Classificação de alto nível baseada em meta-grafos de grafos de visibilidade para detecção do Autismo

Ricardo Barbosa Lima Filho

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2026

Ricardo Barbosa Lima Filho

Classificação de alto nível baseada em meta-grafos de grafos de visibilidade para detecção do Autismo

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Murillo Guimarães Carneiro

Uberlândia
2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

L732 Lima Filho, Ricardo Barbosa, 2000-
2026 Classificação de alto nível baseada em meta-grafos de grafos de visibilidade para detecção do Autismo [recurso eletrônico] / Ricardo Barbosa Lima Filho. - 2026.

Orientador: Murillo Guimarães Carneiro.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
DOI http://doi.org/10.14393/ufu.di.2026.160
Inclui bibliografia.
Inclui ilustrações.

1. Computação. I. Carneiro, Murillo Guimarães,1988-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Computação
Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação, 09/2026, PPGCO				
Data:	12 de Fevereiro de 2026	Hora de início:	09:05	Hora de encerramento:	11:35
Matrícula do Discente:	12312CCP031				
Nome do Discente:	Ricardo Barbosa Lima Filho				
Título do Trabalho:	Classificação de alto nível baseada em meta-grafos de grafos de visibilidade para detecção do Autismo				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Inteligência Artificial				
Projeto de Pesquisa de vinculação:	Processo: 408216/2022-0 Vigência: início: 28/11/2022 fim: 30/11/2026 Título: Classificação de Alto Nível para Diagnóstico do Transtorno do Espectro Autista através de Plataforma Biofotônica Instituição de Execução: Universidade Federal de Uberlândia CNPJ: 25648387000118 Ação: Chamada CNPq/SEMPI/MCTI Nº 57/2022 Processo: 445027/2024-0 Vigência: início: 29/01/2025 fim: 31/01/2027 Título: Aprendizado de máquina automático para detecção não invasiva do Transtorno do Espectro Autista Instituição de Execução: Universidade Federal de Uberlândia CNPJ: 25648387000118 Ação: Chamada CNPq/MCTI/FNDCT Nº 22/2024 Processo: 420212/2023-0 Vigência: início: 06/12/2023 fim: 31/12/2026 Título: Arquiteturas híbridas de aprendizado em redes para problemas de classificação mono e multi sequenciais com aplicações em dados de espectroscopia no infravermelho e de Eletroencefalograma Instituição de Execução: Universidade Federal de Uberlândia CNPJ: 25648387000118 Ação: Chamada CNPq/MCTI Nº 10/2023				

Reuniu-se presencial, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Márcia Aparecida Fernandes - FACOM/UFU, Ricardo Marcondes Marcacini - ICMC/USP e Murillo Guimarães Carneiro - FACOM/UFU, orientador do(a) candidato(a).

Os examinadores participaram desde as seguintes localidades: Todos os membros da banca e o aluno(a) participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Murillo Guimarães Carneiro, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(à) Discente a palavra para a exposição do seu trabalho.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o() candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Márcia Aparecida Fernandes**, **Professor(a) do Magistério Superior**, em 19/02/2026, às 11:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **RICARDO MARCONDES MARCACINI**, **Usuário Externo**, em 19/02/2026, às 11:20, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Murillo Guimarães Carneiro**, **Professor(a) do Magistério Superior**, em 19/02/2026, às 15:16, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7047438** e o código CRC **AE664C8D**.

Referência: Processo nº 23117.007017/2026-23 SEI nº 7047438

Dedico este trabalho a todos os que me ajudaram ao longo desta caminhada.

Agradecimentos

A jornada do mestrado foi marcada por desafios, aprendizados e crescimento, e nenhuma parte dela teria sido possível sem o apoio e a colaboração de pessoas muito especiais.

Agradeço, primeiramente, a Deus, por me conceder força, saúde e perseverança ao longo deste caminho.

Ao meu orientador, Dr. Murillo Guimarães Carneiro, pela orientação dedicada, paciência e por sempre acreditar no meu potencial. Seus ensinamentos acadêmicos e éticos foram fundamentais para a construção deste trabalho e para minha formação como pesquisador.

Aos membros do programa de pós-graduação, professores e funcionários, pelo conhecimento compartilhado, pelo suporte acadêmico e por contribuírem para um ambiente de aprendizado inspirador.

À minha família, em especial aos meus pais, por todo amor, apoio incondicional e por sempre me incentivarem a buscar meus sonhos. Sem vocês, nada disso faria sentido.

Aos colegas de pesquisa e amigos que fiz durante esse período, pelo companheirismo, pelas trocas de ideias e pelo apoio mútuo nos momentos difíceis.

Por fim, agradeço às instituições de fomento, como CNPq, pelo apoio financeiro que possibilitou o desenvolvimento desta pesquisa.

A todos que, de alguma forma, contribuíram para esta conquista, deixo aqui o meu mais sincero agradecimento.

"A arte diz o indizível; exprime o inexprimível, traduz o intraduzível."
(Leonardo Da Vinci)

Resumo

Este trabalho explora técnicas avançadas de classificação para a detecção do Transtorno do Espectro Autista (TEA) utilizando espectroscopia ATR-FTIR aplicada a amostras de saliva. A plataforma proposta é rápida, sustentável e permite a coleta não invasiva de amostras. O espectro ATR-FTIR gerado é uma sequência de alta dimensionalidade que representa interações moleculares, mas cuja natureza sequencial é frequentemente ignorada em estudos da área.

Enquanto métodos tradicionais com ATR-FTIR utilizam classificadores de baixo nível baseados em atributos físicos, abordagens de alto nível mais recentes incorporam características topológicas e estruturais. No entanto, esses métodos enfrentam limitações na construção de grafos, especialmente na representação das relações sequenciais dos dados.

Para superar esse desafio, propomos o VisG2, um método de construção de meta-grafos que representa espectros como grafos de visibilidade e calcula a similaridade entre eles para gerar uma estrutura de alto nível. Avaliações com dados reais de TEA mostram que o VisG2 supera classificadores tradicionais e modernos, como SVM, CNN e Transformers, além de outras técnicas de construção de grafos.

Além disso, quando combinado com métodos de baixo nível, o VisG2 compõe um classificador híbrido que melhora ainda mais o desempenho da abordagem de alto nível. Os resultados confirmam a eficácia do VisG2 na captura de relações sequenciais e seu potencial para a detecção do TEA e aplicações similares.

Palavras-chave: Redes Complexas, Classificação de Alto Nível, Medidas de Redes Complexas, Grafo de Visibilidade, Transtorno do Espectro Autista, ATR-FTIR.

Abstract

This study explores advanced classification techniques for the detection of Autism Spectrum Disorder (ASD) using ATR-FTIR spectroscopy applied to saliva samples. The proposed platform is fast, sustainable, and enables non-invasive sample collection. The resulting ATR-FTIR spectrum is a high-dimensional sequence that encodes molecular interactions, yet its sequential nature is often overlooked in related studies.

While traditional ATR-FTIR approaches rely on low-level classifiers based on physical attributes, more recent high-level methods incorporate topological and structural features. However, these high-level methods face challenges in the graph construction phase, particularly in accurately capturing the sequential relationships within the data.

To address this limitation, we propose VisG2, a meta-graph construction method that represents spectra as visibility graphs and computes their similarities to build a high-level structure. Experiments with real ASD data demonstrate that VisG2 outperforms both traditional and modern classifiers such as SVM, CNN and Transformers, as well as other commonly used graph construction techniques.

Furthermore, when combined with low-level methods, VisG2 forms a hybrid classifier that further enhances the performance of the original high-level approach. The results confirm VisG2's effectiveness in capturing sequential relationships and highlight its potential for ASD detection and other applications involving sequential data.

Keywords: Complex Networks, High-Level Classification, Complex Network Measurements, Visibility Graph, Autism Spectrum Disorder, ATR-FTIR.

Lista de ilustrações

Lista de tabelas

Lista de siglas

AA	Acurácia
ADG	Grau Médio, Average Degree
AM	Aprendizado Máquina
AS	Aprendizado Supervisionado
ASS	Assortatividade, Assortativity
ASP	Menor Caminho Médio, Shortest Average Path
ATR-FTIR	Espectroscopia de Infravermelho com Transformada de Fourier e Reflectância Total Atenuada
BET	Intermedialidade, Betweenness
CART	Classification and Regression Trees
CCF	Coeficiente de Agrupamento, Cluster Coefficient
CLO	Proximidade, Closeness
CNN	Redes Neurais Convolucionais
EIR	Espectroscopia de Infravermelho
F1	F1-Score
HL	Alto Nível, High Level
KNN	K-Vizinhos Mais Próximos
LDA	Análise Discriminante Linear
LL	Baixo Nível, Low Level
MLP	Perceptron Multicamadas
NB	Naive Bayes

RF	Floresta Aleatória
SE	Sensibilidade
SP	Especificidade
SVM	Máquina de Vetores de Suporte
TEA	Transtorno do Espectro Autista
VisG2	Visibility Graphs Graph

Sumário

Capítulo 1

Introdução

O Transtorno do Espectro Autista (TEA) afeta áreas do neurodesenvolvimento responsáveis pela interação social, comunicação e comportamento. Dessa forma, intervenções específicas e precoces são essenciais para potencializar o desenvolvimento infantil, reduzir a possibilidade de cronificação e ampliar as opções terapêuticas (ZANON; BACKES; BOSA, 2014; ASSOCIATION et al., 2014; QIN et al., 2024; SHAW, 2025).

Nesse contexto, métodos de diagnóstico eficientes e acessíveis são fundamentais para a detecção precoce do TEA. A espectroscopia infravermelha de biofluidos tem se destacado como uma abordagem promissora para o diagnóstico sem reagentes de diversas doenças e distúrbios, incluindo COVID-19, HIV e TEA (BARAUNA et al., 2021; SILVA et al., 2020; NASEER; ALI; QAZI, 2021). Em particular, a coleta de saliva permite a aplicação dessa técnica em plataformas de triagem e diagnóstico, oferecendo um método não invasivo, de baixo custo, indolor e de fácil autocoleta. A Espectroscopia de Infravermelho com Transformada de Fourier e Reflectância Total Atenuada (ATR-FTIR) já demonstrou eficácia na detecção de diversos tipos de câncer, sendo valorizada por sua rapidez, dispensa de reagentes e abordagem sustentável, característica das chamadas tecnologias verdes (FERREIRA et al., 2020; CAIXETA et al., 2020; CAIXETA et al., 2023; TINTOR et al., 2024; KHALIGHI et al., 2024). O espectro ATR-FTIR gerado consiste em uma sequência de alta dimensionalidade, na qual cada ponto espectral corresponde à intensidade de absorção associada a vibrações moleculares específicas em diferentes números de onda. Essa representação carrega informações ricas sobre a composição química e as interações moleculares da amostra, apresentando uma estrutura naturalmente sequencial, uma vez que os valores estão ordenados ao longo do domínio espectral. No entanto, em grande parte dos estudos da área, essa característica sequencial é frequentemente desconsiderada, sendo o espectro tratado apenas como um vetor de atributos independentes, o que pode levar à perda de informações relevantes relacionadas a padrões locais, correlações entre bandas adjacentes e dependências estruturais ao longo do espectro.

Além dos avanços na espectroscopia, a análise de dados baseada em redes complexas tem se mostrado uma ferramenta poderosa na identificação de padrões em diversas áreas.

O termo Redes Complexas refere-se a estruturas compostas por um conjunto de vértices (nós) interligados por arestas (links), permitindo a representação de diferentes fenômenos do mundo real (CARNEIRO, 2016). Uma das principais vantagens dessa abordagem é a possibilidade de analisar padrões de alto nível dos dados, indo além das características físicas individuais (FERNANDES; OLIVEIRA; CARNEIRO, 2023). Técnicas de classificação que exploram a estrutura da rede são conhecidas na literatura como classificadores de Alto Nível, High Level (HL) (CARNEIRO; ZHAO, 2018). No entanto, esses métodos enfrentam limitações na construção de grafos, especialmente na representação das relações sequenciais dos dados.

Diante desse contexto, este trabalho apresenta uma técnica baseada em meta-grafo de grafos de visibilidade (VisG2, do inglês Visibility Graphs Graph), um método de construção de meta-grafos desenvolvido para representar relações sequenciais tanto dentro de um único espectro quanto entre múltiplos espectros. O Visibility Graphs Graph (VisG2) é composto por dois componentes principais: a conversão de cada espectro individual em um grafo de visibilidade e um procedimento de geração do meta-grafo, baseado na similaridade estrutural entre esses grafos. Para a etapa de classificação de HL, foi adotada a técnica baseada em conformidade padrão, fundamentada em medidas de redes complexas (SILVA; ZHAO, 2012). Essa abordagem tem se mostrado promissora, especialmente na identificação de padrões associados ao TEA e a outras condições, ao realizar a classificação com base na variação das propriedades topológicas da rede após a inserção de um novo elemento (CARNEIRO, 2016). Essa característica permite uma análise mais sensível e detalhada da estrutura dos dados. Adicionalmente, propomos um classificador híbrido que combina as contribuições de alto e baixo nível (SILVA; ZHAO, 2012), visando aprimorar o desempenho preditivo por meio da integração de diferentes perspectivas analíticas sobre os dados.

1.1 Motivação

A literatura apresenta diversas técnicas para a classificação de dados, como redes neurais, árvores de decisão e máquinas de vetores de suporte, entre outras. Em geral, essas abordagens realizam a classificação com base exclusivamente nos atributos físicos dos dados de entrada, considerando fatores como similaridade, distância ou distribuição. No entanto, essa abordagem ignora as relações semânticas e topológicas entre os itens de dados. Por outro lado, os classificadores baseados em redes complexas permitem a extração de informações a partir dos padrões estruturais da rede, indo além das características físicas individuais dos dados. Técnicas de classificação que utilizam a topologia da rede para extrair conhecimento são conhecidas na literatura como classificadores de HL (SILVA; ZHAO, 2012).

Técnicas de classificação baseadas em redes permitem identificar padrões que frequen-

temente passam despercebidos por abordagens de Baixo Nível, Low Level (LL) (CARNEIRO et al., 2019). Um dos aspectos mais críticos da classificação de HL é a modelagem da rede, que possibilita o mapeamento de sistemas heterogêneos e complexos (NEWMAN, 2003). A maioria das técnicas de construção de redes utiliza algoritmos como k-vizinhos mais próximos e vizinhança de raio-ϵ, que baseiam a formação da rede na proximidade dos dados (CARNEIRO; ZHAO, 2018). No entanto, essas abordagens frequentemente falham em capturar características sequenciais dos dados, limitando sua capacidade de representar informações temporais ou estruturais mais profundas.

No trabalho (ARAúJO; SABINO-SILVA; CARNEIRO, 2024), são exploradas propriedades de redes complexas para a classificação de dados sequenciais ATR-FTIR, visando a identificação do TEA. O estudo se baseia em um classificador baseado em redes, utilizando caracterização de importância e variações do método de construção de redes por vizinhos mais próximos. No entanto, ele não explora as propriedades sequenciais dos dados ATR-FTIR, nem combina associações de alto e baixo nível.

Na pesquisa (FERNANDES; SABINO-SILVA; CARNEIRO, 2025), é proposta uma abordagem baseada em redes para a identificação do TEA a partir de dados espectrais ATR-FTIR. O método emprega um algoritmo genético para gerar e otimizar a topologia das redes, as quais são posteriormente caracterizadas por medidas de centralidade, como Grau e PageRank, utilizadas como atributos no processo de classificação. Embora os dados ATR-FTIR apresentem natureza sequencial, o trabalho os trata como vetores de atributos, não explorando explicitamente propriedades sequenciais do espectro. Além disso, a abordagem não contempla a combinação de associações de baixo e alto nível, concentrando-se exclusivamente na caracterização estrutural das redes otimizadas.

No trabalho (ANDRADE; SABINO-SILVA; CARNEIRO, 2024), apresenta um método para diagnóstico não invasivo do diabetes tipo 2 a partir de espectros salivais obtidos por ATR-FTIR. A principal contribuição é a aplicação de uma técnica de busca evolutiva de arquitetura neural (Neural Architecture Search, NAS), usando um algoritmo genético para encontrar arquiteturas de Redes Neurais Convolucionais (CNN) mais adequadas para classificar os espectros e distinguir pacientes diabéticos de não diabéticos. Essa abordagem visa superar a dificuldade de definir manualmente arquiteturas de CNN eficientes para esse tipo de dado espectral, uma lacuna pouco explorada na literatura de ATR-FTIR, ao mesmo tempo em que assume que as janelas de vizinhança da CNN podem contribuir para o mapeamento de características relacionais dos dados. O trabalho assume técnicas profundas de Aprendizado Máquina (AM) para extração automática de características, e não explora de forma explícita representações estruturais ou topológicas dos dados.

Nesse contexto, este trabalho propõe a construção de um classificador de HL baseado em conformidade padrão, que combina classificações de alto e baixo nível para melhorar o desempenho. Para considerar os aspectos sequenciais dos dados de entrada, é introduzida uma heurística de construção de rede VisG2 que é baseada na similaridade entre grafos de

visibilidade. Nessa abordagem, cada instância é representada por um grafo de visibilidade, e cada grafo de visibilidade atua como um super nó na rede complexa associada a cada classe, de modo que relações sequenciais dos dados são mapeadas diretamente pelos grafos de visibilidade e relações de afinidade entre as instâncias pelo meta-grafo.

1.2 Hipótese e Objetivos

A análise de características estruturais e topológicas de dados pode contribuir para melhorar a detecção dos diferentes transtornos do espectros do autista, em relação aos modelos de classificação atuais, essencialmente baseados nas características físicas desses dados (SPORNS, 2016; BULLMORE; SPORNS, 2009; RUDIE et al., 2013; KAZEMINEJAD; SOTERO, 2019). **Ao combinar um framework de HL e um algoritmo de LL, a habilidade preditiva e de detecção de padrões será aperfeiçoada**. Dessa forma, o objetivo geral desta pesquisa é desenvolver um algoritmo de classificação de HL baseado na conformidade de padrões, integrando uma técnica de construção de redes complexas capaz de capturar características sequenciais dos dados ATR-FTIR por meio de grafos de visibilidade. Ao final, o método combina associações de alto e baixo nível por meio de um classificador híbrido. Os objetivos específicos são apresentados a seguir:

- ❏ Investigar maneiras de tratar adequadamente dados FTIR, seja por normalização da banda, suavização dos dados, ou até mesmo considerar somente as bandas mais relevantes para o processo de classificação.

- ❏ Desenvolver uma técnica de construção de redes complexas por meio de um meta-grafos de grafos de visibilidade.

- ❏ Investigar se a incorporação de grafos de visibilidade em uma estrutura de meta-grafo possibilita a exploração explícita do aspecto sequencial dos dados, resultando em uma representação em rede potencialmente mais informativa.

- ❏ Avaliar se redes construídas a partir de meta-grafos de grafos de visibilidade apresentam maior potencial discriminativo em comparação às redes obtidas por abordagens baseadas exclusivamente em medidas de distância.

- ❏ Desenvolver um classificador híbrido capaz de associar características de alto produzidas pelas medidas de rede e LL produzidas por classificadores tradicionais da literatura.

- ❏ Explorar medidas e propriedades de redes complexas que podem ser usadas para melhorar os algoritmos convencionais de classificação.

1.3 Contribuições

As principais contribuições deste trabalho podem ser destacadas da seguinte forma:

- ❏ Um algoritmo de classificação capaz de considerar tanto aspectos físicos quanto topológicos dos dados, para o problema de detecção do TEA via espectroscopia ATR-FTIR de amostras de saliva.

- ❏ Avaliação de diferentes conceitos e métodos de construção de redes complexas, investigando qual abordagem é mais adequada ao problema em estudo e analisando as características estruturais resultantes de cada rede, bem como suas diferenças em relação às demais.

- ❏ Definição um método de construção de rede complexa capaz de considerar as características sequenciais do dados ATR-FTIR através de grafos de visibilidade.

- ❏ Comparação dos resultados com diferentes classificadores do estado da arte presentes na literatura. Para validar os resultados obtidos será feito uma comparação com os resultados obtidos por classificadores como Análise Discriminante Linear (LDA), Máquina de Vetores de Suporte (SVM), K-Vizinhos Mais Próximos (KNN), Naive Bayes (NB), Floresta Aleatória (RF), Classification and Regression Trees (CART), Perceptron Multicamadas (MLP), CNN e Transformer.

1.4 Organização da Dissertação

O trabalho está organizado nos capítulos de fundamentação teórica, método proposto, resultados e considerações finais, brevemente explicados nos tópicos seguintes:

- ❏ O Capítulo 2 apresenta a fundamentação teórica necessária para a compreensão deste trabalho, incluindo pesquisas relacionadas à proposta. Inicialmente, são discutidos alguns classificadores tradicionais da literatura. Em seguida, são abordados os conceitos de redes complexas e estudos correlatos. Por fim, são apresentados os principais fundamentos da ATR-FTIR.

- ❏ O Capítulo 3 detalha a proposta deste trabalho, que consiste em um classificador de HL baseado em uma técnica de construção de redes complexas utilizando grafos de visibilidade, combinando associações de alto e baixo nível.

- ❏ O Capítulo 4 apresenta e discute os resultados obtidos com a aplicação da técnica proposta ao problema de análise de dados salivares ATR-FTIR para a detecção do Câncer Oral.

- O Capítulo 5 apresenta os resultados obtidos com a técnica de HL proposta, incluindo uma análise exploratória de seus parâmetros e das métricas de desempenho preditivo, para a detecção do Autismo.
- O Capítulo 6 traz as considerações finais, destacando os principais achados da pesquisa, os desafios enfrentados e as direções para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Nesse capítulo será apresentado todos os conceitos relevantes para o trabalho de detecção do TEA através da classificação baseada em redes complexas. A Seção 2.1 aborda os conceitos fundamentais sobre a tarefa de classificação no Aprendizado Supervisionado (AS), bem como a explicação de alguns classificadores tradicionais e medidas avaliativas. Na Seção 2.2 aborda os conceitos sobre redes complexas, bem como medidas de rede e métodos de construção de redes complexas, finalizando com os trabalhos relacionados. Na Seção 2.3 é discutido sobre a espetroscopia de infra vermelho ATR-FTIR. Na Seção 2.4 são discutidos os trabalhos relacionados a pesquisa proposta. Por fim, na Seção 2.5 é feito uma recapitulação sobre o capítulo.

2.1 Classificação de dados em aprendizado supervisionado

O AS é uma das principais abordagens do AM, caracterizado pelo uso de um conjunto de dados rotulado para treinar modelos capazes de realizar previsões sobre novos dados. Dentre as tarefas mais comuns nesse contexto, destaca-se a classificação, que consiste em atribuir rótulos predefinidos a instâncias com base em seus atributos. A classificação desempenha um papel fundamental em diversas aplicações, como reconhecimento de padrões, análise de sentimentos, diagnóstico médico e detecção de fraudes (BISHOP; NASRABADI, 2006).

Tarefas de classificação de dados envolvem, geralmente, três etapas principais: pré-processamento, treinamento e teste. Na fase de pré-processamento, busca-se garantir que os dados estejam limpos, organizados e formatados corretamente para o uso em algoritmos de aprendizado. Nessa etapa, são tratados problemas como ruídos, valores ausentes e inconsistências de formato, que podem comprometer a qualidade do modelo. Durante a etapa de treinamento, o modelo é construído a partir de um conjunto de dados rotulados, ou seja, com as classes previamente definidas, permitindo que ele aprenda a reconhecer

padrões para posteriormente classificar novos registros (CARNEIRO, 2016). Por fim, na fase de teste, o modelo tenta prever a classe de instâncias cuja rotulação é desconhecida, validando sua capacidade de generalização. O desempenho obtido é então avaliado por meio de métricas específicas, que serão detalhadas na Seção 2.1.10.

Nas próximas seções, serão apresentadas técnicas de classificação amplamente consolidadas na literatura, selecionadas por contemplarem diferentes paradigmas de aprendizado incluindo abordagens lineares, não lineares, modelos baseados em margens, em probabilidades e em estruturas hierárquicas. A inclusão dessas técnicas justifica-se pela necessidade de comparar o desempenho do modelo proposto com métodos de diferentes níveis de complexidade e capacidades de generalização. Todas serão implementadas sob o mesmo protocolo experimental, utilizando as mesmas bases de dados e métricas avaliativas, de modo a permitir uma análise comparativa justa e consistente dos resultados obtidos.

2.1.1 Análise discriminante linear

A LDA é uma técnica estatística utilizada para redução de dimensionalidade e classificação. A LDA é eficaz em situações onde as frequências dentro de cada classe diferem e seu desempenho é avaliado em dados de teste gerados de forma aleatória. Este procedimento aumenta a relação entre a variância entre classes e a variância dentro das classes em um conjunto de dados específico, assegurando, dessa forma, a máxima segregabilidade (BALAKRISHNAMA; GANAPATHIRAJU, 1998).

A LDA presume que as informações de cada classe seguem uma distribuição normal e que todas as classes possuem a mesma matriz de correlação (BALAKRISHNAMA; GANAPATHIRAJU, 1998). Essas premissas possibilitam que a técnica simule de maneira eficaz as diferenças entre as classes, visando maximizar a relação entre a variância entre classes e a variância dentro delas, assegurando que as classes estejam o mais separadas possível no espaço limitado. Finalmente, a LDA gera combinações lineares de variáveis preditoras conhecidas como funções discriminantes, que atuam como novos eixos no espaço modificado. O número de tais funções é restringido pelo valor mais baixo entre (número de classes - 1) ou (número de variáveis de previsão) (YE; JANARDAN; LI, 2004).

A Figura 1 mostra um gráfico com duas classes: x e bolinha. Temos duas funções discriminantes LD2 e LD1, em que a LD2 é uma função não muito interessante para classificar as duas classes, não gerando separabilidade. Já a função LD1 consegue separar muito bem as duas classes, sendo assim mais interessante no processo de classificação.

2.1.2 Máquina de vetores de suporte

A técnica SVM é uma ferramenta poderosa de AM supervisionado usado principalmente para classificação e regressão. Ele é muito eficaz para separar dados em diferentes categorias, especialmente quando os dados não são facilmente separáveis de forma linear

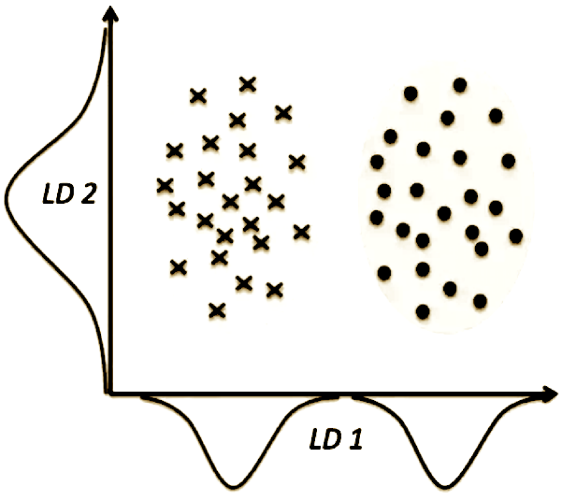


Figura 1 – Funções discriminantes para as duas classes.

Fonte: (NJUKI, 2023)

(PISNER; SCHNYER, 2020). O propósito do SVM é descobrir um hiperplano ótimo que aumente a margem entre as diversas classes no espaço de entrada. Este hiperplano é selecionado de maneira a minimizar a distância entre ele e os pontos mais próximos de cada classe (os vetores de suporte) (JAKKULA, 2006). Isso assegura uma generalização mais eficaz para novos dados.

O conceito central do SVM é descobrir um hiperplano ideal que separe os dados da maneira mais eficiente possível, maximizando a margem entre as classes. Alguns conceitos fundamentais acerca do classificador incluem:

- Hiperplano: Uma linha (para dados bidimensionais), um plano (3D) ou um espaço de dimensões superiores em situações mais complexas.
- Margem: É a distância que separa o hiperplano dos pontos de dados mais próximos de cada classe.
- Vetores de suporte: São os pontos mais próximos do hiperplano que determinam a sua localização.

O SVM usa um conceito matemático chamado margem máxima. Ele busca um hiperplano que maximize a distância entre as classes. Esse hiperplano é definido como:

$$w \cdot x + b = 0 \tag{1}$$

Em que w é o vetor de pesos (direção do hiperplano), x são os dados própriamente ditos (características dos dados) e b é o viés. O objetivo é minimizar $||w||$ enquanto garante que os dados sejam corretamente classificados. A Figura 2 exemplifica o processo realizado pelo SVM, em que o valor minimizado para $||w||$ foi de 2 para duas classes classificadas.

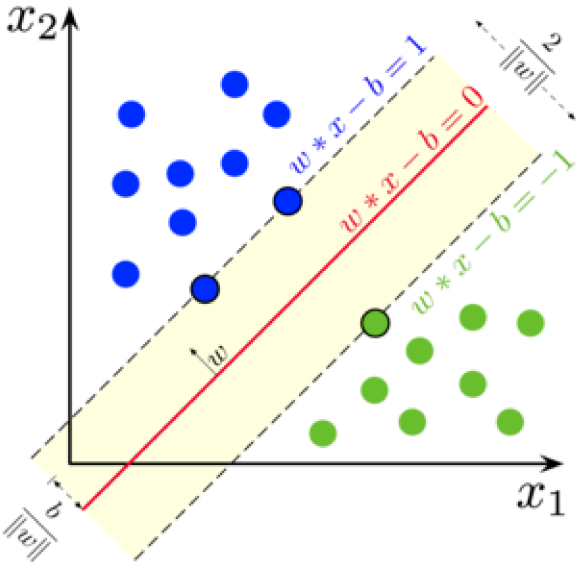


Figura 2 – Hiperplano de margem máxima e margens para um SVM treinado com amostras de duas categorias. Os vetores de suporte são conhecidos como amostras na margem.

Se os dados não forem linearmente separáveis, o SVM usa um truque chamado Kernel Trick para mapear os dados para um espaço de dimensão maior, onde eles se tornam separáveis (Sánchez A, 2003). Alguns kernels comuns são: kernel Linear (para dados linearmente separáveis), kernel Polinomial (para relações não lineares) e kernel Radial Basis Function - RBF (muito usado para dados complexos). O kernel modifica os dados sem a necessidade explícita de ampliar a dimensão.

2.1.3 Naive Bayes

O NB é um algoritmo fundamentado no Teorema de Bayes, que postula que todas as características são independentes dentro de uma classe. O NB provou ser eficiente em várias funções de classificação, incluindo a filtragem de spam, a avaliação de sentimentos e a classificação de documentos (RISH et al., 2001). A sua simplicidade e eficácia fazem dele uma opção popular, particularmente ao lidar com grandes quantidades de dados (RISH et al., 2001).

O Teorema de Bayes é dado pela seguinte equação:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{2}$$

Na equação temos dois eventos A e B, em que P(A) é a probabilidade de A acontecer, e P(B) é a probabilidade de B acontecer. O evento P(A|B) representa a probabilidade de A acontecer sendo que B já ocorreu e P(B|A) é a probabilidade de B acontecer se A já ocorreu. O teorema ilustra matematicamente a necessidade de incorporar essas informações não presentes na amostra, conhecidas como informações a priori, no modelo de previsão. De acordo com o teorema, temos:

- $P(A|B)$: é a Probabilidade a posteriori;
- $P(A)$: é a Probabilidade a priori;
- $P(B|A)$: é a Verossimilhança;
- $P(B)$: é a Marginalização ou Evidência.

