

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

NICOLAS BARBOSA GOMES

**DIAGNÓSTICO AUTOMÁTICO DE AVC PÓS-TRAUMA COM BASE NA  
FACE DE PACIENTES UTILIZANDO REDES NEURAIS EM GRAFOS**

BAURU  
Novembro/2023

NICOLAS BARBOSA GOMES

**DIAGNÓSTICO AUTOMÁTICO DE AVC PÓS-TRAUMA COM BASE NA  
FACE DE PACIENTES UTILIZANDO REDES NEURAIS EM GRAFOS**

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. Assoc. João Paulo Papa  
Coorientador: Dr. Mateus Roder

BAURU  
Novembro/2023

Nicolas Barbosa Gomes

## **Diagnóstico automático de acv pós-trauma com base na face de pacientes utilizando redes neurais em grafos**

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. Assoc. 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

---

**Prof. Dr. Kelton Augusto Pontara da  
Costa**

Universidade Estadual Paulista "Júlio de  
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 16 de Novembro de 2023.

*Primeiramente dou graças a Deus por ter sido comigo durante a realização deste trabalho, e também o dedico à minha família, professores e meus amigos que contribuíram e estiveram comigo durante a jornada.*

# Agradecimentos

Agradeço primeiramente a Deus por ter me dado saúde e a oportunidade de realizar este trabalho. Também agradeço ao meu pai, minha mãe e minha irmã por terem me apoiado durante todos os momentos, principalmente nos mais turbulentos do percurso. Ao professor João Paulo Papa e Mateus Roder por terem me orientado nesta jornada e terem me dado um dos mais bonitos presentes que alguém pode receber: o ‘CONHECIMENTO’.

Agradeço aos meus amigos que tiveram comigo durante a jornada, em especial a Arissa, que teve grande participação neste trabalho e diversos outros durante a graduação, assim como nos artigos publicados em conjunto. Além disso, agradeço ao Cassiano que também participou deste trabalho, ao Davi Neves, Xilsu, Marry, Gustavo, Matheus, Polido, Zastim, Ronaldo, Danilo, Nathan, Renato, Artur, João Pedro e todos os outros por fazerem parte desse curto período de muita alegria, aprendizado e conquistas.

Ainda agradeço imensamente aos professores Andrea, Kelton, Wilson, Paiola, Nilceu, Adriana, Juliana, Simone, Roberta, que não só me ensinaram a programar do zero, mas também trouxeram conhecimento e me apoiaram para que me tornasse uma pessoa melhor.

*"Não reclame e não se explique, apenas se apresente e faça o seu melhor."*  
Bernardo Rocha de Rezende ('Bernardinho')

# Resumo

O acidente vascular cerebral (AVC) pode causar danos significativos aos neurônios, resultando em várias sequelas que afetam negativamente a capacidade do paciente de realizar atividades diárias essenciais, como mastigar, engolir e se comunicar verbalmente. Portanto, é importante que pacientes com essas dificuldades passem por um processo de tratamento e sejam monitorados durante sua execução para avaliar a melhora de sua condição de saúde. O uso de ferramentas computadorizadas e algoritmos que podem detectar rapidamente e de maneira acessível tais sequelas é útil para auxiliar na recuperação do paciente. Devido à morte das células internas do cérebro, um AVC frequentemente leva à paralisia facial, resultando em assimetria entre os dois lados do rosto. Esta monografia se concentra na análise desta assimetria usando um método de aprendizado profundo sem aplicar cálculos de características manuais, introduzindo o modelo “Facial Point Graphs” (FPG), uma abordagem inovadora que se destaca na aprendizagem de informações geométricas e no tratamento eficaz de variações, além do escopo dos cálculos manuais. O FPG permite que o modelo detecte de forma eficaz os distúrbios orofaciais causados por um AVC, a partir de imagens do paciente. A abordagem proposta atingiu acurácia superiores a 90% no conjunto de dados Toronto Neuroface, apontando para avanços substanciais neste domínio.

**Palavras-chave:** Acidente vascular cerebral; Paralisia facial; Aprendizado profundo; Grafos de Pontos Faciais.

# Abstract

Stroke can cause significant damage to neurons, resulting in various sequels that negatively affect the patient's ability to perform essential daily activities such as chewing, swallowing, and verbal communication. Therefore, it's important that patients with these difficulties undergo a treatment process and are monitored during its execution to assess the improvement of their health condition. The use of computerized tools and algorithms that can rapidly and affordably detect such sequels is useful in aiding the patient's recovery. Due to the death of brain's internal cells, a stroke often leads to facial paralysis, resulting in asymmetry between both sides of the face. This monograph focuses on analyzing this asymmetry using a deep learning method without applying manual feature calculations, introducing the "Facial Point Graphs" (FPG) model, an innovative approach that excels in learning geometric information and effectively treating variations beyond the scope of manual calculations. The FPG allows the model to effectively detect orofacial disorders caused by a stroke from patient images. The proposed approach achieved accuracies exceeding 90% in the Toronto Neuroface dataset, indicating substantial advancements in this domain.

**Keywords:** Stroke; Facial paralysis; Deep learning; Facial Point Graph.

# Listas de figuras

Figura 1 – Agregação de mensagens em um grafo bidirecional: (a) grafo de entrada e (b) mecanismo de funcionamento de uma Rede Neural em Grafo. Para maior clareza, a segunda iteração concentra-se exclusivamente no nó ‘A’.	21
Figura 2 – Ilustração do procedimento de extração de características: (a) Utilização do OpenFace 2.0 (BALTRUSAITIS et al., 2018) para detectar, modificar e recortar a imagem facial primária; (b) Imagem recortada e modificada para facilitar a extração de pontos de referência; (c) Extração de pontos de referência utilizando o SPIGA (PRADOSTORREBLANCA; BUENAPOSADA; BAUMELA, 2022); e (d) 16 pontos de referência selecionados como entrada para o modelo.	27
Figura 3 – Processo de formação da entrada da rede neural.	28
Figura 4 – Seleção de frames.	29
Figura 5 – Classificação por Repetição e por Indivíduo.	29
Figura 6 – FPG Modelo	31
Figura 7 – UMAP representando o estado inicial dos dados.	34
Figura 8 – Comparação da acurácia entre as diferentes abordagens utilizadas para a classificação.	35
Figura 9 – Acurácia de modelos mais simples utilizando os mesmos pontos do FPG.	36
Figura 10 – Graph-GradCam resultado.	37
Figura 11 – Diagrama do funcionamento da aplicação.	38
Figura 12 – Modelo da aplicação.	40
Figura 13 – Página de Resumo.	40
Figura 14 – Captura da imagem: (a) tirar a foto e (b) escolher a foto.	41
Figura 15 – Página de demonstração das características extraídas.	41
Figura 16 – Página de inferência.	42

# **Lista de quadros**

Quadro 1 – Tarefas e Sub-tarefas da base de dados Toronto Neuroface. . . . .	25
Quadro 2 – Resultados do modelo FPG-GAT para cada tarefa selecionada. . .	32
Quadro 3 – Resultados do modelo FPG-GCN para cada tarefa selecionada. . .	33

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	Problemática	14
1.2	Justificativa	14
1.3	Objetivos	15
1.3.1	Objetivo Geral	15
1.3.2	Objetivos Específicos	15
1.4	Organização do Trabalho	15
<b>2</b>	<b>TRABALHOS CORRELATOS</b>	<b>17</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
3.1	Grafos	20
3.2	Redes Neurais em Grafos	20
3.3	Redes Neurais Convolucionais em Grafos	21
3.4	Redes Neurais em Grafos com Atenção	22
3.5	Grad-Cam nas Redes Neurais Convolucionais em Grafos	23
<b>4</b>	<b>METODOLOGIA</b>	<b>25</b>
4.1	<b>Conjunto de Dados</b>	<b>25</b>
4.1.1	Pré-processamento dos dados e extração das características	26
4.1.2	Classificação e Validação	28
4.1.3	Modelo Proposto	30
<b>5</b>	<b>EXPERIMENTAÇÃO E RESULTADOS</b>	<b>32</b>
5.1	Experimentos e Resultados	32
5.2	Explicabilidade do FPG-GCN	36
<b>6</b>	<b>APLICAÇÃO DESENVOLVIDA</b>	<b>38</b>
6.1	Flask	38
6.2	Ferramentas de geração de gráficos	38
6.3	Página web	39
6.3.1	Resumo do projeto	39
6.3.2	Captura da imagem	39
6.3.3	Características extraídas	41
6.3.4	Resultados inferidos	42
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>43</b>

7.1	<b>Trabalhos Futuros</b>	43
	<b>REFERÊNCIAS</b>	45

# 1 Introdução

O acidente vascular cerebral (AVC) é uma doença neurológica causada por uma anomalia que altera o fluxo sanguíneo no cérebro. O incidente ocorre quando os vasos que transportam o sangue internamente a este órgão ficam obstruídos, ou se rompem. Neste sentido, existem duas classes principais de AVC: isquêmico e hemorrágico. Os casos de isquemia resultam em um fornecimento inadequado de sangue para a região afetada, ocasionando a carência de nutrientes essenciais e oxigênio. Por outro lado, os acidentes vasculares cerebrais hemorrágicos são caracterizados pela ruptura de vasos sanguíneos, gerando um aumento de pressão sobre os tecidos e nervos adjacentes. A doença é motivo de grande preocupação em todo o mundo e tem se intensificado à medida que a população global envelhece rapidamente (LAPCHAK; ZHANG, 2017). Anualmente, estima-se que de 10,3 a 16,9 milhões de casos, dos quais aproximadamente 5,9 milhões resultam em fatalidades. Ademais, que entre 25,7 e 33 milhões de pessoas necessitam de terapias para lidar com as sequelas do AVC (LAPCHAK; ZHANG, 2017). Nos Estados Unidos, o encargo econômico anual estimado é de 34 bilhões de dólares (BANDINI et al., 2018).

A doença traz uma preocupação global, uma vez que resulta em sérios danos às células da região cerebral afetada, resultando em sequelas como déficits motores, atrofia muscular, e paralisia facial (SCHIMMEL et al., 2017). As células neurais desempenham um papel fundamental na coordenação de muitas funções corporais, e sua deterioração pode afetar profundamente a saúde geral e o bem-estar de um indivíduo. Por exemplo, pacientes com paralisia facial têm os músculos da região enfraquecidos, fato que afeta a capacidade de produzir fala (MULLEN; LOOMIS, 2014), impactando significativamente as habilidades de comunicação social de um indivíduo (SAMSUDIN; SUNDARAJ, 2012). Além disso, atividades cotidianas como comer e beber, tornam-se mais desafiadoras, destacando a importância do tratamento de reabilitação para os pacientes e a criação de processos automatizados para monitorar o tratamento de forma acessível.

Para abordar esse problema, alguns pesquisadores têm proposto técnicas de análise automática de imagens e vídeos para avaliar comprometimentos orofaciais resultantes de derrames e outras doenças. Por exemplo, Bandini et al. (2018) utilizaram marcos faciais tridimensionais extraídos de vídeos, e aplicaram cálculos para extrair características e avaliar a simetria facial, assim como seu movimento. Utilizando uma Máquina de Vetores de Suporte, do inglês *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995), Bandini et al. (2018) atingiram uma acurácia de 87% na classificação de comprometimentos orofaciais em pacientes pós-derrame. Da mesma forma, Parra-

Dominguez, Sanchez-Yanez e Garcia-Capulin (2021) propuseram uma abordagem envolvendo o cálculo de distâncias e ângulos relacionados à simetria facial. Ao extrair vinte e nove características das imagens dos pacientes, eles utilizaram uma Rede de Perceptron de Múltiplas Camadas, do inglês *Multi Layer Perceptron* (MLP) (HAYKIN, 1994) para detectar paralisia facial, alcançando uma acurácia de 94.06% no *Massachusetts Eye and Ear Infirmary Dataset* (MEEI) (GREENE et al., 2020) e 97.22% no *Toronto NeuroFace Dataset* (BANDINI et al., 2020).

Os resultados obtidos nos estudos anteriores demonstraram um desempenho impressionante. No entanto, é importante considerar que todas as características utilizadas foram projetadas manualmente, fato que apresenta algumas limitações. Características elaboradas manualmente dependem fortemente da intuição humana e podem carecer da adaptabilidade necessária para lidar com variações do mundo real. Para eliminar a necessidade de tais cálculos, um modelo chamado Facial Point Graphs (FPGs) foi proposto em (GOMES et al., 2023) para lidar com a Esclerose Lateral Amiotrófica (ELA).

O FPG é um modelo baseado em grafos que utiliza Redes Neurais de Grafos, do inglês *Graph Neural Networks* (GNNs) (SCARSELLI et al., 2008) e tem como entrada grafos formados pelos pontos mais significativos da face para inferir características a partir desses pontos. Neste trabalho, o intuito foi utilizar o modelo para determinar se um indivíduo sofreu um acidente vascular cerebral ou não, ou seja, se há comprometimento orofacial na imagem do paciente ao realizar uma tarefa específica com a face. Até onde se tem conhecimento, não há abordagem existente que utilize um grafo de pontos faciais para detectar esse comprometimento resultante de um acidente vascular cerebral em imagens. A convicção deste trabalho fortemente se baseia na ideia de que os pontos extraídos de imagens faciais podem ser codificados de forma mais eficaz dentro de um espaço não euclidiano, facilitando a definição precisa e a representação de suas características únicas. Além disso, a partir do modelo já treinado, desenvolver uma aplicação que possa auxiliar no tratamento de pacientes que sofreram derrame. Portanto, este trabalho tem três contribuições principais:

- Introduzir o FPG na identificação de comprometimento orofacial resultante de AVC em imagens faciais de indivíduos.
- Empregar uma abordagem de aprendizado profundo dentro do mesmo contexto, eliminando a necessidade de características manualmente elaboradas.
- Desenvolver uma aplicação baseada no modelo para auxílio no tratamento e na verificação da diminuição das sequelas causadas pelo AVC.

