Sinais de Libras para Áudio: primeiros passos de uma ferramenta inovadora

Nicolas Alves Suzuki  
RA: 140712

Isabela Cristina Silva Pedro   
RA: 141658

***Abstract–* O presente artigo tem como objetivo fomentar o aumento das pesquisas em machine learning aplicada à linguagem de sinais, com ênfase na interação entre pessoas surdas e mudas; pois como constatado durante a análise dos trabalhos relacionados, trata-se de um tema que ainda recebe pouca atenção na área.**

**Para tal finalidade, foi desenvolvido um sistema que, conectado à uma webcam, recebe como entrada sinais da Libras correspondentes às letras do alfabeto e a determinadas palavras, os processa e então converte em áudio, que pode ser reproduzido pelo usuário. O banco de dados utilizado foi obtido por meio de fontes online para livre acesso, o modelo escolhido foi a Rede Neural Convolucional (CNN) e para o método de validação foi utilizado a classificação cruzada estratificada com 6 folds.**

**Resultados principais (todo)**

**Conclusões (todo)**

***Palavras-chave–* Libras, machine learning, CNN, linguagem de sinais, interação.**

1. Introdução

A comunicação é o âmago de qualquer sociedade, e existe nas mais variadas formas para além da oral e escrita. A comunicação tátil utiliza o uso de toques para transmitir informação, como o caso do Braille, que possui seu alfabeto definido por padrões em alto relevo e é amplamente utilizado por pessoas com deficiência visual. A linguagem de sinais é mais uma dessas variadas possibilidades de se comunicar, baseia-se em gestos manuais, expressões faciais, movimentos corporais e expressões visuais para transmitir significado e são utilizadas por comunidades surdas ao redor do mundo como principal meio de comunicação.

Apesar da variedade de formas de trocar informações e ideias, as que perpetuam são as formas orais e escritas, o que molda um gap quando se fala sobre inclusão de pessoas com algum tipo de deficiência, já que até os dias atuais as diferentes formas de linguagem não foram consolidadas no dia-a-dia. Muito dificilmente se vê salas de cinema ou até mesmo vídeos nas mais diferentes plataformas com tradução para Libras, por exemplo.

Ao ir para o cenário de pessoas com deficiência auditiva, segundo a OMS 1,5 bilhões de pessoas no mundo possuem algum grau de surdez, que divide-se em casos leves, moderados, severos e profundos em ambos os ouvidos. Cerca de 12,6 milhões de pessoas no mundo são totalmente surdas, incapazes de ouvir qualquer som. Dentro da perspectiva do Brasil, segundo o IBGE, em 2022 o número de pessoas com algum grau de surdez foi de 2,3 milhões.

A conectividade proporcionada pela internet trouxe inúmeras oportunidades, mas também revelou barreiras significativas para a inclusão de pessoas com deficiência auditiva. A Língua Brasileira de Sinais (Libras) é a principal forma de comunicação desses indivíduos mas, como citado por um exemplo anteriormente, não é amplamente utilizada no ambiente digital. Isso resulta na falta de acessibilidade em sites, aplicativos e plataformas digitais, prejudicando a experiência desses indivíduos no acesso às informações e serviços online.

A acessibilidade é um direito básico e essencial de todos, independentemente de suas limitações físicas ou cognitivas. A exclusão de pessoas com deficiência auditiva pode exacerbar as desigualdades existentes na sociedade e limitar suas oportunidades de interação social.

Diante dessa questão, o projeto visa acima de tudo estimular o aprendizado e a inclusão da Libras, por meio de um projeto de uma aplicação utilizando de inteligência artificial que traduzirá as linguagens de sinais em áudio, possibilitando a interação dos usuários da Libras não só com aqueles que ainda não aprenderam a linguagem de sinais, mas também com deficientes visuais, de forma a abrir uma nova possibilidade de interação. Dessa forma, planeja-se dar mais um passo em direção a um futuro mais abrangente quando trata-se de interação social.

