicado as bases CK e BDA através das abordagens A1 e A2 respectivamente.**

Podemos verificar na Tabela 3 que os experimentos realizados na BDA com o classificador RF, obteve um resultado geral de 97.76%, com margem de erro de 2.28% quando aplicado a A1.

Visualizando a Figura 2(c), observamos que o algoritmo conseguiu classificar bem 99% dos dados da expressão alegria e 95% para expressão tristeza. Também que houve um aumento da acurácia em relação aos dados ao experimento anterior.

Isso demonstra que o algoritmo conseguiu coletar mais características, já que a quantidade de dados nesse experimento é o triplo do anterior, isso possibilitou que o algoritmo tivesse mais confiança ao classificar os dados das expressões, gerando assim, um resultado 3% melhor que o anterior.

Aplicando a A2 na BDA, podemos ver que o resultado geral na Tabela 3 para o RF, foi de 68.32%, com margem de erro de 3.45%.

Na Figura 2(d) percebemos que o algoritmo demonstrou algumas mudanças em

relação ao mesmo experimento realizado na base CK, dentre elas podemos destacar a expressão nojo, que conseguiu classificar até 70% dos dados. E ainda houve uma baixa de pouco mais de 10% nos resultados das expressões raiva e desprezo.

As expressões alegria e surpresa tiveram uma queda de 3% e um 1%, respectivamente, em relação ao experimento anterior. E novamente as expressões medo e tristeza obtiveram uma baixa taxa de reconhecimento.

Isso demonstra que mesmo depois de aumentar o conjunto de dados e que utilizando o algoritmo não conseguiu classificar bem essas duas expressões, assim como no experimento anterior.

## 5.2. Extremely Randomized Trees

Empregando a A1 na CK, podemos observar na Figura 3(a) que o ERT conseguiu selecionar bem a expressão de alegria, atingindo uma taxa de acerto de 99%, enquanto a expressão tristeza atingiu 85% de assertividade. No geral obteve-se uma média de 93.90% de acurácia e uma margem de erro de 6.74%.

Os classificadores RF e o ERT obtiveram os mesmos coeficientes para as expressões alegria e tristeza quando utilizado A1, mas ao realizar a média dos resultados gerados nas iterações dos algoritmos o ERT se demonstra levemente inferior ao RF.

Isso deve-se ao fato que ERT estar selecionando dados incorretamente, ou seja, o algoritmo está classificando alguns dados de forma errada. Porém há apenas 0.43% de discrepância entre os classificadores, o que não é considerado uma grande diferença entre os classificadores.

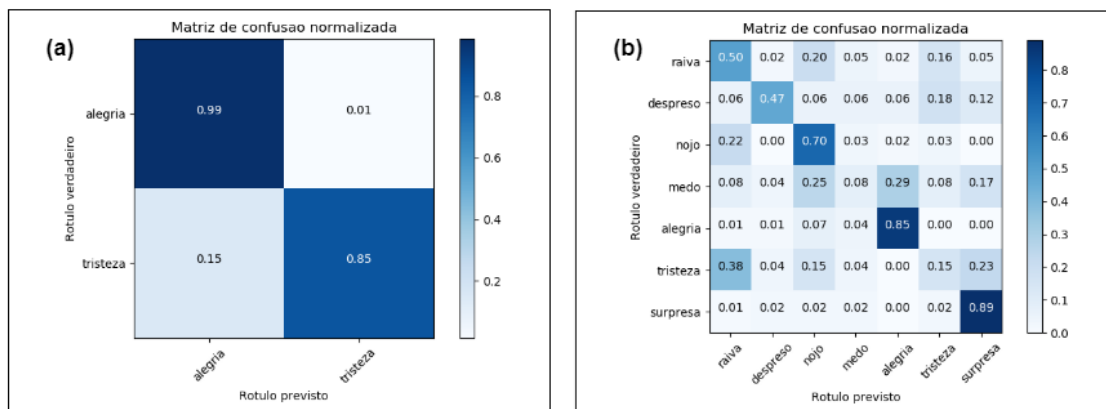
Nas análises feitas com a A2, mostraram algumas diferenças em relação ao modelo anterior. Na CK, o ERT atingiu uma acurácia geral de 63.61%, com uma margem de erro de 6.28%, este valor está apresentado na Tabela 3.

