

Classificação de Palavras da Língua Brasileira de Sinais Utilizando Métricas de Dissimilaridade

Maria Fernanda Gomes da Silva¹, Lailla Milainny Siqueira Bine

¹Ciência da Computação
Universidade Estadual do Paraná (UNESPAR)
Apucarana – Paraná

1. Introdução

A comunicação é fundamental para a interação humana, sendo as línguas de sinais, como a Língua Brasileira de Sinais (Libras), um meio essencial para a comunidade surda no Brasil. Embora a Libras seja reconhecida por lei [1], a barreira de comunicação persiste, já que a maioria dos ouvintes não a domina. Além disso, a comunicação escrita se mostra ineficaz para cerca de 80% da população surda no mundo, que é analfabeta ou semianalfabeta [2], reforçando a necessidade de soluções tecnológicas assistivas. Neste contexto, a Inteligência Artificial (IA) e a Visão Computacional (VC) surgem como ferramentas promissoras para o reconhecimento automático de sinais, visando a interpretação de gestos humanos a partir de dados visuais [3].

O reconhecimento de sinais apresenta um desafio significativo na área de Aprendizado de Máquina (ML): a alta variabilidade gestual e a consequente necessidade de extrair um alto volume de características dos vídeos (como configuração das mãos, velocidade, expressões faciais, etc.). Essa complexidade leva ao problema da “maldição da dimensionalidade”, onde a grande quantidade de características torna os dados esparsos e prejudica a eficácia dos classificadores tradicionais, podendo levar ao sobreajuste (*overfitting*) [4].

O objetivo principal deste trabalho é desenvolver e validar um modelo de reconhecimento de palavras da Libras utilizando uma abordagem baseada em dissimilaridade, transformando a representação dos gestos em um espaço de distâncias para permitir uma classificação robusta por meio de ML.

Os objetivos específicos são:

- Extrair características gestuais (espaciais e temporais) dos sinais de Libras dos vídeos, gerando vetores de características;
- Avaliar o desempenho do modelo na classificação de sinais conhecidos;
- Avaliar a capacidade do modelo de identificar e rejeitar sinais não conhecidos.

2. Trabalhos Relacionados

A pesquisa em reconhecimento automático de Libras é ativa, com a maioria dos trabalhos revisados empregando arquiteturas de Aprendizado Profundo (DL), como Redes Neurais Convolucionais (CNN) e Recorrentes (RNN), ou métodos de classificação supervisionada tradicionais[5].

Entre os trabalhos relacionados, se destacam o de Rezende [6] sendo fundamental, pois desenvolveu a base de dados MINDS-Libras (utilizada neste trabalho) e propôs modelos de referência usando arquiteturas de DL como Redes Neurais Convolucionais

3D e Temporais. Já nos trabalhos de Silva [7] e Santos e Quarto [8], exploraram o uso de Redes Neurais Artificiais (ANN) e CNNs para o reconhecimento de gestos.

Este trabalho se destaca por não focar no desenvolvimento de novas arquiteturas de DL, mas sim na aplicação da metodologia baseada em dissimilaridade. Observa-se que nenhum dos trabalhos revisados emprega essa abordagem para o reconhecimento de sinais em Libras.

Nos trabalhos que utilizaram a metodologia de dissimilaridade, o método se demonstrou eficaz em outros temas afetados pela alta dimensionalidade e variabilidade de dados, como o reconhecimento de drones [9], documentos manuscritos [10] e objetos genéricos [11]. A abordagem é vantajosa porque mapeia problemas de alta dimensão para um espaço de distâncias reduzido e estruturado, permitindo que o sistema reduza os impactos da “maldição da dimensionalidade” [4], faça a classificação e reconheça palavras já aprendidas e identifique e rejeite sinais que não pertencem ao vocabulário conhecido (Detecção de Novidade).

3. Desenvolvimento

O trabalho adota uma pesquisa de caráter exploratório e experimental com abordagem quantitativa, visando validar a eficácia da dissimilaridade no reconhecimento de Libras.

A base de dados utilizada foi a MINDS-Libras [6], composta por 20 palavras diferentes, gravadas por 8 sinalizadores, totalizando 800 amostras de vídeo. A base foi escolhida por sua relevância e pela variabilidade fonológica dos sinais. Para transformar os vídeos brutos em dados numéricos, utilizou-se a biblioteca *MediaPipe* [12]. Como ilustrado na Figura 1, o *MediaPipe* detecta 21 (*landmarks*) tridimensionais (x, y, z) para cada mão, resultando em 126 coordenadas por *frame*.