1. Conceitos Fundamentais
   1. *Língua Brasileira de Sinais (Libras)*

É primordial compreender que as línguas de sinais, assim como as línguas orais, possuem gramática, estrutura, sintaxe próprias e variação regional, ou seja, diferentes países - e até mesmo regiões dentro de um país - podem ter suas próprias línguas de sinais com variações nos gestos, movimentos e expressões. Além disso, elas também são oficialmente reconhecidas. No Brasil a Língua Brasileira de Sinais (Libras) é uma língua oficial desde 2002. Fora do país, são exemplos: American Sign Language (ASL) nos Estados Unidos, British Sign Language (BSL) no Reino Unido e Langue des Signes Française (LSF) na França.

De forma sucinta, a comunicação na Libras utiliza o mesmo princípio da oral, pois envolve a combinação de sinais individuais para formar frases e expressar ideias completas. É possível, com um sinal, formar palavras como “casa” e “comida”, como também conceitos mais amplos, como “felicidade” ou “tristeza”. Ao combinar os sinais para “eu”, “ir” e “escola” na sequência e com expressões corretas, a frase é expressa de maneira completa e compreensível. Portanto, as expressões - combinações de palavras ou sinais para transmitir significado específico ou ideia mais abrangente - também podem ser manifestadas pelos sinais, como “dar uma mãozinha” no sentido de ajudar alguém e são devidamente compreendidas.

* 1. *Inteligência Artificial (IA)*

A Inteligência Artificial é uma área da ciência da computação que busca criar sistemas capazes de simular a inteligência humana. Ela envolve o desenvolvimento de algoritmos e modelos que permitem às máquinas realizar tarefas cognitivas, como reconhecimento de padrões, aprendizado, raciocínio lógico, entre outros.

* 1. *Visão Computacional*

A A Visão Computacional é um subcampo da IA que se concentra em capacitar as máquinas a compreender e interpretar informações visuais. Ela utiliza algoritmos e técnicas para analisar imagens ou vídeos, identificar objetos, reconhecer padrões, realizar rastreamento de movimento, entre outras aplicações.

1. Trabalhos Relacionados

Existem trabalhos na literatura que abordam o objetivo de auxiliar a comunicação em Libras utilizando a IA. Alguns exemplos desses trabalhos são:

* 1. *Design and implementation of CNN for Sign Language recognition*

Nesse trabalho, o dataset foi criado manualmente. Primeiramente, a câmera captura o símbolo ou sinal e cria frames de imagens e então armazenadas, para isso foi utilizada a biblioteca OpenCV. Foi realizado um cálculo do background e posteriormente um novo cálculo com um objeto em frente a ele, para que esses valores pudessem ser utilizados para distinguir se existe ou não uma mão na região de interesse e, se existe, calcular seu contorno. A partir disso, foram armazenadas as imagens nos arquivos de treino e teste, para que posteriormente pudessem virar letras ou números. Foram armazenadas 300 fotos de cada número na base de dados para treino e 40 fotos de cada número para a base de dados de teste. Keras foi utilizado para importar o dataset com os nomes das pastas correspondentes ao nome da classe das fotos. SGD e Adam foram as técnicas de otimização utilizadas e foi constatado que SGD teve melhor acurácia. O modelo foi treinado com taxa de 100% e uma validação de acurácia de 81%.

* 1. *Sign Language recognition using machine learning*

No seguinte trabalho, a proposta é reconhecer os sinais por meio de transmissão de vídeo e traduzi-los para a sua respectiva letra do alfabeto americano. O input é um vídeo, processado frame por frame, utilizando OpenCV. O contorno é encontrado ao escurecer a imagem e identificar a borda branca que delimita a mão.

O dataset utilizado foi composto por 26 imagens correspondentes aos sinais da linguagem de sinais americana, uma imagem para cada letra.

