

Classificação de emoções por meio de expressões faciais em sala de aula

Marcus V. S. Maziero¹ e Paulo R. K. Nakaima²

Abstract—Procura-se neste trabalho realizar breve revisão bibliográfica sobre problema de reconhecimento de expressões faciais em sala de aula para melhorar a qualidade de ensino. Na sequência será realizado um laboratório para demonstrar a aplicação de técnicas de aprendizado de máquina para a solução do problema. Busca-se abordar de forma introdutória e simplificada o tema para fundamentar trabalhos futuros.

Keywords—Inteligência Artificial, Emoções, Imagens, Machine Learning.

I. INTRODUÇÃO

O reconhecimento de expressões faciais são um campo específico e instigante da área de Inteligência Artificial (IA), conforme descreve [4], pois possibilita encontrar e rastrear movimentos incomuns na face das pessoas.

Com o rastreamento desses dados é possível por intermédio classificar as mesmas e compreender como as emoções afetam as tomadas de decisões, conforme [2] é um ponto instigante a ser aprofundado. Em uma perspectiva de utilizar e melhorar o desempenho das aplicações que reconhecem e classificam expressões faciais é utilizado já por algumas empresas o Machine Learning (ML) termo em inglês que remete ao aprendizado de máquina, conforme apresenta [8], a utilização do ML já uma prática familiar em algumas empresas.

Sendo assim a presente pesquisa foi desenvolvida com o intuito de auxiliar os campos pesquisados e apresentar resultados de práticas e técnicas de reconhecimento facial.

II. CONTEXTO

A classificação de emoções por reconhecimento facial é algo que pode ser utilizado para favorecer

a área escolar, conforme [9], as emoções podem descrever a maturidade de aprendizado do discente e como o mesmo está diante dos ensinamentos propostos. Com isso essa classificação auxilia o docente para aumentar o desempenho da sua turma de alunos, melhorando assim o ensino.

Esse tipo de classificação também pode ser usado para fins sociais e comerciais para aumentar desempenhos de vendas e melhorar a abordagem das empresas relacionadas as frentes descritas como apresenta [10] em sua pesquisa. Dessa forma é possível dizer que a classificações de expressões faciais podem auxiliar diversas áreas e que com o avanço desta resulta em diversas melhorias para a educação, comércio, medicina e atividades sociais.

III. OBJETIVOS

Classificar expressões faciais a partir de imagens utilizando técnicas de processamento de imagens. Apresentar protótipo de classificador de expressões faciais visando desenvolver aplicação para professores avaliarem o engajamento dos alunos durante a aula.

IV. TRABALHOS RELACIONADOS

Em [6] é apresentada uma abordagem para extração de características de imagens visando o melhoramento da acurácia para o reconhecimento facial. Esta abordagem é baseada na teoria de redes complexas a qual fornece as medidas para remoção de conexões menos significativas para o processo de aprendizado de máquina. Com esse procedimento foi possível obter acurácia de 95% e 96.3% nas bases *Caltech Face Dataset* (CFD) e *Head Pose Image base* (HPI).

Em [9] criam um sistema para analisar emoções em sala de aula. Aplicam redes neurais convolucionais (*Convolutional Neural Network* - CNN) para classificar as expressões faciais dos alunos

¹Marcus V. S. Maziero e ²Paulo R. K. Nakaima estão vinculados à Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai, 1640, Cornélio Procopio, Brasil. marcus.maziero@outlook.com, nakaima@alunos.utfpr.edu.br

e professor. A aula é gravada em vídeo, na sequência esse é convertido em *frames*; utilizam o algoritmo *Viola-Jones* para detecção facial, ao resultado é aplicado o descritor de texturas *Local Binary Pattern* (LBP), por fim, é utilizado CNN para classificação das emoções. O sistema foi testado com 105 imagens retiradas das bases CK, JAFFE e *google images* as quais incluem fotos de pessoas sozinhas ou em grupos. As expressões foram classificadas como raiva, nojo, felicidade, neutralidade, surpresa, medo e tristeza. Obteve, respectivamente percentual de acerto de: 33,3%, 46,6%, 60%, 46,6%, 0%, 33,3%, 53,3%. No teste com a gravação em vídeo o sistema acertou 80,9% em média e erro de 19,1%.

