



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Aplicação do método de quantis no estudo e diagnóstico do Transtorno do Espectro Autista através de sinais de EEG

Mariane Vieira Maximo

Botucatu
2023

Mariane Vieira Maximo

Aplicação do método de quantis no estudo e diagnóstico do Transtorno do Espectro Autista através de sinais de EEG

Dissertação de Mestrado apresentada ao Curso de Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Mestre em Biometria.

Orientador: Profa. Dra. Andriana S. L. O. Campanharo

Botucatu
2023

M464a Maximo, Mariane Vieira
Aplicação do método de quantis no estudo e diagnóstico do
Transtorno do Espectro Autista através de sinais de EEG / Mariane
Vieira Maximo. -- Botucatu, 2024
29 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP),
Instituto de Biociências, Botucatu
Orientadora: Andriana Susana Lopes de Oliveira Campanharo

1. Redes complexas, Transtorno do Espectro Autismo, Método de
Quantis, Séries Temporais. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade
Estadual Paulista (UNESP), Instituto de Biociências, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE MARIANE VIEIRA MAXIMO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 20 dias do mês de fevereiro do ano de 2024, às 14:30 horas, no(a) Sala da pós-graduação - Central de Aulas, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de MARIANE VIEIRA MAXIMO, intitulada **Aplicação do método de quantis no estudo e diagnóstico do Transtorno do Espectro Autista através de sinais de EEG**. A Comissão Examinadora foi constituída pelos seguintes membros: Profa. Dra. ANDRIANA SUSANA LOPES DE OLIVEIRA CAMPANHARO (Orientador(a) - Participação Presencial) do(a) Biodiversidade e Bioestatística / Universidade Estadual Paulista "Júlio de Mesquita Filho" - Instituto de Biociências, Prof. Dr. LUIZ EDUARDO GOMES GARCIA BETTING (Participação Presencial) do(a) Neurologia e Psiquiatria / Universidade Estadual Paulista "Júlio de Mesquita Filho", Prof. Dr. JOEL MESA HORMAZA (Participação Presencial) do(a) Departamento de Biofísica e Farmacologia / Universidade Estadual Paulista "Júlio de Mesquita Filho" - Instituto de Biociências. Após a exposição pela mestranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: Aprovado. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Profa. Dra. ANDRIANA SUSANA LOPES DE OLIVEIRA CAMPANHARO

Documento assinado digitalmente
 ANDRIANA SUSANA LOPES DE OLIVEIRA CAMPA
Data: 30/04/2024 10:55:06-0300
Verifique em <https://validar.itu.gov.br>

À minha família...

Agradecimentos

Agradeço a todos que contribuíram para a conclusão deste trabalho e realização de mais um passo, especialmente:

Aos meus pais, Lilian e Milton, a meu irmão Luis Henrique.

A minha orientadora, Dr^a Andriana Susana Lopes de Oliveira Campanharo, pela oportunidade de pesquisa, orientação, paciência e compreensão durante esses anos.

Aos meus colegas do departamento Mario Lucas, Juliana Gualberto e todos os amigos obtidos ao longo destes 7 anos em Botucatu pela paciência e cumplicidade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

*Always choose the path that seems best, even it is the most difficult, future custom will become
easy an pleasant.
Pitágoras*

Resumo

O Transtorno do Espectro Autista (TEA) é uma condição complexa que afeta o neurodesenvolvimento, envolvendo áreas como comportamento social, comunicação e linguagem. Devido à sua variabilidade de sintomas, o diagnóstico do TEA é desafiador e propenso a erros, podendo afetar significativamente a qualidade de vida a curto e longo prazo. A identificação precoce e intervenções terapêuticas eficazes podem melhorar o seu prognóstico. A eletroencefalografia (EEG), uma técnica que registra a atividade elétrica cerebral, é uma ferramenta valiosa para investigar o TEA. Diversas técnicas computacionais de análise de sinais de EEG têm sido exploradas na literatura para detectar automaticamente o TEA, tais como a Entropia, a Transformada de Wavelet, a Análise de Componentes Independentes (ICA), entre outras. No entanto, novos estudos e métodos são necessários para aprofundar a compreensão dos mecanismos subjacentes a essa condição.

Neste estudo, empregamos o método de quantis, uma abordagem inovadora, para analisar séries temporais de EEG de pacientes com TEA e indivíduos neurotípicos de duas bases de dados distintas. Nosso objetivo foi realizar uma comparação entre os resultados obtidos para ambas as bases de dados após o mapeamento. Embora seja uma novidade para o TEA, o método de quantis já demonstrou eficácia na classificação de outras condições, como Epilepsia e Alzheimer. O mapeamento usando o método de quantis, combinado com caracterizadores topológicos, mostrou-se eficaz na classificação do TEA. A utilização de combinações de caracterizadores topológicos em classificadores, como SVM (Support Vector Machine), com três caracterizadores diferentes, também demonstrou eficácia em ambas as bases de dados analisadas neste estudo. Além disso, identificamos duas regiões cerebrais de interesse: o lobo occipital e um eletrodo específico no lobo frontal. A análise da colaboração das ondas cerebrais na identificação do transtorno revelou que as ondas de frequência α apresentaram os melhores resultados na diferenciação dos grupos para ambas as bases de dados, fortalecendo ainda mais a compreensão dos mecanismos subjacentes ao TEA. Em resumo, este estudo demonstra a promissora aplicação de técnicas de análise de séries temporais, como o método de quantis, na investigação do TEA por meio de EEG. Essas abordagens têm o potencial de contribuir significativamente para uma compreensão mais profunda dessa complexa condição e para o desenvolvimento de métodos de diagnóstico mais precisos.

Palavras-chave: Grafos, Séries temporais, Redes Complexas, Transtorno do Espectro Autista, Classificação.

Abstract

Autism Spectrum Disorder (ASD) is a complex condition that affects neurodevelopment, involving areas such as social behavior, communication, and language. Due to its variability of symptoms, diagnosing ASD is challenging and prone to errors, potentially significantly affecting short and long-term quality of life. Early identification and effective therapeutic interventions can improve prognosis. Electroencephalography (EEG), a technique that records brain electrical activity, is a valuable tool for investigating ASD. In research, various computational techniques for analyzing EEG signals have been explored in the literature to automatically detect ASD, such as Entropy, Wavelet Transform, Independent Component Analysis (ICA), among others. However, new studies and methods are needed to deepen understanding of the underlying mechanisms of this condition.

In this study, we employed the quantile method, an innovative approach, to analyze EEG time series from patients with ASD and neurotypical individuals from two distinct databases. Our goal was to perform a comparison between the results obtained. Although novel for ASD, the quantile method has already demonstrated efficacy in classifying other conditions, such as Epilepsy and Alzheimer's disease. Mapping using the quantile method, combined with topological descriptors, proved effective in classifying ASD. The use of combinations of topological descriptors in classifiers, such as SVM (Support Vector Machine), with three different descriptors, also demonstrated efficacy in both databases analyzed in this study. Additionally, we identified two brain regions of interest: the occipital lobe and a specific electrode in the frontal lobe. Analysis of the collaboration of brain waves in identifying the disorder revealed that alpha frequency waves yielded the best results in differentiating groups for both databases, further enhancing understanding of the mechanisms underlying ASD. In summary, this study demonstrates the promising application of time series analysis techniques, such as the quantile method, in ASD investigation through EEG. These approaches have the potential to significantly contribute to a deeper understanding of this complex condition and to the development of more accurate diagnostic methods and more effective therapeutic interventions.

