



UNIDADE DE ENSINO SUPERIOR DE FEIRA DE SANTANA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

AURINO GUILHERME XAVIER DE ARAUJO FILHO  
ODael LOPES DE SIQUEIRA NETO  
YANN SILVA AMARANTE

**REDES NEURAIIS CONVOLUCIONAIS *ALEXNET*, *GOOGLNET* E *RESNET* APLICADAS  
AO RECONHECIMENTO DE EXPRESSÕES FACIAIS**

**FEIRA DE SANTANA  
Julho de 2021**

AURINO GUILHERME XAVIER DE ARAUJO FILHO  
ODAEL LOPES DE SIQUEIRA NETO  
YANN SILVA AMARANTE

**REDES NEURAIIS CONVOLUCIONAIS *ALEXNET*, *GOOGLNET* E *RESNET* APLICADAS  
AO RECONHECIMENTO DE EXPRESSÕES FACIAIS**

Trabalho de conclusão de curso da Unidade de Ensino Superior de Feira de Santana, como requisito para obtenção do título de Bacharel em Engenharia Elétrica pelos discentes Aurino Guilherme Xavier de Araujo Filho, Yann Silva Amarante, e do título de Bacharel em Engenharia Mecânica pelo discente Odael Lopes de Siqueira Neto.

**Orientador:** Prof. Me. Márcio da Silva Pereira Bove.

**Co-orientador:** Prof. Me. Reginaldo Barnabé Gonzalez Grimaldi.

**FEIRA DE SANTANA**

**Julho de 2021**

AURINO GUILHERME XAVIER DE ARAUJO FILHO  
ODAEL LOPES DE SIQUEIRA NETO  
YANN SILVA AMARANTE

**REDES NEURAIS CONVOLUCIONAIS ALEXNET, GOOGLNET E RESNET APLICADAS AO  
RECONHECIMENTO DE EXPRESSÕES FACIAIS**

Feria de Santana, 05 de julho de 2021

Banca Examinadora:

---

**Prof. Me. Márcio da Silva Pereira Bove**  
Unidade de Ensino Superior de Feira de Santana  
Orientador

---

**Prof. Me. Reginaldo Barnabé Gonzalez Grimaldi**  
Unidade de Ensino Superior de Feira de Santana  
Avaliador

---

**Prof. Me. Flávio Couvo Teixeira Maia**  
Unidade de Ensino Superior de Feira de Santana  
Avaliador

---

**Prof. Me. João Filipe Lacerda Cruz**  
Unidade de Ensino Superior de Feira de Santana  
Avaliador

## **AGRADECIMENTOS**

Agradecemos a Deus, em primeiro lugar, pela nossa vida e pelo dom da perseverança, que nos permitiu concluir este trabalho.

Agradecemos também aos nossos pais por terem se esforçado tanto para nos proporcionar uma boa educação, por terem nos alimentado com saúde, força e coragem, as quais foram essenciais para a superação de todas as adversidades ao longo desta caminhada.

Agradecemos também a nossa família que, com todo carinho e apoio, não mediu esforços para nos motivar a chegar nesta etapa de nossas vidas.

Agradecemos ao nosso orientador Prof. Me. Márcio da Silva Pereira Bove pela paciência e conhecimentos passados para a execução deste trabalho.

Enfim, agradecemos a todos que, de alguma forma, passaram por nossas vidas e contribuíram para a construção de quem somos hoje.

*“A persistência é o caminho do êxito.”*  
(Charles Chaplin)

## LISTA DE ILUSTRAÇÕES

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1- Arquitetura da rede <i>AlexNet</i> . Fonte: Autores.....                                                                                              | 12 |
| Figura 2 - Módulo <i>Inception</i> . Fonte: Autores.....                                                                                                        | 13 |
| Figura 3 - Conexão de salto da rede <i>ResNet</i> . Fonte: Autores.....                                                                                         | 14 |
| Figura 4 - Seleção dos atores da base de dados RAVDESS. Fonte: Autores.....                                                                                     | 15 |
| Figura 5 - Modelo proposto para classificação dos estados afetivos da face com as redes <i>AlexNet</i> , <i>GoogleNet</i> e <i>ResNet</i> . Fonte: Autores..... | 15 |
| Figura 6 - Processo de treinamento da rede <i>AlexNet</i> . Fonte: Autores. ....                                                                                | 16 |
| Figura 7 - Matriz de confusão da rede <i>AlexNet</i> para etapa de classificação. Fonte: Autores. ....                                                          | 16 |
| Figura 8 - Matriz de confusão da rede <i>AlexNet</i> para o teste <i>holdout</i> . Fonte: Autores. ....                                                         | 17 |
| Figura 9 - Progresso de treinamento da rede <i>GoogLeNet</i> . Fonte: Autores.....                                                                              | 17 |
| Figura 10 - Matriz de confusão da rede <i>GoogLeNet</i> para etapa de classificação. Fonte: Autores. ....                                                       | 17 |
| Figura 11 - Matriz de confusão da rede <i>GoogLeNet</i> para o teste <i>holdout</i> . Fonte: Autores. ....                                                      | 18 |
| Figura 12 - Progresso de treinamento da <i>ResNet18</i> . Fonte: Autores.....                                                                                   | 18 |
| Figura 13 - Matriz de confusão da rede <i>ResNet18</i> para etapa de classificação. Fonte: Autores. ....                                                        | 18 |
| Figura 14 - Matriz de confusão da rede <i>ResNet18</i> para o teste <i>holdout</i> . Fonte: Autores. ....                                                       | 18 |
| Figura 15 - Classificações corretas com a rede <i>AlexNet</i> . Fonte: Autores.....                                                                             | 19 |
| Figura 16 - Classificações corretas com a rede <i>GoogLeNet</i> . Fonte: Autores. ....                                                                          | 19 |
| Figura 17 - Classificações corretas com a rede <i>ResNet</i> . Fonte: Autores. ....                                                                             | 19 |
| Figura 18 - Classificações incorretas com as redes <i>AlexNet</i> , <i>GoogLeNet</i> e <i>Resnet18</i> . Fonte: Autores.....                                    | 19 |

## LISTA DE TABELAS

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Tabela 1- Imagens utilizadas nas etapas de treinamento e teste. Fonte: Autores. .... | 15 |
|--------------------------------------------------------------------------------------|----|

## LISTA DE ABREVIATURAS E SIGLAS

|         |                                                                   |
|---------|-------------------------------------------------------------------|
| CNN     | <i>Convolutional Neural Network</i>                               |
| ILSVRC  | <i>ImageNet Large-Scale Visual Recognition Challenge</i>          |
| IHC     | Interface Homem-Máquina                                           |
| RAVDESS | <i>Ryerson Audio-Visual Database of Emotional Speech and Song</i> |



