



DEEP LEARNING PARA RECONHECIMENTO DE SINAIS DA LIBRAS COMO TECNOLOGIA ASSISTIVA

Samuel França da Costa Pedrosa

Projeto Final de Curso para Engenharia de Computação

EMC

ESCOLA DE ENGENHARIA ELÉTRICA,
MECÂNICA E DE COMPUTAÇÃO



DEEP LEARNING PARA RECONHECIMENTO DE SINAIS DA LIBRAS COMO TECNOLOGIA ASSISTIVA

Samuel França da Costa Pedrosa

Orientador: Prof. Dr. Adriano César Santana

Goiânia, 2024

INTRODUÇÃO

Apesar de ofuscada, nacionalmente a Língua Brasileira de Sinais (LIBRAS) é o principal meio de comunicação entre surdos e é a segunda língua oficial do país.

Devido à marginalização social de surdos, busca-se tecnologias assistivas para reduzir as barreiras comunicativas entre línguas orais-auditivas e gestuais.

Os recentes avanços em **Deep Learning** aplicados ao Processamento de Linguagem Natural (NLP) englobam principalmente linguagens escritas ou faladas, mas destaca-se, também, suas possíveis aplicações para línguas de sinais como a LIBRAS.

No Brasil, empresas como a **Hand Talk** e **Lenovo Brasil** pioneirizam no desenvolvimento de ferramentas tradutoras de LIBRAS.

**“Os limites da minha linguagem
são os limites do meu mundo”**

— Ludwig Wittgenstein

FILÓSOFO DA LINGUAGEM



Fonte: Imagem de Autoria Própria

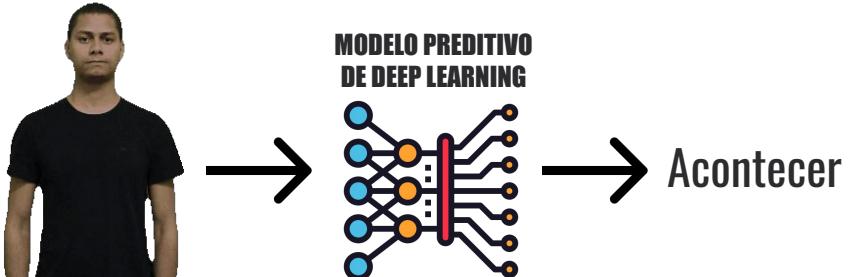
OBJETIVO

Objetivo Geral

Utilizar abordagens modernas para analisar o desempenho em reconhecimento de sinais e contribuir com desenvolvimento de tecnologias assistivas tradutoras de LIBRAS.

Áreas de conhecimento envolvidas:

- Ciência de Dados
- Visão Computacional
- Redes Neurais Artificiais Profundas
- Modelagem Preditiva
- Processamento Digital de Imagens e Sinais
- Aprendizado de Máquina



Fonte: Imagem de Autoria Própria

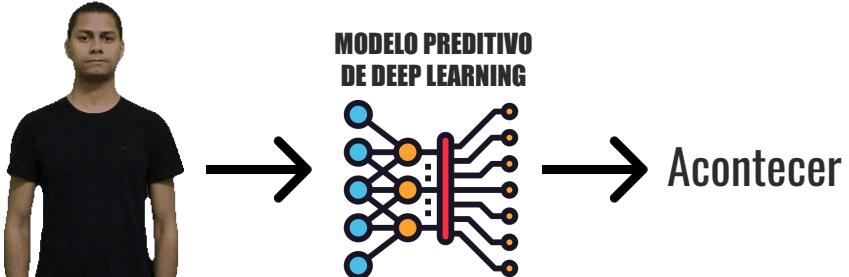
OBJETIVO

Objetivo Específico

Utilizar modelos preditivos classificatórios de **deep learning** (aprendizado profundo) para treinamento em uma **base de dados** contendo vídeos de pessoas gesticulando sinais em LIBRAS com intuito de analisar seus respectivos desempenhos em reconhecimento de sinais.

Etapas:

1. **Formação de uma base de dados** pública em conformidade com a LGPD.
2. **Extração de pontos-chave** corporais a partir de modelos pré-treinados.
3. **Tratamento dos dados** para uma formatação apropriada como entrada dos modelos.
4. Desenvolvimento e modelagem da aplicação em **arquiteturas de redes neurais profundas**.
5. **Avaliação e comparação das métricas** de desempenho dos modelos aplicados.
6. Discussão do impacto do modelo proposto como contribuição para **tecnologias assistivas**.



