UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS - CAMPUS BAURU
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ARISSA YOSHIDA

AUXÍLIO AO DIAGNÓSTICO DE DOENÇA NEURODEGENERATIVA UTILIZANDO EXPRESSÕES FACIAIS

BAURU
Novembro/2023

ARISSA YOSHIDA

AUXÍLIO AO DIAGNÓSTICO DE DOENÇA NEURODEGENERATIVA UTILIZANDO EXPRESSÕES FACIAIS

Trabalho de Conclusão de Curso do Curso de Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Campus Bauru.
Orientador: Prof. Dr. João Paulo Papa
Coorientador: Guilherme Camargo de Oliveira

BAURU
Novembro/2023

Arissa Yoshida

Auxílio ao diagnóstico de doença neurodegenerativa utilizando expressões faciais

Trabalho de Conclusão de Curso do Curso de Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Campus Bauru.

Banca Examinadora

Prof. Dr. João Paulo Papa
Orientador
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Ciência da Computação

Profa. Dra. Simone das Graças Domingues Prado
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Ciência da Computação

Dr. Mateus Roder
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências

Bauru, 14 de Novembro de 2023.

Dedico esse trabalho à minha família e meus amigos, pois sem eles nada na minha vida seria possível.

Agradecimentos

Desejo expressar minha profunda gratidão a todas as pessoas que desempenharam um papel fundamental em minha jornada acadêmica e pessoal até o presente momento. Primeiramente, quero agradece a minha família, e em especial, da minha mãe, do meu pai, da minha vó e dos meus queridos irmãos Kenzo e Maki. Vocês representam um dos principais motivos que me impulsionam a sempre buscar o meu melhor. Também não posso deixar de mencionar meus animais de estimação, em particular, Loris, Chibi e meu maior reprodutor de assobios, que sempre estiveram lá trazendo alegria, conforto e leveza na minha vida. Espero sinceramente que, com meu contínuo esforço e dedicação, possa retribuir a confiança e o amor que todos vocês sempre depositaram em mim. Amo profundamente cada um de vocês e sou imensamente grata por tudo o que fizeram por mim.

A minha jornada na graduação foi incrível graças a muitos indivíduos. Aos veteranos, aos amigos de classe, amigos de longa data, e aos vários amigos que fiz pelo caminho e que influenciaram diretamente a pessoa em que me tornei, não vou conseguir citar todos, mas quero deixar aqui registrados alguns nomes: Flangodile, Livão, Ana Crara, Toad e BiO; Akimi e Diego; Salocin, Leo Anjos, Leo Hardt, Isadora, Lucas Radaelli, Decmou, Manu, Kainhu, Mavi, Tranquila, Rafinha, Iza Marry, Kirua, Natan e Jovem; Davizaum, Xilsu, Giovani, Luize, Joãozinho, Jodas, Gui, Tusco e Linão; Arthur, Renatinho, Marry, Nih, Modscleo4, Gladiador, Mamado, Kszinhu, CtrlSt4rk, Sr. Z, Mathews16_XD, eu2012, Nick146 e PCaliman. Obrigado por fazerem parte dessa etapa tão especial e por serem simplesmente vocês, trazendo alegria a tantos momentos importantes para mim.

Gostaria de agradecer aos grupos dos quais fiz parte: Ao Laboratório Recogna pelo apoio e conhecimento compartilhado, ao time do Protiva pelas inesquecíveis experiências na computação e à minha equipe de Futebol Feminino da UNESP Bauru por todo o acolhimento e jogos inesquecíveis (*"Sou Bauru 14 vezes campeão!"*).

Um agradecimento especial a João Paulo Papa, Mateus Roder, Pedro Henrique Paiola e Wilson Massashiro Yonezawa, cujas orientações e amizade foram fundamentais para o meu desenvolvimento como estudante pela UNESP-Bauru.

E por fim, mas não menos importante, agradeço a todos os membros da banca, cujo apoio e avaliação foram essenciais para tornar este trabalho possível.

A todos vocês, o meu mais profundo agradecimento. Cada um de vocês desempenhou um papel importante em minha jornada, e eu serei eternamente grata por isso.

間
(MIYAZAKI, 1992)

Resumo

O diagnóstico precoce da Esclerose Lateral Amiotrófica (ELA) é fundamental para a determinação do início dos tratamentos, auxiliando tanto no aumento da expectativa de vida quanto na melhora da qualidade de seus pacientes. A análise de movimentos faciais fornece informações importantes para o reconhecimento dos sintomas iniciais da ELA; entretanto, detectar esses sinais não é uma tarefa fácil. Com o advento da visão computacional e dos modelos de aprendizado de máquina, métodos computacionais de auxílio a diagnóstico de doenças neurodegenerativas por meio de vídeos vêm se tornando uma realidade, gerando diferentes abordagens para detecção dos biomarcadores da ELA. Por utilizarem métodos interpretativos, grande parte dos estudos acabam por não explorar a dimensão temporal durante o processo de classificação. Este trabalho propõe introduzir modelos sequenciais de Redes Neurais Recorrentes (*Recurrent Neural Network* - RNN) em dados sequenciais (vídeos) de tal forma que seja investigada a relevância da dinâmica temporal das unidades de ação (*Action Units* - AUs) na identificação da ELA. Concluindo com o desenvolvimento de uma ferramenta de auxilio ao diagnóstico por computador (computer-aided diagnosis, CAD).

Palavras-chave: Esclerose Lateral Amiotrófica; Redes Neurais Recorrentes; Unidades de Ação; CAD.

Abstract

Early diagnosis of Amyotrophic Lateral Sclerosis (ALS) is crucial for determining the onset of treatments, aiding in increasing life expectancy and improving the quality of life for patients. The analysis of facial movements provides valuable information for recognizing the initial symptoms of ALS; however, detecting these signs is a challenging task. With the advent of computer vision and machine learning models, computational methods for assisting in diagnosing neurodegenerative diseases through videos are becoming a reality, generating different approaches to detect ALS biomarkers. Since many of these studies use interpretive methods, a significant portion needs to explore the temporal dimension during the classification process. This work proposes the introduction of sequential models of Recurrent Neural Networks (RNN) in sequential data (videos) in such a way that the relevance of the temporal dynamics of Action Units (AUs) in identifying ALS is investigated. It concludes with the development of a computer-aided diagnosis tool (CAD).

Keywords: Amyotrophic Lateral Sclerosis; Recurrent Neural Networks; Action Units; CAD.

Lista de figuras

Lista de quadros

Lista de abreviaturas e siglas

ADALINE	*Adaptive Linear Neuron*
ALS	*Amyotrophic Lateral Sclerosis*
AU	*Action Units*
CAD	*computer-aided diagnosis*
CNN	*Rede Neural Convolucional*
DN	Doença Neurodegenerativa
ELA	Esclerose Lateral Amiotrófica
FACS	*Facial Action Coding System*
GRU	*Gatet Recurrent Unit*
HC	*Healthy Control*
HOG	*Histogram of Oriented Gradient*
IA	*Inteligência Artificial*
KNN	*K-Nearest Neighbors*
LIME	*Local Interpretable Model-agnostic Explanations*
LOSO	*Leave-one-Subject-out*
LSTM	*Long-Short Term Memory*
MADALINE	*Multiple Adaptive Linear Neuron*
PCA	*Principal Components Analysis*
RNN	*Recurrent Neural Network*
SACI	*Sistemas Adaptativos e Computação Inteligente*
SVM	*Support Vector Machines*
XAI	*eXplainable AI*

Sumário

1 Introdução

Doenças Neurodegenerativas (DNs) são caracterizadas por uma progressiva degeneração da estrutura e da função do sistema nervoso central, fazendo parte de um grupo de doenças heterogêneas debilitantes e que até o momento não possuem cura, chegando a afetar mais de 30 milhões de indivíduos ao redor do mundo (MARCHI et al., 2021). Muitas dessas doenças podem afetar significativamente a musculatura orofacial, levando a dificuldades significativas em várias funções, como fala, deglutição e habilidades oro-motoras. Além disso, tais condições também podem ter impacto na capacidade de um indivíduo expressar emoções por meio de suas expressões faciais (ZIMMERMAN et al., 2007).

Com os avanços computacionais de armazenamento e de processamento de imagens, a utilização do diagnóstico auxiliado por computador (*computer-aided diagnosis*, CAD), tem se tornado cada vez mais promissor, configurando uma forte opção como ferramenta de segunda opinião para especialistas de outros ramos na medicina, como na radiologia (WINKEL et al., 2021). O CAD tem mostrado bons resultados, tanto na melhora da acurácia como na consistência da interpretação dos dados coletados mediante a classificação por análise quantitativa automatizada feita pelo computador.

Como foco deste trabalho, será abordada a esclerose lateral amiotrófica (ELA), uma DN que afeta o sistema nervoso motor, levando a perda gradativa das funções motoras. Em escala mundial, a prevalência calculada é de 4 a 6 casos em cada 100.000 habitantes, chegando no Brasil a incidência de 1,5 casos por 100.000 habitantes, o que totaliza 2.500 novos casos anuais (SCHLINDWEIN-ZANINI et al., 2015).

Embora já existam documentos, como El escorial, publicada pela Federação Mundial de Neurologia (BROOKS et al., 2000), abordando critérios essenciais a respeito do diagnóstico da ELA (do inglês *Amyotrophic Lateral Sclerosis* - ALS), os conceitos atuais e as definições de ELA ainda não foram unificadas ou padronizadas na prática clínica, e às vezes são vagas ou imprecisas, o que pode causar dificuldades para os neurologistas no tratamento clínico da ELA. Em geral, os pacientes com ELA costumam ser diagnosticados em média 14 meses após o surgimento dos primeiros sintomas e têm uma sobrevida média de 3 a 5 anos após o diagnóstico (XU; YUAN, 2021).

Visando identificar biomarcadores da ELA, observou-se que a detecção dos sinais bulbares (sintomas iniciais caracterizados pela disfagia e dificuldade das funções de fala) apresentam forte contribuição para o diagnóstico e entendimento do progresso da doença, chamando atenção para trabalhos como de Bandini et al. (2018) e Gomes et al. (2023) que conseguiram atingir resultados significativos ao explorar modelos de

aprendizado de máquina interpretativos e de aprendizagem profunda para classificação da ELA por meio de expressões faciais.

Por intermédio de técnicas como o Sistema de Codificação da Ação Facial (*Facial Action Coding System - FACS*) criado pelos psicólogos americanos Paul Ekman e Wallace V. Friesen, que consiste em 46 Unidades de Ação (*Actions Units* - AU) associadas aos músculos responsáveis pelos movimentos da face; e métodos de estimação automática de FACS como criado por Hamm et al. (2011), as FACS têm sido um objeto de estudo para detecção automática de DN como no trabalho de Ali et al. (2021). Com o objetivo de reconhecer a doença de Parkinson, que possui dentre seus sintomas prevalentes a hipomimia (expressões faciais reduzidas), ele realiza um estudo de análise de micro-expressões e movimento dos músculos faciais em vídeo por meio de AUs junto de cálculo estatístico da variância na ação bruta das AUs ativas em cada quadro, empregadas como característica para a classificação com Máquina de Vetores de Suporte (*Support Vector Machines - SVM*). Apesar das fortes evidências sobre a análise dos movimentos faciais no diagnóstico de DNs, muitos estudos ainda não exploram a dimensão temporal, o que limita o grau de liberdade e consequentemente a eficácia do classificador.

Tendo em vista o déficit nas abordagens com relação a análise temporal na inferência do diagnóstico, constatou-se uma abertura para explorar as capacidades das Redes Neurais Recorrentes (*Recurrent Neural Network* - RNN) na análise temporal e espacial dos conjuntos de dados em vídeo, com o objetivo de detectar os biomarcadores bulbares relacionados ao diagnóstico da ELA. As RNNs por sua vez já têm sido amplamente exploradas em vários campos que utilizam dados sequenciais, tais como texto (LIU; GUO, 2019), áudio (WRIGHT et al., 2020) e vídeo (OGAWA et al., 2018). Com os avanços advindos da necessidade de compreensão semântica de seus dados, as RNNs vêm ganhando ainda mais espaço, sendo um modelo promissor, também, em redes neurais profundas, através de suas variações, como a LSTM, do inglês *Long Short-Term Memory* (HOCHREITER; SCHMIDHUBER, 1997).

Portanto, este trabalho tem como objetivo preencher essa lacuna ao explorar as capacidades das RNNs na análise temporal e espacial de dados de vídeo para detectar biomarcadores relacionados ao diagnóstico da ELA. Como resultado, o desenvolvimento prático de um sistema CAD baseado na web que funcione como uma segunda opinião, contribuindo como ferramenta de estudo para diagnósticos mais rápidos, automáticos e de baixo custo.

1.1 Problemática

Em geral, os pacientes com ELA costumam ser diagnosticados em média 14 meses após o surgimento dos primeiros sintomas e têm uma sobrevida média de 3 a 5 anos após o diagnóstico (XU; YUAN, 2021). Devido às diferenças nos sintomas apresentados por cada paciente, a identificação dos biomarcadores relacionados ao diagnóstico da ELA é um grande desafio, sendo responsável por atrasos e erros no diagnóstico, como apresentado pela ALS Association (ASSOCIATION, 2020).

1.2 Justificativa

A detecção precoce da ELA é fundamental para prolongar a expectativa de vida dos pacientes e melhorar sua qualidade de vida. Embora não haja cura ou tratamento efetivo para a doença, o diagnóstico precoce permite o início imediato de ensaios terapêuticos e a reabilitação das funções afetadas pelos sintomas bulbares (HESTERLEE, 2022). Ademais, outro fator que pode ser melhorado é a qualidade de vida do paciente. Um exemplo é a pesquisa realizada por Pontes et al. (2010), na qual conclui-se que a detecção precoce permite aos fonoaudiólogos avaliar objetivamente e traçar diferentes manobras fonoaudiológicas importantes para a reabilitação das funções afetadas pelos sintomas bulbares.

Nos últimos anos, os CADs têm se tornado cada vez mais populares, permitindo o uso de vídeos na análise de movimentos faciais para o diagnóstico de doenças neurodegenerativas. Modelos interpretativos de aprendizado de máquina, como aqueles de Bandini et al. (2018) e Oliveira (2022) e até mesmo modelos de aprendizagem profundas como apresentado por Gomes et al. (2023), que introduz as Redes Neurais em Grafo, têm mostrado que a análise de movimentos faciais é viável e pode gerar análises consistentes para auxiliar no diagnóstico da ELA.

A exploração de modelos que aprendam a dinâmica temporal dos movimentos faciais, como as RNNs, é uma grande vantagem para a classificação de doenças por expressões faciais em vídeos. Embora esses modelos ainda não sejam amplamente explorados para o diagnóstico de DNs por expressões faciais em vídeo, eles têm sido amplamente utilizados para a classificação de outros domínios tangentes em dados em vídeo, como no reconhecimento de emoções (LI et al., 2019) dada à sua capacidade de considerar a dependência temporal entre os quadros sequenciais.

Dentre os principais motivos para o uso limitado de modelos profundos na tarefa de diagnóstico de DNs por meio de expressões faciais em vídeo incluem a priorização de modelos interpretativos. No entanto, a análise mais abrangente proporcionada pelo uso de redes capazes de analisar os dados temporal e espacialmente com menos

viés humano, representa uma grande oportunidade. Isso se torna ainda mais relevante quando se considera a mitigação desse problema por meio do uso de modelos de Inteligência Artificial Explicável. Além disso, a utilização de um modelo pré-treinado que requer apenas vídeos em formato 2D como entrada para inferência oferece uma abordagem prática e de baixo custo, podendo ser integrado como uma ferramenta de segunda opinião, como uma opção viável e acessível.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo desta pesquisa é desenvolver um sistema CAD por meio de aplicação web que introduza modelos sequenciais de aprendizado de máquina para auxílio ao diagnóstico da ELA por meio da análise de expressões faciais. Além disso, o projeto pretende ser uma ferramenta útil para pesquisas futuras em auxílio ao diagnóstico de outras DNs, como Parkinson e Alzheimer, que também podem se beneficiar do reconhecimento dos movimentos faciais (QIANG et al., 2022).

1.3.2 Objetivos Específicos

- Estruturar e implementar modelos sequencias de aprendizado de máquina para análise de expressões faciais a partir de dados em vídeo.
- Realizar testes e analisar a eficiência do modelo, por meio da utilização de métricas comparativas como a matriz de confusão.
- Estruturar e implementar uma aplicação CAD que possa inferir o diagnóstico de ELA utilizando técnicas de aprendizado de máquina, fornecendo ferramentas para tratamento dos dados em vídeo; e modelos de IA explicável, com o intuito de tornar a interpretação e os resultados mais compreensíveis e transparentes.

1.4 Organização do Trabalho

O restante do trabalho está organizado da seguinte forma: no Capítulo 2 são revisados os principais conceitos de Aprendizado de Máquina relevantes para a compreensão deste estudo, bem como os principais componentes de pesquisa. O Capítulo 3 descreve a metodologia, detalhando os materiais e métodos adotados. No Capítulo 4, são descritos os resultados experimentais obtidos, bem como a discussão e decisão do modelo adotado na aplicação CAD. Capítulo 5, são detalhadas as etapas de desenvolvimento e o resultado final da aplicação desenvolvida. Por fim, no Capítulo 6

são apresentadas as considerações finais e discussões sobre possíveis direções para trabalhos futuros.