Keywords: Graphs, Time Series, Complex Networks, Autism Spectrum Disorder, Classification.

Lista de figuras

Figura 1 – Representação do TEA e sua natureza heterogênea.	1
Figura 2 – Representação da captação do sinal de EEG e suas subdivisões em ondas δ (0, 5 a 4 Hz) , θ (4 a 8 Hz), α (8 a 13 Hz), β (13 a 30 Hz)e γ (30 a 40 Hz). . .	3
Figura 3 – Ilustração do problema das Sete Pontes. a) Representação esquemática da área com as pontes. b) Representação do problema por um grafo proposto por Euler. Fonte: (BRITANNICA, 2003).	6
Figura 4 – Exemplos de redes complexas com diferentes topologias. A intensidade de cada aresta é dada por sua espessura e a direção da conexão é dada por sua seta. . .	7
Figura 5 – a) Rede não-direcionada e matriz de adjacência correspondente. Rede não-direcionada e matriz de adjacência correspondente; b) Rede direcionada e matriz de adjacência correspondente.	8
Figura 6 – Rede direcionada com peso e matriz de adjacência W correspondente. . . .	9
Figura 7 – Ilustração do mapeamento de uma série temporal em grafos de quantis. . . .	10
Figura 8 – Ilustração da disposição dos 16 eletrodos no escalpo referente à base de dados de King Abdulaziz.	16
Figura 9 – Ilustração da disposição dos 8 eletrodos no escalpo referente à base de dados de Sheffield.	17
Figura 10 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, T3, O1 e F8.	22
Figura 11 – Exemplificação dos sinais do paciente do grupo TEA representados na Figura 10 após aplicação do filtro passa-banda (0,5 a 30 Hz).	22
Figura 12 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo NTEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, T3, O1 e F8.	23
Figura 13 – Exemplificação dos sinais do paciente do grupo NTEA representados na Figura 12 após aplicação do filtro passa-banda (0,5 a 30 Hz).	23
Figura 14 – Espectro de potência dos sinais dos eletrodos Fp1, T3, O1 e F8 dos grupos TEA e NTEA.	24
Figura 15 – Autocorrelação da amostra: a) grupo TEA; b) grupo NTEA.	25
Figura 16 – Gráficos com os valores de variação da média em relação a k para cada caracterizador com cálculo do erro.	26
Figura 17 – Boxplots construídos para os maiores valores atingidos de k na área abaixo da curva ROC.	28
Figura 18 – Ranqueamento dos eletrodos, coloridos de acordo com seu valor abaixo da área de curva ROC para cada um dos caracterizadores utilizados.	29

Figura 19 – Classificação usando SVM com 3 caracterizadores testando todas as combinações possíveis, com Acu de 96,86%.	30
Figura 20 – Exemplos dos sinais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelo eletrodo Fp1 e subdivididos nas ondas δ , θ , α e β	31
Figura 21 – Espectro de potência dos sinais do eletrodo Fp1 para o grupo TEA nas faixas de frequências correspondentes a δ , θ , α e β	31
Figura 22 – Boxplots construídos para cada caracterizador, de cada frequência de onda estudada com base nos seus valores de área abaixo da curva ROC.	32
Figura 23 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, C3, O2 e P10.	35
Figura 24 – Exemplificação dos sinais do paciente do grupo TEA representados na Figura 23 após aplicação do filtro passa-banda (0,5 a 30 Hz).	35
Figura 25 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo NTEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, C3, O2 e P10.	36
Figura 26 – Exemplificação dos sinais do paciente do grupo NTEA representados na Figura 25 após aplicação do filtro passa-banda (0,5 a 30 Hz).	36
Figura 27 – Espectro de potência dos sinais dos eletrodos Fp1, C3, O2 e P10 dos grupos TEA e NTEA.	37
Figura 28 – Autocorrelação da amostra: a) grupo TEA; b) grupo NTEA.	37
Figura 29 – Gráficos com os valores de variação da média em relação a k para cada caracterizador com cálculo do erro.	38
Figura 30 – Boxplots construídos para os maiores valores atingidos de k na área abaixo da curva ROC.	40
Figura 31 – Ranqueamento dos eletrodos, coloridos de acordo com seu valor de Acu para cada um dos caracterizadores utilizados.	42
Figura 32 – Classificação usando SVM com 3 caracterizadores testando todas as combinações possíveis, com Acu de 70,83%.	43
Figura 33 – Exemplos dos sinais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelo eletrodo Fp1 e subdivididos nas ondas δ , θ , α e β	44
Figura 34 – Espectro de potência dos sinais do eletrodo Fp1 para o grupo TEA nas faixas de frequências correspondentes a δ , θ , α e β	44
Figura 35 – Boxplots construídos para cada caracterizador com base nos seus valores de área abaixo da curva ROC.	45

Lista de tabelas

Tabela 1 – Descrição dos níveis do TEA de acordo com suas características e necessidades de suporte.	2
Tabela 2 – Filtros utilizados para gerar as ondas cerebrais e suas frequências correspondentes.	19
Tabela 3 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.	27
Tabela 4 – Maiores valores de áreas abaixo de curva ROC dos caracterizadores para cada eletrodo.	27
Tabela 5 – Tabela de valores das medianas representadas nos boxplots da Figura 17 para os Grupos TEA e NTEA.	28
Tabela 6 – Matriz de confusão acumulada.	30
Tabela 7 – Tabela de valores das medianas representadas nos boxplots da Figura 17 para os Grupos TEA e NTEA.	33
Tabela 8 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.	39
Tabela 9 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.	39
Tabela 10 – Tabela de valores das medianas representadas nos boxplots da Figura 30 para os grupos TEA e NTEA.	41
Tabela 11 – Matriz de confusão acumulada.	43

Sumário

	1 INTRODUÇÃO	1
1.1	Objetivo	4
1.2	Estrutura do texto	5
	2 REDES COMPLEXAS	6
2.1	Contexto histórico	6
2.2	Grafo de quantis	9
2.3	Caracterizadores topológicos	10
2.3.1	Centralidade do autovetor	11
2.3.2	Centralidade de proximidade	11
2.3.3	Coefficiente de agrupamento	12
2.3.4	Comprimento do menor caminho	12
2.3.5	Comprimento do salto médio	13
2.3.6	Energia do grafo	14
2.3.7	Modularidade	14
	3 BASES DE DADOS	15
3.1	Base de dados de King Abdulaziz	15
3.2	Base de dados de Sheffield	17
	4 METODOLOGIA	18
4.1	Pré processamento dos dados	18
4.2	Extração de características	19
4.3	Classificação	19
	5 RESULTADOS	21
5.1	Base de dados de King Abdulaziz	21
5.1.1	Estudo da divisão das ondas cerebrais	30
5.2	Base de dados de Sheffield	34
5.2.1	Estudo da divisão das ondas cerebrais	43
	6 CONCLUSÕES	46
	Referências	48

1 Introdução

O Transtorno do Espectro Autista (TEA) representa uma condição neurológica do desenvolvimento que impacta a forma como uma pessoa interage, comunica-se e percebe o ambiente circundante. Nos últimos 20 anos houve a conceituação de um espectro de transtornos relacionados ao autismo, incluindo Transtorno Desintegrativo da Infância, Transtorno de Asperger, Transtorno Invasivo do Desenvolvimento Sem Outra Especificação (TID-NOS) e síndrome de Rett (FILIPEK et al., 2000), variando em abrangência, gravidade e início. O diagnóstico do TEA é frequentemente estabelecido na infância e seus sintomas podem persistir ao longo da vida (LORD et al., 2018).