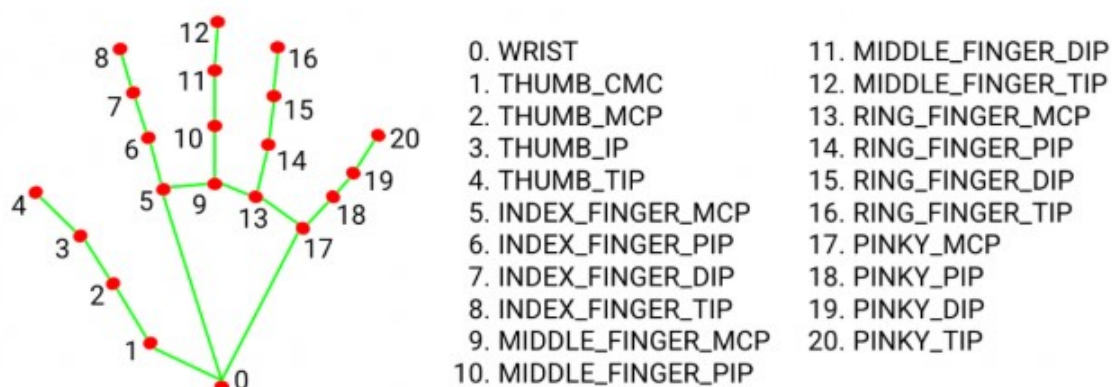


Figura 1. Mapa das Landmarks do MediaPipe.

Para garantir um vetor de tamanho fixo, foram calculadas quatro estatísticas para cada uma das 126 coordenadas ao longo de todos os *frames* do vídeo: Média, Desvio Padrão, Valor Máximo e Valor Mínimo. Cada vídeo foi transformado em um único vetor de 504 características ($126 \times 4 = 504$). Essa técnica captura a “pose média” (Média), a “dinâmica” (Desvio Padrão) e a “amplitude espacial” (Máximo e Mínimo) do gesto [13].

O vetor de 504 dimensões é então transformado no espaço de dissimilaridade, conforme as seguintes etapas:

1. Seleção de Protótipos (*K-Means*): O algoritmo *K-Means* [14] (não supervisionado) foi aplicado separadamente aos dados de treinamento de cada classe para selecionar amostras representativas, chamadas de “Palavras-Chave”. Foram selecionadas 2 protótipos para cada uma das 20 classes conhecidas, totalizando 40 Palavras-Chave.
2. Cálculo de Dissimilaridade [15]: Cada amostra de dados é transformada em um novo vetor onde cada posição representa a Distância Euclidiana da amostra para um dos 40 protótipos. O novo vetor de dissimilaridade passa a ter apenas 40 dimensões (uma para cada protótipo).

$$Z_{ij} = \sqrt{\sum_{k=1}^{504} (F_i - F_j)^2}$$

Onde F_i e F_j são os vetores de atributos originais.

3. Classificação (SVM) [16]: O classificador *Support Vector Machine* (SVM) foi treinado neste novo espaço de 40 dimensões. Utilizou-se o *GridSearch* com validação cruzada para otimizar os hiperparâmetros, resultando na melhor combinação: *kernel* 'rbf', *C*: 100 e *gamma*: 0.001.

Na etapa de configuração do experimento, o conjunto de dados (800 amostras) foi dividido em 70% para treino (560 amostras) e 30% para teste (240 amostras), utilizando o parâmetro *stratify* para garantir a proporção de classes.

O primeiro cenário, Classificação de Sinais Conhecidos, avalia o modelo em um cenário padrão onde ele é treinado e testado com todas as 20 classes do vocabulário. O objetivo é medir a acurácia geral do classificador em distinguir entre os sinais que ele aprendeu. Cenário 1: Classificação de Sinais Conhecidos. O modelo foi treinado e testado com todas as 20 classes para medir a acurácia geral e a capacidade de distinção.

