

O INSTITUTO DE CIÊNCIA E
TECNOLOGIA APRESENTA



1º Workshop de Inteligência Artificial
do ICT-UNIFESP

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

Isabela Cristina Silva Pedro
Nicolas Alves Suzuki

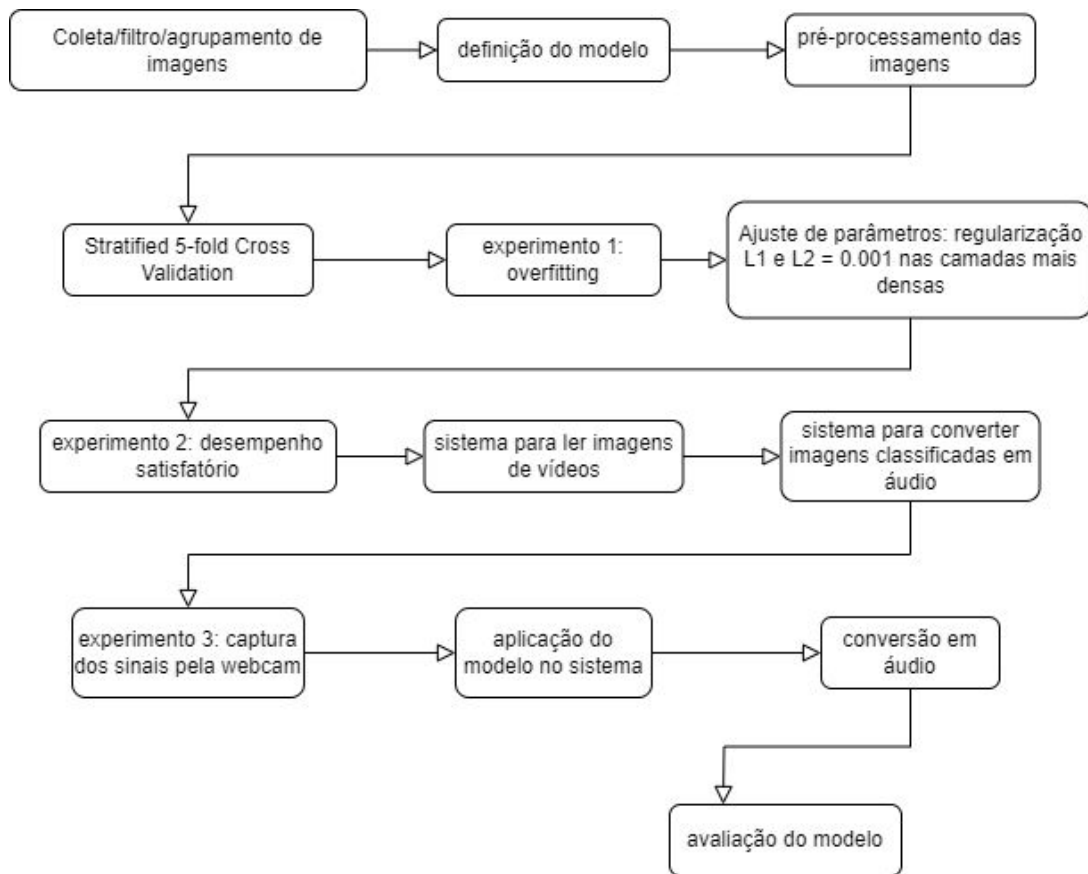


- Segundo a OMS, 1,5 bilhão de pessoas no mundo possuem algum grau de surdez
- Segundo o IBGE, 2,3 milhões de pessoas com algum grau de surdez no Brasil
- Sociedade globalmente conectada, a comunicação oral e escrita é facilitada pelos tradutores
 - Onde estão os tradutores para as linguagens de sinais?

- Desenvolver um sistema que receba uma transmissão em vídeo e seja capaz de transformar sinais da Libras em áudio



- Estimular os estudos da área aplicados à Libras
- Aumentar a inclusão e ampliar as formas de comunicação