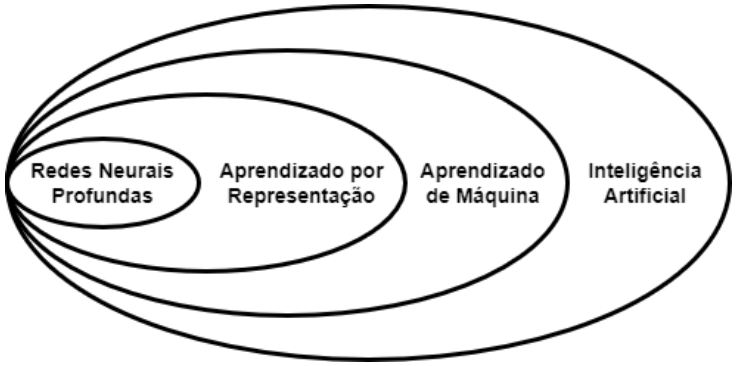
2 Fundamentação Teórica

Neste capítulo, abordaremos conceitos essenciais, incluindo Aprendizado de Máquina, Modelos de Redes Neurais e, em particular, as RNNs e suas variações, que desempenham um papel central neste trabalho. Além disso, exploraremos a importância das FACS na construção da abordagem proposta para a análise de expressões faciais.

2.1 Aprendizado de Máquina

O Aprendizado de Máquina (*Machine Learning*, em inglês) é um subcampo da Inteligência Artificial (IA) que se concentra no desenvolvimento de algoritmos e modelos que permitem a um sistema aprender e melhorar seu desempenho em tarefas específicas a partir da experiência adquirida com os dados. Em outras palavras, o Aprendizado de Máquina tem como principal objetivo a construção de programas capazes de aprimorar seu desempenho a partir de exemplos (MITCHELL, 1997). Diferentemente de outras abordagens dentro do grande campo de IA, ao invés de serem explicitamente programados para realizar uma tarefa, os sistemas de Aprendizado de Máquina usam dados para identificar padrões, fazer previsões ou tomar decisões.

Figura 1 – Diagrama de Venn representando os subconjuntosdo Aprendizado de Máquina até as Redes Neurais Profundas.



Fonte: Adaptada de (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em seus estágios iniciais, a aprendizagem de máquina dependia principalmente de abordagens tradicionais, baseadas em regras, as quais exigiam que especialistas humanos elaborassem recursos complexos e definissem regras de decisão. Esses sistemas, enraizados em técnicas estatísticas, desempenharam um papel fundamental em várias aplicações, incluindo classificação de dados, regressão e reconhecimento de padrões.

Métodos tradicionais de aprendizado de máquina, como árvores de decisão, SVM, vizinhos mais próximos (*K-Nearest Neighbors* - KNN) e regressão linear, formaram os pilares fundamentais do campo. Apesar de terem um papel fundamental em diversos domínios, incluindo a área médica, os métodos tradicionais de aprendizado de máquina têm limitações significativas devido, sendo grande parte dela voltada à sua natureza interpretativa.

Uma dessas limitações é a "maldição da dimensionalidade", a qual se refere a uma série de problemas que surgem ao trabalhar com conjuntos de dados de alta dimensão. Lidar com dados de alta dimensão pode ser desafiador, e muitas vezes requer a extração manual de características, o que não só introduz viés humano, mas também limita a necessidade de especialistas no domínio para realizar essa extração. Além disso, as abordagens tradicionais às vezes têm dificuldade em capturar relações complexas nos dados, o que pode resultar em desempenho subótimo em tarefas específicas.

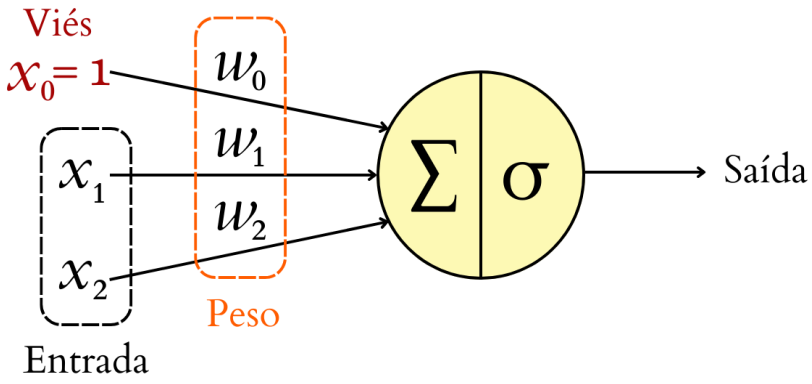
Um ponto crucial na evolução da Aprendizagem de Máquina ocorreu com o advento das redes neurais. Inspiradas na estrutura e na função do cérebro humano, as redes neurais representaram uma reviravolta fundamental na concepção atual da Aprendizagem de Máquina. Elas introduziram uma mudança significativa na abordagem de aprendizado, caracterizada pela capacidade de extrair recursos com maior grau de liberdade e se adaptar a padrões de dados mais complexos.

2.2 Do Perceptron às Redes Neurais Profundas

Em um trabalho intitulado *"A logical calculus of the ideas immanent in nervous activity"* (Um cálculo lógico das ideias inerentes à atividade nervosa) (MCCULLOCH; PITTS, 1943), o neurofisiologista Warren McCulloch e o matemático Walter Pitts concretizaram o que viria a ser o primeiro modelo de uma rede neural. Com o objetivo de explicar o funcionamento dos neurônios, eles desenvolveram algoritmos baseados na lógica de limiar (*threshold logic*, do inglês) e construíram um circuito elétrico capaz de reconhecer duas categorias diferentes de entrada quando $f(x, w)$ é positiva ou negativa. E, para funcionar, os pesos iniciais do modelo linear precisavam ser carregados corretamente, como por exemplo configurado por um operador humano. O funcionamento desse modelo linear dependia da configuração manual dos pesos iniciais, não sendo considerado um modelo capaz de aprender. No entanto, seu sucesso marcou o início da pesquisa em redes neurais, que se dividiu em duas linhas: uma focada na compreensão dos processos cerebrais e como poderiam ser replicados em modelos de redes neurais, e outra, destacada na área de IA, buscando aplicar redes neurais para resolver problemas complexos.

Avançando para 1958, Frank Rosenblatt apresentou o Perceptron, um modelo matemático de um neurônio biológico com a capacidade de aprender conforme uma função de erro. Em neurônios reais, os dendritos recebem sinais elétricos dos axônios de outros neurônios, ao passo que no Perceptron esses sinais, esses sinais elétricos são representados numericamente. Nas sinapses entre dendritos e axônios, os sinais eletrônicos são modulados em várias quantidades, resultando em um sinal de saída quando a força total dos sinais de entrada ultrapassa um limite específico. Esse fenômeno é modelado no Perceptron, por meio do cálculo da soma ponderada das entradas, $z = \sum_i w_i x_i$ para representar a força total dos sinais de entrada e aplicamos uma função de ativação $\sigma(z)$ para determinar a saída, conforme ilustrado na Figura 2. Na representação ilustrada temos $X = x_0, x_1, x_2$ como as características de entrada, sendo $x_0 = 1$ a característica de viés; $W = w_0, w_1, w_2$ os pesos do modelo; e σ, a função de ativação.

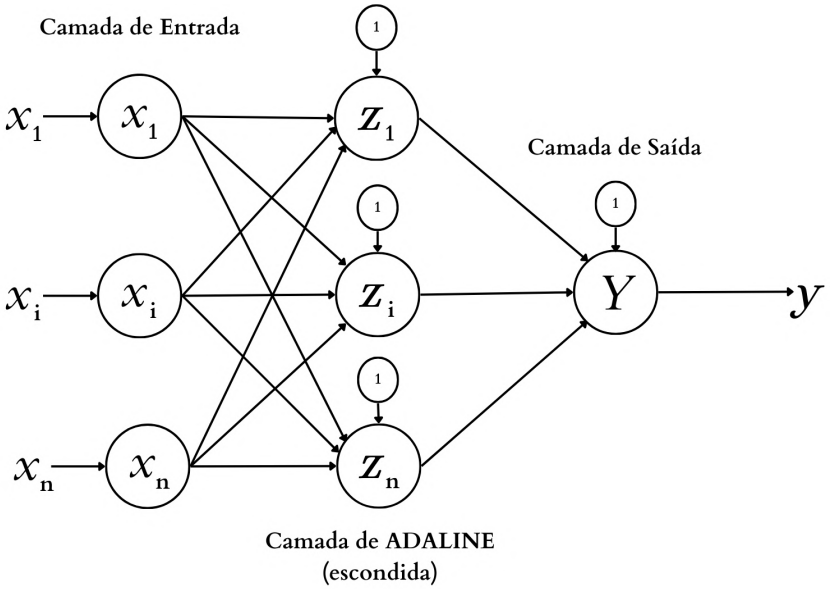
Figura 2 – Representação de um Perceptron



Fonte: Elaborada pelo autor.

Essa arquitetura permite que uma única camada de Perceptron atue como um classificador linear binário, podendo representar operações lógicas condicionais, como AND e OR. Próxima a época, o desenvolvimento do ADALINE (*Adaptive Linear Neuron*), uma rede neural que possui uma única unidade linear, tendo como sua função de aprendizado é baseada no algoritmo *"least mean squares"* e que trouxe a desenvolvimento, MADALINE (*Multiple Adaptive Linear Neuron*), a primeira aplicação de rede neural em um problema do mundo real, resolvendo o problema de ecos nas linhas telefônicas por meio de um filtro adaptativo (WIDROW; HOFF et al., 1960). Sua arquitetura é apresentada na Figura 3.

Figura 3 – Arquitetura MADALINE.



Fonte: Elaborada pelo autor.

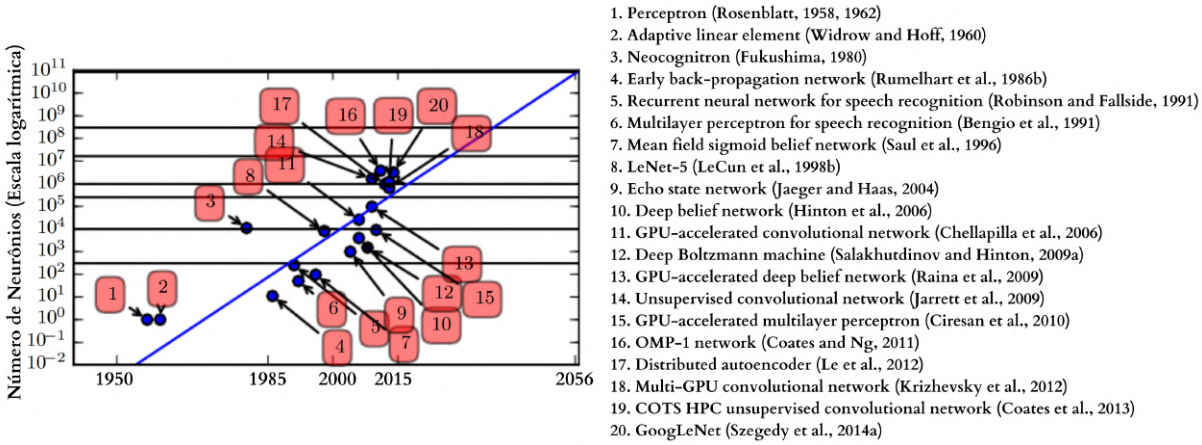
No entanto, uma única camada de Perceptron mostrou-se inadequado para resolver problemas não lineares, como a função lógica XOR, onde $f([0,1],w) = 1$ e $f([1,0],w) = 1$, mas $f([1,1],w) = 0$ e $f([0,0],w) = 0$. Essa limitação levou a uma redução significativa no financiamento da pesquisa em redes neurais, um período conhecido como *"dark age"* ou inverno da IA, que durou até meados de 1981.

A segunda onda de pesquisas em redes neurais surgiu em grande parte por meio de um movimento chamado conexionismo. Uma abordagem interdisciplinar que surgiu dentro do contexto da ciência cognitiva com o intuito de entender a mente, por meio da combinação de várias camadas diferentes de análise, indo contrário a maioria dos cientistas cognitivos que estudavam modelos de raciocínio simbólico. Apesar de sua popularidade, os modelos simbólicos eram difíceis de explicar em termos de como o cérebro poderia realmente implementá-los usando neurônios, desta forma trazendo aos conexionistas começaram a estudar modelos de cognição que poderiam de fato ser fundamentados em implementações neurais, revivendo muitas ideias que remontam ao trabalho do psicólogo Donald Hebb na década de 1940 (HEBB, 1949).

O conceito central do conexionismo é de que um grande número de unidades computacionais simples pode resultar em comportamento inteligente quando interconectadas. Isso se aplica tanto aos neurônios nos sistemas nervosos biológicos quanto às unidades ocultas em modelos computacionais. Entre as principais realizações do movimento conexionista, destacam-se a representação distribuída, a qual implica que cada entrada em um sistema deve ser representada por muitos atributos, e cada atributo deve contribuir para a representação de muitas entradas possíveis (HINTON, 1984); e o uso bem-sucedido da retropropagação (do inglês, *backpropagartion*) para

treinar redes neurais profundas com representações internas (RUMELHART; HINTON; WILLIAMS, 1986; FOGELMAN-SOULIé; CUN, 1987). Apesar dos altos e baixos em termos de popularidade, mas ainda é a abordagem dominante para treinar modelos com cada vez mais neurônios e camadas escondidas. Embora a popularidade da retropropagação tenha tido altos e baixos, ela ainda é a abordagem dominante para treinar modelos com cada vez mais neurônios e camadas escondidas. A Figura 4 ilustra o aumento significativo na pesquisa de modelos mais profundos após essa segunda onda proporcionada pelo movimento conexionista.

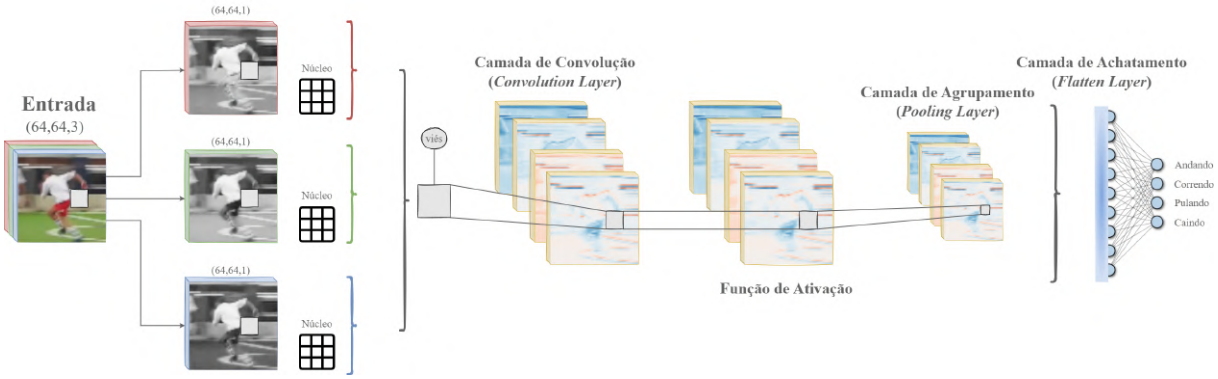
Figura 4 – Crescimento no número de neurônios em aprendizado de máquina pelo tempo



Fonte: Adaptada de (GOODFELLOW; BENGIO; COURVILLE, 2016).

Hoje, redes neurais são amplamente utilizadas em diversas aplicações. Aliado a grande profusão de dispositivos interconectados, gerou-se um enorme volume de dados, proporcionando um rico campo de análise para modelos que requerem grandes conjuntos de dados. O aumento do poder computacional, incluindo o uso de GPUs e TPUs, acelerou o treinamento de redes neurais profundas, tornando-as viáveis em áreas que vão desde previsão de mercado até auxílio em diagnósticos médicos. Na Figura 5, um exemplo de arquitetura de rede neural profunda ainda muito que apresentou como um grande marco no campo de classificação por imagem.

Figura 5 – Arquitetura de uma Rede Neural Convolucional (CNN).



Fonte: Elaborada pelo autor.

Essa transição da aprendizagem de máquina tradicional para redes neurais destaca a importância do processamento eficiente de grandes conjuntos de dados e demonstra o papel essencial das redes neurais no cenário atual de IA e aprendizagem de máquina.

2.3 Dados Sequenciais e as Redes Neurais Recorrentes

Quando se trata de aplicar aprendizado de máquina a dados sequenciais, como texto, fala ou vídeo, uma abordagem seria a do uso de uma rede neural de propagação direta (*feedforward*), como uma CNN, e apresentá-la com toda a sequência de dados. No entanto, essa abordagem tem limitações práticas, como lidar com um tamanho fixo de entrada, e pode não capturar eventos cruciais que envolvam a relação temporal ou a dependência em dados sequenciais. Por exemplo, essa abordagem poderia ter dificuldade em discernir em um vídeo o momento em que uma pessoa tem a intenção de se sentar e quando tem a intenção de se levantar.