## 1.1 Problemática

O AVC é uma condição cerebral grave que pode causar diversos impactos físicos, emocionais e cognitivos nos indivíduos afetados. A reabilitação pós-AVC é um processo complexo e desafiador, exigindo uma avaliação precisa e contínua das condições do paciente. No contexto do diagnóstico de pós-AVC, os grafos oferecem uma representação completa dos pontos mais importantes da face. O estudo destes com redes neurais permite a detecção de sinais de paralisia facial deixada como sequela do AVC, enquanto o indivíduo realiza determinados movimento com os músculos da face, modificando sua expressão facial.

Com a extração de pontos da face humana, e a formação de grafos a partir deles, torna-se possível a utilização de um modelo GNN capaz de inferir características da face humana, e com isso inferir o estado de saúde do paciente. Dessa forma, este trabalho de conclusão de curso (TCC) consistirá em desenvolver um modelo para auxílio médico no tratamento de pessoas que sofreram AVC, avaliando a diminuição da paralisia existente no rosto do paciente. Neste sentido, a classificação do estado de saúde será analisado de acordo com sua expressão facial enquanto realiza determinadas ações, como dar um grande sorriso.

## 1.2 Justificativa

O acidente vascular cerebral está entre as causas mais comuns de morte a nível mundial, e a principal causa de incapacidade permanente nos países desenvolvidos, com quase 800.000 casos por ano nos EUA, e um ônus econômico anual estimado em cerca de US\$ 34 bilhões (BENJAMIN et al., 2017). Um dos aspectos mais debilitantes da doença está relacionado às habilidades de fala onde mais de 40% dos pacientes apresentam distúrbios motores da fala após o AVC (FLOWERS et al., 2013).

Devido a grande incidência deste tipo de doença na população mundial é necessário um meio eficaz de verificação e monitoramento da melhora do paciente durante a realização do tratamento. Neste sentido, o desenvolvimento de uma ferramenta automática que analisa a melhoria na paralisação facial causada pelo AVC é crucial. Além disso, o uso das GNNs se justifica pelo fato das características extraídas da face terem um formato de grafo. A análise desse tipo de estrutura no domínio não euclidiano permite a inferência de características da expressão facial sem a necessidade de cálculos manuais relacionados a extensão do movimento dos músculos.

## 1.3 Objetivos

Esta seção tem como propósito declarar e definir o objetivo geral, assim como os específicos, para o desenvolvimento do TCC.

### 1.3.1 Objetivo Geral

O objetivo do trabalho vislumbra o estabelecimento de um modelo eficaz que possa processar as imagens faciais de pacientes enquanto estes realizam uma ação facial específica. As imagens extraídas dos vídeos mostram o movimento dos músculos orofaciais<sup>1</sup> e, a partir da extração dessas características, a aplicação junta ao modelo têm como objetivo auxiliar no diagnóstico das sequelas do AVC com base na inferência da rede neural. Assim, a eficácia do modelo será determinada pela acurácia com que prediz se o paciente está saudável ou não.

### 1.3.2 Objetivos Específicos

Os objetivos específicos são:

- Preparar e processar o banco de dados *Toronto Neuroface Dataset* (BANDINI et al., 2020);
- Cortar as imagens de forma que o rosto do indivíduo fique centralizado;
- Extrair as características geométricas mais importantes da expressão facial do indivíduo, modelando-as como um grafo;
- Treinar e ajustar um modelo de GNN capaz de inferir as características do grafo formado;
- Avaliar a acurácia do modelo para classificar o estado de saúde do paciente.
- Comparar os modelos propostos ao métodos já utilizados na literatura para analisar o comprometimento dos músculos orofaciais causado pelo AVC.
- Desenvolver uma aplicação automática de fácil utilização e que consiga inferir características da paralisia facial através dos grafos da face.

## 1.4 Organização do Trabalho

A estrutura deste trabalho é a seguinte: O Capítulo 2 fornece uma revisão abrangente da literatura relevante ao tema, o Capítulo 3 oferece uma explicação

---

<sup>1</sup> Musculatura relacionada à comunicação e crítica para funções como mastigação, deglutição e respiração.

detalhada do embasamento teórico. Além disso, o Capítulo 4 delineia o conjunto de dados utilizado, descreve os modelos empregados para recorte de imagens e extração de características faciais, apresenta também a abordagem proposta e explica o método de classificação. O Capítulo 5 exibe os resultados experimentais e discussão dos mesmos, enquanto que no Capítulo 7 há a formulação de conclusões e a indicação de direções para pesquisas futuras.

## 2 Trabalhos Correlatos

Na área de visão computacional, a análise de sinais faciais em imagens tem despertado grande interesse e motivado diversos estudos com o objetivo de automatizar a detecção da paralisia facial (KAEWMAHANIN et al., 2022; PARRA-DOMINGUEZ; SANCHEZ-YANEZ; GARCIA-CAPULIN, 2021; KIM et al., 2015). Essa paralisia que é causada por danos nos nervos, pode surgir devido a condições congênitas, traumas ou doenças como a paralisia de Bell, tumor cerebral ou acidente vascular no cérebro. Nesse contexto, métodos para a detecção da paralisia são desenvolvidos considerando a assimetria facial como um indicador-chave.

Parra-Dominguez, Sanchez-Yanez e Garcia-Capulin (2021) classificaram a paralisia facial sem levar em consideração as variações simétricas entre diferentes expressões faciais realizadas pelo indivíduo, propondo um modelo baseado em MLP para concluir a tarefa. A abordagem tem início com a extração de características dos pontos da face com um modelo denominado MEEshape (GUARIN et al., 2020), que é público e capaz de marcar 68 desses pontos inatos à expressão facial. Como alguns dos pacientes ficaram com a cabeça inclinada nas imagens do conjunto de dados, os autores realizaram uma correção na angulação dos pontos extraídos da face. Em seguida foram elaborados cálculos relacionados a simetria do rosto, como a razão entre a distância do canto externo do olho esquerdo até o meio do lábio e a distância do canto externo do olho direito até o mesmo ponto do lábio. Vinte e nove cálculos manuais deste tipo serviram como entrada à MLP que atingiu uma acurácia de 94.06% no *Massachusetts Eye and Ear Infirmary Dataset* (MEEI) (GREENE et al., 2020) e 97.22% no *Toronto NeuroFace Dataset* (BANDINI et al., 2020).

Com o intuito de analisar o comprometimento da musculatura orofacial causado especificamente pelo derrame, Bandini et al. (2018) demonstraram novas descobertas para a avaliação de maneira automática desta musculatura. Nos experimentos realizados foram capturados vídeos de doze pacientes com AVC (7 homens e 5 mulheres com a idade média de 62,0 anos e desvio padrão de 14,5). Além disso, o conjunto de dados é composto de vídeos do grupo de controle, isto é, indivíduos saudáveis com relação a doença, entre eles 7 homens e 4 mulheres com a idade média de 55,8 anos e desvio padrão de 15,7. Neste contexto, para a gravação dos vídeos foi sugerido a cada participante a realização das seguintes ações: posição de descanso por 20 segundos com os dentes em mordida normal e expressão facial neutra (REST); abertura máxima da mandíbula repetida por 5 vezes (OPEN); fazer um "bico" com os lábios (como se fosse beijar um bebê) um total de 4 vezes (KISS); fingir que está soprando uma vela 5 vezes (BLOW); sorrir com os lábios fechados por 5 vezes (SPREAD); repetição da

sílaba ‘PA’ em uma única respiração o mais rápido possível (PA); repetição da palavra ‘pataka’ o mais rápido possível (PATAKA); e dez repetições da sentença “*Buy Bobby a puppy*” em tom e velocidade de fala habitual.

Após a gravação destes vídeos de maneira padronizada, os autores cortaram manualmente cada repetição de uma ação. Por exemplo, o vídeo da ação SPREAD foi cortado em 5 partes já que o movimento foi realizado por 5 vezes. Para normalização e centralização dos pontos da face em cada *frame*, foi utilizado um algoritmo denominado no inglês *Supervised Descent Method* (SDM) (XIONG; TORRE, 2013) este método foi responsável por minimizar a diferença entre as características extraídas de cada imagem. Para cada uma das repetições foram calculadas dez características geométricas, baseando-se nas coordenadas tridimensionais dos pontos da face. Vale ressaltar que, muitos dos cálculos tinham relação com a simetria do rosto, justamente para capturar sinais da paralisia facial. Por fim, com um total de 23 pacientes, foi realizado o método de validação "Deixar um para fora", do inglês *Leave One Out* (LOSO-CV). Esta validação considera um paciente como teste e os demais como treino para o método de aprendizado. O modelo utilizado neste artigo foi a SVM que atingiu uma acurácia de 87% no diagnóstico dos pacientes.

Uma abordagem semelhante foi realizada por Pecundo, Abu e Alampay (2022). Neste caso, os autores classificaram o estado de saúde dos indivíduos com base em 3 classes: Indivíduos com Esclerose Lateral Amiotrófica (ELA), que sofreram comprometimento Facial devido ao AVC e pacientes saudáveis. O banco de dados para a inferência foi também o Toronto Neuroface (BANDINI et al., 2020). Neste caso, os vídeos deste conjunto representam pessoas realizando determinadas tarefas como a abertura máxima da mandíbula (‘OPEN’), entre outras já descritas anteriormente. Dessa maneira, os autores realizaram a segmentação manual dos vídeos já que um indivíduo realiza a mesma tarefa diversas vezes.

Por conseguinte, foram extraídos pontos específicos da face utilizando a Árvore de Regressão de Conjunto, do inglês *Ensemble Regression Tree* (KAZEMI; SULLIVAN, 2014) implementada na dlib library (KING, 2009). Este modelo permitiu a extração de 68 pontos da face de indivíduos que foram utilizados para o cálculo de 14 características relacionadas a Assimetria Facial, Alcance do Movimento e Forma e Geometria da Boca, isto é, características influentes na inferência da ELA e AVC. Dessa forma, algoritmos de aprendizado de máquina como a Random Forest (HO, 1995), SVM e K-Nearest Neighbors (FIX; HODGES, 1989), alcançaram uma acurácia de 86%.

Até onde se tem conhecimento, a classificação de acidentes vasculares cerebrais usando um modelo de grafo formado por pontos faciais é inovadora. No entanto, modelos de GNNs já foram usados para capturar informações emocionais. De acordo com Ngoc, Lee e Song (2020), os marcos faciais formulados como grafos podem

resultar em um método apropriado para extrair informações valiosas das expressões humana. Em seu modelo, os autores inicialmente mapearam os pontos do rosto a partir de uma Rede Agregada de Estilos, do inglês *Style Aggregated Network* (SAN) (DONG et al., 2018). Vale ressaltar que, cada ponto extraído constitui um vértice do grafo; posteriormente, o método de triangulação de Delaunay (DELAUNAY, 1934) foi empregado para estabelecer as arestas. Por fim, foi utilizado um modelo conhecido como Rede Neural de Grafos Direcionados para classificar expressões faciais nos conjuntos de dados CK+ (LUCEY et al., 2010), MMI (PANTIC et al., 2005) e AFEW (DHALL et al., 2012), alcançando acurárias de 96,02%, 69,4% e 32,64%, respectivamente. É importante ressaltar que, os danos causados por um acidente vascular cerebral no nervo facial alteram a expressão facial, já que esse nervo supre os músculos que controlam o movimento do rosto (LOU; YU; WANG, 2019). Com base neste conhecimento e na validação de GNNs em bancos de expressões faciais, foi empregada uma abordagem utilizando grafos construídos a partir de pontos faciais para analisar as sequelas deixadas no rosto de um indivíduo que sofreu AVC.

Uma proposta semelhante a deste TCC foi apresentada por Gomes et al. (2023), com o objetivo de identificar a Esclerose Lateral Amiotrófica (ELA) com base em um grafo formado a partir dos pontos faciais. No respectivo trabalho, os vídeos que compõem os dados do Toronto Neuroface foram manualmente segmentados com o intuito de separar as diversas repetições de ações executadas pelo indivíduos da amostra. Inicialmente, a abordagem consistiu em processar os *frames* dos vídeos usando o OpenFace 2.0 (BALTRUSAITIS et al., 2018) para obter imagens recortadas e centralizadas, o que simplifica a análise das características faciais. Em seguida, empregou-se uma rede neural conhecida como "*Facial Alignment Network*" (FAN) (BULAT; TZIMIROPOULOS, 2017) para extrair 68 pontos de referência faciais. Com base em 31 deles, foi gerado um grafo utilizando o método de triangulação Delaunay (DELAUNAY, 1934).

Em sequência, para cada experimento conduzido, um entre os 22 pacientes foi designado como teste, enquanto dois foram atribuídos à validação, sendo um identificado como ELA e o outro como saudável. Este último conjunto foi usado como referência para o ajuste de hiperparâmetros do modelo e para execução da técnica parada antecipada, do inglês *early stopping*. Com as 19 pessoas restantes, o modelo foi treinado para classificar cada *frame* sem utilizar características calculadas manualmente. Portanto, a partir da classe mais recorrente na classificação desses *frames*, determinou-se a categoria do segmento de vídeo, utilizando os resultados de cada segmento para identificar a categoria predominante e atribuí-la ao indivíduo correspondente. Como resultado desse processo, alcançou-se uma acurácia de aproximadamente 82% nas tarefas ‘OPEN’ e ‘SPREAD’.

