

## INTELIGÊNCIA ARTIFICIAL PARA RECONHECIMENTO DE EMOÇÕES EM UM AMBIENTE EMPRESARIAL

### ARTIFICIAL INTELLIGENCE FOR RECOGNITION OF EMOTIONS IN A BUSINESS ENVIRONMENT

*Bianca Cavalcante Basso*

<https://orcid.org/0000-0001-8536-488X>

*Centro Paula Souza – Fatec Indaiatuba/SP*

*bianca.basso@fatec.sp.gov.br*

*Orientador: Profa Dra. Maria das Graças J. M. Tomazela*

<https://orcid.org/0000-0002-5471-2658>

*Centro Paula Souza – Fatec Indaiatuba/SP*

*gtomazela@fatecindaiatuba.edu.br*

*Co-orientador: Prof. Dr. Aldo Pontes*

<https://orcid.org/0000-0002-2667-7701>

*Centro Paula Souza – Fatec Indaiatuba/SP*

*aldopontes@hotmail.com*

**RESUMO:** Na rotina de trabalho das pessoas, repara-se que o ser humano precisa tomar uma série de decisões, além de buscar soluções para problemas propostos e impostos a ele, tudo isso com o máximo de rapidez. Isso pode influenciar direta ou indiretamente na vida do ser humano, ocasionando assim a pressão decorrente do dia a dia laboral. Essas mudanças podem causar um mau desempenho decorrente da insatisfação do colaborador, que às vezes não é percebida pelos gestores. Nesse contexto, surgiu a ideia de utilizar algoritmos de inteligência artificial e rede neural convolucional para armazenar e reconhecer expressões faciais para interpretação de emoções dos colaboradores por meio de seus rostos, em tempo real, construindo dessa maneira, um histórico de emoções obtidas dos seus colaboradores. Para tanto, a realização deste estudo abordou os seguintes conceitos chave, que sustentam a pesquisa: Inteligência artificial, redes neurais artificiais, reconhecimento facial e inteligência emocional. Para a realização deste trabalho foi feita uma pesquisa experimental, que resultou em uma aplicação de reconhecimento facial utilizando sete emoções: raiva, nojo, medo, feliz, triste, surpresa e neutro, fazendo uso da técnica de Rede Neural Convolucional, além de ferramentas e bibliotecas de apoio para a utilização do modelo proposto. A acurácia do modelo foi melhor para reconhecimento de expressões neutras, com valores variando entre 95% e 99%; felizes, com acurácia variando entre 58% e 61%; e raiva, com acurácia variando entre 80% e 90%, enquanto as demais emoções resultaram em muitos falsos positivos.

**ABSTRACT:** In the work routine of people, it is noticed that the human being needs to make a series of decisions, in addition to seeking solutions to problems proposed and imposed on him, all of this with maximum speed. It can influence directly or indirectly in the life of the human being, thus causing the pressure resulting from the daily work. These changes can cause poor performance due to employee dissatisfaction, which is sometimes not noticed by managers. In this context, the idea arose to use artificial intelligence algorithms and convolutional neural network to store and recognize facial

expressions for interpreting employees' emotions through their faces (in real time thus building a history of emotions obtained from their employees). To this end, this study addressed the following key concepts, which support the research: Artificial intelligence, artificial neural networks, facial recognition and emotional intelligence. Other related academic researches were also explored, to assist in the idealization of the proposed theme, in addition to verifying tools, software and means of implementation that could assist in the development of this work. To carry out this work, an experimental research was conducted, which resulted in an application of facial recognition (anger, disgust, fear, happy, sad, surprise, neutral), using Convolutional Neural Network, in addition to tools and support libraries for the use of the proposed model. The model's accuracy was better for recognizing neutral expressions, with values ranging between 95% and 99%; happy, with accuracy ranging from 58% to 61%; and anger, with accuracy varying between 80% and 90%, while the other emotions resulted in many false positives.

**PALAVRAS-CHAVE:** Inteligência Artificial, Reconhecimento Facial, Redes Neurais.

**KEYWORDS:** Artificial Intelligence, Facial Recognition, Neural Networks.

## 1 INTRODUÇÃO

Algumas emoções são alteradas dependendo da situação a que o ser humano é exposto, podendo variar perante uma simples mensagem recebida, uma conversa, situações vistas, dentre outras diversas situações existentes.