Para a classificação foi necessário um fundo branco, com a palma da mão do intérprete em direção à câmera, e o algoritmo utilizado foi o modelo SVM (Support Vector Machine) fornecido pela biblioteca scikit-learn. A limitação do dataset obteve como consequência a incorreta classificação de alguns sinais, principalmente os que necessitam de movimento das mãos.

* 1. *Deepsign: Sign Language Detection and Recognition Using Deep Learning*

Neste trabalho, foi utilizado um modelo de rede neural recorrente (RNN) com unidades de memória LSTM e GRU e arquitetura InceptionResNetV2. Um vídeo com a sequência de sinais é utilizado como input e é segmentado, por meio de amostragem, em vídeos menores que contêm, cada um, uma palavra que foi expressa pelos sinais. Assim, quadros individuais são extraídos para que possam ser analisados individualmente. Para isso, o sistema identifica quando um sinal começa e quando ele termina, e assim as palavras são separadas.

Os frames tiveram o tamanho ajustado e fornecidos ao modelo InceptionResNetV2 com pesos pré-treinados do MobileNet (RNC) para processar cada quadro do vídeo e extrair as características relevantes dos gestos. Isso gerou uma matriz com vetor de características. Essas características extraídas dos gestos foram então passadas para a rede neural recorrente (RNN) para prever a palavra correta. O dataset para treinar e testar foi criado pelos autores, os gestos foram realizados por 16 pessoas diferentes, homens e mulheres entre 20 e 25 anos. Os vídeos continham duração de aproximadamente 2 segundos e constituíram um dataset de 11 palavras, onde para cada palavra foram criadas 1100 amostras de vídeo.

Durante os estudos o modelo que teve acurácia mais alta foi o LSTM-GRU model, e foi desenvolvido utilizando Keras/Tensorflow. Esse modelo teve uma acurácia de aproximadamente 95% em analisar os sinais de variados conjuntos de dados.

* 1. *Audio to Sign Language Converter*

O objetivo desse trabalho é detectar o sinal em audio, converter em string por meio das bibliotecas do Python e então checá-las com o dataset. A imagem ou GIF resultante é mostrada(o) na tela na linguagem de sinais indiana. Pyaudio é utilizado para gravação de áudio, então SPHINX é utilizado para o reconhecimento dele.

PIL é utilizado para manipular, filtrar e abrir imagens. Um vetor do tipo NumPy é utilizado para manipular os pixels das imagens. Matplotlib possibilita plotar as imagens ou GIFS.

Foi utilizado um dataset que contém todas as letras em inglês e GIFS para as palavras e expressões mais frequentes. Assim, o áudio é captado e convertido em string, a string é então processada e classificada na sua respectiva representação em ISL. Caso essa classificação seja bem sucedida, a imagem é plotada por meio de técnicas de extração e plot de imagens.

Para facilitar a utilização, foi desenvolvido um GUI fornecido por meio do easygui e Tkinter.

* 1. *Libras Recognition using Deep Learning*

Neste trabalho, os pesquisadores propuseram um sistema de reconhecimento de gestos em Libras utilizando redes neurais convolucionais (CNN). Eles coletaram um grande conjunto de dados de gestos em Libras e treinaram a CNN para classificar corretamente os sinais. O sistema alcançou uma taxa de reconhecimento de alta precisão e demonstrou promessa na comunicação em Libras.

* 1. *Aplicativo para facilitação da comunicação em Libras*

Através de técnicas de processamento de linguagem natural (PLN), foi desenvolvido um aplicativo com o intuito de expandir a comunicabilidade dos surdos e servir como alternativa para um diálogo baseado na paridade entre Português e Libras.

O aplicativo utiliza visão computacional para capturar os gestos em Libras e, em seguida, exibe o correspondente texto e áudio para o usuário receptor.

O objetivo do projeto é facilitar a interação e a comunicação inclusiva.