No segundo cenário, ocorre a identificação de sinais não conhecidos. Utilizou-se o método *Leave-One-Class-Out* (LOCO), e foi repetido 20 vezes. Em cada iteração, uma classe era removida do treino (tornando-se a “novidade”) e usada para testar a capacidade de rejeição. A regra de detecção de novidade foi baseada em um Limiar de Dissimilaridade de 4.0. Se a distância mínima de uma amostra (desconhecida) para o protótipo conhecido mais próximo fosse maior que 4.0, ela era classificada como “Desconhecida”.

4. Resultados

No primeiro cenário, o modelo baseado em dissimilaridade demonstrou um desempenho robusto na classificação das 20 classes conhecidas, atingindo a acurácia geral de 71,67%. Este resultado indica que o modelo foi capaz de classificar corretamente cerca de 7 em cada 10 sinais. Em outras métricas de desempenho, os valores de Precisão (0,7285), *Recall* (0,7167) e *F1-Score* (0,7179) mostraram que o classificador opera de forma equilibrada, o que é crucial em um sistema de comunicação.

Como mostra a Figura 2 na matriz de confusão, sinais como ‘Acontecer’, ‘América’ e ‘Esquina’ apresentaram acurácia superior a 90%. Os principais erros se deram entre sinais com características visuais e dinâmicas semelhantes, como a confusão entre ‘Bala’ e ‘Banco’ (5/12 erros) e entre ‘Vontade’ e ‘Medo’ (4/12 erros).

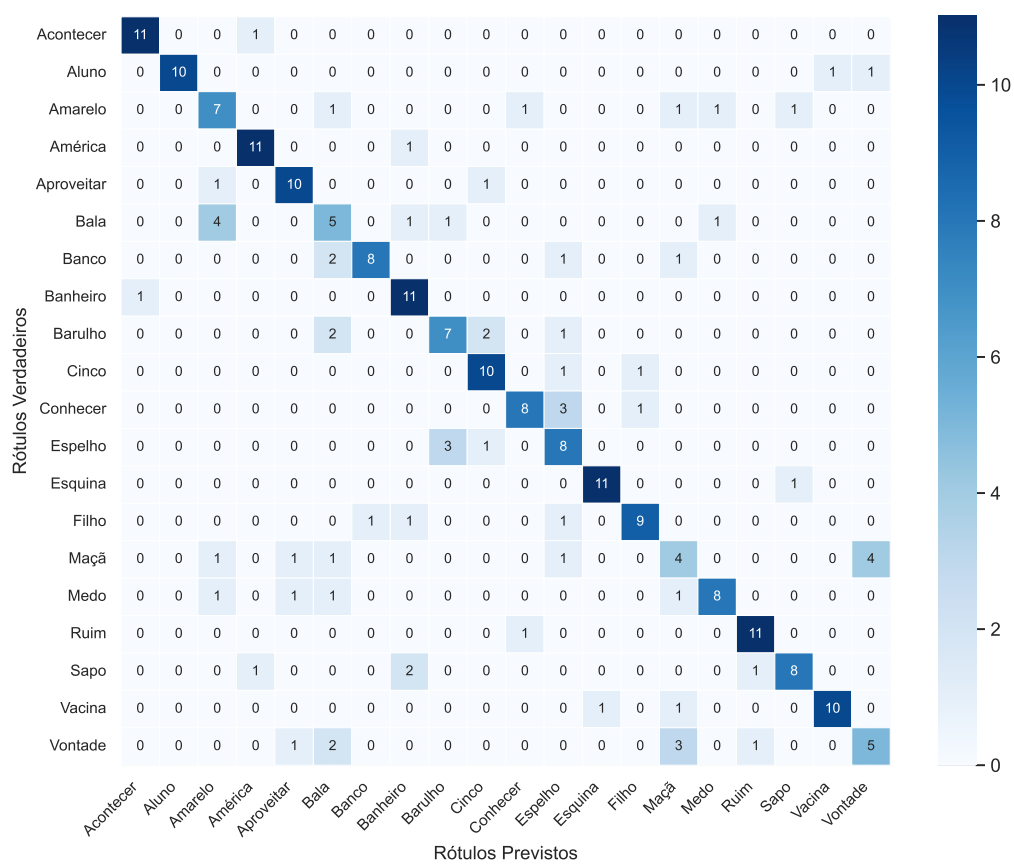


Figura 2. Matriz de confusão do modelo para os sinais conhecidos.