Fonte: Imagem de Autoria Própria

FUNDAMENTAÇÃO TEÓRICA

LIBRAS

Linguagem visuo-espacial de movimentos expressivos e sequenciais das **mãos e articuladores**.

Assim como qualquer língua, possui estrutura léxica, sintaxe e gramática própria, sendo cada sinal parametrizado pela **configuração, posição, orientação, movimento e entonação** da ação dos gestos.

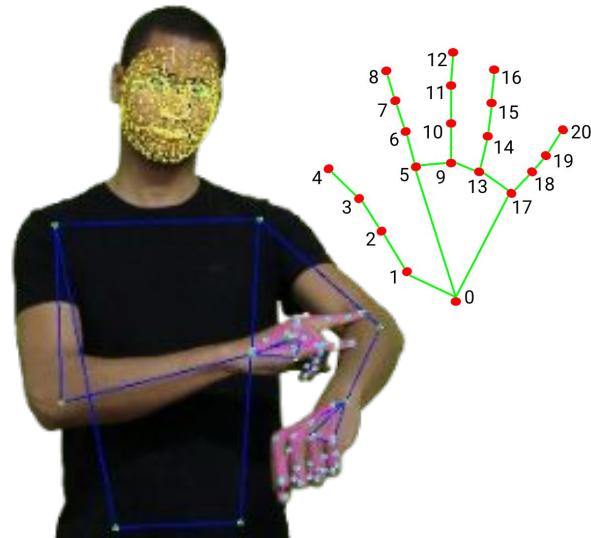
Landmarks

Conceito técnico de **Aprendizado de Modelos de Ação e Detecção Corporal por Visão Computacional**.

Os landmarks ou pontos de referência são **posições-chaves específicas do corpo humano**.

Primordial para **rastreamento do movimento** em vídeos.

Processamento computacional de landmarks é significantemente menor do que de imagens.



Fonte: Imagem de Autoria Própria

FUNDAMENTAÇÃO TEÓRICA

Modelagem Preditiva por Aprendizado de Máquina

Um **modelo preditivo** gera previsões com base nos aprendizados das características e padrões extraídos de um conjunto de dados durante sua etapa de treinamento.

A previsão ou característica aprendida é a saída (**output**) gerada pelos dados de entrada (**input**).

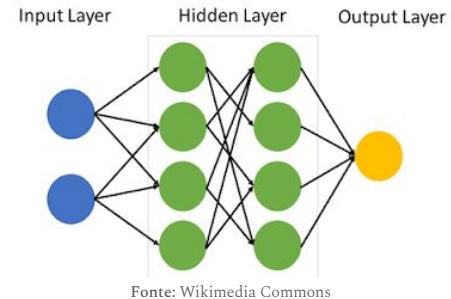
O aprendizado de máquina ou machine learning pode ser considerado:

- **Supervisionado:** Treinado com input rotulado, seja classificatório ou regressivo.
- **Não Supervisionado:** Treinamento com input sem rótulos, seja por agrupamento ou associação.

Redes Neurais Artificiais

É um modelo preditivo inspirado no aprendizado do cérebro humano.

As características do input passam diversas vezes por **camadas de neurônios artificiais de entrada, intermediárias e de saída que ajustam os seus pesos** entre as conexões, gerando uma percepção de padrões como aprendizado.



Há diversos tipos para aplicações em inputs variados, como as **Diretas (FFNN)** e **Recorrentes (RNN)**.

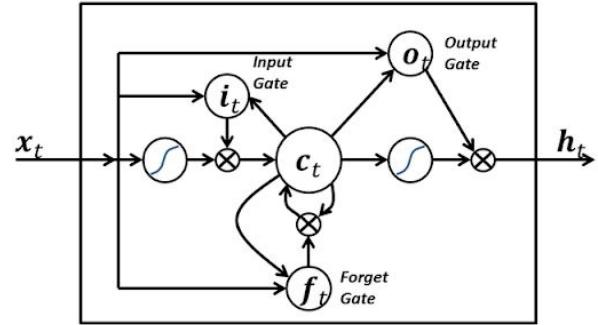
Quando existem muitas camadas intermediárias, denomina-se **Deep Learning** (Aprendizado Profundo), aumentando sua complexidade de aprendizado.

FUNDAMENTAÇÃO TEÓRICA - DEEP LEARNING

O Long Short-Term Memory (Memória Longa de Curto Prazo) é uma arquitetura de Rede Neural Recorrente (RNN).