As características essenciais do TEA englobam desafios na comunicação verbal e não verbal, déficits na reciprocidade social (como dificuldade na formação de amizades ou compreensão das emoções alheias) e a manifestação de comportamentos repetitivos ou restritivos (tais como interesses intensos e restritos ou a execução de rituais). Contudo, é importante notar que essas características apresentam variações significativas de pessoa para pessoa, conferindo ao TEA uma natureza heterogênea assim como representado na figura a seguir (LORD et al., 2018).



Figura 1 – Representação do TEA e sua natureza heterogênea.

Diversos fatores têm sido associados ao desenvolvimento do TEA, caracterizado por uma etiologia e fenótipo complexos, com sobreposições a doenças genéticas monogênicas, como a

síndrome do X frágil (WANG et al., 2014). Estudos indicam que o TEA não é uma doença única, mas sim uma condição que surge devido a fatores genéticos, com importantes contribuições de fatores ambientais. Os pacientes com o TEA exibem diferenças marcantes na expressão de RNA entre o córtex frontal e temporal em comparação com cérebros típicos, além de vias de sinalização de cálcio e proteínas quinases ativadas por mitógeno (PQAM) sendo consideradas biomarcadores do TEA. Estudos de gêmeos têm sido uma ferramenta valiosa para estimar a herdabilidade do TEA. Esses estudos mostram que a similaridade para o TEA é maior entre gêmeos idênticos (gêmeos monozigóticos) que compartilham todos os seus genes, em comparação com gêmeos não idênticos (gêmeos dizigóticos) que compartilham cerca de 50% de seus genes (COUTINHO; BOSSO, 2016; LAVOR et al., 2021). Essa diferença sugere que a genética desempenha um papel importante na compreensão da doença. As estimativas da herdabilidade do TEA em estudos variam, mas geralmente estão na faixa de 70% a 90%.

Os níveis do TEA podem ser categorizados de diferentes maneiras, no entanto, um sistema comum envolve a descrição de três níveis (PSIQUIATRIA, 2014):

Tabela 1 – Descrição dos níveis do TEA de acordo com suas características e necessidades de suporte.

Nível	Características	Suporte
Nível 1 (Leve)	- Dificuldades sociais e de comunicação, mas menos evidentes do que em níveis mais elevados. - Pode ter dificuldade em iniciar e manter conversas, compreender nuances sociais e expressar empatia. - Funciona relativamente de forma independente em atividades diárias.	Suporte social ou de trabalho pode ser necessário.
Nível 2 (Moderado)	- Dificuldades sociais e de comunicação mais pronunciadas. - Pode ter maior dificuldade em se comunicar verbalmente, expressar necessidades e compreender emoções dos outros.	Requer apoio significativo e intervenção para o funcionamento diário.
Nível 3 (Grave)	- Sintomas mais graves, incluindo déficits significativos na comunicação verbal e não verbal, bem como comportamentos repetitivos e restritos.	Necessita de apoio substancial em todas as áreas de funcionamento.

O processo diagnóstico e o tratamento do TEA (níveis 1, 2 e 3) normalmente envolvem uma equipe multidisciplinar de profissionais da saúde, incluindo psicólogos, terapeutas ocupacionais, fonoaudiólogos e pediatras. O tratamento pode incorporar terapias comportamentais, terapia da fala, terapia ocupacional e, em determinados casos, a administração de medicamentos destinados a abordar sintomas específicos, como hiperatividade, ansiedade e até mesmo epilepsia (TUCHMAN; RAPIN, 2002). Salienta-se a importância de reconhecer que cada indivíduo portador do TEA é único, com suas próprias habilidades e desafios. A implementação de abordagens

personalizadas e um apoio adequado precoces podem contribuir significativamente para que pessoas com o TEA alcancem seu potencial máximo e melhorem sua qualidade de vida (SACREY; BENNETT; ZWAIGENBAUM, 2015).

No diagnóstico do TEA, são usados especificadores para registrar as características individuais, como comprometimento intelectual, linguagem e sintomas autistas. Isso permite diagnósticos mais precisos e individualizados, por exemplo, pessoas anteriormente diagnosticadas com Transtorno de Asperger podem agora receber um diagnóstico do TEA sem comprometimento linguístico ou intelectual (PSIQUIATRIA, 2014). Segundo a Organização Mundial da Saúde (OMS) o número de casos vem aumentando drasticamente visto que uma em cada 160 crianças são diagnosticadas com TEA, sendo cinco vezes mais comum em meninos do que em meninas (ZEIDAN et al., 2022).

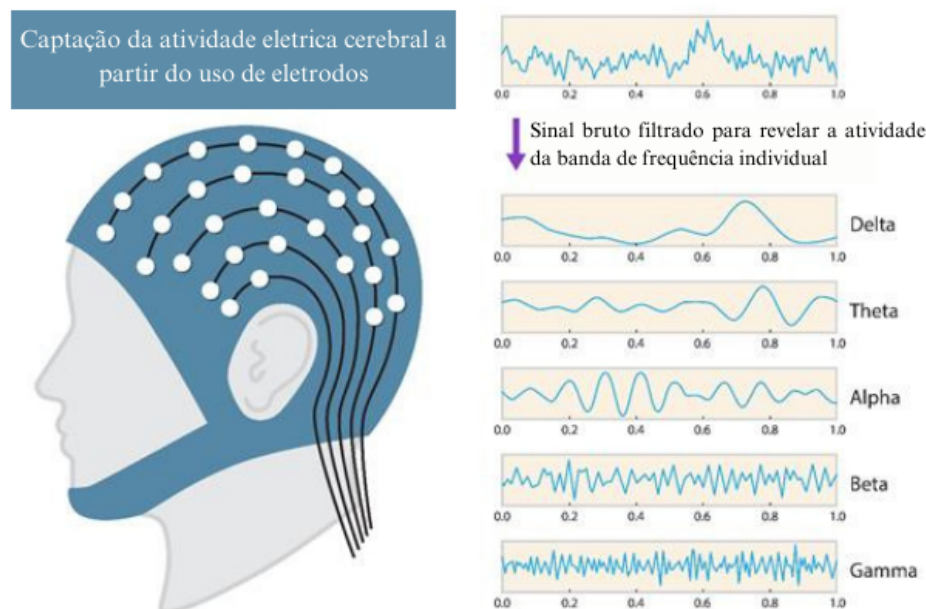


Figura 2 – Representação da captação do sinal de EEG e suas subdivisões em ondas δ (0,5 a 4 Hz), θ (4 a 8 Hz), α (8 a 13 Hz), β (13 a 30 Hz) e γ (30 a 40 Hz).

A eletroencefalografia (EEG) é uma técnica que registra a atividade elétrica do cérebro ao longo do tempo (Figura 2) e tem sido aplicada no contexto do TEA. Estudos com EEG buscam identificar possíveis biomarcadores da atividade cerebral em sinais neurofisiológicos, como padrões anormais de atividade elétrica em regiões específicas do cérebro. Além disso, o EEG é usada para explorar a conectividade cerebral, revelando diferenças na sincronização das ondas cerebrais e na conectividade funcional entre diferentes áreas do cérebro em indivíduos com o TEA (MONTENEGRO et al., 2018).

O EEG também desempenha um papel importante na avaliação de como as pessoas com o TEA respondem a estímulos sensoriais, como luzes intermitentes ou sons, uma vez que

muitas pessoas com o TEA têm sensibilidades sensoriais incomuns (SARMUKADAM et al., 2019). Além disso, a técnica é empregada em estudos de desenvolvimento cerebral, permitindo a observação de como os padrões de atividade cerebral evoluem ao longo do tempo, particularmente em crianças com autismo (SACREY; BENNETT; ZWAIGENBAUM, 2015).