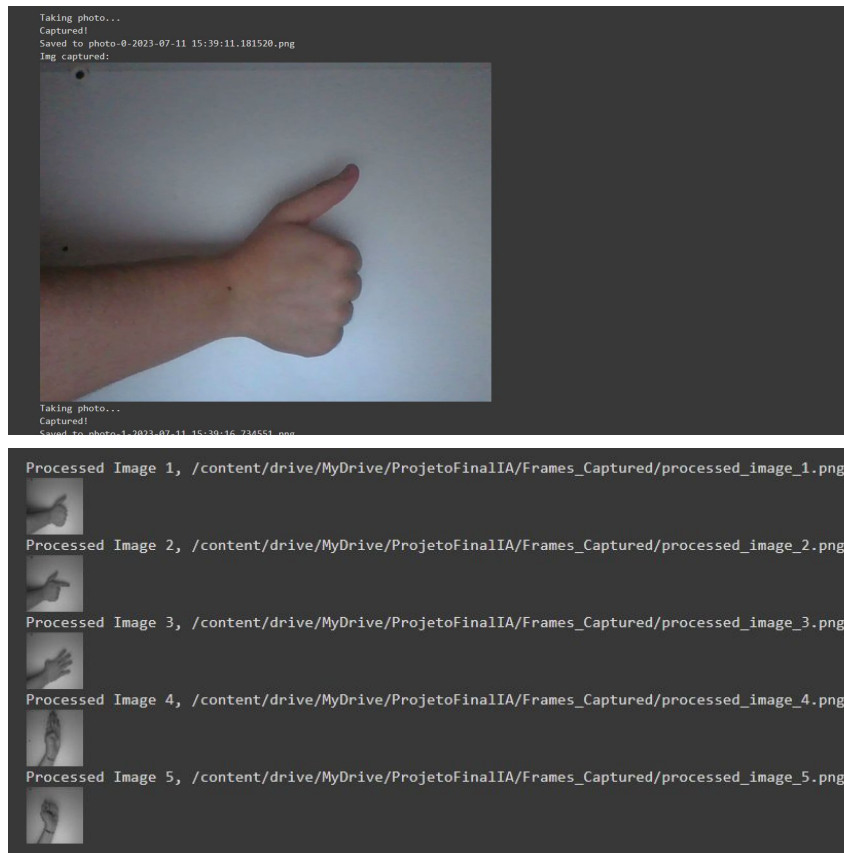
- Treino e validação com Stratified 5 Fold Cross Validation,
- Base de dados vinda de um artigo base,
- Utilização de CNN (Convolutional Neural Network)
 - 1ª e 2ª camada: 64 filtros 3x3,
 - função de ativação: ReLU,
 - otimizador: Adam,
 - regularização L1 e L2 = 0.001 nas camadas mais densas,
 - taxa de dropout: 0.25,
 - função de perda: categorical_crossentropy,
 - função de ativação final: Softmax
 - batch size: 32

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, <u>f1 Score</u> : 0.9464
Fold 2	Accuracy : 0.9832, Precision : 0.9839, <u>f1 Score</u> : 0.9831
Fold 3	Accuracy : 0.9731, Precision : 0.9768, <u>f1 Score</u> : 0.9728
Fold 4	Accuracy : 0.9899, Precision : 0.9906, <u>f1 Score</u> : 0.9898
Fold 5	Accuracy : 0.9865, Precision : 0.9882, <u>f1 Score</u> : 0.9862

- CNN: (98.77%) x MLP (96.77%)
- Captura e conversão realizadas
- Acurácia no experimento 2: 99.62%

- Coleta de 5 frames
- Converter para escala de cinza
- Redimensionar para 50x50 pixels
- Salvar a imagem processada



- Verifica se está no padrão
- Classificação
- Google Text-to-Speech(gTTS)

```
1/1 [=====] - 0s 149ms/step  
Phrase: 1  
1/1 [=====] - 0s 37ms/step  
Phrase: 1 2  
1/1 [=====] - 0s 38ms/step  
Phrase: 1 2 4  
1/1 [=====] - 0s 31ms/step  
Phrase: 1 2 4 E  
1/1 [=====] - 0s 31ms/step  
Phrase: 1 2 4 E E
```

0:00 / 0:02

Duration of audio: 2.904 seconds

- O sistema idealizado foi iniciado, mas é só o começo:
 - Pouca disponibilidade de base de dados de sinais da Libras
 - Trabalhos relacionados abrangem em sua maioria especialmente a ASL (American Sign Language)
- O modelo apresentado demonstrou desempenho superior ao artigo utilizado como base durante a avaliação
- No futuro:
 - aumentar a base de dados de treino e teste,
 - desenvolver melhor o algoritmo para a captação dos frames
 - classificar melhor sinais estáticos e em movimento

- Nossos agradecimentos:
 - Ao Professor Fábio Augusto Faria;
 - À Sara Bueno da Silva;
 - Aos professores presentes na banca;
 - À todos os presentes nessa sessão.