# 3 Fundamentação Teórica

## 3.1 Grafos

Grafos são estruturas de dados que representam computacionalmente as relações entre entidades, sejam elas concretas ou abstratas. De acordo com Hamilton (2020), os grafos podem ser entendidos como uma linguagem universal para descrever sistemas complexos. Formalmente, um grafo  $G = (V, E)$  é definido por um conjunto de nós  $V$  e um conjunto de arestas  $E$  entre esses nós. Uma aresta que vai de um nó  $u \in V$  para um nó  $v \in V$  é representada por  $(u, v) \in E$ . Além disso, essas estruturas podem ser categorizadas como direcionadas ou não direcionadas. No caso dos grafos direcionados, cada aresta está associada a um par ordenado de vértices. Se a aresta  $(u, v) \in E$ , isso não significa necessariamente que  $(v, u) \in E$ . No entanto, no caso dos grafos não direcionados, a relação é a seguinte:  $(u, v) \in E \Rightarrow (v, u) \in E$  e  $(v, u) \in E \Rightarrow (u, v) \in E$ , portanto  $(u, v) \in E \Leftrightarrow (v, u) \in E$ .

Uma representação comumente usada para grafos não esparsos<sup>1</sup> é a matriz de adjacência, denotada como  $A \in \mathbb{R}^{|V| \times |V|}$ . Nesta representação, são atribuídos índices numéricos aos nós do grafo, permitindo que cada nó corresponda a uma linha e coluna específica na matriz  $A$ . Para grafos não ponderados<sup>2</sup>, a relação pode ser expressa da seguinte forma: se  $(u, v) \in E$ , então  $A[u][v] = 1$ , e se  $(u, v) \notin E$ , então  $A[u][v] = 0$ . Um ponto importante a mencionar é que, no caso de grafos ponderados, a matriz de adjacência pode conter valores reais em vez de apenas zeros e uns, ou até mesmo vetores se as arestas possuírem múltiplas características.

## 3.2 Redes Neurais em Grafos

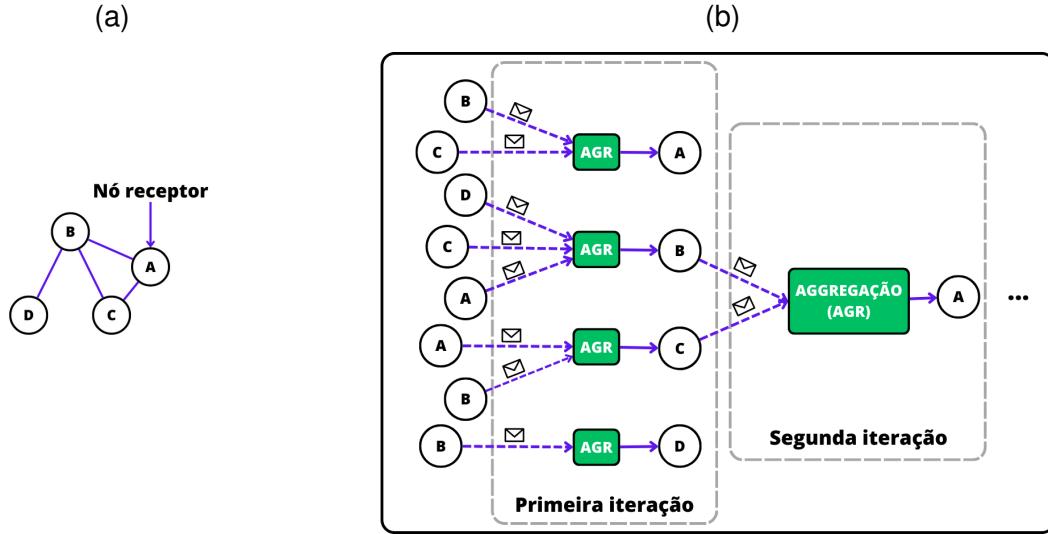
As GNNs revolucionam a forma como os padrões são aprendidos a partir de um conjunto de dados, aproveitando o poder das representações de grafos. No processo iterativo dessa rede neural, cada nó coleta e integra uma agregação das mensagens vindas de seus nós vizinhos. Para isso, utiliza de funções especializadas de agregação e atualização conforme as características dos nós. É importante observar que cada nó transmite informações aos seus nós adjacentes antes de atualizar suas características. Na iteração subsequente, essas novas informações (mensagens) são novamente

<sup>1</sup> um grafo é esparsos quando o número de arestas é aproximadamente igual ao de vértices.

<sup>2</sup> alguns grafos podem ter pesos em suas arestas, caracterizando a relação entre os nós conectados.

repassadas aos nós vizinhos. Esse processo de troca de mensagens é visualmente demonstrado na Figura 1.

Figura 1 – Agregação de mensagens em um grafo bidirecional: (a) grafo de entrada e (b) mecanismo de funcionamento de uma Rede Neural em Grafo. Para maior clareza, a segunda iteração concentra-se exclusivamente no nó ‘A’.



Fonte: Elaborada pelo autor.

Em cada iteração  $k$  da rede neural, um vetor oculto  $\mathbf{h}_u^{(k)} \in \mathbb{R}^n$  é gerado para capturar as características do nó  $u \in V$ , onde  $n$  representa o número de características de entrada. É importante mencionar que, o vetor oculto  $\mathbf{h}_u^{(0)}$  representa a codificação inicial das características antes de qualquer treinamento ocorrer. O processo envolve dois passos: primeiro, uma função invarianta à ordem dos nós agrupa as características da vizinhança  $\mathcal{N}(u)$  do nó  $u$ , permitindo a incorporação de informações relevantes dos nós vizinhos. Em segundo lugar, as características agrupadas são usadas para atualizar as informações do nó, levando à geração de  $\mathbf{h}_u^{(k+1)}$ . O mecanismo específico para essa atualização é descrito da seguinte forma:

$$\mathbf{h}_u^{k+1} = \text{Atualiza}^k (\mathbf{h}_u^k, \text{Agrega}_u^k (\{\mathbf{h}_v^k, \forall v \in \mathcal{N}(u)\})) . \quad (3.1)$$

A Equação 3.1 revela a presença de funções de atualização e agregação. Embora vários modelos possam ser empregados para a função de agregação, este trabalho adota uma formulação baseada em atenção, conforme descrito posteriormente.

### 3.3 Redes Neurais Convolucionais em Grafos

A Rede Neural Convolucional em grafos, do inglês *Graph Convolutional Neural Networks* (GCN) é similar a uma GNN básica, nela a mensagem enviada para os

demais nós será normalizada de acordo com a Equação 3.2. Esta normalização foi proposta pelos pesquisadores Kipf e Welling (2016).

$$\mathbf{m}_{\mathcal{N}(u)} = \sum_{v \in \mathcal{N}(u)} \frac{\mathbf{h}_v}{\sqrt{|\mathcal{N}(u)||\mathcal{N}(v)|}}. \quad (3.2)$$

Outra diferença das GCNs perante a GNN básica é em que a cada iteração há apenas uma matriz de pesos  $\mathbf{W}$ , de forma que tanto as informações existente no nó, tanto as que chegam pela agregação da mensagem são multiplicadas por ela, de acordo com a Equação 3.3:

$$\mathbf{h}_u^{(k)} = \sigma \left( \mathbf{W}^{(k)} \sum_{v \in \mathcal{N}(u) \cup \{u\}} \frac{\mathbf{h}_v}{\sqrt{|\mathcal{N}(u)||\mathcal{N}(v)|}} \right). \quad (3.3)$$

### 3.4 Redes Neurais em Grafos com Atenção

Redes Neurais em Grafos com Atenção, do inglês *Graph Attention Networks* (GATs) oferecem uma técnica poderosa para aprimorar a função de agregação ao atribuir diferentes prioridades às informações da vizinhança. Esse conceito foi introduzido pela primeira vez por Veličković et al. (VELIČKOVIĆ et al., 2017), que lideraram o desenvolvimento do modelo com a seguinte formulação:

$$\mathbf{h}_u^{k+1} = \sigma \left( \sum_{v \in \mathcal{N}(u)} \alpha_{v \rightarrow u}^k \mathbf{W}^k \mathbf{h}_v^k \right). \quad (3.4)$$

Na equação 3.4,  $\mathbf{W} \in \mathbb{R}^{n' \times n}$  representa a matriz de pesos, que é um parâmetro treinável. Neste caso,  $n'$  denota o número de características na saída, e  $\sigma$  representa a função sigmoide. Além disso,  $\alpha_{v \rightarrow u} \in \mathbb{R}$  denota a atenção dada do nó  $v$  para o nó  $u$ , indicando o grau de influência que  $v$  tem na atualização das características de  $u$ . Um valor mais alto de  $\alpha_{v \rightarrow u}$  indica maior influência de  $v$  no processo de atualização das características de  $u$ . Matematicamente, pode-se definir  $\alpha_{v \rightarrow u}$  da seguinte forma:

$$\alpha_{v \rightarrow u}^k = \frac{\exp \left( \text{LeakyReLU} \left( [\mathbf{a}_u^k]^T [\mathbf{W}^k \mathbf{h}_u^k \| \mathbf{W}^k \mathbf{h}_v^k] \right) \right)}{\sum_{v' \in \mathcal{N}(u)} \exp \left( \text{LeakyReLU} \left( [\mathbf{a}_u^k]^T [\mathbf{W}^k \mathbf{h}_u^k \| \mathbf{W}^k \mathbf{h}_{v'}^k] \right) \right)}. \quad (3.5)$$

Na equação 3.5,  $\mathbf{a}_u \in \mathbb{R}^{2*n'}$  define um parâmetro treinável conhecido como vetor de atenção. O símbolo  $\|$  denota o operador de concatenação.

### 3.5 Grad-Cam nas Redes Neurais Convolucionais em Grafos

Esse algoritmo tem como objetivo identificar características importantes dentro do modelo para inferir uma classe específica. Nas Redes Neurais Convolucionais, do inglês *Convolutional Neural Network* (CNN) (LECUN et al., 1998). O intuito é identificar características da última camada que influenciam na determinação da classe, visto que as informações nela são mais completas e significantes (POPE et al., 2019). Além disso, este método exige que a penúltima camada do modelo seja uma convolução.

Seja um grafo com  $N$  nós com seus atributos definidos por  $X \in \mathbb{R}^{N \times d_{in}}$ , onde  $d_{in}$  é a dimensão das características de cada nó. Ademais, considera-se a matriz de adjacências  $A \in \mathbb{R}^{N \times N}$ , podendo esta conter o peso da ligação entre os nós, ou somente, se existe ligação. Outro fator necessário para descrever tal técnica é a matriz de grau do grafo, determinada como  $D_{ii} = \sum_j A_{ij}$ , ou seja, todos os valores que compõem uma única linha da matriz de adjacência são somados e colocados na diagonal, os demais espaços são zerados.

Seguindo o trabalho proposto por (KIPF; WELLING, 2016) pode-se definir a função de uma camada de rede convolucional em grafos como:

$$F^l(X, A) = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} F^{l-1}(X, A) W^l), \quad (3.6)$$

onde  $F^l$  é a ativação da camada  $l$  da rede. É válido ressaltar que,  $F^0 = X$  e  $\tilde{A} = A + I_N$ , é a matriz de Adjacência com adição das auto conexões, onde  $I_N$  é tratada como matriz identidade de dimensão  $N$ . Por fim,  $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$  e  $\sigma$  a função de ativação.

Tomando  $V$  como a expressão  $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ , a camada convolucional  $l$  tem a saída para a  $k$ -ésima característica dos nós, definida pela equação 3.7.

$$F_k^l(X, A) = \sigma(V F^{l-1}(X, A) W_k^l), \quad (3.7)$$

onde  $W_k^l$  denota a  $k$ -ésima coluna da matriz  $W^l$ . Considerando essa notação,  $F_{k,n}^l$  representa a  $k$ -ésima característica para o nó  $n$  na camada  $l$ .

Similarmente ao processo que ocorre na grad-cam das CNNs, é utilizada uma camada global de agrupamento após a última de convolução. A formulação para este processo é descrito na equação 3.8.

$$e_k = \frac{1}{N} \sum_{n=1}^N F_{k,n}^L(X, A). \quad (3.8)$$

Ademais,  $y^c = \sum_k w_k^c e_k$ , sendo o  $y$  a pontuação obtida para a classe  $c$  antes da camada de softmax. A partir deste valor, calcula-se, os diferentes pesos dados a cada

nó na camada  $l$  referente a influência da classe  $c$ , com a relação na Equação 3.9.

$$\beta_k^{l,c} = \frac{1}{N} \sum_{n=1}^N \frac{\delta y^c}{\delta F_{k,n}^l}. \quad (3.9)$$

Por fim, o mapa de calor é passado por uma função ReLU. Isto faz com que os pesos que influenciem a inferência da classe  $c$  de maneira negativa não sejam considerados. A relação para o cálculo do mapa de calor é descrito na Equação 3.10 e com estes valores obtidos pode-se inferir os nós mais ativos na rede, ou seja, aqueles que mais influenciaram para determinação da classe  $c$ .

$$L_{Grad-CAM}^c[l, n] = \text{ReLU}\left(\sum_k \beta_k^{l,c} F_{k,n}^l(X, A)\right). \quad (3.10)$$

# 4 Metodologia

## 4.1 Conjunto de Dados

Apresentado por Bandini et al. (BANDINI et al., 2020), o Toronto Neuroface é um conjunto de dados publicamente disponível projetado para avaliar distúrbios neurológicos. O conjunto é composto por 261 vídeos, cada um validado clinicamente por dois especialistas. É válido ressaltar que, os vídeos incluem avaliações oro-faciais, fato que torna o conjunto de dados um recurso valioso para rastrear e analisar os movimentos faciais em populações clínicas afetadas pela ELA e condições pós-AVC.

Para validar o desempenho do modelo proposto nesse trabalho, foram utilizadas somente as tarefas não faladas de um grupo composto por 11 indivíduos no grupo de controle saudável (HC) e 14 indivíduos no grupo de AVC. Cada participante no conjunto de dados executou múltiplas repetições de algumas tarefas que exigiam a produção de fala e outras que não. É importante notar que, nem todos os participantes no conjunto de dados completaram as tarefas<sup>1</sup>. Após cortar manualmente os vídeos das repetições das tarefas por cada participante, foram obtidos 2.597 repetições. Os detalhes de cada tarefa e o número de repetições em vídeo são apresentados no Quadro 1.