RNNs (ROBINSON; FALLSIDE, 1987) são uma família de redes neurais que conseguem justamente lidar com esse desafio, para o processamento de dados sequenciais. Enquanto uma CNN é uma rede neural especializada no processamento de uma grade de valores, como uma imagem representada por uma matriz X, uma RNN é projetada para o processamento de uma sequência de valores representados como $x(1), ..., x(t)$.

Da mesma forma que as redes convolucionais podem facilmente dimensionar para imagens com largura e altura variáveis, e algumas até podem processar imagens de tamanhos diferentes, as RNNs são altamente flexíveis para lidar com sequências de comprimentos diversos, o que as torna ideais para análise de dados sequenciais com durações variadas, sendo uma escolha poderosa em tarefas que envolvem processa-

mento de linguagem natural, análise de séries temporais, reconhecimento de dados em vídeo e muitos outros domínios onde a natureza sequencial dos dados é fundamental.

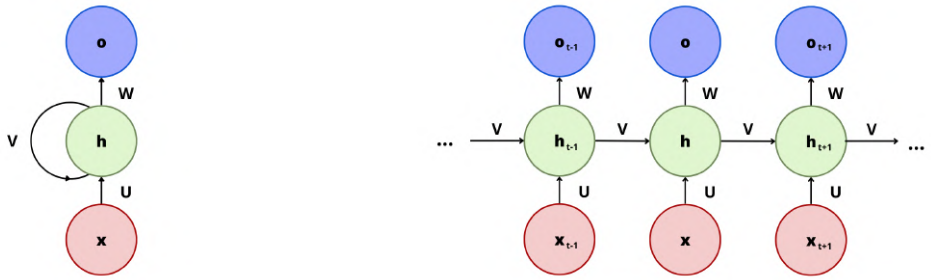
O compartilhamento de parâmetros é um aspecto de extrema importância no funcionamento das RNNs, pois permite a extensão e aplicação do modelo a exemplos com diferentes formas, incluindo sequências de comprimentos variados e posições no tempo. Por exemplo, em frases diferentes, mas com a mesma informação temporal, como "Eu fui ao Nepal em 2009" e "Em 2009, eu fui ao Nepal", podemos identificar a informação do ano em que o narrador visitou o Nepal independente da posição em que ele aparece. Para uma rede de propagação direta com parâmetros separados, seria necessário aprender uma grande gama de regras separadas em cada posição da frase. Por outro lado, as RNNs compartilham pesos entre passos de tempo, permitindo aprendizado eficaz, economia de recursos e melhor desempenho em tarefas sequenciais (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.3.1 Arquitetura das Redes Neurais Recorrentes

As RNNs tradicionais refere a uma célula básica de uma RNN que não possui mecanismos adicionais, como os encontrados em arquiteturas mais complexas, como LSTM (*Long Short-Term Memory*) ou GRU (*Gated Recurrent Unit*), posteriormente explicadas. A célula recorrente padrão geralmente possui apenas uma camada simples de neurônios e é usada para modelar normalmente dependências de curto prazo em sequências de dados.

As RNNs tradicionais são mais simples, porém podem apresentar desafios ao lidar com dependências de longo prazo em sequências, o que as torna menos adequadas para tarefas mais complexas. Uma visão geral de sua arquitetura pode ser vista na Figura 6, que mostra tanto sua representação em grafo compactado quanto expandido.

Figura 6 – Arquitetura da RNN.



(a) Representação Compacta (b) Representação Expandida

Fonte: Elaborada pelo autor.

De forma matemática, em alinhamento com as formulações fundamentais detalhadas no trabalho referenciado (YU et al., 2019), as equações matemáticas da célula tradicional da rede neural recorrente são escritas da seguinte forma:

$$\begin{aligned} h_t &= \sigma(W_h h_{t-1} + W_x x_t + b), \\ y_t &= h_t, \end{aligned} \tag{2.1}$$

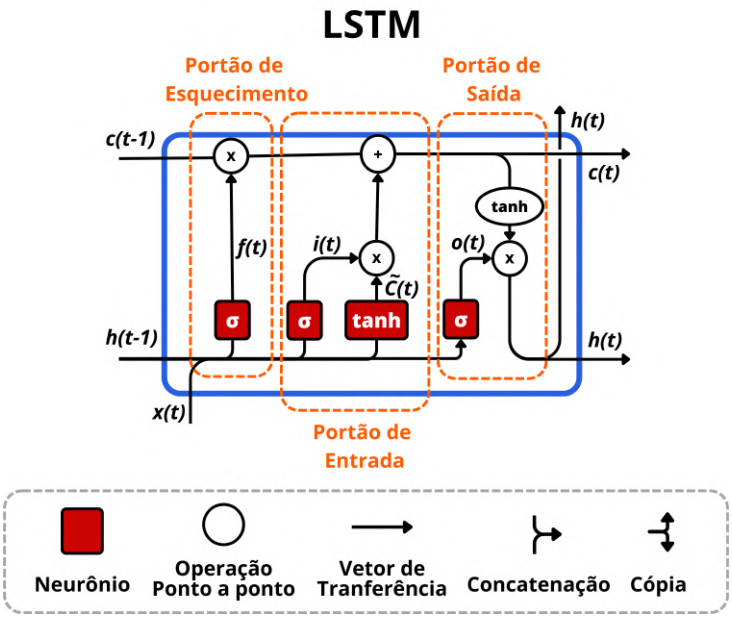
em que x_t, h_t, e y_t são as entradas, informações recorrentes (camada escondida) e saídas, respectivamente, da célula em um determinado momento t. Os pesos da célula são representados por W_h e W_x, enquanto b é o viés da célula, e σ denota uma função de ativação.

2.3.1.1 Memória de Longo Prazo e Curto Prazo

Utilizando memória e introduzindo a ideia de portões (mais conhecidos pelo termo inglês *"gates"*), as RNNs tiveram uma melhora em sua capacidade com o que foi chamado de Memória de Longo Prazo e Curto Prazo (LSTM) (HOCHREITER; SCHMIDHUBER, 1997).

Inicialmente, como o primeiro modelo que veio para suprir o problema de dependências de longo prazo em sequências de dados, Hochreiter e Schmidhuber (1997) desenvolveram uma célula que possuía dois portões, um de entrada e um de saída, cujo objetivo era guiar quais informações deveriam ser guardadas pela célula e quais informações poderiam se tornar a saída da célula. Adicionando mais um portão, o "portão de esquecimento"(do inglês, *Forget-gate*), Gers, Schmidhuber e Cummins (2000) introduziram uma nova variação da LSTM, sendo essa versão a mais popular quando se trata do termo célula LSTM. Esse portão tem por objetivo determinar quais informações devem ser transferidas para a próxima célula, de forma que informações irrelevantes possam ser esquecidas da função de memória. A célula LSTM com o Porão de esquecimento é ilustrada na Figura 7.

Figura 7 – Diagrama de uma célula LSTM



Fonte: Elaborada pelo autor.

Baseado nas conexões da célula LSTM com o Porão de esquecimento ilustrada na Figura 7, as equações de atualizações pode ser matematicamente expressa da seguinte forma, conforme Yu et al. (2019):

$$
\begin{aligned}
f_t &= \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f), \\
i_t &= \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i), \\
\tilde{c}_t &= \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}), \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \\
o_t &= \sigma(W_{oh}h_{t-1} + W_{ox}x_t + bo, \\
h_t &= o_t \cdot \tanh(c_t).
\end{aligned}
\tag{2.2}
$$

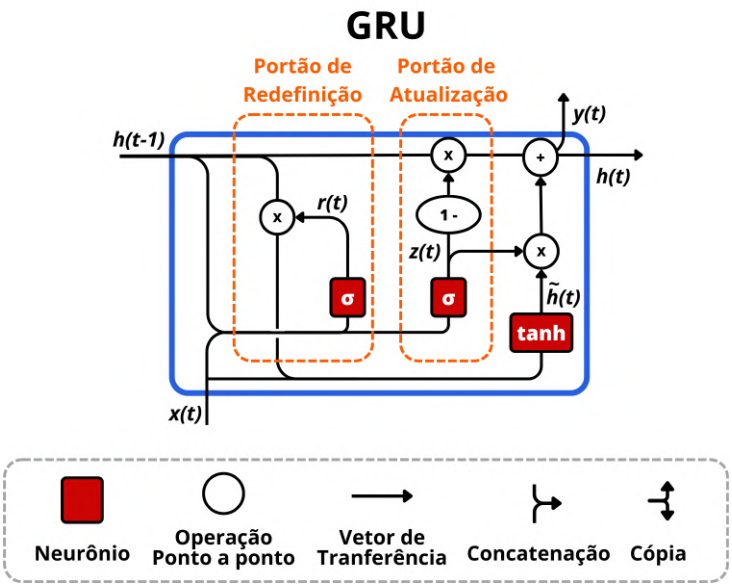
onde c_t denota o estado da célula do LSTM. W_i, $W_{\tilde{c}}$ e W_o são os pesos, e o operador '$\cdot$' denota a multiplicação ponto a ponto de dois vetores. Ao atualizar o estado da célula, são utilizados o portão de entrada i_t, o portão de saída o_t e o portão de esquecimento f_t. O portão de esquecimento f_t determina quais informações serão mantidas e quais serão descartadas. Quando o valor do portão de esquecimento, f_t, é 1, a informação é mantida; enquanto um valor de 0 significa que a informação será descartada.

2.3.1.2 Unidade Recorrente com Portões

Outra variação de RNNs bastante conhecida é a Unidade Recorrente com Portões (GRU) (CHO et al., 2014).

Desenvolvida essencialmente como uma variante da LSTM, apesar de menos poderosa, sua arquitetura é capaz de economizar uma grande quantidade de parâmetros associados. Para reduzir o número de parâmetros, a célula GRU, representada na Figura 8, integra o portão de esquecimento e o portão de entrada da célula LSTM em um único portão de atualização.

Figura 8 – Diagrama de uma célula GRU



Fonte: Elaborada pelo autor.

Assim, possuindo apenas dois portões: um portão de atualização (z_t) e um portão de redefinição (r_t). As equações de atualização são formuladas da seguinte maneira (YU et al., 2019):

$$\begin{aligned} r_t &= \sigma(W_{rh}h_{t-1} + W_{rx}x_t + b_r), \\ z_t &= \sigma(W_{zh}h_{t-1} + W_{zx}x_t + b_z), \\ \tilde{h}_t &= \sigma(W_{\tilde{h}h}h_{t-1} + W_{\tilde{h}x}x_t + b_z), \\ h_t &= (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}. \end{aligned} \tag{2.3}$$

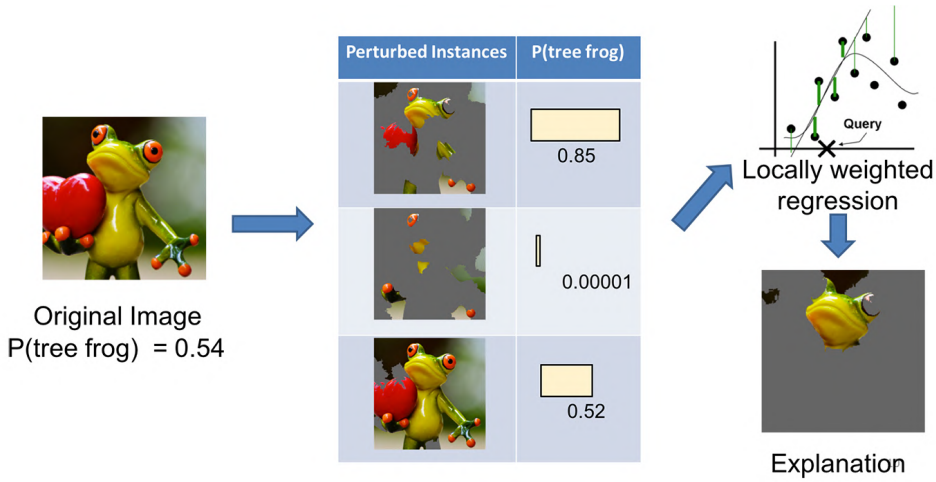
2.4 Inteligência Artificial Explicável

A crescente adoção de soluções de aprendizado de máquina trouxe à tona um desafio significativo: a falta de explicações claras sobre como esses modelos tomam decisões. Isso é particularmente crítico em setores onde transparência e responsabilidade desempenham um papel fundamental, como na área de saúde com ferramentas de auxílio ao diagnóstico. Esse desafio é resultado da transformação de todo o processo

de cálculo em algo frequentemente descrito como uma "caixa-preta", um sistema complexo e opaco, tornando suas decisões difíceis de interpretar. A Inteligência Artificial Explicável (*eXplainable AI* - XAI) é um campo de pesquisa dedicado a tornar os resultados dos sistemas de IA mais compreensíveis para os seres humanos, tornando os mecanismos internos mais transparentes e fornecendo explicações detalhadas sobre suas decisões em vários níveis de profundidade.

Devido à natureza recorrente e às dependências de longo prazo das RNNs em dados sequenciais, a interpretação de suas decisões é desafiadora e requer modelos XAI específicos. A técnica LIME, do inglês *Local Interpretable Model-agnostic Explanations*, é parte de um conjunto de modelos XAI que se baseia apenas nas entradas e saídas para interpretar modelos complexos. É comumente usada em modelos considerados "caixas pretas", tornando-os mais aproximáveis por meio de modelos locais e interpretáveis.

Figura 9 – Visão geral do processo de interpretação de resultados do LIME.



Fonte: (RIBEIRO; GUESTRIN, 2016).

A ideia por trás do LIME surgiu em um artigo de Ribeiro, Singh e Guestrin (2016), no qual os dados foram perturbados e passados por um modelo "caixa preta" para observação dos resultados gerados nas saídas. Esse processo de perturbação é independente do modelo, tornando-o compatível para RNNs e dados sequenciais. O método atribui pesos aos novos pontos de dados criados com base em sua proximidade com os dados originais (sendo capaz de lidar com diferentes dimensões) e, em seguida, ajusta um modelo substituto, como a regressão linear, a esse conjunto de dados ponderado (Figura 9). Esse modelo substituto fornece explicações para pontos de dados individuais, melhorando a interpretabilidade das previsões do modelo.

2.5 Reconhecimento de Expressões Faciais

Desenvolvido por Paul Ekman e Wallace Friesen, o Sistema de Codificação de Expressões Faciais (FACS) (EKMAN; FRIESEN, 1978) é o sistema mais popular para quantificar a intensidade de grupos de músculos faciais referidos como unidades de ação (AUs). Esse sistema de codificação facial permite medir dados objetivos e quantificáveis, proporcionando uma análise não invasiva e natural das expressões faciais com alta precisão temporal. Isso torna possível a análise de micro-expressões e mudanças sutis nas expressões faciais, o que é valioso no campo do reconhecimento de emoções para compreender a dinâmica das respostas emocionais.

Embora o FACS seja um sistema robusto, a extração manual desses códigos quantitativos é uma tarefa demorada e pode introduzir vieses, uma vez que os pesquisadores precisam codificar manualmente gravações de vídeo com base nas unidades de ação definidas. Com os avanços na visão computacional e no aprendizado de máquina, várias técnicas foram desenvolvidas para automatizar esse processo. Neste trabalho, para obter uma extração de características mais robusta, alinhado com as abordagens mais recentes, utilizou-se a ferramenta Py-feat (JOLLY et al., 2021).

Baseando-se no modelo da ferramenta OpenFace 2.0 (BALTRUSAITIS et al., 2018), para reconhecimento das AUs a ferramenta Py-feat extrai características de suas imagens de entrada por Histograma de Gradientes Orientados(*Histogram of Oriented Gradient* - HOG) a partir das coordenadas dos pontos de referência utilizando um algoritmo de *convex hull* para delinear as regiões de interesse. Posteriormente, a representação HOG é comprimida por meio da técnica de Análise de Componentes Principais (*Principal Components Analysis* - PCA) e, por fim, utiliza essas características para prever individualmente cada uma das AUs, fazendo uso de métodos populares de aprendizado superficial baseados em kernels, como a SVM e estratégias de aprendizado em conjunto.

No trabalho de Jolly et al. (2021) é realizada uma comparação das duas ferramentas em relação a identificação das AUs no conjunto de dados DisfaPlus (MAVADATI; SANGER; MAHOOR, 2016). Os resultados dessa comparação em relação a 12 AUs que identificados pelos dois modelos é apresentado no Quadro 1.

Quadro 1 – Comparação das ferramentas Py-feat e OpenFace 2.0 no conjunto de dados DisfaPlus.

Modelo	**Unidades de Ação**											
	01	**02**	**04**	**05**	**06**	**09**	**12**	**15**	**17**	**20**	**25**	**26**
OpenFace 2.0	.71	.52	.69	.49	.81	.54	.83	.34	.43	.13	.72	.67
Py-feat	.60	.54	.64	.57	.71	.39	.78	.25	.26	.35	.84	.56

Fonte: (JOLLY et al., 2021).

3 Materiais e Métodos

Esta capítulo apresenta as bases de dados e a metodologia utilizada nos experimentos realizados para avaliar os desempenhos dos métodos.