Pode ser visto na Figura 3(b) que os valores para a maioria das classes ficaram baixos em relação aos valores do mesmo teste realizado no RF. Podemos ver que a emoção raiva, alegria caíram cerca de 5% e que surpresa caiu apenas 1%. Já a expressão desprezo que antes possuía um dos melhores resultados do modelo anterior, apresenta-se nesse como um dos mais baixos resultados. A expressão desprezo caiu cerca de 41% se comparado ao RF.

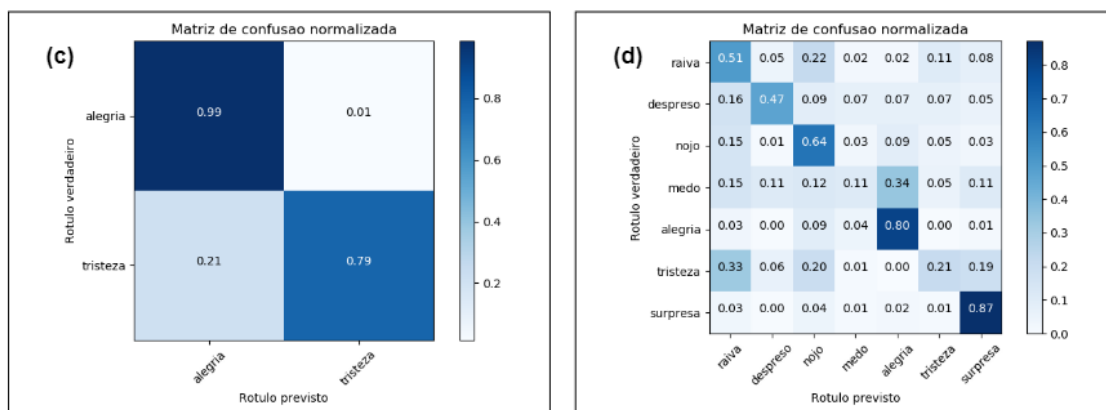
Esse valor expressivo, revela que ERT possui dificuldades na extração das características da expressão desprezo e selecionar os dados corretamente. Ainda podemos ver que as expressões medo e tristeza mantiveram sua baixa de assertividade, logo percebemos a dificuldade que os algoritmos encontram para identificar as características referentes a essas expressões. Contudo, a expressão nojo teve um acréscimo de 5% em comparação ao RF quando aplicado A2 na base CK.

A análise realizada com a BDA utilizando a A2, apresentou poucas diferenças em relação aos experimentos feitos na CK aplicando a A1. Na Figura 3(c), o resultado da previsão da expressão tristeza que obteve acurácia de 79%, revelando um frequência de erro maior em relação a CK. Porém, o ERT obteve um desempenho geral de 95.32% de precisão e com uma margem de erro de 3.90% o que é 1.42% superior ao resultado do mesmo experimento feito na CK.

### Base de dados Cohn Kanade (CK)



### Base de dados Aumentados (BDA)



**Figura 3. Matrizes de Confusão do algoritmo *Extremely Randomized Trees* aplicado as bases CK e BDA através das abordagens A1 e A2 respectivamente. Matriz de Confusão do algoritmo.**

Esse aumento é causado pela média das iterações feitas na BDA, logo quando observamos o resultado geral podemos perceber que os resultados dessas iterações foram melhores que no experimento com CK utilizando a A1.

Por sua vez a A2 aplicado na BDA, o ERT atingiu 62.12% de acerto, com uma margem de erro de 4.30%. Esse resultado se concretizou como o mais baixo entre os classificadores quando aplicado na BDA.

Podemos ver na Figura 3(d) que a previsão do conjunto de expressões obteve resultados diferentes em relação ao mesmo experimento feito no modelo anterior. As expressões alegria e surpresa obtiveram 80% e 87% de acurácia, respectivamente. A expressão raiva teve um acréscimo de 7%, enquanto nojo teve uma queda de 6% em relação ao modelo anterior. Já a expressão desprezo teve uma queda de 31% em relação ao RF. Isso se assemelha com o que aconteceu quando utilizamos o ERT aplicado a A2 na CK.

O classificador ERT possui dificuldade para identificar as características da expressão desprezo, nesse sentido, ele se torna um classificador tendencioso, classificando os dados de maneira incorreta. Novamente as expressões medo e tristeza não obtiveram

um bom resultado, podemos ver que na Figura 3(d) que o ERT identifica 34% dos dados da expressão medo com sendo alegria e 33% dos dados da expressão tristeza com sendo raiva. Esse comportamento deve ao fato de que existe muitas semelhanças entre essas expressões, logo o algoritmo encontra muitas ambiguidades o que ocasiona na classificação equivocada.

