

IMPACTO DO BALANCEAMENTO DE DADOS NO DESEMPENHO DE MODELOS DE RECONHECIMENTO DA LÍNGUA DE SINAIS

Letícia M. G. de Moraes¹, Náthalee Cavalcanti A. Lima², Rosana C. B. Rego²

¹Universidade Federal Rural do Semi-Árido (UFERSA), Rio Grande do Norte, Brasil (le-ticia.moraes30463@alunos.ufersa.edu.br)

²Departamento de Engenharias e Tecnologia, Universidade Federal Rural do Semi-Árido (UFERSA), Rio Grande do Norte, Brasil

Resumo: Este trabalho propõe uma análise sobre o uso de técnicas de balanceamento de dados para séries temporais da Língua Brasileira de Sinais (Libras), visando facilitar a junção de diferentes conjuntos de dados para o treinamento de um modelo de rede neural recorrente. Foram testadas as técnicas SMOTE, ROS, RUS e atribuição de pesos. O modelo treinado sob a técnica ROS foi o que obteve as melhores métricas de avaliação, com uma acurácia de 91,05%, F1-score de 91,11%, precisão de 91,56% e Recall 91,05%.

Palavras-chave: Redes neurais profundas; Visão Computacional; Balanceamento de dados; Língua de Sinais.

INTRODUÇÃO

Para a comunidade surda brasileira, a Língua Brasileira de Sinais (Libras) é utilizada como principal forma de comunicação, sendo uma das primeiras línguas utilizadas pelas pessoas com surdez ou baixa audição. Apesar disso, há uma baixa difusão do uso da Libras entre as pessoas ouvintes, acarretando em barreiras na comunicação (Junior et al., 2024; Moraes et al, 2025a).

Uma das formas de facilitar a comunicação dos ouvintes e falantes de Libras, é a detecção automática de sinais da Libras para o Português através de visão computacional e algoritmos de inteligência artificial (IA) (Moraes et al, 2025b).

Um dos principais desafios, entretanto, está na ausência de um conjunto de dados público e completo que contemple todos os sinais da Libras, o que dificulta o treinamento de algoritmos de IA (Almeida et al, 2022).

Ainda assim, existem conjuntos de dados de Libras, que, mesmo com quantidades reduzidas de sinais, permitem a criação de um modelo capaz de reconhecer e traduzir os sinais para Português, como por exemplo, o Libras-10 (Almeida et al., 2019a) e o MINDS-Libras (Almeida et al., 2019b) utilizados neste trabalho.

Entretanto, o uso de dois conjuntos de dados diferentes pode gerar desbalanceamentos na quantidade de amostras de cada sinal, o que ocorreu ao juntar o Libras-10 e o MINDS-Libras. O desbalanceamento de classes é um ponto crítico no treinamento de algo-

ritmos de IA, uma vez que pode acarretar no sobreajuste do modelo (onde este tem um viés na classe majoritária) (Géron, 2021; Altalhan et al., 2025; Ying et al., 2019; Kulkarni et al., 2020).

Para lidar com dados desbalanceados no treinamento de algoritmos de IA, são utilizadas técnicas que envolvem desde a adição ou remoção de dados no conjunto até atribuição de pesos na função perda (Abhishek e Abdelaziz, 2023; Salman e Liu., 2019).

Para este estudo, foram analisadas as técnicas de nível de dados: SMOTE (*Synthetic Minority Over-sampling Technique*) proposta por Chawla et al. (2002), ROS (*Random Over Sampler*) e RUS (*Random Under Sampler*). Também foi testada a técnica de nível algorítmico do ajuste de pesos das classes na função perda.

Dessa forma, esse trabalho teve como principal objetivo analisar e avaliar a melhor abordagem de balanceamento de dados para séries temporais de sinais dinâmicos da Libras. Além de analisar o impacto que cada tipo de abordagem para resolver o desbalanceamento de dados faz sobre o desempenho de modelos de reconhecimento da língua de sinais.