Todos os sinais usados no aplicativo provêm do Dicionário da Língua Brasileira de Sinais V3, que é fornecido publicamente pelo Ministério da Ciência, Tecnologia e Inovação. A base de dados foi criada manualmente com essas informações disponíveis, que contém sobre cada um dos sinais catalogados na plataforma online do ministério, informações como: assunto, forma da mão, acepção, exemplo de frases. Além disso, todos os dados mostrados em tela são obtidos através de chamadas REST, o que facilitou a criação da base de dados utilizada para formar o sistema e treinar a inteligência artificial.

O autor utilizou a spaCy, que é uma biblioteca gratuita e de código aberto para processamento de linguagem natural em Python que permitiu que ele pudesse utilizar de vários algoritmos como tokenização, lematização, classificação de texto, entre outras, para aplicar no desenvolvimento completo, já que utiliza o processamento de imagem e texto por exemplo.

* 1. *Real-Time Sign Language Recognition Using Machine Learning Techniques*

Este trabalho propõe um sistema de reconhecimento de linguagem de sinais em tempo real usando técnicas de aprendizado de máquina. Os autores utilizam o RTS - Real-Time Sign para reconhecer e interpretar esses gestos.

O objetivo é facilitar a comunicação entre pessoas surdas e ouvintes em tempo real.

Foi utilizado o dataset American Sign Language disponível com traduções e imagens para todas as linguagens presentes na américa. Esse dataset é categorizado por línguas, como inglês, português e então individualmente fornecem as imagens e traduções relativas da linguagem que foi selecionada, como o objetivo do autor era demonstrar a funcionalidade em algumas linguagens ele utilizou os datasets das seguintes linguagens: Inglês Americano, Inglês Britânico, Indiano, Italiano e Turco.

Como algoritmo de utilização principal, o autor utiliza o “Support Vector Machine” juntamente com o framework MediaPipe. Esse algoritmo funciona mapeando os dados de entrada em um espaço dimensional superior e, em seguida, encontrar um hiperplano que melhor separa as diferentes categorias.

* 1. *Sign Language Recognition Using Convolutional Neural Networks and Temporal Features*

Neste estudo, os autores exploram o uso de redes neurais convolucionais (CNNs) e características temporais para o reconhecimento da língua de sinais. Eles propõem uma abordagem que combina informações espaciais e temporais dos gestos para melhorar a precisão do reconhecimento da imagem.

O CNN funciona aplicando filtros de convolução nas camadas convolucionais para realizar a detecção de padrões em diferentes regiões da imagem. Cada filtro é uma matriz numérica que é deslizada pela imagem original, calculando a convolução em cada posição. Isso gera um mapa de características que destaca regiões relevantes, que no caso, o autor utiliza para identificar as mãos das fotos e calcular baseado na matriz qual sinal está presente na imagem.

Para a realização desse projeto, foi utilizado o dataset “ChaLearn Looking at People 2014” disponível publicamente na internet, que contém mais de 27 pessoas que realizam gestos da linguagem de sinais com 20 variações cada, com roupas, com filtros de calor e indicadores de movimento*.*

* 1. *Mobile Applications for Deaf and Hard-of-Hearing Individuals: A Review of Features and Accessibility*

Neste artigo de revisão, a autora examina as características e a acessibilidade de aplicativos móveis desenvolvidos para pessoas surdas e com deficiência auditiva. Ela analisa diferentes aplicativos disponíveis no mercado e destaca as funcionalidades que podem auxiliar na comunicação, educação e acessibilidade para esse grupo de usuários.

* 1. *HandArch: A deep learning architecture for LIBRAS hand configuration recognition*

Neste trabalho, os pesquisadores apresentaram o HandArch, uma nova arquitetura para reconhecimento da mão em tempo real a partir de vídeo para acelerar o desenvolvimento de aplicativos de reconhecimento de linguagem de sinais.