## SUMÁRIO

|      |                                                                                             |    |
|------|---------------------------------------------------------------------------------------------|----|
| 1.   | Introdução.....                                                                             | 10 |
| 2.   | Fundamentação Teórica .....                                                                 | 11 |
| 2.1. | Reconhecimento de Expressões Faciais.....                                                   | 11 |
| 2.2. | Redes Convolucionais.....                                                                   | 12 |
| 2.3. | Rede Convolucional AlexNet.....                                                             | 12 |
| 2.4. | Rede Convolucional GoogLeNet .....                                                          | 13 |
| 2.5. | Rede Convolucional ResNet .....                                                             | 13 |
| 3.   | Métodos e Materiais .....                                                                   | 14 |
| 3.1. | Método de Avaliação do Algoritmo .....                                                      | 14 |
| 3.2. | Recursos do Trabalho.....                                                                   | 14 |
| 3.3. | Banco de Dados RAVDESS.....                                                                 | 14 |
| 3.4. | Base de Dados de Treinamento .....                                                          | 15 |
| 3.5. | Modelo Proposto .....                                                                       | 15 |
| 3.6. | Treinamento Rede AlexNet.....                                                               | 16 |
| 3.7. | Treinamento Rede GoogLeNet.....                                                             | 16 |
| 3.8. | Treinamento Rede ResNet .....                                                               | 16 |
| 4.   | Resultados .....                                                                            | 16 |
| 4.1. | Performance Rede AlexNet .....                                                              | 16 |
| 4.2. | Performance Rede GoogLeNet.....                                                             | 17 |
| 4.3. | Performance ResNet18.....                                                                   | 18 |
| 4.4. | Classificação dos Estados Afetivos da Face.....                                             | 19 |
| 4.5. | Comparação dos resultados para classificação dos estados afetivos da face com imagens ..... | 20 |
| 5.   | Conclusão .....                                                                             | 20 |

# REDES NEURAIIS CONVOLUCIONAIS ALEXNET, GOOGLNET E RESNET APLICADAS AO RECONHECIMENTO DE EXPRESSÕES FACIAIS

AURINO GUILHERME XAVIER DE ARAUJO FILHO, ODAEL LOPES DE SIQUEIRA NETO, YANN SILVA AMARANTE

Departamento de Engenharia Elétrica, Unidade de Ensino Superior de Feira de Santana,  
Feira de Santana, Bahia, Brasil.

**Abstract:** This work investigates the performance of convolutional networks AlexNet, GoogLeNet and ResNet applied to the recognition of facial expressions using image resources from The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). The networks are trained with eight basic affective states of the human face considering four different actors from the RAVDESS database. The performance presented by the AlexNet, GoogLeNet and ResNet networks for the classification step are 99%, 98.1% and 99.3% and for the generalization test based on the Holdout cross validation method a success rate of 91.7%, 95.5% and 88.5%. These results confirm the effectiveness of AlexNet, GoogLeNet and ResNet networks for correct prediction of facial expressions that are present in the database and for new facial expressions similar to those used in the training phase.

**Key-words:** Facial Expressions, Convolutional Neural Network, AlexNet, GoogleNet, ResNet.

**Resumo:** Este trabalho investiga a performance das redes convolucionais AlexNet, GoogLeNet e ResNet aplicadas ao reconhecimento de expressões faciais usando os recursos de imagens do banco de dados The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). As redes são treinadas com oito estados afetivos básicos da face humana considerando quatro atores diferentes do banco de dados RAVDESS. A performance apresentada pelas redes AlexNet, GoogLeNet e ResNet para etapa de classificação são de 99%, 98,1% e 99,3% e para o teste generalização com base no método de validação cruzada Holdout uma taxa de sucesso de 91,7%, 95,5% e 88,5%. Estes resultados confirmam a eficácia das redes AlexNet, GoogLeNet e ResNet para previsão correta de expressões faciais que estão presentes no banco de dados e para novas expressões faciais similares as usadas na fase de treinamento.

**Palavras-chave:** Expressões Faciais, Rede Neurais Convolucionais, AlexNet, GoogleNet, ResNet.

## 1. Introdução

As Redes Neurais Convolucionais, do inglês *Convolutional Neural Network* (CNN) são uma categoria de rede neural que se destacam em aplicações com o reconhecimento da voz (SZE et al., 2017) e no campo da visão computacional com o reconhecimento de imagens (WANG et al., 2020), potencializando o desenvolvimento de soluções tecnológicas inteligentes (BASBRAIN et al., 2018). A partir do avanço nesses dois campos de pesquisa houve um aumento na utilização das CNN, resultando em aplicações diversas que vão desde de carros com direção automática, detecção de câncer, reconhecimento facial, até jogabilidade de jogos complexos (SZE et al., 2017).

A arquitetura geral das CNN considera as etapas de extração de atributos das

imagens e em seguida a sua respectiva classificação. Em síntese, as CNN extraem características centrais das imagens fornecidas para serem usadas em sua classificação. Durante esse processo, há o treinamento da rede neural, objetivando aprender a classificar corretamente as imagens (WANG et al., 2020).

O desafio de reconhecimento visual em grande escala da *ImageNet*, do inglês *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) avalia algoritmos para detecção de objetos e classificação de imagens em grande escala, visando contribuir com o progresso de tecnologias no campo da visão computacional, sendo as redes AlexNet, GoogLeNet e ResNet vencedoras das edições 2012, 2014 e 2015. Além dessas CNN outras arquiteturas obtiveram bons resultados na competição ao longo dos anos (DONG, LIANG, 2019).

A face humana está relacionada a diversas atividades cotidianas, e por meio dela ocorre a comunicação entre as pessoas, sendo possível identificar e verificar um indivíduo além de observar suas expressões faciais (PANTIC, PATRAS, 2006). O reconhecimento facial passa a ser empregado com maior frequência em áreas de identificação biométrica, baseada no escaneamento da face humana, na qual se compara as imagens armazenadas em um banco de dados com a imagem apresentada na biometria, possibilitando o reconhecimento de informações faciais (PREMA, SHANMUGAPRIYA, 2017).

Reconhecer a face humana pode proporcionar benefícios em diversas áreas da vida, podendo ser usada para medidas de segurança, como vigilância em áreas privadas e públicas, biometria, controle de acesso e na investigação de desaparecidos ou criminosos. Além de ter utilidade no *marketing*, Interação Homem-Máquina (IHM), redes sociais, entre outras aplicações (RESHMA, KANNAN, 2019).

O banco de dados Audiovisual Ryerson de Fala Emocional e Canção, do inglês *The Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS) usado neste trabalho disponibiliza um conjunto de vídeos de 24 atores profissionais (12 mulheres e 12 homens) com os estados afetivos básicos da face (calmo, feliz, triste, raiva, medo, surpreso e nojo) (LIVINGSTONE, 2018).

Assim, esse trabalho investiga as redes *AlexNet*, *GoogLeNet* e *ResNet* na tarefa de classificação, por meio de imagens, dos estados afetivos básicos da face humana para uma abordagem multiclases. As redes neurais são treinadas com as imagens de expressões faciais básicas de 4 atores profissionais do banco de dados RAVDESS. As métricas de avaliação computacional consideram o desempenho de treinamento das redes neurais, performance de classificação dos estados

afetivos da face e capacidade de generalização das redes neurais através do método de validação cruzada *Holdout*. Logo, esta pesquisa é motivada pelo desejo de contribuir com o desenvolvimento de sistemas inteligentes aplicados ao reconhecimento de expressões faciais com o uso das redes *AlexNet*, *GoogLeNet* e *ResNet*.