A análise do EEG com o auxílio de ferramentas computacionais é fundamental para aprimorar a precisão dos diagnósticos, minimizando as margens de erro associadas à intervenção humana (ABDULBAQI et al., 2022; GROSSI; VALBUSA; BUSCEMA, 2021). O uso de técnicas computacionais permite processar grandes volumes de dados EEG de maneira eficaz, identificando padrões sutis e produzindo resultados mais confiáveis. Além disso, essas técnicas automatizadas oferecem uma abordagem mais objetiva e rápida, o que é de grande auxílio para os profissionais de saúde no diagnóstico de condições neurológicas. Na literatura, existem diversas técnicas para sua análise que são empregadas no estudo de sistemas dinâmicos, tais como, métodos de tempo e frequência (VARANIS; PEDERIVA, 2011), métodos provindos da Teoria da Informação, métodos baseados em fractais (Higuchi, Katz e Multi-resolução de Contagem de Caixas) (FRAGA; MONDRAGÓN, 2017) e métodos baseados na Teoria de Redes Complexas, como os grafos de visibilidade (LACASA et al., 2009; LACASA et al., 2008). Cada uma dessas abordagens contribui para uma compreensão mais profunda das atividades cerebrais e desempenha um papel significativo na identificação e diagnóstico de distúrbios neurológicos.

Recentemente, um novo método baseado em redes complexas deu origem aos grafos de quantis, como descrita por Campanharo et al. (2016). Essa técnica já se mostrou eficaz em estudos relacionados a condições médicas, como a epilepsia (TOMANIK; CAMPANHARO, 2017; CAMPANHARO; DOESCHER; RAMOS, 2020) e a doença de Alzheimer (PINEDA et al., 2020), e também em um estudo comparativo (VICCHIETTI et al., 2023). Até o momento, a utilização dos grafos de quantis para a avaliação do TEA não havia sido feita, tornando este estudo pioneiro na avaliação de condições neurológicas relacionadas ao TEA.

1.1 Objetivo

Os principais objetivos do trabalho proposto são listados a seguir:

- Empregar o método de quantis pela primeira na classificação de pacientes com TEA e NTEA em diferentes bases de dados, visando discernir-los quantitativamente e comparar seus resultados.
- Investigação das regiões cerebrais relevantes para o TEA.
- Realizar um estudo de comparação entre diferentes frequências de ondas no TEA.

1.2 Estrutura do texto

Os capítulos a seguir estão organizados da seguinte forma:

- Capítulo 2: Breve descrição da teoria de redes complexas, da metodologia abordada e as técnicas de análise de sinais de EEG utilizadas;
- Capítulo 3: Descrição das duas bases de dados utilizadas neste trabalho;
- Capítulo 4: Explicação dos métodos realizados para obtenção dos resultados, dividida em três fases: pré processamento, extração de características e classificação;
- Capítulo 5: Apresentação dos resultados obtidos na diferenciação e na classificação dos pacientes;
- Capítulo 6: Conclusão com base nos resultados alcançados e perspectivas para trabalhos futuros.

2 Redes complexas

2.1 Contexto histórico

A Teoria de Grafos teve início no século XVIII, quando o matemático Leonhard Euler fez contribuições significativas nesse campo. Em 1736, Euler resolveu um problema famoso conhecido como o problema das Sete pontes de Königsberg, localizada na antiga Prússia (EULER, 1741). Königsberg era dividida em várias regiões pelo rio Pregel. Ao todo, essas regiões de terra se conectavam através de sete pontes. O desafio consistia em alcançar cada região de terra atravessando cada ponte apenas uma vez. A Figura 3 representa o problema. Para chegar na solução do problema, Euler simplificou a cidade de Königsberg, ignorando suas características físicas, como a localização exata das pontes e as curvas do rio, representando o problema em um grafo. Cada região de terra foi representada por um vértice (ou nó), e cada ponte que as conectava foi considerada uma aresta (AMARAL; OTTINO, 2004).

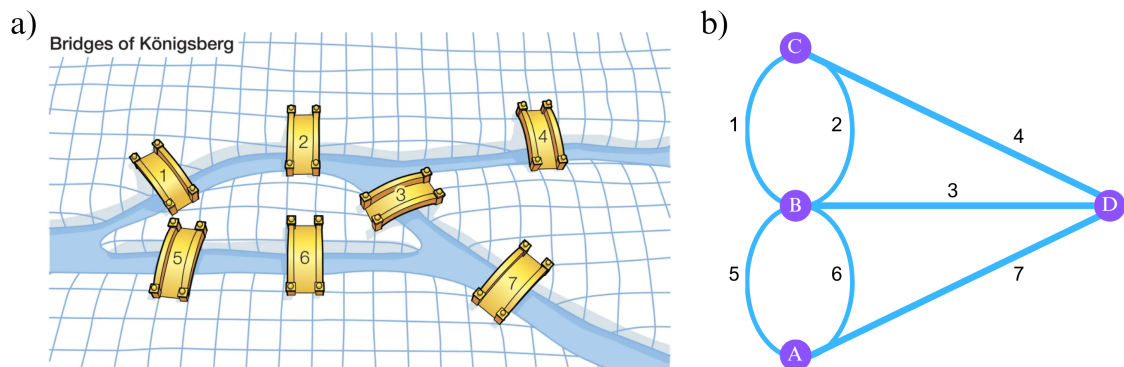


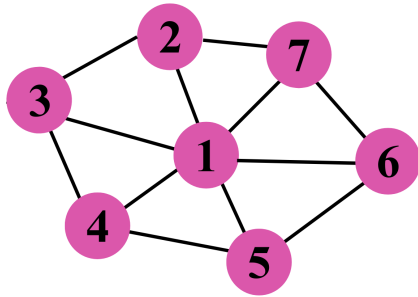
Figura 3 – Ilustração do problema das Sete Pontes. a) Representação esquemática da área com as pontes. b) Representação do problema por um grafo proposto por Euler. Fonte: (BRITANNICA, 2003).

Ao analisar o problema através do grafo, Euler provou que era impossível encontrar um passeio que atravessasse cada ponte apenas uma vez se houvesse mais de duas regiões de terra com um número ímpar de pontes conectando-as. Demonstrando que:

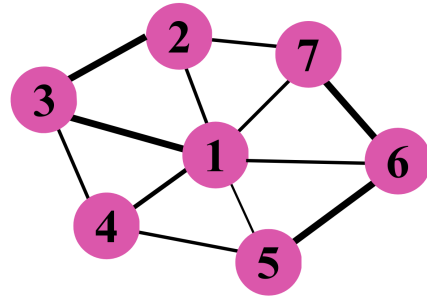
- A soma dos graus dos nós de um grafo é par;
- Todo grafo deve ter um número par de nós ímpares.

Essas descobertas permitiram demonstrar que a existência de percursos de Euler¹ depende do número de nós ímpares: se houver mais de dois nós ímpares, não existe um percurso de Euler; se houver exatamente dois nós ímpares, existem percursos de Euler iniciando em qualquer um deles; e se não houver nós ímpares, os percursos de Euler podem começar em qualquer nó. Esses resultados foram fundamentais para o desenvolvimento da Teoria de Grafos e suas aplicações em diversos campos da Ciência (AMARAL; OTTINO, 2004).