Segundo Bergamini (1996), as pessoas se comportam de maneira diferente dentro das organizações de acordo com uma série de determinantes. A autora descreve como esses determinantes afetam o desenvolvimento do indivíduo em uma empresa, afirmando a existência de duas variáveis: variáveis individuais, que dizem respeito à sua infância, adolescência e à fase adulta; e as variáveis ambientais, que se caracterizam pela forma que o ser humano se adapta ao ambiente social, cultural, aos fatores físicos e muitos outros.

O ambiente empresarial envolve fatores internos como também externos, nos quais observa-se a disputa acirrada por clientes e lucros, o que faz com que os gestores requeiram excessivo empenho e produtividade de seus colaboradores. É essa condição que viabiliza este projeto, que trata da possibilidade de verificar as emoções dos mesmos no meio empresarial, auxiliando gestores a terem uma atenção a mais com seus colaboradores, evitando situações de estresse que venham a prejudicar o desempenho.

Sistemas que usam Inteligência Artificial podem auxiliar os humanos a serem mais precisos em suas decisões, realizar funções com menos chances de errar, desenvolver sistemas que aprendem com o passado e tomar decisões com base nesses fatos passados, para ter uma assertividade maior.

Assim, o objetivo deste trabalho foi utilizar algoritmos de inteligência artificial para armazenar e reconhecer expressões faciais para interpretar emoções de colaboradores por

meio de suas expressões faciais, em tempo real, facilitando assim a tomada de decisão nas avaliações e possíveis intervenções e encaminhamentos desses colaboradores para profissionais especializados.

## 2 METODOLOGIA

Para atingir os objetivos desta pesquisa, optou-se pelo campo das pesquisas de natureza explicativa, que têm como preocupação central identificar os fatores que determinam ou que contribuam para a ocorrência dos fenômenos.

Neste campo, a referência foi a abordagem experimental, que consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto (GIL, 2002). Para o desenvolvimento desta pesquisa, as seguintes variáveis de controle foram definidas: qualidade do banco de imagem, parâmetros de treinamento da rede neural, acurácia da rede neural, eficácia de registro das emoções.

A partir disso, as seguintes etapas foram definidas para a execução deste trabalho:

- Analisar as informações e elaborar um *dataset* para realizar o treinamento e os testes de uma rede neural;
- Analisar e definir quais são os pontos principais para o treinamento da rede neural, usando ferramentas como: a linguagem de programação do *Python*, *Anaconda Navigator*, *Open Source Computer Vision (OpenCV)* e *keras*;
- Realizar os testes de acuracidade, usando algumas imagens separadas do *dataset* primária;
- Analisar e avaliar os resultados obtidos no processo. A avaliação desta pesquisa foi realizada com três pessoas que trabalham em desenvolvimento de sistemas, que se voluntariaram para captura dos rostos em seu dia-a-dia no ambiente de trabalho.

## 2 DESENVOLVIMENTO

A Inteligência Artificial (IA) tem por finalidade criar sistemas inteligentes que possam executar tarefas de forma semelhante ao comportamento humano, é uma área que foi influenciada por diversas áreas como filosofia, matemática, economia, neurociência, psicologia, engenharia de computadores e linguística. (RUSSEL e NORVIG, 2003).

As redes neurais é um dos paradigmas da Inteligência Artificial (IA), a base principal dessas é a simulação de neurônios do cérebro humano para neurônios sistêmicos, que se comunicam por meio de ligação (por isso o conceito de “rede”) (ZANELATO et al, 2017). Para que a rede neural funcione é necessária a utilização de algoritmos que aprendam

com base em exemplos passados. Conforme a rede neural recebe as informações necessárias, é possível desenvolver máquinas mais inteligentes (MARSLAND, 2014).

Uma Rede Neural Convolucional é um algoritmo de Aprendizado Profundo que pode captar uma imagem de entrada, atribuir importância (pesos e bias que podem ser aprendidos) a vários aspectos / objetos da imagem e ser capaz de diferenciar um do outro. (GÉRON, 2017). As redes neurais convolucionais são comumente utilizadas na classificação de imagens e na detecção de objetos. (GÉRON, 2017).