A organização do artigo está estruturada do seguinte modo: no segundo capítulo apresentam-se as características técnicas acerca do reconhecimento de expressões faciais e das redes *AlexNet*, *GoogLeNet* e *ResNet*. O terceiro capítulo apresenta o método de avaliação do algoritmo, recursos do trabalho, bancos de dados, modelo proposto e método de treinamento das redes neurais. No quarto capítulo são apresentados os resultados obtidos nas fases de treinamento, classificação das expressões faciais e teste de validação cruzada, com as redes *AlexNet*, *GoogLeNet* e *ResNet*. Ainda no quarto capítulo compara-se os resultados obtidos nesta pesquisa com outros trabalhos. Por fim, no quinto capítulo apresenta-se a conclusão.

## **2. Fundamentação Teórica**

### **2.1. Reconhecimento de Expressões Faciais**

Os seres humanos podem se expressar de diferentes formas e uma delas é através da face, a qual pode apresentar sinais relacionados as emoções e intenções de uma pessoa. O estudo dos sinais provenientes da face pode ser realizado através de imagens ou vídeos e exigirão uma avaliação das informações faciais em tempo real objetivando o reconhecimento de informações que possam identificar algum padrão (LIU et al., 2014).

A facilidade e velocidade em que o cérebro humano encontra em discernir as expressões faciais não pode ser reproduzida por máquinas, apesar dos avanços

tecnológicos nos últimos anos. Assim, ainda existem diversos obstáculos a serem vencidos, tais como: oclusão, iluminação, posição da face e interferências do plano de fundo. Logo, essas variáveis tornam o reconhecimento facial por sistemas computacionais algo ainda instável, porém com grande utilidade e importância para o desenvolvimento de sistemas tecnológicos inteligentes (LI, 2019).

As emoções básicas (raiva, medo, desprezo, tristeza, felicidade e surpresa) e neutralidade podem ser transmitidas por expressões faciais. Levando isso em consideração, numerosas pesquisas nessa área observaram que existem regiões específicas da face humana, as quais mudam de acordo com as expressões e estão localizadas no entorno da boca, olhos e nariz, facilitando assim o reconhecimento de expressões faciais através da visão computacional (KHORRAMI; PAINE; HUANG, 2015).

## 2.2. Redes Convolucionais

A pesquisa de Hubel e Wiesel (1959) com estudos acerca dos córtices visuais de macacos e pássaros contribuíram para os primeiros passos no desenvolvimento das CNN. Esse trabalho inspirou Kunihiko Fukushima, na década de 80, a introduzir o processo de convolução na área das redes neurais, chamado de *neocognitron*. Contudo, Yann Le Cunn (1989) apresentou um trabalho com uma rede convolucional de sete níveis usando o algoritmo de retropropagação, chamado *LeNet-5*, sendo a mais significativa contribuição para o desenvolvimento das CNN. As principais arquiteturas de CNN atuais são derivações da *LeNet-5* (AJIT; ACHARYA; SAMANTA, 2020).

As CNN são constituídas por três tipos de camadas: camada de convolução, *pooling* e a camada totalmente conectada. Na camada de convolução ocorre a multiplicação de cada *pixel* da imagem

fornecida por um filtro (*kernel*), gerando assim um mapa de ativação, tendo este mapa a função de armazenar características distintas da imagem de forma resumida e como consequência reduzir os dados a serem processados. A camada *pooling* consiste em reduzir a dimensão espacial do mapa de ativação mantendo as informações mais relevantes, ou seja, passar uma informação sucinta sem perder a eficiência no resultado. Por último, tem-se a camada totalmente conectada em que transcorre todo o raciocínio computacional (AJIT; ACHARYA; SAMANTA, 2020).

## 2.3. Rede Convolucional AlexNet

A rede *AlexNet* foi projetada por Alex Krizhevsky (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A arquitetura da rede *AlexNet* é composta pelas camadas de *Convolution*, *ReLU*, *Max Pooling*, *Fully Connected* e *Softmax*. A rede *AlexNet* possui 630 milhões links, 60 milhões de parâmetros para serem treinados e 650.000 neurônios (WANG; HAN, 2020). Entre as camadas, existem camadas de entrada e saída, camadas de unidades lineares retificadas (*ReLU*), camadas de normalização, camadas de agrupamento, camadas de eliminação (*Dropout*), camada *Softmax* (*Probabilistic*), camadas convolucionais (*Convolution*) e camadas totalmente conectadas (*Fully Connected*). Na Figura 1 é apresentada a arquitetura da rede *AlexNet* para entrada de imagens com dimensão  $227 \times 227 \times 3$ .

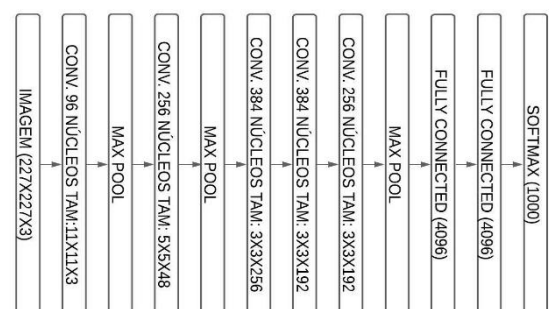


Figura 1- Arquitetura da rede *AlexNet*. Fonte: Autores.

A primeira camada convolucional é responsável por filtrar a imagem de entrada que possui 96 núcleos de tamanho  $11 \times 11 \times 3$  com uma distância de quatro *pixels*. A segunda camada convolucional é empregada para normalizar e agrupar a resposta de saída da primeira camada convolucional e filtrar com 256 núcleos de tamanho  $5 \times 5 \times 48$ . A terceira camada convolucional possui 384 núcleos de tamanho  $3 \times 3 \times 256$  que normaliza e agrupa a resposta de saída da segunda camada convolucional. A quarta camada convolucional possui 384 núcleos de tamanho  $3 \times 3 \times 192$ , e a quinta camada convolucional tem 256 núcleos de tamanho  $3 \times 3 \times 192$ . Já as camadas totalmente conectadas possuem 4096 neurônios. Por fim, a camada *softmax* fornece níveis de ativação probabilístico na saída da rede (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

A camada *ReLU* executa uma operação de limite para cada elemento da entrada, onde qualquer valor menor que zero é definido como zero, sua utilização acelera a velocidade do treinamento e soluciona o problema de dispersão de gradiente quando a rede é profunda. Na camada *Max Pooling* executa-se a redução da resolução dividindo a entrada em regiões de agrupamento retangulares e calculando o máximo de cada região. A técnica de *Dropout* elimina aleatoriamente elementos da matriz de entrada impondo um nível de ativação zero com uma determinada probabilidade, a fim de reduzir a complexidade da adaptação dos neurônios para evitar *overfitting* (DONG; LIANG, 2019), problema comum em redes com ajuste excessivo de parâmetros (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

#### 2.4. Rede Convolucional GoogLeNet

A arquitetura da rede *GoogLeNet* é formada por 22 camadas profundas e aproximadamente 5 milhões de parâmetros.