Para a realização do projeto, os pesquisadores apresentam o Libras91, um novo conjunto de dados de configurações manuais da Língua Brasileira de Sinais (LIBRAS) contendo 91 classes e 108.896 amostras.Utilizando como parâmetro diversas bases disponíveis publicamente, eles criaram a própria base de dados para desenvolver e treinar a IA.

Para criar a própria arquitetura de reconhecimento de imagens da mão em tempo real, os pesquisadores utilizaram os algoritmos SLR e suas aplicações em diversos cenários e unificaram para realizarem o reconhecimento independente das condições de iluminação e fundo da imagem

1. Objetivo

O presente trabalho tem como objetivo desenvolver um sistema que receba como entrada uma transmissão em vídeo e seja capaz de transformar sinais da Libras captados nos frames deste vídeo para descrição em áudio das respectivas palavras em português.

1. Metodología Experimental

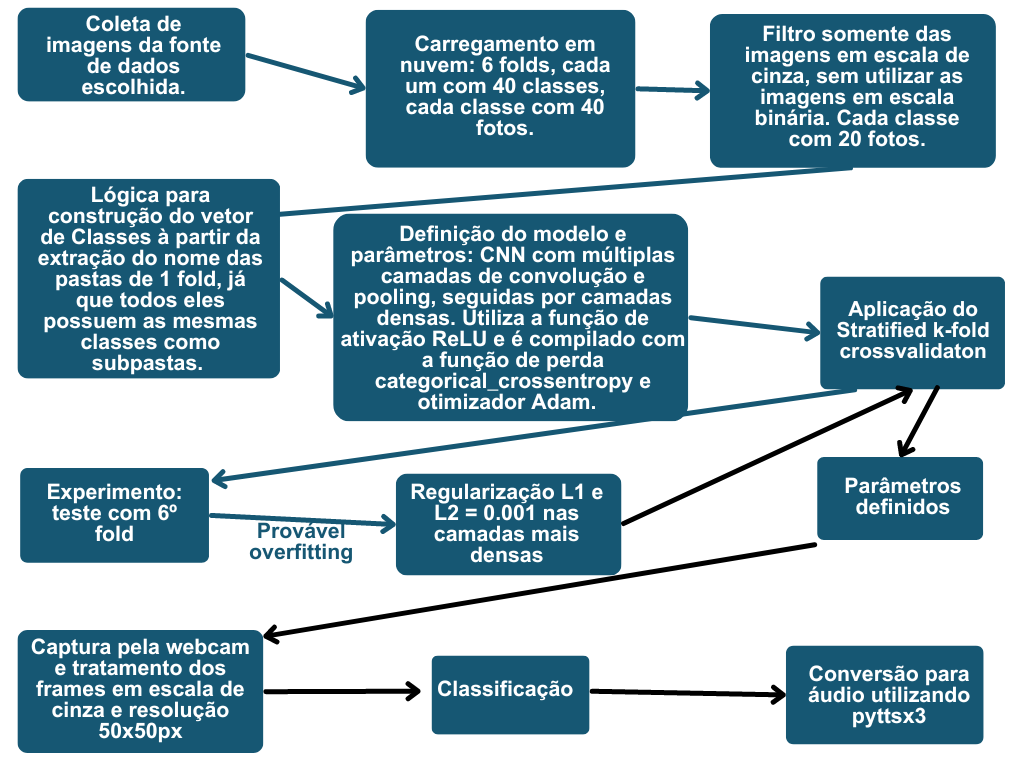


Fig. 1. Pipeline experimental

A coleta de imagens da fonte de dados foi feita por meio do trabalho [11], onde são de autoria dos próprios autores do artigo. No trabalho em questão, os autores dividiram o *dataset* em 6 *folds* para a aplicação do *6-fold cross validation*. No caso desse trabalho, nós utilizamos os dados contidos apenas em 5 desses 6 folds no treinamento por meio do *stratified 5-fold cross validation*, pois optamos por deixar os dados do 6º *fold* para testes.