Quadro 1 – Tarefas e Sub-tarefas da base de dados Toronto Neuroface.

Tarefa	Sub-Tarefa	Descrição	HC	AVC
Falada	PA	Repetição da sílaba '/pa/' o mais rápido possível em uma única respiração.	111	104
	BBP	Repetição da sentença "Buy Bobby a Puppy".	884	533
	PATAKA	Repetição da palavra '/pataka/' o mais rápido possível em uma única respiração.	275	163
Não Falada	SPREAD	Sorrir com os lábios fechados.	59	62
	KISS	Fingir que vai beijar um bebê.	57	62
	OPEN	Máxima abertura da mandíbula e da boca.	55	61
	BLOW	Fingir que vai soprar uma vela.	39	44
	BIGSMILE	Abrir um sorriso largo.	15	16
	BROW	Erguer as sobrancelhas.	20	37

Fonte: Elaborado pelo autor.

<sup>1</sup> Para a sub-tarefa BLOW, houve 15 participantes, consistindo de 8 pacientes de AVC e 7 saudáveis (HC). Para a sub-tarefa BROW, houve 11 participantes entre eles 7 do grupo de AVC e 4 do grupo de controle. Além disso, para a sub-tarefa BIGSMILE houve 6 participantes entre eles 3 do grupo que sofreu AVC e 3 do grupo de controle.

#### 4.1.1 Pré-processamento dos dados e extração das características

Na etapa de pré-processamento, foi utilizada a ferramenta OpenFace 2.0 (BALTRUSAITIS et al., 2018) para filtrar elementos visuais do rosto da pessoa na imagem. Essa técnica isola a região de interesse para análise dos músculos orofaciais, ao mesmo tempo que incorpora a estimativa da posição da cabeça para centralizar o rosto. Portanto, a saída da fase de pré-processamento é uma imagem em tons de cinza recortada, representando a região do rosto, com dimensões de 200 X 200 *pixels* para cada *frame* no conjunto de dados experimental.

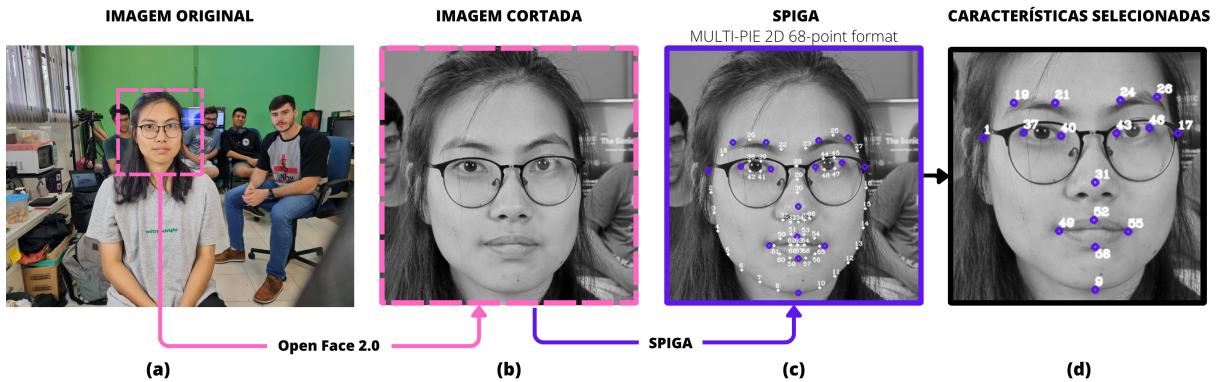
Para criação da estrutura do grafo foram utilizadas técnicas de pré-processamento afim de extrair as características necessárias das imagens. Em primeiro lugar, para capturar pontos da face, foi utilizada um modelo nomeado SPIGA (PRADOS-TORREBLANCA; BUENAPOSADA; BAUMELA, 2022), do inglês *Shape Preserving Facial Landmarks Graph Attention Networks*. Esta abordagem é uma combinação de duas técnicas. A primeira delas baseada em uma CNN *encoder-decoder* onde é feita a extração de mapas de calor (*heatmap*) da imagem. Por conseguinte cada coordenada é tipicamente estimadas no ponto máximo de calor deste *heatmap*.

Do ponto de vista matemático, a forma inicial da face, denotada como  $\mathbf{x}_0 \in \mathbb{R}^{L \times 2}$ , foi definida através da projeção de  $L$  pontos de uma face genérica no espaço tridimensional, na mesma direção da pose 3D estimada pela *backbone*. Além disso, em cada passo  $t$ , a GAT computa um vetor de variação de posição  $\delta\mathbf{x}_t$  que atualiza a posição dos pontos,  $\mathbf{x}_t = \mathbf{x}_{t-1} + \delta\mathbf{x}_t$ . Através desta técnica, foi possível determinar com precisão as posições de cada ponto nos conjuntos de imagens obtidos a partir do banco de dados Toronto Neuroface.

Uma vez que a paralisia facial causada por um acidente vascular cerebral pode afetar um ou ambos os lados do rosto, foram selecionados pontos do rosto com o intuito de capturar as posições espelhadas de simetria, abrangendo os dois lados da região da sobrancelha, olhos e boca. Como resultado, dos 68 pontos de referência no rosto detectados pelo SPIGA, foram selecionados 16, com o intuito de captar um grafo com informações focadas na inferência do pós AVC, isto é, na diferença de simetria entre os lados da face. O processo de extração de características é ilustrado na Figura 2.

Cada nó do grafo tem duas características que representam as coordenadas geométricas [x, y] do ponto na imagem. No entanto, essas coordenadas não são as originais extraídas com o método SPIGA. Todos os demais pontos foram rotacionados com base no vetor que sai do ponto 1 e vai até o ponto 16. Nesse sentido, calculou-se o ângulo formado entre esse vetor e o eixo x do plano cartesiano, com a base na Equação 4.1.

Figura 2 – Ilustração do procedimento de extração de características: (a) Utilização do OpenFace 2.0 (BALTRUSAITIS et al., 2018) para detectar, modificar e recortar a imagem facial primária; (b) Imagem recortada e modificada para facilitar a extração de pontos de referência; (c) Extração de pontos de referência utilizando o SPIGA (PRADOS-TORREBLANCA; BUENAPOSADA; BAUMELA, 2022); e (d) 16 pontos de referência selecionados como entrada para o modelo.



Fonte: Elaborada pelo autor.

$$\angle(P_1, P_{16}) = \text{arctan}2(\Delta x, \Delta y) \times 180/\pi \quad (4.1)$$

$\Delta x$  e  $\Delta y$  representam a variação das coordenadas cartesianas x e y entre os dois pontos, descritas como  $\Delta x = P_{1_x} - P_{16_x}$  e  $\Delta y = P_{1_y} - P_{16_y}$ . Vale ressaltar que, a rotação dos pontos foi obtida por uma transformação linear utilizando da matriz de rotação  $M$  disponibilizada na biblioteca opencv do Python. Dessa maneira, as coordenadas do ponto rotacionado  $(P'_{r_x}, P'_{r_y})$  são obtidas pela formulação descrita na Equação 4.2, onde  $(P_{r_x}, P_{r_y})$  são as coordenadas originais do ponto  $r$ .

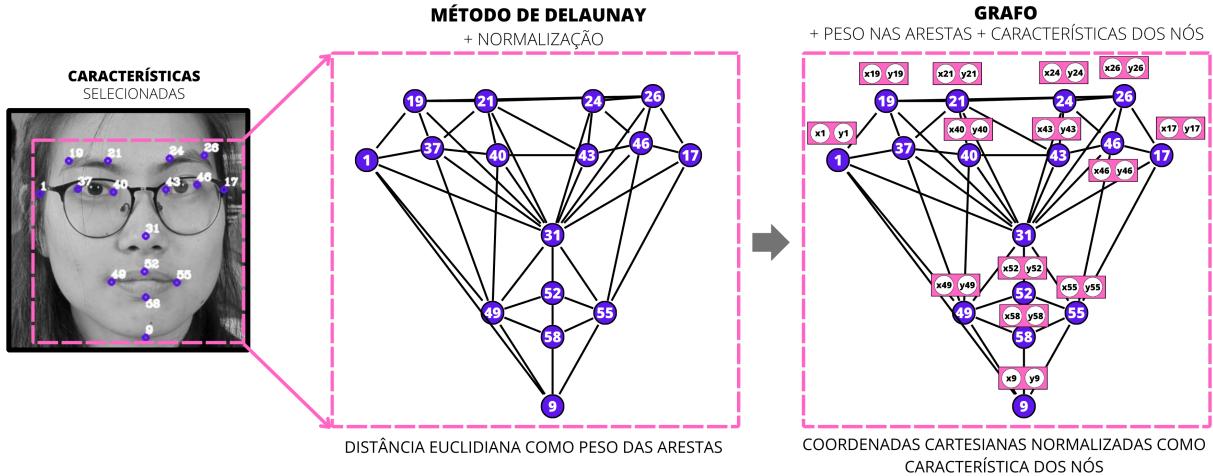
$$\begin{pmatrix} P'_{r_x} \\ P'_{r_y} \end{pmatrix} = \begin{bmatrix} M_{1,1} & M_{1,2} & M_{1,3} \\ M_{2,1} & M_{2,2} & M_{2,3} \end{bmatrix} \begin{bmatrix} P_{r_x} \\ P_{r_y} \end{bmatrix} \quad (4.2)$$

As conexões entre os pontos do grafo foram criadas usando a técnica de triangulação de Delaunay (DELAUNAY, 1934). Este método gera uma malha triangular que captura relações espaciais entre os pontos. Para isso, se baseia no critério de que dentro da circunferência circunscrita ao triângulo formado por três pontos, não há outro ponto do conjunto.

A estrutura final do grafo garante que todos os nós estejam conectados ao ponto 31 da notação MULTI-PIE (ponta do nariz). Esse ponto de normalização serviu como referência para estabelecer um padrão de comunicação consistente e padronizado entre os nós no modelo. Por fim, a distância euclidiana entre cada par de nós foi

calculada, servindo como o peso para cada aresta no grafo. Uma visão geral de todo o processo está apresentada na Figura 3.

Figura 3 – Processo de formação da entrada da rede neural.



Fonte: Elaborada pelo autor.

#### 4.1.2 Classificação e Validação

O método de classificação usado neste trabalho deixa sempre uma pessoa para teste em cada um dos experimentos, e denomina-se no inglês *leave-one-subject-out cross-validation* (LOSO-CV), seguindo o método proposto por Bandini et al. (2018). A fim de aumentar a confiabilidade das previsões em cenários do mundo real e mitigar problemas como o sobreajuste ou a memorização dos dados de treino, o conjunto de dados foi separado em três partes: **(i) Teste** - Composto por um único indivíduo, **(ii) Validação** - Composto por dois indivíduos escolhidos aleatoriamente, um categorizado como saudável e o outro categorizado como AVC, e **(iii) Treinamento** - Conjunto composto pelos indivíduos restantes do conjunto de dados.

Inicialmente, o modelo proposto utiliza quinze *frames* no momento do pico de movimento da repetição, semelhante ao processo que ocorre na imagem 4. Primeiramente foi feita uma classificação baseada nesses *frames* em específico. A partir da média obtida na inferência dos *frames* categorizou-se a classe do indivíduo.

- 1. Classificação da Repetição:** Para cada iteração do LOSO-CV, os *frames* das repetições produzidas por um participante foram tratadas como amostras individuais durante os experimentos. No modelo de validação utilizado, o teste de cada experimento foi composto pelas imagens escolhidas de todas as repetições de um único indivíduo, e o modelo foi responsável por rotular cada uma dessas imagens que compunham as repetições, como mostrado em 5. Ao mesmo tempo, os

Figura 4 – Seleção de frames.

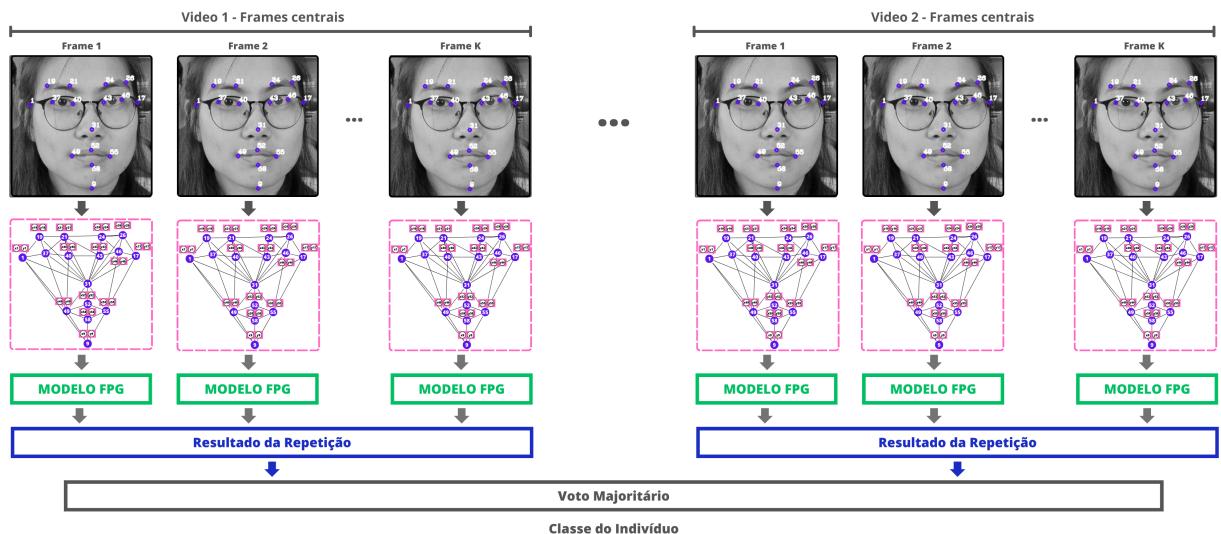


Fonte: Elaborada pelo autor.

indivíduos restantes foram divididos em dois conjuntos, o conjunto de validação e o conjunto de treino, ou seja, ambos os conjuntos constituídos por repetições das ações executadas por dois ou mais indivíduos. Neste caso, o erro do modelo foi calculado com base no rótulo atribuído a cada imagem pertencente as repetições do conjunto de treino, levando em consideração a ação específica em análise.