Em comparação com outras redes neurais, tornou-se mais profunda que as arquiteturas anteriores. Entretanto, o número de parâmetros é considerado relativamente baixo. Um diferencial da rede *GoogLeNet* é a utilização dos chamados módulos *Inception* que operam com diferentes tipos de filtros visando a redução do custo computacional na etapa de treinamento (AJIT; ACHARYA; SAMANTA, 2020). Na Figura 2 é apresentado o fluxograma com o módulo *Inception*.

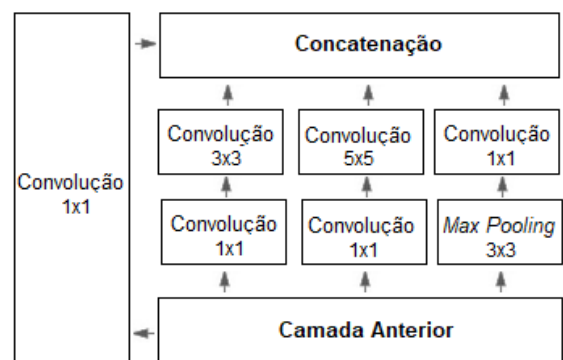


Figura 2 - Módulo *Inception*. Fonte: Autores

Dentro de um módulo *Inception* haverá filtros 1x1, 3x3 e 5x5, camadas *max pooling* 3x3 e, por fim, uma camada de concatenação, sendo esta responsável pelo encaminhamento de um único sinal para camada da rede posterior, sinal este proveniente da captação e união de todas as saídas das camadas de convolução existentes no módulo. Resumidamente, a rede *GoogLeNet* é formada por camadas de convolução, módulos *Inception*, *Max Pooling*, camadas de *Local Normalization Response*, *Softmax* e camadas *Fully Connected* (CRUZ, 2019).

#### 2.5. Rede Convolucional ResNet

A rede *ResNet* foi primeiramente apresentada com 152 camadas, mostrando ser muito profunda e teoricamente suscetível a eventuais erros de degradação. Com o intuito de contornar essa complicação ocorre a introdução dos blocos residuais. A inovação traz a possibilidade de os neurônios de uma camada propagarem

sinais para outra camada distante sem seguir os modelos anteriores, onde essa passagem de informação só é possível para camadas vizinhas. Esse método é chamado de conexão de salto (CRUZ, 2019).

Existem várias versões da rede *ResNet* com diferentes tamanhos sendo capaz de ter versões com 18, 34, 50, 101, 152 ou 200 camadas (SZE et al., 2017). Em relação ao número de parâmetros a rede dispõe aproximadamente de 65 milhões. Além disso, o método de *Batch Normalization* é usado após cada convolução (AJIT; ACHARYA; SAMANTA, 2020). Na Figura 3 é apresentada o fluxograma da arquitetura para rede *ResNet* com ênfase na conexão de salto.

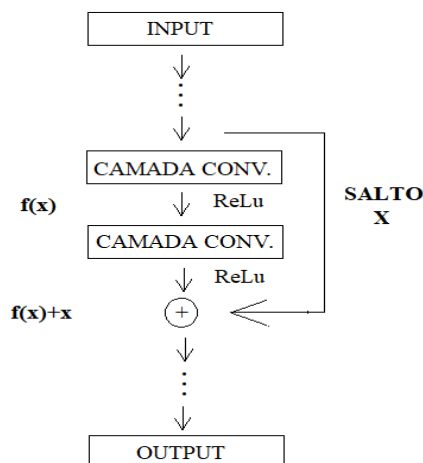


Figura 3 - Conexão de salto da rede *ResNet*. Fonte: Autores.

### 3. Métodos e Materiais

Neste capítulo descreve-se sobre os métodos empregados para a avaliação das redes *AlexNet*, *GoogleNet* e *ResNet* acerca da acurácia de classificação e teste de validação cruzada *Holdout*. Apresentam-se, também, os recursos utilizados para realização dos experimentos, assim como a seleção das imagens para as etapas de treinamento e teste das redes CNN.

#### 3.1. Método de Avaliação do Algoritmo

Para avaliação da acurácia utiliza-se o gráfico de evolução do treinamento juntamente com a matriz de confusão, sendo a matriz de confusão uma métrica frequentemente utilizada para relatar a performance de sistemas de classificação. A interpretação de seus resultados ocorre através da observação dos valores contidos na diagonal principal, a qual informa as quantidades de acertos, e o restante dos valores fora dessa diagonal representa o total de erros ou confusões (ARIZA-LOPEZ; RODRIGUEZ-AVI; ALBA-FERNANDEZ, 2018).

A técnica de validação cruzada *Holdout* avalia a performance do algoritmo em relação a capacidade de generalização, permitindo estimar a precisão do algoritmo acerca da tarefa de classificação para novos dados (MATHWORKS, 2021). Deste modo, a base de dados utilizada foi organizada em imagens de treinamento e teste de validação cruzada.

#### 3.2. Recursos do Trabalho

No desenvolvimento deste trabalho os experimentos foram realizados com o uso do software *MATLAB 2018.b* e um notebook HP *probook* com Processador: Intel Core i5-4300M CPU @ 2.60 GHz com 6 GB de RAM, Placa de vídeo: Intel HD Graphics 4600 e Sistema Operacional: 64 bits.

#### 3.3. Banco de Dados RAVDESS

O RAVDESS é um banco de dados multimodal validado de fala e música emocional. O banco de dados é equilibrado por gênero, sendo constituído com recursos de vídeos e áudio de 24 atores profissionais, vocalizando lexicalmente correspondentes declarações em um sotaque norte-americano neutro (LIVINGSTONE, 2018).

As expressões faciais incluem os estados afetivos calmo, alegria, tristeza, raiva, medo, surpresa, nojo e neutro em

dois níveis de intensidade emocional (normal e forte). O banco de dados RAVDESS fornece aos pesquisadores um conjunto de estímulos validado, gratuito e acessível sob uma licença *Creative Commons Attribution-NonCommercial-ShareAlike 4.0* (LIVINGSTONE, 2018).

### 3.4. Base de Dados de Treinamento

A base de dados do trabalho conta com as expressões faciais calma, feliz, medo, neutro, nojo, raiva, surpreso e triste proveniente da seleção de quatro atores profissionais do banco de dados RAVDESS. O conjunto de dados possui 815 imagens, das quais 658 imagens são destinadas para fase de treinamento e 157 imagens reservadas para o teste de validação cruzada *Holdout*. Na Tabela 1 são apresentadas as respectivas quantidades de imagens destinadas para fase de treinamento e teste de validação cruzada *Holdout* com uso das redes *AlexNet*, *GoogleNet* e *ResNet*.