3.1 Conjunto de Dados

Introduzido por Bandini et al. (2020), Toronto Neuroface é a primeira base de dados pública para análise orofacial de doenças neurológicas. Mesmo sendo uma base de dados pública, a utilização da mesma foi realizada com total conformidade às permissões de acesso, em estrito cumprimento das restrições éticas relacionadas à proteção dos direitos de imagem dos pacientes e às suas respectivas condições clínicas. O conjunto é composto por 261 vídeos com avaliação clínica por vídeo, contendo também mais de 3300 anotações de quadros faciais de indivíduos com ELA e pós-AVC, bem como um grupo de controle (*Healthy Control* - HC). No contexto do problema, foram utilizados apenas os vídeos realizados pelos participantes dos grupos HC e ELA, ambos com 11 indivíduos cada, totalizando 22 participantes.

Quadro 2 – Numero de repetições manualmente recortadas de cada tarefa

Tarefas	**Descrição**	**ELA**	**HC**
SPREAD	Repetições fingindo sorrir com os lábios fechados	55	59
KISS	Repetições fingindo beijar uma criança	59	57
OPEN	Repetições com abertura máxima da mandíbula	54	55
BLOW	Repetições fingindo assoprar uma vela	31	39
BBP	Repetições da frase *“Buy Bobby a Puppy”*	95	111
PA	Repetições das sílabas */pa/* o mais rápido possível em uma única respiração	100	110
PATAKA	Repetições das sílabas */pataka/* o mais rápido possível em uma única respiração	88	108

Fonte: Adaptado de (GOMES et al., 2023).

Cada participante do conjunto experimental foi gravado em um ambiente controlado, realizando um conjunto de tarefas oro-faciais verbais e não verbais, típicas de avaliação clínica. Cada tarefa foi realizada com um número específico de repetições, sendo coletadas em um único vídeo. Para a segmentação de cada repetição,

foi realizada a segmentação manual do vídeo utilizando a ferramenta de edição de vídeo gratuita e de código aberto, Shotcut[1]. O número de repetições, juntamente com a descrição de cada tarefa utilizada no conjunto de dados experimental, é apresentado no Quadro 2.

O critério de corte para as repetições levou em consideração o ciclo completo da movimentação da tarefa. Nesse sentido, para as tarefas não verbais, foram realizados cortes capturando o início da repetição, o ponto em que o movimento atingia o máximo, geralmente próximo ao centro do vídeo, e o encerramento da repetição, quando voltava para o movimento próximo a iniciação do próxima repetição. Por conta da natureza sensível dos dados, a divulgação de imagens dos pacientes do conjunto de dados não é permitida, desta forma, para fins didáticos e com a colaboração dos membros do laboratório de Sistemas Adaptativos e Computação Inteligente (SACI, UNESP-Bauru), foram geradas algumas imagens representativas. Estas imagens foram criadas de acordo com as regras do ambiente controlado descritas pelo trabalho que compõem o conjunto de dados original (BANDINI et al., 2020). Uma visão geral do critério de corte das tarefas não verbais é apresentada na Figura 10.

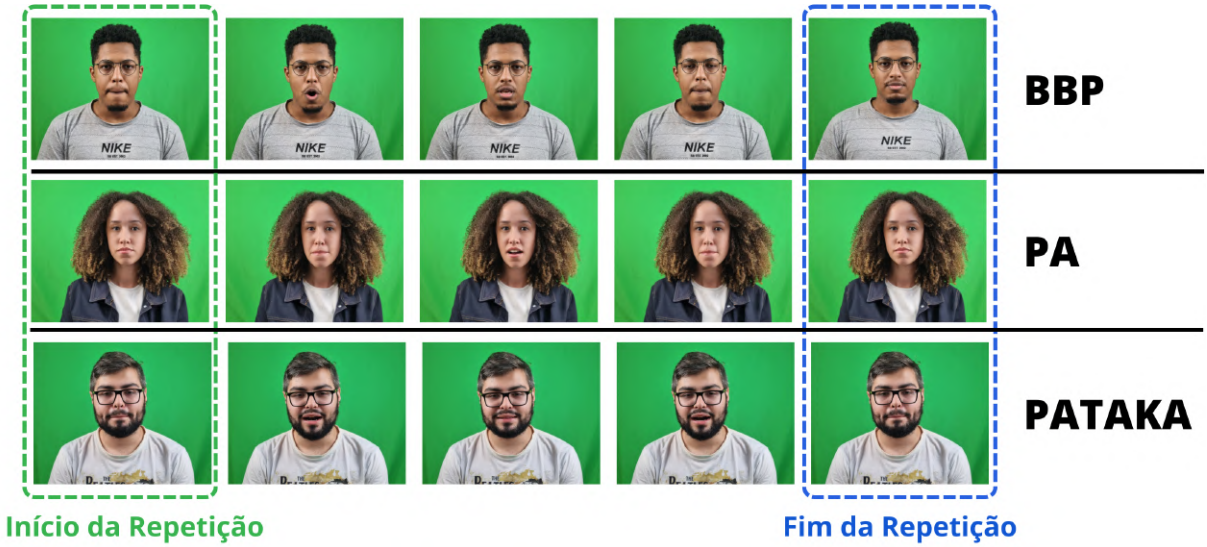
Figura 10 – Representação dos cortes de repetição das tarefas não verbais.

Fonte: Elaborada pelo autor.

No caso das tarefas verbais, os cortes foram efetuados tanto no início quanto no final da repetição, alinhados com o início de um novo ciclo da repetição da palavra, conforme representado na Figura 11.

[1] Ferramenta de edição de vídeo: https://shotcut.org/.

Figura 11 – Representação dos cortes de repetição das tarefas verbais.



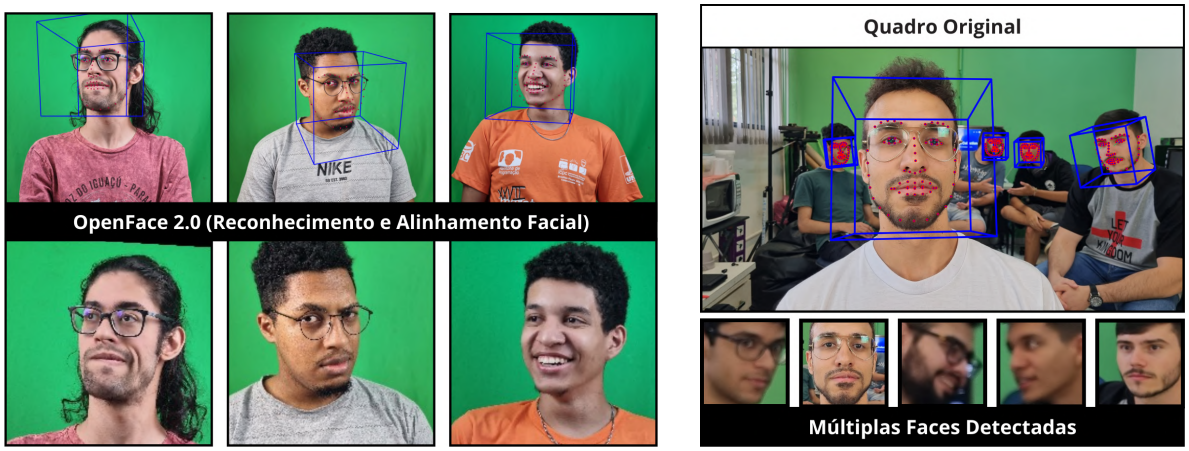
Fonte: Elaborada pelo autor.

3.2 Pré-processamento e Extração de Características

Apesar de as imagens terem sido gravadas em um ambiente controlado, alguns vídeos do conjunto de dados aparecem pessoas ao fundo não relevantes para a análise, além de cenários diferentes. Para mitigar esse viés, todos os quadros de cada vídeo foram processados usando a ferramenta OpenFace 2.0 (BALTRUSAITIS et al., 2018). O processo envolve o recorte e centralização do rosto do indivíduo principal em cada quadro.

O OpenFace 2.0 é capaz de lidar tanto com a detecção de um único rosto quanto com a detecção de múltiplas faces em uma cena, como apresentado na Figura 12. Para escolher o rosto principal, foi considerado utilizar um critério baseado na proximidade em relação à câmera, ou seja, o rosto que ocupa uma área maior e que está mais proeminente em primeiro plano é selecionado como o rosto principal. Em seguida, a ferramenta aplica uma transformação com base na estimativa de posição da cabeça e realiza uma operação de recorte em todos os quadros.

Figura 12 – Ferramenta OpenFace 2.0 para detecção e alinhamento facial.



(a) Detecção de face individual. (b) Detecção de múltiplas faces.

Fonte: Elaborada pelo autor.

Dessa forma, todos os quadros foram padronizados em imagens em escala de cinza com tamanho de 200×200 pixels com a face do indivíduo principal centralizada. Esse processo de pré-processamento garante uma uniformidade nas imagens e facilita a análise subsequente.

Após esse pré-processamento, todos os quadros foram submetidos à ferramenta Py-feat (JOLLY et al., 2021), extraindo um total de 20 unidades de ação, como descritas pela Quadro 3.

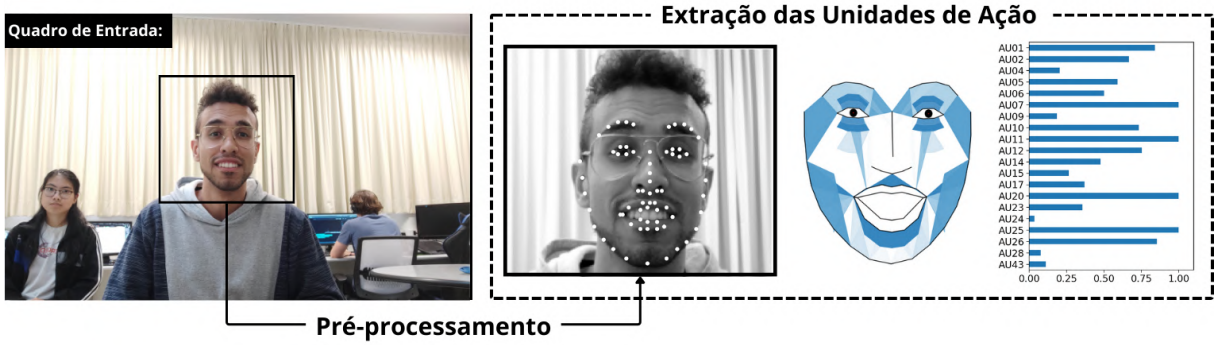
Quadro 3 – Unidades de Ação Extraídas pelo Py-feat (JOLLY et al., 2021)

Unidade de Ação	Descrição
AU01	Elevador Interno da Sobrancelha
AU02	Elevador Externo da Sobrancelha (unilateral, lado direito)
AU04	Abaixador da Sobrancelha
AU05	Elevador da Pálpebra Superior
AU06	Elevador da Bochecha
AU07	Apertador da Pálpebra
AU09	Enrugador do Nariz
AU10	Elevador do Lábio Superior
AU11	Aprofundador Nasolabial
AU12	Puxador do Canto dos Lábios
AU14	Formador de Covinhas
AU15	Depressor do Canto dos Lábios
AU17	Elevador do Queixo
AU20	Esticador do Lábio
AU23	Apertador do Lábio
AU24	Pressionador do Lábio
AU25	Separação dos Lábios
AU26	Abaixador da Mandíbula
AU28	Sucção dos Lábios
AU43	Olhos Fechados

Fonte: Adaptado de (FARNSWORTH, 2022).

Assim, para cada quadro de vídeo, obtivemos um conjunto de vetores, sendo cada valor do vetor uma medida escalar entre $[0, 1]$, representando as AUs pela ferramenta. Uma visão completa do processos é apresentada na Figura 13.

Figura 13 – Visão geral do processo de cada quadro do vídeo antes de entrar nos modelos

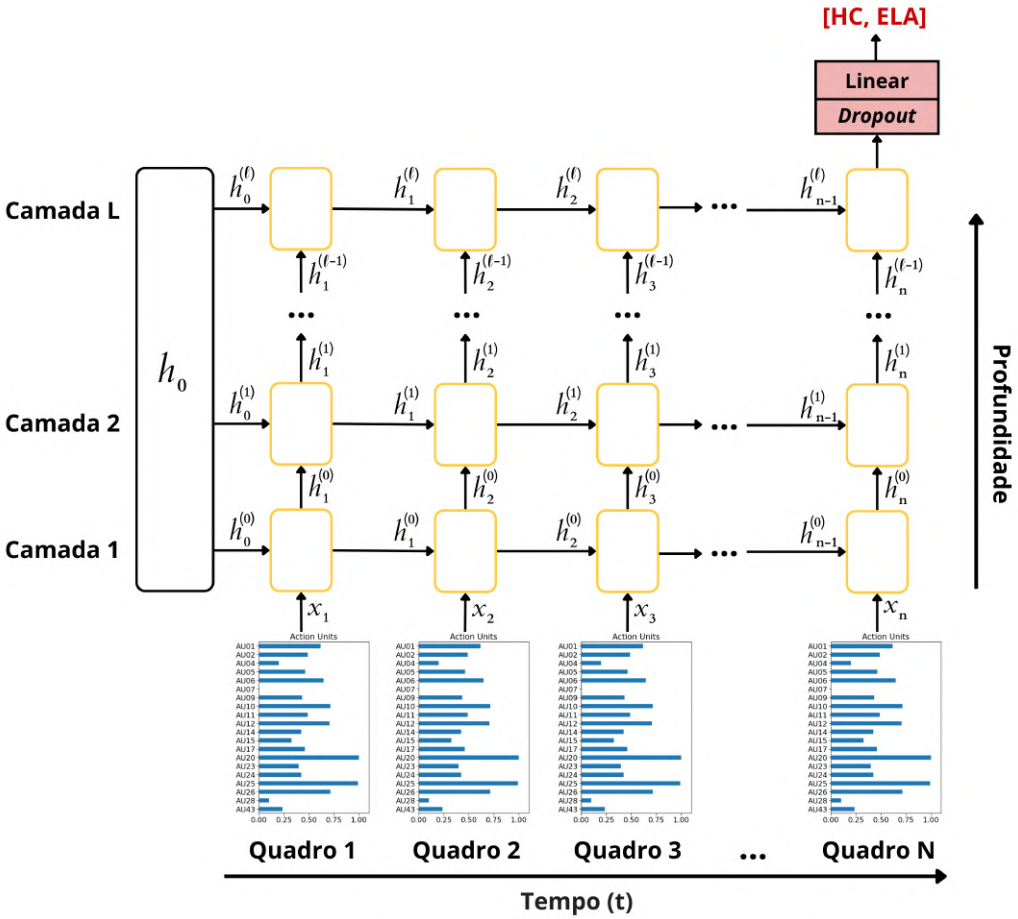


Fonte: Elaborada pelo autor.

3.3 Modelo Proposto

Para realizar uma análise temporal do problema, selecionou-se 15 quadros de cada vídeo de repetição, utilizando cada um deles como entrada para a recorrência do modelo em cada interação. Essa arquitetura do modelo é caracterizada como um framework *Many-to-One* (Muitos para Um), conforme ilustrado na Figura 14. Nessa configuração, cada um dos 15 quadros serve como entrada para o modelo e retorna uma única saída com a classificação.

Figura 14 – Abordagem proposta



Fonte: Elaborada pelo autor.

Cada modelo sequencial incorporou camadas empilhadas para alcançar representações hierárquicas de dados de vídeo sequencial, aumentando a profundidade do modelo. Cada camada tem a capacidade de capturar diferentes níveis de abstração, permitindo que a rede compreenda tanto detalhes de baixo nível quanto padrões de alto nível nos dados. Por meio de validações experimentais, configurou-se quatro camadas empilhadas para os modelos GRU e LSTM e três camadas empilhadas para o RNN tradicional. O número de características ocultas para todos os modelos foi definido como o triplo do número de características na camada de entrada, totalizando 60 características em cada camada oculta.

No processo de múltiplas camadas, a entrada $x_t^{(l)}$ da l-ésima camada é o estado oculto $h_t^{(l-1)}$ da camada anterior, modificado pela aplicação do *dropout*, uma variável aleatória de Bernoulli que assume seu valor conforme uma probabilidade de *dropout* defina como hiperparâmetro, sendo 0.3 para o RNN e 0.4 para os modelos LSTM e GRU.

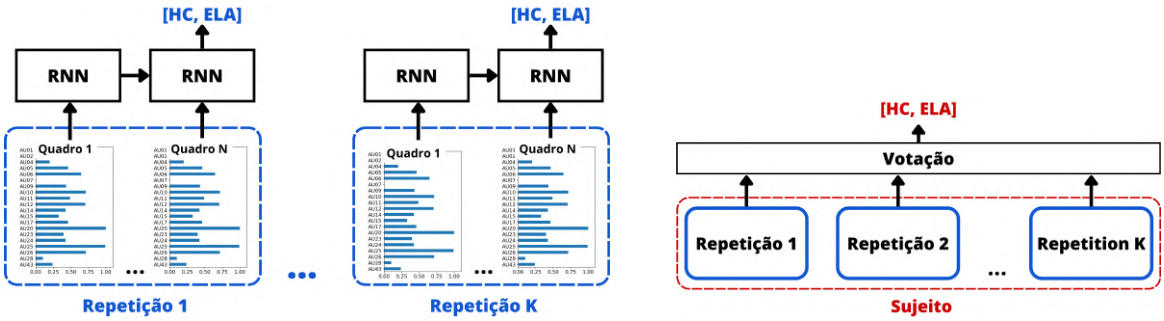
3.4 Classificação e Avaliação

Para a avaliação do modelo foi utilizado o método *Leave-one-Subject-out* (LOSO), o que garante que todos os participantes do conjunto de dados sejam usados como parte dos testes. Além disso, para facilitar a generalização do modelo em cada iteração do LOSO, o conjunto de treinamento foi dividido em dois subconjuntos: um para treinamento e um para validação. Essa abordagem possibilita a realização de busca por hiperparâmetros durante o treinamento, evitando assim a manipulação dos hiperparâmetros de forma a enviesar os resultados em relação ao conjunto de teste. Seguindo a mesma abordagem de Bandini et al. (2018), foram realizados os seguintes tipos de classificação:

- **Classificação baseada na repetição**: A cada interação do LOSO, todos os vídeos de repetição de um participante entravam como conjunto de teste, sendo os demais como conjunto de treino. Ao final, cada video repetição do participante da interação é classificado individualmente obtendo a acurácia em cima dessas repetições, conforme ilustrado pela Figura 15a.
- **Classificação baseada no sujeito**: A classificação é feita levando em consideração todo o conjunto de repetições do indivíduo, através de uma votação. Em caso de empate, o indivíduo é considerado como HC, para obter uma classificação mais conservadora. Assim, para cada interação obtendo apenas uma classificação binária de um único indivíduo, conforme apresentado na Figura 15b.

