

# Rede Neural Artificial Convolutacional Aplicada ao Reconhecimento de Configuração de Mão nos Símbolos de 0 a 9 da Língua Brasileira de Sinais (LIBRAS)

Adriel Vieira Santos<sup>1</sup>, Iago Franco Bacurau<sup>1</sup>, Jayne de Moraes Silva<sup>1</sup>,  
Talles Brito Viana<sup>1</sup>, Robson Gonçalves Fechine Feitosa<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Ceará  
Rodovia CE 292, KM 15, Gisélia Pinheiro – CEP 63115-500 - Crato – Ceará – Brasil

adriel.klt@gmail.com, {tallesbrito, robsonfeitosa}@ifce.edu.br

**Abstract.** *Brazilian Law number 10.436, of April 24th, 2002, recognizes the Brazilian Sign Language (LIBRAS) as a legal way for communication and expression. However, the non-listening community, which makes use of this language, has difficulties to communicate with the listening community. In order to decrease such a difficulty, in this work is described an application of a convolutional neural network for the LIBRAS static symbols recognition. To validate the model, a database with about 2640 images was used, including symbols from 0 to 9 for training; and, 1360 symbols for testing. In addition to an extra set of 1000 symbols for validating more test cases. As a result, a recognition rate ranging from 82.5 % to 98.57 % for different symbols was obtained.*

**Resumo.** *A Lei brasileira número 10.436, de 24 de abril de 2002, reconhece a Língua Brasileira de Sinais (LIBRAS) como meio legal de comunicação e expressão. Entretanto, a comunidade não ouvinte, que faz uso dessa língua, possui muita dificuldade de comunicação com a comunidade ouvinte. Visando reduzir tal dificuldade, o presente trabalho descreve uma aplicação das redes neurais convolucionais para o reconhecimento dos símbolos estáticos da LIBRAS. Para validar o modelo, foi utilizada uma base de dados com cerca de 2640 imagens, com símbolos de 0 a 9 para treinamento; e, 1360 símbolos para teste. Além de um conjunto extra de 1000 símbolos para validação de mais casos de testes. Como resultados, obteve-se uma taxa de acerto que varia entre 82.5% e 98.57% para os diferentes símbolos.*

## 1. Introdução

De acordo com [de Freitas et al. 2017], o censo do IBGE (realizado em 2010) mostra que 45,6 milhões de brasileiros possuem algum tipo de deficiência auditiva e, devido a lei nº 10.436, a Língua Brasileira de Sinais (LIBRAS) foi reconhecida como uma língua oficial no Brasil. Dessa forma, o tema de acessibilidade e suas tecnologias é abordado em vários trabalhos, como: [Brandt 2015] e [Koroishi and Silva 2015]. Diante disso, o presente trabalho possui como objetivo demonstrar a viabilidade do uso das redes neurais convolucionais para o reconhecimento de alguns dos símbolos estáticos da LIBRAS (números de 0 a 9) capturados por vídeo.

As RNAs (Redes Neurais Artificiais) são modelos matemáticos inspirados nos neurônios biológicos. Elas são aplicadas na resolução de problemas que envolvem:

desempenho, poder de generalização, robustez e tolerância a falhas. Devido suas características inspiradas nos neurônios biológicos, sua aplicação na identificação de padrões em imagens foi um caminho natural, em que, por exemplo (de acordo com [Haykin 2007]), uma pessoa consegue identificar um rosto familiar em uma cena qualquer em cerca de 100 à 200 milissegundos.

Outro fator interessante para uso das redes neurais artificiais, como solução ao reconhecimento de padrões em imagens, é a aplicação em problemas em que não é possível descrever em detalhes um modelo matemático formal (algorítmico) como solução. Por conta dessa característica, [Barreto 2002] visualiza as RNAs como um paradigma para resolução de problemas.

Para fins de organização, o presente trabalho está dividido nas seguintes seções: a Seção 2 apresenta a fundamentação teórica que embasa o trabalho; na Seção 3 descreve-se a metodologia e materiais utilizados; e, por fim, a Seção 4 ilustra os resultados e conclusões encontradas.