Tabela 1- Imagens utilizadas nas etapas de treinamento e teste. Fonte: Autores.

| Estados afetivos da face | Imagens de treino | Imagens para teste de validação cruzada |
|--------------------------|-------------------|-----------------------------------------|
| Calmo                    | 88                | 23                                      |
| Feliz                    | 90                | 16                                      |
| Medo                     | 78                | 20                                      |
| Neutro                   | 64                | 21                                      |
| Nojo                     | 79                | 20                                      |
| Raiva                    | 83                | 18                                      |
| Surpreso                 | 89                | 21                                      |
| Triste                   | 87                | 18                                      |

Na Figura 4 é apresentado a seleção dos atores e estados afetivos da face da base de dados RAVDESS.



Figura 4 - Seleção dos atores da base de dados RAVDESS. Fonte: Autores.

### 3.5. Modelo Proposto

Na Figura 5 é apresentado o modelo empregado para avaliar a performance das redes *AlexNet*, *GoogleNet* e *ResNet*.

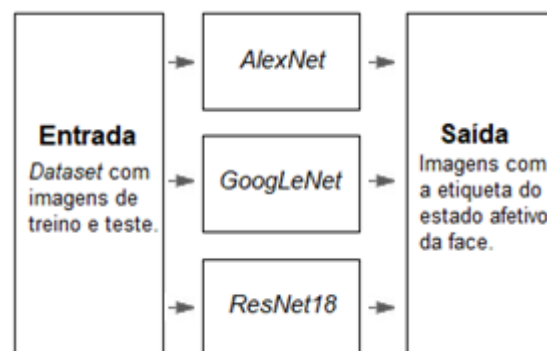


Figura 5 - Modelo proposto para classificação dos estados afetivos da face com as redes *AlexNet*, *GoogleNet* e *ResNet*. Fonte: Autores.



Neste modelo as redes *AlexNet*, *GoogleNet* e *ResNet* já treinadas são responsáveis por classificar os estados afetivos básicos da face conforme em imagens da base de dados RAVDESS que são apresentadas em suas respectivas entradas para as etapas de classificação e teste de validação cruzada *Holdout*. As imagens de saída são identificadas com a descrição do estado afetivo da face juntamente com o percentual de acerto.

### 3.6. Treinamento Rede *AlexNet*

O treinamento da rede *AlexNet* é realizado uma única vez, sendo a taxa de aprendizado de 0.0001 e o *minibatch* de 40. O quantitativo de 658 imagens destinada a fase de treinamento são particionadas em 80% para treino e 20% para teste. As duas camadas ocultas são parametrizadas cada uma com 1000 neurônios. A última camada é parametrizada com 8 classes.

### 3.7. Treinamento Rede *GoogLeNet*

A rede *GoogleNet* é treinada uma única vez, com uma taxa de aprendizagem de 0.0001 e *minibatch* de 40. O total de 658 imagens são organizadas em 80% para treino e 20% para teste. A última camada é parametrizada com 8 classes.

### 3.8. Treinamento Rede *ResNet*

No único treinamento da rede *ResNet*, a taxa de aprendizagem é parametrizada para 0.0001 e o *minibatch* para 40. O quantitativo de 658 imagens destinada a fase de treinamento são particionadas de em 80% treino e 20% teste. A última camada é parametrizada com 8 classes.

## 4. Resultados

Neste capítulo apresentam-se os resultados para as fases de treinamento, classificação e teste de validação cruzada com as redes *AlexNet*, *GoogLeNet* e *ResNet*

considerando o quantitativo de imagens da Tabela 1 e atores da Figura 4.

### 4.1. Performance Rede *AlexNet*

Na Figura 6 é apresentado a performance do treinamento da rede *AlexNet* com as respectivas curvas de acurácia para o desempenho de treinamento e validação de acordo com a evolução das iterações.

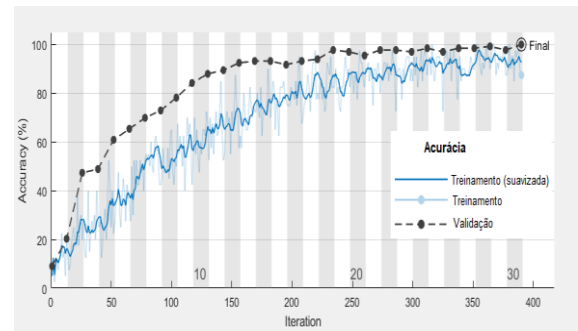


Figura 6 - Processo de treinamento da rede *AlexNet*.  
Fonte: Autores.

No decorrer do processo de treinamento da rede *AlexNet* a curva de validação apresenta uma melhor performance para a estabilidade e previsão acerca da acurácia comparado com as curvas de treinamento. Ao longo das iterações, progressivamente as curvas de validação e treinamento mostram o aumento no desempenho de acurácia. Na Figura 7 é apresentada a matriz de confusão obtida na etapa de classificação com a rede *AlexNet*.

|          |       |       |      |        |      |       |          |        |
|----------|-------|-------|------|--------|------|-------|----------|--------|
| calmo    | 88    | 0     | 0    | 0      | 0    | 0     | 0        | 0      |
| feliz    | 0     | 90    | 0    | 0      | 0    | 0     | 0        | 0      |
| medo     | 0     | 0     | 76   | 0      | 0    | 0     | 0        | 2      |
| neutro   | 0     | 0     | 0    | 64     | 0    | 0     | 0        | 0      |
| nojo     | 0     | 0     | 0    | 0      | 79   | 0     | 0        | 0      |
| raiva    | 2     | 1     | 1    | 0      | 0    | 79    | 0        | 0      |
| surpreso | 0     | 0     | 0    | 0      | 0    | 0     | 89       | 0      |
| triste   | 0     | 0     | 0    | 0      | 0    | 0     | 0        | 87     |
|          | calmo | feliz | medo | neutro | nojo | raiva | surpreso | triste |

Figura 7 - Matriz de confusão da rede *AlexNet* para



etapa de classificação. Fonte: Autores.

A matriz de confusão da fase de treinamento, apresentada na Figura 7, confirma a taxa de sucesso de 99% para acurácia de classificação das imagens. O estado afetivo medo é confundido com triste. Por fim, o estado afetivo raiva é confundido com calmo, feliz e medo. Na Figura 8 é apresentado os resultados obtidos com o teste de validação cruzada *Holdout* com a rede *AlexNet*.

|          |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|
| calmo    | 22 | 0  | 0  | 0  | 0  | 0  | 0  | 1  |
| feliz    | 0  | 14 | 0  | 2  | 0  | 0  | 0  | 0  |
| medo     | 0  | 0  | 20 | 0  | 0  | 0  | 0  | 0  |
| neutro   | 1  | 0  | 0  | 20 | 0  | 0  | 0  | 0  |
| nojo     | 0  | 0  | 0  | 0  | 20 | 0  | 0  | 0  |
| raiva    | 0  | 0  | 3  | 0  | 0  | 12 | 0  | 3  |
| surpreso | 0  | 0  | 0  | 0  | 0  | 3  | 18 | 0  |
| triste   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 18 |

Figura 8 - Matriz de confusão da rede *AlexNet* para o teste *holdout*. Fonte: Autores.