MATERIAL E MÉTODOS

Conjunto de Dados

Para este trabalho, foram utilizados dois conjuntos de dados: Libras-10 e MINDS-Libras. Ambos compostos por vídeos de sinais da Língua Brasileira de Sinais. O primeiro conjunto é composto por 10 sinais,

cada um registrado 10 vezes por uma única sinalizadora, totalizando 100 vídeos. O segundo conjunto, por sua vez, contém registros de 12 sinalizadores, cada um repetindo 5 vezes um conjunto de 20 sinais, totalizando 1.200 vídeos.

Extração de atributos

Cada vídeo das bases de dados passou algoritmo de extração utilizando a ferramenta MediaPipe (Lugaresi, 2019). Esse algoritmo realiza a coleta dados das mãos e do corpo utilizando, respectivamente, os módulos de detecção *Hands* e *Pose*.

Ambos os módulos são formados por *pipelines* com funcionamento semelhante. Cada um é composto por dois modelos de inteligência artificial: o primeiro identifica o objeto de interesse no quadro e define a região de interesse (ROI), que é então enviada ao segundo modelo, responsável por extrair coordenadas tridimensionais de pontos de referência específicos (Zhang et al., 2020; Bazarevsky et al., 2020).

O módulo *Pose* possui 33 pontos de referência (ver Figura 1), enquanto que o *Hands* possui 21 pontos de referência (ver Figura 2). Uma característica particular do módulo *Hands*, é que o modelo de detecção também classifica a lateralidade de uma mão binariamente (esquerda ou direita).

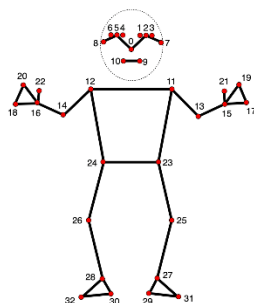


Figura 1: Pontos de referência extraídos do módulo *Pose* (Bazarevsky et al., 2020).

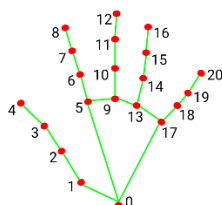


Figura 2: Pontos de referência extraídos do módulo *Hands* (Zhang et al., 2020).

Dessa forma, o algoritmo de extração foi aplicado a cada frame dos vídeos dos conjuntos de dados, coletando as coordenadas tridimensionais de todos os pontos de referência das mãos dos sinalizadores, bem como dos 17 primeiros pontos de referência do corpo. O não uso dos pontos restantes deve-se ao fato de que, na Libras, o ponto de articulação, isto é, a região

onde os sinais são realizados, concentra-se na parte superior do corpo, não sendo necessária a extração de dados dos membros inferiores.

Visando também contemplar sinalizadores canhotos, os vídeos foram espelhados para que o modelo generalize melhor. Com isso, o algoritmo de extração processa o vídeo uma vez no formato original e outra no formato espelhado.

Assim, foi gerado um conjunto de dados formados de aproximadamente 2600 séries temporais, das quais 80% foram separadas para o treinamento do modelo de aprendizado de máquina, como mostra a Figura 3, enquanto que as 20% restantes foram deixadas para o teste dele (ver Figura 4).

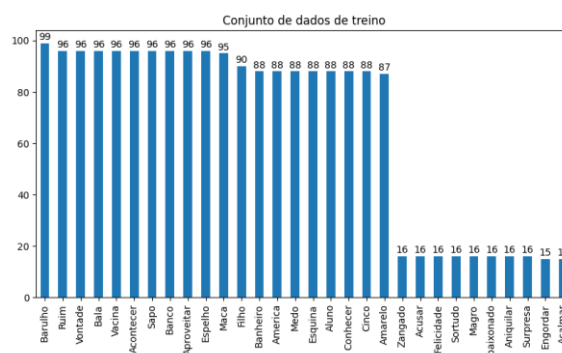


Figura 3: Total de amostras por classe no conjunto de dados de treino.