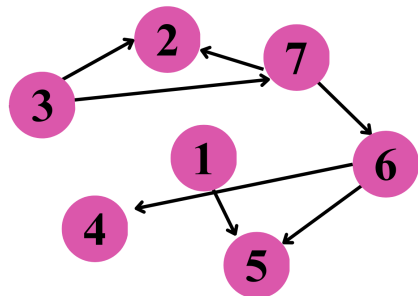
Uma rede complexa é composta por um conjunto de nós, representados como $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$, e um conjunto de arestas, representadas como $\mathcal{H} = \{h_1, h_2, \dots, h_M\}$, que descrevem as conexões entre esses elementos. Essas arestas podem ser direcionadas, indicando uma orientação específica, ou não-direcionadas, quando a orientação não é relevante. Em casos em que as arestas possuem atributos de intensidade, cada uma é associada a um valor de peso. Nesse contexto, a rede ponderada é representada por $g = (\mathcal{N}, \mathcal{H}, W)$, em que $W = \{w_1, w_2, \dots, w_M\}$ corresponde aos pesos atribuídos às M arestas (NEWMAN, 2018; LATORA; NICOSIA; RUSSO, 2017).



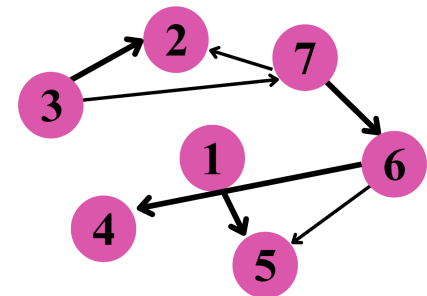
a) Rede sem peso e sem direção.



b) Rede com peso e sem direção.



c) Rede sem peso e com direção.



d) Rede com peso e com direção.

Figura 4 – Exemplos de redes complexas com diferentes topologias. A intensidade de cada aresta é dada por sua espessura e a direção da conexão é dada por sua seta.

¹ Um caminho entre dois nós arbitrários no qual cada aresta do grafo aparece exatamente uma vez.

A Figura 4 mostra exemplos de redes complexas, que são sistemas compostos por elementos interconectados, como nós e arestas. Existem dois tipos de redes destacadas na figura:

- **Redes Direcionadas:** Nelas, as conexões entre os nós têm uma direção específica, representada por setas que indicam o sentido da conexão;
- **Redes Ponderadas:** Nestas redes, a intensidade das conexões é representada pela espessura das arestas. Isso significa que algumas conexões podem ser mais fortes do que outras.

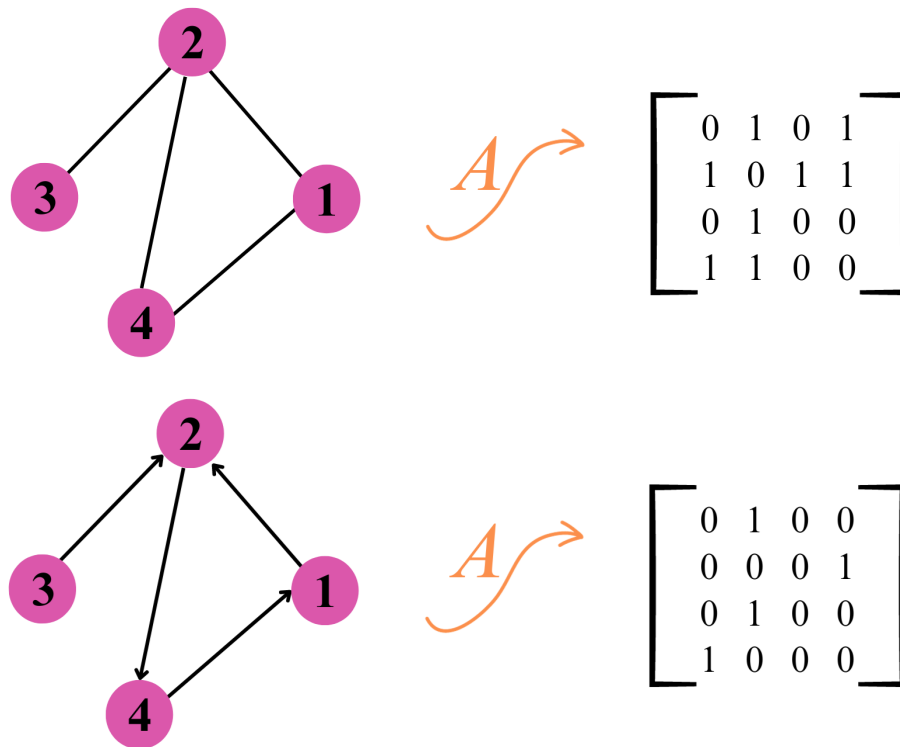


Figura 5 – a) Rede não-direcionada e matriz de adjacência correspondente. Rede não-direcionada e matriz de adjacência correspondente; b) Rede direcionada e matriz de adjacência correspondente.

A partir de uma rede complexa, podemos derivar uma estrutura de dados conhecida como "matriz de adjacência", como exemplificado na Figura 5. Quando dois vértices n_i e n_j estão interconectados, a entrada a_{ij} na matriz de adjacência A é atribuída o valor 1, indicando a presença da conexão. Quando não existe conexão entre n_i e n_j , a entrada a_{ij} na matriz de adjacência é estabelecida como 0. Em redes ponderadas, em vez de utilizar 0 ou 1, a matriz de adjacência W emprega valores numéricos que refletem os pesos das arestas (Figura 6). Por exemplo, se uma conexão entre n_i e n_j tem um peso de 6, o elemento w_{ij} na matriz W será definido como (NEWMAN, 2018; LATORA; NICOSIA; RUSSO, 2017).

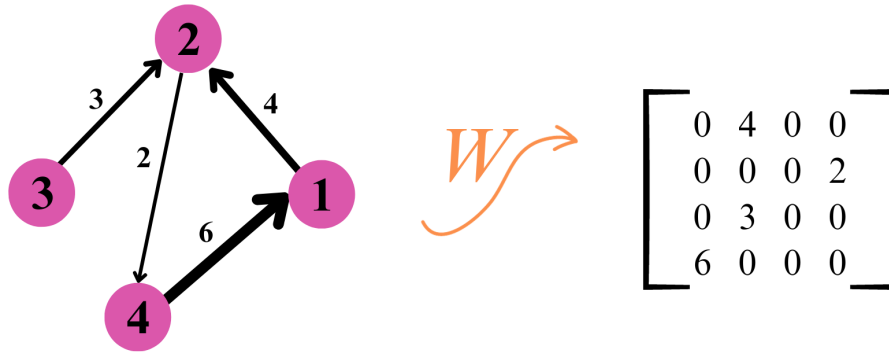


Figura 6 – Rede direcionada com peso e matriz de adjacência W correspondente.

Essa representação em formato matricial é uma maneira eficaz de armazenar informações relativas à estrutura e características de uma rede complexa, tornando mais simples a análise e processamento computacional dessas informações.

Em suma, a Teoria de Grafos lida com relações básicas entre elementos fornecendo os fundamentos matemáticos para a representação e análise de relações em sistemas discretos (FEOFILOFF; KOHAYAKAWA; WAKABAYASHI, 2011), enquanto a teoria de redes complexas explora relações intrincadas e padrões em redes do mundo real, com ênfase em propriedades complexas e emergentes, sendo uma evolução natural da teoria de grafos. À medida em que a teoria em redes complexas avançou, houve um crescente interesse no estudo de dinâmicas dos processos que ocorrem nas redes, como a propagação de informações, doenças e falhas. Foram desenvolvidos modelos e métodos para simular e analisar esses processos, incluindo a modelagem de epidemias, doenças neurológicas, difusão de inovações, influência social, entre outros. Esses estudos permitiram uma compreensão mais profunda dos mecanismos subjacentes a esses fenômenos e a possibilidade de prever e controlar seu comportamento nas redes complexas (JIN; GIRVAN; NEWMAN, 2001; COSTA; BARRETO; SAMPAIO, 2021).