**2. Classificação do Indivíduo:** Seguindo o mesmo método proposto por Bandini et al. (BANDINI et al., 2018), foi utilizado o LOSO-CV tratando cada sujeito como um caso de teste e classificando-o com base no voto majoritário entre as classes obtidas para as repetições. Em casos de empate, para garantir uma predição mais conservadora, o sujeito foi considerado HC, ou seja, saudável.

Figura 5 – Classificação por Repetição e por Indivíduo.



Fonte: Elaborada pelo autor.

Tanto na classificação baseada na repetição, como na classificação baseada no indivíduo, foi utilizado o conjunto de validação para mitigar viés nos hiperparâmetros do modelo. Neste sentido, a partir do erro neste conjunto, foi possível aplicar a técnica

ReduceLROnPlateau<sup>2</sup>, com o intuito de aumentar a estabilidade, a velocidade de convergência e a generalização do modelo durante o treinamento.

Com base na divisão executada no conjunto de dados, o modelo de FPG foi estruturado de duas formas diferentes, uma delas utilizando camadas GAT (FPG-GAT) e outra camadas de GCN (FPG-GCN). É válido notar que, os hiperparâmetros para estes modelos foram escolhidos com base na avaliação empírica de seus desempenhos no conjunto de validação. Por conseguinte, o modelo FPG-GAT foi treinado por um total de 30 épocas, utilizando um tamanho de lote de 16 amostras e com taxa de aprendizagem de  $10^{-4}$  para as camadas do modelo. As camadas ocultas têm um tamanho de 64, e uma taxa de *dropout* de 0,2 foi aplicada após algumas das funções LeakyRelu posteriores as camadas GAT. Já o modelo FPG-GCN foi treinado por 150 épocas, com uma taxa de *dropout* de 0,1, camadas ocultas de dimensão 64 e taxas de aprendizagem de  $10^{-3}$ <sup>3</sup>.

#### 4.1.3 Modelo Proposto

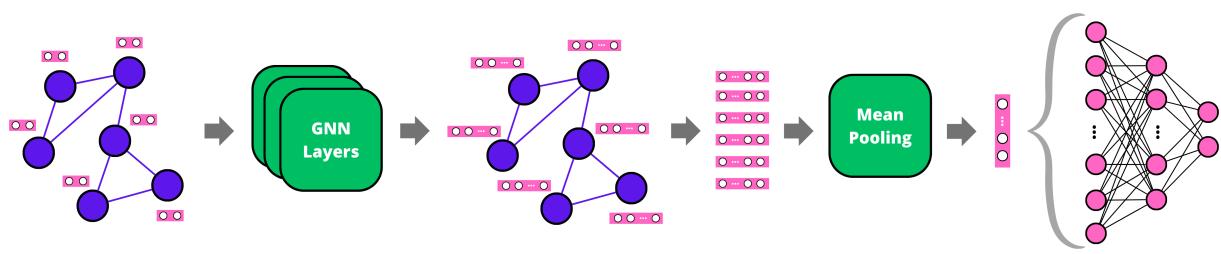
Para cada *frame* analisado foi formado um grafo de dezesseis nós representando 16 pontos concebidos pela etapa de extração de características, conforme explicado na Seção 4.1.1. Nesse processo, cada um dos quinze grafos formados a partir das imagens dos vídeos passaram por seis camadas GAT ou GCN, dependendo do modelo, e duas camadas lineares. Conforme ilustrado na Figura 6, um passo crucial antes que as informações entrem nas camadas lineares envolve o agrupamento, do inglês *pooling*, onde todos os vetores de características do grafo são combinados em um vetor único. A operação de *pooling* é realizada por meio do algoritmo *global\_mean\_pool* disponível na biblioteca PyTorch Geometric (FEY; LENSSEN, 2019). Este que seleciona a média de cada uma das características dos nós e as incorpora em um único vetor que representa o grafo.

Após o agrupamento, o resultado passa por duas camadas lineares adicionais, gerando a saída do modelo. É importante ressaltar que o cálculo do erro é baseado na classe de cada grafo (*frame*), visto que neste trabalho é feita a classificação das imagens dos pacientes nos picos de movimento. Para a classificação da repetição, os resultados finais são obtidos pela identificação da média dos quinze *frames* utilizados. Entretanto, ao classificar o sujeito, o voto majoritário é derivado das classificações de

<sup>2</sup> Para esta otimização foi utilizada uma paciência de valor 5, a mínima taxa de aprendizado com valor  $10^{-5}$ , o limiar como 0.1 e o fator pelo qual a taxa de aprendizado seria dividida como 0.1

<sup>3</sup> Nota: A avaliação experimental dos modelos de FPG foi realizada usando PyTorch Geometric (FEY; LENSSEN, 2019) em uma GPU. Um ponto importante a mencionar é que, até o presente trabalho, a implementação de Redes Neurais em Grafos usando a operação de dispersão em uma GPU introduz um comportamento não determinístico.

Figura 6 – FPG Modelo



Fonte: Elaborada pelo autor.

cada repetição. A classe que mais aparece na repetição das ações realizadas pelos indivíduos, representará a classe dada para o mesmo.

# 5 Experimentação e Resultados

Neste capítulo são apresentados os experimentos e os resultados obtidos para cada um deles, assim como uma breve explicação do que foi realizado.

## 5.1 Experimentos e Resultados

Os primeiros experimentos foram realizados, considerando as tarefas selecionadas e os recortes. A análise neste trabalho baseou-se na classificação utilizando o modelo de FPG proposto com duas redes em grafos: a GCN (FPG-GCN) e a GAT (FPG-GAT). A entrada desses modelos foram os pontos faciais extraídos dos *frames* de maior amplitude das ações realizadas nos vídeos. Dessa forma, os resultados experimentais foram obtidos para cada tarefa utilizando-se ambos os modelos, conforme mostrado nos Quadros 2 3.

Quadro 2 – Resultados do modelo FPG-GAT para cada tarefa selecionada.

TAREFA	Classificação (GAT)	Acurácia	Especificidade	Sensibilidade
SPREAD	Repetição	38,04%	40,13%	36,45%
	Indivíduo	40,00%	45,45%	35,71%
KISS	Repetição	53,02%	38,88%	63,56%
	Indivíduo	52,00%	36,36%	64,29%
OPEN	Repetição	52,99%	46,86%	57,43%
	Indivíduo	64,00%	54,54%	71,43%
BLOW	Repetição	42,53%	25,56%	53,74%
	Indivíduo	44,44%	28,57%	54,54%
BIGSMILE	Repetição	72,01%	87,74%	57,26%
	Indivíduo	83,33%	100,00%	66,67%
BROW	Repetição	87,94%	86,77%	88,57%
	Indivíduo	90,91%	75,00%	100,00%

Fonte: Elaborado pelo autor.

Critérios como acurácia, especificidade e sensibilidade foram considerados para avaliar os resultados obtidos pelo modelo proposto. Uma análise mais específica dos resultados foi possível devido aos valores de sensitividade e especificidade calculados. O primeiro deles indica a proporção entre as pessoas classificadas com sequelas de AVC e o total de indivíduos com AVC no conjunto de dados. Já o segundo indica a proporção entre as pessoas classificadas como HC e todas os indivíduos saudáveis do banco.

É possível observar na tabela que apesar de terem sido usadas técnicas para tentar diminuir o número de falsos positivos, a sensibilidade foi maior que a especificidade

Quadro 3 – Resultados do modelo FPG-GCN para cada tarefa selecionada.

TAREFA	Classificação (GCN)	Acurácia	Especificidade	Sensibilidade
SPREAD	Repetição	31,19%	23,67%	36,95%
	Indivíduo	32,00%	18,18%	42,86%
KISS	Repetição	50,60%	37,86%	60,10%
	Indivíduo	48,00%	36,36%	57,14%
OPEN	Repetição	51,88%	36,48%	63,03%
	Indivíduo	52,00%	36,36%	64,29%
BLOW	Repetição	45,23%	26,80%	57,41%
	Indivíduo	50,00%	14,29%	72,73%
<b>BIGSMILE</b>	<b>Repetição</b>	<b>91,78%</b>	<b>88,82%</b>	<b>94,56%</b>
	<b>Indivíduo</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>
BROW	Repetição	36,36%	25,00%	42,86%
	Indivíduo	37,20%	29,19%	41,54%

Fonte: Elaborado pelo autor.

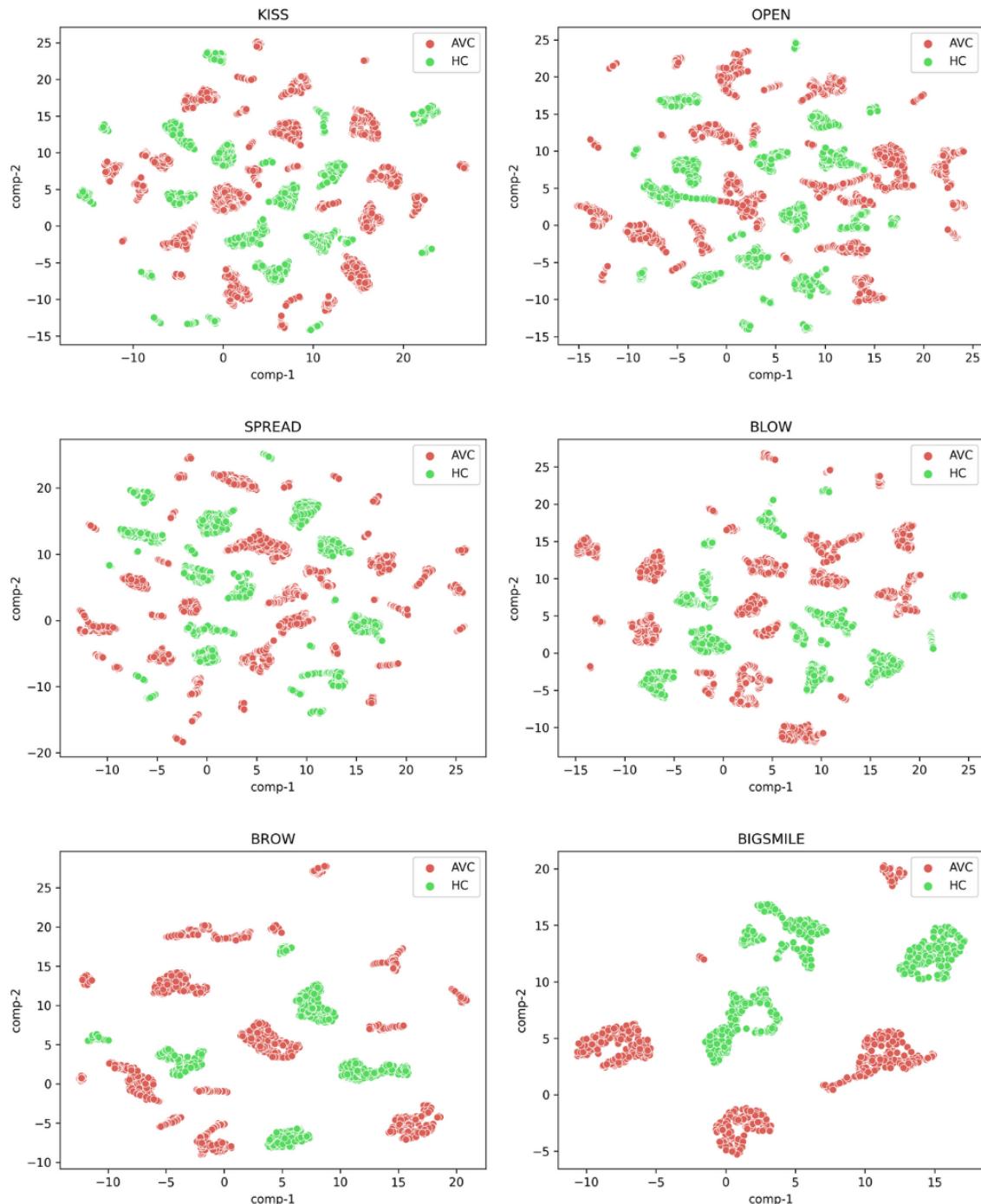
na maioria dos casos. Acredita-se fortemente que a menor especificidade tenha relação com o fato do banco ter maior quantidade de pacientes com AVC do que pessoas saudáveis. Portanto, o modelo pode encontrar mínimos locais no cálculo do gradiente que favorecem o rótulo da pessoa como doente. Apesar disto, o modelo FPG-GAT proposto alcançou uma acurácia de aproximadamente 91%, rotulando apenas uma pessoa como falso positivo para a tarefa 'BROW'. A tarefa de classificação binária que se destacou para este modelo foi a repetição do movimento de erguer as sobrancelhas. Essa tarefa possui apenas onze pacientes, porém como cada um executa a mesma tarefa múltiplas vezes foram obtidos um total de 855 frames, quantidade suficiente para o modelo aprender e alcançar uma acurácia elevada.

É válido notar que, no geral, o modelo FPG-GCN não alcançou o mesmo desempenho que o FPG-GAT, este fato é esperado e acontece devido ao modelo de GAT possuir atenção, isto é, são dadas diferentes importâncias aos nós da rede durante o treinamento, geralmente apresentando melhores resultados que modelos de GCN. Contudo, na tarefa 'BIGSMILE', o modelo FPG-GCN inferiu todas as pessoas com a classe correta, isto é, teve uma acurácia de 100% na classificação individual. Nota-se que, este modelo inferiu corretamente todas as pessoas e não todos os frames do teste. 91% destes últimos foram inferidos de maneira correta, e a partir do voto majoritário das acurácia obtidas nas repetições inferiu-se todos os indivíduos apropriadamente. Além disso, o modelo FPG-GAT inferiu os dados desta mesma tarefa com uma acurácia de 83%. Para as demais tarefas, os modelos não atingiram resultados expressivos.