Desenvolvida no final do século XX com intuito de minimizar, apesar de persistir, as limitações das RNNs relacionadas ao **Problema de Dissipação dos Gradientes**.

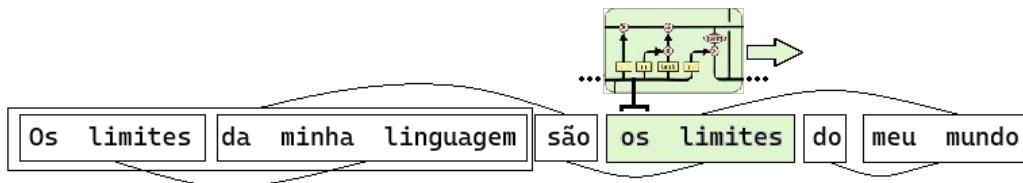
Além das de entrada e saída, utiliza **portas** de esquecimento para descartar informações que perdem importância.



Fonte: Wikimedia Commons

De forma simplificada, o LSTM é uma variação avançada de RNN que aprende, a partir de um input sequencial, **quando considerar ou desconsiderar características recorrentes** para as próximas sequências.

Além de séries temporais condicionais, é muito aplicado, devido às **estruturas sintáticas serem sequências recorrentes**, em **Processamento de Linguagem Natural (NLP)**. Cenários, esses, condizentes com o projeto.



Fonte: Imagem de Autoria Própria

FUNDAMENTAÇÃO TEÓRICA - DEEP LEARNING TRANSFORMERS

É uma arquitetura de deep learning desenvolvida pela Google no artigo “Attention is All You Need (2017)”.

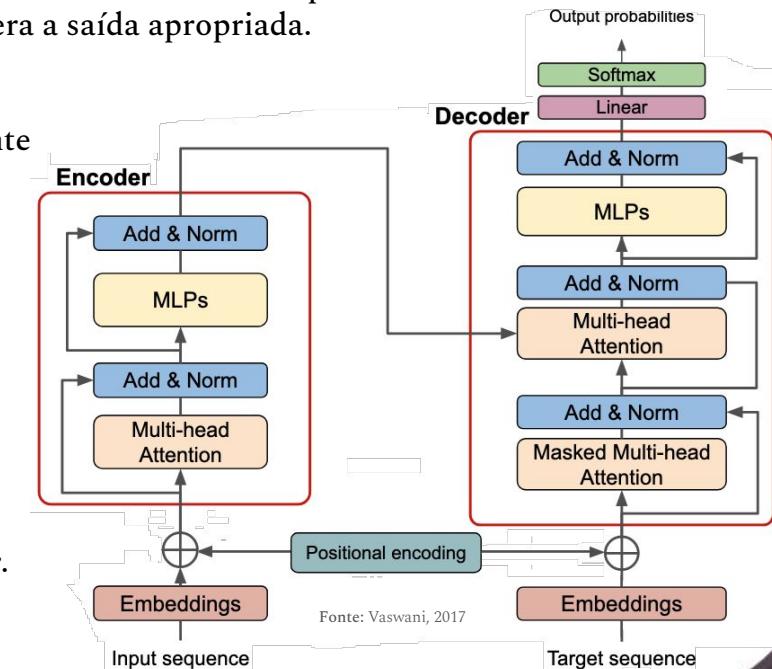
Sua estrutura de Encoder-Decoder revolucionou o campo das ferramentas de **tradução** e **IAs generativas**.

- **Encoder:** Recebe a entrada para extrair e codificar seu contexto e dependências.
- **Decoder:** Recebe a codificação do encoder e gera a saída apropriada.

Apesar de primordialmente ser projetada para NLP, suas **áreas de aplicação já são diversas**, principalmente as que envolvem dados sequenciais.

Grande vantagem em relação às RNNs por usar **mecanismos de atenção (self-attention)** em vez de unidades recorrentes. Ou seja, relaciona todos os elementos sequenciais entre si identificando padrões e dependências de forma independente.

Consequência de uma estrutura mais complexa e elevada eficácia, comparado com as RNNs, possui **custo computacional quadraticamente maior**.



Fonte: Vaswani, 2017

METODOLOGIA

Formação da Base de Dados

Etapa desafiadora, pois, ao contrário das linguagens orais e escritas, **há pouco volume de dados disponíveis abertamente e de fácil coleta** para formação de uma base de dados robusta. As poucas bases de dados contendo vídeos de pessoas sinalizando ainda não condizem com a magnitude e complexidade que essa língua possui.

