

## 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. \*

Arthur Cesar da Silva Almeida \* Tamires Martins Rezende \*\*,\*\*\*  
Sílvia Grasiella Moreira Almeida \*\*\*\*

\* *Departamento de Engenharia de Controle e Automação, Universidade Federal de Ouro Preto, Campus Universitário Morro do Cruzeiro, Bauxita, Ouro Preto Minas Gerais, Brasil (e-mail: arthur.cesar@aluno.ufop.edu.br).*

\*\* *Machine Intelligence and Data Science (Minds) Laboratory, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627 - Escola de Engenharia - Bloco I, Sala 2200, Belo Horizonte, Minas Gerais, Brasil (e-mail: inc.tamires@gmail.com)*

\*\*\* *Fundação de Inovações Tecnológicas (FITec), Av Cristóvão Colombo, 485, 8º Andar - Belo Horizonte, MG (e-mail: tamiresrezende@fitec.org.br).*

\*\*\*\* *Instituto Federal de Minas Gerais (IFMG) - Campus Ouro Preto, Rua Pandiá Calógeras 898, Bauxita, Ouro Preto, Minas Gerais, Brasil (e-mail: silvia.almeida@ifmg.edu.br).*

---

**Abstract:** This paper aim the study and sign Libras cataloging for further search for the MINDS-Libras dataset. Although there is progress in the availability of sign bases in Libras for use by the scientific community, as is the case with MINDS-Libras, it is still a challenge for a Libras sign recognition area to have access to a sample of signs for training computer systems that can recognize signals automatically. It intends to expand to a MINDS-Libras base of two: evaluating the generation of synthetic data and analyzing the large mass of data publicly formed during the remote activities that take place in the period of the Covid-19 pandemic. The initiative for this work is in the contribution that they intend to force for the Computational vision of recognition that impacts directly in the social and cultural areas for the emerged community.

**Resumo:** O presente trabalho tem como objetivo o estudo e a catalogação de sinais em Libras para posterior ampliação da base de sinais MINDS-Libras. Embora haja um avanço na disponibilização de bases de sinais em Libras para uso pela comunidade científica, como é o caso da MINDS-Libras, ainda é um desafio para a área de reconhecimento de sinais de Libras ter acesso a amostras de sinais para treinamento de sistemas computacionais que possam reconhecer os sinais de forma automática. Pretende-se ampliar a base MINDS-Libras de duas formas: avaliando a geração de dados sintéticos e analisando a grande massa de dados disponibilizada publicamente durante as atividades remotas que aconteceram no período da pandemia da Covid-19. A relevância deste trabalho encontra-se na contribuição que pretende fornecer para a área de reconhecimento de gestos por meio de Visão Computacional, atuando na ampliação de canais de comunicação que possam impactar diretamente nas áreas social e cultural para a comunidade surda.

*Keywords:* Libras dataset; dataset expansion; Libras sign recognition.

*Palavras-chaves:* Base de dados de Libras; ampliação de base de dados; reconhecimento de sinais de Libras.

---

## 1. INTRODUÇÃO

A comunicação consiste em um processo que envolve troca de informações entre seres que utilizam sistemas simbólicos como suporte para este fim. É um processo fundamental e natural da vida e surge da necessidade do ser humano de se comunicar como parte de sua sobrevivência em sociedade. A comunicação é uma condição para que se fortaleça e se consolide uma cultura e ambas, comunicação e cultura, tornam-se indissociáveis nesse processo, uma vez que o ato de comunicar pressupõe um contexto cultural. Todo esse processo faz parte da evolução humana e é um componente essencial da educação, podendo ser realizado de forma verbal ou não-verbal (Perles, 2007). Em geral é atribuída maior relevância à comunicação verbal, que ocorre quando um grupo de indivíduos com interesses comuns se reúnem e comunicam-se por meio da linguagem falada. A comunicação de natureza oral-auditiva tende a ser a mais utilizada para a troca de informações, pois a maior parte da população mundial utiliza os sentidos de audição e fala para comunicar-se.

Contudo, dados recentes do Instituto Brasileiro de Geografia Estatística (IBGE) informam que há uma grande parcela da sociedade brasileira, cerca de 2,3 milhões de pessoas com 2 anos ou mais de idade, que possuem algum tipo de deficiência auditiva. Esta quantidade de pessoas com limitação de audição aumenta proporcionalmente com a idade (IBGE, 2021) e são, portanto, privadas da forma plena do uso da comunicação em sua forma oral-auditiva. De acordo com as estimativas para o futuro do primeiro relatório mundial sobre audição lançado pela Organização Mundial da Saúde (OMS) em 2021, um quarto da população global, cerca de 2,5 bilhões de pessoas, viverá com algum grau de perda auditiva em 2050. O estudo também estima que, sem medidas para prevenir e tratar este tipo de perda, pelo menos 700 milhões de pessoas precisarão de acesso a cuidados auditivos e outros serviços de reabilitação (ONU, 2021).

Para essas pessoas, privadas parcial ou totalmente da audição, tarefas simples do cotidiano tornam-se desafiadoras, em especial se o poder público não planeja e executa políticas de inclusão consistentes. A ausência destas políticas de forma regular acaba por contribuir para uma redução das interações sociais entre as pessoas privadas de audição.