A fim de explicar essa diferença existente entre os resultados das distintas tarefas, utilizou-se a técnica denominada Aproximação e Projeção de *Manifolds Uniformes*, do inglês *Uniform Manifold Approximation and Projection* (UMAP). O UMAP , é uma

técnica de redução de dimensionalidade não linear. Sua finalidade é mapear dados em um espaço de dimensões elevadas para um de menor dimensão, geralmente duas ou três dimensões, preservando as relações complexas entre os dados. Os resultados obtidos para tal técnica utilizada no conjunto de dados utilizado podem ser vistos na Figura 7.

Figura 7 – UMAP representando o estado inicial dos dados.



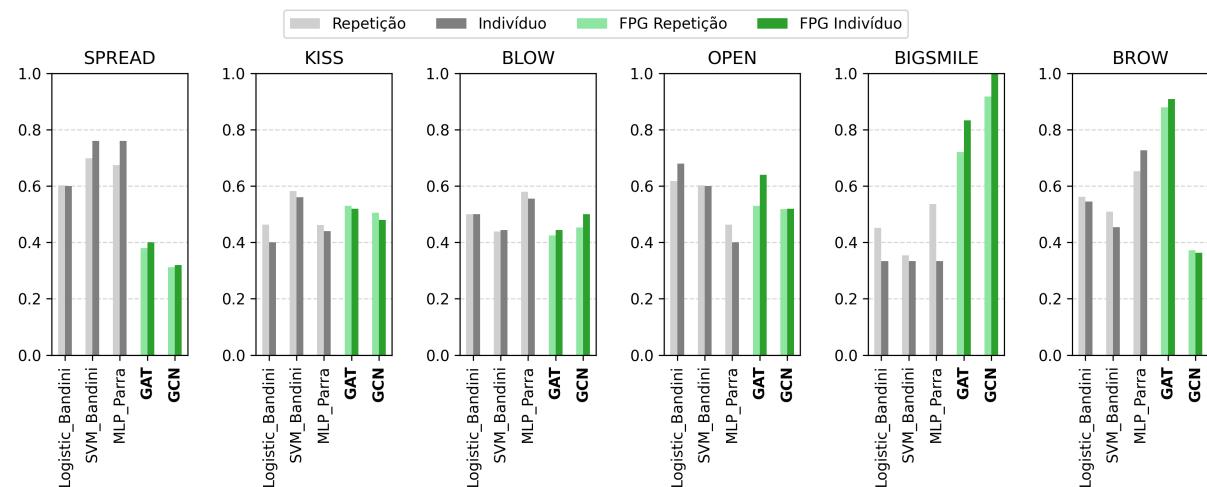
Fonte: Elaborada pelo autor.

O UMAP mostrou visualmente que, os dados extraídos das tarefas ‘KISS’,

‘OPEN’, ‘BLOW’ e ‘SPREAD’ se organizam de maneira mais complexa, quando comparados aos dados das tarefas ‘BROW’ e ‘BIGSMILE’. Essa diferença da organização inicial é um dos fatores que justificam visualmente, tanto as acuráncias mais baixas alcançadas pelos modelos de FPG no primeiro grupo de tarefas, quanto as acuráncias elevadas alcançadas pelos mesmos modelos nas tarefas ‘BROW’ e ‘BIGSMILE’. É válido ressaltar que, seria necessário um volume maior de dados nas tarefas ‘BROW’ e ‘BIGSMILE’ para validar os bons resultados no diagnóstico de pós-AVC nessas duas tarefas.

Ademais, o mesmo padrão experimental foi estabelecido para comparar os modelos de referência com o FPG. Como os recortes foram executados de maneira manual, achou-se necessário aplicar os métodos utilizados por Bandini et al. (2018) e Parra-Dominguez, Sanchez-Yanez e Garcia-Capulin (2021) no mesmo conjunto de dados utilizado neste trabalho. Para isso foram encontradas as mesmas características calculadas nos dois artigos (BANDINI et al., 2018; PARRA-DOMINGUEZ; SANCHEZ-YANEZ; GARCIA-CAPULIN, 2021). Além disso, foram implementadas as mesmas técnicas de aprendizado de máquina utilizadas pelos autores. Os resultados obtidos para as tarefas são ilustrados na Figura 8.

Figura 8 – Comparação da acurácia entre as diferentes abordagens utilizadas para a classificação.



Fonte: Elaborada pelo autor.

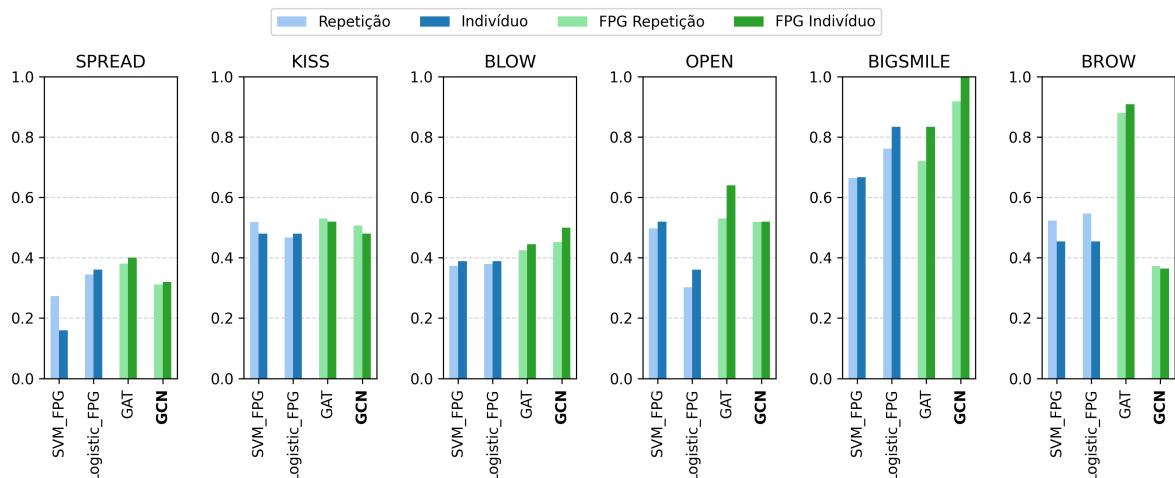
A melhor tarefa dos modelos comparativos foi a ‘SPREAD’ e nas demais teve resultados semelhantes aos dos modelos propostos neste trabalho, com exceção do ‘BIGSMILE’ e do ‘BROW’. Esta diferença deve-se ao fato dos dados utilizados para a inferência serem distintos entre os modelos. É válido ressaltar que, as características para os modelos comparativos foram ‘manualmente’ selecionadas e calculadas, seguindo a abordagem proposta pelos autores. Além disso, os resultados para a SVM e a

Regressão Logística, realizando uma busca em grade para os hiperparâmetros <sup>1</sup>, foram diferentes dos encontrados por Bandini et al. (2018), da mesma forma que os resultados para a MLP foram diferentes dos demonstrados por Parra-Dominguez, Sanchez-Yanez e Garcia-Capulin (2021).

É importante ressaltar que, foram seguidos os mesmos procedimentos introduzidos pelos autores. Entretanto, a diferença nos resultados obtidos deve-se ao fato de neste trabalho terem sido utilizados dados diferentes, dos utilizados pelo trabalhos citados. Além disso, na abordagem deste TCC há um conjunto de validação, o que diminui o conjunto de treino, influenciando também na diferença de resultados.

Outra comparação foi realizada com modelos mais leves como a Regressão Logística e a SVM <sup>2</sup>, utilizando exatamente os mesmos dados que os modelos de FPG. Dada a ilustração da Figura 9, pode-se notar que, no geral, os modelos de FPG alcançaram melhores resultados. É valido observar, a possibilidade de atingir resultados consideráveis na inferência de sequelas de AVC analisando a face e sem utilizar características extraídas a mão, isto é, procurando ao máximo diminuir o viés humano na seleção dos dados.

Figura 9 – Acurácia de modelos mais simples utilizando os mesmos pontos do FPG.



Fonte: Elaborada pelo autor.

## 5.2 Explicabilidade do FPG-GCN

Há um certo fato que dificulta a aplicação de modelos profundos na medicina. Quanto mais camadas existem no modelo, mais difícil torna-se explicar o seu comportamento. Para contornar tal situação, foi utilizada a técnica denominada grad-cam em

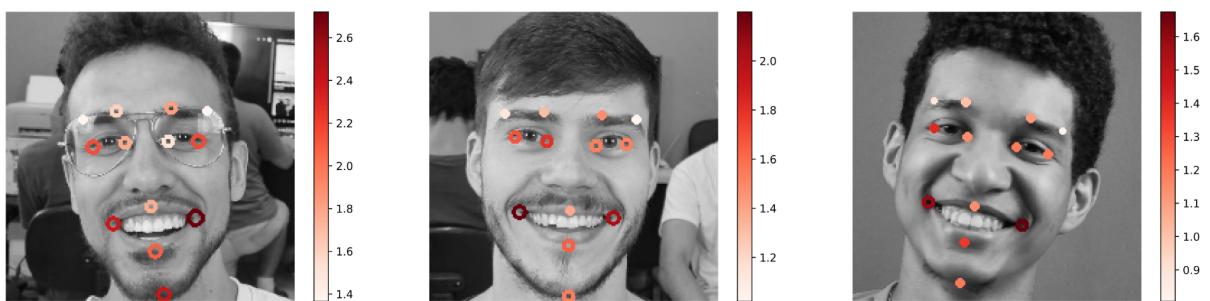
<sup>1</sup>  $C = [2^{-5}, 2^{-3}, \dots, 2^{15}]$ , e  $\gamma = [2^{-15}, 2^{-13}, \dots, 2^3]$ .

<sup>2</sup>  $C = [10^{-3}, 10^{-2}, \dots, 10^3]$ , e  $\gamma = [10^{-4}, 10^{-3}, \dots, 10^0]$ .

grafos a fim de verificar quais pontos foram mais influentes na predição da tarefa ‘BIGSMILE’. Esta técnica, foi inicialmente desenvolvida para Redes Neurais Convolucionais e adaptada recentemente para Redes Neurais Convolucionais em Grafos (POPE et al., 2019). O algoritmo utiliza o erro calculado no final da rede para a classe desejada e, a partir disso, consegue extrair os nós mais importantes para a camada selecionada, como já descrito na Seção 3.5.

Esta técnica foi aplicada somente para este modelo, visto que, as GATs possuem um algoritmo diferente e até o momento atual da escrita deste trabalho, não foi desenvolvido um método de grad-cam para esta rede. Por fim, a técnica foi aplicada na tarefa em que o FPG-GCN alcançou melhor acurácia, sendo o intuito usá-la na aplicação desenvolvida neste trabalho, para inferir a classe de indivíduos de maneira automática. Os resultados da aplicação desta técnica podem ser vistos na Figura 10.

Figura 10 – Graph-GradCam resultado.



Fonte: Elaborada pelo autor.

A escala ao lado direito das imagens é gerada com base nos pesos devolvidos pelo algoritmo de Grad-CAM em grafos. Quanto mais alto o valor, maior a influência do nó em inferir a classe. Neste sentido, é possível observar que, os pontos que tiveram maior influência na inferência da classe foram os pontos das laterais da boca. Este fato é relevante, visto que a tarefa em questão é um sorriso largo. Os pontos das extremidades laterais da boca são pontos de máxima amplitude nas ações dos vídeos, fato que indica a análise de características relacionadas ao movimento sugerido por esta ação, o que valida a caracterização feita pelo modelo para tais experimentos.

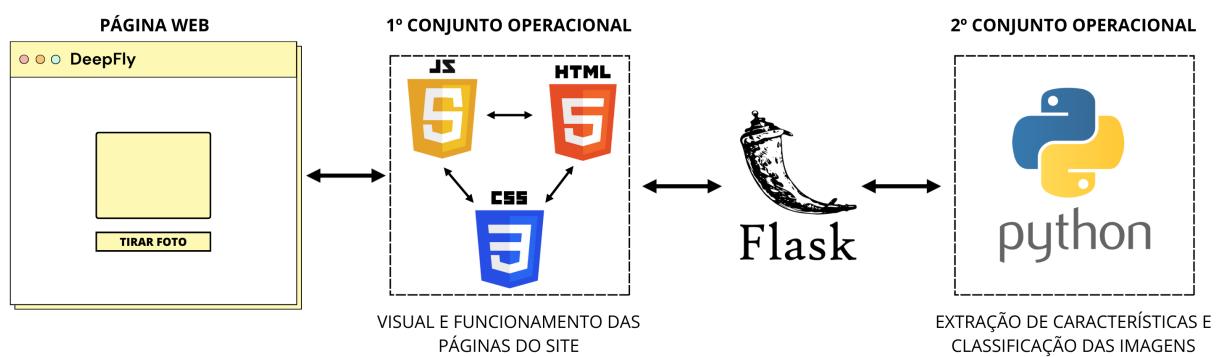
# 6 Aplicação Desenvolvida

Por fim, após validar os resultados experimentais, os modelos foram salvos e puderam ser usados para inferir o estado de saúde de indivíduos de maneira simples e intuitiva dentro da aplicação desenvolvida. É válido ressaltar que, o software não será disponibilizado para uso público e o objetivo deste não é afirmar o estado de saúde do indivíduo. Contudo, servir como método de auxílio a profissionais que visam acompanhar as sequelas faciais deixadas pelo AVC em seus pacientes.