O segundo experimento, que visou a Detecção de Novidade, validou a principal contribuição metodológica do trabalho. Mostrando estabilidade na classificação conhecida, o modelo apresentou uma acurácia média de 73,14% para as classes conhecidas se manteve estável em todas as 20 iterações do método LOCO (variância entre 70,61% e 75,44%), conforme apresentados na Tabela 1. Isso demonstra a robustez do modelo, que não teve sua capacidade de aprendizado prejudicada pela remoção de uma classe. Além disso, teve uma taxa de detecção perfeita na identificação de sinais não conhecidos. A taxa de detecção de novidade de 100,00% em todos os cenários confirma a eficácia do método de dissimilaridade. Isso significa que nenhum sinal desconhecido foi incorretamente classificado como conhecido.

A distância mínima das amostras desconhecidas para os protótipos conhecidos sempre excedeu significativamente o limiar de dissimilaridade, validando a abordagem como um método altamente confiável para rejeitar sinais fora do vocabulário treinado.

Esses resultados significam que o sistema aprendeu não apenas a reconhecer padrões específicos, mas também a identificar quando uma amostra não pertence ao vocabulário previamente aprendido, uma habilidade essencial para aplicações reais de reconhecimento de Libras. Essa propriedade é particularmente relevante em contextos dinâmicos, onde novos sinais podem surgir, pois garante que o sistema opere de forma segura e confiável.

Tabela 1. Resultados da Identificação de Sinais Não Conhecidos.

| Sinal “Desconhecido” | Acurácia (Conhecidos) | Deteção (Novidade) |
|-----------------------------|------------------------------|---------------------------|
| Acontecer (palavra1) | 73,68% | 100,00% |
| Aluno (palavra2) | 72,81% | 100,00% |
| Amarelo (palavra3) | 71,93% | 100,00% |
| América (palavra4) | 71,93% | 100,00% |
| Aproveitar (palavra5) | 70,61% | 100,00% |
| Bala (palavra6) | 73,68% | 100,00% |
| Banco (palavra7) | 70,61% | 100,00% |
| Banheiro (palavra8) | 72,37% | 100,00% |
| Barulho (palavra9) | 72,37% | 100,00% |
| Cinco (palavra10) | 72,81% | 100,00% |
| Conhecer (palavra11) | 72,37% | 100,00% |
| Espelho (palavra12) | 72,81% | 100,00% |
| Esquina (palavra13) | 73,68% | 100,00% |
| Filho (palavra14) | 74,12% | 100,00% |
| Maça (palavra15) | 75,00% | 100,00% |
| Medo (palavra16) | 73,25% | 100,00% |
| Ruim (palavra17) | 75,00% | 100,00% |
| Sapo (palavra18) | 75,44% | 100,00% |
| Vacina (palavra19) | 73,68% | 100,00% |
| Vontade (palavra20) | 74,56% | 100,00% |
| MÉDIA GERAL | 73,14% | 100,00% |

5. Conclusão

O trabalho propôs e validou uma metodologia de reconhecimento automático de palavras da Libras baseada em dissimilaridade, visando superar os desafios impostos pela alta dimensionalidade e pela necessidade de deteção de vocabulário em constante evolução.

O sistema combinou a extração de características dinâmicas do *MediaPipe* (vetor de 504 dimensões) com a transformação para um espaço de dissimilaridade (vetor de 40 dimensões baseado em protótipos *K-Means*), culminando na classificação por um SVM otimizado.

Para o cenário de classificação de sinais conhecidos, o modelo atingiu uma acurácia robusta de 71,67%, mantendo a performance de classificação em um nível estável mesmo ao simular cenários de deteção de novidade (média de 73,14%).

A performance na identificação de sinais não conhecidos demonstrou a contribuição mais significativa. Atingir 100% de sucesso na deteção e rejeição de sinais desconhecidos através do método LOCO comprova que a abordagem de dissimilaridade é altamente eficaz em criar uma separação clara entre o vocabulário aprendido e sinais novos. Esta característica é essencial para a segurança e confiabilidade em aplicações práticas de reconhecimento de Libras, onde novos sinais são inevitáveis.

Para os trabalhos futuros, com o objetivo de expandir esta pesquisa, sugere-se:

1. Explorar o uso de Redes Neurais (DL) para aprender métricas de similaridade de forma otimizada;

2. Adaptar a metodologia para o reconhecimento de sinais contínuos (frases completas);
3. Expandir o vocabulário para testar a escalabilidade da abordagem com um número muito maior de palavras;
4. Otimizar o sistema para implementação em tempo real em dispositivos móveis ou web.