Considerando modelos interpretativos como referência de comparação para o nosso trabalho, os experimentos também foram realizados considerando mais dois modelos de classificação para fins de comparação: SVM com função linear e de base radial (RBF) e Regressão Logística. Ambos os modelos foram treinados com as mesmas 20 unidades de ação ao longo dos 15 quadros de entrada, considerando a variação quadro a quadro. Para a busca dos hiperparâmetros, foi realizada uma busca em grade para otimizar os valores do parâmetro de confiança C e do parâmetro de escala do kernel RBF γ nos modelos de referência.

Figura 15 – Modos de Classificação



(a) Classificação baseada na repetição (b) Classificação baseada no sujeito

Fonte: Adaptada de (GOMES et al., 2023).

3.4.1 Matriz de Confusão

Para avaliar o desempenho dos modelos propostos, utilizou-se a matriz de confusão, que mostra a frequência de classificações para cada classe. A acurácia, a precisão e a sensibilidade são métricas que fornecem uma maneira de avaliar a utilidade e a integridade do classificador. Abaixo, são descritas essas medidas:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \tag{3.1}$$

$$\text{Precisão} = \frac{VP}{VP + FN} \tag{3.2}$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \tag{3.3}$$

em que, VP, VN, FN e FP são verdadeiro positivo, verdadeiro negativo, falso negativo e falso positivo, respectivamente. Considerando também a Pontuação F1(do inglês, F1-Score), a média harmônica entre precisão e a sensibilidade, descrita pela seguinte equação:

$$\text{Pontuação F1} = \frac{2 \cdot \text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \tag{3.4}$$

3.4.2 Testes Pós-Hoc

Para comparar os resultados gerados por cada modelo, realizou-se o teste Pós-Hoc de Friedman, utilizando as medidas extraídas das matrizes de confusão obtidas por cada modelo em cada tarefa. O teste de Friedman é uma ferramenta estatística apropriada para avaliar a diferença de desempenho entre vários modelos ou métodos quando os dados não seguem uma distribuição normal ou as suposições

de homogeneidade de variâncias não são atendidas. Nesse contexto, ele nos permite determinar se há diferenças significativas entre os modelos sequenciais implementados em relação à classificação das amostras.

Após a aplicação do teste de Friedman e a obtenção de resultados que rejeitem a hipótese de que os modelos são iguais, é relevante realizar testes pós-hoc específicos, como o teste de Nemenyi, para identificar quais modelos apresentam diferenças estatisticamente significativas entre si. Assim, ampliando o entendimento das diferenças estatísticas no desempenho dos diferentes modelos.

3.5 Máquina do Experimento

As especificações da máquina utilizada para os experimentos podem ser vistas pelo Quadro 4.

Quadro 4 – Máquina do Experimento

CPU	**GPU**	**RAM**	**Memória**	**Sistema Operacional**
Intel i7 4790K @4.00GHz	Titan V (12 GB) CUDA: 12	32 GB	1 TB	Linux

Fonte: Elaborado pelo autor.

4 Resultados Experimentais

Neste capítulo, apresenta-se os resultados da experimentação para validar os modelos sequenciais. Realizou-se uma análise detalhada do desempenho dos modelos sequenciais (RNN, LSTM e GRU) na tarefa de classificação de tarefas verbais e não verbais relacionadas à expressão facial no diagnóstico da ELA, conforme descrito no Capítulo 3. Além disso, discute-se as possíveis razões para os resultados obtidos, faremos comparações entre os modelos e explicaremos o critério utilizado para selecionar o modelo e a tarefa a ser usada no desenvolvimento da aplicação CAD.

Quadro 5 – Melhores resultados obtidos em cada tarefa

Tarefa	Classificação	Acurácia	Precisão	Sensibilidade	Pontuação F1
BBP	Repetição	38,90%	40,00%	23,70%	26,50%
	Sujeito	42,10%	42,10%	42,10%	42,10%
PA	Repetição	56,20%	52,40%	43,80%	46,10%
	Sujeito	60,00%	60,00%	60,00%	60,00%
PATAKA	Repetição	46,80%	42,90%	27,80%	31,80%
	Sujeito	45,00%	45,00%	45,00%	45,00%
SPREAD	Repetição	62,20%	68,20%	56,90%	60,30%
	Sujeito	61,90%	61,90%	61,90%	61,90%
KISS	**Repetição**	**78,00%**	**68,20%**	**66,00%**	**66,90%**
	Sujeito	**81,00%**	**81,00%**	**81,00%**	**81,00%**
OPEN	Repetição	75,20%	72,70%	65,20%	68,10%
	Sujeito	**81,00%**	**81,00%**	**81,00%**	**81,00%**
BLOW	Repetição	60,00%	61,50%	60,00%	60,70%
	Sujeito	66,70%	66,70%	66,70%	66,70%

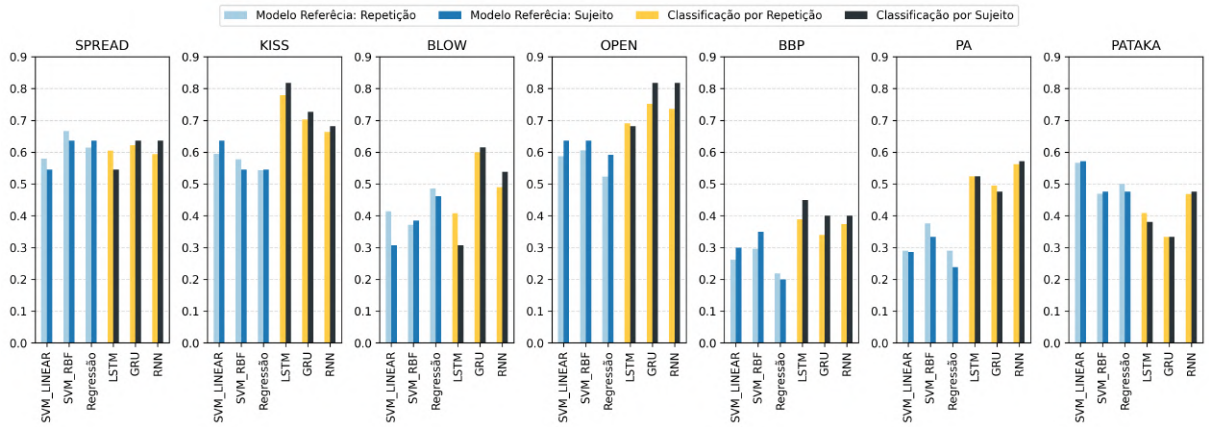
Fonte: Elaborado pelo autor.

Conforme destacado pelo Quadro 5, os resultados obtidos revelam que as tarefas "KISS"e "OPEN" se destacam como as mais distintivas entre as diferentes classes, alcançando uma acurácia de aproximadamente 81% na classificação por sujeito, o que implica na correta identificação de 18 indivíduos em um total de 22. Além disso, ambas as tarefas apresentaram o desempenho mais robusto na classificação por vídeo, com acurácias de cerca de 78% e 75%, respectivamente.

Por outro lado, as classes "BBP", "BLOW"e "PATAKA" demonstraram maior complexidade, embora tenham obtido resultados próximos aos modelos de referência. Com a exceção da tarefa "PATAKA" os modelos sequenciais revelaram-se mais eficazes em comparação com os modelos de referência.

Uma visão geral do desempenho dos modelos é apresentada pela Figura 16.

Figura 16 – Resultados obtido pelos modelos nas tarefa orofaciais não verbais e verbais

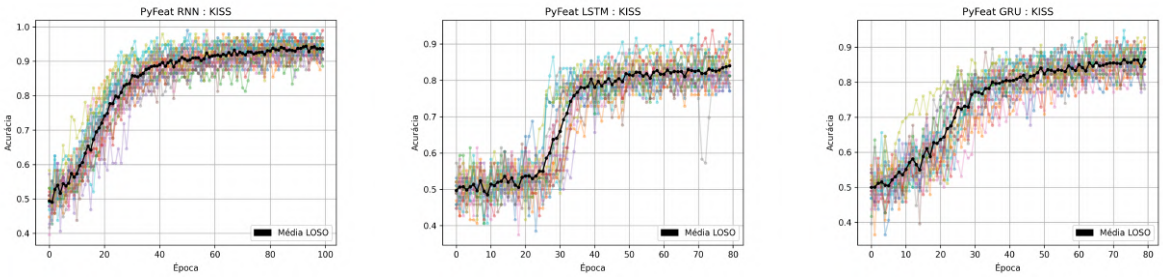


Fonte: Elaborada pelo autor.

4.1 Treinamento dos Modelos

Ao analisar a acurácia e a perda dos modelos ao longo do treinamento, levando em consideração a variação do grupo de validação, observa-se um padrão distinto na curva de aprendizado dos modelos que obtiveram os melhores resultados, como "OPEN", "KISS" e "SPREAD" em comparação aos demais modelos.

Figura 17 – Acurácia no treinamento dos modelos na tarefa "KISS".



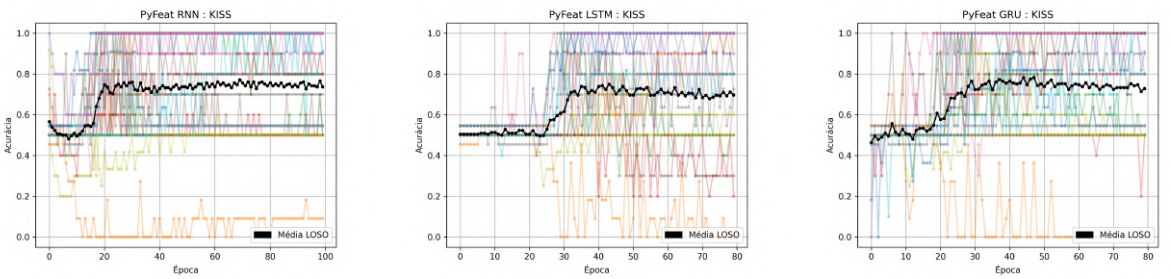
(a) RNN (b) LSTM (c) GRU

Fonte: Elaborado pelo autor.

Esses modelos, que se destacaram, demonstraram uma tendência interessante durante o treinamento. Notou-se que, em grande parte, eles mantiveram uma alta acurácia no conjunto de treinamento ao longo das épocas e, ao mesmo tempo, conseguiram manter desempenho consistente no conjunto de validação. Isso se reflete no gráfico de perda, que apresentou uma leve tendência decrescente, indicando que esses modelos não apenas melhoravam seu desempenho no conjunto de treinamento, mas também conseguiram aprimorar a acurácia no conjunto de validação. Esses resultados são congruentes com os testes realizados e confirmam a qualidade desses modelos.

As tarefas com resultados menos precisos, como a tarefa "PATAKA" demonstraram boa capacidade de aprendizado no conjunto de treinamento, conforme indicado

Figura 18 – Acurácia na validação dos modelos na tarefa "KISS".

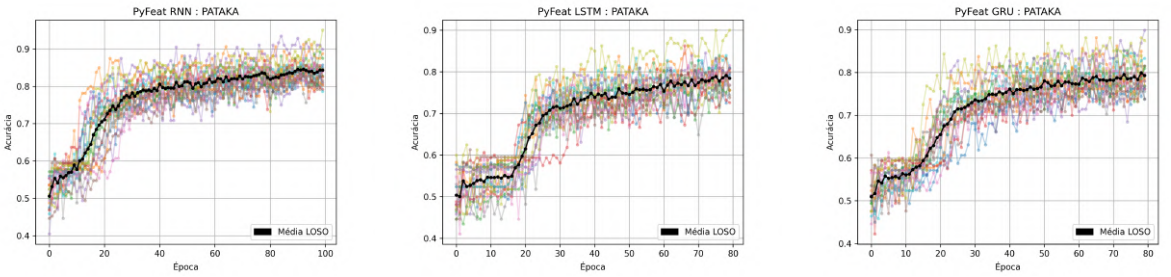


(a) RNN (b) LSTM (c) GRU

Fonte: Elaborado pelo autor.

pelos gráficos da Figura 19, mas essa capacidade não se refletiu na validação, como apresentado pela Figura 20, sugerindo um desafio na generalização do modelo para novos dados.

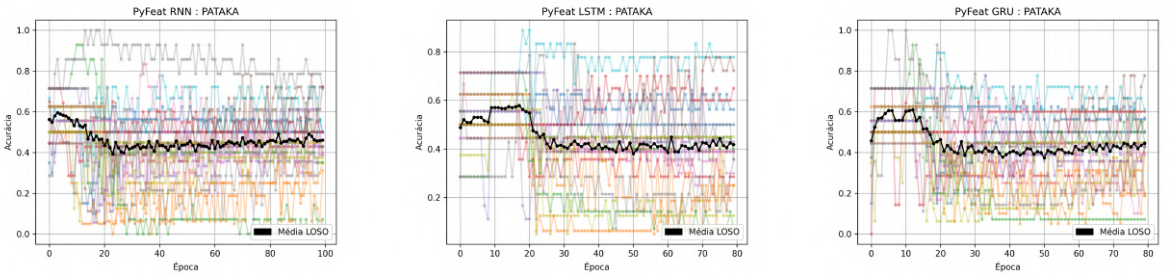
Figura 19 – Acurácia no treinamento dos modelos na tarefa "PATAKA".



(a) RNN (b) LSTM (c) GRU

Fonte: Elaborado pelo autor.

Figura 20 – Acurácia na validação dos modelos na tarefa "PATAKA".



(a) RNN (b) LSTM (c) GRU

Fonte: Elaborado pelo autor.

Esse mesmo padrão de comportamente se refletiu nas tarefas verbais "PA"e "BBP" e não verbal "BLOW". Várias razões podem explicar esses resultados durante a fase de treinamento. Isso inclui a possibilidade de sobreajuste, onde os modelos

se ajustam excessivamente aos dados de treinamento e não generalizam bem para novos dados. Além disso, pode ser que essas tarefas não sejam tão discriminativas quanto o esperado, o que pode limitar a capacidade dos modelos em identificar padrões significativos.

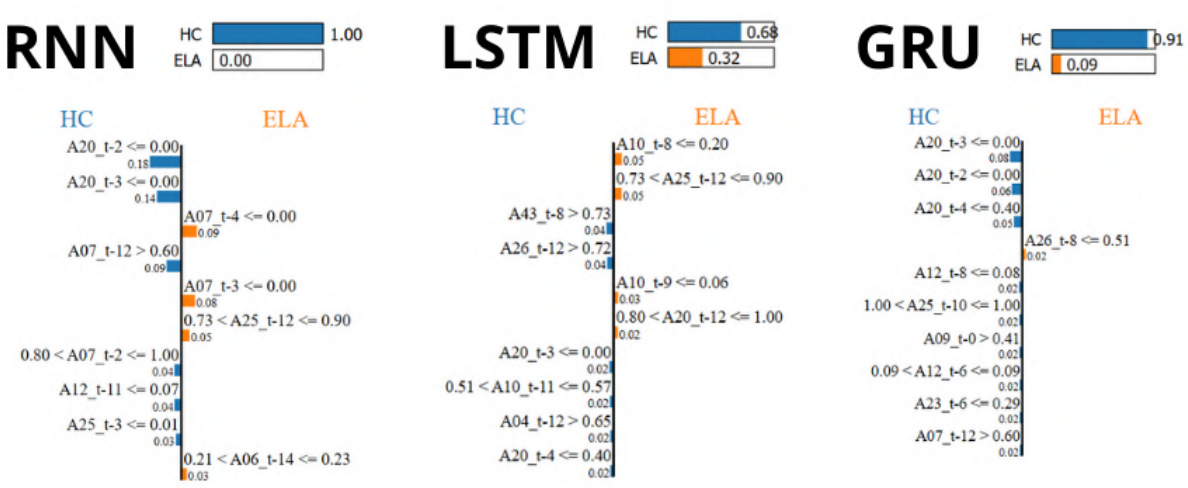
No entanto, é fundamental ressaltar que os resultados apresentados por Bandini et al. (2018) indicam que a tarefa "BBP" deveria ser uma das mais informativas. Portanto, existe a possibilidade de que a abordagem baseada em unidades de ação não seja capaz de extrair informações discriminativas da mesma maneira que os métodos manuais utilizados por eles, especificamente para essa tarefa. Isso é evidenciado pelo fato de os modelos que utilizam métodos tradicionais utilizados como referência terem alcançado resultados semelhantes aos dos modelos sequenciais.