### 5.3. Gradient Tree Boosting

O classificador GTB ao aplicada a A1 aos dados da CK, demonstra na Figura 4(a) que conseguiu selecionar corretamente todos os dados da expressão alegria submetidos, obtendo uma um resultado de 100% para a mesma. Já a expressão tristeza obteve 88% de assertividade, relevando que o GBT classificou 12% dos dados dessa expressão como sendo a expressão alegria. Neste experimento o classificador atingiu um resultado geral de 93.19%, com uma margem de erro de 7.18%.

Utilizando A2 nos dados da CK, o algoritmo apresentou bons resultados. O GTB conseguiu atingir o resultado geral de 74.19%, com a margem de erro de 6.70%. Pode-se destacar ainda na Figura 4(b) que as expressões raiva, nojo, alegria e surpresa tiveram os melhores resultados dentre os mesmos experimento realizados nos modelo preditivos anteriores. As expressões atingiram 61%, 80%, 94% e 98% respectivamente. No mais, o classificador manteve os baixos resultados para as expressões medo e tristeza, sendo que para a expressão medo o algoritmo identificou que 38% dos dados pertencem a expressão alegria e para tristeza, 42% dos dados foram identificado como sendo raiva. Já a expressão desprezo teve o mesmo resultado se comparado com o ERT, mas se comparado ao RF teve um queda de até 41%.

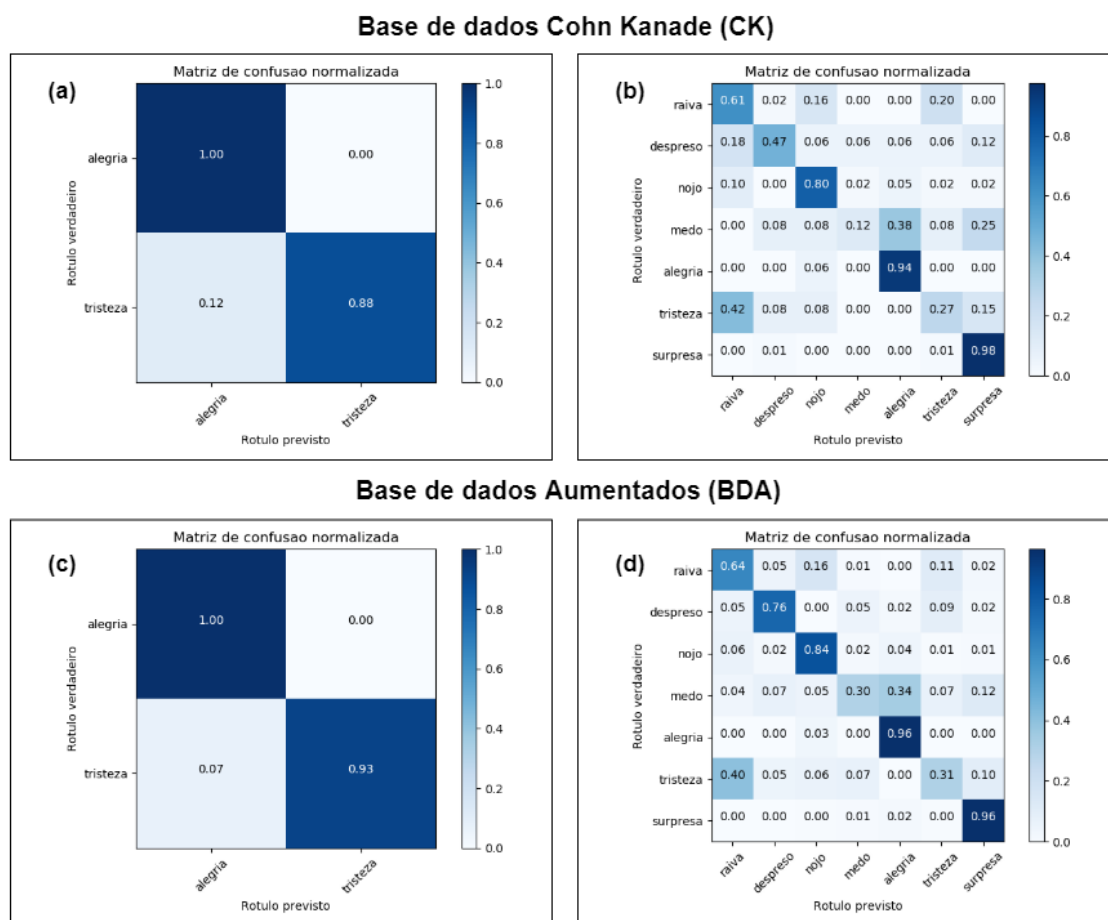
As taxas apresentadas na Figura 4(c) demonstra que o GTB ao aplicar a A1 aos dados BDA, podemos ver que ele obteve valores superiores em relação aos experimentos realizados com CK e que também teve os melhores resultados se comparado aos modelos anteriores. Também pode-se observar que o classificador teve o melhor desempenho em prever a expressão tristeza dentre todos os modelos usados nesse trabalho, conseguindo alcançar 93% de acertos. O GTB atingiu ainda a acurácia geral de 98.43% com margem de erro de 2.52%, como é demonstrado na Tabela 3. Dessa forma se caracterizou como sendo o resultado mais significativo dentre todas as técnicas que utilizaram a A1 aplicada aos dados da BDA.