2.1.4 k-vizinhos mais próximos

O KNN é um dos algoritmos de AS mais simples e intuitivos. É comumente empregado em classificações e regressões, fundamentado na noção de que objetos parecidos tendem a estar próximos em um espaço de características (ARAúJO, 2018).

O KNN na etapa de treinamento apenas armazena os dados, sendo assim o modelo não constrói uma função matemática (HART et al., 2000). Na etapa de testes é calculado a distância entre o novo dado e os dados armazenados e os ordena e maneira ascendente. Já na etapa de classificação é selecionado os k vizinhos mais próximos e para classificar basta escolher a classe mais frequente.

A seleção do valor de k para um número ímpar previne possíveis empates em problemas binários. Quando há igualdade de classes (k é par), existem duas táticas possíveis: a classe da instância anterior, ou a seleção de um rótulo de classe aleatório. (ARAúJO, 2018).

A Figura 3 exemplifica o processo de classificação do KNN em que os valores de vizinhos mais próximos foram respectivamente três e seis. Quando a quantidade de vizinhos selecionada foi igual a três, a instância foi classificada como pertencente à classe "bolinha". Por outro lado, ao considerar seis vizinhos, a instância passou a ser classificada como "quadrado", conforme a classe majoritária entre os objetos mais próximos. Esse exemplo mostra como é importante a escolha da quantidade de vizinhos (k) de maneira adequada, pois ela influencia diretamente no resultado.

2.1.5 Floresta aleatória

O RF é um dos algoritmos mais utilizados de AS por sua exatidão, confiabilidade e simplicidade de utilização. Ele se fundamenta em um método de aprendizado em grupo (ensemble learning), que combina diversas árvores de decisão para aprimorar a precisão e diminuir a possibilidade de overfitting (BREIMAN, 2001a).

O modelo pode ser explicado nas seguintes etapas:

- São gerados múltiplos subconjuntos do conjunto de treinamento por meio de amostragem com reposição (bootstrap).
- Para cada subconjunto amostrado, é construída uma árvore de decisão de forma independente.

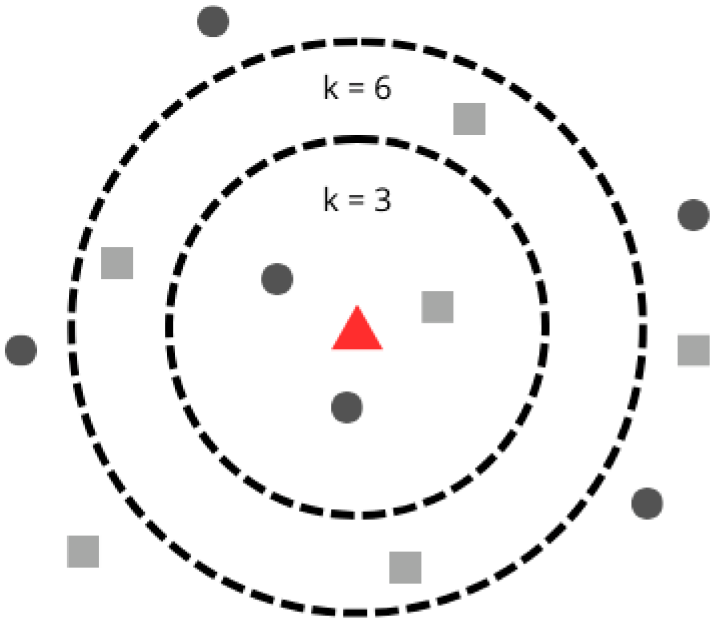


Figura 3 – Exemplo da escolha de diferentes valores k no classificador KNN.

- Durante a construção de cada árvore, a seleção das variáveis em cada nó é realizada a partir de um subconjunto aleatório de atributos, e não do conjunto completo de características disponíveis.
- A predição final do modelo é determinada por meio de votação majoritária (no caso de classificação) entre as árvores que compõem a floresta.

O RF apresenta: alta precisão que reduz *overfitting* em comparação com uma única árvore de decisão (BREIMAN, 2001a; LOUPPE, 2014), funciona com dados numéricos e categóricos (BREIMAN, 2001a), lida bem com valores faltantes e ruído nos dados (BREIMAN, 2001a; LOUPPE, 2014), e permite análise de importância das variáveis (mostra quais atributos mais influenciam na predição) (BREIMAN, 2001a).

2.1.6 Árvore de decisão - CART

O CART é um algoritmo de árvore de decisão amplamente utilizado tanto em tarefas de classificação quanto de regressão, tendo sido originalmente proposto em (BREIMAN et al., 1984). Seu funcionamento baseia-se na divisão recursiva do conjunto de dados em subconjuntos progressivamente mais homogêneos, por meio de regras simples do tipo seentão. Em cada etapa do processo, o algoritmo seleciona a variável e o ponto de corte que maximizam a pureza dos dados, de acordo com critérios específicos, como o índice de Gini, resultando em uma estrutura hierárquica em forma de árvore.

De forma geral, o CART inicia o processo de aprendizagem com todos os dados concentrados em um único nó raiz e, a partir desse ponto, realiza divisões binárias sucessivas.

Esse procedimento é repetido até que um critério de parada seja atingido, como a profundidade máxima da árvore ou um número mínimo de amostras por nó. O modelo final é altamente interpretável, permitindo acompanhar explicitamente o caminho de decisões que conduz a uma determinada predição (JAMES et al., 2013; QUINLAN, 1986).

O CART é amplamente empregado em problemas de classificação supervisionada, incluindo diagnóstico médico, análise de crédito, detecção de falhas e reconhecimento de padrões, bem como em tarefas de regressão quando o objetivo é a predição de valores contínuos (ROKACH; MAIMON, 2015). Uma de suas principais vantagens reside na simplicidade e na facilidade de interpretação, características que tornam o algoritmo particularmente atrativo em contextos nos quais a explicação das decisões do modelo é tão relevante quanto o seu desempenho preditivo.

Entretanto, o CART apresenta limitações importantes. Árvores de decisão tendem a sofrer com sobreajuste quando crescem excessivamente, tornando-se sensíveis a ruídos e variações no conjunto de treinamento. Além disso, pequenas mudanças nos dados podem resultar em estruturas de árvores significativamente distintas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Em razão dessas limitações, o CART é frequentemente empregado como modelo base em métodos ensemble, como Random Forests (BREIMAN, 2001b) e Gradient Boosting (FRIEDMAN, 2001), os quais combinam múltiplas árvores para melhorar a capacidade de generalização.

De modo geral, o CART é mais indicado para problemas que demandam modelos simples, eficientes e interpretáveis, especialmente em cenários nos quais há relações não lineares entre as variáveis. Contudo, sua aplicação em conjuntos de dados complexos ou altamente ruidosos requer cuidados adicionais, como estratégias de poda ou o uso de métodos ensemble, a fim de evitar o sobreajuste e garantir um desempenho mais robusto.

2.1.7 Perceptron Multicamadas

O MLP é um modelo de aprendizado de máquina pertencente à classe das redes neurais artificiais, amplamente utilizado em tarefas de classificação e regressão. Ele é composto por múltiplas camadas de neurônios artificiais organizadas de forma sequencial, geralmente incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (HORNIK; STINCHCOMBE; WHITE, 1989). Cada neurônio realiza uma combinação linear das entradas, seguida da aplicação de uma função de ativação não linear, permitindo ao modelo aprender relações complexas entre os dados (PARK; LEK, 2016).

O funcionamento do MLP baseia-se em um processo de treinamento supervisionado, no qual os parâmetros da rede (pesos e vieses) são ajustados iterativamente com o objetivo de minimizar um erro entre as saídas previstas e os valores reais (PARK; LEK, 2016). Esse ajuste é realizado por meio do algoritmo de retropropagação do erro (backpropagation), que propaga o erro da saída em direção às camadas anteriores, permitindo a correção gradual dos parâmetros (HORNIK; STINCHCOMBE; WHITE, 1989). Gra-

ças à presença de funções de ativação não lineares, o MLP é capaz de modelar relações altamente complexas e não lineares presentes nos dados.

O MLP é amplamente aplicado em diversos domínios, como reconhecimento de padrões, processamento de sinais, análise de imagens, diagnóstico médico, previsão de séries temporais e classificação de dados espectrais (BISHOP; NASRABADI, 2006; HAYKIN, 2009; ORRÙ et al., 2020). Sua flexibilidade o torna especialmente indicado para problemas em que as relações entre as variáveis não são triviais ou facilmente representadas por modelos lineares. Além disso, o MLP é frequentemente utilizado como modelo base em sistemas mais complexos de aprendizado profundo.

Apesar de suas vantagens, o MLP apresenta algumas limitações. O modelo pode sofrer com sobreajuste quando treinado com conjuntos de dados pequenos ou ruidosos, além de demandar um ajuste cuidadoso de hiperparâmetros, como o número de camadas, neurônios e taxa de aprendizado (LAWRENCE; GILES, 2000). Outro aspecto relevante é a sua baixa interpretabilidade quando comparado a modelos mais simples, como árvores de decisão. Ainda assim, quando bem configurado e treinado, o MLP apresenta excelente desempenho em uma ampla variedade de problemas de aprendizado de máquina.

2.1.8 Redes Neurais Convolucionais

A CNN é uma classe de modelos de aprendizado profundo especialmente desenvolvida para lidar com dados que possuem estrutura espacial ou local, como imagens, sinais e séries espectrais. A principal ideia da CNN é aprender automaticamente padrões relevantes diretamente dos dados brutos, reduzindo a necessidade de extração manual de características (LECUN et al., 1998).

De forma geral, uma CNN é composta por camadas que aplicam operações de convolução, responsáveis por extrair padrões locais (por exemplo, bordas, picos ou variações espectrais), seguidas de etapas de redução de dimensionalidade e camadas finais de decisão (LECUN et al., 1998). Ao empilhar várias camadas, a rede consegue aprender representações cada vez mais abstratas dos dados, indo de padrões simples para estruturas mais complexas (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Esse mecanismo torna a CNN muito eficiente para capturar relações locais e invariâncias, como deslocamentos ou pequenas deformações.

A CNN é amplamente utilizada em visão computacional, incluindo reconhecimento de imagens, classificação de objetos, detecção de rostos e análise médica por imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; GOODFELLOW; BENGIO; COURVILLE, 2016). Além disso, elas também têm sido aplicadas com sucesso em análise de sinais unidimensionais, como espectros (por exemplo, FTIR e ATR-FTIR), séries temporais e sinais biomédicos, explorando dependências locais ao longo do eixo espectral ou temporal (LENG et al., 2023; JUNIOR et al., 2025).

Entre as principais vantagens da CNN estão a alta capacidade de aprendizado automático de características, o bom desempenho em dados de alta dimensionalidade e a eficiência computacional quando comparadas a redes totalmente conectadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Por outro lado, elas exigem grandes volumes de dados para treinamento eficaz, podem apresentar baixa interpretabilidade e demandam maior custo computacional, especialmente em arquiteturas profundas (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.1.9 Transformer

O Transformer é uma arquitetura de aprendizado profundo introduzida em (VASWANI et al., 2017), originalmente proposta para tarefas de modelagem de sequências, como tradução automática. Diferentemente de modelos recorrentes, o Transformer processa toda a sequência de entrada de forma paralela, utilizando mecanismos de atenção para modelar dependências entre os elementos dos dados. Essa característica permite capturar relações tanto locais quanto globais, independentemente da distância entre os elementos na sequência.

O funcionamento geral do Transformer baseia-se no mecanismo de atenção, que permite ao modelo atribuir diferentes pesos às partes da entrada conforme sua relevância para a tarefa. Por meio da chamada self-attention, cada elemento da sequência pode interagir diretamente com todos os outros, aprendendo representações contextuais ricas. Para preservar a noção de ordem dos dados, informações de posição são incorporadas às entradas, possibilitando o tratamento adequado de sequências ordenadas (VASWANI et al., 2017).

Os Transformers tornaram-se amplamente utilizados em Processamento de Linguagem Natural (PLN), sendo a base de modelos consagrados como BERT (DEVLIN et al., 2019) e GPT (RADFORD et al., 2018). Além disso, essa arquitetura vem sendo aplicada com sucesso em outras áreas, como visão computacional, análise de séries temporais e processamento de sinais, incluindo dados espectrais, nos quais o modelo explora dependências ao longo do eixo temporal ou espectral (DOSOVITSKIY et al., 2021).

Entre as principais vantagens do Transformer destacam-se a capacidade de modelar dependências de longo alcance, o paralelismo no treinamento e a flexibilidade para diferentes tipos de dados sequenciais (VASWANI et al., 2017). Entretanto, o mecanismo de autoatenção apresenta complexidade computacional e de memória proporcional ao quadrado do comprimento da sequência, uma vez que calcula interações par-a-par entre todos os elementos. Esse comportamento pode limitar sua aplicação em sinais muito extensos ou em cenários com restrições de memória e processamento, além de aumentar significativamente o custo computacional (TAY et al., 2020).

2.1.10 Validação e desempenho preditivo

Nesta seção serão apresentadas as medidas avaliativas escolhidas para a análise de resultados neste projeto.

2.1.10.1 Acurácia

A Acurácia (AA) é a proporção de todos os acertos, ou seja, de verdadeiros positivos e verdadeiros negativos. Quanto maior a AA de um teste, mais preciso em determinar quem tem ou não a doença.

$$AA = \frac{VP + VN}{VP + FN + FP + VN} \tag{3}$$

2.1.10.2 Sensibilidade

A Sensibilidade (SE) representa a proporção de indivíduos que têm a doença e apresentam teste positivo, ou seja, a capacidade do teste em identificar os verdadeiros positivos. A principal limitação está na possibilidade de falsos-positivos.

$$SE = \frac{VP}{VP + FN} \tag{4}$$

2.1.10.3 Especificidade

A Especificidade (SP) é a proporção de indivíduos que não têm a doença e apresentam teste negativo, ou seja, a capacidade de identificar os verdadeiros negativos. A principal limitação é na possibilidade de falso-negativo quanto mais específico um teste.

$$SP = \frac{VN}{FP + VN} \tag{5}$$

2.1.10.4 F1-Score

O F1-Score (F1) representa a média harmônica entre a SE e SP.

$$F1 = 2 \cdot \frac{SE \cdot SP}{SE + SP} \tag{6}$$

2.2 Redes Complexas para classificação de dados

As redes complexas são estruturas matemáticas empregadas para representar sistemas que são formados por múltiplas interações entre componentes individuais. Esses sistemas são encontrados em várias áreas do saber, incluindo Biologia, Sociologia, Ciência da Computação e Física. Diferentemente dos grafos tradicionais, nos quais as conexões seguem

padrões regulares ou distribuições aleatórias simples, as redes complexas apresentam propriedades estruturais emergentes resultantes da organização não trivial das ligações entre os nós. Entre essas propriedades destacam-se o comportamento small-world, caracterizado por curtos caminhos médios entre os nós, e a estrutura scale-free, marcada pela presença de poucos nós altamente conectados (hubs) e muitos nós com baixo grau de conexão (BARABÁSI; ALBERT, 1999).

Uma rede consiste em um conjunto de vértices (ou nós) e um conjunto de arestas (ou arcos) que conectam pares de vértices. As arestas definem algum vínculo entre dois vértices, podendo ter ou não pesos. Essas arestas com pesos podem estabelecer a capacidade ou intensidade com que o tráfego se move pela rede (CARNEIRO, 2016). Adicionalmente, o grafo pode ser orientado ou não. No grafo direcionado, cada aresta possui uma direção específica que liga um vértice de origem a um vértice de destino (VERA, 2011).

O estudo de redes complexas tem se tornado um tema central em diversas áreas do conhecimento devido à sua capacidade de modelar sistemas reais com um alto grau de interconectividade e dinamicidade. Com o crescimento exponencial da disponibilidade de dados e do avanço das tecnologias computacionais, compreender a estrutura e a dinâmica das redes se tornou essencial para resolver problemas na ciência da computação (PÓSFAI; BARABÁSI, 2016; NEWMAN, 2018)

A literatura apresenta várias medidas de rede, que espelham características estruturais distintas da rede, empregadas em várias aplicações (NEWMAN, 2003; RUBINOV; SPORNS, 2010; COSTA et al., 2007). Seis medidas foram escolhidas para este estudo, considerando o desafio de detecção do TEA, explicadas na seção 2.2.1.

2.2.1 Medidas de redes

Essa seção aborda as medidas de rede escolhidas para o projeto, que são utilizadas para extrair conhecimento sobre a formação e topologia da rede, trazendo então a definição e exemplos de redes para cada medida.

2.2.1.1 Assortatividade

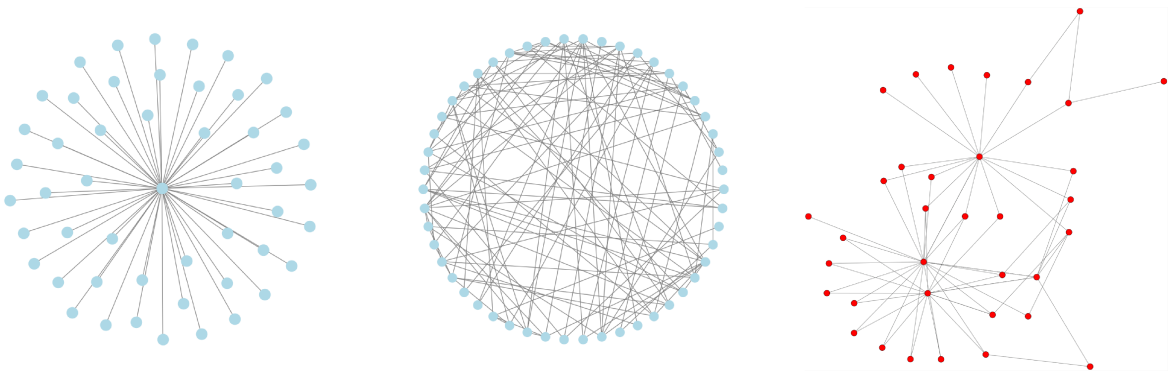
A Assortatividade, Assortativity (ASS) é uma métrica que evidencia a predileção por conexões entre os vértices de uma rede, baseada em algum atributo específico. Em redes complexas, a ASS de grau refere-se à propensão de vértices com o valor de graus próximos se ligarem (CARNEIRO, 2016; NEWMAN, 2003; NOLDUS; MIEGHEM, 2015). O coeficiente de ASS r é a medida da correlação de Pearson de grau entre pares de nós interligados. Os valores tendem a ser em torno de $[-1, 1]$, onde valores positivos sinalizam uma correlação entre nós de grau similar e valores negativos indicam relações entre nós de grau distinto (CARNEIRO, 2016). Dizemos que redes com valores negativos ($r \to -1$) apresentam um padrão de mistura disassortativo, enquanto valores positivos ($r \to 1$)

apresentam um padrão de mistura assortativo e por fim, valores que estão próximos de 0 as conexões ocorrem de maneira aleatória, sem preferência de grau.

A ASS é expressa na Equação 7, onde L representa a quantidade de conexões na rede e i_u, k_u representam os graus dos vértices i e k, que juntos formam uma aresta u.

$$r = \frac{L^{-1}\sum_u i_u k_u - \left[L^{-1}\sum_u \frac{1}{2}(i_u + k_u)\right]^2}{L^{-1}\sum_u \frac{1}{2}(i_u^2 + k_u^2) - \left[L^{-1}\sum_u \frac{1}{2}(i_u + k_u)\right]^2} \tag{7}$$

A Figura 4 mostra três redes com caracteristicas diferentes em termo de ASS. A primeira rede, Figura 4(a), apresenta uma rede disassortativa, com todos os vértices se conectando a um mesmo vértice central, apresentando pouca variabilidade de graus. Já a segunda rede, Figura 4(b), apresenta conexões aleatórias, demonstrando uma ASS de zero. Por fim, a terceira e última rede, Figura 4(c), apresenta um padrão mais próximo do assortativo, com valor de $r = 0.38$.



(a) $r = -1$ (b) $r = 0$ (c) $r = 0{,}38$

Figura 4 – Valores de ASS para diferentes tipos de redes: disassortativa (a), neutra (b) e assortativa (c).

2.2.1.2 Coeficiente de agrupamento

O Coeficiente de Agrupamento, Cluster Coefficient (CCF) é um indicador utilizado em redes complexas para avaliar o nível de interconexão entre os vizinhos de um nó (CARNEIRO, 2016). Ele avalia a chance de dois vizinhos de um nó estarem interligados, demonstrando a "aglomeração" da rede localmente.

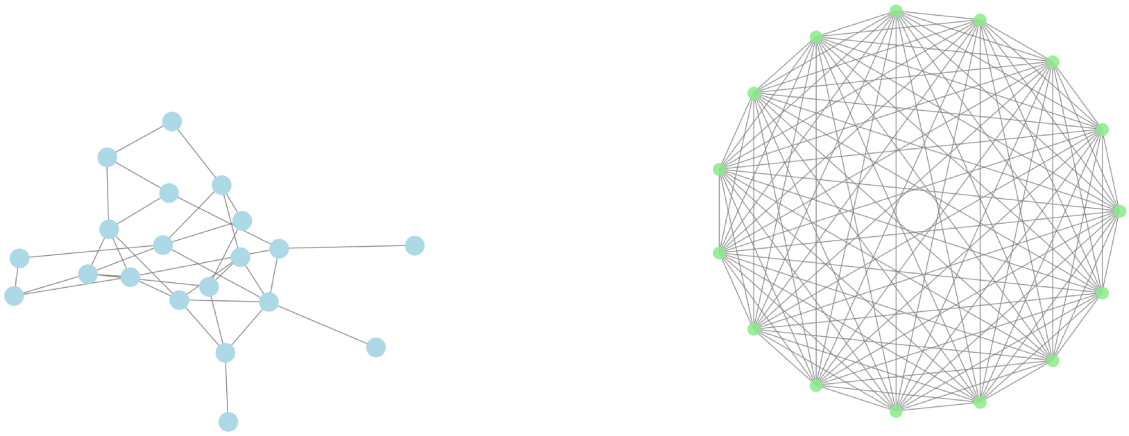
Considerando que o vértice i possui k_i vizinhos, o número máximo de arestas que podem existir entre esses vizinhos, em um grafo não direcionado, é dado por $\frac{k_i(k_i-1)}{2}$. Sempre que dois vizinhos u e s do vértice i estiverem conectados, forma-se uma aresta entre eles (CARNEIRO, 2016). Assim, o coeficiente de agrupamento local (CCF) do vértice i é definido pela Equação 8.

$$CC_i = \frac{2|e_{us}|}{k_i(k_i - 1)} \tag{8}$$

O CCF médio da rede é dado pela Equação 9. Onde $CC_i \in [0, 1]$.

$$CCF = \frac{1}{|V|} \sum_{i \in V} CC_i \tag{9}$$

A Figura 5 mostra duas redes para exemplificar o CCF. A primeira rede, Figura 5(a), consiste em uma rede aleatória de Erds-Rényi com a probabilidade de conexões baixa ($p \to 0$), gerando assim, um grafo com o valor do CCF bem baixo, próximo a zero. A segunda rede, Figura 5(b), mostra uma rede com todos os vértices conectados entre si, gerando um CCF de 1.



(a) $CC = 0{,}15$ (b) $CC = 1$

Figura 5 – Valores de CCF para diferentes tipos de redes.

2.2.1.3 Grau médio

O Grau Médio, Average Degree (ADG) é uma medida que representa o número médio de conexões estabelecidas entre os vértices da rede. Primeiramente, o grau de um vértice da rede é dado pela Equação 10.

$$k_i = \sum_{j=0}^{n} a_{ij} \tag{10}$$

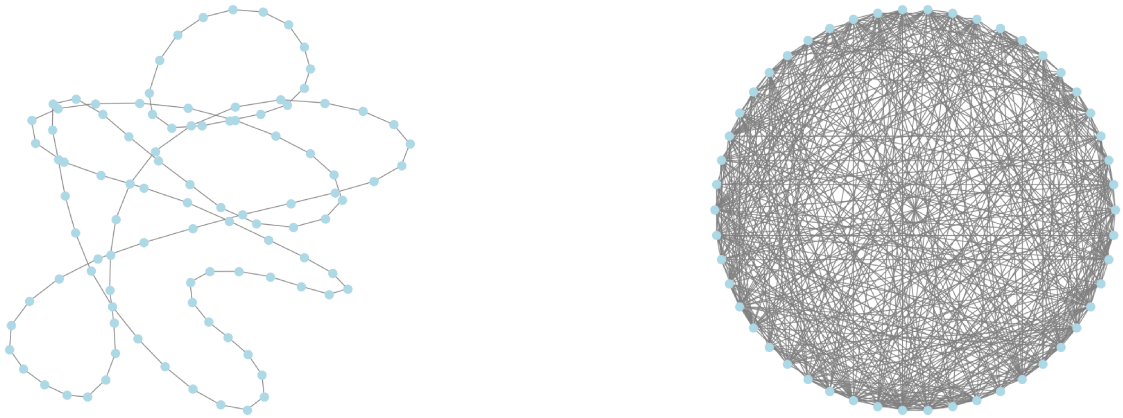
o grau do vértice i, denotado por k_i, corresponde ao número de conexões associadas a esse vértice, sendo obtido a partir da soma dos elementos da linha i da matriz de adjacência da rede, conforme apresentado na Equação 10. Nessa equação, a_{ij} representa o elemento da matriz de adjacência A, assumindo valor igual a 1 caso exista uma aresta entre os vértices i e j, e 0 caso contrário, enquanto n corresponde ao número total de vértices da rede.

O ADG é dado pela média dos graus da rede, dado pela Equação 11.

$$ADG = \frac{1}{n}\sum_{i=0}^{n} k_i \tag{11}$$

a quantidade de total de vértices da rede é representada pela variável n.

A Figura 6 mostra exemplos de duas redes em relação a medida ADG. A primeira rede, Figura 6(a), representa um grafo regular em que o grau de todos os vértices são iguais. Já a segunda rede, Figura 6(b), apresenta uma rede aleatória com os graus dos vértices distintos entre si. Perceba que o medida ADG demonstra a densidade da rede, pois quando o valor é baixo representa uma rede esparsa (poucas conexões por nó), porém quando o valor é alto representa uma rede densa (muitos nós interligados).



(a) $k = 2$ (b) $k = 19{,}60$

Figura 6 – Valores de ADG para diferentes tipos de redes (regular e aleatória).

2.2.1.4 Menor caminho médio

O Menor Caminho Médio, Shortest Average Path (ASP) de uma rede complexa mede a distância média entre todos os pares de nós na rede também chamado de distância geodésica (CARNEIRO, 2016). Esta ação simboliza o método mais eficaz para percorrer a rede, reduzindo o tempo e o custo para a troca de informações (BORGWARDT; KRIEGEL, 2005; VERA, 2011).

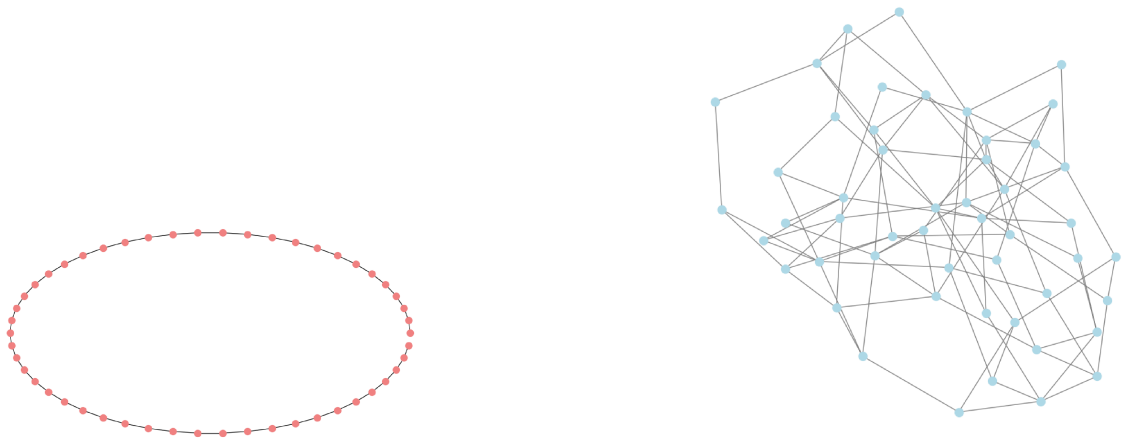
A métrica do ASP é crucial para compreender a eficácia da transmissão de dados em redes complexas. Redes como a Internet, redes sociais e redes neurais costumam ter valores reduzidos de ASP, o que as torna extremamente eficazes na transmissão e disseminação de informações (VERA, 2011).

Seja $D_{i,j}$ a menor distância entre i até j, a medida ASP é dada pela Equação 12.

$$ASP = \frac{1}{n(n-1)}\sum_{i \neq j} D_{i,j} \tag{12}$$

a quantidade de total de vértices da rede é representada pela variável n.

A Figura 7 mostra dois exemplos de redes com valores de ASP. A Figura 7(a), mostra um grafo de linha que apresenta pouca interconectividade, logo o valor de ASP é elevado. Porém a segunda rede, Figura 7(b) apresenta um grafo de Small-World (mundo pequeno), que tem por caracteristica valores de ASP reduzidos, apresentando boa fluidez na rede.



(a) $MC = 17$ (b) $MC = 4{,}35$

Figura 7 – Valores de ASP para diferentes tipos de redes (Grafo de linha e Small world).

2.2.1.5 Proximidade

A medida de Proximidade, Closeness (CLO) indica a proximidade de um vértice em relação aos demais nós da rede, ou seja, essa medida representa o inverso da distância geodésica média de um ponto em relação aos demais (CARNEIRO, 2016; FREEMAN et al., 2002). Ela demonstra a velocidade com que a informação pode se propagar de um nó para todos os outros.

A equação da CLO é dada pela Equação 13.

$$CLO = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{n-1}{\sum_{j=1}^{n} d(i,j)} \right), \tag{13}$$

onde $P \in [0,1]$ e $d(i,j)$ é a menor distância entre i e j, e o valor de n representa a quantidade de vértices da rede.

A Figura 8 mostra uma rede aleatória com os valores da medida CLO nos vértices. É possível identificar que os vértices mais periféricos possuem valores de CLO baixos, já os pontos centrais possuem valores elevados.