A base de dados **MINDS-LIBRAS** (IFMG, 2020) se destaca pela qualidade dos vídeos como coleção pública de LIBRAS. Possui 1200 vídeos gravados em 1080p com **12 sinalizadores** gravando, cada um, 5 amostras de **20 sinais**.

Para enriquecimento da base do MINDS-LIBRAS, realiza-se gravações de autoria própria acrescentando **mais dois sinalizadores** gravando, cada um, **50 amostras** dos mesmos 20 sinais, resultando em mais 2000 vídeos.

A formação da base de dados utilizada consiste nos vídeos do MINDS-LIBRAS com o seu enriquecimento de autoria própria, **totalizando 64 GB em 3200 vídeos** com 30 FPS, tendo **192 mil imagens** a serem processadas.

Frames do sinal “Amarelo” (cima) e “Vontade” (baixo)



Fonte: Imagem de Autoria Própria

METODOLOGIA

Mapeamento dos Landmarks

Utilização da biblioteca aberta da Google **MediaPipe Holistic** que disponibiliza métodos de rastreamento de landmarks das **mãos, postura e rosto integrados simultaneamente** na mesma detecção.

Implementação de um algoritmo de mapeamento dos landmarks que, além do MediaPipe, utiliza bibliotecas do **Python**, como **OpenCV** para visão computacional e **Pandas** para manipulação de dados.

Função **recebe um vídeo** como parâmetro, mapeia **86 landmarks** úteis e **retorna o mapa** como estrutura tabular. Simplificando 64 GB de vídeos em 700 MB (**99% de redução**).

O formato dos dados do mapa representa, para cada landmark, uma função espacial-temporal discreta, denominada **Ação Sígnica**. $\text{Landmark} \equiv F(x(t), y(t))$

Tratamento e Alinhamento dos Dados

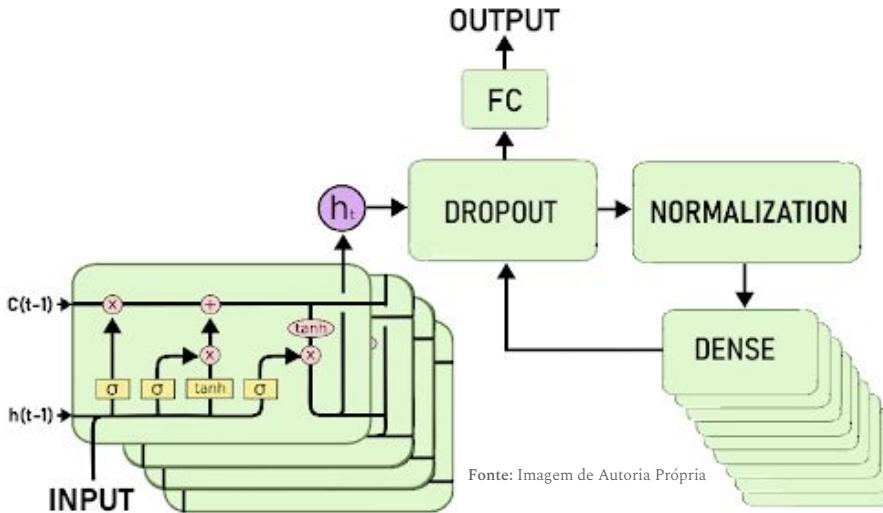
Como o tempo dos vídeos não é o mesmo, os mapas precisam passar por algum **método de alinhamento** para devida formatação padronizada do input:

- **Interpolação:** Aproximação linear dos valores pela média dos mapas.
- **Padding:** Ajusta os mapas para o maior tamanho com preenchimento de zeros.
- **DTW:** Dynamic Time Warping deforma partes sequenciais que melhor se alinham pela média.

METODOLOGIA - MODELAGEM

Arquitetura do Modelo usando LSTM

1. **Forward LSTM Layer:** Camada direta de LSTM, sua pluralidade gera deep learning.
2. **First Dropout Layer:** Primeira camada de desativação.
3. **Normalization Layer:** Camada que normaliza a saída do LSTM.
4. **Dense Layer:** Camada de redimensionamento.
5. **Second Dropout Layer:** Segunda camada de desativação.
6. **Fully Connected Layer:** Última camada para redimensionar a saída.



Fonte: Imagem de Autoria Própria

METODOLOGIA - MODELAGEM

Arquitetura do Modelo usando Transformers

1. Embedding:

Camada de incorporação.

1. Positional Embedding:

Camada de aprendizado posicional.