## 6.1 Flask

Para desenvolvimento de uma aplicação web intuitiva, foi necessário criar uma comunicação entre linguagens. Mais especificamente, uma conexão entre dois conjuntos operacionais, o primeiro deles operando em JavaScript, HTML e CSS e o segundo em Python. O Flask é um *framework* que permite a criação de uma ‘comunicação’ entre dois conjuntos operacionais embutidos no site, como representado na Figura 11. Dessa forma, o primeiro grupo, responsável pela parte visual e funcionamento da página web, pode trocar informações com a linguagem python. Neste caso, o python foi utilizado por dois motivos principais: (i) Extrair as características das imagens capturadas pelo usuário e (ii) Inferir com os modelos salvos e já treinados as características extraídas a chance do usuário ter ou não sequelas do AVC.

Figura 11 – Diagrama do funcionamento da aplicação.



Fonte: Elaborada pelo autor.

## 6.2 Ferramentas de geração de gráficos

Com o objetivo de melhorar a experiência do usuário na análise das características de entrada do modelo, a integração da biblioteca D3.js desempenhou um papel

fundamental. D3.js é uma biblioteca JavaScript altamente flexível e de fácil implementação, que possibilitou a criação de gráficos interativos para representar informações complexas de forma clara e dinâmica. Especificamente, essa biblioteca permitiu a geração de representações visuais para grafos, tornando a visualização das conexões entre elementos e a identificação de padrões de associação mais acessíveis, assim como a formação da matriz de adjacência do mesmo. A inclusão desses recursos visuais não apenas enriqueceu a apresentação das informações, mas também facilitou a compreensão e interpretação dos dados do site.

Com o intuito de gerar uma visualização do processo de extração de características, foi utilizada a biblioteca Plotly. O Plotly é uma biblioteca aberta e versátil, útil para a visualização de dados e compatível com Python, JavaScript e R. No contexto deste trabalho, ela foi utilizado em conjunto com JavaScript, o que permitiu a aplicação apresentar gráficos interativos da imagem original com os pontos extraídos.

## 6.3 Página web

A estrutura da página foi projetada para armazenar as técnicas utilizadas e validadas durante a elaboração de artigos e projetos relacionados a área de aprendizado de máquina e por isso a língua utilizada dentro dela é o inglês. Ao acessar a página inicial o usuário tem outras opções de diagnóstico de doença. Contudo, toma-se como foco deste trabalho as páginas relacionadas a caracterização do pós-AVC. Um modelo completo do funcionamento da página é ilustrado na Figura 12.

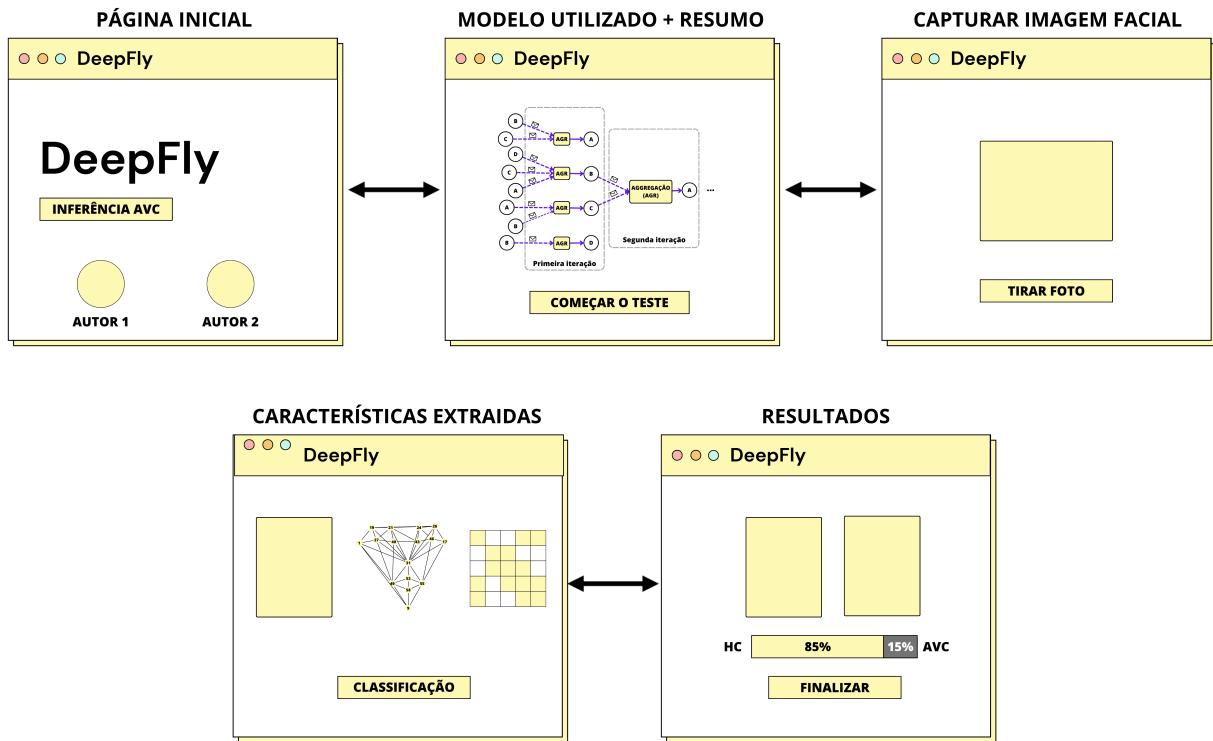
### 6.3.1 Resumo do projeto

Assim que o usuário escolhe o teste de pós-AVC será encaminhado a uma página de explicação do procedimento que será realizado. Nesta página há uma breve explicação onde o usuário pode entender um pouco melhor do modelo de grafos utilizado, assim como um resumo sobre a abordagem proposta neste trabalho, como pode ser visto na Figura 13.

### 6.3.2 Captura da imagem

A captura da imagem é feita por uma *webcam* simples do dispositivo conectado ao *website*. Neste sentido, o JavaScript recebe a imagem em formato binário e a transforma em uma cadeia de caracteres (*string*), que envia para o Python pela conexão criada. Nessa operação, é feita uma verificação se há faces na imagem ou não, através da ferramenta opencv disponível dentro do próprio Python. É valido ressaltar que, se o usuário tirar uma foto que não contenha uma face, surgirá um aviso na tela informando-o da necessidade de capturar uma imagem facial. Além disso, na página que realiza

Figura 12 – Modelo da aplicação.

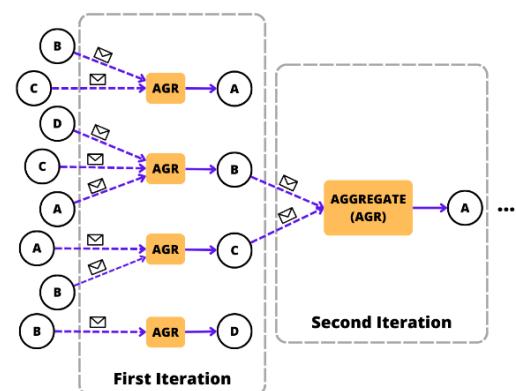


Fonte: Elaborada pelo autor.

Figura 13 – Página de Resumo.

#### Abstract:

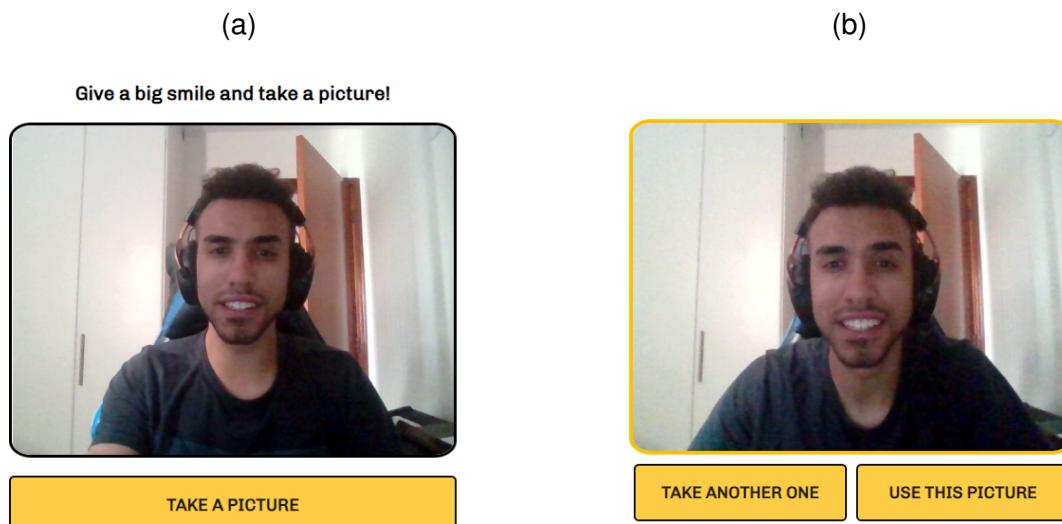
Stroke can cause significant damage to neurons, resulting in various sequelae that negatively impact the patient's ability to perform essential daily activities such as chewing, swallowing, and verbal communication. Therefore, it is important for patients with such difficulties to undergo a treatment process and be monitored during its execution to assess the improvement of their health condition. The use of computerized tools and algorithms that can quickly and affordably detect such sequelae proves helpful in aiding the patient's recovery. Due to the death of internal brain cells, a stroke often leads to facial paralysis, resulting in certain asymmetry between the two sides of the face. This paper focuses on analyzing this asymmetry using a deep learning method without relying on handcrafted calculations, introducing the Facial Point Graphs (FPG) model, a novel approach that excels in learning geometric information and effectively handling variations beyond the scope of manual calculations. FPG allows the model to effectively detect orofacial impairment caused by a stroke using video data. The experimental findings on the Toronto Neuroface dataset revealed the proposed approach surpassed state-of-the-art results, promising substantial advancements in this domain.



Fonte: Elaborada pelo autor.

a captura, o usuário pode visualizar a sua imagem e a foto que foi capturada após apertar o botão ‘Take a picture’, como ilustrado na Figura 14. Dessa forma, o usuário pode ainda, selecionar se deseja usar a imagem capturada ou tirar outra foto.

Figura 14 – Captura da imagem: (a) tirar a foto e (b) escolher a foto.

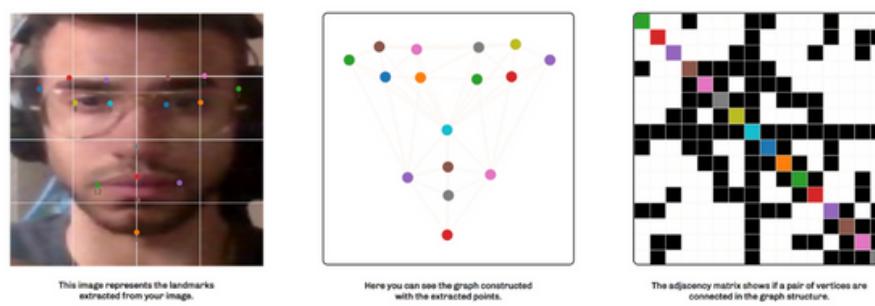


Fonte: Elaborada pelo autor.

### 6.3.3 Características extraídas

No caso do usuário capturar uma foto facial com sucesso, o processo continua para uma tela onde há a possibilidade de análise das características extraídas da imagem utilizada. No centro desta página, o usuário consegue verificar o grafo formado a partir da face e na parte mais a esquerda pode verificar os pontos extraídos juntos a imagem original. Ademais, na parte a direita da página é possível visualizar a matriz de adjacências que representa todas as ligações do grafo de forma intuitiva, como ilustrado na Figura 15.

Figura 15 – Página de demonstração das características extraídas.

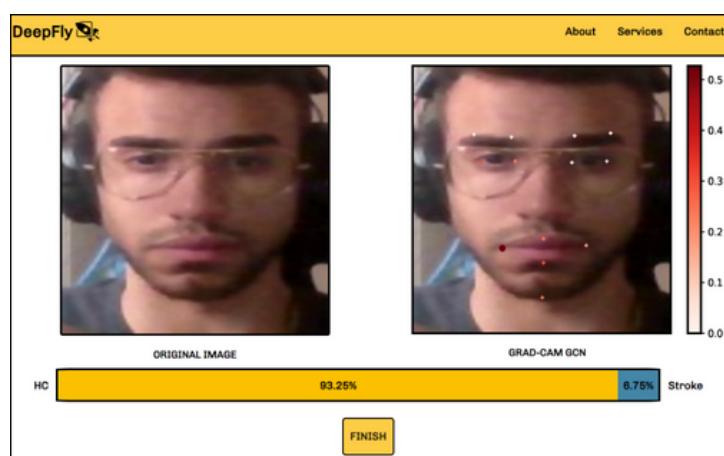


Fonte: Elaborada pelo autor.

### 6.3.4 Resultados inferidos

Por fim, o usuário tem o acesso a área de inferência quando este utiliza o botão de classificação da página anterior. O intuito desta página não é mostrar se o indivíduo tem ou não sequelas do AVC, mas sim disponibilizar as porcentagens de saída do modelo para ambas as classes. Cabe a um profissional da máquina fazer o diagnóstico, novamente, o aplicativo desenvolvido não tem o intuito de dar o diagnóstico de paciente e sim auxiliar profissionais no diagnóstico. Nesta página podem ser vistas a imagem original e uma com os pontos coloridos de acordo com os pesos do Grad-CAM, demonstrando os pontos mais relevantes na inferência do modelo, como ilustrado na Figura 16.

Figura 16 – Página de inferência.



Fonte: Elaborada pelo autor.

# 7 Considerações Finais

Este estudo representa o esforço pioneiro de avaliar GNNs na identificação de deficiências decorrentes de AVC por meio da análise de imagens faciais e expressões correspondentes. A aplicação das GNNs na extração de características representa uma via promissora para mitigar vieses humanos, abrindo novas perspectivas na detecção de deficiências resultantes de AVC. As sequelas resultantes de um AVC podem impactar significativamente as atividades diárias, como alimentação, interação social e outras tarefas simples, podendo levar à reclusão do indivíduo. A restauração das funções comprometidas se torna fundamental para mitigar esse isolamento social e recuperar a qualidade de vida do paciente. Neste sentido, novas abordagens como os modelos de FPG surgiram como meios promissores para identificar as deficiências orofaciais causadas pelo AVC, o que é crucial para um tratamento eficaz dos pacientes. A implementação da identificação automatizada por meio de métodos computacionais apresenta vantagens significativas, tais como o aprimoramento do tratamento para reduzir as sequelas, além de representar uma abordagem mais econômica para avaliação.