A arquitetura de uma Rede Neural Convolucional é análoga àquela do padrão de conectividade de neurônios no cérebro humano e foi inspirada na organização do Visual Cortex. Os neurônios individuais respondem a estímulos apenas em uma região restrita do campo visual conhecida como Campo Receptivo. (GÉRON, 2017).

De acordo com Kumar (2018), a camada de convolução é composta por 3 operações: campos receptivos locais, pesos compartilhados e *pooling*.

Na camada de campos receptivos locais cada pixel de uma imagem é um valor numérico que representa a intensidade de cor de acordo com a escala de cor utilizada, como RGB (*Red, Green, Blue*), por exemplo, ou apenas intensidade em escala de cinza para imagens em preto e branco.

Os pesos compartilhados são auxiliados para realizar os cálculos com base nos resultados das outras camadas.

De acordo com Géron (2017) na Rede Neural Convolucional existem camadas de *pooling*, que são geralmente usadas imediatamente após camadas convolucionais e o que fazem é simplificar as informações na saída da camada convolucional, ou seja, uma camada de *pooling* que recebe cada saída do mapa de características da camada e prepara um mapa de características condensadas.

Baseando-se nesses aspectos, a rede neural convolucional é ideal para utilizar no reconhecimento facial. Segundo Taigman et al (2014) o reconhecimento facial pode ser dividido em quatro partes, sendo elas: detecção, alinhamento, descrição e classificação:

- a) Detecção: É a utilização de um algoritmo para identificar faces em uma imagem, basicamente se passa um filtro pela imagem e utiliza a classificadores para determinar se a imagem apresentada representa uma face ou não.
- b) Alinhamento: Consiste em estimar a orientação da face e a posição dos olhos, da boca, das sobrancelhas e do nariz.

- c) Descrição: Consiste no desenvolvimento um descritor para a face humana de tal maneira que ao utilizar um classificador (próximo passo), possa ter uma acurácia alta.
- d) Classificação: Essa etapa existe um aprendizado de máquina profundo para encontrar a melhor forma de descrever essas faces.

A junção do reconhecimento das emoções faciais e a compreensão sobre inteligência emocional podem ajudar empresas a conseguir identificar algum tipo de mudança repentina de humor com um respectivo colaborador e assim notificar um superior sobre essa alteração, auxiliando esse superior em alguma tomada de decisão ou até mesmo uma conversa para identificar um possível problema que possa ter ocorrido.

Assim, o experimento desta pesquisa consistiu na aplicação de um algoritmo de reconhecimento de expressões e emoções faciais treinados a partir de Redes Neurais. Foram utilizadas as sete emoções básicas raiva, nojo, medo, feliz, triste, surpresa e neutro. Utilizou-se uma Rede Neural Convolucional supervisionada que passou por um processo de aprendizagem/treinamento com entradas e saídas definidas.

#### 4 RESULTADOS OBTIDOS

Para o desenvolvimento da pesquisa, foi utilizada uma aplicação pronta que foi extraída no repositório do *GitHub*.

Para realizar o treinamento da rede neural, validando as expressões, utilizou-se a base de dados *Challenges in Representation Learning: Facial Expression Recognition Challenge*. A base de dados utilizada para o treinamento, concentra um total de 35.888 exemplos de imagens de rostos em escala de cinza de 48x48 pixels. Os rostos foram registrados automaticamente para que ficassem mais ou menos centralizado e ocupassem aproximadamente a mesma quantidade de espaço em cada imagem.

O arquivo em CSV (*Comma Separated Values*), denominado “Fer2013.csv”, contém duas colunas, emoção e pixels. A coluna “emoção” contém um código numérico que varia de 0 a 6 (0 = Raiva, 1 = Nojo, 2 = Medo, 3 = Feliz, 4 = Triste, 5 = Surpresa, 6 = Neutro), esse código serve para a identificação da emoção que está sendo expressa pela imagem. A coluna “pixels” contém uma sequência de números que são os pixels que cada imagem contém. Após a organização e definição da base de dados, o algoritmo transformava os pixels em imagens para que assim fosse possível uma melhor identificação dos rostos.