## **2. Fundamentação Teórica**

### **2.1. Redes Neurais Artificiais**

Segundo [Haykin 2007], uma RNA é um processador maciçamente paralelamente distribuído consistindo de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso.

As redes neurais artificiais podem ser visualizadas em duas arquiteturas diferentes: redes alimentadas para frente e redes recorrentes. O primeiro tipo de rede pode ser visualizada como um grafo<sup>1</sup> acíclico dirigido e o segundo tipo como um grafo cíclico dirigido em que a saída de um neurônio pode ser utilizada para alimentar o próprio neurônio novamente ou toda a rede. A rede neural pode conter várias camadas. Quanto maior a quantidade de camadas, maior a capacidade de resolução de problemas da rede.

A rede neural Multilayer Perceptron (MLP) é um tipo de rede neural que deriva do Perceptron desenvolvido por Rosenblatt em 1958. De acordo com [Haykin 2007], este é o modelo mais simples de rede neural para tratar problemas linearmente separáveis como os operadores *OU* e *E* da lógica. Além disso, segundo o mesmo autor, Rosenblatt demonstra com sua pesquisa um teorema que garante a separação de problemas lineares por redes do tipo Perceptron simples. O treinamento das redes do tipo MLP utiliza o algoritmo Backpropagation. Segundo [Haykin 2007], este algoritmo é categorizado como aprendizagem por correção de erro.

Segundo [Vargas et al. 2016], uma Rede Neural Convolutiva (RNC) é uma variação das redes MLP, inspirada no processo biológico de processamento de dados visuais. A estrutura desse tipo de rede pode ser dividida em duas partes: camadas para extração de características e camadas de classificação.

De acordo com [Vargas et al. 2016], uma convolução é uma operação entre duas funções que produz uma terceira função. O objetivo da aplicação da convolução na rede é uma forma de integrar sobre o modelo a capacidade de extrair informações dos campos

---

<sup>1</sup>Um grafo é uma entidade formada por um conjunto de arestas e um conjunto de vértices, sendo as arestas uma relação binária sobre o conjunto de vértices.

perceptíveis contidos nas imagens, assim como ocorre na visão humana. Acredita-se que quando uma pessoa visualiza uma imagem, ela não a interpreta como um todo, mas sim em partes que a caracterizam.

Baseado em [Mansano 2017] e [Vargas et al. 2016], para entender o funcionamento da rede é necessário entender a etapa de convolução. Dado uma imagem binarizada, representada em uma matriz bidimensional, pode-se aplicar sob esta uma outra matriz que representa um padrão que se deseja extrair, também chamada de kernel. A aplicação da convolução entre essas duas funções produz uma nova matriz que fornece evidências do padrão contido na imagem. Após a convolução, aplica-se uma técnica de pooling (máximo, mínimo, média ou mediana) no conjunto de elementos contidos dentro da área de um kernel. A utilização dessa última técnica permite que as alterações ocorridas na imagem como rotação, escala e translação se tornem invariantes.

O algoritmo de aprendizagem Backpropagation é utilizado para treinar a rede. Por conta da grande quantidade de camadas que podem existir, é necessário utilizar nas camadas de extração de características a função de ativação ReLU (unidade linear retificada), que evita o problema do desaparecimento do gradiente. Diferente das funções sigmoide e tangente hiperbólica, como mostrado em [Academy 2018].

### 3. Metodologia

Para a construção do modelo proposto neste trabalho, foi necessária a utilização do WEKA. Segundo [Frank and Mark 2016], o WEKA é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Esta contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização. A escolha desta ferramenta foi motivada por conter a implementação do modelo apresentado, ser *open-source* e dispor de métodos de avaliação dos classificadores e sua visualização.

Tendo como objetivo testar o modelo, foi necessário produzir uma base de dados contendo 4000 imagens com tamanhos 28x28 possuindo os símbolos de 0 a 9 da língua brasileira de sinais (LIBRAS). As imagens foram obtidas de uma câmera VGA integrada em um notebook Positivo, em um ambiente com fundo branco. A imagem foi binarizada usando a classe BufferedImage (biblioteca disponível na API da linguagem de programação Java)<sup>2</sup>.