Embora as Redes Neurais Profundas frequentemente enfrentem desafios na generalização de conjuntos de dados menores, os modelos FPG, baseados em grafos, têm demonstrado resultados notáveis, alcançando taxas de acurácia acima de 90% na classificação de deficiências orofaciais decorrentes de acidente vascular cerebral. Além disso, o uso de dados bidimensionais para extração de características da face permite que a coleta dos dados seja feita por câmeras simples, ampliando o escopo para aplicações suportadas por dispositivos prontamente acessíveis, como celulares.

## 7.1 Trabalhos Futuros

Em trabalhos futuros, é válido que seja feita uma análise mais profunda das tarefas ‘BROW’ e ‘BIGSMILE’, aprimorando a capacidade do modelo em diferenciar pessoas saudáveis das que sofreram AVC, através de características faciais. Além disso, seria interessante realizar a análise em outras bases de dados mais extensas, com o objetivo de inferir deficiências faciais causadas por paralisias provenientes de diferentes doenças, como a paralisia de Bell ou tumores cerebrais. Essa expansão contribuiria para aumentar a robustez do modelo. Outra abordagem potencial poderia envolver a análise apenas dos quadros onde ocorre o movimento mais extenso, classificando especificamente a imagem da face. Estas explorações poderiam aperfeiçoar e otimizar o desempenho dos modelos FPG, fornecendo informações valiosas para a melhoria e a

continuidade do desenvolvimento da ferramenta desenvolvida neste trabalho. Tornando-a mais precisa e clinicamente aplicável.

Outro fator relevante a pesquisas na área da saúde é ter profissionais médicos juntos ao projeto. Neste sentido, torna-se possível realizar uma análise técnica, junto de um estudo mais complexo e profundo das tarefas realizadas e seus respectivos potenciais para diagnosticar a doença. Uma melhoria do projeto, pode ser feita contatando profissionais com especialidade em doenças neurológicas e suas possíveis sequelas, para que estes se juntem ao projeto.

# Referências

- 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: 17 jun. 2023.
- BANDINI, A.; GREEN, J. R.; RICHBURG, B.; YUNUSOVA, Y. Automatic detection of orofacial impairment in stroke. In: *Interspeech*. [s.n.], 2018. p. 1711–1715. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8373824>. Acesso em: 30 set. 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/abstract/document/9177259>. Acesso em: 20 mar. 2023.
- BENJAMIN, E. J.; BLAHA, M. J.; CHIUVE, S. E.; CUSHMAN, M.; DAS, S. R.; DEO, R.; FERRANTI, S. D. D.; FLOYD, J.; FORNAGE, M.; GILLESPIE, C. Heart disease and stroke statistics—2017 update: a report from the american heart association. *circulation*, Am Heart Assoc, v. 135, n. 10, p. e146–e603, 2017. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28122885/>. Acesso em: 20 set. 2023.
- BULAT, A.; TZIMIROPOULOS, G. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In: *International Conference on Computer Vision*. [s.n.], 2017. Disponível em: <https://ieeexplore.ieee.org/document/8237378>. Acesso em: 18 jun. 2023.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Kluwer Academic Publishers, USA, v. 20, p. 273–297, 1995. ISSN 0885-6125. Disponível em: <https://link.springer.com/article/10.1007/BF00994018>. Acesso em: 20 set. 2023.
- DELAUNAY, B. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, v. 7, n. 793-800, p. 1–2, 1934. Disponível em: [https://galiulin.narod.ru/delaunay\\_.pdf](https://galiulin.narod.ru/delaunay_.pdf). Acesso em: 25 jul. 2023.
- DHALL, A.; GOECKE, R.; LUCEY, S.; GEDEON, T. Collecting large, richly annotated facial-expression databases from movies. *IEEE multimedia*, v. 19, n. 3, p. 34, 2012. Disponível em: <https://ieeexplore.ieee.org/document/6200254>. Acesso em: 26 set. 2023.
- DONG, X.; YAN, Y.; OUYANG, W.; YANG, Y. Style aggregated network for facial landmark detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [s.n.], 2018. p. 379–388. Disponível em: <https://ieeexplore.ieee.org/document/8578145>. Acesso em: 30 out. 2023.
- FEY, M.; LENSSSEN, J. E. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019. Disponível em: <https://arxiv.org/abs/1903.02428>. Acesso em: 10 out. 2023.

- FIX, E.; HODGES, J. L. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, v. 57, n. 3, p. 238–247, 1989. Disponível em: <https://www.jstor.org/stable/1403797>. Acesso em: 25 out. 2023.
- FLOWERS, H. L.; SILVER, F. L.; FANG, J.; ROCHON, E.; MARTINO, R. The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. *Journal of communication disorders*, Elsevier, v. 46, n. 3, p. 238–248, 2013. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0021992413000117>. Acesso em: 5 set. 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: 5 set. 2023.
- GREENE, J. J.; GUARIN, D. L.; TAVARES, J.; FORTIER, E.; ROBINSON, M.; DUSSELDORP, J.; QUATELA, O.; JOWETT, N.; HADLOCK, T. The spectrum of facial palsy: The meeij facial palsy photo and video standard set. *The Laryngoscope*, Wiley Online Library, v. 130, n. 1, p. 32–37, 2020. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31021433/>. Acesso em: 12 jul. 2023.
- GUARIN, D. L.; YUNUSOVA, Y.; TAATI, B.; DUSSELDORP, J. R.; MOHAN, S.; TAVARES, J.; VEEN, M. M. van; FORTIER, E.; HADLOCK, T. A.; JOWETT, N. Toward an automatic system for computer-aided assessment in facial palsy. *Facial Plastic Surgery & Aesthetic Medicine*, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New ..., v. 22, n. 1, p. 42–49, 2020. Disponível em: <https://www.liebertpub.com/doi/abs/10.1089/fpsam.2019.29000.gua>. Acesso em: 16 jul. 2023.
- HAMILTON, W. L. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Morgan & Claypool Publishers, v. 14, n. 3, p. 1–159, 2020. Disponível em: [https://www.cs.mcgill.ca/~wlh/grl\\_book/](https://www.cs.mcgill.ca/~wlh/grl_book/). Acesso em: 18 mar. 2023.
- HAYKIN, S. *Neural networks: a comprehensive foundation*. [S.I.]: Prentice Hall PTR, 1994.
- HO, T. K. Random decision forests. In: IEEE. *Proceedings of 3rd international conference on document analysis and recognition*. 1995. v. 1, p. 278–282. Disponível em: <https://ieeexplore.ieee.org/document/598994>. Acesso em: 25 abr. 2023.
- KAEMAHANIN, W.; RASSAMEECHAROENCHAI, T.; JUTHAREE, W.; TONGSKUL-ROONGRUANG, T.; WIPHUNAWAT, P.; JENNAWASIN, T.; KAEWKAMNERDPONG, B. Automatic facial asymmetry analysis for elderly stroke detection by using cosine similarity. In: IEEE. *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2022. p. 1–4. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9795508>. Acesso em: 6 set. 2023.
- KAZEMI, V.; SULLIVAN, J. One millisecond face alignment with an ensemble of regression trees. In: *Proceedings of the IEEE conference on computer vision and pattern*

- recognition.* [s.n.], 2014. p. 1867–1874. Disponível em: [https://openaccess.thecvf.com/content\\_cvpr\\_2014/html/Kazemi\\_One\\_Millisecond\\_Face\\_2014\\_CVPR\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2014/html/Kazemi_One_Millisecond_Face_2014_CVPR_paper.html). Acesso em: 25 out. 2023.
- KIM, H. S.; KIM, S. Y.; KIM, Y. H.; PARK, K. S. A smartphone-based automatic diagnosis system for facial nerve palsy. *Sensors*, MDPI, v. 15, n. 10, p. 26756–26768, 2015. Disponível em: <https://www.mdpi.com/1424-8220/15/10/26756>. Acesso em: 20 jul. 2023.
- KING, D. E. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, JMLR.org, v. 10, p. 1755–1758, 2009. Disponível em: <https://www.jmlr.org/papers/volume10/king09a/king09a.pdf>. Acesso em: 25 jun. 2023.
- KIPF, T. N.; WELLING, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016. Disponível em: <https://arxiv.org/abs/1609.02907>. Acesso em: 10 out. 2023.
- LAPCHAK, P. A.; ZHANG, J. H. The high cost of stroke and stroke cytoprotection research. *Translational stroke research*, Springer, v. 8, p. 307–317, 2017. Disponível em: <https://link.springer.com/article/10.1007/s12975-016-0518-y>. Acesso em: 28 set. 2023.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <https://ieeexplore.ieee.org/abstract/document/726791>. Acesso em: 15 out. 2023.
- LOU, J.; YU, H.; WANG, F.-Y. A review on automated facial nerve function assessment from visual face capture. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, IEEE, v. 28, n. 2, p. 488–497, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8945478>. Acesso em: 18 abr. 2023.
- LUCEY, P.; COHN, J. F.; KANADE, T.; SARAGIH, J.; AMBADAR, Z.; MATTHEWS, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: IEEE. *2010 ieee computer society conference on computer vision and pattern recognition-workshops*. 2010. p. 94–101. Disponível em: <https://ieeexplore.ieee.org/document/5543262>. Acesso em: 26 out. 2023.
- MULLEN, M.; LOOMIS, C. Differentiating facial weakness caused by bell's palsy vs. acute stroke. *JEMS*, 2014. Disponível em: <https://www.jems.com/patient-care/differentiating-facial-weakness-caused-by-bells-palsy-vs-acute-stroke>. Acesso em: 25 set. 2023.
- NGOC, Q. T.; LEE, S.; SONG, B. C. Facial landmark-based emotion recognition via directed graph neural network. *Electronics*, MDPI, v. 9, n. 5, p. 764, 2020. Disponível em: <https://www.mdpi.com/2079-9292/9/5/764>. Acesso em: 27 out. 2023.
- PANTIC, M.; VALSTAR, M.; RADEMAKER, R.; MAAT, L. Web-based database for facial expression analysis. In: IEEE. *2005 IEEE international conference on multimedia and Expo*. 2005. p. 5–pp. Disponível em: <https://ieeexplore.ieee.org/abstract/document/1521424>. Acesso em: 26 jul. 2023.
- PARRA-DOMINGUEZ, G. S.; SANCHEZ-YANEZ, R. E.; GARCIA-CAPULIN, C. H. Facial paralysis detection on images using key point analysis. *Applied Sciences*, MDPI,

v. 11, n. 5, p. 2435, 2021. Disponível em: <https://www.mdpi.com/2076-3417/11/5/2435>. Acesso em: 22 set. 2023.

PECUNDO, A. M.; ABU, P. A.; ALAMPAY, R. Amyotrophic lateral sclerosis and post-stroke orofacial impairment video-based multi-class classification. In: *Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference*. [s.n.], 2022. p. 150–157. Disponível em: <https://dl.acm.org/doi/10.1145/3582099.3582123>. Acesso em: 2 nov. 2023.

POPE, P. E.; KOLOURI, S.; ROSTAMI, M.; MARTIN, C. E.; HOFFMANN, H. Explainability methods for graph convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [s.n.], 2019. p. 10772–10781. Disponível em: [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Pope\\_Explainability\\_Methods\\_for\\_Graph\\_Convolutional\\_Neural\\_Networks\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Pope_Explainability_Methods_for_Graph_Convolutional_Neural_Networks_CVPR_2019_paper.html). Acesso em: 2 set. 2023.

PRADOS-TORREBLANCA, A.; BUENAPOSADA, J. M.; BAUMELA, L. Shape preserving facial landmarks with graph attention networks. In: *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press, 2022. Disponível em: <https://arxiv.org/abs/2210.07233>. Acesso em: 16 jun. 2023.

SAMSUDIN, W. W.; SUNDARAJ, K. Image processing on facial paralysis for facial rehabilitation system: A review. In: *IEEE. 2012 IEEE International Conference on Control System, Computing and Engineering*. 2012. p. 259–263. Disponível em: <https://ieeexplore.ieee.org/abstract/document/6487152>. Acesso em: 25 ago. 2023.

SCARSELLI, F.; GORI, M.; TSOI, A. C.; HAGENBUCHNER, M.; MONFARDINI, G. The graph neural network model. *IEEE transactions on neural networks*, IEEE, v. 20, n. 1, p. 61–80, 2008. Disponível em: <https://ieeexplore.ieee.org/abstract/document/4700287>. Acesso em: 5 set. 2023.

SCHIMMEL, M.; ONO, T.; LAM, O.; MÜLLER, F. Oro-facial impairment in stroke patients. *Journal of oral rehabilitation*, Wiley Online Library, v. 44, n. 4, p. 313–326, 2017. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/joor.12486>. Acesso em: 22 jul. 2023.

VELIČKOVIĆ, P.; CUCURULL, G.; CASANOVA, A.; ROMERO, A.; LIO, P.; BENGIO, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017. Disponível em: <https://arxiv.org/abs/1710.10903>. Acesso em: 25 out. 2023.

XIONG, X.; TORRE, F. De la. Supervised descent method and its applications to face alignment. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [s.n.], 2013. p. 532–539. Disponível em: [https://openaccess.thecvf.com/content\\_cvpr\\_2013/html/Xiong\\_Supervised\\_Descent\\_Method\\_2013\\_CVPR\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2013/html/Xiong_Supervised_Descent_Method_2013_CVPR_paper.html). Acesso em: 20 jul. 2023.