Comparando o treinamento das três redes em cima da tarefa "KISS" e "PA", conseguiu-se compreender como os modelos se desenvolveram conforme o passar das épocas e como estavam conseguindo lidar com a generalização do aprendizado levando em consideração a validação.

4.2 Análise e Comparação entre Modelos

Comparando os resultados obtidos pelo modelo XAI, foi possível compreender quais eram as principais características observadas por cada modelo, conforme apresentado pela Figura 21.

Figura 21 – Interpretação do LIME em cada modelo de um exemplar da tarefa OPEN



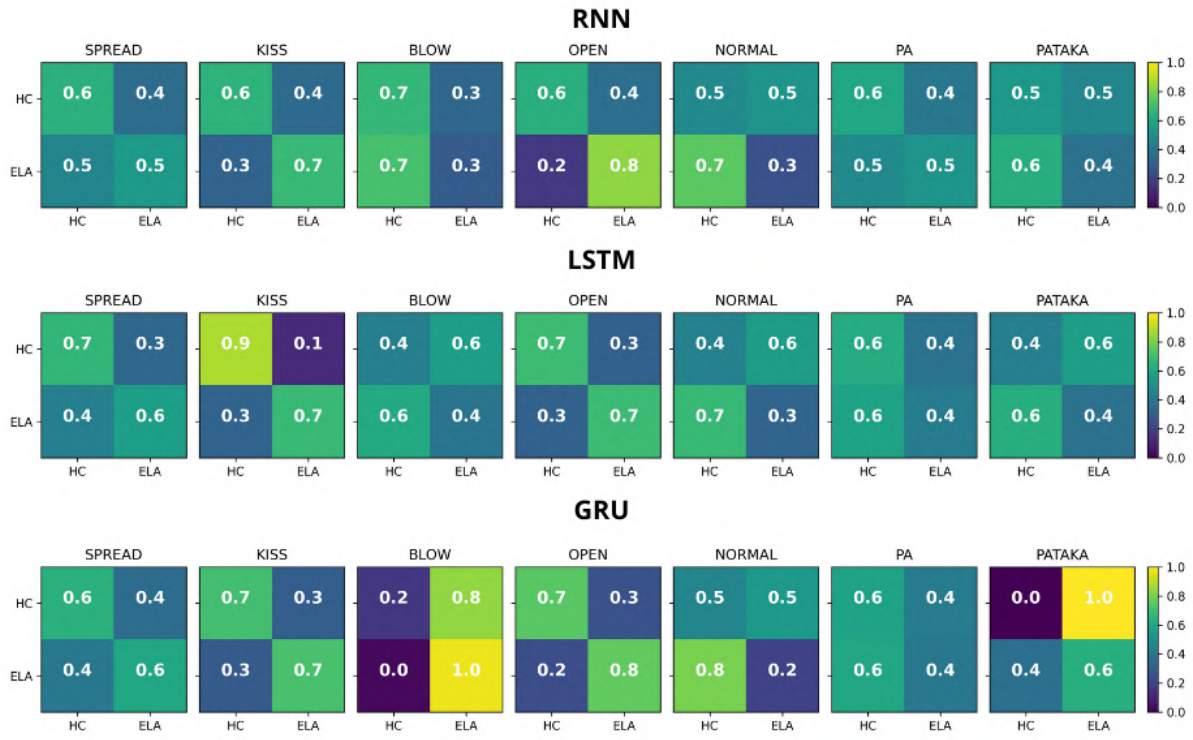
Fonte: Elaborada pelo autor.

É notável que, de maneira geral, os modelos compartilham uma similaridade na escolha dos pontos considerados na tomada de decisões. Um exemplo disso pode

ser observado na Figura 21, na qual todos os gráficos destacam a importância da AU20 (Alongador dos Lábios) no terceiro quadro para a classificação de indivíduos HC. Além disso, é interessante notar que os modelos RNN e GRU, os quais obtiveram resultados mais próximos dos valores reais, consideram características semelhantes, como a AU25 (Separação dos Lábios) e a AU12 (Elevador do Canto do Lábio). Ambas essas AUs fazem parte da região de interesse mencionada por Bandini et al. (2018) e Gomes et al. (2023), que se concentram na região dos lábios e da mandíbula. Essa convergência na seleção de características sugere a importância dessa região na diferenciação para a classificação.

Analisando a matriz de confusão da classificação por repetição de cada modelo temos para cada modelo:

Figura 22 – Matriz de confusão normalizada obtida por cada modelo



Fonte: Elaborado pelo autor.

Como forma de comparação, foi realizado o teste de Friedman. No entanto, ao aplicá-lo às diferentes medidas, nenhum deles atingiu um nível de significância menor ou igual a 0,05. Isso sugere que não há evidência estatística suficiente para concluir que existem diferenças significativas entre os modelos analisados. No entanto, ao adotar um critério de confiança mais baixo, com um nível de 10%, foi observado que duas tarefas, "BLOW" e "KISS", demonstraram diferenças significativas entre os modelos, com um valor de p em aproximadamente 0,097. Para uma avaliação mais aprofundada dessas diferenças, o teste de Nemenyi foi aplicado, resultando nos seguintes diagramas:

Figura 23 – Diagrama de diferença crítica em relação a Pontuação F1.

(a) Classificação por Repetição KISS (b) Classificação por Sujeito BLOW

Fonte: Elaborado pelo autor.

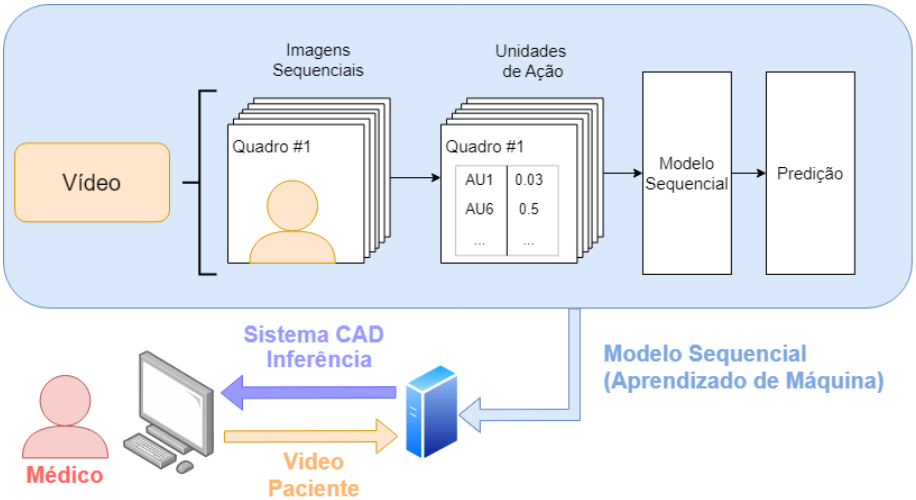
Comparando com os outros modelos sequenciais, a RNN demonstrou maior consistência em termos de acurácia. Vários fatores podem explicar esse desempenho, incluindo o curto contexto temporal (apenas 15 quadros por vídeo), o que não exige uma memória extensa, bem como a menor profundidade do modelo, o que o torna menos suscetível ao sobreajuste. No entanto, ao considerar os resultados dos testes pós-hoc, chegou-se a conclusão da implementação dos três modelos para a etapa final deste trabalho. Isso se deve ao fato de que os resultados variaram consideravelmente de acordo com a tarefa e não apresentaram uma diferença significativa real (conforme indicado pelo teste de Friedman). A inclusão dos três modelos proporcionará uma interpretação mais abrangente dos critérios de decisão, uma vez que levará em conta a análise de três modelos. Por fim, ao considerar os resultados apresentados pelos modelos de referência e sequenciais, e ao levar em consideração a escolha de uma tarefa mais discriminativa, concluiu-se que a tarefa "OPEN" seria a mais apropriada, uma vez que obteve o melhor desempenho.

5 Aplicação de Auxílio ao Diagnóstico

O objetivo deste estudo não se limita à análise dos modelos sequenciais (RNN, LSTM e GRU) para a classificação da doença neurodegenerativa ELA. Além disso, visa a construção de uma aplicação prática de auxílio ao diagnóstico que seja capaz de fornecer uma segunda opinião ao profissional de saúde. Essa aplicação será baseada nos modelos desenvolvidos e treinados neste trabalho, conforme apresentado no Capítulo 4.

A aplicação foi desenvolvida como uma plataforma *offline*, com o intuito de oferecer maior segurança aos dados sensíveis fornecidos pelos usuários, garantindo o acesso somente por meio de um servidor local. A escolha de uma aplicação web visa facilitar o acesso a profissionais de saúde em diversos dispositivos, permitindo o uso por meio de navegadores padrão, sem a necessidade de instalações complexas ou requisitos de hardware específicos. A visão geral do processo da aplicação pode ser vista na Figura 24.

Figura 24 – Visão Geral Aplicação CAD



Fonte: Elaborada pelo autor.

A proposta é que os profissionais de saúde gravem e carreguem vídeos de pacientes na aplicação. A partir disso, eles poderão obter uma melhor visualização do paciente por meio da extração e análise dos modelos de aprendizado de máquina. Os resultados incluirão a visualização das características extraídas e a interpretação das inferências feitas pelo modelo LIME em relação aos critérios de decisão mais relevantes para o diagnóstico. É crucial ressaltar que essa aplicação não tem a intenção de substituir a avaliação clínica de um profissional de saúde qualificado, mas sim de oferecer informações adicionais destinadas a auxiliar na tomada de decisões. A confirmação

final do diagnóstico ainda dependerá da avaliação de um médico especializado.

Dessa forma, a criação da aplicação de auxílio ao diagnóstico é uma extensão prática deste estudo, com o potencial de impactar positivamente a detecção automática da ELA e servir como base para estudos sobre a aplicação em outras doenças neurodegenerativas. Além disso, a aplicação oferece ferramentas integradas e de fácil manuseio para o processamento de dados em vídeo e a extração de AUs, fornecendo material para uso além do aqui abordado. As etapas de desenvolvimento, bem como a interface final do programa desenvolvido, são apresentadas nas Seções 5.1 e 5.2.

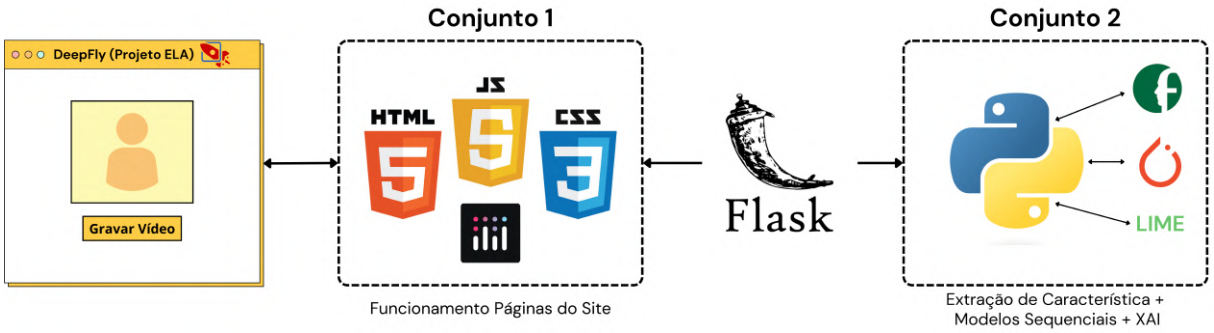
5.1 Integração de Modelos Sequenciais e Visualização Interativa

O desenvolvimento da aplicação pode ser dividido em três partes principais. A primeira parte lida com a comunicação entre os modelos sequenciais e a página web. A segunda parte aborda a geração de gráficos para tornar os resultados e a visualização das etapas de extração de características intuitivos e bem explicativos. Por fim, a terceira etapa envolve a combinação de todos os módulos desenvolvidos em um único sistema de fácil instalação e uso.

Para a captura de vídeos diretamente no site, utilizou-se a funcionalidade incorporada do JavaScript, conhecida como MediaRecorder. Essa abordagem foi escolhida devido ao alto custo computacional envolvido na tentativa de enviar cada quadro do vídeo diretamente para o Python por meio do *socket* implementado no Flask. Inicialmente, planejou-se capturar os vídeos no código Python e, em seguida, processar os quadros, porém, percebeu-se que o processo de envio de todos os quadros capturados pela câmera para o Python era excessivamente exigente em termos de recursos computacionais. Portanto, a solução mais eficaz foi gravar o vídeo usando o MediaRecorder e, posteriormente, transmiti-lo como uma string codificada em *base64* para o Python. Essa abordagem provou ser mais eficiente e viável do ponto de vista computacional.

Dado que os modelos foram implementados usando a biblioteca PyTorch, foi necessário estabelecer a troca de informações de entrada e saída entre os modelos e a aplicação. Optou-se pelo uso do Flask para facilitar essa intermediação. O Flask é um framework que permite a comunicação entre duas linguagens operacionais incorporadas no site, JavaScript e Python, por meio de sockets. No site (Figura 25), há dois componentes operacionais: o primeiro desempenha um papel fundamental na parte visual do site e na resposta às interações dos usuários nas páginas, enquanto o segundo é amplamente controlado pela linguagem Python para lidar com os modelos de aprendizado de máquina e extração de características.

Figura 25 – Comunicação entre bibliotecas e linguagens do sitema



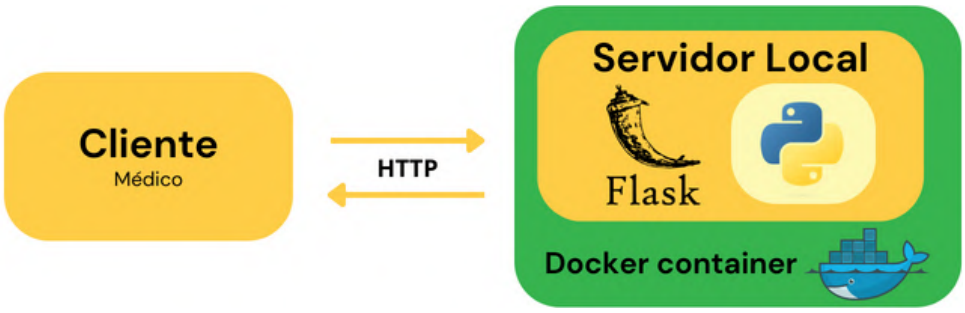
Fonte: Elaborada pelo autor.

É importante destacar que tanto a extração das características usadas no processo de inferência quanto a própria inferência são executadas por meio da linguagem Python. A linguagem Python também foi utilizada no treinamento dos modelos desenvolvidos neste trabalho. Durante o treinamento, os pesos dos modelos são salvos e, posteriormente, carregados para a avaliação dos vídeos gravados, garantindo a consistência das análises e a coerência dos resultados.

Com o intuito de fornecer retorno da informação coletada ao usuário, durante o processo de extração de características, utilizou-se a biblioteca Plotly. O Plotly é uma biblioteca de visualização de dados versátil, compatível com Python, JavaScript e R. Nossa aplicação a utiliza em conjunto com JavaScript para criar gráficos interativos que exibem as AUs ao longo do tempo. Além disso, o Plotly é empregado na página de inferência, facilitando a interação e visualização das AUs consideradas relevantes para a classificação da doença.

Para simplificar a distribuição e a execução da aplicação web, incorporou-se o Docker. O Docker consolida todos os arquivos e recursos necessários, incluindo as bibliotecas instaladas no ambiente de execução, em uma única imagem de contêiner. Os contêineres isolam o software de seu ambiente, garantindo que ele funcione de maneira uniforme, independentemente das diferenças entre ambientes, como o de desenvolvimento e o de produção. O Docker usa um formato padronizado para empacotar esses elementos e está disponível tanto para aplicativos baseados em Linux quanto em Windows. Isso garante que o software em contêiner sempre será executado da mesma forma, independentemente da infraestrutura, proporcionando consistência na execução da aplicação, independentemente do ambiente em que é executada. O resultado final é a construção implementada da aplicação, conforme mostrado na Figura 26.

Figura 26 – Estapa final do processo de implantação dos modelos sequenciais na aplicação CAD.