Referências

- [1] Hand Talk. Libras: o que é, quais os principais sinais, alfabeto e números? Blog da Hand Talk, n.d. URL <https://www.handtalk.me/br/blog/libras/>. Acesso em: 18 Nov. 2025.
- [2] World Federation of the Deaf (WFD). Contribution to the united nations enable on promoting the rights of persons with disabilities. UN Enable - United Nations Department of Economic and Social Affairs (DESA), n.d. Documento de posição/declaração, sem data de publicação específica. A WFD possui status consultivo na ONU desde a década de 1950.
- [3] Danilo de Milano and Luciano Barrozo Honorato. Visão computacional, 2010. Acessado em 15 de novembro de 2025.
- [4] Mario Köppen. The curse of dimensionality. *Fraunhofer IPK Berlin*, 2000. E-Mail: mario.koeppen@ipk.fhg.de. Documento acessado em 15 de novembro de 2025.
- [5] César Augusto Rodrigues Pacheco and Natasha Sophie Pereira. Deep learning conceitos e utilização nas diversas Áreas do conhecimento. *Revista Ada Lovelace*, 2: 34–49, 2018. ISSN 2596-3058.
- [6] Tamires Martins Rezende. *Reconhecimento automático de sinais da Libras: desenvolvimento da base de dados MINDS-Libras e modelos de redes convolucionais*. PhD thesis, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, 2021. URL <https://repositorio.ufmg.br/handle/1843/39785>. Acesso em: 11 abr. 2025.
- [7] Brunna Carolinne Rocha Silva. Desenvolvimento de tecnologia baseada em redes neurais artificiais para reconhecimento de gestos da língua de sinais. Master's thesis, Universidade Federal de Goiás, Goiânia, GO, 2018. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica e de Computação.
- [8] Márcio Fabiano Oliveira de Moura Santos and Cícero Costa Quarto. Classificação de libras em imagens através de redes neurais convolucionais. *Brazilian Journal of Development*, 8(12):80897–80915, 2022. doi: 10.34117/bjdv8n12-269.
- [9] Alisson R. Svaigen, Azzedine Boukerche, Linnyer B. Ruiz, and Antonio A. F. Loureiro. Dissident: A dissimilarity-based approach for improving the identification of unknown uavs. In *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023. doi: 10.1109/PIMRC56721.2023.10293825.
- [10] Guilherme Pereira Noya. Identificação de escritores usando dissimilaridade em bases multi-script, 2017. Orientador: Prof. Dr. Diego Bertolini Gonçalves.
- [11] Elżbieta Pełalska, Pavel Paclík, and Robert P.W. Duin. A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research*, 2: 175–211, 2001.
- [12] Google AI. Mediapipe vision – gesture recognizer. Site oficial do MediaPipe, 2025. URL https://ai.google.dev/edge/mediapipe/solutions/vision/gesture_recognizer?hl=pt-br.

- [13] Diego Mariano. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score. *BIOINFO – Revista Brasileira de Bioinformática e Biologia Computacional*, 2021. URL https://www.researchgate.net/profile/Diego-Mariano/publication/353001655_Metricas_de_avaliacao_em_machine_learning_acuracia_sensibilidade_precisao_especificidade_e_F-score/links/60e854a930e8e50c01f2e86a/Metricas-de-avaliacao-em-machine-learning-acuracia-sensibilidade-pr.pdf. Publicado em ResearchGate. DOI: 10.51780/978-6-599-275326-15.
- [14] Diego Addan. K-means: Algoritmos não supervisionados e de clusterização. Aula da disciplina de Inteligência Artificial, n.d. URL <http://inf.ufpr.br/diego/IA06.pdf>. Acesso em: 19 Nov. 2025.
- [15] Sarah. Medidas de similaridade e dissimilaridade. Medium, feb 2022. URL <https://medium.com/@sarahmbs/medidas-de-similaridade-e-dissimilaridade-479f3d5f64c6>. Acesso em: 18 Nov. 2025.
- [16] Vikramaditya Jakkula. Tutorial on support vector machine (SVM). Material de curso, Khoury College of Computer Sciences, Northeastern University, n.d. URL <https://course.khoury.northeastern.edu/cs5100f11/resources/jakkula.pdf>. Acesso em: 19 Nov. 2025.