Em [1] apresenta uma pesquisa de reconhecimento de emoções humanas por imagens faciais, com a abordagem *Single Shot Facial Expression Recognition* (SSFER), demonstra que o método MMOD-CNN foi o que apresentou melhor acurácia 91.89%, para comprovar é apresentado o resultado de experimentos combinando CNN e classificador, com cinco CNN e quatro classificadores, são eles VGGNet, InceptionResNetV2, InceptionV3, MobileNetV2, ResidualNet, Softmax, SVM, Random Forest e KNN respectivamente, a pesquisa também demonstra um experimento real com vinte e sete estudantes do ensino médio utilizando do MMOD-CNN para reconhecer as emoções dos alunos.

V. METODOLOGIA

A. Ambiente de Software

Para realizar o laboratório foi utilizado de algumas tecnologias e ferramentas são elas:

- *Scikit*: é uma ferramenta de código aberto, simples e eficaz na análise de dados, acessível e reutilizável por vários contextos [5]. O projeto foi iniciado em 2007 com por intermédio de um projeto da Google Summer of Code, atualmente é mantido pela comunidade
- *Python*: é uma linguagem de programação em script's, interpretada, de código aberto e rápida, mantido pela comunidade e administrada pela Python Software Foundation. Diversas empresas já utilizam dela, como Google, YouTube e EVE Online.

B. Base de dados

Foi selecionada a base de imagens CKPLUS48. A qual pode ser encontrada em: <https://www.kaggle.com/shawon10/ckplus>. Contém 750 imagens de rostos em escalas de cinza com as expressões de raiva, nojo, desprezo, medo, felicidade, tristeza e surpresa. Para este laboratório foram selecionadas apenas as expressões raiva, medo, felicidade, tristeza e surpresa.

C. Extração de características

Para extração de características foi utilizado os descritores Local Binary Patterns (LBP) e o Gabor.

Sendo que o primeiro extrai 256 características de textura é não paramétrico, possui baixo custo de processamento [7]. Seu funcionamento consiste em dividir a imagem em sub-blocos sendo que para cada sub-bloco é calculado o histograma, o vetor de características é formado com a concatenação desses histogramas [7]. Sendo que o segundo gera filtros com ondas senoidal, com os eixos desta onda busca encontrar a Gaussiana e assim manter uma proporção constante na imagem, para no final do processo gerar contornos e texturas para obter dados[3]. Neste laboratório o Gabor extrai 60 características de textura como linhas e contornos nas imagens.

Os arquivos finais podem ser encontrados em: encurtador.com.br/gnDOZ.

D. Seleção dos classificadores e técnica de amostragem

Os classificadores foram selecionados com base no trabalho de [7], o qual indica quais foram os mais utilizados em problemas de reconhecimento de expressões faciais (Facial Expression Recognition - FER), sendo eles:

- *Regressão Logística (Logistic Regression - LR)*: produz a partir de um conjunto de dados, modelos que permitem a predição por variáveis categóricas
- *K-Nearest Neighbors - (KNN)*: procura o vizinho K mais próximo de um dado ponto no espaço do conjunto de dados.
- *Máquina de Vetores Suporte (Support Vectors Machine - SVM)*: é uma máquina de entrada e saída onde os dados que são enviados como

entrada são mapeadas em um espaço multidimensional e encontra um hiperplano que separa os dados de entrada.

- *Rede Perceptron Multicamadas (Multi-Layer Perceptron - MLP)*: é uma rede neural com camadas onde os números de neurônios são indeterminados, sendo assim não é possível prever a saída intermediária entre as camadas.

Para os experimentos executados nas seções seguintes o conjunto de dados foram divididos em 80% para treino e 20% para teste. A seleção desses subconjuntos foram feitos de modo aleatório com o número gerador (*seed*) 327 para possível reprodução dos experimentos.

E. Seleção da técnica de normalização

A partir dos arquivos gerados com os descritores foram alternadas as técnicas de normalização para verificar qual delas atinge melhor resultado para acurácia, métrica que calcula a razão entre quantidade de predições corretas e a quantidade de predições realizadas.

As técnicas de normalização testadas foram:

- *Min-Max*: consiste em transformar cada característica com valor mínimo em 0 e os valores máximos em 1, sendo que o restante é transformado em um valor decimal entre 0 e 1.
- *Standard*: são calculados a média e o desvio padrão da conjunto de amostras, em seguida é subtraída de cada amostra a média, o resultado então é dividido pelo desvio padrão.
- *Max Absolute*: é uma técnica parecida com o Min-Max porém, somente os valores absolutos e positivos são mapeados entre 0 e 1.
- *Robust*: as estatísticas de mapeamento ocorrem pela escala de percentil, sendo assim aumenta a margem de onde estão os valores dos dados podendo ser até negativo, exemplo: -3 a 2.