2.2.1.6 Intermedialidade

A métrica de rede Intermedialidade, Betweenness (BET) ou Betweenness é crucial na análise de redes, pois mede a relevância de um nó (ou aresta) com base na sua posição

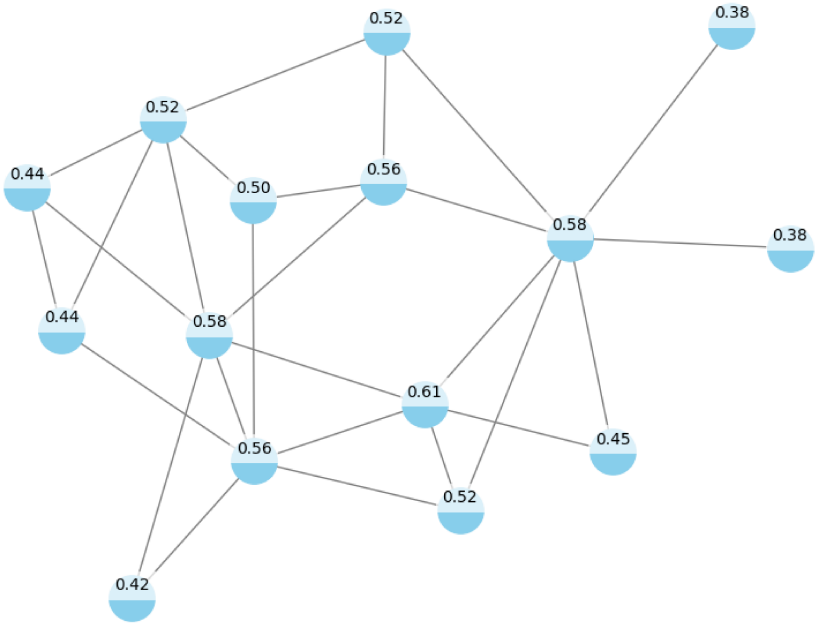


Figura 8 – Exemplo de uma rede com os valores de CLO nos vértices.

estratégica no trajeto mais curto entre outros nós na rede. Ela serve para reconhecer nós (ou vértices) que funcionam como "pontes" ou "mediadores" de informações entre outros pares de nós, sendo essencial em pesquisas sobre a dinâmica de redes, como o fluxo de informações ou recursos (FREEMAN et al., 2002; CARNEIRO, 2016).

Dado um vértice i, considera-se o número de caminhos geodésicos que passam por esse vértice. Dessa forma, avalia-se a participação de i nos menores caminhos entre pares de vértices u e v, em que $u, v \in V - \{i\}$ (CARNEIRO, 2016). Assim, a centralidade de intermediação (BET) é definida conforme a Equação 14.

$$b_i = \sum_{u,v \in V-i} n_{uv}^i \tag{14}$$

A Figura 9 apresenta os valores de BET para cada vértice em uma rede que um vértice e conectados com todos os demais, grando vários vértices periféricos em que o vértice central da rede possui maior importância.

2.2.2 Métodos de construção de rede

Um dos principais desafios da classificação baseada em redes complexas é justamente a construção de uma rede representativa, capaz de mapear os dados de maneira adequada para o melhor desempenho na classificação. Nesse contexto existem duas grandes categorias de heurísticas de rede: rede KNN e rede de raio ϵ. A primeira cria uma conexão com seus k vizinhos mais próximos, já a segunda cria uma conexão com os vértices presentes

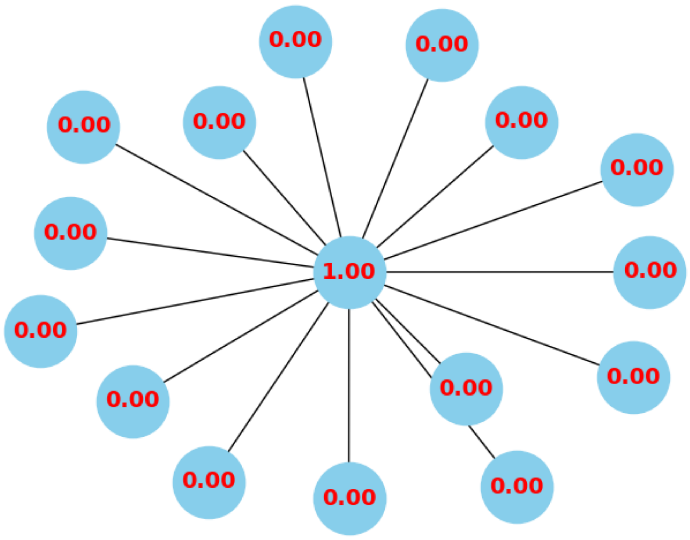


Figura 9 – Exemplo de uma rede com os valores de BET nos vértices.

em um raio ϵ. Nas próximas sessões, serão discutidas as métodos de construção de rede tradicionais a fim de comparar com a técnica de construção proposta neste projeto.

2.2.2.1 Rede de k vizinhos mais próximos

A técnica de construção de rede baseada no algoritmo KNN é mais utilizada na literatura (JR et al., 2011; CARNEIRO; ZHAO, 2018; CARNEIRO; GAMA; RIBEIRO, 2021), e consiste em criar uma conexão do item de dados com seus k vizinhos mais próximos, desde que eles pertençam a mesma classe. Formalmente a matriz de adjacência da rede A, pode ser obtida pela Equação 15.

$$A_{ij} = \begin{cases} 1, & \text{se } x_j \in k\text{-NN}(x_i) \text{ e } y_i = y_j, \\ 0, & \text{senão.} \end{cases} \tag{15}$$

Nessa equação, A_{ij} assume valor igual a 1 quando o padrão x_j pertence ao conjunto dos k vizinhos mais próximos do padrão x_i, isto é, $x_j \in k$-NN(x_i), e ambos os padrões possuem o mesmo rótulo de classe ($y_i = y_j$). Caso contrário, A_{ij} assume valor 0, indicando a ausência de conexão entre os vértices i e j.

A partir desta técnica surgiram algumas variações que serão explicadas na Seção 2.2.2.2 até a Seção 2.2.2.4.

2.2.2.2 Rede Seletiva de k vizinhos mais próximos

Inicialmente proposta em (CARNEIRO; ZHAO, 2018), a técnica Seletiva de KNN consiste em uma variação da heurística de construção de redes baseada em KNN. Nessa abordagem, cada amostra é conectada aos seus k vizinhos mais próximos considerando apenas elementos pertencentes à mesma classe. Diferentemente da técnica tradicional, na

qual a busca pelos vizinhos mais próximos é realizada sobre todo o conjunto de dados e posteriormente pode ser aplicada uma restrição de classe, a versão seletiva restringe previamente essa busca aos padrões com rótulo igual. Dessa forma, evitam-se comparações entre amostras de classes distintas. Formalmente, a matriz de adjacência é definida pela Equação 16.

$$A_{ij} = \begin{cases} 1, & \text{se } x_j \in \text{Sel-}k\text{-NN}(x_i), \\ 0, & \text{senão.} \end{cases} \tag{16}$$

2.2.2.3 Rede Mútua de k vizinhos mais próximos

Inicialmente proposta em (BRITO et al., 1997), a técnica Mútua de KNN é outra variação do método de construção de rede KNN, que consiste em criar uma conexão em um dado de item i com os demais k vizinhos mais próximos, desde que i pertença aos vizinhos mais próximos de j e vice versa (OZAKI et al., 2011). Formalmente a matriz de adjacência é dado pela seguinte regra:

$$A_{ij} = \begin{cases} 1, & \text{se } x_j \in k\text{-NN}(x_i),\ x_i \in k\text{-NN}(x_j) \text{ e } y_i = y_j, \\ 0, & \text{senão.} \end{cases} \tag{17}$$

2.2.2.4 Rede Seletiva e Mútua de k vizinhos mais próximos

A rede seletiva e mútua de KNN (CARNEIRO; GAMA; RIBEIRO, 2021) consite em criar um ligação de um item de dado i com seus vizinhos mais próximos da mesma classe, desde que i pertença aos vizinhos mais próximos de j e vice versa. Diferentemente da heurística mútua, comparamos apenas os vizinhos de mesma classe igualmente no método seletivo. Formalmente a matriz de adjacência é dado pela seguinte regra:

$$A_{ij} = \begin{cases} 1, & \text{se } x_j \in \text{Sel-}k\text{-NN}(x_i),\ x_i \in \text{Sel-}k\text{-NN}(x_j), \\ 0, & \text{senão.} \end{cases} \tag{18}$$

2.2.2.5 Rede vizinhança de raio ϵ

Outra categoria de construção de rede é a baseada em vizinhança de raio ϵ (WAN; YI, 2004; CARNEIRO; GAMA; RIBEIRO, 2021; CARNEIRO; ZHAO, 2018), que consiste em criar uma ligação de um item de dado com os vértices presentes dentro do raio ϵ, desde que o vértice pertença a mesma classe do vértice de referência. Formalmente a matriz de adjacência é dado pela seguinte regra:

$$A_{ij} = \begin{cases} 1, & \text{se } D_{i,j} <= \epsilon \text{ e } y_i = y_j, \\ 0, & \text{senão.} \end{cases} \tag{19}$$

Onde $D_{i,j}$ é a distância entre o vértice X_i e X_j.

2.2.3 Classificação por conformidade padrão

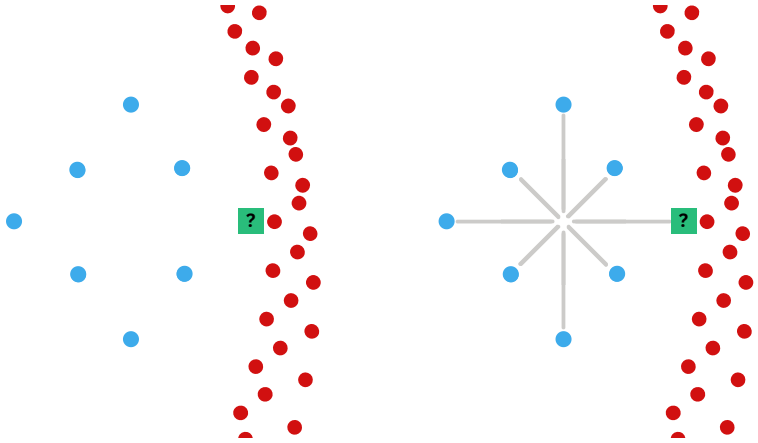
O estudo baseado em redes complexas permite extrair conhecimento de uma outra perspectiva ao analisar padrões de formação da rede. Diferentemente de classificadores tradicionais da literatura, o classificador baseado em redes complexas, também conhecidos como classficadores de HL, não classificam novos dados com relação direta a seus atributos, como similaridade e proximidade, mas sim com padrões topológicos da rede extraídos pelas medidas de rede. Cada medida de rede permite extrair informações distintas tornado a técnica de HL robusta e adequada tanto para problemas mais simples e problemas complexos. A técnica de HL abordada será a classificação baseada em conformação de padrões.

Essa técnica pode ser segmentada em duas etapas principais, o treinamento e a classificação. A fase de treinamento envolve a formação da rede com base nos dados de entrada, representados como um vetor de atributos. A formação da rede é gerada por uma das heurísticas explicadas na Seção 2.2.2. A partir da rede gerada, são calculadas as medidas de rede que servirão como padrão de conformidade para a inclusão de qualquer novo atributo na rede (CARNEIRO, 2016). Na fase de classificação, a cada elemento novo adicionado à rede, todas as medidas serão recalculadas e o efeito resultante será examinado no componente (classe) da rede. Assim, a classe com a menor variação se tornará o rótulo da instância devido à sua maior aderência aos padrões previamente estabelecidos.

A Figura 10 mostra a vantagem da classificação HL. A primeira Figura 10(a) mostra uma instância marcada com um ponto de interrogação em um quadrado que deve ser classificada ou na classe azul ou na classe vermelha. Classificadores tradicionais tendem a classificar na classe com maior proximidade, ignorando padrões dos dados. Já o classificador de HL, Figura 10(b), consegue classificar na classe correta (azul), justamente por identificar o padrão com a classe azul em questão.

Em (CARNEIRO; GAMA; RIBEIRO, 2021), é conduzido um estudo minucioso sobre a capacidade preditiva de oito medidas de rede escolhidas na literatura, cujos resultados são confirmados por meio de conformidade padrão. Os achados sugerem que indicadores como o caminho médio mais curto e a ASS, além de oferecerem resultados preditivos satisfatórios, também são mais resistentes a alterações estruturais na rede.

Em (SILVA; ZHAO, 2012) sugere um método híbrido de classificação de dados que mescla aprendizado de nível baixo (low-level) e alto (high-level). Ao contrário dos métodos convencionais que se concentram apenas em características físicas dos dados de entrada, o método proposto também inclui a avaliação de padrões estruturais e topológicos dos dados apresentados na forma de rede. Esta metodologia possibilita ao classificador avaliar a aderência das amostras de teste à formação de padrões dos dados, aprimorando a performance em cenários complexos de classificação. Os pesquisadores evidenciam que,



(a) Base de dados “Toy” (b) Classificação de HL

Figura 10 – Objeto verde para classificação nas categorias vermelha ou azul. (a) Classificadores convencionais enfrentam desafios devido à relação próxima do objeto de teste com a classe vermelha. (b) A classificação de HL é apta a categorizar o objeto de teste na categoria azul, ao analisar as características estruturais e topológicas dos dados recebidos na rede.

Fonte: (LIMA FILHO et al., 2024)

à medida que a complexidade na configuração das classes cresce, é imprescindível um aumento na contribuição do termo de HL para uma correta classificação. Ademais, a pesquisa demonstra como a metodologia sugerida pode ser empregada em questões práticas, como a detecção de variações e distorções em imagens de números manuscritos, levando a um aumento na taxa global de identificação de padrões.

No trabalho (CARNEIRO et al., 2014), temos uma abordagem conhecida como classificação de HL em K-associados ótimos, que combina com a classificação por conformidade padrão. A metodologia sugerida, além de mesclar termos de baixo e alto nível, categoriza os dados não somente com base em suas características físicas (atributos), mas também verifica a conformidade com o padrão da rede. Os achados da pesquisa sugerem que o método proposto supera os classificadores convencionais em várias situações. A técnica se mostrou eficiente ao levar em conta informações locais e globais dos dados de treinamento, possibilitando uma avaliação mais completa das estruturas e relações existentes nos dados.

2.2.4 Classificação por caracterização de importância

Outro classificador baseado em redes complexas que temos na literatura é o baseado na caraterização de importância. Esta metodologia foi sugerida em (CUPERTINO; ZHAO; CARNEIRO, 2015) e emprega o método de passeio aleatório para categorizar os dados com base na facilidade de acesso. Tecnicamente, quanto mais fácil for o acesso a um vértice, mais conexões ele tem e, consequentemente, sua importância aumenta (CARNEIRO,

2016). Formalmente, a técnica é dividida em duas etapas: treinamento e classificação.

No treinamento os dados rotulado em forma de vetor de atributos são utilizados para formar a rede por algum método de construção de rede como o KNNG, explicado na Seção 2.2.2. A dados relevantes para o porcesso de classificação é dado pelas medidas de eficiência diferencial espaço-estrutural e PageRank, a partir da rede G construída (FERNANDES et al., 2023). Quando a rede for construída e cada instância de teste é conectada na rede, os atributos físicos dos dados se tornam irrelevantes para o processo de classificação (CARNEIRO, 2016).

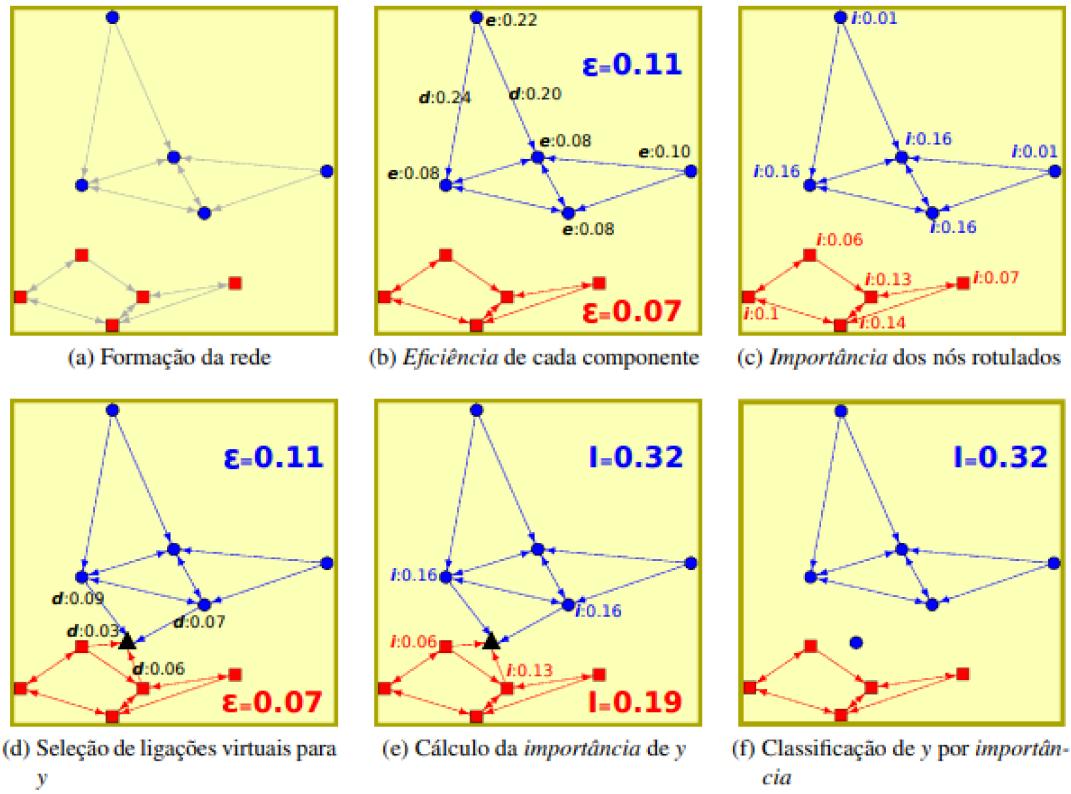
Já na fase de classificação, um vértice de teste não rotulado é temporariamente inserido em cada rede (cada rede representa uma classe da base de dados) pela medida de eficiência diferencial espaço-estrutural, e então seu valor de importância (dado pela medida de PageRank) é calculado. O item é classificado na rede em que ele recebe maior valor de importância.

A Figura 11 exemplifica o passo a passo do processo de classificação baseado em importância. O primeiro passo é a construção da rede, em seguida são cálculadas os valores de eficiência e importância de cada componente, para todas as redes construídas. Já no processo de classficação um dado y é temporariamente conectado em cada rede através da medida de eficiência. O valor de importância atribuído ao vértice y é calculado, e para finalizar, o vértice y recebe a classe com o maior valor de importância atribuído, no caso foi a classe do circulo azul.

Em (CARNEIRO; ZHAO, 2018), propôs-se uma metodologia de classificação baseada na caracterização de importância, na qual uma nova ocorrência não rotulada é categorizada com base no conceito de relevância definido pelo PageRank do Google das redes de dados subjacentes. Adicionalmente à metodologia, introduziu-se uma nova métrica de rede, a eficiência diferencial espaço-estrutural, que busca harmonizar as propriedades físicas e topológicas dos dados de entrada.

No artigo (FERNANDES et al., 2023) é explorado medidas de centralidade de rede para descrever a importância de cada um dos nós. Na classificação foi utilizado o classificador baseado em caracterização de importância com a rede construída pelo grafo de k-vizinhos mais próximos (KNNG). Na etapa de classificação cada nó de amostra recebe uma pontuação de eficência e de importância baseada nas medidas de centralidade, o dado é classificado na rede que apresentar o maior valor de importância para o nó de amostra. Os resultados são promisores superando classificadores tradicionais em algumas bases de dados, mostrando que as medidas de centralidade capturam informações relevantes para a classificação. A medida de centralidade de autovetor foi a que demonstrou o melhor desempenho.

No trabalho (CARNEIRO; ZHAO, 2018) é explorado diferentes métodos de contrução de rede baseados na vizinhança de raio ϵ e em k-vizinhos mais próximos na classificação por caracterização de importância. Os resultados são promissores principalmente para as



(a) Formação da rede (b) *Eficiência* de cada componente (c) *Importância* dos nós rotulados

(d) Seleção de ligações virtuais para *y* (e) Cálculo da *importância* de *y* (f) Classificação de y por *importância*

Figura 11 – Passo a passo da classificação por caracterização de importância. (a) Processo de criação de rede para cada classe l. (b) Cálculo da eficiêcia de cada componente da rede G. (c) Cálculo da importância de cada componente da rede G. (d) O vértice de teste y é temporariamente conectado em cada rede de treino, de acordo com os valores de eficiência. (e) Cálculo da importância atribuída pelo vértice y em cada classe. (f) A classificação é dada pela rede em que o vértice y recebeu maior importâmcia.

Fonte: (CARNEIRO; ZHAO, 2018)

heurísticas de construção seletivas, que predominaram nos melhores resultados obtidos no reconhecimento de imagens.

2.3 ATR-FTIR: Espectroscopia de Infravermelho com Transformada de Fourier e Reflectância Total Atenuada

Nesta seção será abordado sobre a ATR-FTIR, desde uma breve introdução, principios, funcionamento até a sua aplicação no AM.

2.3.1 Introdução à Espectroscopia no Infravermelho (EIR)

A Espectroscopia de Infravermelho (EIR) é um método de análise que se baseia na absorção de radiação infravermelha por moléculas, gerando transições vibracionais que fornecem informações estruturais minuciosas sobre os sistemas analisados (GAIGEOT; MARTINEZ; VUILLEUMIER, 2007). Essa absorção é resultado das mudanças vibracionais dos átomos em uma molécula quando exposta à radiação eletromagnética na faixa do infravermelho (cerca de $4000cm^{-1}$ a $400cm^{-1}$ em comprimento de onda) (GAIGEOT; MARTINEZ; VUILLEUMIER, 2007).

A EIR é categorizada de diversas formas, com base na área do espectro eletromagnético em que é aplicada. Cada uma dessas áreas proporciona dados adicionais sobre as características moleculares e estruturais dos materiais examinados:

- Espectroscopia no Infravermelho Médio (MIR, $4000 - 400cm^{-1}$): A MIR é o método de EIR mais comumente empregado. As absorções neste intervalo espelham as vibrações básicas das ligações químicas, sendo extremamente valiosas para a identificação molecular. O espectro produzido possibilita a identificação precisa de compostos (BELLON-MAUREL; MCBRATNEY, 2011).

- Espectroscopia no Infravermelho Próximo (NIR, $14.000 - 4000cm^{-1}$): A NIR se fundamenta em sobretons e combinações de vibrações essenciais identificadas na faixa do infravermelho médio. Apesar de ser menos precisa que a MIR, ela é frequentemente empregada em análises rápidas e não destrutivas, ganhando popularidade em setores industriais e biomédicos (BELLON-MAUREL; MCBRATNEY, 2011).

- Espectroscopia no Infravermelho Distante (FIR, $400 - 10cm^{-1}$): A FIR é eficaz na análise de vibrações de baixa energia, como rotações moleculares e modos vibracionais em sólidos. É comumente empregada para analisar redes cristalinas e fenômenos estruturais em materiais de alta tecnologia (CASEY, 2012).

2.3.2 Princípios do Espectroscopia Infravermelha com Transformada de Fourier (FTIR)

A Espectroscopia Infravermelha com Transformada de Fourier (FTIR) é um método frequentemente empregado na análise química, possibilitando a identificação e caracterização de compostos através da interação da radiação infravermelha com a matéria (BERTHOMIEU; HIENERWADEL, 2009). Esta metodologia investiga a absorção de radiação infravermelha por moléculas, examinando as oscilações das ligações químicas e fornecendo um espectro específico para cada composto (MOVASAGHI; REHMAN; REHMAN, 2008).

Ao contrário da espectroscopia infravermelha tradicional, que avalia a absorção em apenas uma frequência de cada vez, a espectroscopia infravermelha FTIR emprega um

interferômetro de Michelson para recolher simultaneamente dados de todas as frequências do espectro infravermelho (MOHAMED et al., 2017). Este sinal é matematicamente tratado através da Transformada de Fourier, que converte as informações para o domínio da frequência, resultando em um espectro minucioso (BERTHOMIEU; HIENERWADEL, 2009).

Moléculas captam radiação IR em frequências particulares, que se alinham com as vibrações de suas ligações químicas. Esta absorção é influenciada pela alteração do momento dipolar da molécula, isto é, quando uma ligação química vibra e modifica a distribuição de cargas elétricas, ela interage com a radiação IR (MOHAMED et al., 2017). Cada vínculo químico (como C-H, O-H, C=O, N-H, etc.) possui frequências de absorção particulares, facilitando a identificação dos elementos de uma amostra.

O FT no FTIR ocorre porque o aparelho não mede a absorção diretamente em cada frequência. Ao invés disso, ele capta um sinal no tempo através de um interferômetro de Michelson (MOHAMED et al., 2017). A Transformação de Fourier (FT) transforma essas informações no domínio da frequência, resultando no espectro FTIR.

O espectro FTIR é um diagrama de transmitância ou absorbância em relação ao número de onda (cm^{-1}). Os picos do espectro sinalizam as frequências que foram assimiladas, evidenciando a existência de grupos funcionais particulares na amostra (MOVASAGHI; REHMAN; REHMAN, 2008).

Principais regiões do espectro:

- Região de grupo funcional ($4000 - 1500cm^{-1}$): Identifica ligações como O-H, N-H, C=O, C≡C, etc.

- Região de impressão digital ($1500 - 400cm^{-1}$): Contém padrões únicos da molécula, permitindo sua identificação.

A espectroscopia infravermelha (FTIR) é amplamente utilizada em vários campos, incluindo química, biomedicina, ciência de materiais e meio ambiente, graças à sua elevada SE, rapidez e habilidade de análise sem danificar (MOVASAGHI; REHMAN; REHMAN, 2008). Os seus princípios básicos envolvem a absorção seletiva da radiação IR, a transformação dos sinais brutos através da Transformada de Fourier e a criação de um espectro que desvenda a estrutura molecular da amostra (BERTHOMIEU; HIENERWADEL, 2009).

2.3.3 ATR-FITR: princípios e funcionamento

A Técnica de Reflectância Total Atenuada (ATR) é uma ferramenta complementar à Espectroscopia de Infravermelho com Transformada de Fourier (FTIR), utilizada para examinar amostras sem a exigência de uma preparação intrincada (GRDADOLNIK, 2002).

Na técnica de ATR, a amostra é comprimida contra um cristal com alto índice de refração, normalmente diamante, germânio ou ZnSe. A luz infravermelha penetra no cristal e passa por uma reflexão interna total, gerando uma onda evanescente que se estende levemente sobre a superfície da amostra. A energia que a amostra absorve nessa interação é documentada, dando origem ao espectro infravermelho (BIEBERLE-HüTTER et al., 2021).

A Figura 12 mostra o processo feito pela espectroscopia ATR-FTIR. A fonte emite um feixe de luz infravermelha (IR), que é direcionado ao cristal ATR, que apresenta um elevado índice de refração. Ao atingir um ângulo específico no cristal, a luz passa por múltiplas reflexões internas totais, resultando em uma onda evanescente que se estende levemente pela amostra em contato com o cristal. A interação da onda evanescente com os elementos da amostra leva à absorção seletiva de certas frequências de infravermelho, que varia conforme as propriedades moleculares existentes. Então, um detector capta o feixe em ascensão, que o analisa e produz um espectro infravermelho característico da amostra.

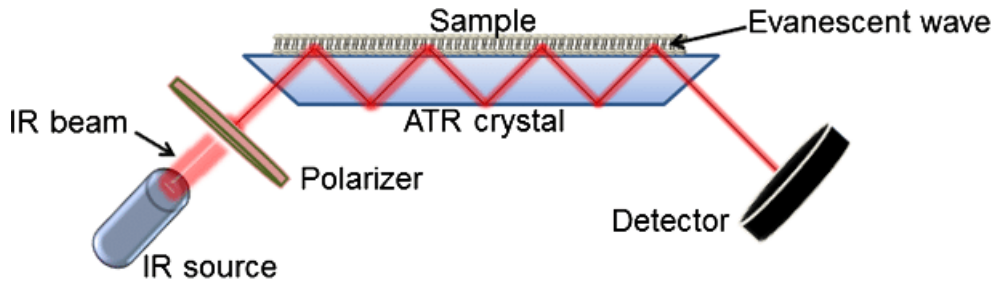


Figura 12 – Esquema de um sistema ATR-FTIR. O feixe de infravermelho (IR beam) emitido pela fonte (IR source) é direcionado através de um polarizador (Polarizer) e entra no cristal ATR (ATR crystal), onde sofre múltiplas reflexões internas. Durante esse processo, a onda evanescente (Evanescent wave) interage com a amostra (Sample) depositada na superfície do cristal. A luz resultante é então detectada pelo detector (Detector), permitindo a obtenção do espectro infravermelho da amostra.

Fonte: (BIEBERLE-HüTTER et al., 2021)

Os componentes principais do sistema ATR-FTIR mostrados na Figura 12 são:

- Produtor de Luz Infravermelha (IR): Produz um raio de luz IR que é direcionado ao cristal ATR.
- Cristal ATR: Normalmente produzido a partir de materiais de alta refração, como diamante, germânio ou seleneto de zinco (ZnSe). O feixe de infravermelho penetra no cristal e é refletido internamente.
- Amostra: Inserida diretamente na superfície do cristal ATR.

- ❏ Reflexão Total Interna: O feixe de luz IR passa por reflexões internas completas no cristal, originando uma onda evanescente que se propaga superficialmente pela amostra.
- ❏ Onda Evanescente: Interage com a amostra, sendo absorvida em partes, dependendo das propriedades moleculares da mesma.
- ❏ Detector: Após várias reflexões, o feixe IR que emerge é recolhido e examinado para produzir o espectro infravermelho da amostra.

Alguns benefícios do uso da ATR associada à FTIR incluem a ausência de preparação complicada da amostra, como moagem ou dissolução. Além disso, é adequado para diversos estados da matéria, tais como sólidos, líquidos e géis. A ATR-FTIR é perfeita para análises biológicas, já que a água tem menos influência na região do infravermelho médio, possibilitando análises rápidas e repetitivas sem prejudicar a amostra (GRDADOLNIK, 2002).

2.3.4 ATR-FTIR salivar em dados sequenciais e aprendizado de máquina

A ATR-FTIR tem se destacado como um instrumento eficaz na avaliação de amostras biológicas, como a saliva, graças à sua habilidade de identificar assinaturas moleculares específicas de diversas condições fisiológicas e patológicas (MARTINEZ-CUAZITL et al., 2021). Devido à sua facilidade de obtenção e abundância de biomarcadores, a saliva tem sido extensivamente pesquisada como uma opção não invasiva para diagnósticos médicos (LIMA FILHO et al., 2024). Contudo, para interpretar os espectros produzidos pelo ATR-FTIR, são necessárias técnicas avançadas de processamento de dados, especialmente por sua característica sequencial e complexa (LIMA FILHO et al., 2024).

Neste cenário, têm-se utilizado métodos de AM para aprimorar a análise e classificação de espectros ATR-FTIR salivares. Modelos supervisionados e não supervisionados são aptos a reconhecer padrões complexos nos dados, possibilitando a identificação de enfermidades e a distinção entre amostras saudáveis e patológicas com grande exatidão.

A Figura 13 mostra algumas regiões de interesse do espectro, sendo assim conseguimos filtrar melhor o que há em cada região. Algumas dessa bandas de infravermelho são repletas de lipídios, proteínas, carboídratos e ácidos nucleicos. Em especial temos o pico da Amida I (região de infravermelho entre as bandas 1660cm^{-1} and 1630cm^{-1}), que geralmente é utiliza para normalizar o espectro .