A utilização da língua de sinais como forma de comunicação busca permitir que as pessoas se comuniquem por meio de gestos, os chamados sinais, que são formados por expressões faciais e corporais. Esta forma de comunicação é a forma natural de comunicação da população surda devido à sua característica visual e espacial, sendo uma importante ferramenta de inclusão social ao transcender as barreiras do silêncio impostas pela surdez (Castro, 2020).

Ao contrário do que muitos acreditam, a língua de sinais não é universal. Pelo contrário, cada país possui sua própria língua de sinais, com características próprias. No Brasil, a Língua Brasileira de Sinais (LIBRAS) é a segunda língua oficial do país e, da mesma forma como ocorre com as línguas orais-escritas, ela é composta por

uma estrutura linguística e gramatical próprias, além de apresentar regionalismos (Almeida, 2014).

Devido a estas características visuais e espaciais inerentes à língua de sinais, técnicas de visão computacional aplicadas a reconhecimento de gestos têm sido desenvolvidas com a intenção de que sistemas possam auxiliar a tradução entre as línguas. Uma barreira para os sistemas de reconhecimento automático da língua de sinais tem sido a dificuldade de obter conjuntos de dados (*datasets*) públicos de sinais dinâmicos com informações multimodais consistentes para o uso da comunidade científica (Cárdenas, 2020). O principal motivo deve-se pela existência de diversas ferramentas e técnicas disponíveis para a criação desses conteúdos, e que acabam por proporcionar métodos de aquisição de dados e protocolos de gravação das bases de dados muito diferenciados entre si (Almeida, 2014).

Avaliando o recente cenário da pandemia da Covid-19 com relação à disponibilização de material digital em vídeos, a sociedade viu-se forçada a adotar estratégias de enfrentamento à mesma, não apenas ao vírus, mas também aos obstáculos causados pelo distanciamento social. No meio acadêmico, por exemplo, essa circunstância contribuiu para que os alunos e professores optassem pela utilização das tecnologias digitais interativas, mudando e transferindo metodologias pedagógicas tradicionais das salas de aula (Oliveira et al., 2020). Essa estratégia resultou numa disponibilização em massa de conteúdos digitais em diversas áreas do conhecimento.

Diante deste contexto, o objetivo deste trabalho consiste no estudo e catalogação das bases de sinais da Libras disponíveis na literatura, visando a identificação de vídeos que possam contribuir na ampliação da base MINDS-Libras (Minds, 2019), que já encontra-se disponível publicamente e tem sido utilizada como ferramenta para treinamento em sistemas inteligentes. Além disso, técnicas de aumento sintético de dados também tem sido analisadas com este objetivo. Dessa forma, contabilizar vídeos diversos e buscar ferramentas para aumentar a base torna-se uma tarefa importante para a aplicação de metodologias comumente utilizadas em reconhecimento.

O presente artigo está organizado da seguinte forma: a seção 2 apresenta uma investigação de bases de sinais e glossários disponíveis ao público. A seção 3 é composta de uma descrição dos materiais utilizados para o cadastramento/registo de trabalhos e *datasets*, além das metodologias empregadas na ampliação da MINDS-Libras. Na seção 4 apresentam-se os resultados alcançados até a atual etapa da pesquisa. Finalmente, na seção 5, são realizadas reflexões a respeito da metodologia adotada e possibilidades para trabalhos futuros são discutidas.

## 2. REVISÃO BIBLIOGRÁFICA

Um importante passo no processo de aprendizado de máquina (*machine learning*) é criar e/ou encontrar dados que reflitam o problema a ser resolvido, para uso em seu treinamento e teste do algoritmo. Dessa forma, nesta seção serão investigados conjuntos de dados da aplicação apresentada, que é o reconhecimento de Libras, baseando-se em duas abordagens. A primeira, tem como enfoque a busca por *datasets* disponíveis na literatura. Já na segunda abordagem, a pesquisa foi direcionada para os conteúdos

\* Este trabalho foi desenvolvido sob financiamento do Instituto Federal de Minas Gerais - Campus Ouro Preto - Edital de Pesquisa 27/2021.

de mídia disponíveis em plataformas online de streaming e que possam ser utilizados na expansão da MINDS-Libras.

### 2.1 Datasets encontrados na literatura

No trabalho de Rodrigues (2021) é proposta a criação de uma robusta base de dados de vídeos, denominada V-Librasil. Esta base tem vídeos gravados com intérpretes nativos na Libras e tradutores profissionais, os quais sinalizam, ao todo, 1.364 palavras e expressões em Libras. Cada sinal é executado por três sinalizadores, totalizando 4.092 amostras. Algumas palavras ou expressões apresentam uma variação regional por parte dos sinalizadores durante a realização dos sinais e, em alguns casos, apresentam uma baixa resolução devido à forma de gravação do vídeo. Em relação às técnicas de gravação, todos os vídeos foram gravados em um ambiente controlado com *chroma key*. Desta forma, permite-se a adição de diferentes cenários de fundo a cada sinal gravado, o que possibilita o treinamento de sistemas mais eficazes contra ruídos. Por fim, uma plataforma online, chamada *V-Librasil Vídeo Dataset*, foi desenvolvida para permitir a pesquisa e o download dos vídeos com o intuito de fomentar o desenvolvimento de novas pesquisas.