A escolha deu-se pelo seguinte motivo: Os testes em [11] após o treinamento e validação do modelo foram realizados em um *dataset* não específico de Libras, enquanto que o objetivo do grupo foi testá-lo apenas com sinais da Libras. Portanto, o treino e validação do modelo ocorreram com 5 *folds,* sendo o último utilizado nos experimentos.

Outra filtragem foi realizada no *dataset* referente às imagens. O *dataset* presente em [11] possui imagens em escala de cinza e em máscara binária, entretanto o grupo optou por treinar o modelo apenas com as primeiras mencionadas.

O modelo escolhido foi a Rede Neural Convolucional (RNC) - conhecida como *Convolutional Neural Network* (CNN) - não correspondendo ao modelo escolhido por [11]. Em [11] foi utilizado o *Multilayer Perceptron,* portanto também foi necessário utilizar mais recursos para capturar os dados das imagens: os descritores de forma (HOG e ZIM) e um vetor de características associado. Já nesse trabalho, ao utilizar a CNN, essas etapas de pré-processamento foram eliminadas.

A rede neural convolucional desse experimento possui ReLU como função de ativação e Adam como otimizador (mais detalhes presentes no código fonte em [13]) e inicialmente não contou com regularização.

Após o primeiro experimento com o *6º fold* (separado para somente testes) foi observado uma alta acurácia na validação, e uma significativa queda da mesma no teste, o que o grupo considerou como provável *overfitting*. foram adicionadas então regularizações L1 e L2 nas camadas mais densas, e o modelo apresentou melhora nos resultados obtidos no teste.

Com os parâmetros do modelo definidos, foi iniciado o experimento com os dados obtidos por *webcam*. Para a captura e processamento de imagens para as proporções e tonalidades requisitadas, foram utilizadas as bibliotecas *IPython.display, google.colab.output, base64, time, cv2 e os*. Foi desenvolvido um algoritmo para capturar 5 frames, um frame a cada 2 segundos, para que os sinais pudessem ser classificados.



Fig. 2. Imagens capturadas pela *webcam*

Após essas imagens serem armazenadas, foram classificadas pelo modelo e então um vetor foi gerado com as classificações, que por sua vez foi convertido ao áudio. Para essa conversão, a biblioteca *gTTS* foi utilizada.

* 1. *Base de Dados*

O dataset utilizado neste trabalho, intitulado *Recognition of Static Gestures Applied to Brazilian Sign Language* (Libras) [11], consiste em uma coleção de imagens de gestos estáticos das mãos que representam diferentes sinais em Libras. Para a sua formulação, 3 especialistas em Libras e dois alunos surdos se ofereceram como modelos para compor o conjunto de dados, que contém 9600 imagens [12].

O conjunto de signos é composto por: letras do alfabeto Libras: A, B, C, D, E, F, G, I, L, M, N, O, P, Q, R, S, T , U, V, W, X, Y, números: 1, 2, 4, 5, 7 e 9; e palavras: Avião, Palavra, Adulto, América, Casa, Gás, Lei, Identidade, Junto, Pedra, Pequeno e Verbo, totalizando 40 sinais diferentes. Para cada sinal realizado pelos autores foram geradas 40 imagens no formato PNG com resolução de 50x50 pixels, das quais 20 correspondem a imagens em tons de cinza e a outra metade a imagens em preto e branco. Para a formulação do *dataset* desse trabalho, foram somente consideradas as imagens em tons de cinza, que correspondem à 4.800 das 9.600 imagens presentes no trabalho baseado pelo grupo.



Fig. 4. Diferentes posturas para o mesmo sinal (Bastos et al., 2015)

Além disso, diferentemente de [11], o grupo decidiu por treinar e validar o modelo apenas com *5 folds*, utilizando o último apenas para testes, já que o *dataset* utilizado para testar o modelo em [11] não é composto por sinais da Libras.