1. Normalization Layer:

Camada de normalização pro encoder.

1. Encoder:

Principal camada de aprendizado.

1. First Dropout Layer:

Camada de desativação.

1. Feedforward Attention:

Camada direta com multi-head attention.

1. Second Dropout Layer:

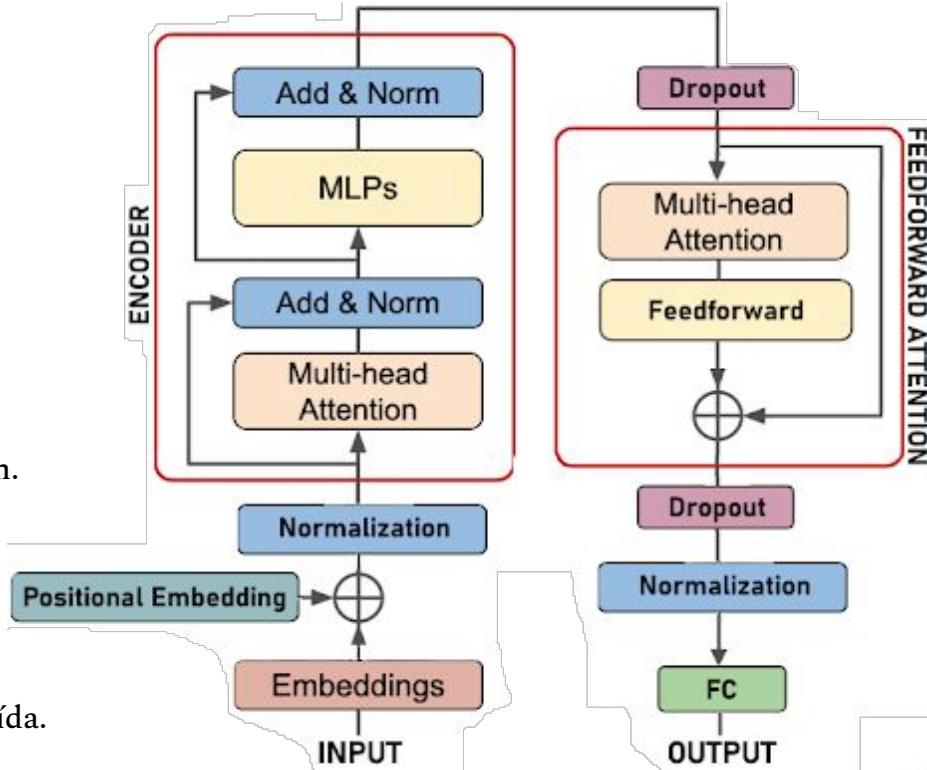
Segunda camada de desativação.

1. Normalization Layer:

Camada de normalização do attention.

1. Fully Connected Layer:

Última camada para redimensionar a saída.



Fonte: Imagem de Autoria Própria

METODOLOGIA

Hiperparâmetros para Treinamento

HIPERPARÂMETRO	VALOR	MODELO
Batches	16	Ambos
Epochs	100	Ambos
Dropout Rate	0.1 (10%)	Ambos
Hidden Size	256	LSTM
Recurrent Deep Layers	4	LSTM
Encoder Deep Layers	8	Transformer
Heads	86	Transformer
Feedforward Dimension	2752	Transformer

Métricas para Avaliação

Como forma de avaliação, separa **80%** dos dados para treino e **20%** para validação, então utiliza-se as métricas:

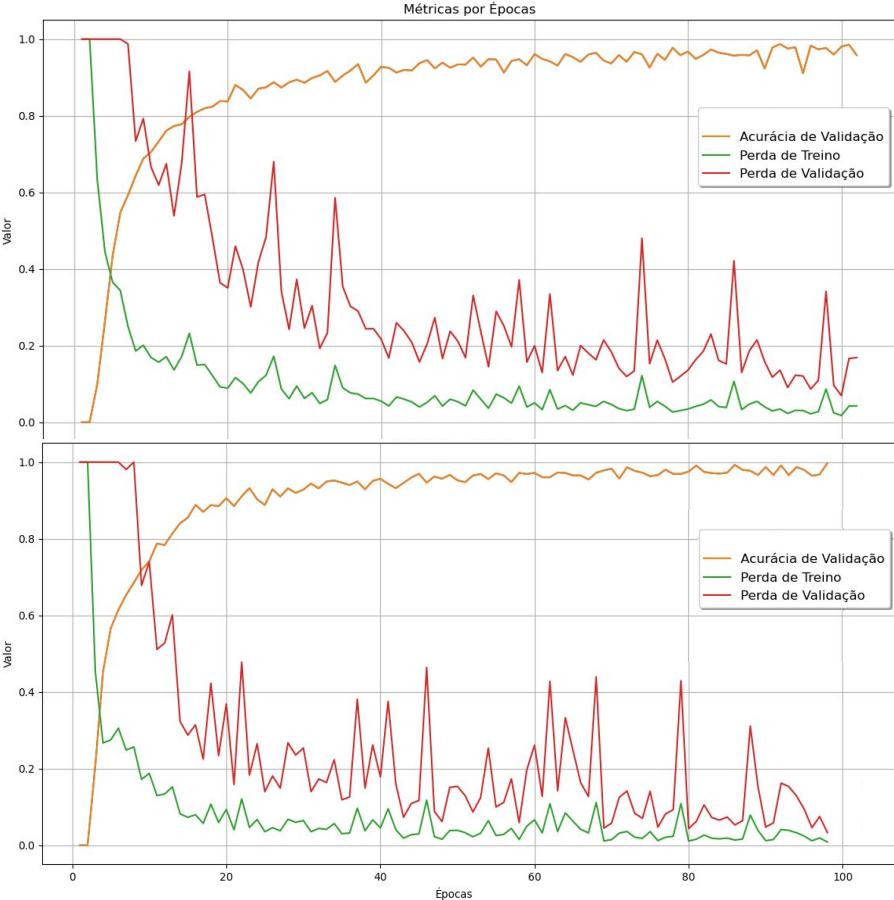
- **Acurácia de Treino**
- **Função de Perda de Treino**
- **Acurácia de Validação**
- **Função de Perda de Validação**

Avaliação visual por **gráficos** das métricas e pela **Matriz de Confusão**.

Tempo de processamento (Intel i7 9°) para os dois modelos, cada um usando os três métodos de alinhamento:
Interpolação, Padding e DTW.

RESULTADOS - INTERPOLAÇÃO

LSTM



Matriz de Confusão

O gráfico mostra a matriz de confusão para o modelo Transformer, com Sinal Real na vertical e Sinal Previsto na horizontal. Os valores nas células representam a contagem de acertos.

Sinal Real \ Sinal Previsto	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	0	32	0	5	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	1	31	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0
Vontade	0	1	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26

Matriz de Confusão (Segunda)

O gráfico mostra a matriz de confusão para o modelo Transformer, com Sinal Real na vertical e Sinal Previsto na horizontal. Os valores nas células representam a contagem de acertos.

Sinal Real \ Sinal Previsto	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0
Vontade	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0

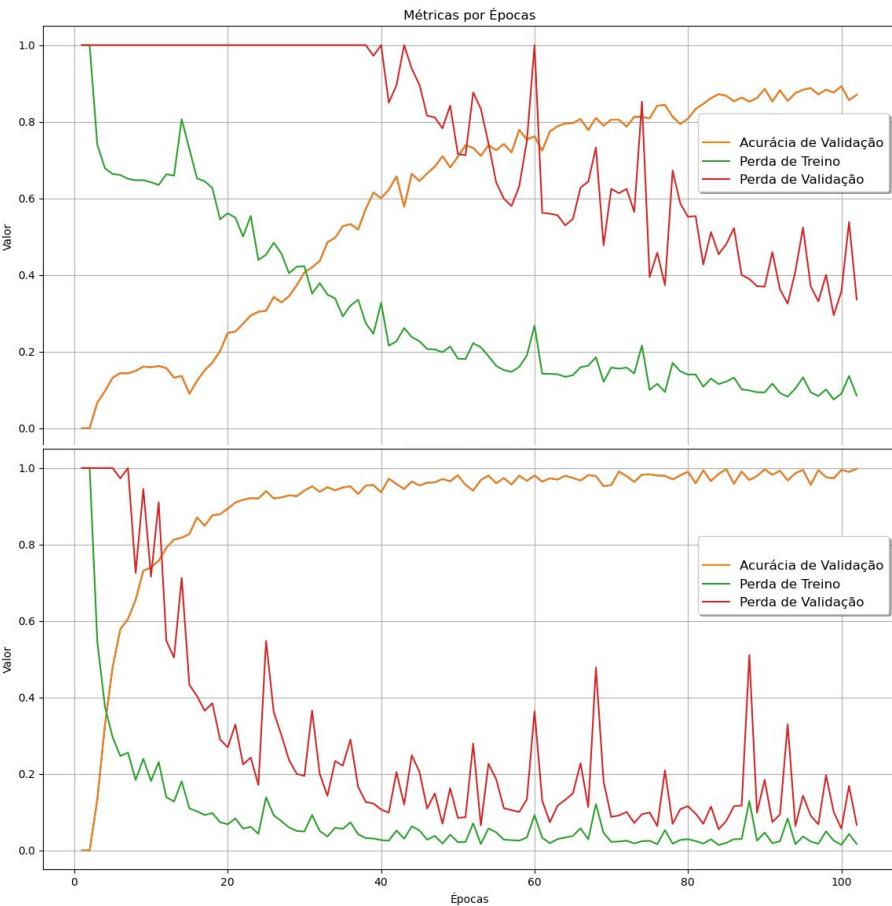
Matriz de Confusão (Terceira)