Em Cárdenas (2020) foi desenvolvido um dataset público da Libras chamado LIBRAS-UFOP, que teve sua criação baseada no conceito de pares mínimos. Esse conceito indica que, caso haja alteração em apenas um dos três parâmetros primários de língua de sinais (configuração da mão, ponto de articulação e movimento) em um determinado sinal, haverá alteração na mudança de seu significado. Para a construção da base, todos os sinais foram selecionados, categorizados e validados por uma profissional proficiente em Libras. Além disso, foram consideradas diferentes variações de luminosidade, velocidade de realização do sinal, altura e vestimenta dos sinalizadores, visando garantir uma alta variação dentro das classes. Como método de aquisição de dados, foi utilizado o dispositivo *Microsoft Kinect V1* para obter dados de vídeo RGB, profundidade e de posições de esqueleto dos sinalizadores. Contudo, para a validação da base criaram-se duas versões do LIBRAS-UFOP. A primeira, contendo sinais isolados e a segunda contendo sinais contínuos. O conjunto com os sinais isolados é composto por 56 classes agrupadas em quatro categorias. Os sinais foram realizados por cinco sujeitos e cada um deles executou cada sinal dez vezes, em média. Já o conjunto com os sinais contínuos possui 37 classes. Eles foram executados por dez sinalizadores, e cada um deles executou cada sinal doze vezes em média. No trabalho de Cerna et al. (2021), os autores propuseram, para o conjunto de sinais isolados, um sistema de reconhecimento baseado em três redes neurais convolucionais (CNN) multi-stream funcionando em paralelo. Os resultados experimentais conseguiram atingir uma taxa de reconhecimento de sinais de 74,25%.

Em Gameiro et al. (2020) os autores elaboraram o CEFET Libras, um *dataset* composto por 547 seqüências de vídeos com 24 entradas, entre palavras e frases. Foi utilizada uma câmera RGB localizada em um ambiente controlado e com três tipos possíveis de fundos estáticos como método de aquisição de dados. Os sinais foram registrados em média por 20 sinalizadores não surdos. Este material encontra-se em processo de publicização por parte dos autores. Com relação a validação da base, os pesquisadores propuseram

um método de reconhecimento usando os k-vizinhos mais próximos (k-nearest neighbors) e o método *random forest*. Os resultados dos testes conseguiram alcançar uma precisão média de 65,81% na classificação multiclasse de sinais da Libras.

Diferente dos trabalhos anteriores, no artigo de Silva et al. (2020) os autores estruturaram uma base de expressões faciais da Língua Brasileira de Sinais denominada SILFA. As expressões faciais têm um papel importante na comunicação no contexto da Libras. Elas fornecem adjetivação, modificam a semântica dos sinais e também fornecem a coesão do que está sendo sinalizado, expressando negações, questionamentos, dúvidas, dentro outros sentimentos. A ausência de expressão facial, quando se realiza um sinal, pode produzir uma mensagem confusa. A base SILFA contém amostras de vídeos de expressões faciais de um grupo bastante diversificado composto por 10 indivíduos, surdos e ouvintes, fluentes em Libras. Esses indivíduos sinalizam 23 frases, que foram elaboradas para obter expressões faciais específicas. Parte das frases foram projetadas para expressar múltiplas expressões faciais. Já o protocolo de gravação é bem sólido ao detalhar as tecnologias envolvidas e o posicionamento dos articuladores frente à câmera. Para mais, todos os vídeos foram gravados em resolução de 1440x1080 pixels, que é adequada em termos de detalhamento, e foram capturados sob iluminação controlada. O SILFA também é uma base em processo de publicização pelos autores.

Outro trabalho relevante para a pesquisa é o de Stumpf et al. (2020), na qual é apresentado o SIGNBANK de Libras e uma ferramenta online de busca acessível ao público nacional e internacional. A primeira versão desse *dataset* começou a ser estabelecida em 2008 por meio do Identificador de Sinais. Com o tempo, sistema foi migrado para o Signbank de Libras e este foi consolidado com objetivo de integrar o Global Signbank, uma vez que a integração ao Global Signbank permitiria o acesso internacional ao mesmo e a possibilidade de desenvolvimento de estudos comparativos entre diferentes línguas de sinais. Atualmente o Signbank encontra disponível online para consulta de sinais, seus identificadores, traduções e outras informações linguísticas. Este banco de dados está em contínuo desenvolvimento, contando com pesquisadores do país para a sua ampliação e detalhamento linguístico. No que se refere a estrutura linguística do sinal, ela é descrita neste sistema considerando três aspectos: fonológicos, morfológicos e semânticos. Quanto ao acervo, o Signbank é composto por 3.061 sinais. Todas as amostras possuem uma boa resolução de vídeo, um plano de fundo uniforme e um enquadramento acima da cintura do articulador. Vale ressaltar que para algumas palavras, existem mais de um tipo de sinalização, que são relacionadas a variações regionais da Libras. No entanto existe apenas uma única seqüência de vídeo para cada palavra ou expressão.