Dado que um sinal de EEG é uma série temporal, uma forma eficaz de derivá-lo em redes complexas é utilizando o método de Grafo de Quantis.

2.2 Grafo de quantis

Seja $\mathcal{M}$ um mapeamento de uma série temporal X com $\mathcal{T}$ pontos em uma rede g , onde $X = \{x(t) | t \in \mathbb{N}, x(t) \in \mathbb{R}\}$ e $g = \{\mathcal{N}, \mathcal{H}\}$ é um conjunto de N nós e M arestas. O Grafo de quantis (CAMPANHARO et al., 2011) pressupõe uma discretização simples de X que não leva em consideração a distribuição de seus valores, mas sim seus $\mathcal{Q}$ quantis, calculados como $\mathcal{Q} \approx 2 \times \mathcal{T}^{\frac{1}{3}}$. Assim, a série temporal é dividida em $\mathcal{Q}$ quantis, onde, cada ponto dessa série será atribuído a um quantil correspondente q_i , de forma que:

$$q_i = x\left(\frac{\mathcal{T} * i}{\mathcal{Q}}\right) \quad i = 1, 2, \dots, \mathcal{Q}. \quad (2.1)$$

Após isso, $\mathcal{M}$ associa cada quantil q_i a um nó $n_i \in \mathcal{N}$ na rede correspondente. Dois vértices n_i e n_j estarão conectados na rede com uma aresta w_{ij} , cujo peso de cada é dado pelo número de vezes que um ponto $x(t) \in q_i$ é seguido por um ponto $x(t+k) \in q_j$ na série temporal, com $t = 1, 2, \dots, \mathcal{T} - k$ e k sendo a distância temporal entre dois pontos quaisquer $k = 1, 2, \dots, k_{max}$ (CAMPANHARO; RAMOS, 2016), possibilitando a construção de diferentes redes que representam a ligação dos pontos da série temporal com k pontos futuros da série representado na figura a seguir.

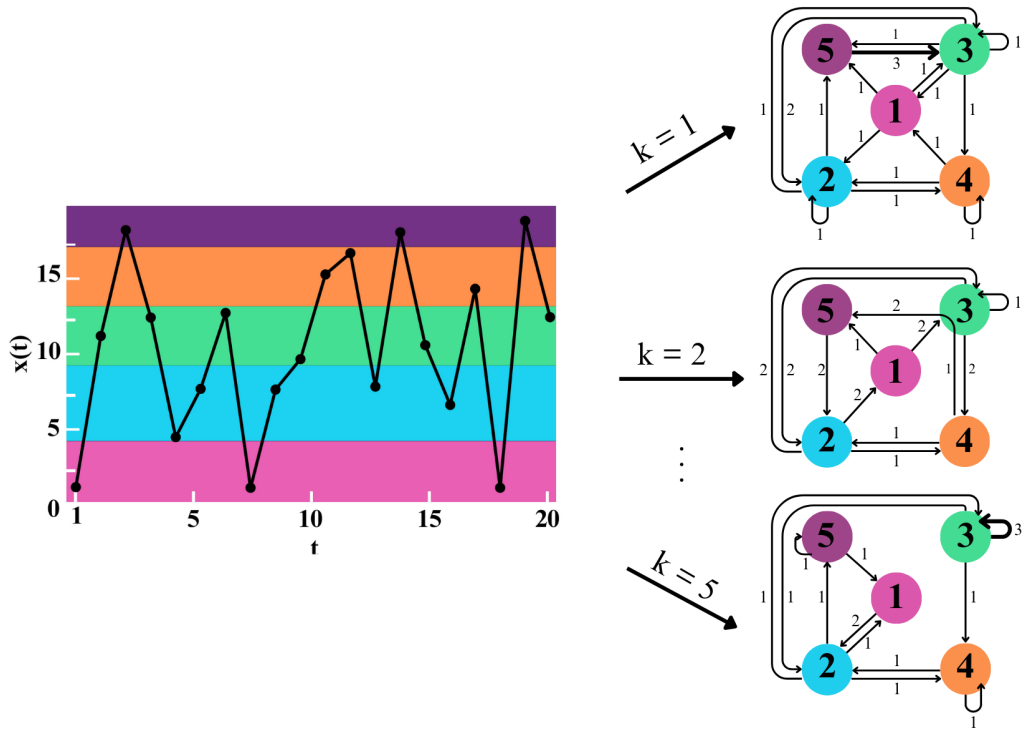


Figura 7 – Ilustração do mapeamento de uma série temporal em grafos de quantis.

Através de uma rede podemos representá-la por uma matriz de transição de Markov P , obtida pela matriz de adjacência de pesos W , sendo:

$$p_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}. \quad (2.2)$$

2.3 Caracterizadores topológicos

As propriedades de uma série temporal podem ser quantificadas com base nas propriedades estruturais de sua rede associada. Os caracterizadores topológicos são medidas essenciais para entender a organização e o comportamento dessas redes por meio de suas conexões (JIN; GIRVAN; NEWMAN, 2001). Neste trabalho, as medidas de redes utilizadas serão descritas a seguir, sendo essas, a centralidade do autovetor (ec), a centralidade de proximidade (C), o

coeficiente de agrupamento (Ca), o comprimento do salto médio (Δ), a eficiência global (L), a energia do grafo (E) e a modularidade ($\mathcal{M}o$).

2.3.1 Centralidade do autovetor

A medida de centralidade do autovetor de um vértice pode ser expressa de duas maneiras equivalentes: como uma equação matricial e como uma soma. Ela define que a centralidade de um vértice está relacionada à soma das centralidades dos vértices aos quais ele está conectado. O valor próprio λ representa o maior autovalor de A , e N é o número de vértices no grafo, podendo ser calculado como (RUHNAU, 2000; BONACICH, 2007):

$$\lambda ec_i = \sum_{j=1}^N a_{ij} ec_{n_j}, \quad (2.3)$$

com ec_j centralidade do autovetor do nó n_j , com $j = 1, \dots, N$.

$$Aec = \lambda ec. \quad (2.4)$$

Esse tipo de equação é amplamente conhecido e resolvido pelos autovalores e autovetores de A .

Definimos a medida de centralidade do autovetor média ec como a média aritmética da medida de centralidade do autovetor de todos os vértices n_i ($i = 1, 2, \dots, N$):

$$ec = \frac{1}{N} \sum_{i=1}^N \lambda ec_i. \quad (2.5)$$

2.3.2 Centralidade de proximidade

A centralidade de proximidade é uma medida que busca avaliar o quão próximo um vértice está dos outros vértices em uma rede. Se os caminhos mais curtos entre o vértice n_i e os demais vértices na rede são curtos, então o vértice n_i é considerado central em termos de proximidade (KIM et al., 2011). Isso reflete o quão fácil e conveniente é para o vértice em foco se conectar com os demais vértices na rede (ZHANG; LUO, 2017). Seu cálculo é representado pela seguinte equação:

$$C' = \sum_{j=1}^N d(n_i, n_j), \quad (2.6)$$

onde d representa a distância (ou comprimento do caminho) entre dois vértices específicos na rede.