No artigo (MARTINEZ-CUAZITL et al., 2021) a ATR-FTIR foi utilizada para detectar assinaturas biológicas específicas em amostras de saliva de pacientes infectados pelo COVID-19. Os cientistas examinaram amostras salivares de 255 pacientes diagnosticados com COVID-19 e 1.209 pessoas saudáveis. As avaliações indicaram mudanças notáveis nas

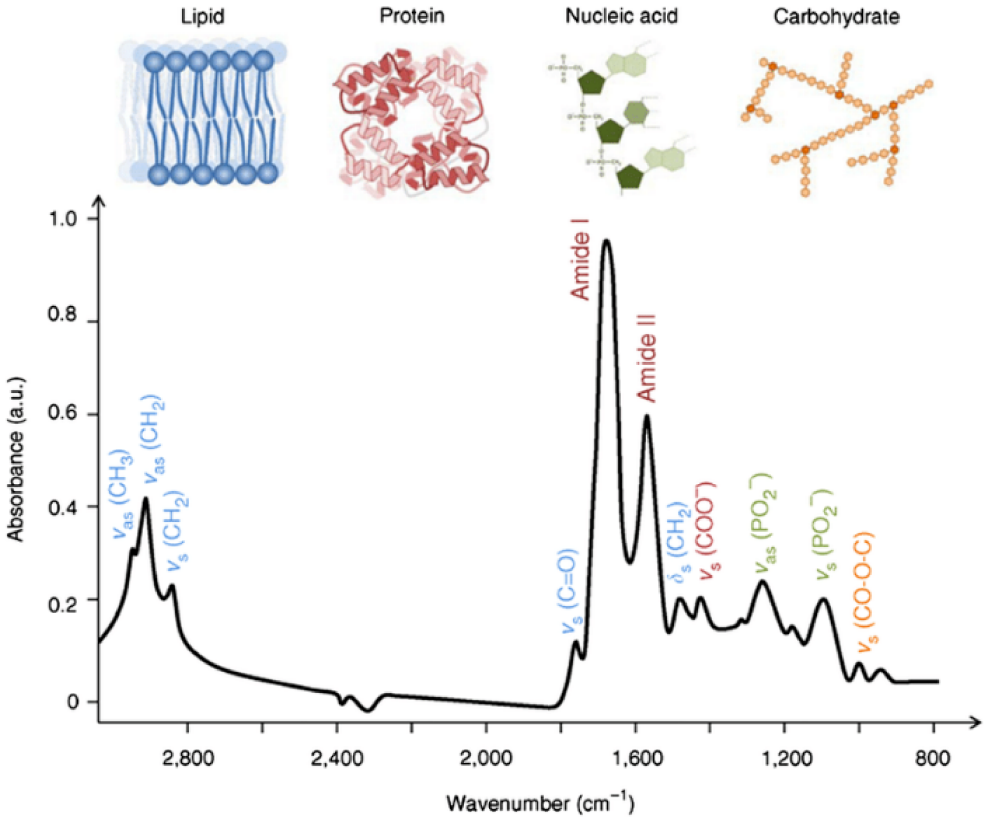


Figura 13 – Regiões de interesse de um espectro ATR-FTIR.

Fonte: (ROHMAN et al., 2020)

áreas relacionadas às amidas I, II e III, além de bandas ligadas a moléculas fosforiladas, carboidratos e ácidos nucleicos. Adicionalmente, notaram-se diferenças nas áreas relacionadas às imunoglobulinas, com maior absorção nas amostras de pacientes infectados pelo COVID-19. Por meio de um modelo multivariado de regressão linear, conseguimos distinguir claramente entre os grupos sadios e infectados. Estes resultados indicam que a ATR-FTIR pode ser um método promissor para a detecção rápida e não invasiva da COVID-19 através da avaliação da saliva.

O estudo (RODRIGUES et al., 2019) investigou ATR-FTIR para detectar marcas moleculares particulares na saliva de pacientes com doença renal crônica (DRC). Os cientistas compararam amostras de saliva de pacientes com DRC com indivíduos sadios, empregando o método ATR-FTIR para identificar variações nos elementos salivares. A avaliação indicou que quatro elementos apresentaram variações notáveis ($p<0{,}05$) entre os grupos. Em particular, os modos vibracionais do tiocianato (SCN-, 2052 cm^{-1}) e dos fosfolipídios/carboidratos (924 cm^{-1}) revelaram potencial como biomarcadores salivares para distinguir pacientes com Doença Renal Crônica de pessoas saudáveis. A fusão dos espectros originais e derivados nesses modos vibracionais resultou em uma SE de $92,8\%$ e SP de $85,7\%$ na identificação da DRC. Estes resultados indicam que a avaliação salivar através do ATR-FTIR pode ser um método promissor para o diagnóstico não invasivo da

doença renal crônica.

O trabalho (LIMA FILHO et al., 2024) investigou ATR-FTIR salivar para detectar câncer de boca, através da classificação de HL analisando propriedades e medidas de rede. Os resultados se mostram promissores quando a medida de dissimilaridade utilizada na construção da rede é o cosseno, especialmente em conjunto com o CCF como medida topológica, apresentando 71% de AA e 81% de SE, superando todos os classificadores analisados, até os de última geração como a CNN. Esses achados sugerem que a análise salivar por meio do ATR-FTIR pode ser uma técnica promissora para o diagnóstico antecipado e não invasivo do câncer bucal.

2.4 Trabalhos Relacionados

Existem diversos trabalhos relacionados à classificação de dados ATR-FTIR (BARAUNA et al., 2021; FERREIRA et al., 2020; CAIXETA et al., 2023; MARTINEZ-CUAZITL et al., 2021; ANDRADE; SABINO-SILVA; CARNEIRO, 2024), os quais apresentam avanços e resultados significativos na tarefa de classificação desses dados. No entanto, tais abordagens concentram-se predominantemente na exploração de relações de similaridade entre os espectros, deixando de considerar características topológicas inerentes aos dados. Nesse sentido, diversas pesquisas têm investigado a classificação de dados de ATR-FTIR por meio de redes complexas (LIMA FILHO et al., 2024; FERNANDES; SABINO-SILVA; CARNEIRO, 2025; ARAúJO; SABINO-SILVA; CARNEIRO, 2024). Esses trabalhos apresentam resultados promissores, competitivos e, em muitos casos, superiores às técnicas do estado da arte consideradas nas análises. Contudo, uma limitação relevante dessas abordagens reside no fato de que os aspectos sequenciais dos dados espectrais não são explicitamente explorados. Por outro lado, estudos recentes têm proposto modelos capazes de incorporar as dependências sequenciais dos dados de ATR-FTIR em seus classificadores (ANTONIOU et al., 2023; JUNIOR et al., 2025). Essas pesquisas alcançam resultados expressivos, evidenciando a importância de se considerar a natureza sequencial dos espectros no desenvolvimento de métodos de classificação mais eficazes para dados de ATR-FTIR. A seguir, são descritos os principais trabalhos relacionados.

2.4.1 Classificação de dados ATR-FTIR

A espectroscopia ATR-FTIR tem sido amplamente explorada na literatura como uma ferramenta eficiente para diagnóstico e triagem de diferentes condições clínicas, especialmente quando associada a algoritmos de aprendizado de máquina. De modo geral, esses trabalhos tratam os espectros como vetores de características, explorando principalmente relações de similaridade entre amostras para realizar a classificação.

Um dos estudos pioneiros nessa linha é apresentado em (BARAUNA et al., 2021), no qual os autores propõem um método ultrarrápido para a detecção in loco da infecção

por SARS-CoV-2 a partir de espectros ATR-FTIR. O trabalho combina uma aquisição espectral simples com algoritmos de análise estatística e de classificação supervisionada, alcançando elevados valores de sensibilidade e especificidade. Os resultados demonstram o potencial da espectroscopia ATR-FTIR como alternativa rápida, não invasiva e de baixo custo para diagnóstico, embora a abordagem trate os espectros de forma vetorial, sem considerar explicitamente sua natureza sequencial.

De forma semelhante, o trabalho (FERREIRA et al., 2020) investigam o uso da espectroscopia ATR-FTIR aplicada à análise de saliva para o diagnóstico de Câncer de Mama. Nesse trabalho, os espectros são pré-processados e utilizados como entrada para algoritmos de classificação tradicionais, evidenciando diferenças bioquímicas relevantes entre grupos saudáveis e patológicos. Os autores reportam resultados promissores, reforçando a aplicabilidade da técnica em contextos clínicos, porém novamente limitando-se à exploração de medidas de similaridade no espaço espectral.

No contexto do diagnóstico metabólico, o trabalho (CAIXETA et al., 2023) propõem a utilização de espectros salivários ATR-FTIR combinados com classificadores SVM para a triagem de Diabetes Mellitus tipo 2. O estudo demonstra que a combinação entre pré-processamento espectral e métodos clássicos de aprendizado supervisionado pode alcançar desempenho competitivo, indicando a presença de padrões discriminativos nos espectros. Ainda assim, o método assume independência entre as variáveis espectrais, não explorando dependências estruturais ou sequenciais entre os comprimentos de onda.

Mais recentemente, o trabalho (ANDRADE; SABINO-SILVA; CARNEIRO, 2024) avançam na automatização do processo de classificação ao empregar técnicas de Neural Architecture Search (NAS) evolutivo aplicadas a dados ATR-FTIR de saliva para o diagnóstico de Diabetes Mellitus tipo 2. O trabalho busca identificar arquiteturas neurais otimizadas diretamente a partir dos dados, obtendo resultados superiores a modelos manuais em determinados cenários. Apesar do avanço metodológico, os espectros continuam sendo tratados predominantemente como vetores de entrada, sem uma modelagem explícita das dependências sequenciais intrínsecas aos dados espectrais.

De forma geral, os trabalhos apresentados nesta subseção evidenciam o grande potencial da espectroscopia ATR-FTIR aliada a métodos de aprendizado de máquina para tarefas de classificação e diagnóstico. Contudo, observa-se que a maioria das abordagens se concentra na exploração de relações de similaridade entre os espectros, deixando em segundo plano aspectos estruturais, topológicos ou sequenciais que podem carregar informações relevantes para o processo de classificação.

2.4.2 Classificação de dados ATR-FTIR utilizando Redes Complexas

Nos últimos anos, abordagens baseadas em redes complexas têm sido amplamente investigadas para a classificação de dados ATR-FTIR, devido à sua capacidade de representar relações estruturais e padrões globais a partir de sinais espectrais. Diferentemente dos métodos tradicionais, que exploram predominantemente medidas de similaridade entre amostras ou atributos espectrais individuais, os modelos baseados em grafos permitem a extração de descritores topológicos de HL, os quais têm se mostrado eficazes em problemas de diagnóstico biomédico.

No contexto da detecção do TEA, o trabalho (ARAúJO; SABINO-SILVA; CARNEIRO, 2024) propõem uma abordagem de classificação de HL por caracterização de importância baseada em redes complexas aplicadas a dados salivares ATR-FTIR. Nesse trabalho, os espectros de cada classe são mapeados em um grafo a partir de estratégias de construção de redes baseados em vizinhos mais próximos, possibilitando a extração de métricas topológicas que representam propriedades estruturais dos dados. Os resultados apresentaram um desempenho competitivo e, em alguns cenários, superior às técnicas tradicionais aplicadas ao problema de classificação de dados ATR-FTIR.

Ainda no domínio da detecção de TEA, o modelo proposto em (FERNANDES; SABINO-SILVA; CARNEIRO, 2025) apresentam uma metodologia que combina redes complexas com algoritmos genéticos para a classificação de dados ATR-FTIR de saliva. O algoritmo genético é empregado com o objetivo de otimizar o processo de construção e seleção das redes utilizadas na etapa de caracterização de importância, buscando configurações de grafos que maximizem o desempenho do classificador. Os resultados obtidos demonstram que a otimização evolutiva contribui para a melhoria da capacidade discriminativa do modelo, reforçando o potencial da integração entre redes complexas e técnicas evolucionárias no contexto de biomarcadores espectrais.

No problema de detecção de Câncer Oral, o trabalho (LIMA FILHO et al., 2024) propõem um modelo de classificação de HL por conformidade de padrões baseado em redes complexas aplicado a dados ATR-FTIR salivares. A abordagem consiste na transformação dos espectros de cada classe em um grafo, que serão avaliados através de seis medidas redes selecionadas pelos autores. O trabalho avalia o desempenho de duas medidas de proximidade (cosseno e euclidiana), no processo de construção da rede MKNNG. Os resultados obtidos indicam que, ao utilizar a proximidade do cosseno, o classificador de HL apresenta desempenho superior a todos os classificadores analisados, incluindo métodos de última geração baseados em CNN.

Apesar dos resultados relevantes apresentados por essas pesquisas, observa-se que as abordagens baseadas em redes complexas para a classificação de dados ATR-FTIR concentram-se majoritariamente na exploração de propriedades topológicas dos grafos gerados. De modo geral, esses trabalhos não consideram explicitamente a natureza se-

quencial dos espectros ao longo dos números de onda, tratando os dados de forma essencialmente estática. Essa limitação evidencia uma lacuna na literatura, motivando o desenvolvimento de modelos que integrem simultaneamente aspectos topológicos e sequenciais dos dados ATR-FTIR, como proposto neste trabalho.

2.4.3 Classificação de dados ATR-FTIR considerando dependências sequenciais

Diferentemente das abordagens tradicionais de classificação de dados ATR-FTIR, que tratam os espectros como vetores estáticos de alta dimensionalidade, alguns trabalhos recentes passaram a considerar explicitamente a natureza sequencial desses dados ao longo dos números de onda. Essa perspectiva permite modelar dependências locais e globais presentes nos espectros, explorando a ordem espectral como uma fonte adicional de informação discriminativa.

Nesse contexto, o trabalho (ANTONIOU et al., 2023) propõem o uso de redes neurais recorrentes (RNNs) para a modelagem de espectros ATR-FTIR no domínio do tempo, aplicadas ao problema de detecção de Tumores Cerebrais. Os autores tratam os espectros como sequências ordenadas, permitindo que o modelo capture dependências sequenciais entre bandas espectrais adjacentes. Os resultados demonstram que a abordagem recorrente é capaz de extrair padrões relevantes associados a alterações bioquímicas, superando métodos tradicionais que desconsideram a estrutura sequencial dos dados.

De forma complementar, o modelo proposto em (JUNIOR et al., 2025) apresentam um modelo híbrido baseado em redes neurais convolucionais e redes LSTM bidirecionais (CNN-BiLSTM) para a detecção sustentável e não invasiva da COVID-19 a partir de dados salivares ATR-FTIR. Nessa abordagem, as camadas convolucionais são empregadas para a extração de padrões locais ao longo do espectro, enquanto as camadas BiLSTM modelam dependências sequenciais de longo alcance entre os números de onda. Os resultados obtidos indicam desempenho elevado em termos de acurácia, sensibilidade e especificidade, evidenciando a importância de explorar explicitamente as dependências sequenciais inerentes aos dados ATR-FTIR.

Embora esses trabalhos demonstrem que a consideração da natureza sequencial dos espectros ATR-FTIR pode resultar em melhorias significativas no desempenho de classificação, observa-se que tais abordagens baseiam-se exclusivamente em modelos neurais profundos, sem incorporar representações estruturais ou topológicas dos dados. Dessa forma, permanece em aberto a investigação de métodos que integrem simultaneamente informações sequenciais e topológicas, como aquelas obtidas por meio de grafos, motivando a proposta apresentada neste trabalho.

2.5 Considerações finais

Neste Cápitulo foi abordado os conceitos fundamentais do trabalho, como a classificação em AM, bem como alguns classificadores. Redes complexas e suas medidas de rede, processos de construção de redes complexas, classficação baseadas em rede: por conformidade padrão e caracterização de importância. Uma breve contextualização sobre a espectroscopia ATR-FTIR, princípios, funcionamento e dados sequênciais de espectros salivares para o AM. Alguns trabalhos recentes como (LIMA FILHO et al., 2024) utiliza espectros ATR-FTIR salivares para classificação baseada em redes complexas, mas ainda não foi explorado métodos de construção de rede baseados em grafos de visibilidade com a classificação híbrida, que é a proposta deste trabalho, detalhada no próximo Cápitulo.

Capítulo 3

VisG2: classificação de alto nível baseada em meta-grafo de grafos de visibilidade

Os classificadores tradicionais utilizados em tarefas de AM comumente denominados classificadores de LL realizam a categorização de dados com base em características físicas e estatísticas diretamente observáveis, como distância entre vetores, grau de similaridade ou padrões de distribuição dos atributos (CARNEIRO, 2016; SILVA; ZHAO, 2012). Embora eficientes em muitos cenários, esses modelos não levam em consideração o contexto estrutural ou relacional em que os dados estão inseridos. Por outro lado, os classificadores de HL oferecem uma abordagem alternativa: em vez de se apoiarem apenas nos valores individuais dos atributos, eles exploram as conexões e padrões topológicos formados entre os dados em uma estrutura de rede. Através de medidas de redes complexas, como ASP, CCF e ASS, esses modelos são capazes de capturar comportamentos coletivos e interações implícitas entre instâncias (SILVA; ZHAO, 2012; CARNEIRO et al., 2014).

Esse paradigma de classificação vem se consolidando como uma promissora linha de pesquisa, especialmente em domínios onde padrões emergentes entre dados desempenham um papel crucial. Um exemplo recente dessa tendência está nos trabalhos que envolvem espectroscopia salivar com a técnica ATR-FTIR, em que classificadores de HL demonstraram superioridade na identificação de condições médicas complexas, como o TEA e Câncer Bucal (LIMA FILHO et al., 2024; ARAúJO; SABINO-SILVA; CARNEIRO, 2024).

Apesar dos avanços, ainda persistem desafios importantes na aplicação de classificadores de HL a dados espectroscópicos. Um dos principais obstáculos é a ausência de métodos eficazes para construção de redes que incorporem características sequenciais dos dados, típicas de domínios como espectroscopia e séries temporais. Em contextos como esse, os atributos apresentam um encadeamento semântico relevante ou seja, a ordem e a relação entre os pontos carregam informações fundamentais que não podem ser ignoradas. No entanto, os métodos convencionais de construção de redes geralmente tratam os dados

como pontos independentes no espaço, perdendo essa rica estrutura sequencial.

Para suprir essa lacuna, este trabalho propõe o uso de grafos de visibilidade (visibility graphs), originalmente concebidos para representar séries temporais (LACASA et al., 2008). Essa técnica transforma sequências de dados em grafos cujas arestas são definidas com base em critérios geométricos de visibilidade entre pontos, preservando a estrutura sequencial e permitindo que padrões topológicos emergentes sejam analisados em uma representação gráfica.

Neste contexto, o capítulo apresenta uma nova abordagem para classificação de dados espectroscópicos: um modelo de HL baseado na técnica de conformidade padrão em redes complexas, combinada com grafos de visibilidade para representar instâncias individuais. O modelo proposto, denominado VisG2, constrói um meta-grafo, onde cada nó representa um espectro individual, modelado como um grafo de visibilidade, e as conexões entre os nós são estabelecidas com base em medidas de similaridade entre esses grafos. Essa estrutura permite capturar tanto a dinâmica interna dos espectros quanto suas relações intergrupais, favorecendo uma análise mais robusta e sensível de padrões.

Além disso, propõe-se uma estratégia de classificação híbrida, que combina os pontos fortes das abordagens de baixo e alto nível, potencializando a capacidade discriminativa do modelo.

Este capítulo está organizado da seguinte forma: Na Seção 3.1, é apresentada uma visão geral da técnica proposta, incluindo um fluxograma que ilustra todas as etapas do modelo. Em seguida, a Seção 3.2 introduz os conceitos relacionados aos grafos de visibilidade, enquanto a Seção 3.3 detalha o processo de construção da rede a partir desses grafos. A Seção 3.4 descreve o funcionamento do classificador de HL baseado na conformidade padrão, e a Seção 3.5 apresenta a abordagem híbrida de classificação. A complexidade computacional do modelo é discutida na Seção 3.6. Por fim, a Seção 3.7 recapitula os principais pontos abordados.

3.1 Descrição da técnica

A técnica proposta baseia-se em um modelo híbrido de classificação, que integra dois paradigmas distintos: o de LL e o de HL. A ideia central é combinar, por meio de uma função de combinação linear, as probabilidades geradas por ambos os classificadores, buscando aproveitar as vantagens de cada abordagem em um processo de decisão mais robusto.

As probabilidades de LL são provenientes de classificadores tradicionais amplamente utilizados em tarefas de AS, como KNN, SVM e LDA. Esses modelos se baseiam exclusivamente nos atributos físicos dos dados como distâncias, ângulos ou distribuições para realizar a categorização. Por serem eficientes e consolidados, fornecem uma base sólida de classificação fundamentada na similaridade direta entre as amostras.

Em contrapartida, o classificador de HL adota uma abordagem fundamentada em redes complexas. Ele constrói uma rede para cada classe presente nos dados e analisa como as medidas de rede variam com a inserção de uma nova instância de teste. Essas medidas como o CCF, a ASS ou a BET capturam padrões estruturais que emergem da organização dos dados como um sistema interconectado, permitindo identificar conformidades e anomalias em relação às classes já estabelecidas.

Durante a fase de teste, ambos os classificadores operam de maneira independente. O classificador de LL, como o KNN, simplesmente armazena os dados de treinamento e utiliza critérios de proximidade para classificar novas amostras. Já o classificador de HL constrói grafos para cada classe com base nos dados já conhecidos e, ao inserir a nova instância de teste, avalia como sua presença altera as características topológicas dessas redes. A partir dessa análise, ele gera uma pontuação que reflete a "conformidade"da instância com cada classe.

A combinação dos dois modelos é feita por uma função linear ponderada, cuja equação será detalhada em seções posteriores. O parâmetro λ é responsável por regular a influência relativa de cada componente no resultado final. Sua escolha pode variar conforme o classificador de LL adotado. Por exemplo, no caso do KNN, experimentos mostraram melhor desempenho ao se atribuir um peso maior ao componente de HL, dada sua capacidade de capturar relações topológicas mais refinadas em dados sequenciais (LACASA et al., 2008).

A Figura 14 ilustra o funcionamento dessa técnica híbrida em um cenário binário, com duas classes representadas visualmente pelas formas triângulo/laranja e quadrado/vermelho. O classificador de LL (neste exemplo, o KNN) considera apenas a proximidade dos pontos no espaço de atributos, o que o leva a favorecer a classe vermelha. Em contraste, o classificador de HL observa a inserção do novo ponto na rede e, com base no padrão topológico formado, identifica que ele está mais alinhado com a classe laranja. A decisão final é então obtida pela combinação das probabilidades dos dois modelos, ponderadas por λ, resultando em uma classificação mais equilibrada e informada.

A proposta central deste trabalho consiste na construção de redes baseadas em grafos de visibilidade, técnica que dá origem ao modelo denominado VisG2. Esse processo é ilustrado na Figura 15 e representa um avanço na estruturação de dados sequenciais obtidos por meio da espectroscopia ATR-FTIR para fins de classificação.

Durante a fase de treinamento, cada instância dos dados sequenciais é convertida em um grafo de visibilidade, onde os pontos da sequência são transformados em vértices conectados conforme critérios de visibilidade definidos na literatura (LACASA et al., 2008). Cada grafo gerado representa um nó dentro de uma rede complexa, sendo que uma rede distinta é construída para cada classe presente no conjunto de treinamento. Assim, cada classe é representada por uma rede composta por múltiplos grafos de visibilidade.

Com as redes por classe devidamente estruturadas, calcula-se um conjunto de medidas

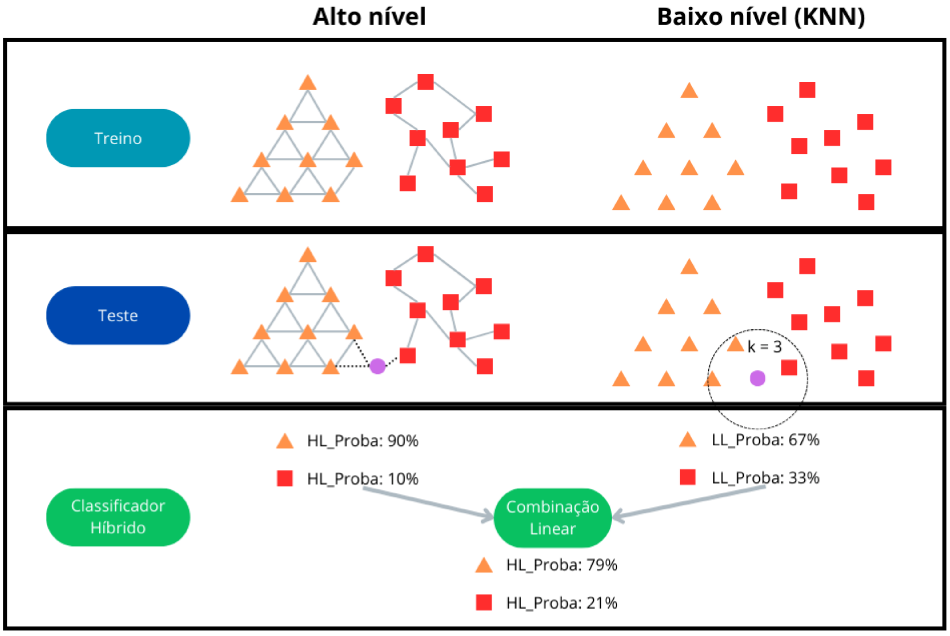


Figura 14 – Panorama da técnica híbrida de classificação de dados em treino teste e composição das probabilidades.

de rede para capturar as propriedades topológicas de cada uma. Essas medidas como o CCF, o ADG, ou outras métricas estruturais formam a base para avaliar a conformidade de novos dados com os padrões previamente aprendidos.

Na fase de teste, um novo dado sequencial é também transformado em um grafo de visibilidade. Esse grafo é então temporariamente conectado a cada uma das redes de classe, e as medidas topológicas são novamente calculadas. A ideia é avaliar o impacto da inserção desse novo vértice (grafo) sobre a estrutura original da rede. Quanto menor for a variação nas medidas de rede após a inserção, maior será a compatibilidade do novo dado com os padrões da classe correspondente.

A classificação de HL é então realizada com base nesse critério de conformidade estrutural. A classe atribuída ao novo dado será aquela cuja rede apresentar a menor alteração em suas propriedades topológicas, indicando que o grafo de visibilidade do teste se alinha melhor com o padrão topológico daquela classe.

3.2 Grafo de visibilidade

A ideia dos grafos de visibilidade emergiu como uma abordagem inovadora para representar séries temporais ou sequências numéricas sob a forma de estruturas de redes complexas. Inicialmente proposta por Lacasa em (LACASA et al., 2008), essa técnica transforma uma sequência numérica em um grafo no qual cada ponto da sequência é representado por um nó, e as conexões entre os nós são estabelecidas com base em um critério geométrico de visibilidade. A inspiração para esse modelo vem de problemas

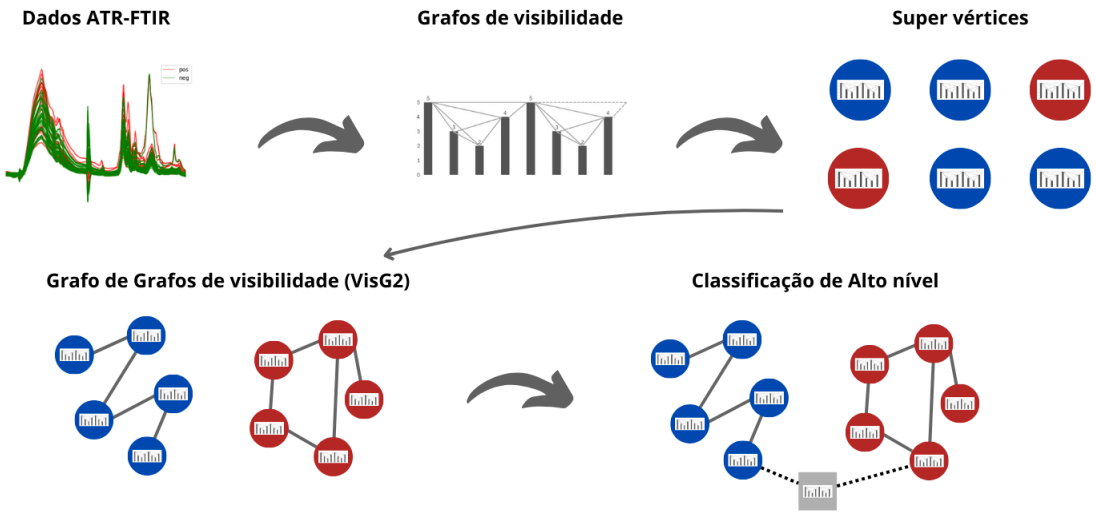


Figura 15 – Panorama da técnica de construção do meta-grafo de grafos de visibilidade.

clássicos das áreas de computação gráfica e visão computacional, onde o conceito de visibilidade é comumente utilizado para definir relações espaciais entre objetos.

Desde sua introdução, os grafos de visibilidade se destacaram por sua capacidade de preservar propriedades essenciais dos dados originais, permitindo não apenas uma representação alternativa da série, mas também a identificação de padrões estruturais ocultos. Por meio dessa transformação, torna-se possível aplicar métodos de análise de redes complexas à série, abrindo novas possibilidades para o entendimento de seus comportamentos dinâmicos e estruturais. Diferentes variantes do grafo de visibilidade foram desenvolvidas, como o grafo de visibilidade horizontal (HVG) e o grafo de visibilidade natural, a fim de atender aos requisitos específicos de diferentes tipos de séries temporais e sequências (ZOU et al., 2019).

Atualmente, os grafos de visibilidade são amplamente reconhecidos como uma ferramenta poderosa na análise de dados sequenciais, pois permitem a extração de características topológicas relevantes de maneira eficiente e interpretável. Sua adoção em contextos como a detecção de transições de fase, classificação de sinais fisiológicos ou análise de instabilidades financeiras reforça seu valor metodológico e prático.

Neste trabalho, utilizamos o grafo de visibilidade tradicional proposto por Lacasa (LACASA et al., 2008), que é particularmente eficaz para converter dados espectroscópicos sequenciais, como os obtidos por meio de ATR-FTIR, em redes complexas. A construção desse grafo se apoia em dois princípios fundamentais:

- ❏ Cada ponto da sequência numérica é representado por um nó no grafo.
- ❏ Dois nós t_i e t_j (com $i < j$) são conectados por uma aresta se não houver obstrução visual direta entre os pontos correspondentes na sequência.

Mais formalmente, dois pontos da série temporal (t_a, y_a) e (t_b, y_b) estão conectados no

grafo de visibilidade se, para todo ponto intermediário (t_c, y_c) onde $a < c < b$, a seguinte condição for satisfeita:

$$y_c < y_b + (y_a - y_b)\frac{t_c - t_a}{t_b - t_a}, \tag{20}$$

Essa equação assegura que a linha reta traçada entre os pontos t_a e t_b não é interceptada por nenhum ponto entre eles, ou seja, não há obstrução visual direta. Isso define a visibilidade geométrica entre os dois nós e, por consequência, sua conexão no grafo.

Essa técnica oferece uma maneira natural de capturar a estrutura interna de sequências numéricas, respeitando a ordem e a relação entre seus elementos. No contexto deste projeto, ela é a base para a construção de redes mais expressivas, nas quais padrões de classe podem ser analisados com maior profundidade utilizando métricas de redes complexas.