* 1. *Protocolo de Validação*

Devido ao tamanho consideravelmente limitado do *dataset*, mesmo sendo uniforme, o grupo optou por utilizar a validação cruzada estratificada com 6 *folds*.

Essa escolha deu-se pelo objetivo de garantir a mesma proporção de exemplos de cada classe em cada *fold*, para que alguma classe não seja dominante em algum deles e uma maior robustez seja alcançada.

* 1. *Medidas de avaliação*

Durante o protocolo de validação foram utilizadas as medidas: acurácia, f1 *score* e precisão para cada *fold*.

No fold separado apenas para teste onde ocorreu o experimento do modelo, os valores obtidos foram: acurácia = 0.8919; precisão = 0.8868; f1Score = 0.8819.

1. Resultados e Discussões

Ao utilizar CNN o código de pré-processamento e uma única etapa de validação, diferentemente de [11], o algoritmo envolvido no pré-processamento de dados e na classificação ficou mais sucinto. Ainda assim, não foi observado uma redução do desempenho do modelo utilizado no grupo em relação ao MLP, utilizado por [11]. Em [11], a acurácia média do modelo após a validação por *6-fold cross validation* foi de 96,77%, enquanto que a acurácia alcançada pela CNN ao utilizar *stratified 5-fold cross validation* foi de aproximadamente 97.58%.

Tabela I

Acurácia, precisão e f1 score por fold durante o *stratified 5-fold cross validation*

| **Folds** | **Medidas de avaliação** |
| --- | --- |
| Fold 1 | Accuracy : 0.9461, Precision: 0.9555, f1Score: 0.9464 |
| Fold 2 | Accuracy : 0.9832, Precision : 0.9839, f1Score : 0.9831 |
| Fold 3 | Accuracy : 0.9731, Precision : 0.9768,  f1Score : 0.9728 |
| Fold 4 | Accuracy : 0.9899, Precision : 0.9906,  f1Score : 0.9898 |
| Fold 5 | Accuracy : 0.9865, Precision : 0.9882,  f1Score : 0.9862 |

Durante a etapa de testes, o modelo alcançou uma acurácia de 89.2%, o que indica que está generalizando bem os padrões aprendidos durante o treinamento para amostras não vistas anteriormente. Apresentou também uma precisão de 88.69% e um f1Score de 88.2%.

(tabela dessas validacoes por sinal)

Ao aplicar o modelo aos dados obtidos por *webcam,* é possível classificá-los e convertê-los na palavra correspondente descrita em áudio, concretizando o objetivo do grupo. Além do desenvolvimento do sistema em questão, o grupo pôde também comparar metodologias, principalmente com o trabalho utilizado como base, o que possibilitou um desenvolvimento considerável nos conhecimentos relacionados à sobre como trabalhar, de fato, com algoritmos de machine learning, principalmente quando se refere à maneira de moldá-lo: a liberdade de escolha e implementação é imensurável.

Quanto ao desempenho do algoritmo, como demonstrado anteriormente, foi satisfatório. No entanto, é importante ressaltar que o desempenho do sistema pode ser influenciado por diversos fatores, como a qualidade da câmera, condições de iluminação e variações individuais nos gestos de Libras.

Devido à isso, o grupo constata que o algoritmo empregado é apenas o ponto inicial de uma longa e necessária trajetória, pois é necessário expandir o *dataset,* as técnicas de deteção de objetos, o processamento de imagens e os parâmetros para melhor classificar imagens mais diversificadas, em vídeos de maior duração e com extração de maior quantidade de frames, no objetivo de tornar o sistema capaz de traduzir vídeos de longa duração e até mesmo chamadas de vídeo, em tempo real. Ao decorrer dessas contínuas melhorias, os algoritmos de machine learning surtirão um importante efeito nos quesitos de inclusão social, principalmente na realidade presente integralmente conectada e globalizada.