1) *Descritor LBP*: A base CKPLUS com o descritor LBP apresentou pequena variação nos resultados como podem ser vistas na Figura 1.

Os classificadores MLP, SVM obtiveram os melhores resultados, 97% de acurácia. Os piores resultados foram obtidos pelo classificador KNN, 77%.

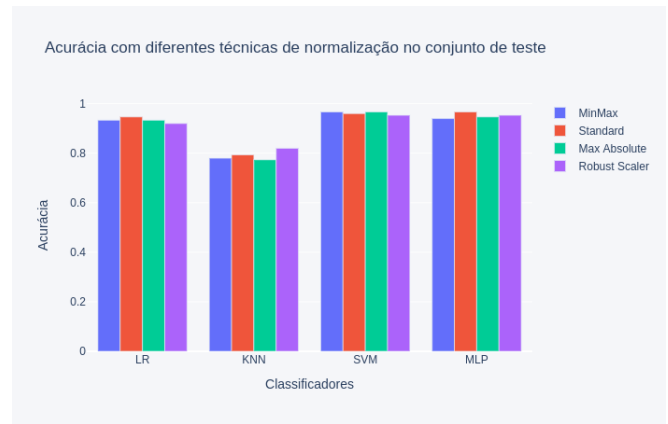


Fig. 1. Comparação de técnicas de normalização com descritor LBP.

2) *Descritor Gabor*: Foram mantidos os classificadores e métrica utilizados no último experimento. Nenhum dos classificadores obtiveram 50% de acurácia. Os resultados podem ser vistos na Figura 2.

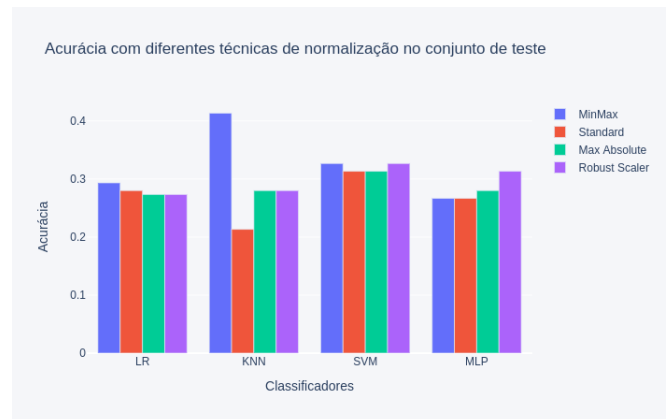


Fig. 2. Comparação de técnicas de normalização com descritor Gabor.

Pode-se observar que a técnica de normalização Min-Max, atrelado ao classificador KNN obteve o melhor resultado deste experimento, 41%. Observa-se também que ocorre uma alta discrepância sobre os resultado obtidos entre os descritores LBP e Gabor.

F. Redução de dimensionalidade

1) *Descritor LBP*: Em seguida buscou-se reduzir a dimensão do conjunto de dados de modo a reter 90% de variância. Primeiramente os dados foram normalizados utilizando a técnica *Standard*,

pois obteve resultado levemente superior no experimento anterior.

Com a técnica PCA para redução de dimensionalidade foi selecionado o menor número de componentes que retinha 90% de variância. Foi possível reduzir de 256 para 133 característica. Ver Figura 3.

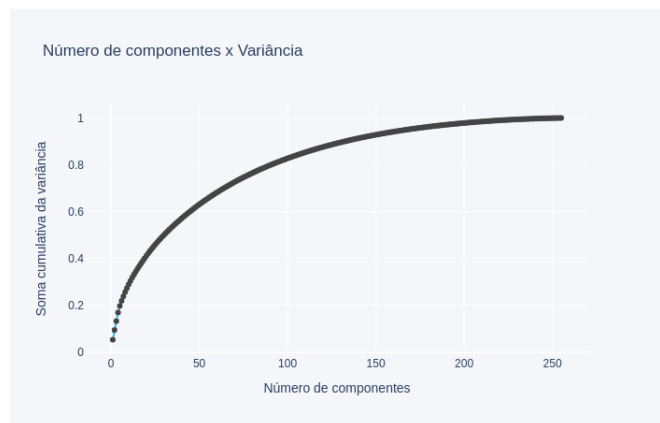


Fig. 3. Selecionar o menor número de componentes retendo 90% de variância.