O gráfico mostra a matriz de confusão para o modelo Transformer, com Sinal Real na vertical e Sinal Previsto na horizontal. Os valores nas células representam a contagem de acertos.

Sinal Real \ Sinal Previsto	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0
Vontade	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0

RESULTADOS - PADDING

LSTM
TRANSFORMER



Matriz de Confusão

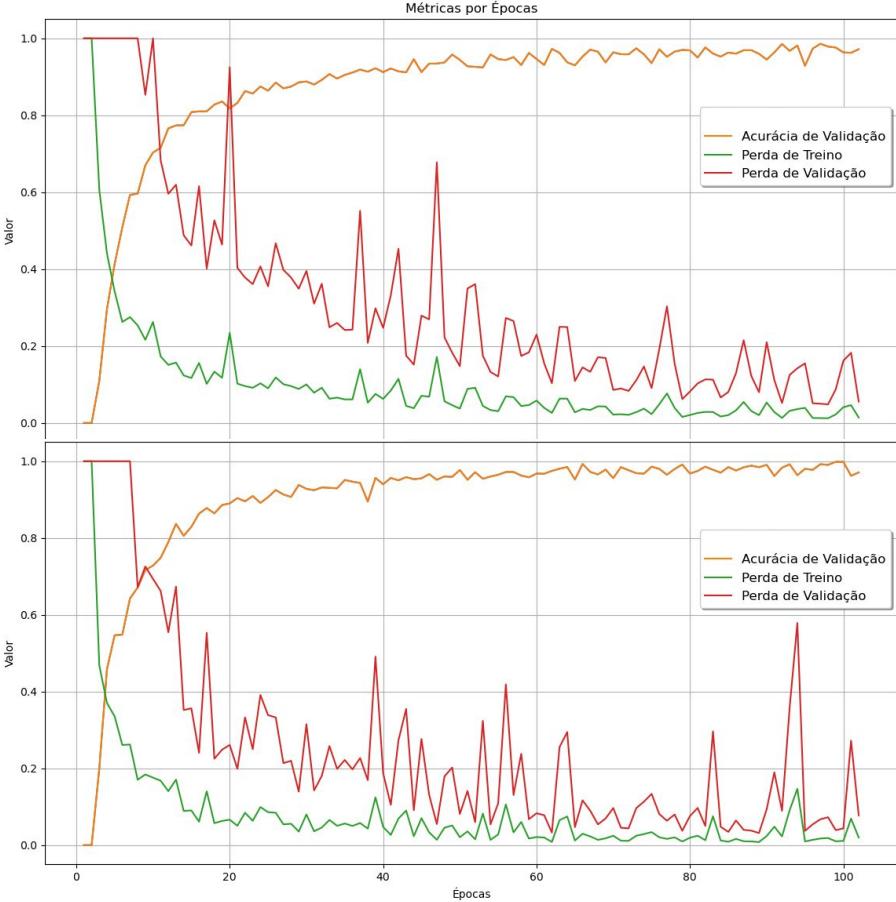
As matrizes de confusão mostram a taxa de acerto das classificações para cada sinal, divididas entre Sinal Real (vertical) e Sinal Previsto (horizontal). As cores das células representam a taxa de acerto, com escala de 0 a 40%.

Sinal Real	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	24	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	25	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	25	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	19	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	3	1	25	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	1	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	1	1	0	30	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0
Maca	0	1	0	0	0	0	4	0	32	0	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	1	0	2	0	0	1	0	24	1	0	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	1	0	0	0	25	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0
Vontade	0	1	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	1	0	26	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0

Sinal Real	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0	0
Vontade	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0

RESULTADOS - DTW

LSTM



Matriz de Confusão

O gráfico mostra a matriz de confusão para o modelo Transformer, com as classes de sinal real na vertical e as classes de sinal previsto na horizontal. As contagens de acerto estão escritas nas células da diagonal principal.