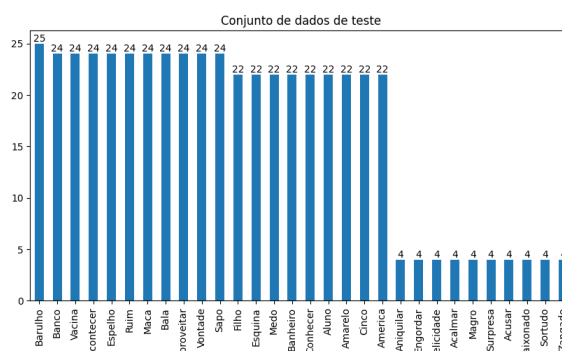


Figura 4: Total de amostras por classe no conjunto de dados de teste.

Balanceamento de dados

Como foram utilizadas duas bases de dados diferentes para gerar o conjunto de dados das séries temporais, existe um desbalanceamento entre a quantidade de amostras que também é perpetuada no conjunto de dados de treino.

Conjunto de dados desbalanceados podem contribuir para o sobreajuste do modelo, uma vez que este tende a aprender padrões das classes majoritárias,

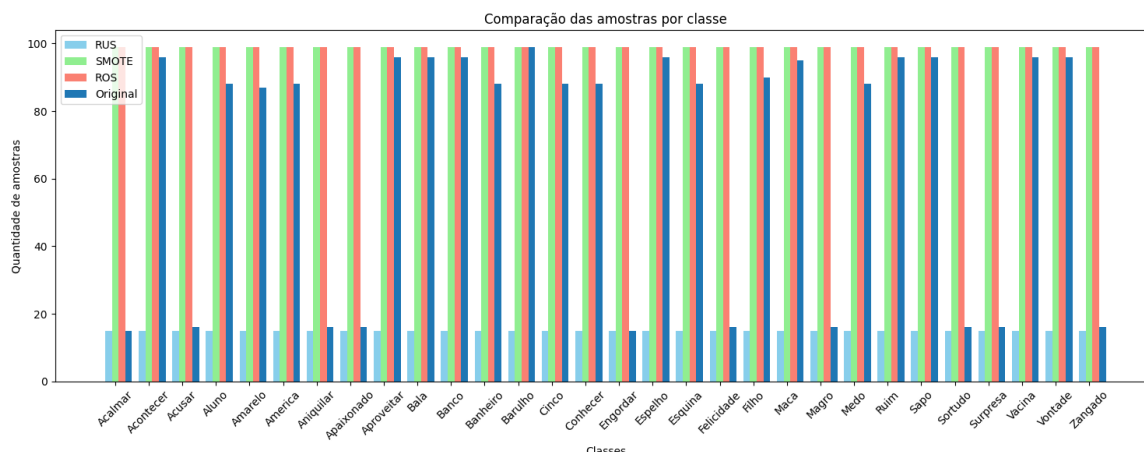


Figura 5: Comparação da quantidade de amostras por classe em cada abordagem.

resultando em um bom desempenho nos dados de treino, mas uma baixa generalização para dados novos (Ying et al., 2019; Kulkarni et al., 2020).

Com o objetivo de encontrar a melhor técnica de balanceamento de séries temporais para sinais da Libras, o conjunto de dados de treino foi testado em 4 abordagens diferentes.

Synthetic Minority Oversampling Technique (SMOTE) (Chawla, 2002): Nesta técnica, são criadas amostras sintéticas das classes minoritárias, aumentando a quantidade até o número de amostras da classe majoritária.

Random Over Sampler (ROS): Esta técnica também consiste no aumento das amostras das classes minoritárias. Entretanto, diferentemente da SMOTE, os dados gerados pelo ROS não são sintéticos, mas sim advindos da duplicação das amostras existentes.

Random Under Sampler (RUS): Diferentemente das técnicas anteriores, o RUS resolve o desbalanceamento do conjunto de dados ao transformá-lo em um conjunto de dados artificialmente balanceado. Ou seja, ao invés de aumentar os dados do conjunto, o RUS realiza a diminuição das amostras das classes majoritárias. Além disso, o fator de diminuição utilizado é adicionado como peso na função perda do modelo para classes que tiveram amostras removidas.