Por fim, há o dicionário internacional *SPREAD THE SIGN*, um projeto sem fins lucrativos da organização *European Sign Language Centre*. De acordo com o trabalho de Hilzensauer e Krammer (2015), esse projeto tem o objetivo de oferecer, gratuitamente, um dicionário multilíngue na qual os sinais de países diferentes podem ser consultados e/ou comparados. A iniciativa começou como um dicionário vocacional direcionado a estudantes surdos que

viajavam para outros países e que queriam melhorar sua comunicação. Contudo, tornou-se uma ferramenta para qualquer pessoa que queira pesquisar e aprender, não se limitando mais a temas vocacionais. Atualmente, o Spread the Sign é o maior dicionário online de língua de sinais do mundo, abrangendo grande parte da Europa e outros países como o Brasil, EUA, Japão, entre outros. Ele está continuamente em crescimento e cada país parceiro tem o objetivo de alcançar a marca de 15000 sinais registrados na plataforma.

## 2.2 Conteúdos de mídia disponíveis em plataformas online de streaming

Nesta abordagem, foi encontrado o dicionário virtual de Libras da Universidade Federal de Viçosa (UFV, 2017), que pertence ao projeto Inovar +. Esse dicionário permite ao usuário buscar sinais a partir da identificação de categorias e também por configurações de mão, sendo constituído por aproximadamente 1.180 sinais regionais básicos. Todos os vídeos seguem o mesmo padrão de gravação com um plano de fundo uniforme de cor clara e um enquadramento acima da cintura do articulador. Todavia, pelo fato de ser um projeto voltado para o ensino e aprendizagem da Libras, existe apenas uma sequência de vídeo para cada palavra ou expressão.

Ao explorar um dos maiores sites de compartilhamento de vídeos do mundo, o YouTube, vários canais se destacam por possuírem desde vídeos amadores até produções sofisticadas acerca da Libras. Dentre eles, há o canal oficial do Instituto Federal de São Paulo - Câmpus Registro (IFSP, 2014), destinado a divulgação de conteúdos educacionais. Um desses conteúdos é referente a um glossário da Libras, que foi desenvolvido para servir como material de apoio. Esse glossário possui diversos sinais classificados em 25 categorias. Além disso, todos os vídeos são padronizados com um plano de fundo azul uniforme, possuindo um protocolo de gravação sólido, que estabelece uma posição de início e fim para cada sinal e um enquadramento do sinalizador acima da cintura.

O canal do departamento de Letras-Libras da Universidade Federal do Rio de Janeiro (UFRJ, 2016) também se destaca ao disponibilizar diversos conteúdos relevantes. Um desses conteúdos é o projeto de extensão “Libras para todos: ensino a distância”, que apresenta um conjunto de 45 vídeos de alta resolução, iluminação controlada, plano de fundo uniforme e protocolo de gravação bem definido. Neste projeto, são apresentados diversos glossários que, ao todo, somam cerca de 250 palavras ou expressões em Libras. Além dos glossários, são apresentados alguns exemplos práticos de diálogos simulados.

## 3. MATERIAIS E MÉTODOS

Neste tópico será apresentada a ferramenta desenvolvida para o cadastramento e análise dos *datasets*, visando uma maior praticidade para o momento da definição de estratégias de expansão da base MINDS-Libras. Além disso, também foram exploradas técnicas para aumentar artificialmente o conjunto de dados. Esse conceito é muito importante, pois modelos de aprendizado de máquina terão melhor desempenho e serão generalistas se o conjunto de dados for representativo e em proporção significativa.

### 3.1 Expansão da base MINDS-Libras

Uma planilha em Microsoft Excel foi estruturada para auxiliar no cadastro, visualização de informações e na busca por novos *datasets*. O software Microsoft Excel foi escolhido pelo fato de integrar-se facilmente a vários sistemas e por ser uma linguagem universal e conhecida. Além disso, a planilha foi desenvolvida com propósito de ser uma das contribuições deste trabalho para a comunidade científica, auxiliando a mesma na consulta e pesquisa por bases de dados da Libras. Sua interface de navegação pode ser observada na Figura 1.



Figura 1. Menu de Navegação  
Fonte: Elaborado pelo autor.

A ferramenta é composta por uma área de cadastramento de trabalhos acadêmicos que apresentam *datasets* (Figura 2). Neste campo é possível descrever tais características: (i) método de aquisição de dados, (ii) número de amostras, (iii) material disponibilizado, (iv) protocolo de gravação, (v) trabalhos relacionados, (vi) informações para contato, (vii) autores e (viii) observações. Além disso, há a possibilidade de registrar somente a base de dados, caso não seja encontrado, inicialmente, nenhum trabalho relacionado a ela.

Figura 2. Interface para cadastramento de trabalhos  
Fonte: Elaborado pelo autor.

Um campo para a pesquisa também foi criado. Nele, disponibilizam-se atalhos/hiperlinks para sites de publicação de livros de referência acadêmica e periódicos de artigos com foco em ciência, tecnologia, matemática e áreas afins com o intuito de oferecer uma maior praticidade. Em síntese, essa planilha visa centralizar as pesquisas da área, tornando-se um instrumento facilitador para a busca de estudos no reconhecimento automático da Libras.