A Figura 16 ilustra a construção de um grafo de visibilidade, no qual a Equação 20 é utilizada para determinar quais elementos da sequência são visíveis entre si a partir de um referencial geométrico. Nesse cenário, cada barra vertical representa um vértice do grafo, enquanto as linhas traçadas entre elas indicam as arestas, estabelecidas sempre que houver visibilidade direta entre dois pontos da sequência, ou seja, quando nenhum ponto intermediário bloqueia a visão entre eles.

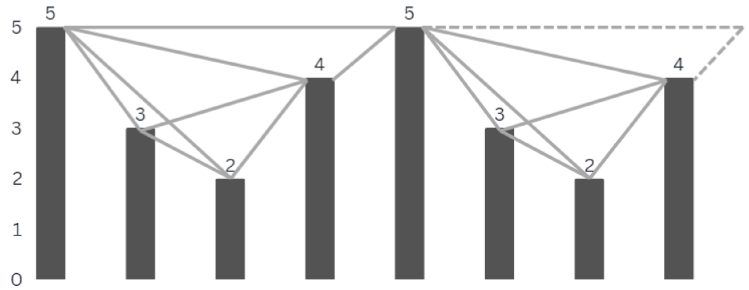


Figura 16 – Representação de um grafo de visibilidade, onde as barras correspondem aos vértices e as linhas conectando as barras visíveis entre si representam as arestas.

3.3 Construção de meta-grafos baseados em grafos de visibilidade

A construção de redes complexas representativas é uma etapa fundamental na aplicação de classificadores de HL, uma vez que toda a lógica de conformidade de padrões depende diretamente da topologia resultante dessas redes. Para cada classe presente na base de dados, é necessário construir uma rede complexa que traduza, de forma adequada,

as relações internas entre os elementos da classe. Assim, o desempenho do classificador HL está diretamente vinculado à qualidade da estrutura de rede gerada.

Tradicionalmente, os métodos mais utilizados para construção de redes em contextos supervisionados são baseados em estratégias como os k-vizinhos mais próximos ou a vizinhança por raio-ϵ. Tais abordagens conectam os elementos de acordo com sua proximidade em termos de distância (euclidiana, cosseno, etc.), formando grafos não direcionados. Com o tempo, surgiram diversas variações e melhorias sobre esses métodos como o *Mutual KNN* e *Selective KNN* que buscam aperfeiçoar a conectividade e eliminar conexões espúrias. No entanto, apesar de eficientes para dados com estrutura vetorial convencional, essas técnicas falham em capturar propriedades sequenciais ou estruturais mais complexas, como as encontradas em dados espectroscópicos ou séries temporais.

Diante dessa limitação, os grafos de visibilidade emergem como uma alternativa promissora. Essa técnica, como apresentado na seção anterior, é capaz de converter sequências numéricas em grafos com base em critérios geométricos de visibilidade. Sua grande vantagem reside na capacidade de preservar relações sequenciais implícitas nos dados originais, algo que os métodos convencionais de proximidade não conseguem realizar de forma satisfatória.

Neste trabalho, propõe-se um novo modelo de construção de redes baseado em grafos de visibilidade, em que cada sequência é transformada em um grafo individual e, posteriormente, cada um desses grafos é tratado como um supernó em uma rede mais ampla. O resultado é uma estrutura de HL que pode ser interpretada como um meta-grafo de grafos de visibilidade. Essa rede complexa preserva tanto a informação local (dentro de cada sequência, por meio de seu grafo de visibilidade) quanto a estrutura global (interações entre as sequências).

A técnica proposta, denominada VisG2, consiste em construir um meta-grafo a partir de grafos de visibilidade gerados individualmente para cada instância sequencial da base de dados. O objetivo principal é capturar, de forma estruturada, a similaridade entre essas instâncias com base em suas representações topológicas ou seja, suas conexões em termos de visibilidade. Essa abordagem é especialmente eficaz para dados espectroscópicos, como os espectros de ATR-FTIR, que possuem forte coerência sequencial e padrões locais bem definidos.

O processo de construção do meta-grafo VisG2 começa com a transformação de cada instância sequencial da base de dados em um grafo de visibilidade, conforme detalhado na seção anterior. Cada grafo gerado a partir dessa etapa é tratado como um super nó (sv_i) dentro de uma rede complexa. A ideia central é que a conexão entre esses super nós seja determinada pela similaridade estrutural entre seus respectivos grafos de visibilidade.

Essa similaridade é calculada com base no índice de Jaccard, uma métrica bastante conhecida em tarefas de comparação de conjuntos, e que, neste contexto, é aplicada sobre as arestas dos grafos. Formalmente, a similaridade entre dois grafos de visibilidade G_i e

G_j é dada por:

$$J(G_i, G_j) = \frac{|E_i \cap E_j|}{|E_i \cup E_j|}, \tag{21}$$

onde E_i e E_j representam os conjuntos de arestas dos grafos G_i e G_j, respectivamente. O numerador da equação quantifica as arestas compartilhadas (interseção), enquanto o denominador considera o total de arestas distintas (união). O valor de $J(G_i, G_j)$ varia entre 0 e 1, indicando, respectivamente, ausência total ou sobreposição completa entre os grafos comparados.

Alternativamente, essa métrica pode ser expressa em termos das matrizes de adjacência A^i e A^j dos grafos, onde $A^i, A^j \in \{0,1\}^{|V|\times|V|}$. Essa formulação matricial permite uma implementação computacional mais eficiente, especialmente em ambientes com grandes volumes de dados:

$$J(G_i, G_j) = \frac{\sum_{z,k} A^i_{zk} A^j_{zk}}{\sum_{z,k} \max(A^i_{zk}, A^j_{zk})}. \tag{22}$$

Após o cálculo de similaridade entre todos os pares de grafos de visibilidade, inicia-se a construção do meta-grafo VisG2. Para isso, define-se um número k, que representa o número de conexões que cada super nó deverá manter com os grafos mais similares dentro de sua própria classe. O meta-grafo é então formalizado como $mG = (sV, sE)$, onde:

- sV é o conjunto de super nós, sendo cada sv_i correspondente a um grafo de visibilidade G_i;
- sE é o conjunto de arestas, estabelecido com base em regras de vizinhança e rótulo de classe.

As conexões no meta-grafo são realizadas de acordo com o seguinte critério:

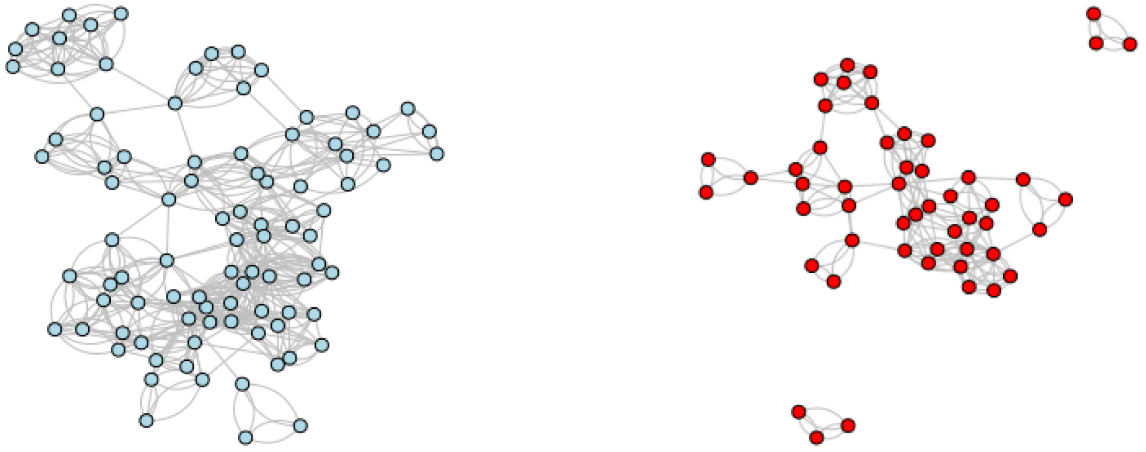
$$se_{ij} = \begin{cases} 1, & \text{se } sv_j \in \text{top-}k(sv_i) \textbf{ e } y_i = y_j, \\ 0, & \text{caso contrário.} \end{cases} \tag{23}$$

Em outras palavras, dois super nós sv_i e sv_j estarão conectados se:

1. sv_j estiver entre os k vizinhos mais similares de sv_i (ou vice-versa); e
2. ambos pertencerem à mesma classe (isto é, possuírem o mesmo rótulo y).

Esse critério garante que a estrutura da rede final preserve a coerência semântica dos dados, agrupando apenas grafos estruturalmente similares e semanticamente compatíveis. O resultado é uma rede de HL capaz de refletir padrões topológicos relevantes entre instâncias da mesma classe, servindo como base para a classificação baseada em conformidade de padrões, conforme será discutido nas seções seguintes.

A Figura 17 mostra duas redes geradas pela heurística de construção de rede VisG2 com valor de $k = 8$. A primeira rede, Figura 17(a)(vértices de azul), representam a rede de controle, já a segunda rede, Figura 17(b)(vértices de vermelho), representa a rede de TEA.



(a) Rede de amostras controle

(b) Rede de amostras TEA

Figura 17 – Exemplos de redes de controle e TEA, gerados pela técnica VisG2.

3.4 Classificador de alto nível via conformidade padrão

A classificação de HL propõe uma abordagem baseada na análise da conformidade estrutural de um item de teste em relação aos padrões topológicos previamente aprendidos durante o treinamento. Em vez de utilizar apenas características estatísticas locais ou globais dos dados, essa técnica avalia como a adição de um novo item afeta a estrutura de redes complexas previamente construídas para cada classe. A hipótese central é que, quanto menor a perturbação causada por um novo item em uma rede de classe, maior é sua compatibilidade com os padrões dessa classe, e, consequentemente, maior a probabilidade de pertencimento.

A fase de treinamento consiste na construção das redes de HL, conforme descrito na seção anterior, onde cada classe da base de dados é representada por uma rede complexa. Já a fase de teste envolve a inserção de uma nova instância nas redes de cada classe, e a posterior avaliação da variação de determinadas medidas topológicas causadas por essa inserção.

Formalmente, dada uma medida de rede u, e considerando $m^{(c)}$ como o valor da medida antes da inserção de um item de teste y em uma rede de classe $G^{(c)}$, e $m'^{(c)}$ como o valor após a inserção, a variação relativa dessa medida é definida pela Equação 24.

$$\Delta G_y^{(c)}(u) = \frac{\left|m^{(c)} - m'^{(c)}\right|}{\sum_{q \in L} \left|m^{(q)} - m'^{(q)}\right|} \tag{24}$$

em que L é o conjunto de todas as classes. Essa equação permite comparar o impacto da inserção do item em diferentes redes. Quanto menor o valor de $\Delta G_y^{(c)}(u)$, menor a alteração estrutural provocada, o que indica maior compatibilidade com a rede.

Na prática, valores elevados de variação indicam que o item é estruturalmente diferente dos demais componentes da rede, sugerindo baixa conformidade com a classe. Por outro lado, valores próximos de zero indicam alta compatibilidade. No entanto, um problema pode surgir quando o item de teste não gera nenhuma conexão com a rede o que ocorre, por exemplo, quando ele é topologicamente muito distinto. Nesses casos, a variação se torna nula, o que poderia ser interpretado erroneamente como alta conformidade. Para mitigar esse problema, atribui-se nesses casos o valor máximo de variação, penalizando corretamente a ausência de conexões relevantes.

A probabilidade de pertencimento de um item de teste y à classe c é então definida pela Equação 25.

$$H_y^{(c)} = \frac{\sum_{u=1}^{Z} \alpha(u) \left[1 - f_y^{(c)}(u)\right]}{\sum_{g \in L} \sum_{u=1}^{Z} \alpha(u) \left[1 - f_y^{(g)}(u)\right]} \tag{25}$$

em que:

- Z é o número total de medidas de rede utilizadas;
- $\alpha(u) \in [0, 1]$ representa o peso (ou influência) atribuído à medida u na decisão final, com $\sum_{u=1}^{Z} \alpha(u) = 1$;
- $f_y^{(c)}(u)$ é a variação normalizada da medida u para a classe c, considerando o balanceamento das classes.

A normalização de $f_y^{(c)}(u)$ é feita da seguinte forma:

$$f_y^{(c)}(u) = \Delta G_y^{(c)}(u) \cdot p^{(c)}, \tag{26}$$

em que $p^{(c)}$ representa a proporção de instâncias pertencentes à classe c, definida por:

$$p^{(c)} = \frac{1}{V} \sum_{u=1}^{V} \nVdash_{[y_u = c]}, \tag{27}$$

sendo V o número total de vértices da rede e $\nVdash_{[y_u=c]}$ a função indicadora que retorna 1 se o vértice y_u pertence à classe c, e 0 caso contrário.

Ao final desse processo, o item de teste y é atribuído à classe c que apresentar o maior valor de $H_y^{(c)}$, ou seja, àquela cuja estrutura sofreu a menor perturbação relativa e ponderada após a inserção do novo item.

3.5 Combinação de alto e baixo nível

Com os vetores de pertinência de uma medida de rede, que representam o classificador de HL, e o vetor de probabilidades do classificador de LL, é possível combinar ambos os vetores, gerando assim a classificação híbrida. Dessa forma, a probabilidade de um item de teste y pertencer à classe c é dada pela Equação 28.

$$M_y^{(c)} = \lambda H_y^{(c)} + (1 - \lambda) C_y^{(c)}, \tag{28}$$

em que $H_y^{(c)}$ é a probabilidade fornecida pelo classificador de HL e $C_y^{(c)}$ é a probabilidade do classificador de LL. O parâmetro λ representa uma combinação linear convexa entre ambos os classificadores, priorizando o classificador de HL para valores mais altos de λ, enquanto para valores mais baixos, a prioridade recai sobre o classificador de LL.

3.6 Algoritmo e complexidade

Nessa seção será discutido sobre a complexidade do algoritmo proposto. Inicialmente será tetalhado os principais passo do código e em cada trecho terá uma explicação de sua complexidade. No final será avalidado a complexidade total da classificação híbrida baseada em grafos de visibilidade.

1. Construir as redes para cada classe da base de dados, a partir de dados sequenciais ATR-FTIR:

   - ❏ Transformar todas as instâncias em grafos de visibilidade.

     Complexidade: Para cada instância X, é construído um grafo de visibilidade. Considerando uma série temporal com V pontos, a construção do grafo de visibilidade natural requer a verificação da condição de visibilidade entre pares de vértices, resultando em complexidade computacional média $O(V^2)$, podendo atingir $O(V^3)$ no pior caso. Dessa forma, a complexidade total para a construção dos grafos associados a todas as n instâncias do conjunto de dados é dada por $O(nV^2)$.

   - ❏ Calcular a matriz de similaridade de JACCARD.

     Complexidade: Para calcular a matriz de similaridade, é necessário analisar cada par de grafos de visibilidade. No entanto, um aspecto importante é que percorremos apenas metade da matriz, já que os valores nas diagonais superior e inferior são idênticos. Embora essa abordagem não reduza a complexidade no pior caso, ela melhora a eficiência do algoritmo, tornando sua execução mais rápida. Assim, a complexidade do algoritmo é: $O(n^2)$ em que n é a quantidade de grafos. Alternativamente, a mesma complexidade se aplica para comparação das matrizes de adjacência.

Embora o cálculo exato da matriz de similaridade de Jaccard apresente complexidade $O(n^2)$ no pior caso, diversas estratégias podem ser empregadas para reduzir o custo computacional na prática. Entre elas destacam-se o uso de representações compactas dos grafos, técnicas de aproximação da similaridade, exploração da esparsidade dos grafos de visibilidade, bem como estratégias de filtragem e paralelização. Tais abordagens permitem diminuir significativamente o tempo de execução sem comprometer a capacidade discriminativa do modelo.

Como melhoria futura do modelo, pretende-se adotar uma estratégia para reduzir o número de comparações entre as redes, evitando a análise direta entre grafos que, com base em critérios simples, já apresentam baixa probabilidade de similaridade. Dessa forma, em vez de realizar comparações exaustivas entre todos os pares de redes, o modelo passa a considerar apenas candidatos plausíveis. Inicialmente, as redes são organizadas em clusters compostos por grafos potencialmente semelhantes, utilizando descritores estruturais de baixo custo computacional, tais como grau médio, densidade da rede e coeficiente de agrupamento médio. Esse agrupamento pode ser realizado por meio de estruturas de indexação ou algoritmos de particionamento com custo aproximado $O(n \log n)$, decorrente do processo de ordenação ou busca hierárquica dos descritores. Esse procedimento garante que grafos estruturalmente distintos não sejam comparados diretamente. Em seguida, o cálculo da similaridade de Jaccard é realizado apenas entre grafos pertencentes ao mesmo cluster, reduzindo significativamente o número de comparações necessárias. Enquanto a abordagem exaustiva apresenta complexidade quadrática $O(n^2)$, a restrição das comparações a subconjuntos candidatos faz com que o número médio de avaliações cresça de forma quase linear, resultando em uma complexidade média aproximada de $O(n \log n)$, tornando o método mais eficiente na prática.

- ❑ Ordenar as top-k similaridades.

  Complexidade: para cada linha da matriz (ou seja, para cada grafo), a função realiza uma ordenação, que tem complexidade $O(n \log n)$. Como isso é feito para todos os objetos (n objetos), a complexidade total dessa operação é: $O(n^2 \log n)$.

  Embora o cálculo exato das top-k similaridades para todos os grafos apresente complexidade quadrática no pior caso, é possível reduzir significativamente o custo computacional ao restringir o número de comparações. Ao empregar estratégias de clusterização prévia baseadas em descritores simples das redes, cada grafo passa a ser comparado apenas com um subconjunto reduzido de candidatos. Sob a hipótese de clusters balanceados com tamanho médio $O(\log n)$, a complexidade esperada do processo torna-se próxima de $O(n \log n)$.

- ❏ Separar os dados por classe.

  Complexidade: a operação para verificar se os objetos pertencem à mesma classe tem complexidade $O(k)$ para cada linha. Como temos n objetos, a complexidade total é: $O(n \cdot k)$.

- ❏ Criação dos grafos de cada classe.

  Complexidade: considerando l como a quantidade total de classes distintas, e a criação de cada subgrafo sendo feita em um tempo $O(V + E)$, onde V é o número de vértices e E o número de arestas, a complexidade total da tarefa é: $O(l \cdot (V + E))$.

2. Calcular as medidas de rede (6 medidas) para cada rede.

   Os grafos gerados pelo modelo proposto apresentam natureza esparsa, uma vez que o processo de construção das redes a partir dos grafos de visibilidade resulta em um número de arestas proporcional ao número de vértices, isto é, $m \ll n^2$. Essa característica impacta diretamente a complexidade computacional do cálculo das medidas de rede, tornando o processo mais eficiente quando comparado a grafos densos. Considerando essa propriedade, a complexidade das medidas utilizadas é descrita a seguir:

   - ❏ **Assortatividade**: possui complexidade de $O(n)$, sendo computacionalmente eficiente mesmo para grafos de maior porte, especialmente em cenários esparsos.
   - ❏ **Coeficiente de Agrupamento**: apresenta complexidade de $O(n \cdot k^2)$, onde k representa o grau médio dos nós. Em grafos esparsos, k é pequeno e limitado, o que reduz significativamente o custo computacional dessa medida na prática.
   - ❏ **Grau médio**: possui complexidade de $O(n)$, uma vez que depende apenas da contagem das arestas associadas a cada vértice.
   - ❏ **Intermedialidade (Betweenness)**: apresenta complexidade de $O(n^2)$ quando calculada de forma exata. Entretanto, em grafos esparsos, o uso de algoritmos otimizados, como o de Brandes (BRANDES, 2001), reduz o custo para $O(nm)$, tornando a medida viável para o cenário considerado.
   - ❏ **Menor caminho médio**: possui complexidade de $O(n(n + m))$. Para grafos densos, esse custo se aproxima de $O(n^3)$, enquanto que, para grafos esparsos, como os gerados neste trabalho, o custo efetivo é reduzido para $O(nm)$.
   - ❏ **Proximidade (Closeness)**: apresenta complexidade semelhante à do menor caminho médio, isto é, $O(n(n+m))$, sendo computacionalmente mais eficiente em grafos esparsos, onde $m \ll n^2$.

Dessa forma, a esparsidade inerente aos grafos de visibilidade gerados pelo modelo proposto contribui para a redução do custo computacional no cálculo das medidas de rede, viabilizando a aplicação do método mesmo em conjuntos de dados com maior número de vértices.

3. Inserir temporariamente um vértice de teste nas redes com base nas top-k similaridades.

   - ❏ Cria o grafo de visibilidade e calcula a matriz de similaridade de JACCARD para o item de teste x.

     Complexidade: a complexidade no pior caso é $O(nm)$. Sendo n o número de amostras de teste, e m o número de amostras de treinamento.

   - ❏ Seleciona as top-k similaridades.

     Complexidade: é $O(nm)$, com n sendo o número de amostras de teste, e m o número de amostras de treinamento.

4. Calcular os valores de pertinência para cada instância x em cada classe l com base na variação das medidas de rede (Dado nas eqs. (24) to (27)) .

   Complexidade: considerando C como o número de classes distintas, V o número de vértices em cada rede e n o número de amostras de teste, a complexidade teórica original dessa etapa seria $O(nCV^2)$. No entanto, no modelo proposto, as redes associadas a cada classe permanecem fixas durante a fase de teste, permitindo que as medidas de rede sejam pré-computadas previamente. Dessa forma, durante a classificação, calcula-se apenas o impacto incremental da inserção da instância de teste, cujo custo é linear no número de vértices. Além disso, como os grafos gerados são esparsos, com $m \ll V^2$, o custo efetivo das operações reduz-se para $O(V + m)$, que na prática é aproximadamente $O(V)$. Assim, a complexidade final dessa etapa é reduzida para: $O(nCV)$.

5. Complexidade total do classificador de HL, com o método de construção de rede baseados em grafos de visibilidade é dada por:

   - ❏ Construção da rede: $O(nV^2 + n\log n + n\log n + nk + l(V + E))$ Termo de ordem mais alta é $O(n\log n)$
   - ❏ Cálculo das medidas de rede: $O(n + nk^2 + n + nm + nm + n(n + m))$ Termo de ordem mais alta é $O(n^2)$
   - ❏ Conexão temporária nas redes: $O(nm + nm)$ Termo de ordem mais alta é $O(nm)$
   - ❏ Cálculo dos valores de pertinência: a complexidade é dada por $O(nCV)$

Sendo a complexidade final dada por $O(nV^2 + n^2 + nm + nCV)$ com o termo de ordem mais alta sendo $O(nV^2)$.

6. Combinar as probabilidades geradas pelos classificador de baixo e alto nível (Dado na Equação 28).

   Complexidade: é $O(nV^2 + h(n, d, |L|))$, em que h é o classificador de LL, n é o número de amostras, d são os atributos e $|L|$ são as classes.

3.7 Considerações finais

Neste Capítulo 3, foram detalhados todos os passos da técnica de classificação de HL baseada em grafos de visibilidade. Inicialmente, a Seção 3.1 apresenta uma visão geral da técnica desenvolvida. Em seguida, a Seção 3.2 discute o conceito de grafo de visibilidade, enquanto a Seção 3.3 aborda sua aplicação na construção da rede. A proposta é fundamentada na similaridade entre grafos de visibilidade (VisG2). A Seção 3.4 apresenta o classificador de HL por conformidade padrão, seguido da Seção 3.5, que descreve a combinação das classificações de alto e baixo nível. O capítulo se encerra com a análise da complexidade do algoritmo, detalhada na Seção 3.6.

Capítulo 4

Experimentos Câncer Oral

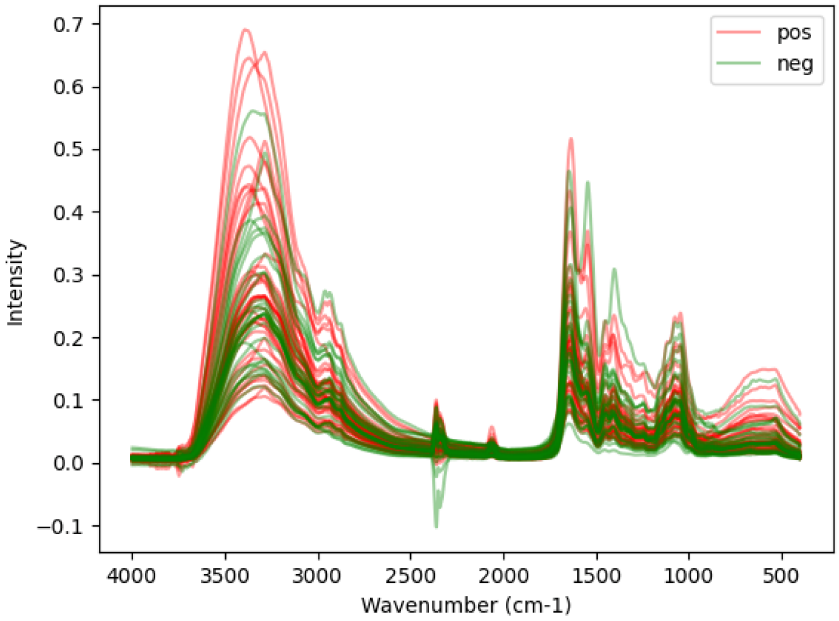
Neste capítulo serão apresentadas algumas simulações para evidenciar o potencial da técnica VisG2 na solução de problemas de sequências derivados de dados ATR-FTIR. Inicialmente, será realizado uma breve descrição da base de dados, bem como as técnicas de pré-processamento adotadas. Em seguida, serão analisados os resultados obtidos pela técnica VisG2 base (sem combinação com classificadores de LL). Por fim, esses resultados serão comparados com outros classificadores do estado da arte aplicados ao problema de detecção de Câncer Oral.

4.1 Base de dados Câncer Oral e preparação de dados

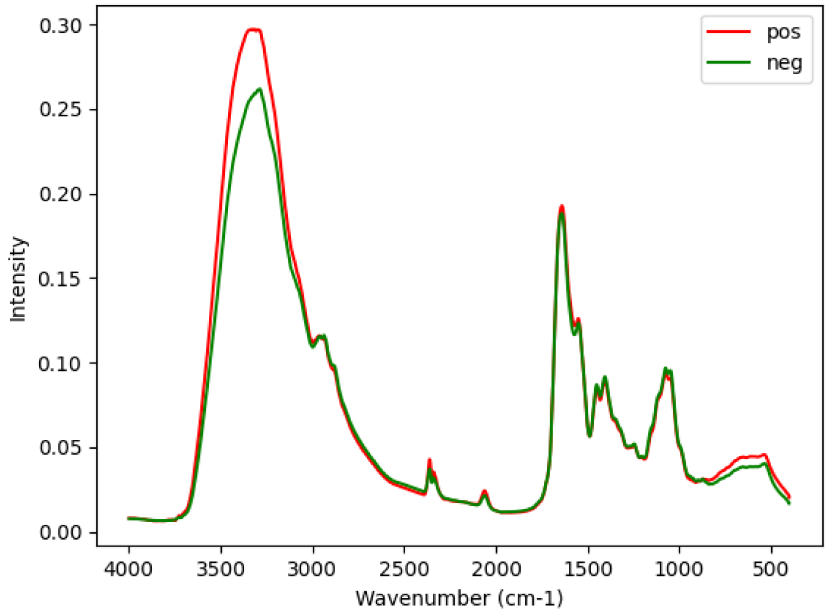
Os dados utilizados neste estudo foram coletados com a aprovação do Comitê de Ética em Pesquisa da Universidade Federal de Uberlândia (UFU), sob o protocolo nž 249.200.9, no Hospital de Clínicas da UFU. Para o diagnóstico dos pacientes pertencentes ao grupo com câncer oral, utilizou-se a classificação TNM de tumores malignos, proposta pela União Internacional para o Controle do Câncer (UICC). Já os pacientes do grupo controle foram selecionados de forma a manter correspondência de idade e sexo em relação ao grupo-alvo, além de apresentarem histórico negativo para outros tipos de câncer.

O conjunto de dados é composto por 65 instâncias de espectros, cada uma representando um paciente diferente. Tais espectros obtidos por ATR-FTIR caracterizam os componentes e estruturas moleculares presentes em amostras de saliva (por exemplo, proteínas, lipídios, entre outros), descritos em termos de 1868 bandas no infravermelho (atributos), com números de onda variando de $4000cm^{-1}$ a $400cm^{-1}$.

Dentro do conjunto, temos 26 instâncias pertencentes ao grupo controle (negativo) e 39 instâncias referentes ao grupo com câncer oral (positivo). As amostras individuais estão ilustradas na Figura 18(a), a qual evidencia o desafio do problema, principalmente devido ao elevado grau de sobreposição entre os grupos. Além disso, a média espectral de cada grupo é apresentada na Figura 18(b), oferecendo uma visão comparativa adicional.



(a) Amostras ATR-FTIR



(b) Espectros ATR-FTIR calculados em média por grupo

Figura 18 – Amostras de ATR-FTIR de Câncer Oral (pos) e amostras de controle (neg).

Com relação à preparação e às etapas de pré-processamento dos dados, nesta investigação procedeu-se da seguinte forma:

1. Correção da linha de base dos espectros utilizando os métodos asymmetric least squares(BOELENS et al., 2004) e rubber-band(BAEK et al., 2015);

2. Normalização dos espectros pelo pico da amida I (faixa espectral no infravermelho

entre 1660cm^{-1} e 1630cm^{-1}), o que evita a influência de sinais indesejados entre as amostras (MORAIS et al., 2019);

3. Truncamento dos espectros para a região entre 1800cm^{-1} e 900cm^{-1}, com o objetivo de evitar ruídos e valores discrepantes (outliers) (BAKER et al., 2014).

4.2 Resultados do VisG2 na base de dados Câncer Oral

A Tabela 1 apresenta os resultados obtidos pelo classificador de HL equipado com o VisG2. Foram adotadas as medidas avaliativas convencionais para problemas relacionados à detecção de doenças AA, SE, SP e F1 sendo o F1 a métrica principal para a avaliação do modelo.

Observa-se que o melhor desempenho foi obtido pelas medidas de rede BET e ASS, ambas alcançando 69% em termos de F1. Essas métricas apresentaram resultados equilibrados de SE e SP, com valores aproximados de 64% e 75%, respectivamente, resultando em uma AA global de aproximadamente 71%. Além disso, as medidas de CCF, ADG e ASP também exibiram resultados consistentes e próximos ao desempenho das melhores métricas de rede, demonstrando relevância para a análise. Em contrapartida, a medida de CLO apresentou desempenho insatisfatório, não oferecendo contribuições significativas para a interpretação dos resultados.

Em conclusão, os resultados evidenciam que a escolha das medidas de rede exerce impacto direto no desempenho do classificador de HL. Nesse cenário, as métricas BET e ASS destacam-se como as mais promissoras para a tarefa de detecção de câncer oral, ao passo que a CLO se mostrou pouco informativa. Esses achados reforçam a importância da seleção criteriosa de atributos de rede para potencializar a eficácia do VisG2 em aplicações biomédicas. A medida de rede CLO apresentou um desempenho ruim, não contribuindo para alguma análise mais profunda de desempenho.