2) *Descritor Gabor*: Primeiramente foram normalizados os dados com a técnica *MinMax* pois obteve o melhor resultado no experimento anterior. Ver Figura 4.

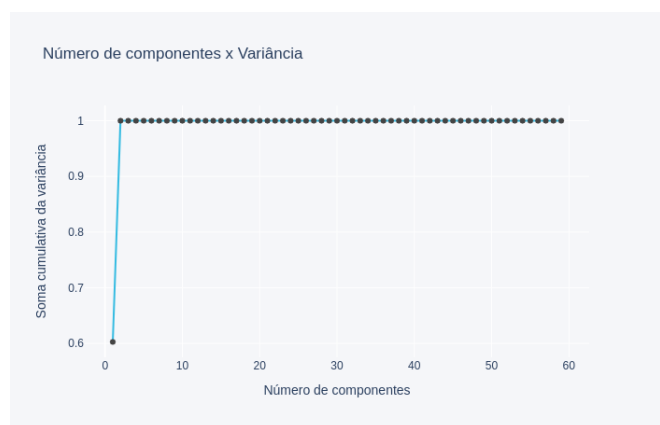


Fig. 4. Selecionar o menor número de componentes retendo 100% de variância.

Observa-se melhores resultados comparado com o experimento anterior. Pois se fez possível reduzir as características de 60 para duas retendo 100% de variância. Embora o desempenho dos classificadores atinjam no máximo 69.3%, nota-se grande melhoria quando é reduzida a quantidade

de características quando compara-se os resultados da Figura 2 com a Figura 5

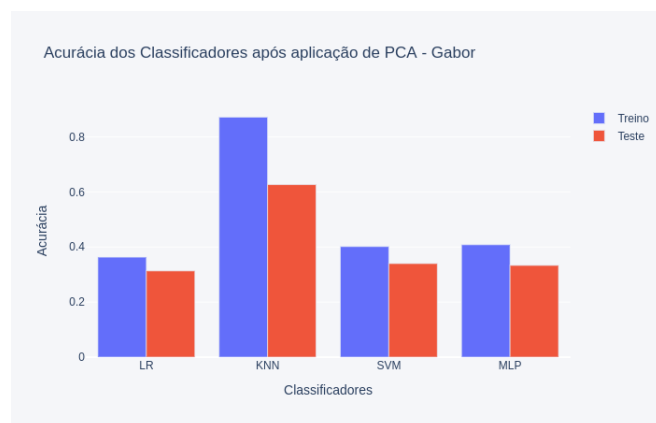


Fig. 5. Acurácia com o descritor Gabor após redução de dimensionalidade

VI. RESULTADOS

Nas seções anteriores foram coletados alguns indícios para seleção dos descritores combinadas com a melhor técnica de normalização, menor número de características. Por isso, para consolidar os dados anteriores e evitar superestimação dos resultados encontrados foi utilizado a técnica de validação cruzada estratificada em 10 partições.

1) *Descritor LBP*: Com relação a acurácia as médias e desvio padrão foram respectivamente: LR: 69%, 16%; KNN: 46%, 24%; SVM: 64%, 22%; MLP: 71%, 18%. Estes dados podem ser vistos na Tabela I e Figura 6.

TABLE I
ACURÁCIA DOS CLASSIFICADORES UTILIZANDO O DESCRITOR LBP

	Média	Desvio Padrão
MLP	71%	18%
LR	69%	16%
SVM	64%	22%
KNN	46%	24%

2) *Descritor Gabor*: Com o descritor Gabor a média e desvio padrão foram respectivamente: LR: 28%, 26%; KNN: 25%, 13%; SVM: 29%, 30%; MLP: 27%, 28%. Estes dados podem ser vistos na Tabela II e Figura 7.