### 3.2 Aumento sintético de dados

Os modelos de aprendizado de máquina são capazes de desempenhar diversas tarefas a partir das bases de dados.

Porém, muitas vezes, a quantidade de informação disponível é limitada e sem dados de treinamento suficientes. Modelos de aprendizado profundo (*Deep Learning*), quando treinados com exemplos limitados, tendem a sofrer *overfitting*, que acontece quando um modelo tem um desempenho preciso nos dados de treinamento, porém não consegue generalizar e extrapolar os dados durante o teste (Schons, 2018). O aumento de dados (*Data Augmentation* – DA) é um método eficaz para minimizar os problemas mencionados, sendo também indicado para melhorar o desempenho do modelo. Ele é empregado para expandir artificialmente o tamanho de um conjunto de treinamento, gerando dados sintéticos a partir alterações dos dados originais (Castro, 2020). É possível aumentar dados de áudio, texto, imagens e quaisquer outros tipos.

Para compreender o aumento de dados de imagens, primeiramente é necessário entender como uma imagem digital é composta. Imagens monocromáticas são compostas por vários tons de uma única cor e podem ser descritas por uma função discreta  $f(x,y)$ , onde  $x$  e  $y$  são coordenadas espaciais. Essas imagens podem ser representadas por uma matriz bidimensional, onde cada ponto da grade matricial, denominado de pixel, possui um valor que representa a intensidade luminosa no ponto observado (Hanel, 2021). Já para representar imagens coloridas, adota-se como padrão o sistema de cores RGB, embora existam outros. Nesse sistema um pixel é composto por três componentes que representam as intensidades de vermelho (*Red*), verde (*Green*) e azul (*Blue*). Em outras palavras, uma imagem colorida pode ser definida como uma composição de três imagens monocromáticas (de Queiroz e Gomes, 2006). As técnicas de DA para imagens ou quadros de vídeos baseiam-se em transformações geométricas, fotométricas e outros métodos avançados com *deep learning*. Além disso, alterações temporais também podem ser empregadas para aumentar o número de amostras, ao realizar, por exemplo, o escalonamento da duração da sequência de vídeo (Rezende et al., 2021).

*Transformações geométricas* As técnicas de transformações geométricas são indicadas para corrigir vieses posicionais, pois baseiam-se na alteração da posição espacial dos *pixels*. Em geral, não distorcem a figura na imagem. Dentre as técnicas baseadas nessa estratégia, há o espelhamento ou *flipping*, na qual é realizada uma inversão na imagem original, podendo ser uma inversão da mesma na horizontal e/ou na vertical (Hanel, 2021). Outro artifício é a rotação, na qual as imagens são rotacionadas em torno de um eixo entre  $1^\circ$  e  $359^\circ$ . As rotações leves ( $1^\circ$  a  $20^\circ$ ) geralmente são úteis em tarefas de reconhecimento. À medida em que o grau de rotação aumenta, há o risco do rótulo do dado não ser preservado. A transformação de cisalhamento ou *shear* utiliza um conceito semelhante à da técnica de rotação: um eixo da imagem é fixado enquanto ela é esticada em determinado ângulo. O corte de imagens ou *cropping* também tem gerado resultados promissores. Esse corte pode ser realizado aleatoriamente ou na área central de cada imagem. Porém, dependendo da dimensão do corte os rótulos podem não ser preservados. Por fim, temos as técnicas de zoom ou escala que podem ser aplicadas ampliando ou diminuindo o tamanho original da amostra (Claro et al., 2020).

*Transformações fotométricas* As técnicas de transformações fotométricas agem sobre a composição dos *pixels* da imagem, alterando seu valor. Geralmente, elas são empregadas para corrigir vieses de iluminação. A inserção de ruídos é uma das técnicas mais comuns. Nela, valores aleatórios são inseridos em determinados *pixels*. Essa estratégia acaba distorcendo padrões que ocorrem com frequência e que podem não ser úteis para a rede neural. Isso permite que a rede aprenda características mais robustas (Pérez Alonso et al., 2019). Há também o aprimoramento no espaço de cores que atua no histograma de cores resultando, por exemplo, na alteração do brilho, contraste, temperatura de cor, além de muitos outros aspectos (Claro et al., 2020). Por fim, tem-se a filtragem com *kernels*, que se baseia no deslizamento de um filtro sobre a matriz da imagem original realizando, ao mesmo tempo, a operação de convolução. Este procedimento é aplicado diversas vezes até que o filtro tenha percorrido a imagem por completo. As imagens resultantes contêm elementos que facilitam a identificação da classe alvo como bordas realçadas, detalhes mais aguçados sobre o objeto de interesse ou até mesmo um maior desfoque para aumentar a resistência a distorção de movimento. Vale ressaltar que, para estruturas de redes neurais convolucionais (*CNN's*) tal operação não é viável, pois essas estruturas são capazes de implementar esse processo internamente (Shorten e Khoshgoftaar, 2019).