Tabela 1 – Resultados de classificação de alto nível para a base de dados do Câncer Oral em termos de Acurácia (AA), Sensibilidade (EP), Especificidade (SP) e F1-Score (F1).

Algoritmo	Medidas de Rede	AA	SE	SP	F1
Alto Nível (VisG2)	ASS	0.70±0.09	**0.64±0.08**	0.74±0.10	**0.69±0.07**
	CCF	0.67±0.07	0.63±0.06	0.70±0.08	0.66±0.07
	ADG	0.69±0.08	0.63±0.07	0.74±0.09	0.68±0.08
	BET	**0.71±0.06**	**0.64±0.05**	0.75±0.07	**0.69±0.08**
	ASP	0.69±0.08	**0.64±0.07**	0.73±0.09	0.68±0.08
	CLO	0.63±0.12	0.35±0.06	**0.82±0.15**	0.49±0.10

Apresentamos aqui a comparação da nossa técnica de HL com diversos algoritmos de AS. Nossa seleção inclui técnicas amplamente utilizadas em dados ATR-FTIR, como a LDA, e métodos de ponta, como a SVM com kernel gaussiano (RBF). Também consideramos técnicas tradicionais de AM, como o NB e o MLP, além da técnica de aprendizado profundo, como a CNN e o Transformer (Encoder), atualmente consideradas estado da arte. Por fim, também comparamos os resultados com a classificação de HL equipada com a rede mútua de KNN, apresentados no trabalho (LIMA FILHO et al., 2024). Com exceção do LDA e do NB, todas as demais técnicas tiveram seus parâmetros ajustados de acordo com diferentes configurações de hiperparâmetros, conforme descrito em (CARNEIRO et al., 2023).

A Tabela 2 apresenta os resultados preditivos obtidos por cada uma das técnicas em comparação. Nota-se que os melhores desempenhos foram alcançados pela nossa técnica de alto nível VisG2, que superou inclusive o método de alto nível MKNNG, considerado até então o mais eficiente. O diferencial do VisG2 está em sua concepção voltada para dados sequenciais, o que possibilita um mapeamento mais preciso das informações em forma de rede, refletindo em ganhos significativos de desempenho.

No que se refere às técnicas comumente aplicadas na literatura de ATR-FTIR, o LDA apresentou os piores resultados, enquanto o SVM-RBF e o Encoder demonstrou sua robustez ao obter resultados competitivos em relação à nossa classificação de HL baseada na BET e ASS. No entanto, foram superadas em três das quatro métricas preditivas pela nossa classificação de alto nível VisG2, que também superou a CNN, uma técnica de aprendizado profundo considerada estado da arte, em quase todas as métricas avaliadas.

Tabela 2 – Resultados de classificação de alto nível para a base de dados do Câncer Oral em termos de Acurácia (AA), Sensibilidade (SE), Especificidade (SP) e F1-Score (F1) em comparação com técnicas tradicionais e de última geração.

Algoritmo	Medida de Rede	AA	SE	SP	F1
LDA	-	0.53±0.12	0.58±0.10	0.46±0.15	0.51±0.08
NB	-	0.64±0.09	0.58±0.11	**0.80±0.07**	0.67±0.06
KNN	-	0.59±0.10	0.68±0.08	0.48±0.13	0.56±0.09
MLP	-	0.52±0.13	0.41±0.14	**0.80±0.06**	0.54±0.11
CNN	-	0.65±0.08	0.69±0.07	0.60±0.10	0.64±0.05
SVM-RBF	-	0.68±0.07	0.78±0.06	0.54±0.12	0.64±0.07
Transformer	-	0.69±0.06	0.75±0.07	0.60±0.06	0.66±0.06
HL (MKNNG)	CCF	**0.71±0.05**	**0.81±0.04**	0.57±0.09	0.67±0.05
HL (VisG2)	BET	**0.71±0.06**	0.64±0.05	0.75±0.07	**0.69±0.08**
HL (VisG2)	ASS	0.70±0.09	0.64±0.08	0.74±0.10	**0.69±0.07**

4.3 Considerações finais

No Capítulo 4, foram apresentados os experimentos envolvendo o método de construção de redes VisG2, proposto neste trabalho para o problema de detecção do Câncer Oral. Inicialmente, na Seção 4.1, foram descritas as características da base de dados utilizada, bem como as etapas de pré-processamento aplicadas aos sinais ATR-FTIR. Em seguida, na Seção 4.2, foram conduzidos os experimentos com o classificador de HL base equipado com o VisG2, acompanhados de uma análise detalhada do melhor desempenho obtido pelo modelo. Ainda na Seção 4.2, foi realizada uma comparação entre o VisG2 e outra abordagem de construção de redes, a MKNNG, que até então apresentava o melhor desempenho reportado para o problema em estudo, mas que foi superada pela técnica proposta. No mesmo contexto comparativo, o desempenho do classificador de HL com VisG2 também foi analisado em relação a outros classificadores tradicionais amplamente utilizados na literatura. Vale ressaltar que parte desses experimentos foi originalmente publicada em (LIMA FILHO et al., 2024), sendo posteriormente adaptada neste trabalho para considerar uma nova medida objetivo (F1), além da inclusão dos resultados obtidos com o modelo Transformer (Encoder).

Capítulo 5

Experimentos TEA e Análise dos Resultados

Este capítulo tem como objetivo apresentar os experimentos realizados e a análise exploratória conduzida para avaliar o desempenho da abordagem proposta. Inicialmente, investigamos o impacto da variação dos hiperparâmetros e medidas de rede envolvidos no classificador de HL, visando otimizar sua capacidade discriminativa. Em seguida, são apresentados os resultados obtidos com a classificação híbrida, que integra classificadores de LL (tradicionais com SVM) e classificadores de HL (baseados em redes complexas extraídas a partir dos dados).

A fim de validar a eficácia da abordagem híbrida, os resultados obtidos são comparados com aqueles produzidos por diferentes classificadores convencionais amplamente utilizados na literatura, como SVM, RF e KNN. Também é avaliada a contribuição da técnica VisG2, proposta para a construção de redes a partir dos dados, por meio da comparação com outras técnicas de geração de grafos comumente aplicadas no contexto de classificação de HL. Por fim, realiza-se uma investigação sobre grafos de visibilidade com similaridades alta, média e baixa, gerados por meio da técnica VisG2.

5.1 Base de dados TEA e preparação de dados

O conjunto de dados utilizado neste estudo é composto por 159 instâncias espectrais, obtidas com a devida autorização do Comitê de Ética em Pesquisa, garantindo conformidade com os padrões éticos para coleta de amostras biológicas. Esses espectros foram adquiridos por meio da técnica de ATR-FTIR, a qual permite a caracterização detalhada dos componentes moleculares presentes nas amostras de saliva como proteínas, lipídios, ácidos nucleicos e outros metabólitos.

Cada espectro contém 1867 atributos, que correspondem às intensidades de absorção em diferentes números de onda, cobrindo a faixa de $4.000cm^{-1}$ a $400cm^{-1}$. Essa faixa espectral contempla regiões fundamentais para a identificação de grupos funcionais pre-

sentes na amostra, fornecendo uma representação abrangente das assinaturas químicas de interesse.

Em termos de composição, o conjunto de dados é dividido em dois grupos distintos: grupo controle (negativo) e grupo com TEA (positivo). O grupo controle é composto por 102 instâncias espectrais, provenientes de 34 indivíduos, enquanto o grupo TEA conta com 57 instâncias, coletadas de 19 pacientes diagnosticados clinicamente. Cada paciente contribuiu com três amostras distintas, a fim de aumentar a robustez estatística e a variabilidade natural do conjunto.

As amostras individuais de ambos os grupos são visualmente representadas na Figura 19(a), onde é possível observar a alta sobreposição espectral entre os grupos, evidenciando o desafio de discriminar padrões sutis que separam os indivíduos com TEA dos controles apenas com base em seus espectros. Para facilitar a análise comparativa, a média espectral de cada grupo também foi calculada e apresentada na Figura 19(b), permitindo uma visão global das principais diferenças e semelhanças nas assinaturas infravermelhas médias dos dois grupos analisados.

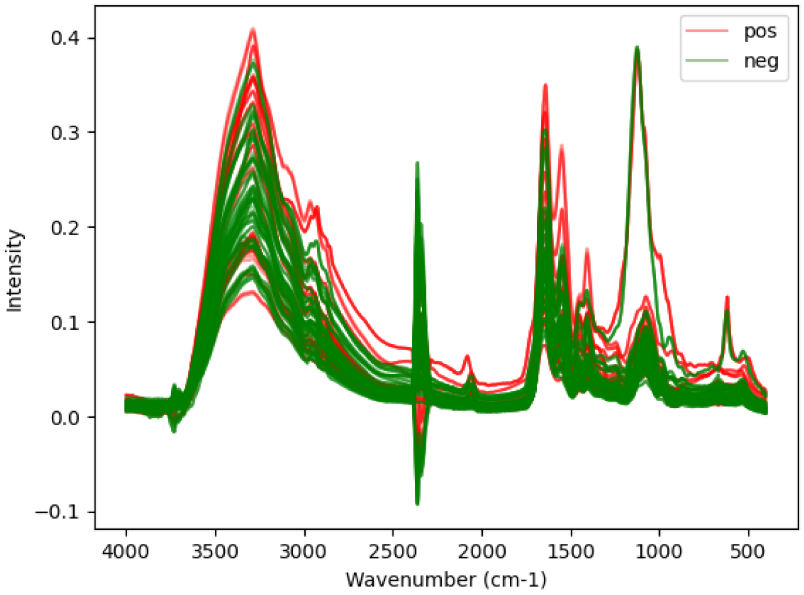
Em relação ao pré-processamento dos dados espectrais, foram aplicadas duas técnicas reconhecidamente eficazes no contexto da espectroscopia de biofluidos, que contribuíram significativamente para a capacidade do modelo de AM em distinguir corretamente entre as classes analisadas. As técnicas empregadas foram: (i) a normalização pelo pico da Amida I localizado na região entre $1660cm^{-1}$ e $1630cm^{-1}$ e (ii) o truncamento dos espectros para a faixa espectral entre $1800cm^{-1}$ e $900cm^{-1}$. A escolha por essas abordagens foi motivada por considerações físico-químicas e metodológicas:

Em relação ao pico da Amida I:

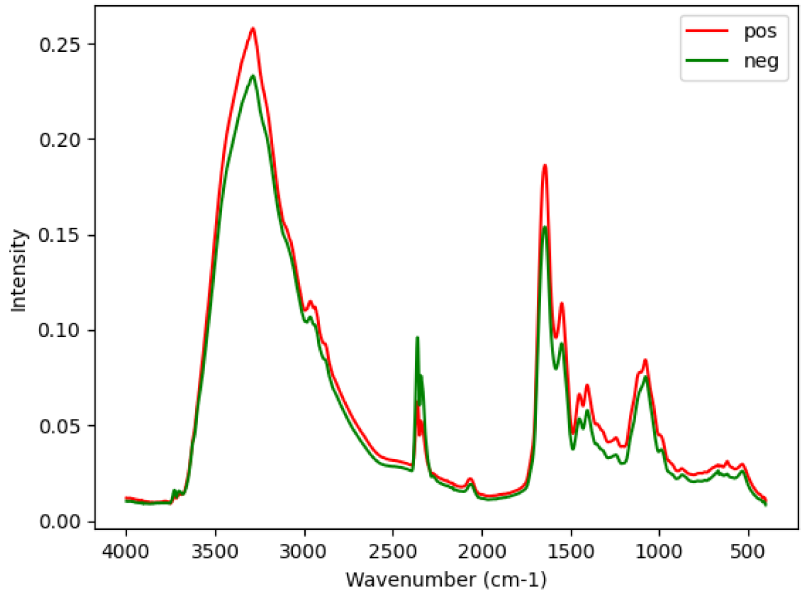
- **Estabilidade e confiabilidade do pico da Amida I:** O pico da Amida I, que geralmente aparece na região de $1660cm^{-1}$ a $1630cm^{-1}$, está relacionado principalmente às ligações $C = O$ de proteínas e é um dos picos mais proeminentes e consistentes nos espectros de amostras biológicas, servindo como uma excelente referência interna para normalização.

- **Correção de variações não-biológicas:** Durante a aquisição dos espectros por ATR-FTIR, podem ocorrer variações instrumentais ou de preparação da amostra. Essas variações introduzem mudanças de escala nos espectros que não estão relacionadas à composição química real da amostra. A normalização pelo pico da Amida I reduz esse problema ao ajustar todos os espectros a uma mesma escala relativa, tornando-os comparáveis.

- **Preservação da informação relativa:** A normalização pelo pico da Amida I preserva as relações entre bandas espectrais dentro de cada espectro. Isso é crucial em aplicações de AM, onde a proporção relativa entre bandas carrega grande parte da informação discriminativa.



(a) Amostras ATR-FTIR



(b) Espectros ATR-FTIR calculados em média por grupo

Figura 19 – Amostras de ATR-FTIR de TEA (pos) e amostras de controle (neg).

❑ **Consistência ao longo de sequências:** Lidando com dados sequenciais (ex: triplicatas para cada paciente), é essencial garantir que as diferenças entre réplicas sejam atribuídas ao fenômeno biológico, e não a variações técnicas. A normalização pela Amida I ajuda a equalizar os espectros das três réplicas, garantindo que o modelo de ML aprenda características reais do sinal bioquímico.

- **Suporte na literatura científica:** A normalização pela banda da Amida I é uma técnica consolidada na literatura espectroscópica, especialmente em estudos de biofluídos (como saliva, sangue, urina). É recomendada em diversas pesquisas para melhorar a reprodutibilidade dos resultados e aumentar a sensibilidade de classificadores ou métodos estatísticos.

Em relação ao truncamento da região espectral:

- **Redução de ruído:** Partes irrelevantes ou ruidosas da série (ex.: início/fim com baixa informação) podem ser descartadas, ajudando o modelo a focar nas regiões mais informativas.
- **Redução da dimensionalidade:** Menos atributos resulta em menor custo computacional, menor risco de overfitting e mais interpretabilidade.
- **Normalização da forma dos dados:** Em problemas com entradas de diferentes comprimentos, o truncamento pode padronizar a dimensão da entrada para que todos os exemplos tenham o mesmo número de atributos.
- **Foco em regiões específicas de interesse:** Em dados espectrais, como FTIR, pode-se truncar para manter apenas regiões onde ocorrem bandas significativas (como Amida I ou II). Isso melhora o desempenho do classificador.
- **Melhoria da generalização:** Ao remover informações redundantes ou irrelevantes, o modelo pode aprender padrões mais generalizáveis.

5.2 Análise exploratória do classificador de alto nível equipado com o VisG2 e seus hiperparâmetros

Esta seção tem como objetivo conduzir uma análise exploratória aprofundada do classificador de HL aliado ao método de construção de redes proposto neste trabalho, denominado VisG2. Essa análise busca compreender o comportamento e o desempenho da técnica a partir da variação de seus principais hiperparâmetros.

A combinação entre o classificador de HL e a técnica VisG2 resulta em um modelo mais adequado para o problema de classificação dos dados ATR-FTIR, uma vez que explora propriedades topológicas extraídas de redes complexas construídas a partir de séries sequenciais. Esse aspecto é especialmente relevante, pois os grafos de visibilidade utilizados pelo VisG2 preservam a estrutura sequencial dos dados, algo que técnicas tradicionais de construção de redes como as baseadas em vizinhos mais próximos (KNNG) ou em vizinhança de raio-ϵ tendem a ignorar. Assim, o VisG2 representa uma abordagem mais coerente com a natureza dos dados espectrais analisados, potencializando a extração de padrões mais representativos.

Para investigar o impacto de diferentes configurações no desempenho do modelo, três hiperparâmetros principais foram sistematicamente avaliados: (i) as medidas de rede utilizadas para caracterizar os super nós, (ii) os valores de k-top similaridades empregados na formação das conexões entre esses super nós, e (iii) a quantidade máxima de conexões permitidas para cada vértice nos grafos de visibilidade construídos a partir das séries temporais originais.

No contexto da técnica de HL combinada com o VisG2, foram conduzidos experimentos variando três hiperparâmetros distintos que influenciam diretamente a construção da rede e, consequentemente, o desempenho da classificação.

5.2.1 Medidas de rede

O primeiro parâmetro analisado refere-se às medidas topológicas utilizadas na representação dos super nós. Ao todo, foram consideradas seis medidas de rede amplamente reconhecidas na literatura de redes complexas, todas previamente apresentadas na Seção 2 de Fundamentação Teórica deste trabalho.

Cada medida captura diferentes aspectos estruturais da rede, permitindo que o classificador de HL explore diferentes perspectivas topológicas dos dados. Essa diversidade torna o modelo mais robusto e adaptável a diferentes tipos de variação nos dados.

5.2.2 Parâmetro top-k de similaridade entre super nós

O segundo hiperparâmetro investigado está relacionado ao número de conexões entre os super nós da rede de HL, definido com base nas k-top similaridades. Esse parâmetro determina o limite máximo de conexões que um super nó poderá estabelecer com os demais, considerando as similaridades mais altas entre suas representações topológicas. Em outras palavras, apenas as k melhores similaridades são utilizadas para formar arestas entre os super nós, o que impacta diretamente a conectividade da rede HL.

Nos experimentos, foram considerados os seguintes valores: top-$k \in \{1, 2, 3, \ldots, 15\}$.

Valores menores de top-k resultam em redes mais esparsas, enquanto valores maiores promovem maior conectividade entre os super nós, o que pode influenciar tanto positivamente quanto negativamente a capacidade discriminativa do modelo, dependendo do domínio dos dados.

5.2.3 Densidade do grafo de visibilidade (w)

O terceiro e último hiperparâmetro analisado é a quantidade máxima de conexões permitidas para cada vértice nos grafos de visibilidade individuais construídos a partir das séries sequenciais. Esse parâmetro é essencial para controlar a densidade do grafo de visibilidade, que serve como base para a extração de características topológicas utilizadas pelo modelo.

A ideia é investigar como o desempenho do classificador se altera quando os grafos de visibilidade são mais esparsos (menos conexões) ou mais densos (mais conexões). Os valores considerados foram: $w \in \{3, 5, 8, 10, 12\}$.

A variação desse parâmetro permite identificar a sensibilidade do modelo à estrutura local das séries temporais, oferecendo insights valiosos sobre a relação entre conectividade da rede e desempenho classificatório.

5.2.4 Análise dos resultados

A Tabela 3 apresenta os resultados obtidos pelo classificador de HL aliado à técnica de construção de redes proposta neste trabalho. As métricas de avaliação consideradas foram AA, SE, SP e F1, sendo esta última adotada como critério principal para a validação do melhor desempenho preditivo, por considerar de forma balanceada os erros de classificação das classes minoritária e majoritária.

Observa-se que o melhor desempenho do classificador foi alcançado utilizando a medida de rede CCF, com valores próximos de 70% em todas as métricas. Esses resultados indicam um desempenho consistente e equilibrado, com destaque para o excelente valor de F1, evidenciando a capacidade dessa medida em capturar padrões topológicos relevantes para a tarefa de classificação. Em contrapartida, a medida de CLO apresentou o pior desempenho entre as avaliadas, sugerindo que não é uma métrica adequada para o problema de detecção de pacientes com TEA com base em dados espectrais ATR-FTIR salivar.

As demais medidas de rede analisadas apresentaram desempenhos intermediários, com boas taxas de SE, porém com quedas perceptíveis na SP. Um destaque especial é a medida de BET, que obteve um F1 superior às demais (exceto ao CCF), mostrando-se competitiva e eficaz para o contexto do estudo.

É importante enfatizar o excelente desempenho do classificador de HL quando associado à medida de CCF. Isso é especialmente relevante considerando a complexidade do problema de classificação do TEA, marcado por forte sobreposição entre os espectros das amostras, conforme discutido na Seção 5.1. Tal característica torna a tarefa de distinguir entre pacientes com e sem o transtorno particularmente desafiadora, reforçando o potencial do uso de propriedades topológicas na identificação de padrões diagnósticos em dados sequenciais complexos.

5.2.5 Análise da variação das top-k similaridades e da quantidade máxima de vizinhos do grafo de visibilidade

Seguimos agora para a análise da variação das top-k similaridades no meta-grafo construído a partir dos grafos de visibilidade, bem como da influência do número máximo

Tabela 3 – Resultados de classificação de alto nível para a base de dados do Autismo em termos de Acurácia (AA), Sensibilidade (EP), Especificidade (SP) e F1-Score (F1).

Algoritmo	Medidas de Rede	AA	SE	SP	F1
Alto Nível (VisG2)	ASS	0.70±0.06	**0.82±0.04**	0.49±0.12	0.62±0.08
	CCF	0.70±0.05	0.70±0.05	**0.71±0.05**	**0.70±0.04**
	ADG	0.69±0.06	0.80±0.05	0.50±0.11	0.62±0.07
	BET	0.70±0.06	0.76±0.05	0.58±0.09	0.66±0.06
	ASP	**0.71±0.05**	0.81±0.04	0.51±0.12	0.63±0.08
	CLO	0.66±0.07	0.76±0.05	0.48±0.13	0.59±0.09

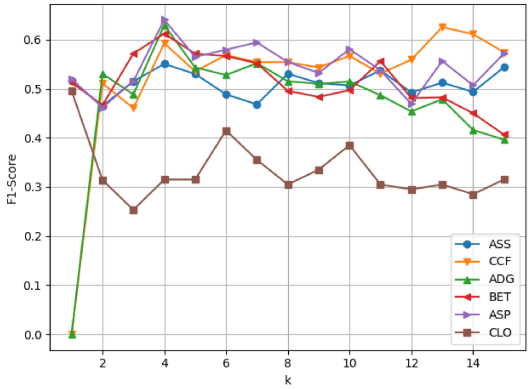
de vizinhos no grafo de visibilidade. Ambas as análises são realizadas considerando as seis medidas de rede descritas anteriormente.

Inicialmente, avaliamos a variação das top-k similaridades. A Figura 20 apresenta cinco gráficos, nos quais cada gráfico representa um valor distinto para o número máximo de vizinhos no grafo de visibilidade. O eixo vertical (y) representa os valores de F1 métrica principal adotada neste trabalho enquanto o eixo horizontal (x) indica os valores de k, ou seja, o número de similaridades consideradas para formar conexões no meta-grafo. As curvas em cada gráfico indicam o desempenho das diferentes medidas de rede utilizadas.

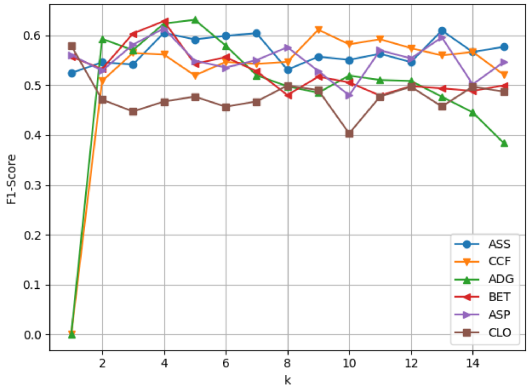
Um aspecto relevante observado é que, à medida que o número máximo de vizinhos no GV aumenta, o desempenho, em termos de F1, tende a diminuir para valores mais altos de top-k similaridades. Essa tendência é verificada na maioria das medidas de rede, com exceção da medida CCF. Para essa medida, o comportamento é inverso: valores mais altos da janela no GV levam a um aumento no desempenho mesmo com top-k maiores.

Outra observação importante refere-se à medida CLO, que apresentou desempenho consistentemente inferior em relação às demais medidas, com piora acentuada para valores mais altos de top-k. Além disso, algumas medidas, como ASS, CCF e ADG, não conseguem gerar resultados satisfatórios para valores muito baixos de top-k. Esse comportamento é explicado pela própria natureza dessas medidas, que dependem de uma conectividade mínima para fornecer informações estruturais úteis, o que não ocorre em redes excessivamente esparsas.

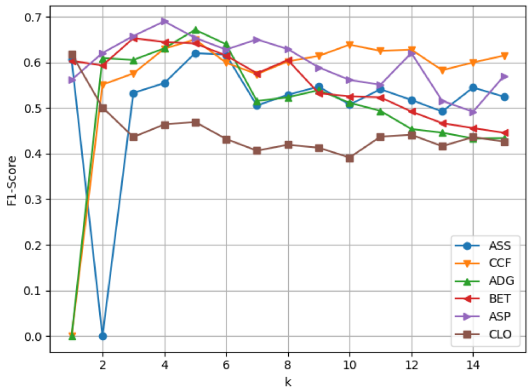
Por fim, observa-se que o classificador de HL, construído com base na técnica proposta VisG2, apresentou melhor desempenho para grafos de visibilidade com janelas maiores entre 8 e 12 vizinhos. Esse resultado indica que, ao analisar dados espectrais do tipo ATR-FTIR, o uso de redes mais densas no GV contribui positivamente para a extração de características topológicas relevantes. Em relação ao parâmetro top-k, verifica-se que medidas como ASS, ADG e CLO obtêm seus melhores desempenhos em valores mais baixos, enquanto a medida CCF se destaca com valores mais altos, mostrando-se mais robusta em cenários com maior número de conexões no meta-grafo.



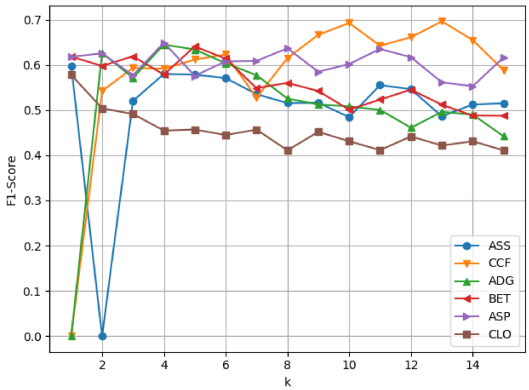
(a) Janela Grafo de Visibilidade 3



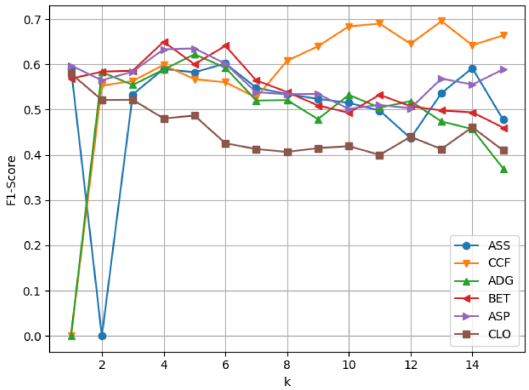
(b) Janela Grafo de Visibilidade 5



(c) Janela Grafo de Visibilidade 8



(d) Janela Grafo de Visibilidade 10



(e) Janela Grafo de Visibilidade 12

Figura 20 – Variação do F1-Score em função das top-k similaridades, para diferentes valores da janela máxima de vizinhos no grafo de visibilidade e para cada medida de rede considerada no estudo.

5.3 Análise exploratória da técnica VisG2 com outros métodos de construção de rede

Dando continuidade aos experimentos, esta seção tem como objetivo comparar o desempenho do método proposto, VisG2, com outras técnicas de construção de redes amplamente utilizadas na literatura. Inicialmente, será analisado apenas o melhor desempenho alcançado pelo classificador de HL quando equipado com diferentes métodos de construção de rede. Em seguida, a análise será aprofundada, considerando a variação do parâmetro top-k para o VisG2 e do número de k-vizinhos mais próximos para as demais abordagens.

Com relação aos parâmetros utilizados, será considerada apenas a quantidade máxima de vizinhos do grafo de visibilidade nos valores $w \in \{3, 5, 8, 10, 12\}$, com base nos melhores resultados observados em experimentos anteriores. Para o parâmetro top-k (ou k-vizinhos mais próximos, conforme a técnica), serão avaliados valores em top-$k \in \{1, 2, \ldots, 15\}$.

Além disso, será considerado, para cada técnica, o melhor desempenho entre as seis medidas de rede empregadas neste estudo, uma vez que o foco da análise é avaliar a eficácia dos diferentes métodos de construção de rede, e não das medidas de rede em si.

5.3.1 Análise dos resultados

A Tabela 4 compara o desempenho de diferentes heurísticas de construção de redes utilizando distintas medidas topológicas. As heurísticas avaliadas incluem KNNG, S-KNNG, M-KNNG, S-MKNNG e VisG2, sendo estas aplicadas com as medidas de rede de ASS, CCF e ASP. As medidas de desempenho consideradas foram AA, SE, SP e F1, com o objetivo de avaliar a eficácia dos modelos de classificação de HL.

Observa-se que a heurística VisG2 equipada com o CCF, apresenta o melhor desempenho geral em todas as métricas analisadas, alcançando 70% em AA, SE e F1, e 71% em SP. Este resultado é notável e está marcado com um asterisco na tabela, sugerindo que a diferença em relação às demais abordagens pode ser estatisticamente significativa. O alto desempenho da técnica proposta evidencia a relevância do CCF na extração de informações, e do VisG2 de mapear os dados de espectroscopia de modo mais eficiente que as demais técnicas.

As heurísticas S-KNNG e S-MKNNG, que também utilizam o CCF, demonstram desempenhos superiores em comparação com suas versões tradicionais KNNG e M-KNNG, respectivamente. Por exemplo, S-KNNG apresenta valores de AA e SE de 67% e 69%, respectivamente, superando KNNG, que possui ambas as métricas em 63%. De forma semelhante, S-MKNNG supera M-KNNG em todas as métricas, sugerindo que técnicas seletivas de construção de redes superam as técnicas mútuas, possivelmente devido às características estruturais das redes geradas: enquanto as seletivas tendem a formar redes mais densas, as mútuas resultam em estruturas mais esparsas.

Em contraste, as heurísticas equipadas com a Assortatividade (KNNG) e Menor Caminho Médio (M-KNNG) apresentam desempenhos mais modestos. Embora o M-KNNG tenha ligeira vantagem em SP (65%), ele não se destaca nas demais métricas. A heurística KNNG, por sua vez, demonstra um desempenho bastante equilibrado (entre 63% e 64%), mas inferior às demais variações aprimoradas.

Em resumo, os resultados sugerem fortemente que o CCF é a medida de rede mais adequada para esta tarefa de classificação, especialmente quando associado a heurísticas avançadas como VisG2. Isso pode estar relacionado à capacidade dessa métrica em capturar a densidade local e o comportamento coletivo entre os nós, aspectos fundamentais para representar relações relevantes entre as instâncias dos dados, e da capacidade do VisG2 em preservar as características sequenciais dos dados ATR-FTIR.

Tabela 4 – Resultados de classificação de alto nível para a base de dados do Autismo em termos de Acurácia (AA), Sensibilidade (SE), Especificidade (SP) e F1-Score (F1) em comparação com diferentes métodos de construção de rede.