Atribuição de pesos na função perda: Nesta abordagem, a quantidade de amostras não aumenta ou diminui. Ao invés disso, esta técnica modifica a função perda do modelo de forma a atribuir pesos maiores para as classes minoritárias do que para as majoritárias (Abhishek e Abdelaziz, 2023). Com isso, a quantidade de amostras utilizadas em cada classe é exatamente igual ao conjunto de dados de treino original.

A Figura 5 apresenta os resultados das quantidades de amostras em cada abordagem, percebe-se que as

abordagens de aumento de dados (SMOTE e ROS) resultaram na mesma quantidade de dados, enquanto que a técnica RUS resultou em uma menor quantidade de dados.

Vale ressaltar que, mesmo que o resultado das técnicas SMOTE e ROS aparente ser o mesmo, a diferença se encontra nos tipos de dados aumentados. Enquanto que no SMOTE são gerados dados sintéticos (novos), na técnica ROS os dados existentes são duplicados.

Modelo de aprendizado profundo

Os quatro conjuntos de treino resultantes das abordagens de balanceamento foram testados em um modelo *Long Short-Term Memory* (LSTM). Este tipo de modelo de rede neural é utilizado para reconhecer dados sequenciais, que é o caso das séries temporais do conjunto de dados gerado pelo algoritmo de extração (Géron, 2021).

Para este trabalho, foi criado um modelo LSTM classificador. Ou seja, ele recebe como entrada uma série temporal de coordenadas tridimensionais e tem como saída o sinal predito para aquela sequência. Dessa forma, cada técnica de balanceamento gerou um modelo LSTM diferente.

A Figura 6 exibe o processo para o treinamento de cada modelo. Inicialmente, os dados de treino são utilizados como base pela técnica de balanceamento. Depois disso, o conjunto de dados balanceado é utilizado como entrada para o treinamento do modelo LSTM. Por fim, o mesmo conjunto de dados de teste é utilizado em todos os processos para avaliar as métricas de desempenho de cada modelo.

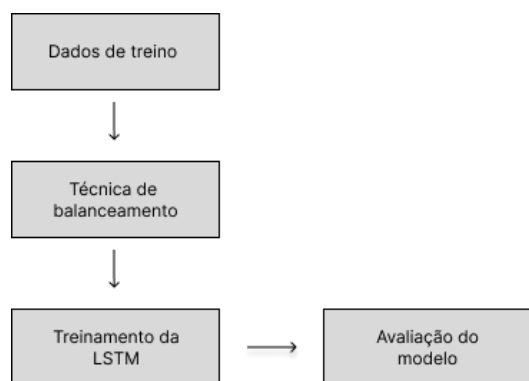


Figura 6: Processo de treinamento de cada modelo.

RESULTADOS E DISCUSSÃO

Como resultado dos treinamentos em cada abordagem, tem-se 4 modelos diferentes: o treinado ajustando os pesos das classes, o com o SMOTE, com o ROS e, por fim, com o RUS.

Os modelos foram testados sob o conjunto de dados de teste resultante da divisão do conjunto total. Para avaliá-los, foram consideradas quatro métricas: acurácia, *F1-score*, *Recall* e precisão. Para todas, valores maiores significam uma melhor generalização, exatidão e consistência do modelo.

A Tabela 1 apresenta os resultados do modelo LSTM para o conjunto de dados de teste em cada abordagem de balanceamento diferente.

Tabela 1. Métricas de avaliação do modelo LSTM

	Acurácia	F1-Score	Recall	Precisão
ROS	91,05%	91,11%	91,05%	91,56%
SMOTE	89,86%	89,84%	89,05%	90,31%
Pesos	82,10%	82,42%	82,10%	83,81%
RUS	19,88%	12,58%	19,88%	11,09%

Dentre os testes, o modelo treinado com o conjunto gerado pelo ROS foi o que obteve as melhores métricas de avaliação. O modelo treinado com o conjunto de dados sintéticos (técnica SMOTE) apresentou resultados próximos, com uma precisão maior que 90%, indicando uma consistência próxima ao do ROS. O modelo RUS foi o que obteve o menor desempenho, o que indica uma baixa confiabilidade em suas predições.