A rede *AlexNet* apresentou um percentual de 91,7% para previsão correta de novas imagens. As previsões incorretas ocorrem entre os estados afetivos calmo com triste, feliz com neutro, neutro com calmo, raiva com medo e triste, por fim, surpreso com raiva.

#### 4.2. Performance Rede GoogLeNet

Na Figura 9 é apresentado o processo de treinamento da rede *GoogLeNet* com as respectivas curvas de acurácia para o desempenho de treinamento e validação de acordo com a evolução das iterações.

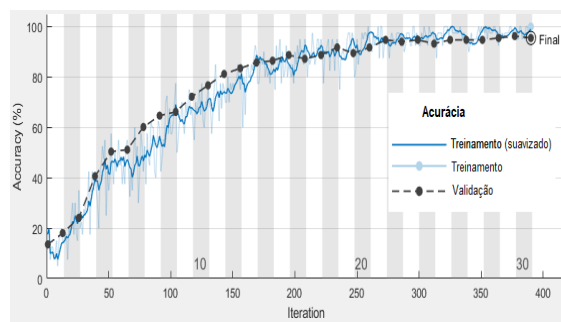


Figura 9 - Progresso de treinamento da rede *GoogLeNet*. Fonte: Autores.

Ao longo do processo de treinamento da rede *GoogLeNet*, as curvas progrediram com estabilidade e performance compatíveis para taxa de acurácia. Na Figura 10 é apresentada a matriz de confusão para etapa de treinamento com a rede *GoogLeNet*.

|          |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|
| calmo    | 88 | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| feliz    | 0  | 90 | 0  | 0  | 0  | 0  | 0  | 0  |
| medo     | 0  | 0  | 75 | 0  | 0  | 0  | 0  | 3  |
| neutro   | 0  | 0  | 0  | 64 | 0  | 0  | 0  | 0  |
| nojo     | 0  | 0  | 0  | 0  | 79 | 0  | 0  | 0  |
| raiva    | 3  | 3  | 0  | 0  | 0  | 76 | 0  | 1  |
| surpreso | 0  | 0  | 0  | 0  | 0  | 0  | 89 | 0  |
| triste   | 1  | 0  | 0  | 0  | 1  | 0  | 1  | 84 |

Figura 10 - Matriz de confusão da rede *GoogLeNet* para etapa de classificação. Fonte: Autores.

A matriz de confusão da Figura 10 apresenta uma taxa de 98,1% para classificação correta das imagens da fase de treinamento. As confusões ocorrem entre os estados afetivos medo com triste, raiva com calmo, raiva com feliz, raiva com triste, triste com calmo, triste com nojo e triste com surpreso. Na Figura 11 é demonstrado os resultados para o teste de validação cruzada *Holdout* com a rede *GoogLeNet*.

|          |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|
| calmo    | 23 | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| feliz    | 0  | 15 | 0  | 1  | 0  | 0  | 0  | 0  |
| medo     | 0  | 0  | 16 | 0  | 0  | 0  | 0  | 4  |
| neutro   | 0  | 0  | 0  | 21 | 0  | 0  | 0  | 0  |
| nojo     | 0  | 0  | 0  | 0  | 20 | 0  | 0  | 0  |
| raiva    | 0  | 0  | 0  | 1  | 0  | 17 | 0  | 0  |
| surpreso | 0  | 0  | 0  | 0  | 0  | 1  | 20 | 0  |
| triste   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 18 |

Figura 11 - Matriz de confusão da rede *GoogLeNet* para o teste *holdout*. Fonte: Autores.

A matriz de confusão da rede *GoogLeNet* apresentou um percentual de sucesso de 95,5% para previsão correta de novas imagens com os estados afetivos básicos da face. As confusões ocorrem entre os estados afetivos medo com triste, raiva com neutro e surpreso com raiva.

### 4.3. Performance ResNet18.

Na Figura 12 é apresentado o processo de treinamento da CNN *ResNet18* com as respectivas curvas de acurácia para o desempenho de treinamento e validação de acordo com a evolução das iterações.

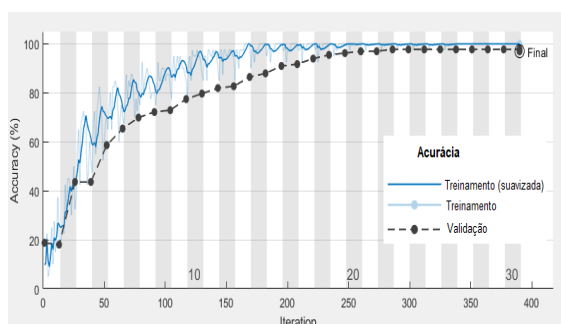


Figura 12 - Progresso de treinamento da *ResNet18*. Fonte: Autores.

Ao decorrer do processo de treinamento da rede *ResNet18* a curva de validação apresentou uma performance inferior para acurácia comparado com as curvas de treinamento. Ao final do processo as curvas

mostram desempenho de acurácia compatíveis.

|          |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|
| calmo    | 88 | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| feliz    | 0  | 90 | 0  | 0  | 0  | 0  | 0  | 0  |
| medo     | 0  | 0  | 76 | 0  | 0  | 1  | 1  | 0  |
| neutro   | 0  | 0  | 0  | 64 | 0  | 0  | 0  | 0  |
| nojo     | 0  | 0  | 0  | 0  | 79 | 0  | 0  | 0  |
| raiva    | 0  | 0  | 0  | 0  | 0  | 82 | 0  | 1  |
| surpreso | 0  | 0  | 0  | 0  | 0  | 0  | 89 | 0  |
| triste   | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 86 |

Figura 13 - Matriz de confusão da rede *ResNet18* para etapa de classificação. Fonte: Autores.

A rede *ResNet18* alcançou uma acurácia de classificação de 99,3%. O estado afetivo medo foi confundido com raiva e surpreso, o estado afetivo raiva com triste, por fim, triste com medo. Na Figura 14 é apresentado os resultados para o teste de validação cruzada *Holdout* com a rede *ResNet18*.

|          |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|
| calmo    | 21 | 1  | 0  | 0  | 0  | 1  | 0  | 0  |
| feliz    | 0  | 16 | 0  | 0  | 0  | 0  | 0  | 0  |
| medo     | 0  | 0  | 19 | 0  | 0  | 0  | 1  | 0  |
| neutro   | 0  | 0  | 0  | 21 | 0  | 0  | 0  | 0  |
| nojo     | 0  | 3  | 0  | 0  | 13 | 2  | 2  | 0  |
| raiva    | 0  | 0  | 0  | 0  | 2  | 14 | 0  | 2  |
| surpreso | 0  | 0  | 0  | 3  | 0  | 0  | 18 | 0  |
| triste   | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 17 |

Figura 14 - Matriz de confusão da rede *ResNet18* para o teste *holdout*. Fonte: Autores.

A matriz de confusão da rede *Resnet18* apresenta uma taxa de sucesso de 88,5% para previsão correta de imagens destinadas ao teste de validação cruzada descritas na Tabela 1.

#### 4.4. Classificação dos Estados Afetivos da Face

Nas Figuras 15, 16 e 17 são apresentadas as previsões corretas para o reconhecimento das expressões faciais em imagens.