Heurísticas	Medidas de Rede	AA	SE	SP	F1
KNNG	ASS	0.63±0.08	0.63±0.07	0.64±0.09	0.64±0.08
S-KNNG	CCF	0.67±0.06	0.69±0.06	0.63±0.08	0.66±0.07
M-KNNG	ASP	0.63±0.07	0.62±0.08	0.65±0.07	0.63±0.07
S-MKNNG	CCF	0.66±0.06	0.67±0.06	0.65±0.07	0.66±0.06
VisG2	CCF	**0.70±0.05**	**0.70±0.05**	**0.71±0.05**	**0.70*±0.04**

5.3.2 Análise da variação de F1-Score em relação ao parâmetro k dos métodos de construção de rede

A análise a seguir foca na variação do parâmetro top-k de similaridade ou, mais especificamente, na quantidade de k vizinhos utilizados na construção das redes, dependendo da técnica empregada. Cada linha no gráfico representa o desempenho de uma técnica de construção de rede em relação ao F1. Os valores reportados correspondem ao melhor desempenho obtido entre as seis medidas de rede consideradas para cada técnica. No caso do VisG2, também é considerado o melhor resultado dentre as diferentes janelas de grafos de visibilidade, conforme discutido na seção anterior.

O primeiro aspecto notável é a queda acentuada no desempenho das técnicas seletivas (S-KNNG e S-MKNNG) à medida que o valor de k aumenta. Isso indica que essas técnicas são mais sensíveis ao acréscimo de vizinhos na rede, resultando em uma degradação significativa do desempenho. Tal comportamento pode estar relacionado à maior densidade introduzida nas redes seletivas com o aumento de k, o que possivelmente reduz sua capacidade discriminativa.

De forma geral, todas as técnicas apresentam desempenho satisfatório para valores baixos de k, com F1 superiores a 60%. Até $k = 4$, os resultados permanecem estáveis e

elevados. A partir desse ponto, observa-se uma leve queda no desempenho das técnicas KNNG, M-KNNG e VisG2, enquanto S-KNNG e S-MKNNG sofrem um declínio mais pronunciado.

Um aspecto interessante é que a técnica básica de vizinhos mais próximos (KNNG), a técnica mútua (M-KNNG) e a proposta baseada em grafos de visibilidade (VisG2) demonstram maior tolerância ao aumento de k. Nessas abordagens, o acréscimo de vizinhos contribui para a densificação da rede, o que, em vez de prejudicar, parece favorecer o desempenho possivelmente por enriquecer a estrutura topológica com conexões informativas adicionais.

Por fim, o VisG2 apresenta o melhor desempenho na maior parte do gráfico, destacando a importância de preservar as características sequenciais dos dados ATR-FTIR na tarefa de detecção do TEA. Embora as técnicas seletivas apresentem bons resultados iniciais, seu desempenho cai substancialmente com o aumento de k, não sendo capazes de superar a técnica proposta. Esse comportamento é esperado, uma vez que tais abordagens priorizam a detecção de padrões estruturais globais, mesmo na presença de ruídos e outliers, o que pode limitar sua capacidade de adaptação à medida que a densidade da rede aumenta. Por outro lado, as técnicas KNNG e M-KNNG demonstram melhora progressiva com redes mais densas; ainda assim, permanecem inferiores ao VisG2, reforçando sua superioridade como uma abordagem robusta e eficaz para o problema em questão.

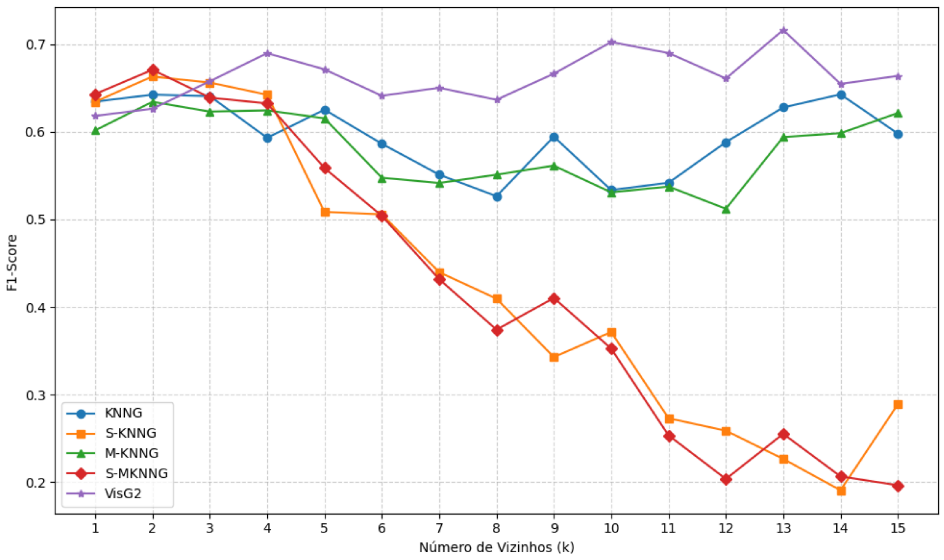


Figura 21 – Variação do F1-Score em função do parâmetro k de vizinhos de cada técnica de construção de rede.

5.4 Análise exploratória da técnica de alto nível equipada com VisG2 em relação a outros classificadores da literatura

Nesta seção, prosseguimos com a análise da técnica proposta, agora em comparação com outras abordagens de classificação de LL. Assim como nas seções anteriores, adotamos um processo de análise estruturado: primeiramente, discutiremos a configuração de cada classificador de LL, garantindo uma base de comparação justa e equilibrada com a técnica de HL. Em seguida, será realizada uma comparação detalhada do desempenho obtido pelo classificador de alto nível (VisG2) em relação às demais abordagens, utilizando métricas padrão de avaliação.

Os classificadores escolhidos para analise foram: como LDA, KNN, NB, MLP, RF, CART, bem como classificadores de última geração na literatura ATR-FTIR, como SVM, CNN e o Transformer (Encoder). A configuração dos classificadores, exceto LDA e NB, foi a seguinte:

- KNN tem dois parâmetros, o valor de k que $k \in \{1, 2, \ldots, 30\}$ e a medida de proximidade que $D_{i,j} \in \{euclidiana, manhattan, cosseno\}$.
- SVM tem dois parâmetros, o kernel $k \in \{RBF, Polynomial\}$ e para o kernel polinomial temos o parâmetro $d \in \{1, 2, \ldots, 15\}$.
- MLP tem dois parâmetros, a taxa de aprendizagem $\alpha \in \{0.01, 0.001, 0.0001\}$ e a camada oculta $n_h \in \{10, 50, 100, 500\}$. ReLU foi escolhida como função de ativação, Adam como o solucionador, e 200 para o número de épocas.
- RF tem dois parâmetros, o número de árvores no conjunto $t \in \{50, 100, 150\}$, e a profundidade máxima de cada árvore $max\_depth \in \{5, \ldots, 15\}$.
- CART tem três parâmetros, a profundidade máxima de cada árvore $max\_depth \in \{5, \ldots, 15\}$, o número mínimo de amostras necessárias para dividir um nó interno $min\_samples\_split \in \{2, \ldots, 5\}$, e o número mínimo de amostras necessárias para estar em um nó folha $min\_samples\_leaf \in \{1, 2, 3\}$. O critério de divisão foi fixado em *entropy*.
- CNN tem três parâmetros, a quantidade de épocas $n_e \in \{10, 50, 100, 200\}$ e a taxa de aprendizagem $\alpha \in \{0.0001, 0.001, 0.01\}$. O dropout foi definido em 0.2 e os filtros eram 16 de entrada e 32 de saída.
- Transformer (Encoder) possui seis hiperparâmetros ajustáveis. O número de camadas do codificador foi definido como $L \in \{1, 2, 3\}$, enquanto o número de cabeças de atenção foi selecionado dentre $H \in \{1, 2, 4\}$. A dimensão da camada feed-forward

foi avaliada em $d_{ff} \in \{64, 128, 256\}$, e a dimensão do espaço de embedding em $d_{emb} \in \{32, 64, 128\}$. A taxa de dropout foi variada no intervalo $[0.0, 0.3]$, enquanto a taxa de aprendizagem foi definida no intervalo $\alpha \in [10^{-4}, 10^{-2}]$, em escala logarítmica.

5.4.1 Análise dos resultados

A Tabela 5 apresenta o desempenho da técnica de classificação de HL em comparação com diversos classificadores tradicionais da literatura. Observa-se que a abordagem de HL supera de forma significativa todas as demais técnicas, incluindo métodos consagrados como o LDA e o SVM, além de classificadores de última geração, como CNN e Transformer (Encoder).

Um aspecto interessante é o desempenho competitivo do KNN, que, apesar de sua simplicidade, alcançou resultados equilibrados e um F1 expressivo de 67%, demonstrando sua eficácia em problemas com características bem definidas de proximidade entre instâncias. No entanto, mesmo esse bom desempenho não foi suficiente para superar a técnica proposta.

Alguns classificadores, como NB, RF e Encoder, apresentaram valores elevados de SE (acima de 77%), o que indica uma boa capacidade de identificar instâncias positivas. No entanto, ambos sofreram uma acentuada queda na SP, principalmente o NB, que apresentou um valor muito baixo nessa métrica. Esse desequilíbrio entre SE e SP comprometeu diretamente o F1, especialmente no caso do NB, cujo F1 foi reduzido para apenas 43%.

Em termos de AA, tanto a técnica de HL quanto o RF atingiram o maior valor, com 70%. No entanto, a técnica de HL se destacou por apresentar também os melhores resultados de F1 (70%) e SP (71%), reforçando sua robustez e equilíbrio na detecção de ambas as classes.

O melhor desempenho da abordagem de HL pode ser atribuído à sua capacidade de incorporar informações estruturais e relacionais por meio da modelagem em grafos, o que permite capturar padrões globais e sequenciais que classificadores tradicionais, baseados apenas em características locais ou vetoriais, não conseguem representar de forma tão eficaz. Essa vantagem se mostra ainda mais relevante em dados espectroscópicos como os de ATR-FTIR, nos quais a preservação da estrutura sequencial e a consideração de dependências entre regiões do espectro são fundamentais para uma classificação precisa.

5.4.2 Análise de redes com difirentes similaridades geradas pelo VisG2

Esta seção é dedicada à análise comparativa entre diferentes redes com distintos níveis de similaridade em relação a um grafo de referência. Esse tipo de análise é essencial para demonstrar a eficácia do algoritmo proposto na identificação de redes com alta, média e

Tabela 5 – Resultados de classificação de alto nível para a base de dados do Autismo em termos de Acurácia (AA), Sensibilidade (SE), Especificidade (SP) e F1-Score (F1) em comparação com técnicas tradicionais e de última geração.

Algoritmo	Medida de Rede	AA	SE	SP	F1
LDA	-	0.57±0.09	0.60±0.08	0.53±0.11	0.56±0.07
NB	-	0.60±0.10	0.77±0.05	0.30±0.15	0.43±0.12
kNN	-	0.67±0.07	0.68±0.06	0.65±0.08	0.67±0.05
SVM	-	0.65±0.08	0.70±0.06	0.58±0.10	0.63±0.07
MLP	-	0.61±0.09	0.69±0.07	0.47±0.12	0.56±0.09
RF	-	**0.70±0.06**	0.77±0.05	0.57±0.09	0.65±0.06
CART	-	0.66±0.07	0.71±0.06	0.56±0.10	0.63±0.07
CNN	-	0.67±0.07	0.75±0.06	0.51±0.11	0.61±0.08
Transformer	-	0.69±0.06	**0.80±0.05**	0.55±0.08	0.65±0.07
HL (VisG2)	CCF	**0.70±0.05**	0.70±0.05	**0.71±0.05**	**0.70±0.04**

baixa similaridade. No contexto do modelo como um todo, as conexões do meta-grafo são estabelecidas apenas com redes de maior similaridade, permitindo uma análise precisa dos aspectos sequenciais dos dados de espectroscopia.

A Figura 22 ilustra três redes: a primeira representa o grafo de referência, enquanto as duas seguintes possuem similaridade de 98% e 97.6% com esse grafo. O grafo de referência possui vértices em azul, enquanto as redes comparadas apresentam vértices em verde-claro. As arestas em vermelho indicam conexões que diferem da estrutura da rede de referência. As Figuras 23 e 24 seguem a mesma lógica de representação: nas redes com similaridade média, os vértices estão em verde; nas de baixa similaridade, em amarelo. Importante destacar que o grafo de referência pode variar entre os diferentes níveis de comparação.

Na Figura 22, observa-se que há poucas arestas em vermelho nas redes comparadas, evidenciando sua alta similaridade com o grafo de referência ou seja, poucas conexões diferem entre as redes. Já na Figura 23, que exibe redes com 76.5% e 75.8% de similaridade, as diferenças tornam-se mais evidentes, com um número significativamente maior de arestas divergentes. Por fim, a Figura 24 apresenta redes com apenas 63% e 61.7% de similaridade, revelando um grande número de arestas vermelhas. Esse padrão indica uma baixa similaridade estrutural com o grafo de referência, refletindo diferenças substanciais nas sequências representadas pelas redes.

5.5 Análise exploratória da classificação de alto nível combinada com classificadores de baixo nível

Nesta seção, analisamos o desempenho do classificador híbrido, no qual a técnica de classificação de HL é integrada ao método proposto neste trabalho, o VisG2. Para

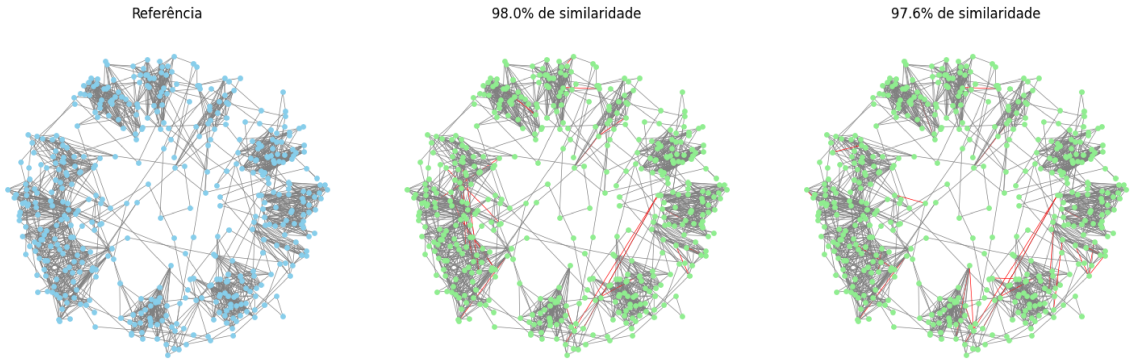


Figura 22 – Redes com alta similaridade.

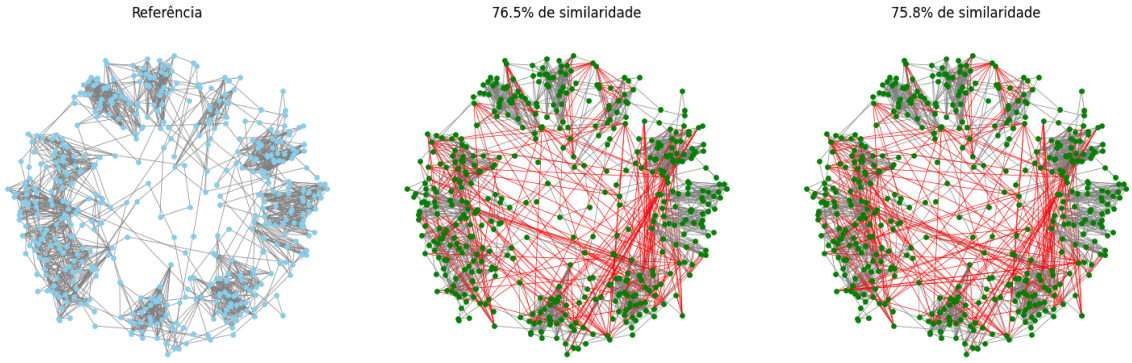


Figura 23 – Redes com média similaridade.

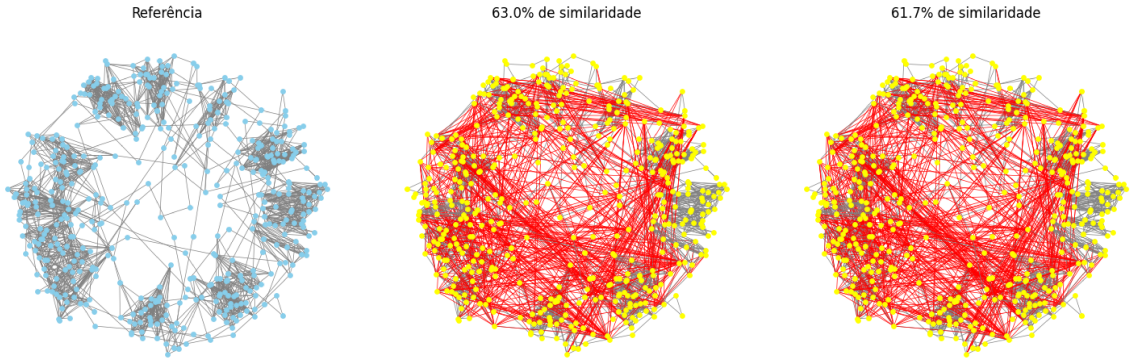


Figura 24 – Redes com baixa similaridade.

compor a parte de classificação de LL do modelo híbrido, foram utilizados os seguintes classificadores: SVM, KNN, LDA, CNN e NB. Esses classificadores foram escolhidos por sua relevância na literatura e diversidade de abordagens (lineares, baseados em instância, probabilísticos e redes neurais profundas).

O parâmetro de combinação entre os níveis, representado por λ, assume os valores $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, onde:

- $\lambda = 0$ indica uso exclusivo do classificador de LL,

- $\lambda = 1$ indica uso exclusivo do classificador de HL.

Além disso, os experimentos consideraram diferentes configurações da janela do grafo de visibilidade, com $w \in \{3, 5, 8, 10, 12\}$, e diferentes quantidades de vizinhos de maior similaridade na construção do metagrafo, com *top-k* $\in \{1, 2, 3, \ldots, 15\}$. Todos os classificadores de LL foram ajustados com suas melhores configurações de hiperparâmetros, conforme detalhado na Seção 5.4.

A Tabela 6 apresenta os resultados obtidos para a classificação híbrida. O primeiro destaque é que o melhor resultado geral foi alcançado pela combinação entre o classificador de HL e a CNN, com $\lambda = 0.8$, ou seja, atribuindo 80% de peso à predição do VisG2. Essa configuração alcançou um F1 de 73%, AA de 72%, SE de 71% e SP de 74%. Embora o ganho em relação ao modelo de HL isolado pareça modesto (cerca de 2 pontos percentuais), ele é expressivo considerando a complexidade da tarefa de classificação em dados espectroscópicos ATR-FTIR.

As demais combinações com os classificadores NB, LDA, KNN e SVM não conseguiram superar o desempenho do classificador de HL individualmente. No entanto, observou-se que o melhor desempenho dessas combinações ocorre consistentemente quando $\lambda = 0.8$, indicando que a maior contribuição do classificador de HL tende a beneficiar a performance global. Esse comportamento corrobora os achados anteriores discutidos na Seção 5.4, em que o VisG2 já demonstrava superioridade em relação aos classificadores de LL isolados.

Essa análise reforça a robustez da abordagem proposta, sugerindo que o VisG2 pode ser utilizado como componente principal em arquiteturas híbridas, com potencial de ganhos adicionais ao ser complementado por classificadores como a CNN, especialmente em cenários com dados complexos e estruturais, como é o caso dos espectros ATR-FTIR.

A Figura 25 apresenta cinco gráficos, cada um correspondente a um valor distinto da janela do grafo de visibilidade $w \in \{3, 5, 8, 10, 12\}$. Em cada gráfico, as curvas representam as combinações da técnica de alto nível (VisG2) com os cinco classificadores de LL utilizados nesta análise: SVM, KNN, LDA, CNN e NB. O eixo horizontal indica os diferentes valores do parâmetro λ, que controla a proporção de influência entre os dois níveis de classificação (de 0 a 1), enquanto o eixo vertical representa o F1 obtido pela classificação híbrida.

De modo geral, é possível observar uma tendência clara em que o desempenho do modelo híbrido melhora consistentemente quando a contribuição do classificador de HL é predominante, especialmente em $\lambda = 0.8$ (ou seja, 80% da predição oriunda do VisG2 e 20% do classificador de LL).

Outro ponto importante revelado pela figura é que os melhores resultados de F1 estão associados à janela 12 do grafo de visibilidade, que também foi a configuração que apresentou o melhor desempenho para o classificador de HL isoladamente. Isso reforça a importância de configurar corretamente os parâmetros da técnica de HL, especialmente o

Tabela 6 – Resultados da classificação híbrida variando o parâmetro λ de influência do classificador de alto e baixo nível. LL Alg. Representa o classificador de LL e HL Grafo, aponta o método de construção de rede (VisG2) em termos de Acurácia (AA), Sensibilidade (SE), Especificidade (SP) e F1-Score (F1).

LL	**HL**	λ	**Dados TEA**			
Alg.	***Grafo***	***Lambda***	***AA***	***SE***	***SP***	***F1***
SVM	VisG2	0	0.65±0.08	0.70±0.06	0.58±0.10	0.63±0.07
SVM	VisG2	0.2	0.65±0.08	**0.72±0.06**	0.53±0.11	0.61±0.08
SVM	VisG2	0.4	0.63±0.09	0.67±0.07	0.57±0.10	0.62±0.08
SVM	VisG2	0.6	0.65±0.07	0.65±0.08	0.66±0.08	0.65±0.06
SVM	VisG2	0.8	**0.70±0.05**	0.69±0.06	**0.72±0.06**	**0.70±0.05**
SVM	VisG2	1	**0.70±0.05**	0.70±0.05	0.71±0.05	**0.70±0.04**
KNN	VisG2	0	0.67±0.07	0.68±0.06	0.65±0.08	0.67±0.05
KNN	VisG2	0.2	**0.70±0.06**	0.73±0.05	0.65±0.08	0.69±0.06
KNN	VisG2	0.4	**0.70±0.06**	**0.74±0.05**	0.65±0.08	0.69±0.06
KNN	VisG2	0.6	**0.70±0.06**	0.73±0.05	0.65±0.08	0.68±0.06
KNN	VisG2	0.8	0.69±0.06	0.66±0.07	**0.74±0.06**	**0.70±0.05**
KNN	VisG2	1	**0.70±0.5**	0.70±0.05	0.71±0.05	**0.70±0.04**
LDA	VisG2	0	0.57±0.09	0.60±0.08	0.53±0.11	0.56±0.07
LDA	VisG2	0.2	0.60±0.08	0.63±0.07	0.54±0.10	0.58±0.07
LDA	VisG2	0.4	0.60±0.08	0.63±0.07	0.54±0.10	0.58±0.07
LDA	VisG2	0.6	0.60±0.08	0.63±0.07	0.54±0.10	0.58±0.07
LDA	VisG2	0.8	0.61±0.07	0.59±0.08	0.64±0.08	0.61±0.06
LDA	VisG2	1	**0.70±0.05**	**0.70±0.05**	**0.71±0.05**	**0.70±0.04**
CNN	VisG2	0	0.67±0.07	0.75±0.06	0.51±0.11	0.61±0.08
CNN	VisG2	0.2	0.71±0.06	**0.77±0.05**	0.61±0.09	0.68±0.06
CNN	VisG2	0.4	0.70±0.06	0.73±0.06	0.63±0.08	0.68±0.06
CNN	VisG2	0.6	0.70±0.06	0.75±0.05	0.63±0.08	0.68±0.06
CNN	VisG2	0.8	**0.72±0.05**	0.71±0.06	**0.74±0.05**	**0.73±0.04**
CNN	VisG2	1	0.70±0.05	0.70±0.05	0.71±0.05	0.70±0.04
NB	VisG2	0	0.60±0.10	0.77±0.05	0.30±0.15	0.43±0.12
NB	VisG2	0.2	0.68±0.07	0.83±0.04	0.40±0.13	0.55±0.09
NB	VisG2	0.4	0.68±0.07	**0.84±0.04**	0.40±0.13	0.55±0.09
NB	VisG2	0.6	0.69±0.06	**0.84±0.04**	0.42±0.12	0.56±0.08
NB	VisG2	0.8	0.68±0.06	0.71±0.06	0.62±0.09	0.66±0.06
NB	VisG2	1	**0.70±0.05**	0.70±0.05	**0.71±0.05**	**0.70±0.04**

tamanho da janela de visibilidade, que influencia diretamente na construção da estrutura topológica dos grafos.

Como já discutido anteriormente, o único cenário em que a combinação híbrida superou o desempenho do VisG2 individual foi com a CNN, utilizando janela 12. Neste caso específico, a combinação com $\lambda = 0.8$ alcançou um F1 de 73%, superando os 70% obtidos pela técnica de HL sozinha. Esse comportamento pode ser observado no gráfico correspondente à janela 12 da Figura 25(e).

Nas demais configurações de janela, mesmo com diferentes classificadores de LL, nenhuma combinação híbrida conseguiu ultrapassar os 70% de F1, evidenciando que a contribuição mais relevante para o desempenho do modelo híbrido provém do classificador de HL especialmente quando operando sob sua melhor configuração estrutural.

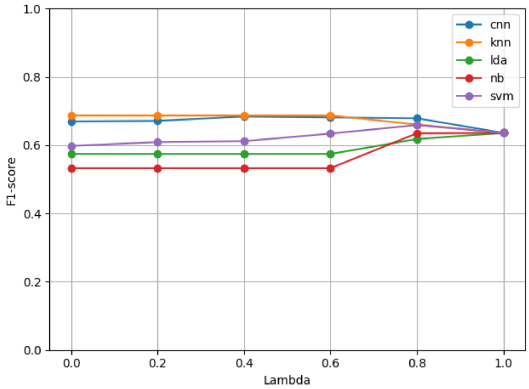
Esses resultados destacam a robustez e a importância do VisG2 na arquitetura híbrida, além de demonstrar que, embora classificadores avançados como a CNN possam contribuir, o nível de abstração topológica proporcionado pela classificação de HL é o fator mais decisivo para o desempenho final do sistema.

5.6 Análise dos atributos mais relevantes da classificação de alto nível VisG2

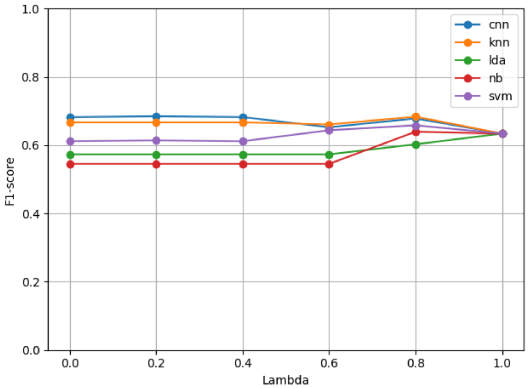
Esta seção é dedicada à análise dos atributos mais relevantes identificados pelo método VisG2 no processo de classificação das amostras salivares. O objetivo é interpretar, sob a perspectiva bioquímica, as regiões espectrais que mais contribuíram para a tomada de decisão do modelo, destacando sua associação com classes de biomoléculas, como proteínas, carboidratos, lipídios e ácidos nucleicos. A interpretação desses atributos permite não apenas compreender o funcionamento interno do classificador, mas também estabelecer correlações entre os padrões espectrais da saliva e potenciais alterações metabólicas ou fisiológicas de interesse clínico.

A análise dos atributos mais relevantes para a classificação com o método VisG2, realizada por meio da explicabilidade com SHAP, apresentado pela Figura 26, evidencia que determinadas regiões espectrais do ATR-FTIR possuem maior impacto na tomada de decisão do modelo. Esses atributos estão diretamente relacionados a componentes bioquímicos fundamentais da saliva, os quais refletem tanto a composição basal quanto possíveis alterações metabólicas associadas a diferentes condições fisiológicas e patológicas.

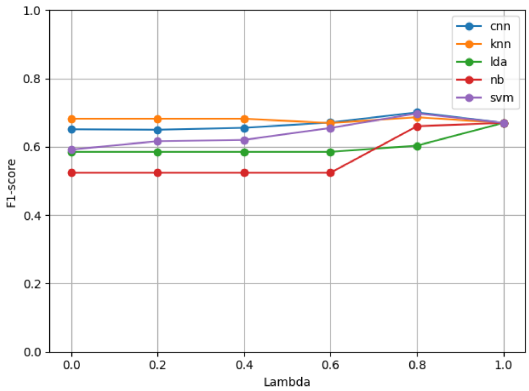
A região em torno de $1768cm^1$ mostrou-se como a de maior contribuição para a classificação. Essa faixa é tipicamente associada ao estiramento da ligação $C = O$ de grupos carbonílicos presentes em lipídios e ésteres. A forte influência dessa região sugere que a variabilidade do perfil lipídico salivar desempenha papel central na discriminação entre as classes analisadas, possivelmente refletindo processos inflamatórios ou estados metabólicos distintos.



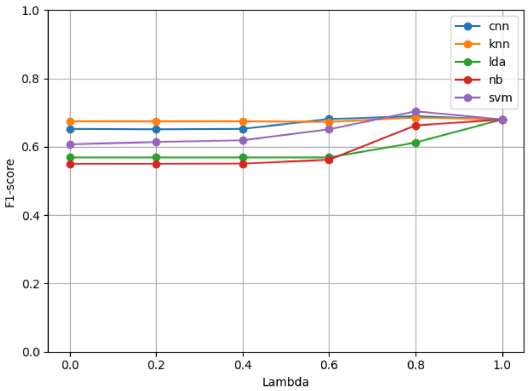
(a) Janela Grafo de Visibilidade 3



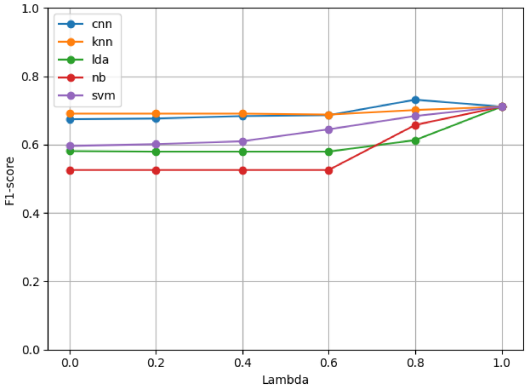
(b) Janela Grafo de Visibilidade 5



(c) Janela Grafo de Visibilidade 8



(d) Janela Grafo de Visibilidade 10



(e) Janela Grafo de Visibilidade 12

Figura 25 – Variação do F1-Score em função do parâmetro λ, para diferentes valores da janela máxima de vizinhos no grafo de visibilidade. Cada linha representa uma combinação com um classificador diferente.

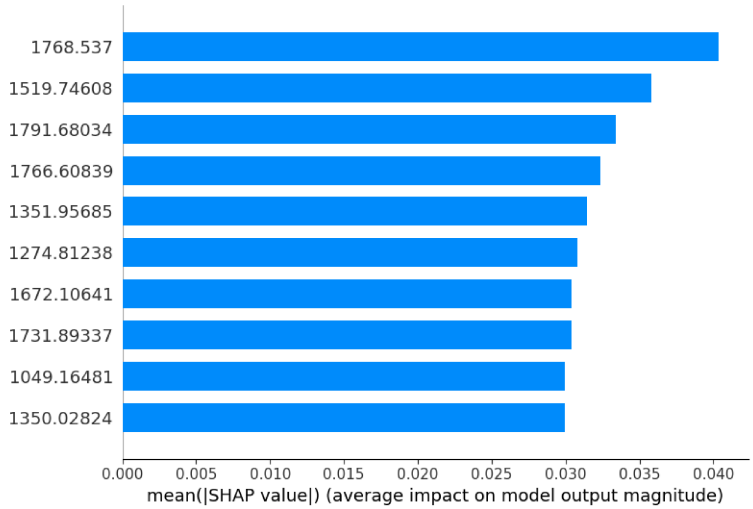


Figura 26 – Análise da explicabilidade da classificação de alto nível VisG2.