1. Conclusão

A conclusão deste artigo é que há uma necessidade de aumentar as pesquisas em machine learning aplicada à linguagem de sinais, especialmente na interação entre pessoas surdas e mudas. O tema ainda recebe pouca atenção na área, e isso é evidenciado pela análise dos trabalhos relacionados. Portanto, o objetivo do artigo é fomentar o aprendizado e a inclusão da Língua Brasileira de Sinais (Libras) por meio de um sistema que traduza os sinais em áudio, permitindo a interação dos usuários da Libras com outras pessoas e também com deficientes visuais.

O sistema desenvolvido utiliza uma webcam como entrada para capturar os sinais da Libras correspondentes às letras do alfabeto e a certas palavras. Esses sinais são processados e convertidos em áudio, que pode ser reproduzido pelo usuário. A Rede Neural Convolucional (CNN) foi escolhida como o modelo para treinamento, e a validação foi feita utilizando a classificação cruzada estratificada com 6 folds.

Os trabalhos relacionados mencionados no artigo abordam o reconhecimento de gestos em Libras e a tradução para texto ou áudio por meio de técnicas de inteligência artificial. Alguns utilizam redes neurais convolucionais (CNN), enquanto outros utilizam modelos de aprendizado de máquina, como o SVM (Support Vector Machine) e redes neurais recorrentes (RNN). Esses trabalhos demonstraram resultados promissores na comunicação em Libras, alcançando altas taxas de reconhecimento e facilitando a interação e a comunicação inclusiva.

Em suma, o artigo destaca a importância de desenvolver soluções baseadas em machine learning para aprimorar a interação entre pessoas surdas e mudas por meio da linguagem de sinais. A aplicação prática dessas tecnologias pode contribuir para a inclusão e acessibilidade desses indivíduos, superando as barreiras existentes no ambiente digital.

References

1. SIGN LANGUAGE RECOGNITION USING MACHINE LEARNING S.Saravana Kumar, Vedant L. Iyangar, 2018
2. Deepsign: Sign Language Detection and Recognition Using Deep Learning,Deep Kothadiya , 3 June 2022
3. Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning, Arpita Halder, Akshit Tayade.
4. DESIGN AND IMPLEMENTATION OF CNN FOR SIGN LANGUAGE RECOGNITION, Jilin Daxue Xuebao, November 2022.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Deafend: aplicativo para facilitação da comunicação com libras, Vieira, João Vitor Veronese, 2020.
7. HandArch: A deep learning architecture forLIBRAS hand configuration recognition, Gabriel Peixoto de Carvalho,Andr ́e Luiz Brandão,Fernando Teubl Ferreira .
8. PORFIRIO, A.; WIGGERS, K.; OLIVEIRA, L. E. S.; WEINGAERTNER, D. “LIBRAS Sign Language Hand Configuration Recognition Based on 3D Meshes”. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2013.
9. Audio to Sign Language Converter, Rishin Tiwari, Saloni Birthare, Mr. Mayank Lovanshi, 2022
10. COMPUTER VISION AND NEURALNETWORKS FOR LIBRAS RECOGNITION, Silas Luiz Furtado, Jauvane de Oliveira.
11. I. L. O. Bastos, M. F. Angelo and A. C. Loula, "Recognition of Static Gestures Applied to Brazilian Sign Language (Libras)," 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, 2015, pp. 305-312. DOI: 10.1109/SIBGRAPI.2015.26
12. I. L. O. Bastos, M. F. Angelo and A. C. Loula, "Recognition of Static Gestures Applied to Brazilian Sign Language (Libras)", LASIC, Disponível em: http://sites.ecomp.uefs.br/lasic/projetos/libras-dataset, Acessado em: 6 de junho de 2023.
13. Silva Pedro, I.C. & Suzuki, N.A. (2023). Sign-Language-Translate. Disponível em: https://github.com/NicolasSuzuki/Sing-Language-Translate.