#### 4. RESULTADOS E DISCUSSÕES

Após uma investigação minuciosa de trabalhos acadêmicos, dicionários online e plataformas de *streaming*, foram contabilizados 24 artigos que retratam métodos de construção de uma base de dados e a implementação de técnicas de validação. Dentre os trabalhos registrados, mais da metade foram desenvolvidos pelo grupo de pesquisa do laboratório MINDS da Universidade Federal de Minas Gerais (UFMG).

Em relação aos conjuntos de dados, foram registrados ao todo 25 itens (Figura 3), desde grandes conjuntos de dados, dicionários/glossários isolados até repositórios pertinentes ao ensino da Libras. Dentre esses itens, 8 são *datasets* robustos, devidamente consolidados e prontos para serem utilizados, sendo necessário apenas solicitar o acesso com os respectivos autores. Esse subgrupo é formado por 6 *datasets* referentes a sinais dinâmicos da Libras, 1 relacionado a expressões faciais e 1 associado às configurações manuais da Libras. Como foi descrito, essa planilha é uma das contribuições deste trabalho e encontra-se disponível em: <https://gitlab.com/tamiresrezende/expansao-da-base-minds-libras>.

Há também 6 itens referentes a repositórios educacionais que, mesmo não tendo dados pertinentes para a expansão do MINDS-Libras, podem ser úteis para outros pesquisadores da área. Por fim, 11 dicionários/glossários da Libras foram identificados. Diferente dos *datasets*, esses dicionários, por serem direcionados ao ensino e aprendizagem, apresentam apenas dados em formato de vídeo, um único registro para cada sinal e protocolos de gravação específicos. Apesar dos dicionários seguirem protocolos próprios, eles apresentam apenas alterações sutis entre si. Como exemplo estas alterações, temos a mudança da cor empregada no plano de fundo dos vídeos. No entanto,

LEGENDA	Nº	NOME	STATUS
Disponível	1	CEFET Libras	✘
Aguardando retorno	2	Grammatical Facial Expressions Data Set	✔
Não disponível	3	Libras - UFOP - Cont	✔
Não utilizável	4	SILFA	✘
	5	UFV - Dicionário de Libras	✔
	6	V - LIBRASIL	✔
	7	LIBRAS SIGN BANCK - UFSC	⚠
	8	Glossários "Curso de Libras Básico - EAD" UFRJ	✔
	9	Libras Movement	✔
	10	Libras do BoRa PESC	⚠
	11	TV INES	✘
	12	Repositório digital Huet	✔
	13	SD - Libras 10	✔
	14	LIBRAS HC - RGBDS	✔
	15	Canal oficial INEP - Glossários	✔
	16	Canal ENAP - Desafio da Libras	✔
	17	Fesai	⚠
	18	Glossário Economia em Libras	⚠
	19	Libras Pernambuco	⚠
	20	Glossário Libras - IFSP	⚠
	21	Hadassa	⚠
	22	Nathália Mendes - Libras Virtual	⚠
	23	Glossário - IFSC - Campus Palhosa	⚠
	24	Kelly Letícia	✔
	25	Glossário USP	⚠

Figura 3. Conjunto de dados cadastrados  
Fonte: Elaborado pelo autor.

uma dificuldade encontrada nesta etapa foi a variação significativa do inventário dos sinais de um dicionário para outro.

As técnicas de *data augmentation* investigadas foram implementadas no *Google Colaboratory*, que é um serviço de nuvem gratuito que visa promover a pesquisa de aprendizado de máquina e inteligência artificial. O *notebook* desenvolvido foi implementado em Python e também encontra-se disponível para consulta, juntamente com a planilha. A título de exemplificação, um frame do sinal de "Bala", cujo vídeo encontra-se disponibilizado online no Dicionário de Libras UFV, foi utilizado como referência em todas as transformações.

As transformações geométricas, ilustradas na Figura 4, são referentes a aplicação de técnicas de rotação aleatória leve ( $0^\circ - 20^\circ$ ), redimensionamento, translação horizontal e corte central. Já nas transformações fotométricas apresentadas na Figura 5, as técnicas de alteração da saturação, tonalidade (hue), ruído gaussiano e a mudança do domínio RGB para a escala de cinza foram aplicadas individualmente nesses exemplos.

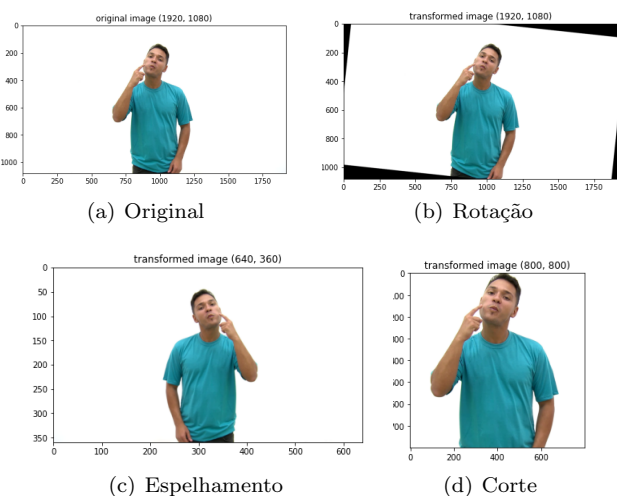


Figura 4. Transformações geométricas.  
Fonte: Adaptado de UFV (2017).