Além das métricas de avaliação, foram levadas em consideração os valores da função erro do modelo durante o treinamento. Que permitem analisar se durante o treino houve um sobreajuste do modelo sobre o conjunto de dados de treinamento. As Figuras

7 e 8 exibem a função perda dos dois melhores modelos.

Os gráficos de perda dos modelos mostram que, ao longo do treinamento, os erros de treino e de validação diminuíram, convergindo para valores próximos. Além disso, as métricas de desempenho também apresentaram resultados semelhantes entre si. Esses fatores indicam a ausência de sobreajuste, uma vez que os modelos demonstram boa capacidade de generalização, conseguindo prever adequadamente os dados dos conjuntos de validação e teste.

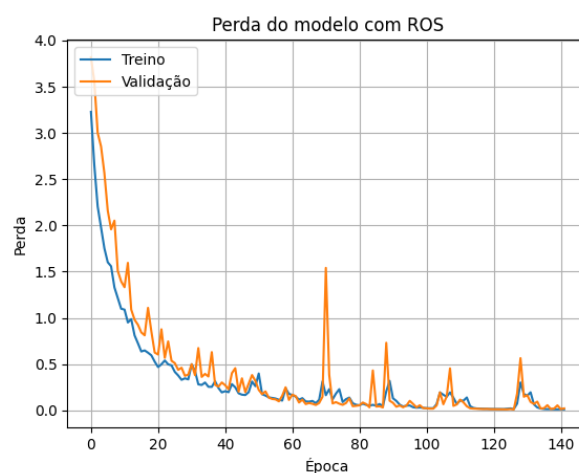


Figura 7: Comparação da quantidade de amostras por classe em cada abordagem.

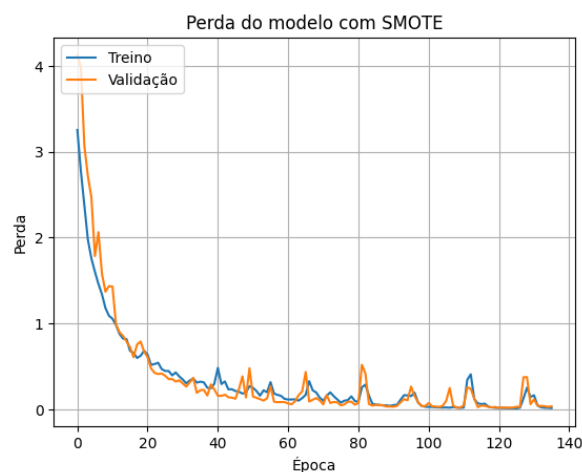


Figura 8: Comparação da quantidade de amostras por classe em cada abordagem.

A Figura 9 apresenta a matriz de confusão do melhor modelo. Maiores resultados na diagonal principal representam uma maior precisão do modelo para as classes correspondentes. Enquanto que resultados fora dela significam predições incorretas. A matriz de