Outro atributo de grande relevância está na faixa de $1519cm^1$, relacionada às vibrações características da banda amida II, que envolve o dobramento do grupo $N-H$ e o estiramento da ligação $C-N$ em proteínas. Essa contribuição reforça a importância do conteúdo proteico da saliva, incluindo enzimas, peptídeos antimicrobianos e proteínas estruturais, que podem sofrer alterações de acordo com o estado clínico ou metabólico do indivíduo.

A banda observada em $1672cm^1$, típica da amida I, também se destacou entre os atributos mais influentes. Essa região é amplamente reconhecida como marcador estrutural de proteínas, refletindo conformações secundárias como α-hélices e folhas β. Sua relevância sugere que não apenas a presença, mas também a organização estrutural das proteínas salivares é determinante na classificação realizada pelo VisG2.

No espectro também emergiram picos associados a carboidratos, como a região de $1049cm^1$, relacionada ao estiramento $C-O$ e vibrações glicosídicas, atribuídas a glicoconjugados como mucinas. A presença desse atributo entre os mais relevantes indica que a matriz de carboidratos da saliva, fundamental para a viscosidade e funções protetoras, desempenha papel importante na diferenciação das amostras.

Por fim, é importante destacar a contribuição das bandas em $1274cm^1$ e $1350cm^1$, ligadas a vibrações de grupos fosfato e cadeias laterais de aminoácidos. Essas regiões sugerem que a composição de ácidos nucleicos e a modificação de proteínas também participam do processo discriminativo, refletindo possíveis variações ligadas a processos celulares ou ao microambiente oral.

5.7 Considerações finais

No Capítulo 5, foram apresentados os experimentos com o método de construção de redes VisG2, proposto neste trabalho para o problema de detecção do TEA. Inicialmente,

na Seção 5.1, foram detalhadas as características da base de dados utilizada, juntamente com as etapas de pré-processamento aplicadas. Em seguida, na Seção 5.2, foram conduzidos os experimentos com o classificador de HL base, equipado com o VisG2, além de uma análise sobre os principais hiperparâmetros envolvidos. Na Seção 5.3, foi realizada uma comparação entre o VisG2 e outras abordagens de construção de redes presentes na literatura, com o objetivo de verificar o diferencial da técnica proposta. Já na Seção 5.4, o foco foi comparar o desempenho do classificador de HL com VisG2 frente a outros classificadores tradicionais amplamente utilizados na literatura, bem como um experimento voltado à análise das redes selecionadas pelo algoritmo proposto, considerando estruturas com alta, média e baixa similaridade em relação a uma rede de referência passo essencial para compreender a sensibilidade e a seletividade do modelo. Posteriormente, na Seção 5.5, foi implementada a estratégia de classificação híbrida, combinando classificadores de alto e baixo nível, com o intuito de potencializar o desempenho da predição. Por fim, a Seção 5.6 foi dedicada à análise interpretável do modelo por meio da técnica SHAP, a qual identificou os atributos que mais contribuíram para o processo de classificação. Esta etapa é fundamental para compreender o comportamento do modelo e garantir maior transparência e interpretabilidade nos resultados obtidos.

Capítulo 6

Conclusão

Este trabalho propõe uma nova técnica de construção de redes capaz de explorar os aspectos sequenciais dos dados, o que, no contexto da detecção do TEA por meio de espectroscopia ATR-FTIR salivar, representa uma abordagem promissora frente às técnicas tradicionais baseadas em k-vizinhos mais próximos. A técnica desenvolvida, denominada VisG2, consiste na criação de um meta-grafo a partir de grafos de visibilidade. A construção da rede representa uma etapa fundamental na classificação de HL, mas não é a única: este trabalho também apresenta um classificador híbrido, capaz de combinar os resultados de classificadores de alto e baixo nível, com o objetivo de aprimorar o desempenho preditivo do modelo de HL. A hipótese central deste estudo "Ao combinar um framework de HL baseado em medidas de redes complexas com um algoritmo de LL, a habilidade preditiva e de detecção de padrões será aperfeiçoada" foi confirmada pelos resultados experimentais. De modo geral, os resultados obtidos foram bastante positivos em todas as métricas de avaliação adotadas, demonstrando que a abordagem proposta é promissora para a detecção do TEA a partir de dados ATR-FTIR, especialmente por considerar os aspectos sequenciais inerentes aos sinais analisados.

Em relação aos objetivos específicos, destacam-se as seguintes contribuições:

- **Desenvolver uma técnica de construção de por meio de um meta-grafos de grafos de visibilidade**

  Neste trabalho, foi desenvolvida uma técnica denominada VisG2, que combina grafos de visibilidade gerados a partir de dados sequenciais obtidos por meio de espectroscopia ATR-FTIR, com o objetivo de construir um meta-grafo utilizado na etapa de classificação de HL. Cada conexão entre os vértices do meta-grafo é estabelecida com base na similaridade entre grafos individuais, analisada por meio de suas respectivas matrizes de adjacência. Os resultados obtidos demonstram que nenhuma das técnicas de construção de redes baseadas em k-vizinhos mais próximos foi capaz

de superar o desempenho da VisG2, evidenciando sua eficácia e potencial para o problema proposto.

- **Desenvolver um classificador híbrido capaz de associar características de alto nível produzidas pelas medidas de rede e baixo nível produzidas por classificadores tradicionais da literatura**

  O trabalho também resultou no desenvolvimento de um classificador híbrido, que combina informações de alto e baixo nível com o objetivo de aprimorar o desempenho preditivo da classificação de HL. A associação de HL é baseada em uma técnica de conformidade padrão, a qual utiliza medidas de rede para quantificar a variação estrutural causada pela inserção de um novo elemento na rede representativa de cada classe. A classe cuja rede sofre a menor variação é atribuída como rótulo pela técnica de HL. Os resultados indicaram uma melhoria modesta, porém estatisticamente significativa, demonstrando que a abordagem híbrida é eficaz em reforçar o desempenho da classificação de HL.

- **Demonstrar que a medidas e propriedades de redes complexas podem ser usadas para melhorar o desempenho algoritmos convencionais de classificação**

  Neste trabalho, foram realizados experimentos comparativos entre o classificador de HL e diferentes classificadores de LL. Os resultados demonstraram que o classificador de HL não apenas obteve os melhores desempenhos absolutos, como também apresentou os resultados mais equilibrados em todas as métricas de avaliação adotadas, culminando no maior valor de F1-Score observado. Esses resultados indicam que a técnica proposta para construção de redes é capaz de preservar características sequenciais relevantes dos dados, contribuindo significativamente para a melhoria do desempenho preditivo na tarefa de classificação.

6.1 Limitações

Entre as principais limitações deste trabalho, destaca-se o grande número de hiperparâmetros envolvidos na técnica de classificação de HL. Essa característica exige um processo de otimização extensa, com múltiplas execuções para buscar a combinação ideal de valores, o que resulta em um tempo computacional elevado e maior esforço experimental. Outro fator limitante é a alta complexidade computacional do algoritmo de HL. Essa complexidade se deve, principalmente, à necessidade de construir um grafo de visibilidade para cada espectro ATR-FTIR e realizar diversas comparações entre matrizes de adjacência durante a etapa de treinamento. Esses procedimentos tornam a execução do algoritmo ainda mais demorada, especialmente em conjuntos de dados maiores.

Apesar dessas limitações, elas também indicam oportunidades promissoras para trabalhos futuros. No que se refere à elevada quantidade de hiperparâmetros do modelo, é possível empregar algoritmos de busca e otimização para a seleção automática de valores mais adequados, como Algoritmos Genéticos ou frameworks mais avançados, a exemplo do Optuna. Quanto à eficiência computacional, foram apresentados insights sobre como reduzir a complexidade do método para valores próximos a $O(n \log n)$ na Seção 3.6, o que abre uma linha de pesquisa relevante para tornar a abordagem mais escalável e aplicável a problemas com grandes volumes de dados, ampliando sua utilidade prática e reprodutibilidade científica.

6.2 Trabalhos Futuros

Como trabalhos futuros, pretende-se aprimorar a técnica proposta visando otimizar seu tempo de execução, possibilitando sua aplicação em cenários com maior volume de dados e, consequentemente, com maior complexidade computacional. Essa otimização pode envolver tanto melhorias no algoritmo de construção dos grafos quanto na forma de cálculo das similaridades entre as redes.

Além disso, seria interessante adaptar a abordagem para outros problemas que envolvem séries temporais ou dados sequenciais, como sinais de EEG (eletroencefalograma), espectroscopia NIR (infravermelho próximo), entre outros. A estrutura sequencial desses sinais é compatível com a filosofia dos grafos de visibilidade, permitindo explorar características temporais e padrões recorrentes em diferentes domínios da ciência e da saúde.

Como extensão desta pesquisa, pretende-se explorar o uso de modelos baseados em Transformers de forma complementar ao classificador fundamentado em redes complexas proposto neste trabalho, o qual já considera explicitamente os aspectos sequenciais dos dados ATR-FTIR. Nessa perspectiva, os Transformers podem ser empregados como mecanismos de aprendizado de representações sequenciais de HL, cujos embeddings serviriam como base para a construção das redes, permitindo a incorporação de dependências espectrais complexas e não lineares previamente aprendidas. Alternativamente, os Transformers podem ser utilizados para o aprendizado de medidas de similaridade adaptativas ou integrados em estratégias de combinação de classificadores, formando modelos híbridos que conciliem informações sequenciais e estruturais. Essas abordagens têm potencial para aumentar a capacidade discriminativa, a robustez a ruídos e a interpretabilidade do modelo, ampliando seu alcance e relevância em aplicações biomédicas.

Outra linha promissora seria investigar novas abordagens para a construção do metagrafo, ainda considerando os aspectos sequenciais dos dados, mas com variações nos modelos de grafos de visibilidade. Por exemplo, o uso do grafo de visibilidade horizontal (Horizontal Visibility Graph HVG) pode trazer novas perspectivas e propriedades topológicas que contribuam para uma representação ainda mais robusta dos dados espectrais.

No contexto específico da espectroscopia ATR-FTIR, um ponto importante a ser aprofundado diz respeito à seleção dos atributos espectrais mais relevantes para a tarefa de classificação. Para isso, técnicas baseadas em CNN podem ser exploradas para extrair automaticamente regiões informativas do espectro. Alternativamente, algoritmos genéticos podem ser utilizados para realizar uma seleção evolutiva de características, permitindo reduzir a dimensionalidade dos dados e, ao mesmo tempo, preservar ou até melhorar o desempenho classificatório do modelo.

6.3 Contribuições em Produção Bibliográfica

Este trabalho resultou nas seguintes contribuições em termos de produção bibliográfica:

- Lima Filho, Ricardo B., and Murillo G. Carneiro. "Diagnóstico do câncer oral através da classificação de alto nível." **Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)**. SBC, 2023.
- Lima Filho, Ricardo B., et al. "High-Level Network-based Detection of Oral Cancer from ATR-FTIR Spectroscopy." **2024 International Joint Conference on Neural Networks (IJCNN)**. IEEE, 2024.
- Lima Filho, Ricardo B., Murillo G. Carneiro and Robison S. Silva. "High-Level classification based on meta-graph of visibility graphs for Autism detection". **2025 International Joint Conference on Neural Networks (IJCNN)**. IEEE, 2025.

Referências

ANDRADE, L.; SABINO-SILVA, R.; CARNEIRO, M. Evolutionary neural architecture search for type 2 diabetes mellitus diagnosis from salivary ATR-FTIR spectroscopy. In: **Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2024. p. 459–470. ISSN 2763-8952. <https://doi.org/10.5753/sbcas.2024.2675>.

ANTONIOU, G. et al. Recurrent neural networks for time domain modelling of FTIR spectra: application to brain tumour detection. **Analyst**, Royal Society of Chemistry, v. 148, n. 8, p. 1770–1776, 2023. <https://doi.org/10.1039/D2AN02041F>.

ARAúJO, J. P. de. **Análise da Taxa de Convergência da Regra de Classificação dos k-Vizinhos Mais Próximos**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, Outubro 2018.

ARAúJO, L.; SABINO-SILVA, R.; CARNEIRO, M. High-level classification using complex networks for autism spectrum disorder detection. In: **Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2024. p. 331–341. ISSN 2763-8952. <https://doi.org/10.5753/sbcas.2024.2218>.

ASSOCIATION, A. P. et al. **DSM-5: Manual diagnóstico e estatístico de transtornos mentais**. [S.l.]: Artmed Editora, 2014.

BAEK, S.-J. et al. Baseline correction using asymmetrically reweighted penalized least squares smoothing. **Analyst**, Royal Society of Chemistry, v. 140, n. 1, p. 250–257, 2015. <https://doi.org/10.1039/C4AN01061B>.

BAKER, M. J. et al. Using fourier transform ir spectroscopy to analyze biological materials. **Nature protocols**, Nature Publishing Group, v. 9, n. 8, p. 1771–1791, 2014. <https://doi.org/10.1038/nprot.2014.110>.

BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis-a brief tutorial. **Institute for Signal and information Processing**, Mississippi, v. 18, n. 1998, p. 1–8, 1998.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999. <https://doi.org/10.1126/science.286.5439.509>.

BARAUNA, V. G. et al. Ultrarapid on-site detection of sars-cov-2 infection using simple ATR-FTIR spectroscopy and an analysis algorithm: high sensitivity and specificity. **Analytical Chemistry**, ACS Publications, v. 93, n. 5, p. 2950–2958, 2021. <https://doi.org/10.1021/acs.analchem.0c04608>.

BELLON-MAUREL, V.; MCBRATNEY, A. Near-infrared (nir) and mid-infrared (mir) spectroscopic techniques for assessing the amount of carbon stock in soils–critical review and research perspectives. **Soil Biology and Biochemistry**, Elsevier, v. 43, n. 7, p. 1398–1410, 2011. <https://doi.org/10.1016/j.soilbio.2011.02.019>.

BERTHOMIEU, C.; HIENERWADEL, R. Fourier transform infrared (ftir) spectroscopy. **Photosynthesis research**, Springer, v. 101, p. 157–170, 2009. <https://doi.org/10.1007/s11120-009-9439-x>.

BIEBERLE-HüTTER, A. et al. Operando attenuated total reflection fourier-transform infrared (ATR-FTIR) spectroscopy for water splitting. **Journal of Physics D: Applied Physics**, v. 54, 01 2021. <https://doi.org/10.1088/1361-6463/abd435>.

BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.

BOELENS, H. F. et al. New background correction method for liquid chromatography with diode array detection, infrared spectroscopic detection and raman spectroscopic detection. **Journal of chromatography A**, Elsevier, v. 1057, n. 1-2, p. 21–30, 2004. <https://doi.org/10.1016/j.chroma.2004.09.035>.

BORGWARDT, K. M.; KRIEGEL, H.-P. Shortest-path kernels on graphs. In: IEEE. **Fifth IEEE international conference on data mining (ICDM'05)**. [S.l.], 2005. p. 8–pp.

BRANDES, U. A faster algorithm for betweenness centrality. **The Journal of Mathematical Sociology**, Routledge, v. 25, n. 2, p. 163–177, 2001. <https://doi.org/10.1080/0022250X.2001.9990249>.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.

_____. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>.

BREIMAN, L. et al. **Classification and Regression Trees**. Belmont, CA: Wadsworth International Group, 1984.

BRITO, M. R. et al. Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. **Statistics & Probability Letters**, Elsevier, v. 35, n. 1, p. 33–42, 1997. <https://doi.org/10.1016/S0167-7152(96)00213-1>.

BULLMORE, E.; SPORNS, O. Complex brain networks: graph theoretical analysis of structural and functional systems. **Nature reviews neuroscience**, Nature Publishing Group UK London, v. 10, n. 3, p. 186–198, 2009. <https://doi.org/10.1038/nrn2575>.

CAIXETA, D. C. et al. Salivary molecular spectroscopy: A sustainable, rapid and non-invasive monitoring tool for diabetes mellitus during insulin treatment. **PLoS One**, Public Library of Science San Francisco, CA USA, v. 15, n. 3, p. e0223461, 2020. <https://doi.org/10.1371/journal.pone.0223461>.

_____. Salivary ATR-FTIR spectroscopy coupled with support vector machine classification for screening of type 2 diabetes mellitus. **Diagnostics**, MDPI, v. 13, n. 8, p. 1396, 2023. <https://doi.org/10.3390/diagnostics13081396>.

CARNEIRO, M.; ZHAO, L. Analysis of graph construction methods in supervised data classification. In: IEEE. **2018 7th Brazilian Conference on Intelligent Systems (BRACIS)**. [S.l.], 2018. p. 390–395. <https://doi.org/10.1109/BRACIS.2018.00074>.

CARNEIRO, M. G. **Redes complexas para classificação de dados via conformidade de padrão, caracterização de importância e otimização estrutural**. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação - USP, São Carlos, 2016.

CARNEIRO, M. G. et al. Particle swarm optimization for network-based data classification. **Neural Networks**, Elsevier, v. 110, p. 243–255, 2019. <https://doi.org/10.1016/j.neunet.2018.12.003>.

CARNEIRO, M. G.; GAMA, B. C.; RIBEIRO, O. S. Complex network measures for data classification. In: **2021 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2021. p. 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9533608>.

CARNEIRO, M. G. et al. High-level classification for eeg analysis. In: **International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2023. p. 1–8. <https://doi.org/10.1109/IJCNN54540.2023.10191823>.

_____. Network-based data classification: combining k-associated optimal graphs and high-level prediction. **Journal of the Brazilian Computer Society**, SpringerOpen, v. 20, n. 1, p. 1–14, 2014. <https://doi.org/10.1186/1678-4804-20-14>.

CARNEIRO, M. G.; ZHAO, L. Organizational data classification based on the importance concept of complex networks. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 29, n. 8, p. 3361–3373, 2018. <https://doi.org/10.1109/TNNLS.2017.2726082>.

CASEY, C. M. Far-infrared spectral energy distribution fitting for galaxies near and far. **Monthly Notices of the Royal Astronomical Society**, Blackwell Science Ltd Oxford, UK, v. 425, n. 4, p. 3094–3103, 2012. <https://doi.org/10.1111/j.1365-2966.2012.21455.x>.

COSTA, L. d. F. et al. Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007. <https://doi.org/10.1080/00018730601170527>.

CUPERTINO, T. H.; ZHAO, L.; CARNEIRO, M. G. Network-based supervised data classification by using an heuristic of ease of access. **Neurocomputing**, Elsevier, v. 149, p. 86–92, 2015. <https://doi.org/10.1016/j.neucom.2014.03.071>.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **Proceedings of NAACL-HLT**, 2019.

DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. **International Conference on Learning Representations**, 2021.

FERNANDES, J. M.; OLIVEIRA, G. M. B.; CARNEIRO, M. G. Network optimization based on genetic algorithm for high-level data classification. **IEEE Latin America Transactions**, v. 21, n. 2, p. 295–301, 2023. <https://doi.org/10.1109/TLA.2023.10015222>.

FERNANDES, J. M.; SABINO-SILVA, R.; CARNEIRO, M. G. **Network-Based Detection of Autism Spectrum Disorder Using Sustainable and Non-invasive Salivary Biomarkers**. 2025. Disponível em: <https://arxiv.org/abs/2509.16126>.

FERNANDES, J. M. et al. Data classification via centrality measures of complex networks. In: IEEE. **2023 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2023. p. 1–8. <https://doi.org/10.1109/IJCNN54540.2023.10192048>.

FERREIRA, I. C. et al. Attenuated total reflection-fourier transform infrared (ATR-FTIR) spectroscopy analysis of saliva for breast cancer diagnosis. **Journal of oncology**, Hindawi, v. 2020, 2020. <https://doi.org/10.1155/2020/4343590>.

FREEMAN, L. C. et al. Centrality in social networks: Conceptual clarification. **Social network: critical concepts in sociology. Londres: Routledge**, v. 1, p. 238–263, 2002.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001. <https://doi.org/10.1214/aos/1013203451>.

GAIGEOT, M.-P.; MARTINEZ, M.; VUILLEUMIER, R. Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides. **Molecular Physics**, Taylor & Francis, v. 105, n. 19-22, p. 2857–2878, 2007. <https://doi.org/10.1080/00268970701724974>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.

GRDADOLNIK, J. ATR-FTIR spectroscopy: Its advantage and limitations. **Acta Chimica Slovenica**, Slovenian Chemical Society, v. 49, n. 3, p. 631–642, 2002.

HART, P. E. et al. **Pattern classification**. [S.l.]: Wiley Hoboken, 2000.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009. <https://doi.org/10.1007/978-0-387-84858-7>.

HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. Upper Saddle River, NJ: Prentice Hall, 2009.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359–366, 1989. <https://doi.org/10.1016/0893-6080(89)90020-8>.

JAKKULA, V. Tutorial on support vector machine (svm). **School of EECS, Washington State University**, v. 37, n. 2.5, p. 3, 2006.

JAMES, G. et al. **An Introduction to Statistical Learning: With Applications in R**. New York: Springer, 2013. <https://doi.org/10.1007/978-1-4614-7138-7>.

JR, J. R. B. et al. A nonparametric classification method based on k-associated graphs. **Information Sciences**, Elsevier, v. 181, n. 24, p. 5435–5456, 2011. <https://doi.org/10.1016/j.ins.2011.07.043>.

JUNIOR, A. P. S. et al. **CNN-BiLSTM for sustainable and non-invasive COVID-19 detection via salivary ATR-FTIR spectroscopy**. 2025. Disponível em: <https://arxiv.org/abs/2509.12241>.

KAZEMINEJAD, A.; SOTERO, R. C. Topological properties of resting-state fmri functional networks improve machine learning-based autism classification. **Frontiers in neuroscience**, Frontiers Media SA, v. 12, p. 1018, 2019. <https://doi.org/10.3389/fnins.2018.01018>.

KHALIGHI, S. et al. Evaluating the impact of data pre-processing methods on classification of ATR-FTIR spectra of bituminous binders. **Fuel**, Elsevier, v. 376, p. 132701, 2024. <https://doi.org/10.1016/j.fuel.2024.132701>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2012. v. 25.

LACASA, L. et al. From time series to complex networks: The visibility graph. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 105, n. 13, p. 4972–4975, 2008. <https://doi.org/10.1073/pnas.0709247105>.

LAWRENCE, S.; GILES, C. L. Overfitting and neural networks: conjugate gradient and backpropagation. In: IEEE. **Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium**. [S.l.], 2000. v. 1, p. 114–119. <https://doi.org/10.1109/IJCNN.2000.857823>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998. <https://doi.org/10.1109/5.726791>.

LENG, H. et al. Raman spectroscopy and ftir spectroscopy fusion technology combined with deep learning: A novel cancer prediction method. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, Elsevier, v. 285, p. 121839, 2023. <https://doi.org/10.1016/j.saa.2022.121839>.

LIMA FILHO, R. B. et al. High-level network-based detection of oral cancer from ATR-FTIR spectroscopy. In: IEEE. **2024 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2024. p. 1–8. <https://doi.org/10.1109/IJCNN60899.2024.10650557>.

LIMA FILHO, R. B.; SABINO-SILVA, R.; CARNEIRO, M. G. High-level classification based on meta-graph of visibility graphs for autism detection. In: IEEE. **2025 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2025. p. 1–8. <https://doi.org/10.1109/IJCNN64981.2025.11228714>.

LOUPPE, G. Understanding random forests. **Cornell University Library**, v. 10, 2014.

MARTINEZ-CUAZITL, A. et al. ATR-FTIR spectrum analysis of saliva samples from covid-19 positive patients. **Scientific Reports**, Nature Publishing Group UK London, v. 11, n. 1, p. 19980, 2021. <https://doi.org/10.1038/s41598-021-99529-w>.

MOHAMED, M. A. et al. Fourier transform infrared (ftir) spectroscopy. In: **Membrane characterization**. [S.l.]: Elsevier, 2017. p. 3–29. <https://doi.org/10.1016/B978-0-444-63776-5.00001-2>.

MORAIS, C. L. et al. Standardization of complex biologically derived spectrochemical datasets. **Nature protocols**, Nature Publishing Group UK London, v. 14, n. 5, p. 1546–1577, 2019. <https://doi.org/10.1038/s41596-019-0150-x>.

MOVASAGHI, Z.; REHMAN, S.; REHMAN, D. I. ur. Fourier transform infrared (ftir) spectroscopy of biological tissues. **Applied Spectroscopy Reviews**, Taylor & Francis, v. 43, n. 2, p. 134–179, 2008. <https://doi.org/10.1080/05704920701829043>.

NASEER, K.; ALI, S.; QAZI, J. ATR-FTIR spectroscopy as the future of diagnostics: a systematic review of the approach using bio-fluids. **Applied Spectroscopy Reviews**, Taylor & Francis, v. 56, n. 2, p. 85–97, 2021. <https://doi.org/10.1080/05704928.2020.1738453>.

NEWMAN, M. E. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003. <https://doi.org/10.1137/S003614450342480>.

_____. Network structure from rich but noisy data. **Nature Physics**, Nature Publishing Group UK London, v. 14, n. 6, p. 542–545, 2018. <https://doi.org/10.1038/s41567-018-0076-1>.

NJUKI, S. Funcionalidades do assistente mql5 que você precisa conhecer (parte 04): Análise discriminante linear. 2023. [Online; accessed 23-January-2025].

NOLDUS, R.; MIEGHEM, P. V. Assortativity in complex networks. **Journal of Complex Networks**, Oxford University Press, v. 3, n. 4, p. 507–542, 2015. <https://doi.org/10.1093/comnet/cnv005>.

ORRÙ, P. F. et al. Machine learning approach using mlp and svm algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. **Sustainability**, MDPI, v. 12, n. 11, p. 4776, 2020. <https://doi.org/10.3390/su12114776>.

OZAKI, K. et al. Using the mutual k-nearest neighbor graphs for semi-supervised classification on natural language data. In: **Proceedings of the fifteenth conference on computational natural language learning**. [S.l.: s.n.], 2011. p. 154–162.

PARK, Y.-S.; LEK, S. Artificial neural networks: Multilayer perceptron for ecological modeling. In: **Developments in environmental modelling**. [S.l.]: Elsevier, 2016. v. 28, p. 123–140.

PISNER, D. A.; SCHNYER, D. M. Support vector machine. In: **Machine learning**. [S.l.]: Elsevier, 2020. p. 101–121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>.

PÓSFAI, M.; BARABÁSI, A.-L. **Network science**. [S.l.]: Citeseer, 2016.

QIN, L. et al. New advances in the diagnosis and treatment of autism spectrum disorders. **European Journal of Medical Research**, Springer, v. 29, n. 1, p. 322, 2024. <https://doi.org/10.1186/s40001-024-01916-2>.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, n. 1, p. 81–106, 1986. <https://doi.org/10.1023/A:1022643204877>.

RADFORD, A. et al. Improving language understanding by generative pre-training. 2018.

RISH, I. et al. An empirical study of the naive bayes classifier. In: SEATTLE, WA, USA;. **IJCAI 2001 workshop on empirical methods in artificial intelligence**. [S.l.], 2001. v. 3, n. 22, p. 41–46.

RODRIGUES, R. P. et al. Differential molecular signature of human saliva using ATR-FTIR spectroscopy for chronic kidney disease diagnosis. **Brazilian dental journal**, SciELO Brasil, v. 30, n. 5, p. 437–445, 2019. <https://doi.org/10.1590/0103-6440201902228>.

ROHMAN, A. et al. The use of FTIR and raman spectroscopy in combination with chemometrics for analysis of biomolecules in biomedical fluids: A review. **Biomedical Spectroscopy and Imaging**, v. 8, p. 1–17, 01 2020. <https://doi.org/10.3233/BSI-200189>.

ROKACH, L.; MAIMON, O. **Data Mining with Decision Trees: Theory and Applications**. Singapore: World Scientific, 2015.

RUBINOV, M.; SPORNS, O. Complex network measures of brain connectivity: uses and interpretations. **Neuroimage**, Elsevier, v. 52, n. 3, p. 1059–1069, 2010. <https://doi.org/10.1016/j.neuroimage.2009.10.003>.

RUDIE, J. D. et al. Altered functional and structural brain network organization in autism. **NeuroImage: clinical**, Elsevier, v. 2, p. 79–94, 2013. <https://doi.org/10.1016/j.nicl.2012.11.006>.

SHAW, K. A. Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 yearsautism and developmental disabilities monitoring network, 16 sites, united states, 2022. **MMWR. Surveillance Summaries**, v. 74, 2025.

SILVA, S. F. d. P. et al. Avaliação de biomarcadores salivares para diagnóstico de transtorno de espectro autista por espectroscopia ATR-FTIR. Universidade Federal de Uberlândia, 2020.

SILVA, T. C.; ZHAO, L. Network-based high level data classification. **IEEE Transactions on Neural Networks and Learning Systems**, v. 23, n. 6, p. 954970, 2012. <https://doi.org/10.1109/TNNLS.2012.2195027>.

SPORNS, O. **Networks of the Brain**. [S.l.]: MIT press, 2016.

Sánchez A, V. Advanced support vector machines and kernel methods. **Neurocomputing**, Elsevier, v. 55, n. 1-2, p. 5–20, 2003. <https://doi.org/10.1016/S0925-2312(03)00373-4>.

TAY, Y. et al. Efficient transformers: A survey. **ACM Computing Surveys**, 2020.

TINTOR, Đ. et al. Gaining insight into protein structure via ATR-FTIR spectroscopy. **Vibrational Spectroscopy**, Elsevier, v. 134, p. 103726, 2024. <https://doi.org/10.1016/j.vibspec.2024.103726>.

VASWANI, A. et al. Attention is all you need. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2017. v. 30.

VERA, A. M. **Propriedades de redes complexas de telecomunicações**. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, Agosto 2011.

WAN, P.-J.; YI, C.-W. Asymptotic critical transmission radius and critical neighbor number for k-connectivity in wireless ad hoc networks. In: **Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing**. [S.l.: s.n.], 2004. p. 1–8. <https://doi.org/10.1145/989459.989461>.

YE, J.; JANARDAN, R.; LI, Q. Two-dimensional linear discriminant analysis. In: SAUL, L.; WEISS, Y.; BOTTOU, L. (Ed.). **Advances in Neural Information Processing Systems**. MIT Press, 2004. v. 17. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2004/file/86ecfcbc1e9f1ae5ee2d71910877da36-Paper.pdf>.

ZANON, R. B.; BACKES, B.; BOSA, C. A. Identificação dos primeiros sintomas do autismo pelos pais. **Psicologia: teoria e pesquisa**, SciELO Brasil, v. 30, p. 25–33, 2014. <https://doi.org/10.1590/S0102-37722014000100004>.

ZOU, Y. et al. Complex network approaches to nonlinear time series analysis. **Physics Reports**, Elsevier, v. 787, p. 1–97, 2019. <https://doi.org/10.1016/j.physrep.2018.10.005>.