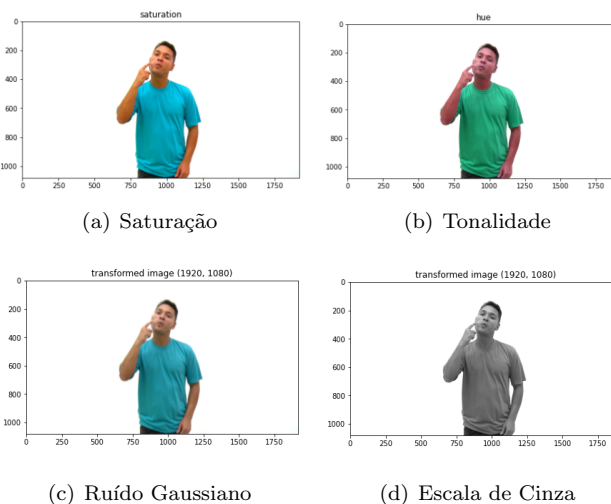


Figura 5. Transformações fotométricas.  
Fonte: Adaptado de UFV (2017).

Além da aplicação individual dessas técnicas, também há a possibilidade de empregar mais de uma técnica para uma mesma amostra. A Figura 6 exemplifica um cenário aonde todas as características de intensidade da imagem (brilho, contraste, saturação e tonalidade) foram modificadas e técnicas de rotação aleatória, redimensionamento e mudança de tonalidade foram aplicadas.

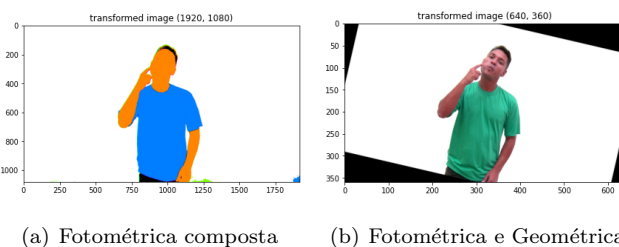


Figura 6. Amostras compostas por alterações fotométricas e geométricas  
Fonte: Adaptado de UFV (2017).

Depois da pesquisa por conteúdos/dados de sinais da Libras e do estudo de técnicas geométricas e fotométricas de aumento sintético de dados, a fase final concentrou-se na elaboração de estratégias sólidas para a expansão da base MINDS-Libras. Contudo, nesta etapa, foi percebido um impasse. Mesmo havendo variação do inventário dos sinais entre os glossários, o fato de não haver repetições do mesmo sinal disponíveis dificulta a expansão da base de dados devido à não padronização de gravação ao criar cada conjunto de dados.

As técnicas de *data augmentation* offline, neste ponto, ajudam a igualar o número de amostras e assim expandir o *dataset*. As técnicas possibilitam criar diversidade entre os dados. Do ponto de vista das abordagens de reconhecimento automático de sinais de libras utilizando técnicas de aprendizado de máquina, há uma demanda por diversidade em relação aos sinalizadores<sup>1</sup>, isto é, é necessário que o algoritmo aprenda o movimento do sinal e não se especifique nas características físicas de quem o executa.

<sup>1</sup> Sinalizador: quem executa o sinal.

Os resultados alcançados abrem espaço para abordagens alternativas para DA, indicando também que somente a utilização das técnicas abordadas neste trabalho podem não serem o suficiente. Nesse sentido, tendo em vista trabalhos futuros, um caminho promissor a ser seguido é o emprego de redes contraditórias generativas (Generative Adversarial Networks – GAN's). Essa ideia consiste num modelo que utiliza os dados da própria base de dados para gerar novas imagens com características similares à base de dados. Alinhando essa nova abordagem com os resultados alcançados neste trabalho, é possível realizar a criação/expansão de uma base de dados mais robusta, pois o número de repetições do sinal por parte dos articuladores não seria um fator impactante.

## 5. CONCLUSÃO

Este trabalho consistiu no estudo e catalogação das bases de dados da Libras disponíveis na literatura. Para isso, foi desenvolvida uma ferramenta acessível ao público para o cadastramento e especificação de trabalhos e/ou bases de dados. Além disso, as vantagens do uso de técnicas geométricas e fotométricas de *data augmentation* foram analisadas. Dentre as contribuições deste trabalho para a comunidade científica estão a planilha de cadastramento e o código implementado no Google Colab para as aquisições dos quadros de vídeo, aplicação das técnicas de transformações geométricas e fotométricas, ambos disponíveis publicamente em: <https://gitlab.com/tamiresrezende/expansao-da-base-minds-libras>.

A importância da utilização de técnicas de *data augmentation* para melhorar o desempenho de modelos de aprendizado de máquina é notória, mas cada problema demanda uma análise específica. Tendo como foco o aprendizado de máquina, mais precisamente o reconhecimento de sinais, as transformações aqui apresentadas podem ser expandidas com técnicas de *deep learning* com foco na diversidade dos sinalizadores. Nesta área há uma gama de possibilidades, que vão sendo definidas de acordo com a capacidade computacional dos sistemas empregados, já que processar vídeos tem um alto custo computacional.