Em uma rede direcionada, as distâncias entre dois nós podem variar em direções opostas. Para evitar isso, a centralidade de proximidade típica de um nó é calculada como a soma das

distâncias entre esse nó e todos os outros, normalizada em relação ao tamanho da rede (KIM et al., 2011):

$$C = \left[\sum_{j=1}^N d(n_i, n_j) \right]^{-1}. \quad (2.7)$$

No máximo, o índice é igual a $(N - 1)^{-1}$, o que ocorre quando o nó está adjacente a todos os outros nós.

2.3.3 Coeficiente de agrupamento

Dada uma rede g , consideremos um vértice n_i que possui h_i vizinhos. O número máximo de arestas que podem existir entre esses vizinhos é dado por $h_i(h_i - 1)/2$, o que ocorre quando todos os vizinhos de n_i estão conectados entre si. O coeficiente de agrupamento Ca_i do vértice n_i é obtido pela razão entre o número E_i de arestas efetivamente existentes entre os vizinhos de n_i e o número total de arestas possíveis. Em outras palavras, o coeficiente de agrupamento Ca_i mede a probabilidade de conexões existentes entre os vizinhos do vértice n_i em relação ao máximo possível de conexões entre eles, ou seja:

$$Ca'_i = \frac{E_i}{(h_i(h_i - 1))/2}. \quad (2.8)$$

O coeficiente de agrupamento para o vértice de uma rede com peso e direcionada pode ser calculado diretamente por meio de sua matriz de adjacência W associada (FAGIOLO, 2007), sendo:

$$Ca'_i(W) = \frac{(W^{\frac{1}{3}})_{ii}^3}{h_i(h_i - 1)}, \quad (2.9)$$

O coeficiente de agrupamento médio Ca é definido como a média aritmética dos Ca'_i , para todos os vértices $n_i (i = 1, 2, \dots, N)$:

$$Ca = \frac{1}{N} \sum_{i=1}^N Ca'_i. \quad (2.10)$$

2.3.4 Comprimento do menor caminho

Em redes não ponderadas e não direcionadas, o menor caminho entre dois vértices, n_i e n_j , é aquele que os conecta com o menor número possível de arestas. Por outro lado, em redes ponderadas, os pesos das arestas podem ter duas interpretações distintas. Primeiramente, podem representar distâncias físicas, como em uma rede de cidades onde os vértices são cidades e as arestas são as distâncias entre elas. Nesse caso, a distância ao longo de um caminho é a soma dos pesos das arestas nesse caminho. Em segundo lugar, os pesos podem refletir a força de conexão

entre os vértices, como em uma rede de internet, onde os vértices são roteadores e os pesos das arestas são larguras de banda. Nesse caso, a distância ao longo de um caminho é a soma dos inversos dos pesos das arestas nesse caminho (COSTA et al., 2007).

O menor caminho médio $\mathcal{L}$ de uma rede conectada é obtido através do cálculo:

$$\mathcal{L} = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (2.11)$$

onde d_{ij} representa os elementos da matriz de distâncias D , representando o valor do menor caminho entre os vértices n_i e n_j .

Entretanto, a fim de evitar divergências nos valores de $\mathcal{L}$, é prática comum limitar a análise a pares de vértices pertencentes à maior componente conectada da rede (WATTS; STROGATZ, 1998). Essa abordagem é adotada para assegurar cálculos mais robustos e significativos, especialmente em cenários onde a rede pode conter componentes desconexas ou infinitas, para isso, usamos a definição de eficiência global (LATORA; MARCHIORI, 2001):

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \quad (2.12)$$

quantificando a eficiência da rede no envio de informações entre vértices, considerando que ela é diretamente proporcional ao inverso da distância entre dois vértices n_i e n_j .

2.3.5 Comprimento do salto médio

Obtida uma matriz de transição de Markov, cada posição p_{ij} representa a probabilidade de transição entre o vértice n_i para o vértice n_j , possibilitando estimar o valor médio de caminhos percorridos por um caminhante aleatório, dado por:

$$\Delta = \frac{1}{S} \sum_{i=1}^S \delta_s(i, j), \quad (2.13)$$

$s = S$ saltos de comprimento $\delta_s(i, j) = |i - j|$, para $i, j = 1, \dots, \mathcal{Q}$ (CAMPANHARO; RAMOS, 2016). O cálculo do comprimento do salto médio, pode ser feito com base na expressão:

$$\Delta = \frac{1}{N} \text{tr}(MpP^T), \quad (2.14)$$

onde, P é a matriz de transição de Markov, Mp é a matriz de posições $N \times N$ com os elementos $Mp_{ij} = |i - j|$ e tr é o traço da matriz resultante.

2.3.6 Energia do grafo

A energia de um grafo g , denotada por E , é definida como a soma dos valores absolutos de seus autovalores (GUTMAN; LI; ZHANG, 2009). Portanto, se A é a matriz de adjacência de g e $\lambda_1, \dots, \lambda_z$ são os autovalores de A , então:

$$E = \sum_{i=1}^z |\lambda_i|, \quad (2.15)$$

onde o conjunto $\{\lambda_1, \dots, \lambda_z\}$ é chamado de espectro de g , representado como $Spec\ g$.

2.3.7 Modularidade

A modularidade é uma medida usada na análise de redes complexas para avaliar a qualidade da divisão da rede em comunidades ou módulos. Uma comunidade é um grupo de nós altamente interconectados, onde os nós dentro do mesmo módulo estão mais densamente conectados entre si do que com nós fora desse módulo (GUIMERA; SALES-PARDO; AMARAL, 2004). Para uma determinada partição pa dos nós de uma rede em módulos, a modularidade $\mathcal{M}_p$ desta partição é definida como:

$$\mathcal{M}_p = \sum_{f=1}^r \left[\frac{\mathcal{A}_f}{\mathcal{A}} - \left(\frac{d_f}{2\mathcal{A}} \right)^2 \right], \quad (2.16)$$

onde r é o número de módulos da rede, $\mathcal{A}$ é o total de arestas da rede, $\mathcal{A}_f$ é o número de arestas entre os vértices pertencentes ao s -ésimo módulo e d_f é a soma dos graus de conectividade dos vértices pertencentes ao s -ésimo módulo. Essa definição de modularidade implica que $\mathcal{M} \geq 0$ e que $\mathcal{M} = 0$ para uma partição aleatória dos nós (JIN; GIRVAN; NEWMAN, 2001). Definimos a modularidade Mo de uma rede como a maior modularidade de todas as possíveis partições da rede, ou seja, $Mo = \max(\mathcal{M}_p)$.

3 Bases de Dados

Depois de uma investigação minuciosa e abrangente em fontes independentes, identificou-se duas bases de dados apropriadas, ou seja, com o mínimo de informações necessárias para ser apto a realizar o estudo. Devido à dificuldade em encontrá-las, optou-se por utilizar ambas, viabilizando uma análise comparativa dos resultados. A escassez de bases de dados com informações acessíveis para o pesquisador ressalta a complexidade associada à procura por recursos relevantes para a pesquisa. A seguir, apresenta-se uma breve descrição de cada base de dados utilizada neste trabalho.