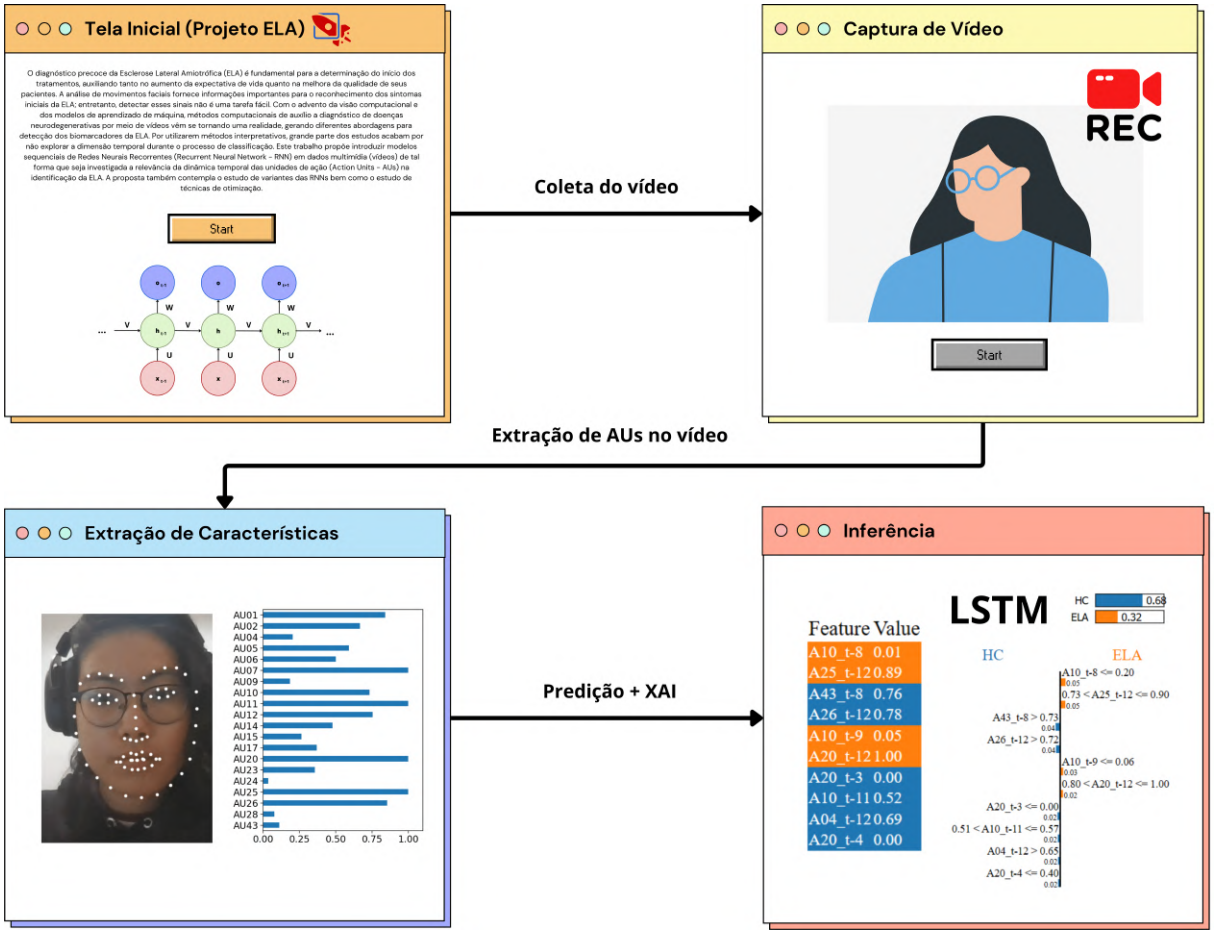


Fonte: Adaptada de (HANDLIN, 2022).

5.2 Produto Final

Como resultado tem-se a implementação do seguinte fluxo de processos, conforme ilustrado pela Figura 27. As interfaces da aplicação se dividem em 4 páginas principais, além de interfaces adicionais, como a tela de carregamento.

Figura 27 – Mockup com o fluxo de eventos da aplicação Web implementada

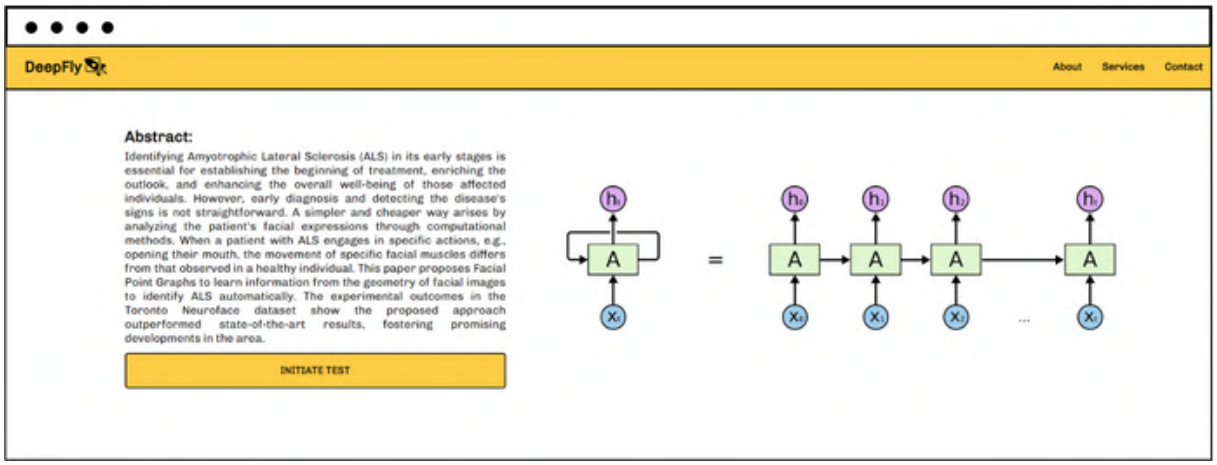


Fonte: Elaborada pelo autor.

As interfaces finais do site podem ser visualizadas nas próximas figuras. A

primeira página (Figura 28) do projeto consiste em uma interface para introdução, apresentando um resumo do problema e fornecendo uma explicação visual do foco do modelo adotado neste trabalho.

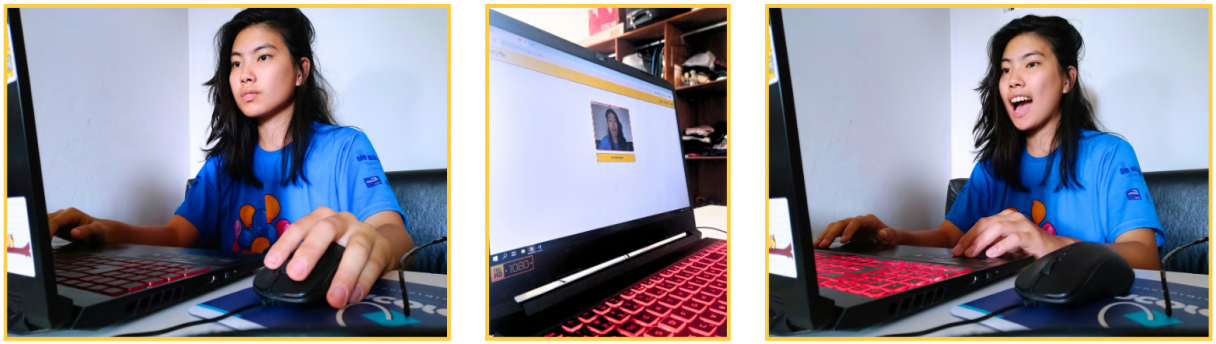
Figura 28 – Interface Tela Principal de Introdução ao Sistema CAD.



Fonte: Elaborada pelo autor.

Como segunda interface temos a tela de captura. Nessa parte, é realizada a gravação de vídeos pelo próprio navegador utilizando a ferramenta MediaRecorder, como mencionado na seção anterior. Uma simulação é apresentada na Figura 29.

Figura 29 – Exemplo de Uso

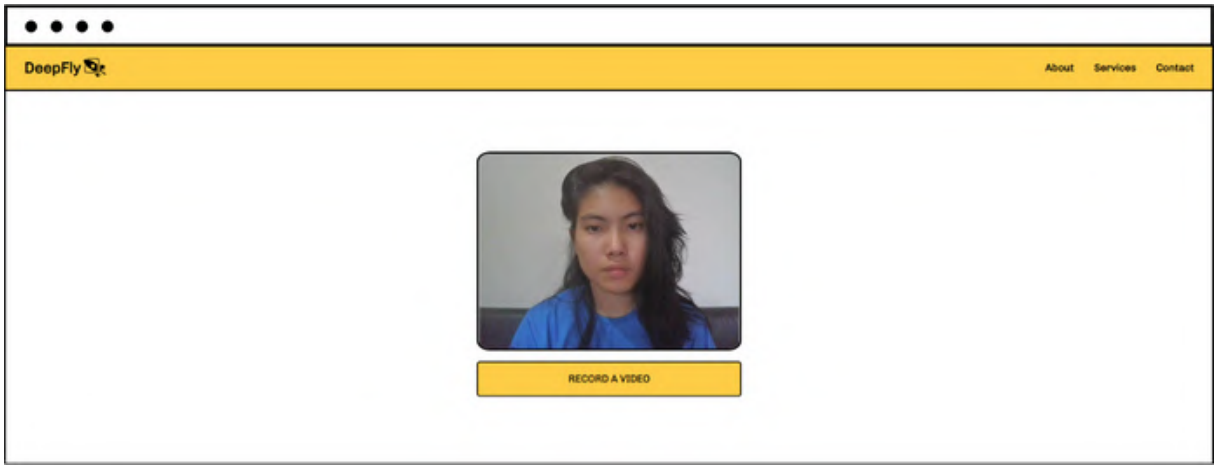


Fonte: Elaborada pelo autor.

A largura e a altura da área de captura foram definidas de acordo com um padrão para facilitar a captura do rosto principal, o que limita as dimensões, conforme ilustrado pela Figura 30. Nessa página, o usuário pode gravar seu vídeo, e ao final é possível visualizar a gravação, tendo a opção de escolher regravar ou prosseguir para a etapa de extração de características.

A interface de extração de características (Figura 31) demonstra a biblioteca Plotly em ação. No centro da interface, há um gráfico gerado por essa biblioteca que exibe todas as unidades de ação coletadas ao longo do tempo, permitindo que o usuário interaja e examine os valores de forma iterativa. Entre o processo de extração

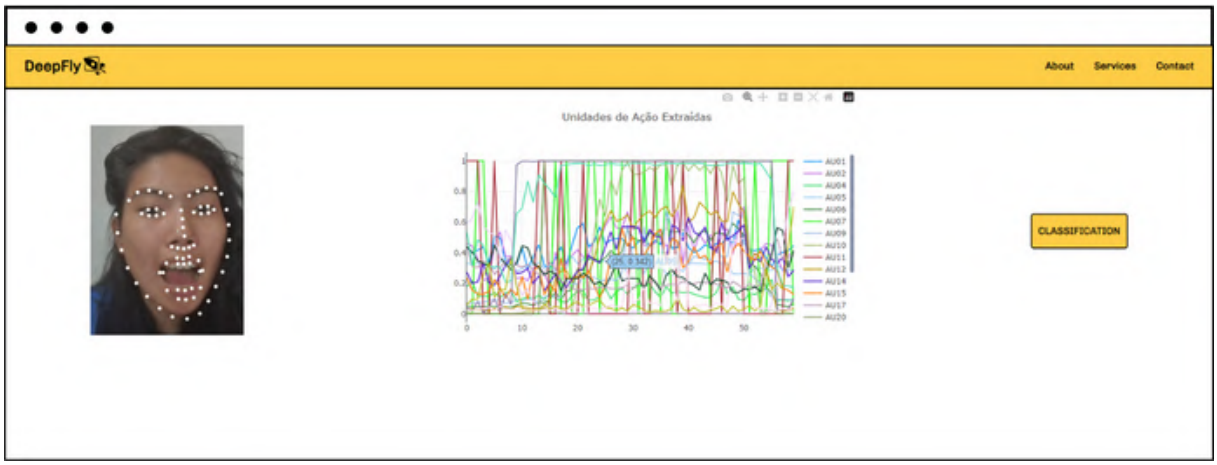
Figura 30 – Interface Tela de Captura de Vídeo.



Fonte: Elaborada pelo autor.

e o carregamento dessa página, os valores de entrada do modelo são salvos em um arquivo *.json*, proporcionando um *backup* no caso de a página ser recarregada. No canto esquerdo, também é apresentada uma animação em formato GIF gerada durante pelo processo de extração, destacando os pontos faciais detectados pela ferramenta Py-feat. Isso permite que o usuário identifique eventuais inconsistências na imagem de referência e opte por voltar a uma página anterior para regravar, se necessário.

Figura 31 – Interface Tela de Extrações de Características.

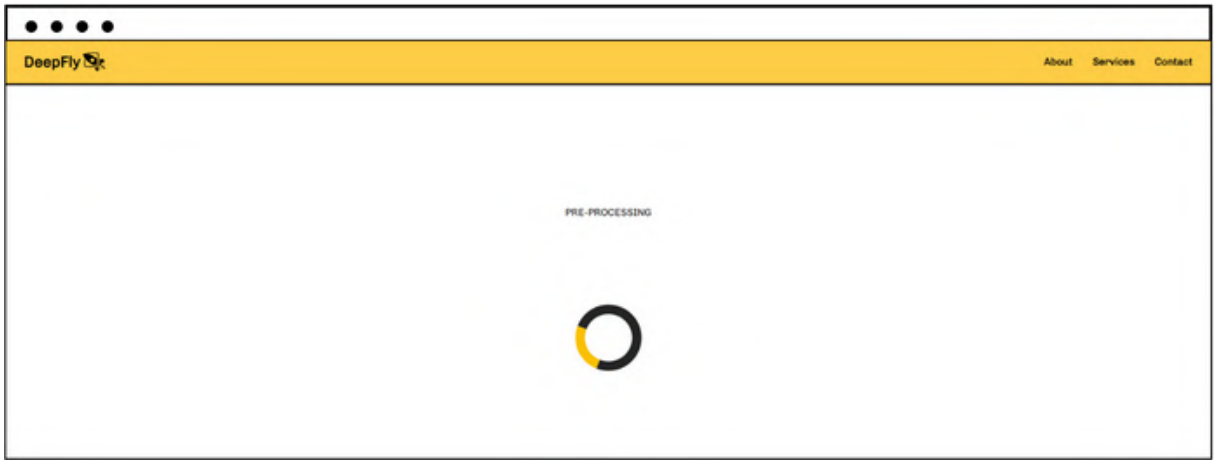


Fonte: Elaborada pelo autor.

Como o processo de extração de características é computacionalmente custoso, uma vez que é analisado quadro a quadro de vídeos coletados, essa etapa pode levar um tempo considerável. Para evitar que o usuário mude de página ou interrompa o processo de extração de características, foi implementada uma página adicional de carregamento (Figura 32) que mantém o processo de extração em execução em segundo plano. Isso garante que o usuário possa aguardar o término da extração de

características sem interrupções.

Figura 32 – Interface Tela de Carregamento.

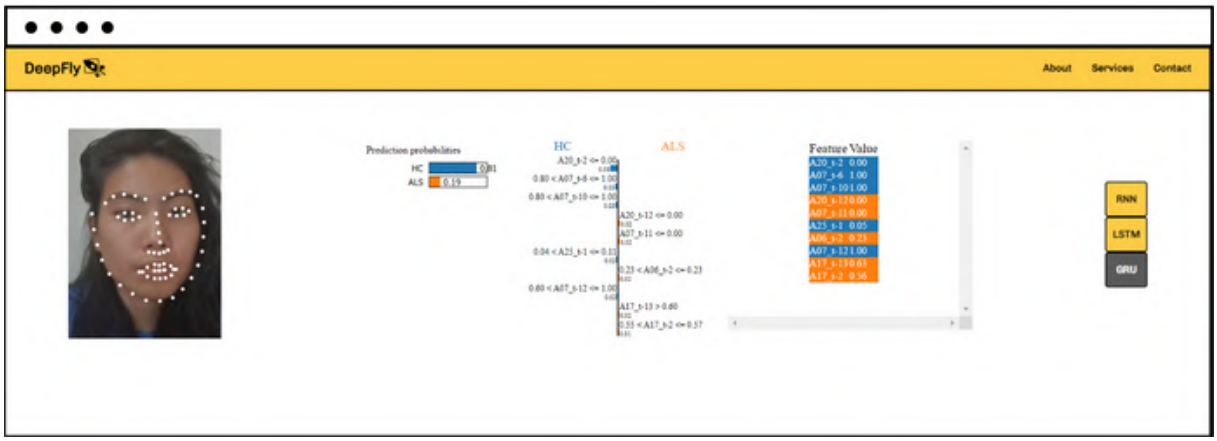


Fonte: Elaborada pelo autor.

Essa tela foi adotada levando em consideração que realizar muitas interrupções de processo podem pesar para o servidor local, atrapalhando no desempenho geral da experiência como usuário. A inclusão dessa tela de carregamento foi uma escolha importante, pois a interrupção frequente do processo de extração de características pode sobrecarregar o servidor local e prejudicar o desempenho geral da experiência do usuário.

Na última página (Figura 33), são apresentados os resultados obtidos por cada modelo sequencial, oferecendo uma visão mais abrangente do que a análise de um único modelo.

Figura 33 – Interface Tela de Predições e Interpretação do modelo XAI.