Fig. 6. Média e Desvio Padrão da Acurácia com o descritor LBP

TABLE II
ACURÁCIA DOS CLASSIFICADORES UTILIZANDO O DESCRITOR
GABOR

	Média	Desvio Padrão
SVM	29%	30%
LR	28%	26%
MLP	27%	28%
KNN	25%	13%

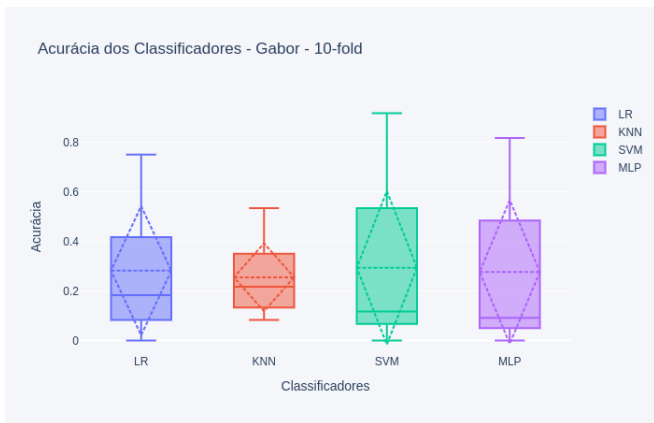


Fig. 7. Média e Desvio Padrão da Acurácia com o descritor Gabor

VII. CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho foi realizado o reconhecimento de expressões faciais com aprendizado de máquina. Foram comparadas com base na acurácia o melhor desempenho para os descritores LBP e Gabor. O primeiro foi que obteve melhor resultado. Foram comparadas diferentes técnicas de normalização sendo que os testes realizados com o descritor Gabor foi o que apresentou maior sensi-

bilidade para este tipo de alteração. Além disto foi aplicada a técnica de redução de dimensionalidade com o PCA, novamente os resultados baseados no descritor Gabor foi o mais afetado.

Por fim, coletou-se os indícios de qual a combinação de descritor, técnica de escalonamento, redução de dimensionalidade realizamos a validação cruzada estratificada para obter estimativas mais precisas. Com isto ocorreu notável queda no desempenho dos classificadores.

Como trabalhos futuros pode-se buscar novas bases de imagens. Identificar regiões das faces específicas e extrair as características manualmente com os filtros utilizados para ser possível a regulação dos parâmetros do filtro Gabor e LBP. Com relação aos classificadores utilizar Grid Search e validação cruzada para selecionar os parâmetros exigidos pelos algoritmos de classificação utilizados. Pode-se também utilizar algoritmos de *Deep learning*.

REFERENCES

- [1] Anderson Araújo da Cruz. Uma abordagem para reconhecimento de emoção por expressão facial baseada em redes neurais de convolução. In *Uma Abordagem para Reconhecimento de Emoção por Expressão Facial baseada em Redes Neurais de Convolução 2019*. Universidade Federal do Amazonas, UFAM, 2019.
- [2] Chie Hieda, Takato Horii, and Nagai Takayuki. Emotion differentiation based on decision-making in emotion model. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 659–665. IEEE, November 2018.
- [3] Eucassio Gonçalves Lima Júnior, Luis Henrique Silva, Cornelia Janayna Passarinho, and Ricardo de Andrade Lira Rabêlo. Um robusto reconhecimento facial por filtro de gabor curvo e entropia. pages 80–89. *Revista de Sistemas e Computação*, January 2016.
- [4] Marryam Murtaza, Muhammad Sharif, Musarrat Abdullah Yasmin, and Tanveer Ahmad. Facial expression detection using six facial expressions hexagon (sfhe) model. In *Facial expression detection using Six Facial Expressions Hexagon (SFEH) model*. School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, 2019.
- [5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [6] Joao Gilberto S. Piotto and Fabricio Martins Lopes. Combining SURF descriptor and complex networks for face recognition. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 275–279. IEEE, October 2016.

- [7] S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian. Facial expression recognition techniques: a comprehensive survey. *IET Image Processing*, 13(7):1031–1040, 2019.
- [8] Susmita Ray. A quick review of machine learning algorithms. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)*, pages 35–39. IEEE, October 2019.
- [9] K. S. Sahla and T. Senthil Kumar. Classroom Teaching Assessment Based on Student Emotions. In Juan Manuel Corchado Rodriguez, Sushmita Mitra, Sabu M. Thampi, and El-Sayed El-Alfy, editors, *Intelligent Systems Technologies and Applications 2016*, volume 530, pages 475–486. Springer International Publishing, Cham, 2016. Series Title: Advances in Intelligent Systems and Computing.
- [10] T. Özcan and A. Baştürk. Static image-based emotion recognition using convolutional neural network. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2019.