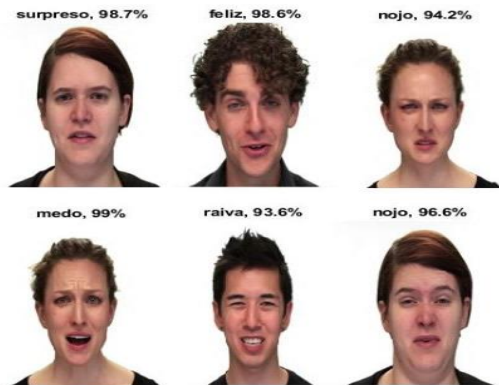


Figura 15 - Classificações corretas com a rede *AlexNet*. Fonte: Autores.

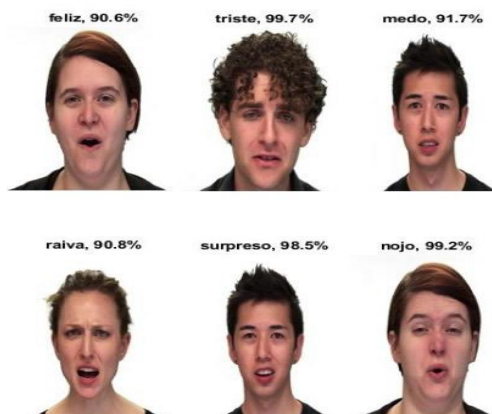


Figura 16 - Classificações corretas com a rede *GoogLeNet*. Fonte: Autores.



Figura 17 - Classificações corretas com a rede *ResNet*. Fonte: Autores.

As etiquetas empregadas nas imagens das Figuras 15, 16 e 17 com a descrição do estado afetivo da face juntamente com as respectivas probabilidades indicam a performance de acerto das CNN *AlexNet*, *GoogLeNet* e *ResNet18* ao classificar os estados afetivos da face em imagens. As imagens selecionadas mostram uma faixa entre 88,5% até 99,7% para previsão correta no reconhecimento das expressões faciais.

Na Figura 18 é apresentado algumas previsões incorretas obtidas nas etapas de classificação e teste de validação cruzada com as redes *AlexNet*, *GoogLeNet* e *Resnet18*.

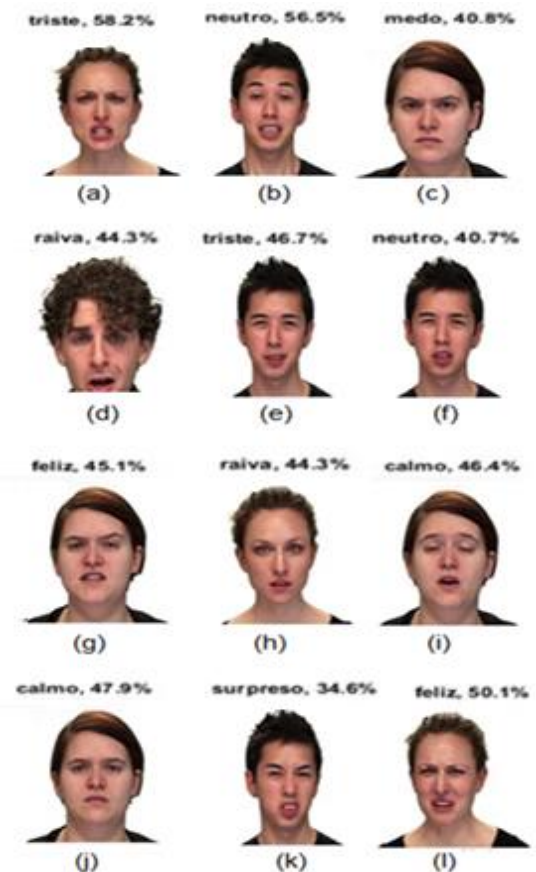


Figura 18 - Classificações incorretas com as redes *AlexNet*, *GoogLeNet* e *Resnet18*. Fonte: Autores.

As classificações incorretas ocorrem devido a semelhanças entre algumas expressões faciais, sendo (a) Raiva classificada incorreta como Triste; (b) Feliz classificada incorreta como Neutro; (c) Raiva classificada incorreta como Medo;

(d) Medo classificada incorreta como Raiva; (e) Calmo classificada incorreta como Triste; (f) Surpreso classificada incorreta como Neutro; (g) Raiva classificada incorreta como Feliz; (h) Surpreso classificada incorreta como Raiva; (i) Neutro classificada incorreta como Calmo; (j) Triste classificada incorreta como Calmo; (k) Nojo com Surpreso; (l) Nojo classificada incorreta como Feliz. As imagens selecionadas mostram resultados para uma faixa entre 34,6% até 58,2% com expressões faciais classificadas de forma incorreta, estes baixos níveis de ativação dos neurônios de saída para classificações incorretas confirmam a forte familiaridade entre as expressões faciais.

Com os resultados obtidos de acurácia com as redes *AlexNet*, *GoogLeNet* e *Resnet18* de 99%, 98,1% e 99,3% para a etapa de classificação e 91,7%, 95,5% e 88,5% para o teste de validação cruzada, constata-se a robustez das redes para classificação correta das expressões faciais básicas em recurso de imagem, como também para previsão correta de novos estados afetivos básicos da face que são similares aos usados na fase de treinamento.

#### **4.5. Comparação dos resultados para classificação dos estados afetivos da face com imagens**

Os pesquisadores *L. S. Videla e P. M. A. Kumar* (2020) utilizam uma CNN de 10 camadas para detectar características importantes da expressão facial sem supervisão humana. Esta pesquisa usou o conjunto de dados *Extended Cohn-Kanade* (CK+) e *Japanese Female Facial Expression* (JAFPE) obtendo as performances de acurácia para cada base de dados de 99,3% e 78,1%.

O trabalho de *G. Shengtao, X. Chao and F. Bo* (2019) investiga as CNN *AlexNet*, *VGGNet* e *ResNet* com uso do banco de dados FER2013, visando o reconhecimento de expressões faciais de forma paralela,

onde as CNN obtiveram o desempenho de 66,6%, 69,4% e 70,7%.

A pesquisa de *S. Shaees, H. Naeem, M. Arslan, M. R. Naeem, S. H. Ali and H. Aldabbas* (2020) compara a performance da rede *AlexNet* com a abordagem de *Support Vector Machine* (SVM) para reconhecimento de expressões faciais usando a base de dados *Cohn-Kanadec CK + Expression*. Os resultados mostram claramente que a CNN produz resultados superiores em faixa de 98,3% até 99,9%.

Considerando os resultados produzidos nesta pesquisa com as redes *AlexNet*, *GoogLeNet* e *ResNet18* comparado a performance obtida pelos outros pesquisadores, constata-se que as redes *AlexNet*, *GoogLeNet* e *ResNet18* produzem respostas robustas e compatíveis para o reconhecimento de expressões faciais em recursos de imagem.