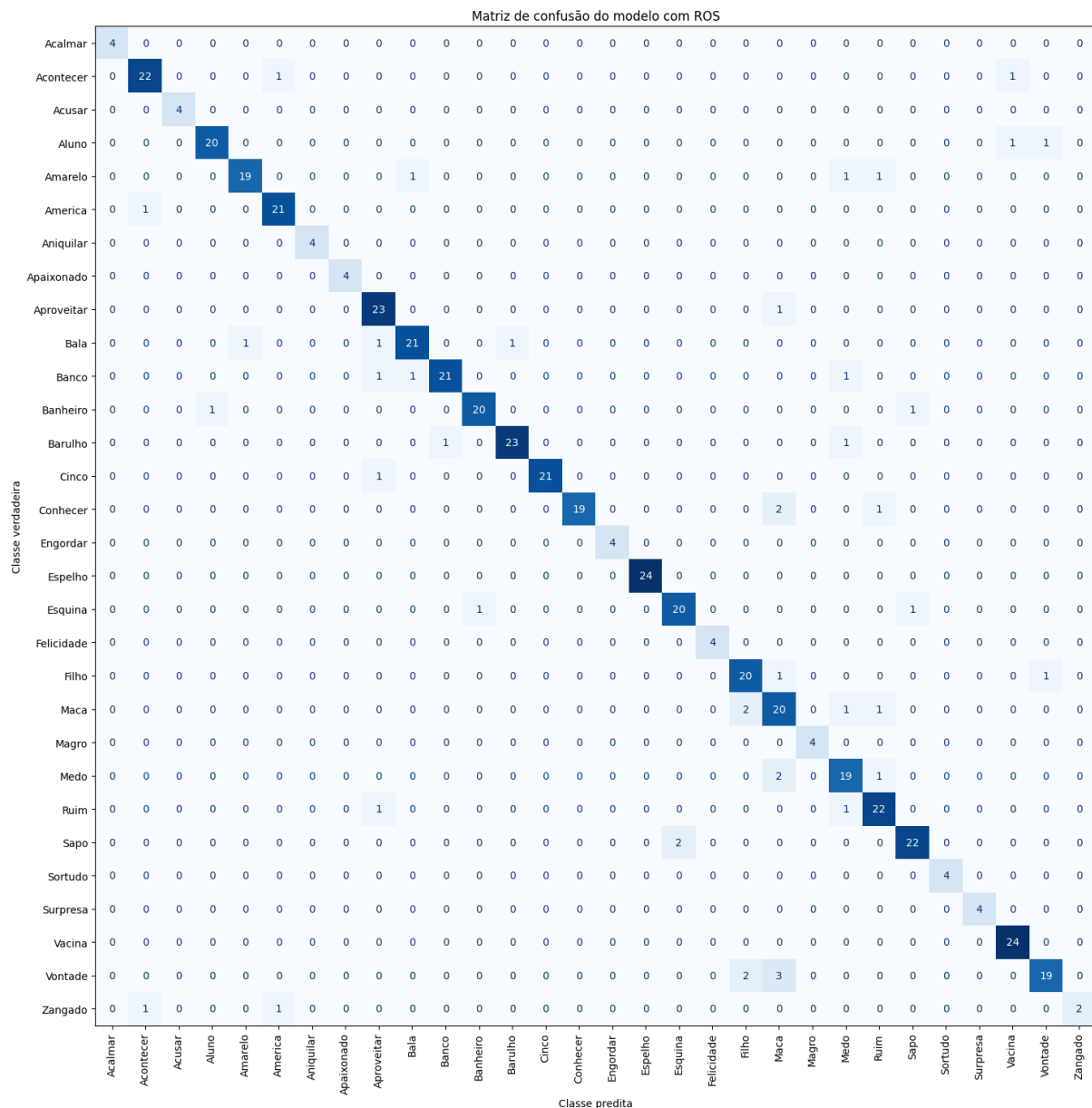


Figura 9: Matriz de confusão do melhor modelo

confusão do modelo ROS, portanto, apresenta um bom resultado, tendo em vista que a maior parte das predições feitas corresponderam com as classes reais.

CONCLUSÃO

Dentre as abordagens de balanceamento investigadas neste trabalho para a Libras, aquelas baseadas no aumento do conjunto de dados apresentaram melhor desempenho. O modelo treinado com a técnica ROS obteve as melhores métricas de avaliação, seguido pelo modelo com SMOTE, que alcançou resultados próximos, embora ligeiramente inferiores.

Por outro lado, os modelos treinados com abordagens de redução de amostras (RUS) e com ajuste de pesos das classes apresentaram desempenho inferior. Assim, entre as técnicas analisadas, a ROS se destaca como a abordagem mais eficaz para o balanceamento de dados em séries temporais de sinais dinâmicos da Libras.

AGRADECIMENTOS

Ao Laboratório de Inteligência Computacional (CI-Lab) pelo suporte à pesquisa e à Fundação de Amparo e Promoção da Ciência, Tecnologia e Inovação do Rio Grande do Norte (FAPERN) pelo apoio financeiro por meio da bolsa de Iniciação Científica.