Sinal Real \ Sinal Previsto	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	31	0	1	0	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0
Vontade	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0

O gráfico mostra a matriz de confusão para o modelo Transformer, com as classes de sinal real na vertical e as classes de sinal previsto na horizontal. As contagens de acerto estão escritas nas células da diagonal principal.

Sinal Real \ Sinal Previsto	Aluno	Amarelo	Aproveitar	Bala	Banco	Barulho	Cinco	Conhecer	Espelho	Filho	Maca	Medo	Ruim	Vacina	Vontade	Acontecer	America	Banheiro	Esquina	Sapo
Aluno	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amarelo	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aproveitar	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bala	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banco	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barulho	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cinco	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Conhecer	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0
Espelho	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
Filho	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0
Maca	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0
Medo	0	0	0	0	0	0	0	0	0	0	32	0	1	0	0	0	0	0	0	0
Ruim	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0
Vacina	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0
Vontade	0	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0
Acontecer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0
America	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0
Banheiro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0
Esquina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0
Sapo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0

RESULTADOS

MODELO COM ALINHAMENTO	ACURÁCIA TREINO	PERDA TREINO	ACURÁCIA VALIDAÇÃO	PERDA VALIDAÇÃO	TEMPO TOTAL
LSTM com Interpolação	0.9806	0.0176	0.9850	0.0695	27 min
LSTM com Padding	0.8762	0.0748	0.8960	0.2952	38 min
LSTM com DTW	0.9786	0.0123	0.9861	0.0483	25 min
Transformer com Interpolação	0.9968	0.0035	0.9965	0.0138	210 min
Transformer com Padding	0.9945	0.0172	0.9861	0.0680	505 min
Transformer com DTW	0.9986	0.0099	0.9896	0.0391	207 min

CONCLUSÃO

Resultados precisos para os dois modelos com diferenças esperadas pela teoria.

Alinhamento com **Interpolação** apresentou melhor desempenho, devido ao alto custo computacional do tratamento DTW e resultados inferiores do Padding.

Apesar do custo-benefício do **LSTM**, é possível que sua **eficácia diminua com o aumento da escala da base de dados e quantidade de sinais**.

Transformers se apresentam como melhor recurso para uma **LLM de LIBRAS**.

A dificuldade no desenvolvimento de uma tecnologia assistiva tradutora para LIBRAS não é a falta de modelos mais avançados, mas as **dificuldades na formação de uma grande base de dados**.

Essa base, com **investimentos e colaborações**, pode ser criada para promover a primeira LLM de LIBRAS capaz de intermediar a comunicação entre surdos e ouvintes.



Fonte: Imagem de Autoria Própria

REFERÊNCIAS BIBLIOGRÁFICAS

VASWANI et al. *Attention is All You Need*. In: *Advances in Neural Information Processing Systems*, 2017.

Disponível em: arxiv.org/abs/1706.03762. Acesso em: 25 nov. 2024.

HOCHREITER, Sepp. *Investigações sobre redes neurais dinâmicas*. – Universidade Técnica de Munique. 1991.

Disponível em: people.idsia.ch/~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf. Acesso em: 25 nov. 2024.

ALMEIDA et al. MINDS-LIBRAS Dataset: A Dataset for Automatic Recognition of Brazilian Sign Language. 2020.

Disponível em: <https://zenodo.org/record/4322984>. Acesso em: 25 nov. 2024.

LECUN et al. *Deep learning*. *Nature*, 2015.

Disponível em: nature.com/articles/nature14539. Acesso em: 25 nov. 2024.

IOFFE et al. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015.

Disponível em: arxiv.org/pdf/1503.04069.pdf. Acesso em: 25 nov. 2024.

CHOLETT et al. *Deep Learning with Python*. Shelter Island: Manning, 2018.

Disponível em: manning.com/books/deep-learning-with-python. Acesso em: 25 nov. 2024.

HARRISON. *Machine Learning: Guia de Referência Rápida*. São Paulo: Novatec Editora, 2019.

Todo o projeto foi desenvolvido por Samuel Pedrosa em 2024. Os códigos utilizados estão disponíveis abertamente em um repositório na plataforma GitHub, visando contribuir para futuras pesquisas e tecnologias assistivas.