Depois da transformação dos pixels em imagens, o modelo da rede neural foi escrito utilizando a biblioteca do *Keras* e a linguagem de programação *Python*. A junção dessas duas ferramentas possibilitou a construção das camadas de identificação da rede neural,

pois cada modelo é composto por um ou mais camadas (a quantidade de camadas varia de acordo com o problema a ser solucionado).

Para este estudo foram escritas quatro camadas do modelo da rede neural. As imagens passam pelas camadas da rede e assim a rede compara os pontos, as semelhanças e analisa as saídas (que são as emoções de cada rosto). Quando o processo de treinamento é passado em cada camada da rede, os pesos são gerados por meio da função de ativação da rede neural, apresentando uma assertividade nos resultados da identificação das imagens e expressões. A quantidade de épocas que a rede neural irá tentar executar é definida no início no código do modelo, essa quantidade de épocas serve para definir quantas vezes essa rede neural passou pelas camadas para gerar o “conhecimento” da aplicação.

Após a etapa de treinamento com o arquivo *Fer2013.csv*, a rede neural utiliza o arquivo de *test.csv* para avaliar os resultados obtidos e gerar o modelo e os pesos. Esses valores são alterados a cada interação com cada época. A cada passagem pelas épocas observa-se ainda a porcentagem de acerto (*acc*) e perda (*loss*) sendo alterada, pois essa é a forma que a rede neural apresenta os resultados.

Quando o algoritmo da rede neural termina de passar por todas as épocas definidas no início da aplicação, o modelo e os pesos que foram definidos por meio do aprendizado, são salvos em um arquivo com o formato de *hdf5* (esse é o formato comum para armazenar modelos e pesos de uma rede neural) por meio da chamada da função de *call-back*. Assim, é possível utilizá-los em outras aplicações. A rede neural convolucional utilizada apresentou uma acurácia em torno de 65% após o aprendizado e os testes.

Para avaliar o modelo e os pesos gerados pela rede neural, foi desenvolvida uma aplicação com *OpenCV* para reconhecimento da face e da expressão em tempo real.

As capturas das expressões faciais em tempo real são armazenadas em uma pasta junto de um arquivo de log (contém cada expressão) e com isso, é possível validar as expressões comparando o arquivo de log com as imagens capturadas pela aplicação.

Para validar a acuracidade da aplicação, foi realizado uma conferência dos frames obtidos com a filmagem dos colaboradores em tempo real. A filmagem foi feita a cada hora e a sua duração foi de 2 segundos. Isso foi o suficiente para adquirir um total de 2 mil imagens de cada colaborador, no decorrer de uma semana de trabalho (trabalhando 8 horas por dia e 5 dias por semana).

Na Figura 1 é possível visualizar um exemplo do arquivo de log com as chances de ser essa emoção, além do frame capturado do colaborador 1.

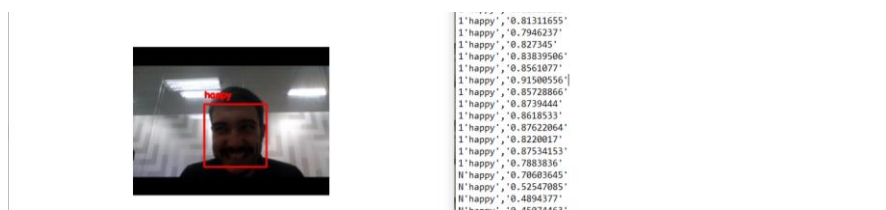


Figura 1: Log do sistema com o Frame no ambiente de trabalho, colaborador 1

Fonte: Elaborado pelo autor

O processo realizado com o colaborador 1 foi realizado com os demais. A partir do exemplo acima, foi possível validar as expressões faciais realizadas pelos colaboradores, com cada frame e com o arquivo de log. Os resultados obtidos pelo arquivo de log foram adicionados à um arquivo do *Excel* para validação dos acertos e erros, além das montagens dos gráficos, baseado na quantidade dos frames.