Como continuação deste trabalho, tais técnicas serão exploradas com o intuito de melhorar a representatividade da base de dados e gerar um modelo de aprendizado generalista.

## AGRADECIMENTOS

Os autores agradecem ao incentivo, direcionamento e financiamento do Instituto Federal de Minas Gerais - Campus Ouro Preto, Edital de Pesquisa 27/2021 - DPIP, e ao laboratório MINDS/UFMG.

## REFERÊNCIAS

Almeida, S.G.M. (2014). *Extração de características em reconhecimento de parâmetros fonológicos da Língua Brasileira de Sinais utilizando sensores RGB-D*. Ph.D. thesis, Universidade Federal de Minas Gerais.

Castro, G.Z. (2020). *Reconhecimento de Línguas de Sinais Utilizando Redes Neurais Convolucionais e Transferência de Aprendizado*. Dissertação de mestrado, Universidade Federal de Minas Gerais - UFMG.

Cerna, L.R., Cardenas, E.E., Miranda, D.G., Menotti, D., e Camara-Chavez, G. (2021). A multimodal libras-ufop brazilian sign language dataset of minimal pairs

using a microsoft kinect sensor. *Expert Systems with Applications*, 167, 114179.

Claro, M., Vogado, L., Santos, J., e Veras, R. (2020). Utilização de técnicas de data augmentation em imagens: Teoria e prática. *Sociedade Brasileira de Computação-SBC. Disponível em:* < <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/48/217/455-1>.

Cárdenas, E.J.E. (2020). *Desenvolvimento de uma abordagem para reconhecimento contínuo da Língua Brasileira de Sinais utilizando imagens dinâmicas e técnicas de aprendizagem profunda*. Tese, Universidade Federal de Ouro Preto. Ciência da Computação.

de Queiroz, J.E.R. e Gomes, H.M. (2006). Introdução ao processamento digital de imagens. *Rita*, 13(2), 11–42.

Gameiro, P.V., Passos, W.L., Araujo, G.M., de Lima, A.A., Gois, J.N., e Corbo, A.R. (2020). A brazilian sign language video database for automatic recognition. In *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*, 1–6. IEEE.

Hanel, M.S. (2021). Análise de métodos de data augmentation para melhoria do desempenho de uma rede neural de detecção de defeitos em superfícies metálicas.

Hilzensauer, M. e Krammer, K. (2015). A multilingual dictionary for sign languages: “spreadthesign”. *ICERI2015 Proceedings*, 7826–7834.

IBGE, E.S. (2021). Pns 2019: país tem 17,3 milhões de pessoas com algum tipo de deficiência. URL <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/31445-pns-2019-pais-tem-17-3-milhoes-de-pessoas-com-algum-tipo-de-deficiencia>. Data de acesso: 18/02/2022.

IFSP (2014). Instituto federal de são paulo - campus registro. <https://www.youtube.com/channel/UC9oAJWdDlPxBKVZ9HnzIrEw/about>.

Minds (2019). Brazilian sign language recognition. <http://minds.eng.ufmg.br/project/4>.

Oliveira, R.M.d., Corrêa, Y., e Morés, A. (2020). Ensino remoto emergencial em tempos de covid-19: formação docente e tecnologias digitais. *Revista Internacional de Formação de professores*, 5, e020028–e020028.

ONU (2021). Cerca de 2,5 bilhões de pessoas sofrem perda auditiva até 2050. URL <https://news.un.org/pt/story/2021/03/1743112>. Data de acesso: 18/02/2022.

Pérez Alonso, D.A. et al. (2019). Análisis de señales de tos para detección temprana de enfermedades respiratorias.

Perles, J.B. (2007). Comunicação: conceitos, fundamentos e história. *Biblioteca on-line de Ciências da Comunicação*, 1–17.

Rezende, T.M. et al. (2021). Reconhecimento automático de sinais da libras: desenvolvimento da base de dados minds-libras e modelos de redes convolucionais.

Rodrigues, A.J. (2021). V-librasil: uma base de dados com sinais na língua brasileira de sinais (libras). *Dissertação de Mestrado, Centro de Informática-CIN, Universidade Federal de Pernambuco-UFPE*.

Schons, T. (2018). *Rede de Convolução para Sistema Biométrico baseado em EEG*. Ph.D. thesis, Universidade Federal de Ouro Preto.

Shorten, C. e Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of*

*big data*, 6(1), 1–48.

- Silva, E.P.d., Costa, P.D.P., Kumada, K.M.O., e De Martino, J.M. (2020). Silfa: Sign language facial action database for the development of assistive technologies for the deaf. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 688–692. IEEE.
- Stumpf, R., Pizzio, A.L., Lucinda, R., Müller de Quadros, R., e Crasborn, O. (2020). Signbank da libras. banco de señas de libras.
- UFRJ, D.d.L.L. (2016). Proj. de extensão “libras para todos: Ead”, dep. de letras-libras. URL <https://www.youtube.com/c/LetrasLibrasUFRJ/about>. Data de acesso: 28/02/2022.
- UFV (2017). Dicionário de libras. URL <https://sistemas.cead.ufv.br/capes/dicionario/#>. Data de acesso: 20/02/2022.