REFERÊNCIAS

- ABHISHEK, K.; ABDELAZIZ, M. Machine Learning for Imbalanced Data: Tackle imbalanced datasets using machine learning and deep learning techniques. Packt Publishing, Birmingham, UK, 2023.
- ALMEIDA, Arthur; REZENDE, Tamires; ALMEIDA, Silvia. Ampliação da base MINDS-Libras: um estudo de aplicação de técnicas de aumento sintético de dados e da inclusão de novos conjuntos de vídeos disponíveis na literatura. Congresso Brasileiro de Automática (CBA), v. 3, 2022.
- ALMEIDA, Silvia Grasiella Moreira; REZENDE, Tamires Martins; TOFFOLO, Andreia Chagas Rocha; CASTRO, Cristiano Leite de. Libras-10 Dataset. Zenodo, 2019a.
- ALMEIDA, Sílvia G. M.; REZENDE, Tamires M.; ALMEIDA, Gabriela T. B.; TOFFOLO, Andreia C. R.; GUIMARÃES, Frederico G. MINDS-Libras Dataset. Zenodo, 2019b.
- ALTALHAN, Manahel; ALGARNI, Abdulmohsen; TURKI-HADJ ALOUANE, Monia. Imbalanced Data Problem in Machine Learning: A Review. IEEE Access, v. 13, p. 13686–13699, 2025.
- BAZAREVSKY, Valentin; GRISHCHENKO, Ivan; RAVEENDRAN, Karthik; ZHU, Tyler; ZHANG, Fan; GRUNDMANN, Matthias. BlazePose: On-device Real-time Body Pose Tracking. CoRR, v. abs/2006.10204, 2020.
- CHAWLA, Nitesh V.; BOWYER, Kevin W.; HALL, Lawrence O.; KEGELMEYER, W. Philip. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, v. 16, p. 321–357, 2002.
- GÉRON, Aurélien. Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras & TensorFlow: conceitos, ferramentas e técnicas para a construção de sistemas inteligentes. 2. ed. Alta Books, Rio de Janeiro, 2021.
- JUNIOR, Nestor Franca de Andrade; PINTO, Anderson; ALMEIDA, Welizangela; REGO, R. C. B. Brazilian Sign Language Translation: AI for the inclusion of deaf people. Anais do Congresso Brasileiro Interdisciplinar em Ciência e Tecnologia, Diamantina, 2024.
- KULKARNI, Ajay; CHONG, Deri; BATARSEH, Feras A. Foundations of data imbalance and solutions for a data democracy. Data Democracy. Academic Press, p. 83–106, 2020.
- LUGARESI, Camillo; TANG, Jiuqiang; NASH, Hadon; MCCLANAHAN, Chris; UBOWEJA, Esha; HAYS, Michael; ZHANG, Fan; CHANG, Chuo-Ling; YONG, Ming Guang; LEE, Juhyun; CHANG, Wan-Teh; HUA, Wei; GEORG, Manfred; GRUNDMANN, Matthias. MediaPipe: A Framework for Building Perception Pipelines. CoRR, v. abs/1906.08172, 2019.
- MORAIS, Letícia; ALMEIDA, Welizangela; REGO, Rosana. Machine learning approaches for efficient recognition of Brazilian Sign Language. Anais do XXI Simpósio Brasileiro de Sistemas de Informação, Recife, p. 49–56, 2025a.
- MORAIS, Letícia; COSTA, Ana Kelly N.; ALMEIDA, Welizangela Moreira de; LIMA, Nathalee Cavalcanti de Almeida; REGO, Rosana C. B. Análise da aplicação de visão computacional e aprendizado de máquina na tradução da LIBRAS para texto. Anais do XVII Congresso Brasileiro de Inteligência Computacional (CBIC 2025), Belo Horizonte, p. 1–8, 2025b.
- SALMAN, Shaek; LIU, Xiuwen. Overfitting Mechanism and Avoidance in Deep Neural Networks. CoRR, v. abs/1901.06566, 2019.
- YING, Xue. An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series, v. 1168, n. 2, p. 022022, 2019.
- ZHANG, Fan; BAZAREVSKY, Valentin; VAKUNOV, Andrey; TKACHENKA, Andrei; SUNG, George; CHANG, Chuo-Ling; GRUNDMANN, Matthias. MediaPipe Hands: On-device Real-time Hand Tracking. CoRR, v. abs/2006.10214, 2020.