A acurácia do modelo foi maior com as expressões de feliz, neutra e raiva. A Tabela 1 sintetiza os resultados obtidos pela aplicação. Percebeu-se também que, para o colaborador que utilizava um volume maior de barba, a identificação das expressões foi mais difícil, isso pode ter ocorrido devido à uma leve cobertura de sua boca.

Tabela 1: Sintetização dos resultados.

Expressão	Acurácia
Raiva	Em torno de 80% a 90%
Feliz	Em torno de 58% a 61%
Neutro	Em torno de 95% a 99%
Triste	Em torno de 18% a 32%
Surpreso	Em torno de 13% a 18%

Fonte: Elaborado pelo autor.

## 5 CONSIDERAÇÕES FINAIS

Por se tratar da primeira versão da pesquisa, a aplicação foi submetida a uma análise no ambiente empresarial. Durante a utilização da aplicação, foram capturados os *frames* em tempo real das expressões realizadas pelos três colaboradores e armazenados tanto os *frames* quanto um arquivo de log, para assim ser possível a identificação das expressões faciais e analisá-las se realmente o modelo teve uma acurácia satisfatória.

Após a análise dos resultados obtidos nesta pesquisa, percebeu-se que, em alguns casos, quando seria acionado a aplicação para captura dos rostos e das expressões, os colaboradores tentaram forçar algumas expressões, reduzindo assim a acurácia dos resultados, pois o ideal era validar as expressões naturais de cada colaborador em seu ambiente de trabalho, por isso as expressões de assustado e triste resultaram em muitos falsos positivos. Entretanto, a acurácia do modelo se torna mais efetiva com as expressões de feliz, neutra e raiva, pois são nessas expressões que o índice de acerto são maiores nos

três casos de teste. Com esses resultados considera-se que o objetivo desta pesquisa foi atingido, sendo possível a captura das expressões faciais em tempo real pela aplicação e a avaliação das expressões e emoções expressas pelos colaboradores.

## REFERÊNCIAS

BERGAMINI, C. W.: **Inteligência Emocional**. RAE-Revista de Administração de Empresas, v. 36, n. 4, out-dez, p.74-77, 1996. Disponível em: <<http://dx.doi.org/10.1590/S0034-75901996000400008>> Acesso: 15 de outubro de 2019.

GÉRON, A.: **Hands-On Machine Learning with Scikit-Learn & TensorFlow - Concepts, Tools, and Techniques To Build Intelligent Systems**. 2017.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

KUMAR, A.: **Achieving 90% accuracy in Object Recognition Task on CIFAR-10 Dataset with Keras: Convolutional Neural Networks**. 2018. Disponível em: <<https://appliedmachinelearning.blog/2018/03/24/achieving-90-accuracy-in-object-recognition-task-on-cifar-10-dataset-with-keras-convolutional-neural-networks/>> Acesso em: 15 de outubro de 2019.

KUMAR, A.: **Demonstration of Facial Emotion Recognition on Real Time Video Using CNN : Python & Keras**. 2018. Disponível em: <<https://appliedmachinelearning.blog/2018/11/28/demonstration-of-facial-emotion-recognition-on-real-time-video-using-cnn-python-keras/>> Acesso em: 09 de setembro de 2019.

MARSLAND, S.: **Machine Learning: An Algorithmic Perspective** (2º Edição) 2009.

RUSELL, S., NORVIG, P.: **Inteligência Artificial (tradução da segunda edição)** Editora campus. 2003.

TAIGMAN, Y., YANG, M., RANZATO, M. A., WOLF L.: **DeepFace: Closing the Gap to Human-Level Performance in Face Verification**. 2014. Disponível em: <[https://www.cs.toronto.edu/~ranzato/publications/taigman\\_cvpr14.pdf](https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf)> Acesso em 27 de agosto de 2019.

ZANELATO, A. L., FERREIRA, C. B.: **Sistema de reconhecimento de expressões faciais**. 2017. Disponível em: <[http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/8459/1/CT\\_COENC\\_2017\\_1\\_6.pdf](http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/8459/1/CT_COENC_2017_1_6.pdf)> Acesso em: 20 de março de 2019.