Já para a A2 aplicada a dados da BDA, podemos analisar na Figura 4(d) que as expressões raiva, desprezo, nojo, alegria e surpresa apresentou os melhores resultados entre os modelos preditivos quando aplicado a BDA. Logo as expressões obtiveram 64%, 76%, 84%, 96% e 96% respectivamente. Isso qualifica o GTB como o modelo preditivo como o melhor desempenho em reconhecer características das expressões faciais empregado ao escopo desse estudo. Olhando detalhadamente a Figura 4(d) vemos que o classificado também encontra dificuldade para identificar as expressões medo e tristeza. No entanto, vale ressaltar que as mesmas obtiveram os melhores resultados em relação aos modelos anteriores, com 30% e 31% para as expressões de medo e tristeza, respectivamente.

## 6. Considerações Finais

Para este estudo foi proposto o uso de duas bases de dados para avaliar o desempenho em diferentes conjuntos técnica de AM em reconhecer expressões. Nesse sentido





**Figura 4. Matrizes de Confusão do algoritmo *Gradient Tree Boosting* aplicado as bases CK e BDA através das abordagens A1 e A2 respectivamente. Matriz de Confusão do algoritmo.**

foi utilizado uma base de dados chamada Cohn Kanade, baseada no estudo FACS e outra que se baseia na própria CK, mas com diferenças nas orientações das imagens e com o triplo de dados que chamada de Base de Dados Aumentados. Além disso foi utilizado duas abordagens, A1 e A2, que visam avaliar como os algoritmos se comportam ao com diferentes grupos de classes submetidas a ele.

Na A1 os classificadores analisaram apenas os dados das expressões alegria e tristeza. Na A2 foi utilizado as emoções básicas incorporadas no estudo FACS juntamente com a expressão desprezo, presente na base de dados CK. Esse agrupamento do conjunto de expressões dos dois cenários gerou o seguinte conjunto de expressões: raiva, desprezo, nojo, medo, alegria, tristeza e surpresa. Ainda nesse trabalho, não foi utilizado a expressão neutra, pois a mesma tinha grande quantidade de dados em relação a quantidade de dados das outras expressões.

Vale ressaltar a importância desse estudo, que pode ser aplicado em diversas áreas. Uma delas é a educação, onde o professor poderá usar, por exemplo, os dados das emoções de alegria e tristeza para saber se um aluno está indo bem ou não na disciplina. Além disso o professor poderá verificar todo o processo de aprendizagem do estudante durante



o semestre (Melo et al. 2014) e montar um perfil de aprendizagem adequado para realizar intervenções e ajudar os alunos que possuem dificuldade e incentivar os que possuem pouca dificuldade a melhorar mais. Outra área seria o *e-commerce*, que poderia aplicar a técnica apenas com a A1, aliado ao sistema de recomendação (SR) e saber, de forma automática, se o que foi recomendado para o cliente o agradou ou não.