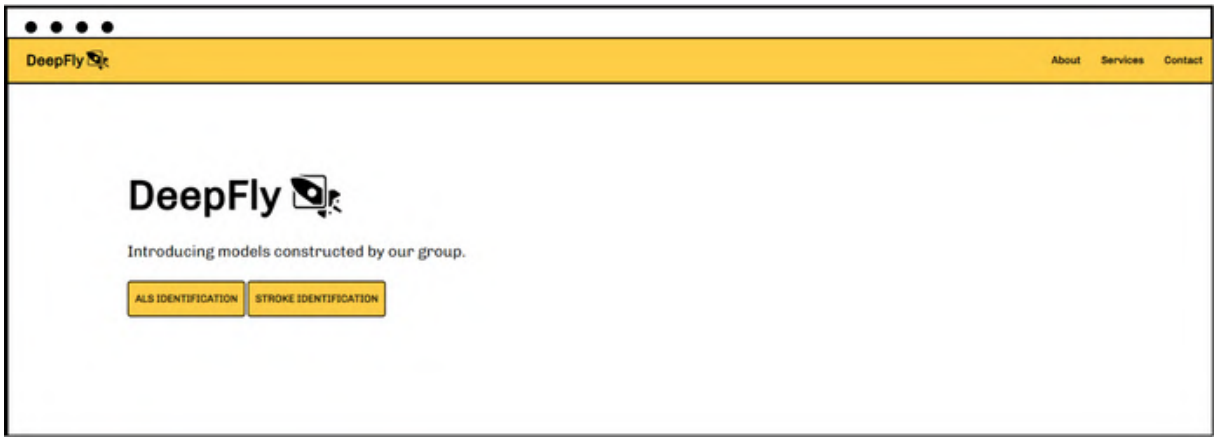
Fonte: Elaborada pelo autor.

Além disso, são exibidos os gráficos gerados pelo LIME, nos quais os valores considerados mais relevantes - ou seja, a AU e o tempo correspondente - indicam

quais foram os principais critérios adotados na decisão de classificação da doença por aquele modelo sequencial.

A estrutura da aplicação foi projetada para acomodar não apenas o diagnóstico da ELA, mas também outras doenças que possam utilizar a análise facial como parte do processo de diagnóstico. Essa estrutura foi concebida de maneira flexível e expansível, permitindo a integração de novos trabalhos e pesquisas à medida que novas metodologias surgirem. Esse grupo de trabalhos no campo de aprendizado de máquina é carinhosamente chamado de "DeepFly", representando os voos de sucesso em direção aos próximos passos na pesquisa no campo do aprendizado profundo.

Figura 34 – Tela Principal DeepFly



Fonte: Elaborada pelos membros do grupo de pesquisa.

6 Considerações Finais

Neste trabalho, abordamos a criação de uma aplicação prática voltada para a detecção da doença neurodegenerativa ELA, uma doença desafiadora, caracterizada por atrasos no diagnóstico e falsos positivos devido à sua natureza complexa e às diversas abordagens subjetivas existentes. Observando pesquisas relacionadas à doença, ficou evidente que o uso de expressões faciais na análise de tarefas orofaciais por meio de vídeos surge como um biomarcador promissor tanto para diagnóstico quanto para acompanhamento médico. O reconhecimento automático dessas expressões pode acelerar o processo de diagnóstico e reduzir custos.

Assim, este trabalho adotou uma abordagem centralizada em modelos de aprendizado profundo sequenciais, com ênfase nas redes neurais recorrentes, para análise das unidades de ação. Isso possibilitou não apenas a análise espacial, mas também a análise temporal na classificação de vídeos de tarefas orofaciais com o objetivo de identificar a ELA.

A aplicação desenvolvida teve como objetivo fornecer uma valiosa ferramenta de segunda opinião para profissionais de saúde. Ela oferece suporte na detecção e classificação da ELA, por meio da inferência dos modelos estudados, bem como informações sobre a interpretabilidade da inferência. A abordagem visa fornecer uma solução de baixo custo e amplamente acessível. No entanto, é essencial salientar que esta aplicação, apesar da estrutura já implementada, representa uma fase inicial no desenvolvimento, já que os modelos empregados não atingiram o estado da arte na classificação da ELA, sendo assim tendo caráter exclusivamente acadêmico e de pesquisa.

É fundamental enfatizar que a aplicação não foi desenvolvida para ser vista como uma substituição para a avaliação clínica de um médico especializado. Em vez disso, ela fornece informações adicionais destinadas a auxiliar na tomada de decisões, enfatizando o papel crucial do profissional de saúde qualificado no processo de diagnóstico.

Ademais, é importante destacar que, devido às limitações do banco de dados utilizado, que possui um número reduzido de dados disponíveis, os resultados obtidos neste estudo podem não ser totalmente representativos e não têm a capacidade de abranger e generalizar os biomarcadores da doença com alta confiabilidade. Essa falta de dados em quantidade suficiente pode afetar a capacidade do modelo de identificar com precisão os padrões relacionados à ELA, uma vez que a generalização só pode ser alcançada com uma base de dados mais extensa e diversificada. Portanto, é essencial

considerar essa limitação ao interpretar os resultados deste trabalho.

A aplicação desenvolvida apresenta grande potencial no campo de divulgação do auxílio ao diagnóstico e acompanhamento de doenças neurodegenerativas, especialmente a ELA. Além disso, as ferramentas de pré-processamento e extração de características de vídeo implementadas podem servir como base para futuros estudos e pesquisas, mesmo em outros domínios em que a análise facial seja considerada.

O projeto foi pensado com modularidade em mente, assim, permitindo futuras expansões e integrações de novas metodologias. O nome "DeepFly"foi adotado para representar um conjunto de trabalhos no campo de aprendizado profundo, e, portanto, esta aplicação faz parte desse grupo.

Como trabalho futuro, destacamos a oportunidade de melhorar a segurança na coleta de dados para disponibilizar a aplicação de forma segura online, além de expandir a gama de ferramentas disponíveis. Também é relevante obter *feedback* de profissionais de saúde para aprimorar ainda mais a aplicação e adaptá-la às necessidades clínicas específicas, bem como obter um maior número de dados. A busca pela melhoria contínua dos modelos e dos métodos adotados é fundamental para que essa aplicação possa contribuir de maneira significativa para o diagnóstico precoce e o acompanhamento da ELA e, possivelmente, de outras doenças neurodegenerativas no futuro.

Referências

ALI, M. R.; MYERS, T.; WAGNER, E.; RATNU, H.; DORSEY, E.; HOQUE, E. Facial expressions can detect parkinson's disease: Preliminary evidence from videos collected online. *NPJ digital medicine*, Nature Publishing Group, v. 4, n. 1, p. 1–4, 2021. Disponível em: https://www.nature.com/articles/s41746-021-00502-8. Acesso em: 02 Novembro 2023.

ASSOCIATION, A. *FACTSHEET from the ALS ASSOCIATION*. 2020. Disponível em: https://www.als.org/navigating-als/resources/fyi-second-opinion-faqs. Acesso em: 31 Outubro 2023.

BALTRUSAITIS, T.; ZADEH, A.; LIM, Y. C.; MORENCY, L.-P. Openface 2.0: Facial behavior analysis toolkit. In: IEEE. *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. 2018. p. 59–66. Disponível em: https://ieeexplore.ieee.org/document/8373812. Acesso em: 02 Novembro 2023.

BANDINI, A.; GREEN, J. R.; TAATI, B.; ORLANDI, S.; ZINMAN, L.; YUNUSOVA, Y. Automatic detection of amyotrophic lateral sclerosis (als) from video-based analysis of facial movements: speech and non-speech tasks. In: IEEE. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 2018. p. 150–157. Disponível em: https://ieeexplore.ieee.org/document/8373824. Acesso em: 02 Novembro 2023.

BANDINI, A.; REZAEI, S.; GUARIN, D. L.; KULKARNI, M.; LIM, D.; BOULOS, M. I.; ZINMAN, L.; YUNUSOVA, Y.; TAATI, B. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 25, n. 4, p. 1111–1119, 2020. Disponível em: https://ieeexplore.ieee.org/document/9177259. Acesso em: 02 Novembro 2023.

BROOKS, B. R.; MILLER, R. G.; SWASH, M.; MUNSAT, T. L. El escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic lateral sclerosis and other motor neuron disorders*, Taylor & Francis, v. 1, n. 5, p. 293–299, 2000. Disponível em: https://pubmed.ncbi.nlm.nih.gov/11464847/. Acesso em: 02 Novembro 2023.

CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014. Disponível em: https://arxiv.org/abs/1406.1078. Acesso em: 02 Novembro 2023.

EKMAN, P.; FRIESEN, W. V. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978. Disponível em: https://doi.org/10.1037/t27734-000. Acesso em: 02 Novembro 2023.

FARNSWORTH, B. *Facial Action Coding System (FACS) – A Visual Guidebook*. 2022. Disponível em: https://engineering.rappi.com/serve-your-first-model-with-scikit-learn-flask-docker-df95efbbd35e. Acesso em: 02 Novembro de 2023.

FOGELMAN-SOULIé, F.; CUN, Y. L. Modèles connexionnistes de l'apprentissage. *Intellectica*, v. 2, n. 1, p. 114–143, 1987. Included in a thematic issue : Apprentissage et machine. Disponível em: https://www.persee.fr/doc/intel_0769-4113_1987_num_2_1_1804. Acesso em: 02 Novembro 2023.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with lstm. *Neural computation*, MIT Press, v. 12, n. 10, p. 2451–2471, 2000. Disponível em: https://direct.mit.edu/neco/article-abstract/12/10/2451/6415/Learning-to-Forget-Continual-Prediction-with-LSTM?redirectedFrom=fulltext. Acesso em: 02 Novembro 2023.

GOMES, N. B.; YOSHIDA, A.; RODER, M.; OLIVEIRA, G. C. de; PAPA, J. P. Facial point graphs for amyotrophic lateral sclerosis identification. *arXiv preprint arXiv:2307.12159*, 2023. Disponível em: https://arxiv.org/abs/2307.12159. Acesso em: 03 Novembro 2023.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. Disponível em: http://www.deeplearningbook.org. Acesso em: 02 Novembro 2023.

HAMM, J.; KOHLER, C. G.; GUR, R. C.; VERMA, R. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of neuroscience methods*, Elsevier, v. 200, n. 2, p. 237–256, 2011. Disponível em: https://pubmed.ncbi.nlm.nih.gov/21741407/. Acesso em: 02 Novembro 2023.

HANDLIN, C. W. *Serve your first model with Scikit-Learn + Flask + Docker*. 2022. Disponível em: https://engineering.rappi.com/serve-your-first-model-with-scikit-learn-flask-docker-df95efbbd35e. Acesso em: 02 Novembro de 2023.

HEBB, D. *The organization of behavior. EmphNew york*. [S.l.]: Wiley, 1949.

HESTERLEE, S. *Early Diagnosis of ALS Benefits Patients*. 2022. Urlhttps://www.contemporaryclinic.com/view/early-diagnosis-of-als-benefits-patients. Acesso em: 31 Outubro 2023.

HINTON, G. E. Distributed representations. Carnegie Mellon University, 1984.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT press, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory?redirectedFrom=fulltext. Acesso em: 02 Novembro 2023.

JOLLY, E.; CHEONG, J. H.; XIE, T.; BYRNE, S.; KENNY, M.; CHANG, L. J. Py-feat: Python facial expression analysis toolbox. *arXiv preprint arXiv:2104.03509*, 2021. Disponível em: https://arxiv.org/abs/2104.03509. Acesso em: 02 Novembro 2023.

LI, T.-H. S.; KUO, P.-H.; TSAI, T.-N.; LUAN, P.-C. Cnn and lstm based facial expression analysis model for a humanoid robot. *IEEE Access*, IEEE, v. 7, p. 93998–94011, 2019. Disponível em: https://ieeexplore.ieee.org/document/8760246. Acesso em: 02 Novembro 2023.

LIU, G.; GUO, J. Bidirectional lstm with attention mechanism and convolutional layer for text classification. *Neurocomputing*, Elsevier, v. 337, p. 325–338, 2019. Disponível em: https://doi.org/10.1016/j.neucom.2019.01.078. Acesso em: 02 Novembro 2023.

MARCHI, F. D.; CONTALDI, E.; MAGISTRELLI, L.; CANTELLO, R.; COMI, C.; MAZZINI, L. Telehealth in neurodegenerative diseases: opportunities and challenges for patients and physicians. *Brain Sciences*, MDPI, v. 11, n. 2, p. 237, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/33668641/. Acesso em: 02 Novembro 2023.

MAVADATI, M.; SANGER, P.; MAHOOR, M. H. Extended disfa dataset: Investigating posed and spontaneous facial expressions. In: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*. [s.n.], 2016. p. 1–8. Disponível em: https://ieeexplore.ieee.org/document/7789672. Acesso em: 02 Novembro 2023.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943. Disponível em: https://link.springer.com/article/10.1007/BF02478259. Acesso em: 02 Novembro 2023.

MITCHELL, T. M. *Machine learning*. 1997.

MIYAZAKI, H. *Porco Rosso*. [S.l.]: Studio Ghibl, 1992.

OGAWA, T.; SASAKA, Y.; MAEDA, K.; HASEYAMA, M. Favorite video classification based on multimodal bidirectional lstm. *IEEE Access*, IEEE, v. 6, p. 61401–61409, 2018. Disponível em: https://ieeexplore.ieee.org/document/8496751. Acesso em: 02 Novembro 2023.

OLIVEIRA, L. S. d. *Auxílio ao diagnóstico de ELA e AVC através de expressão facial*. Dissertação (Monografia) — Universidade Estadual Paulista (Unesp), Bauru/SP, 2022. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/3c42a5c4-3ffe-4d0f-ab44-987797d62411/content. Acesso em: 02 Novembro 2023.

PONTES, R. T.; ORSINI, M.; FREITAS, M. R. de; ANTONIOLI, R. de S.; NASCIMENTO, O. J. Alterações da fonação e deglutição na esclerose lateral amiotrófica: revisão de literatura. *Revista Neurociências*, v. 18, n. 1, p. 69–73, 2010. Disponível em: https://periodicos.unifesp.br/index.php/neurociencias/article/view/8505/6039. Acesso em: 02 Novembro 2023.

QIANG, J.; WU, D.; DU, H.; ZHU, H.; CHEN, S.; PAN, H. Review on facial-recognition-based applications in disease diagnosis. *Bioengineering*, MDPI, v. 9, n. 7, p. 273, 2022. Disponível em: https://pubmed.ncbi.nlm.nih.gov/35877324/. Acesso em: 02 Novembro 2023.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should I trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. [s.n.], 2016. p. 1135–1144. Disponível em: https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf. Acesso em: 02 Novembro 2023.

RIBEIRO, S. S. M. T.; GUESTRIN, C. *Local Interpretable Model-Agnostic Explanations (LIME): An Introduction*. 2016. Disponível em: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/. Acesso em: 02 Novembro de 2023.

ROBINSON, A. J.; FALLSIDE, F. *The utility driven dynamic error propagation network*. University of Cambridge Department of Engineering Cambridge, 1987. v. 1. Disponível em: https://gwern.net/doc/ai/nn/rnn/1987-robinson.pdf. Acesso em: 02 Novembro 2023.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Disponível em: https://www.nature.com/articles/323533a0. Acesso em: 02 Novembro 2023.

SCHLINDWEIN-ZANINI, R.; QUEIROZ, L. P.; CLAUDINO, L. S.; CLAUDINO, R. Aspectos neuropsicológicos da esclerose lateral amiotrófica: relato de caso. *Arquivos Catarinenses de Medicina*, v. 44, n. 1, p. 62–70, 2015. Disponível em: https://revista.acm.org.br/index.php/arquivos/article/view/11. Acesso em: 02 Novembro 2023.

WIDROW, B.; HOFF, M. E. et al. Adaptive switching circuits. In: NEW YORK. *IRE WESCON convention record*. 1960. v. 4, n. 1, p. 96–104. Disponível em: https://www-isl.stanford.edu/~widrow/papers/c1960adaptiveswitching.pdf. Acesso em: 02 Novembro 2023.

WINKEL, D. J.; TONG, A.; LOU, B.; KAMEN, A.; COMANICIU, D.; DISSELHORST, J. A.; RODRÍGUEZ-RUIZ, A.; HUISMAN, H.; SZOLAR, D.; SHABUNIN, I. et al. A novel deep learning based computer-aided diagnosis system improves the accuracy and efficiency of radiologists in reading biparametric magnetic resonance images of the prostate: results of a multireader, multicase study. *Investigative radiology*, LWW, v. 56, n. 10, p. 605–613, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/33787537/. Acesso em: 02 Novembro 2023.

WRIGHT, A.; DAMSKÄGG, E.-P.; JUVELA, L.; VÄLIMÄKI, V. Real-time guitar amplifier emulation with deep learning. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 3, p. 766, 2020. Disponível em: https://www.mdpi.com/2076-3417/10/3/766. Acesso em: 02 Novembro 2023.

XU, R.-S.; YUAN, M. Considerations on the concept, definition, and diagnosis of amyotrophic lateral sclerosis. *Neural Regeneration Research*, Wolters Kluwer–Medknow Publications, v. 16, n. 9, p. 1723, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/33510061/. Acesso em: 02 Novembro 2023.

YU, Y.; SI, X.; HU, C.; ZHANG, J. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 31, n. 7, p. 1235–1270, 2019. Disponível em: https://doi.org/10.1162/neco_a_01199. Acesso em: 02 Novembro 2023.

ZIMMERMAN, E. K.; ESLINGER, P. J.; SIMMONS, Z.; BARRETT, A. M. Emotional perception deficits in amyotrophic lateral sclerosis. *Cognitive and Behavioral Neurology*, LWW, v. 20, n. 2, p. 79–82, 2007. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1905862/. Acesso em: 02 Novembro 2023.