



**XVI
ENFOC
2021**

XVI - Encontro de Iniciação Científica
e Fórum Científico, VII Seminário PIBID
- Programa Institucional de Bolsa de
Iniciação à Docência.

DESENVOLVIMENTO DE TECNOLOGIA INCLUSIVA PARA INTERFACE COM SURDOS

Ana Lícia Santos Braga 1, Janaina Santos da Silva 2, Marina Araújo Almeida 3, Murilo Muchalak de Camargo 4, Jessika Alvares Coppi Arruda Gayer 5, Charles Way Hun Fung 6.

1. Estudante do curso de Análise e Desenvolvimento de Sistemas do Centro Universitário Internacional UNINTER
2. Estudante do curso de Engenharia da Computação do Centro Universitário Internacional UNINTER
3. Estudante do curso de Licenciatura em Educação Especial do Centro Universitário Internacional UNINTER
4. Estudante do curso de Engenharia da Computação do Centro Universitário Internacional UNINTER
5. Professora da UNINTER - Orientadora
6. Professor da UNINTER - Orientador

Grupo de trabalho: ENGENHARIA, TECNOLOGIA E INOVAÇÃO

RESUMO

Este trabalho apresenta uma pesquisa para o desenvolvimento de uma interface para comunicação com surdos. Para tanto, busca-se investigar sobre a Língua Brasileira de Sinais - Libras, fazendo um levantamento de requisitos para o desenvolvimento da aplicação e utilizando técnicas de treinamento de inteligência artificial para que a interface desenvolvida atue como tradutor automático e simultâneo de Libras.

Palavras-chave: Tecnologia assistiva. Tradutor para surdos. Interface assistiva. Inteligência artificial. Tradutor simultâneo.

INTRODUÇÃO

O ser humano, enquanto ser social, utiliza a linguagem para se comunicar com seus pares, transmitindo culturas, valores, e diversos aspectos concernentes àquele grupo social (QUADROS, 2006). A língua falada, ou oral, é um dos tipos de linguagem mais difundidos no mundo, mas esse aspecto cerceia diversos grupos do direito à livre comunicação, como a comunidade surda.

Neste trabalho utiliza-se o termo surdo para caracterizar a pessoa com surdez (total ou parcial) que utiliza a Língua de Sinais Brasileira (Libras) como primeira língua para se comunicar nas situações cotidianas. O termo “não-surdo” é usado para caracterizar a pessoa não falante de Libras.

O problema com o qual se depara, dada a situação apresentada, é: como auxiliar os surdos a se comunicar com os não-surdos através de um recurso computacional?

Segundo o IBGE (2010), quase 10 milhões de pessoas no Brasil tem algum tipo de deficiência auditiva, o que pode dificultar o seu acesso a diversos direitos. De acordo com QUADROS (2006), é preciso que as duas línguas coexistam e ocupem seus lugares no cotidiano da pessoa surda.

Nesse sentido, este projeto pode servir de auxílio na comunicação de Surdos e Ouvintes, através do desenvolvimento de uma interface inclusiva para os comunicação por meio de um programa tradutor de Libras. Isso propiciará aos

PARCEROS



REALIZAÇÃO





indivíduos uma comunicação mútua, sendo útil tanto aos falantes da Libras quanto aos não falantes, através de uma tradução de uma biblioteca de vídeos com sinais correspondentes ao significado pretendido pelo usuário no momento da interação.

Com essa finalidade em mente, o projeto pretende: I - investigar sobre a Língua Brasileira de Sinais - Libras, conferindo as nuances que a divergem de uma língua oral e seus parâmetros; II - pesquisar as interfaces existentes para o usuário, criando um estado da arte sobre os recursos atuais; III - fazer o levantamento de requisitos do problema e gerar uma especificação para o desenvolvimento; IV - gerar banco de vídeos de treinamento, com a utilização de técnicas e recursos de Inteligência Artificial - IA; V - aplicar algoritmos de reconhecimento de padrões, leitura de vídeos e treinamentos de IA; e, por fim, VI - desenvolver a aplicação conforme especificações.

METODOLOGIA

A metodologia utilizada neste estudo foi dividida em três macro-etapas: aquisição de dados, treinamento e teste em ambiente real.