#### **5. Conclusão**

Este artigo investigou as redes *AlexNet*, *GoogLeNet* e *ResNet18* aplicadas ao reconhecimento de expressões faciais em recursos de imagem considerando as etapas de classificação e capacidade de classificação para novas imagens que são similares as usadas na fase de treinamento. Os resultados de acurácia de classificação de 99%, 98,1% e 99,3% e teste de validação cruzada de 91,7%, 95,5% e 88,5% com as redes *AlexNet*, *GoogLeNet* e *ResNet18* para uma abordagem multiclases confirmam que essas redes são uma solução eficaz para o desenvolvimento de soluções inteligentes no campo da visão computacional com reconhecimento de expressões faciais.

Para trabalhos futuros propõem-se um estudo com novas redes convolucionais, afim de investigar a performance de classificação e generalização no processo de classificação dos estados afetivos da face. Também se sugere o reconhecimento de expressões faciais em tempo real com

uso de uma webcam.

## Referências

AJIT, Arohan; ACHARYA, Koustav; SAMANTA, Abhishek. **A Review of Convolutional Neural Networks**. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, Vellore, India: IEEE, 2020. p. 1-5. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9077735>>. Acesso em: 06 abr. 2021.

ARIZA-LOPEZ, F.J.; RODRIGUEZ-AVI, J.; ALBA-FERNANDEZ, M.V. **Complete Control of an Observed Confusion Matrix**. In: IGARSS 2018 - 2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2018, Valencia, Spain: IEEE, 2018. p. 1222-1225. Disponível em: <<https://ieeexplore.ieee.org/document/8517540>>. Acesso em: 11 jul. 2021.

BASBRAIN, Arwa Mohammed et al. **A Neural Network Approach to Score Fusion for Emotion Recognition**. In: 2018 10th Computer Science and Electronic Engineering (CEECE), 10, 2018, Colchester, UK: IEEE, 2018. p. 180-185. Disponível em: <<https://ieeexplore.ieee.org/document/8674191>>. Acesso em: 11 jun. 2021.

CRUZ, Anderson Araújo da. **Uma abordagem para reconhecimento de emoção por expressão facial baseada em redes neurais de convolução**. 2019. 120 f. Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas, Manaus, 2019.

DONG, Yu-Nan; LIANG, Guang-Sheng. **Research and Discussion on Image Recognition and Classification Algorithm Based on Deep Learning**. In: 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), 2019, Taiyuan, China: IEEE, 2019. p. 274-278. Disponível em: <<https://ieeexplore.ieee.org/document/8945665>>. Acesso em: 21 maio 2021.

KHORRAMI, Pooya; PAINE, Tom Le; HUANG, Thomas S. **Do Deep Neural Networks Learn Facial Action Units When Doing Expression Recognition?** In: 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, Santiago, Chile: IEEE, 2015. p. 19-27. Disponível em: <<https://ieeexplore.ieee.org/document/7406361>>. Acesso em: 15 jun. 2021.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON,

Geoffrey E. **Imagenet classification with deep convolutional neural networks**. *Advances in neural information processing systems*, v. 25, p. 1097-1105, 2012.

LI, Qinfeng. **An Improved Face Detection Method Based on Face Recognition Application**. In: 2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), 2019, Nagoya, Japan: IEEE, 2019. p. 260-264. Disponível em: <<https://ieeexplore.ieee.org/document/8936020>>. Acesso em: 04 abr. 2021.

LIU, Ping et al. **Facial Expression Recognition via a Boosted Deep Belief Network**. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, Columbus, Oh, USA: IEEE, 2014. p. 1805-1812. Disponível em: <<https://ieeexplore.ieee.org/document/6909629/authors>>. Acesso em: 30 maio 2021.

LIVINGSTONE, Steven R.; RUSSO, Frank A.. **The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north american english**. *Plos One*, [S.L.], v. 13, n. 5, p. 1-35, 16 maio 2018. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0196391>. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391>>. Acesso em: 05 maio 2021.

MATHWORKS. **Cross-Validation**. Disponível em: <<https://www.mathworks.com/discovery/cross-validation.html>>. Acesso em: 23 jul. 2021.

PANTIC, Maja; PATRAS, Ioannis. **Dynamics of Facial Expression: recognition of facial actions and their temporal segments from face profile image sequences**. *IEEE Transactions on Systems, Man, And Cybernetics, Part B (Cybernetics)*. [S.L.], p. 433-449. 13 mar. 2006. Disponível em: <<https://ieeexplore.ieee.org/document/1605389>>. Acesso em: 04 abr. 2021.

PREMA, R.; SHANMUGAPRIYA, P. **A review: face recognition techniques for differentiate similar faces and twin faces**. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, Chennai, India: IEEE, 2017. p. 2899-2902. Disponível em: <<https://ieeexplore.ieee.org/document/8389985/authors>>. Acesso em: 01 abr. 2021.

RESHMA, M R; KANNAN, B. **Approaches on**

**Partial Face Recognition: a literature review.** In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, Tirunelveli, India: IEEE, 2019. p. 538-544. Disponível em: <<https://ieeexplore.ieee.org/document/8862783/metrics#metrics>>. Acesso em: 04 abr. 2021.

SHAEES, Shamoil et al. **Facial Emotion Recognition Using Transfer Learning.** In: 2020 International Conference on Computing and Information Technology (ICCIT-1441), 2020, Tabuk, Saudi Arabia. IEEE, 2020. Disponível em: <<https://ieeexplore.ieee.org/document/9213757/authors#authors>>. Acesso em: 15 ago. 2021

SHENGTAO, Gu; CHAO, Xu; BO, Feng. **Facial expression recognition based on global and local feature fusion with CNNs.** In: 2019 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), 2019, Dalian, China. IEEE, 2019. Disponível em: <<https://ieeexplore.ieee.org/document/8960765>>. Acesso em: 15 ago. 2021.

SZE, Vivienne et al. **Efficient Processing of Deep Neural Networks: a tutorial and survey.** *Proceedings Of The IEEE. [S.I.]*, p. 2295-2329. 20 nov. 2017. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8114708/authors#authors>>. Acesso em: 06 abr. 2021.

VIDELA, Lakshmi Sarvani; KUMAR, P.M. Ashok. **Facial Expression Classification Using Vanilla Convolution Neural Network.** In: 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, Chennai, India. IEEE, 2020. Disponível em: <<https://ieeexplore.ieee.org/document/9202053>>. Acesso em: 15 ago. 2021.

WANG, Huai; HAN, Zhuo. **An Improved AlexNet Model with Multi-channel Input Images Processing for Human Face Feature Points Detection.** In: 2020 12th International Conference on Communication Software and Networks (ICCSN), 2020, Chongqing, China: IEEE, 2020. p. 246-251. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9139075>>. Acesso em: 17 jul. 2021.

WANG, Zijie J. et al. **CNN EXPLAINER: learning convolutional neural networks with interactive visualization.** *IEEE Transactions on Visualization and Computer Graphics. [S.I.]*, p. 1396-1406. 13 out. 2020. Disponível em: <<https://ieeexplore.ieee.org/document/9222325/authors#authors>>. Acesso em: 05 abr. 2021.