A rede neural convolucional treinada segue a arquitetura LeNet<sup>3</sup>. Tal arquitetura obteve sucesso no reconhecimento de dígitos manuscritos, conforme apresentado em [LeCun et al. 1995]. Outro motivo que embasou a escolha da arquitetura LeNet foi pelo bom desempenho, mesmo em máquinas de baixa capacidade de processamento, como a que foi utilizada para o treinamento do presente trabalho. Os parâmetros usados na rede foram:

- Taxa de aprendizado: 0.001; Tamanho do lote: 120; Função de ativação: ReLU; Número de épocas: 10; Camadas de convolução: 02, sendo: Camada 01 com 20 neurônios e Camada 02 com 50 neurônios; Camada oculta (MLP interna) com 500 neurônios; Camada de saída com 10 neurônios; e, Número de pooling: 02.

---

<sup>2</sup>Disponível em: <https://docs.oracle.com/javase/7/docs/api/java/awt/image/BufferedImage.html>.

<sup>3</sup>Disponível em: <http://yann.lecun.com/exdb/lenet/>.

#### 4. Resultados e Conclusões

A rede neural foi treinada com 2640 imagens, referentes a cerca de 66% da base de dados, sendo que as 1360 imagens restantes foram usadas para testes, cerca de 34%. O classificador conseguiu obter uma taxa de acerto de 98.57% (1343 imagens). Isso demonstra que os resultados foram muito positivos para o conjunto usado.

Devido à baixa capacidade da máquina disponível, não foi usada a técnica de validação cruzada, descrita em [Haykin 2007], durante a avaliação do modelo. Assim, um importante trabalho futuro seria realizar uma análise mais apurada dos resultados, com o intuito de verificar se há indícios de super adaptação. Para contornar este problema, uma nova base de dados foi criada por uma única pessoa. A base foi formada por 1000 imagens e a LeNet ainda classificou corretamente cerca de 82.5% das instâncias de forma correta.

Embora os resultados alcançados sejam relevantes, é importante notar que a produção de uma base de dados mais diversificada com uma alta quantidade de imagens pode, possivelmente, melhorar os resultados. Como a máquina disponível para o treinamento não possuía alta capacidade de processamento, não foi possível avaliar diferentes parâmetros e algoritmos com intuito de justificar a estrutura produzida.

#### Referências

- Academy, D. S. (2018). *Deep Learning Book*, volume 1. Disponível em: [deeplearning-book.com.br](http://deeplearning-book.com.br). Acesso em: 30 nov. 2018.
- Barreto, J. M. (2002). Introdução às redes neurais artificiais. *Florianópolis: UFSC*.
- Brandt, A. H. (2015). Lirann: sistema de reconhecimento de libras baseado em redes neurais artificiais com kinect. *Disponível em: <http://hdl.handle.net/11624/1031>, acesso. Acesso em: 02 abr. 2019.*
- de Freitas, G. R., Maranhão, T. L. G., and Félix, W. M. (2017). A deficiência auditiva e a possibilidade de inclusão no mercado de trabalho. *ID ON LINE REVISTA MULTIDISCIPLINAR E DE PSICOLOGIA*, 11(34):184–218.
- Frank, E. and Mark, A. (2016). Hall, and ian h. witten (2016). the weka workbench. online appendix for "data mining: Practical machine learning tools and techniques. *Morgan Kaufmann, Fourth Edition.*, 78(71):74–5.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. Bookman Editora.
- Koroishi, G. O. and Silva, B. V. L. (2015). Reconhecimento de sinais da libras por visão computacional. *Mecatrone*, 1(1).
- LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Muller, U., Sackinger, E., et al. (1995). Comparison of learning algorithms for handwritten digit recognition. In *International conference on artificial neural networks*, volume 60, pages 53–60. Perth, Australia.
- Mansano, A. F. (2017). O que é uma rede neural convolucional? *Disponível em: <https://www.linkedin.com/pulse/o-que-%C3%A9-um-rede-neural-convolucional-alex-fernandes-mansano>, acesso. Acesso em: 30 nov. 2018.*
- Vargas, A. C. G., Paes, A., and Vasconcelos, C. N. (2016). Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In *Proceedings of the XXIX Conference on Graphics, Patterns and Images*, pages 1–4.