A aquisição de dados foi feita com vídeos gravados e com *WebCam* em tempo real. Utilizou-se a *MediaPipe*, uma das ferramentas de Inteligência Artificial (IA) do Google, de código aberto, que tem como base a biblioteca *TensorFlow*. Com a aplicação do recurso *Holistic*, foram mapeados os pontos de referência (*landmarks*) do rosto, pose (posição do corpo) e das mãos.

O processo extrai pontos de interesse a cada quadro dos vídeos. Para o treinamento, os pontos foram divididos em treinamento e teste, para o treinamento dos modelos e avaliação. Na etapa seguinte o modelo de IA treinado que apresenta melhor desempenho é utilizado no software em testes em tempo real. Com o auxílio da biblioteca *Numpy*, que usa vetores multidimensionais, os pontos obtidos são guardados em um arquivo separado por vírgulas, CSV, seguindo um modelo “x”, “y”, “z” (altura, largura e profundidade), e “v” de visibilidade, que é a probabilidade do ponto de referência estar ou não na imagem (ZHANG et al., 2020).

Na etapa de treinamento, foi adotada a biblioteca *Scikit-learn*, que disponibiliza algoritmos de aprendizado de máquina (Pedregosa et al. 2011). Neste estudo optou-se pelos seguintes algoritmos: *Logistic Regression*, *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)* e *MLP Classifier (Multi-layer perceptron)*.

Antes do treinamento, os pontos-chave coletados são divididos em em dois vetores (X e Y), um contendo as coordenadas e outro contendo as classes, que é o nome do sinal que será treinado. Sendo V a representação do vetor base, equação 1, com as coordenadas estimadas pelo *MediaPipe*. são extraídos 543 marcos (*MediaPipe Holistic*):

$$V_{543} = [x_j, y_j, z_j, v_j], 0 < j < 544; \quad (1)$$

X, equação 2, sendo o vetor que contém todos os vetores de coordenadas, onde n é o número de quadros que foram utilizados na etapa de aquisição de dados.



$$X_n = [V_n], (n = 1, 2, \dots, N) \quad (2)$$

E y é o vetor com os sinais respectivos de cada quadro, y [sinais].

Na sequência os vetores são divididos de forma aleatória, sendo parte destinada ao treinamento e parte para a avaliação. Neste estudo as amostras foram divididas em 80/20, ou seja, 80% das amostras coletadas são para treinamento e 20% para avaliação (teste de performance). Assim, realiza-se o treinamento dos sete modelos, usando os algoritmos supracitados e avaliando-se o desempenho de cada um sobre as amostras de teste. Com os resultados encontrados são montadas as matrizes de confusão e se compara os valores de acurácia dos modelos.

RESULTADOS E DISCUSSÃO

Para a análise de desempenho dos modelos foi realizado o treinamento dos algoritmos com as letras que formam o nome 'UNINTER', onde foi executada primeiramente a etapa de aquisição e armazenamento de dados, utilizando vídeos gravados pelos participantes da pesquisa, seguida pela etapa de treinamento e avaliação dos modelos classificadores. Sendo adotada a amostragem de 20% do total de classes para a avaliação, o total de amostras para esta etapa foi de 1751.

Os algoritmos treinados obtiveram os seguintes resultados para os percentuais de precisão e tempo gasto para o treinamento: *Logistic Regression* - 98,46%, 38.51 segundos, *Ridge Regression* - 98,52%, 1.09 segundos, *Random Forest* - 99,54%, 12.94 segundos, *Gradient Boosting* - 99,66%, 1360 segundos, *K-Nearest Neighbors (KNN)* - 97,03%, 0.35 segundos, *Support Vector Machine (SVM)* - 73,84%, 50.33 segundos e *MLP Classifier (Multi-layer perceptron)* - 89,21%, 109 segundos.

Analisando os valores obtidos realizados, foi possível perceber que a maioria dos métodos obtiveram desempenho acima de 97%, sendo o modelo *Gradient Boosting* o melhor com 99,66% de precisão, em contrapartida o método SVM (*Support Vector Machine*) que obteve o resultado mais baixo, 73,84%. Porém, quando considerado o tempo de treinamento, o algoritmo *Gradient Boosting* foi o que precisou de mais tempo para finalizar o treinamento e o modelo *K-Nearest Neighbors* foi o mais rápido levando apenas 0,35 segundos.

Além da análise da precisão dos modelos treinados, também foi criada a matriz de confusão de cada um dos algoritmos. A partir das mesmas foi possível observar quais letras tiveram maior ocorrência de troca por parte dos modelos e quantas foram as previsões falsas para cada letra.