Portanto, para o escopo desse estudo, o método *ensemble* que apresentou os melhores resultados nas diferentes abordagens e nas duas bases de dados, foi o *Gradient Tree Boosting*. O classificador conseguiu um resultado de 93% em testes com a CK e 98% na BDA, quando usado a A1. Já utilizando a A2 o algoritmo conseguiu atingir 75% aplicado a CK e para a BDA 77%.

Para trabalhos futuros, pretende-se testar os mesmos métodos em conjuntos de dados novos, para assim, ter uma validação que os resultados obtidos vão além do escopo desse trabalho. Ainda vale realizar um estudo mais detalhado sobre técnicas que sejam capazes reconhecer expressões faciais em imagem com precisão, mas com o foco em técnicas que consigam identificar principalmente as classes medo e tristeza.



## Referências

- Aly et al. 2016 Aly, S., Abbott, A. L., and Torki, M. (2016). A multi-modal feature fusion framework for kinect-based facial expression recognition using dual kernel discriminant analysis (dkda). In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–10. IEEE.
- Andradóttir 2006 Andradóttir, S. (2006). An overview of simulation optimization via random search. *Handbooks in operations research and management science*, 13:617–631.
- Bergstra and Bengio 2012 Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.
- Breiman 2001 Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Dantas et al. 2015 Dantas, A. C., de Melo, S., Moura, F., and Fernandes, M. (2015). Reconhecimento dinâmico de emoções através de expressões faciais utilizando árvore de decisão. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, page 1102.
- Daumé III 2012 Daumé III, H. (2012). A course in machine learning. *Publisher, ciml.info*, pages 5–73.
- Dwan et al. 2017 Dwan, F., Oliveira, E., and Fernandes, D. (2017). Predição de zona de aprendizagem de alunos de introdução à programação em ambientes de correção automática de código. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1507.
- Dwan 2018 Dwan, F. P. (2018). Uso de um método preditivo para inferir a zona de aprendizagem de alunos de programação em um ambiente de correção automática de código. Master's thesis. Instituto de Computação.
- Ekman 1989 Ekman, P. (1989). The argument and evidence about universals in facial expressions. *Handbook of social psychophysiology*, pages 143–164.
- Ekman et al. 1978 Ekman, P., Friesen, W. V., and Hager, J. C. (1978). Facial action coding system (facs). *A technique for the measurement of facial action*. Consulting, Palo Alto, 22.
- Ekman and Oster 1979 Ekman, P. and Oster, H. (1979). FACIAL EXPRESSIONS +316 OF EMOTIONI.
- Ekman and Rosenberg 1997 Ekman, P. and Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.
- Friedman 2001 Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.
- Geurts et al. 2006 Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1):3–42.



- Kanade et al. 2000 Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53. IEEE.
- Lei et al. 2010 Lei, S., Weng, M., Ma, X., and Xi, L. (2010). Rough set based decision tree ensemble algorithm for text classification. *J. Comput. Inf. Syst*, 6(1):89–95.
- Libralon 2014 Libralon, G. L. (2014). Modelagem computacional para reconhecimento de emoções baseada na análise facial.
- Lien et al. 1998 Lien, J. J., Cohn, J. F., Kanade, T., and Li, C. C. (1998). Automated facial expression recognition based on FACS action units. *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, pages 390–395.
- Lucey et al. 2010 Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE.
- Melo et al. 2014 Melo, S. L., De Moura, F. F., Macedo, K., Alves, F. S. R., and Fernandes, M. A. (2014). Estudo comparativo de técnicas computacionais para classificação de emoções. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 25(1):456.
- Opitz and Maclin 1999 Opitz, D. and Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11:169–198.
- Pedregosa et al. 2011 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.
- Robotics Institute 2002 Robotics Institute, C. (2002). Facs - facial action coding system. (<https://www.cs.cmu.edu/~face/facs.htm>). Acessado: 08-06-2017.
- Russell and Norvig 2004 Russell, S. J. and Norvig, P. (2004). *Inteligencia Artificial: um enfoque moderno*. Number 04; Q335, R8y 2004.
- Samuel 1959 Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229.
- Tomkins 1980 Tomkins, S. S. (1980). Affect as amplification: Some modifications in theory. *Emotion: Theory, research, and experience*, 1:141–164.
- Wei et al. 2016 Wei, W., Jia, Q., and Chen, G. (2016). Real-time facial expression recognition for affective computing based on kinect. In *Industrial Electronics and Applications (ICIEA), 2016 IEEE 11th Conference on*, pages 161–165. IEEE.