A matriz que menos apresentou trocas foi a do modelo *Gradient Boosting*, onde três das seis letras apresentaram 100% de precisão. E apresentou apenas 6 inversões no total das previsões do sinal realizado.



Os modelos que mais apresentaram trocas, falsos positivos e falsos negativos, foram os modelos *Support Vector Machine (SVM)* e *MLP Classifier (Multi-layer perceptron)*. Este segundo apresentando dificuldade para prever os sinais das letras T, R e N, invertendo-as com a letra U. Sendo 28 inversões entrada T e saída como sendo a letra U, 32 onde o sinal correto era a letra N e 45 onde o sinal correto era a letra R. O primeiro modelo apresentou várias inversões entre todas as letras, na maioria dos casos apresentando como resultado a letra E. Sendo o sinal para a letra U o mais invertido, onde foi previsto 84 vezes como a letra E.

CONCLUSÕES

De acordo com os resultados obtidos é possível perceber que vários dos algoritmos utilizados na pesquisa foram capazes de alcançar valores expressivos para acurácia na previsão dos sinais das letras realizados em LIBRAS.

Apesar do modelo *Gradient Boosting* ser o que apresentou o melhor desempenho na etapa de avaliação, há outros modelos como o *K-Nearest Neighbors* que apresentaram bons resultados sem consumir muito recurso computacional, sendo possível o treinamento em um tempo menor quando comparado com os outros modelos.

Ainda foi possível perceber que os modelos *Support Vector Machine (SVM)* e *MLP Classifier (Multi-layer perceptron)* precisam ser aprimorados para utilização nessa aplicação. Pois apesar de apresentarem os dois piores desempenhos na etapa de testes dessa pesquisa, esses modelos podem ser implementados de diversas maneiras. E o resultado para um número maior de amostras de treinamento pode influenciar no desempenho de todos os modelos.

Outro ponto a ser aprimorado é a etapa de aquisição de dados. Para este trabalho foram utilizados vídeos de 2 participantes, formando um total de 8765 amostras. Porém, para pesquisas futuras, é sugerido a utilização de mais vídeos e com uma quantidade maior de participantes para a formação de um conjunto de treinamento mais heterogêneo, o que pode melhorar o desempenho final da inteligência artificial treinada.

Também é importante salientar que os resultados obtidos pelos modelos treinados nesse estudo, são altamente dependentes do desempenho da previsão da posição corporal da pessoa, no momento da aquisição dos pontos pela IA utilizada nessa etapa (*MediaPipe*). Ou seja, os resultados tanto no treinamento quanto em uma futura utilização em ambiente real, vão depender tanto do modelo treinado quanto do modelo utilizado para detecção da pessoa.

Por fim, outro ponto sugerido para pesquisas futuras é a avaliação do desempenho dos modelos, ou do modelo que apresentou melhor resultado, em ambiente real, realizando a exposição do modelo treinado a dados adquiridos em tempo real, aos quais ele não foi anteriormente exposto. Com isso será possível mensurar qual é o desempenho e utilidade deste no dia-a-dia das pessoas que se comunicam por meio da LIBRAS.



**XVI
ENFOC
2021**

XVI - Encontro de Iniciação Científica
e Fórum Científico, VII Seminário PIBID
- Programa Institucional de Bolsa de
Iniciação à Docência.

REFERÊNCIAS

IBGE. Instituto Brasileiro de Geografia Estatística. **Censo demográfico : 2010 : características gerais da população, religião e pessoas com deficiência.** Rio de Janeiro: IBGE, 2010. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf>. Acesso em 19 de agosto de 2021.

Mediapipe Holistic. Google LLC, 2020. Disponível em: <<https://google.github.io/mediapipe/solutions/holistic.html#mediapipe-holistic>>. Acesso em 16 de Agosto de 2021.

PEDREGOSA *et al.* **Scikit-learn: Machine Learning in Python.** JMLR 12, pp. 2825-2830, 2011. Disponível em <<http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>>. Acesso em 16 de Agosto de 2021.

QUADROS, Ronice Müller de. **Ideias para ensinar português para alunos surdos** / Ronice Muller Quadros, Magali L. P. Schmiedt. Brasília: MEC, SEESP, 2006.

ZHANG, Fan *et al.* **MediaPipe Hands: On-device Real-time Hand Trackin.** Califórnia, p. 1-5, 18 de jun. 2020.

PARCEIROS



REALIZAÇÃO