3.1 Base de dados de King Abdulaziz

O primeiro conjunto de dados utilizado foi fornecido gratuitamente pelo Hospital da Universidade King Abdulaziz (KAU), na Arábia Saudita, em Jeddah ([ALHADDAD et al., 2012](#)). Constituído por 17 pacientes com idades entre 6 e 20 anos, sendo 13 grupo TEA (3 meninas e 9 meninos, 6 – 20 anos) e 4 grupo NTEA (todos meninos, 9-13 anos) sem histórico de distúrbios neurológicos. Os sinais foram registrados usando 16 eletrodos Ag/AgCl com tampa g.tec EEG, amplificadores g.tec USB e o software BCI2000, em estado de repouso, conforme mostrado na Figura 8. Os dados foram filtrados utilizando um filtro passa-banda com uma faixa de frequência de 0,1 a 60 Hz, um filtro de fenda com uma frequência de 60 Hz e digitalizados a uma taxa de amostragem de 256 Hz.

Um exemplo do estudo de ondas pode ser visualizado na Figura 20, onde temos a divisão de um sinal de EEG em ondas δ , θ , α e β .

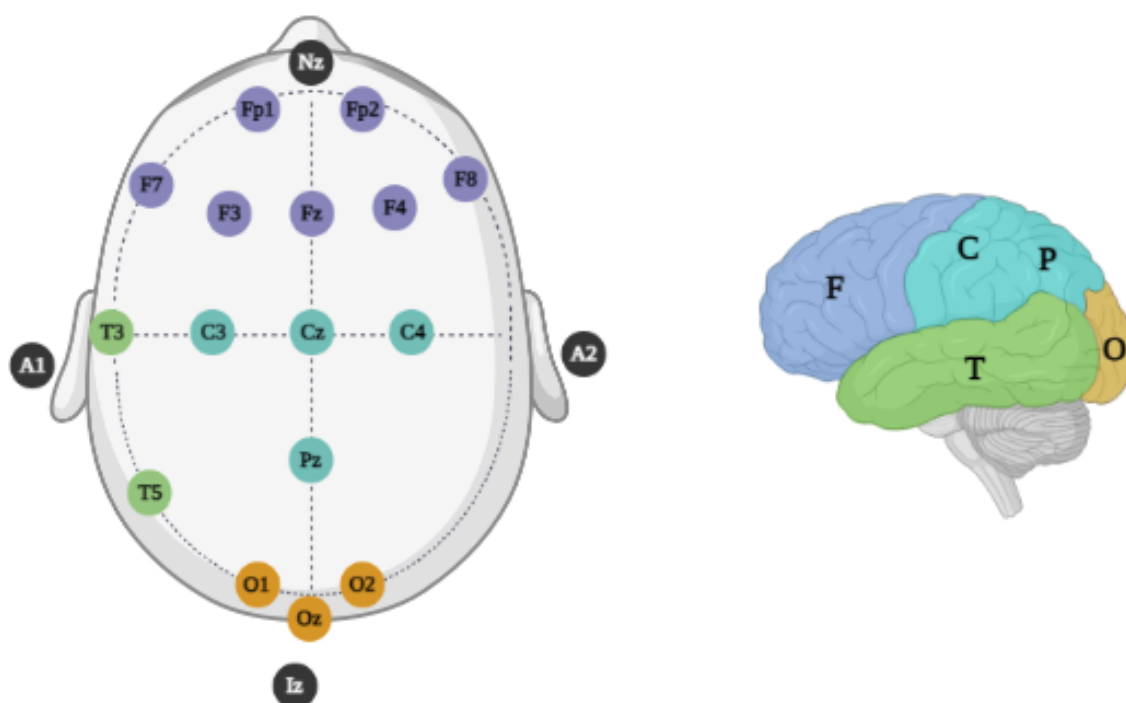


Figura 8 – Ilustração da disposição dos 16 eletrodos no escalpo referente à base de dados de King Abdulaziz.

3.2 Base de dados de Sheffield

O segundo conjunto de dados foi fornecido gratuitamente pela University of Sheffield (MILNE et al., 2019). Os dados foram obtidos de 56 indivíduos com idades entre 18 e 68 anos, sendo 23 do grupo TEA e 23 do grupo NTEA (neurotípicos). Cada participante contribuiu com mais de 2,5 minutos de dados de EEG em repouso com os olhos fechados. Os dados de EEG foram adquiridos por meio do sistema EEG Biosemi ActiveTwo a partir de 64 eletrodos. Os dados foram filtrados utilizando um filtro passa-alta para remover frequências abaixo de 1 Hz e amostrados novamente a 512 Hz.

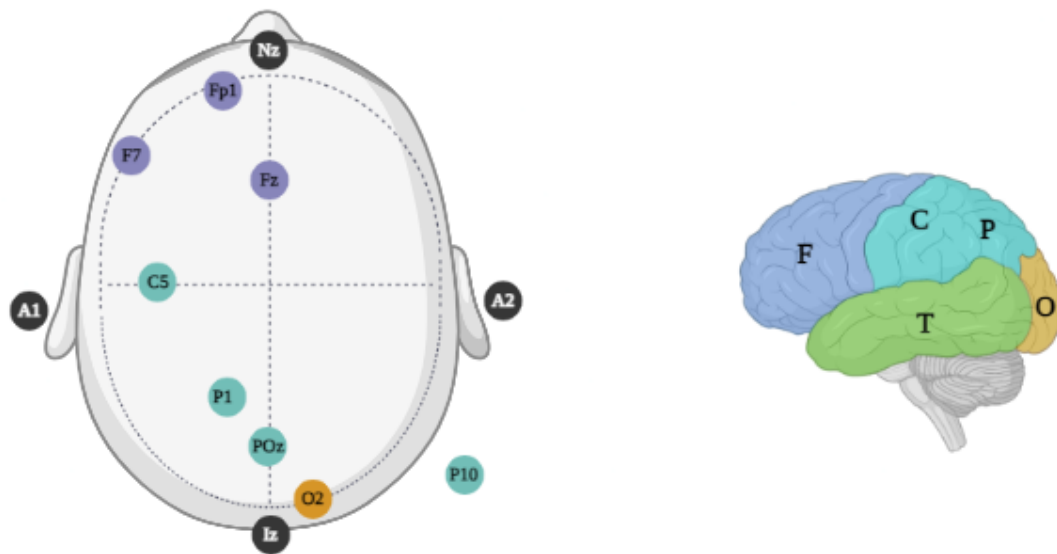


Figura 9 – Ilustração da disposição dos 8 eletrodos no escalpo referente à base de dados de Sheffield.

4 Metodologia

A metodologia deste trabalho foi dividida em três fases distintas.

4.1 Pré processamento dos dados

A base de dados de King Abdulaziz se trata de um conjunto de dados de acesso público. Para acessar o conjunto de dados, foi realizado um pedido formal por e-mail ao proprietário do conjunto de dados (ALBERT; BARABÁSI, 2002). Após esses procedimentos, deu-se início à etapa de extração desses dados, os quais estavam em árabe e continham informações sobre a condição e registros de EEG dos pacientes.

Para a segunda base de dados, de Sheffield, obtida de uma fonte independente e adquirida de um estudo de EEG previamente publicado (MILNE et al., 2019), no qual os participantes consentiram com a disponibilização de seus dados para fins de análise secundária, o processamento foi realizado utilizando o EEGLAB (FAYAZ, 2023) e programas no MATLAB personalizados. Dado que existiam diferentes configurações de eletrodos em uso por pacientes distintos nesta base de dados, um algoritmo de busca foi implementado e aplicado com o intuito de identificar um conjunto de eletrodos comuns, minimizando ao máximo a exclusão de pacientes para fins de análise. Como resultado desse processo, identificou-se um conjunto de 8 eletrodos compartilhados por 40 pacientes (24 TEA, 16 NTEA), a saber: Fp1, F7, Fz, C5, P1, POz, O2 e P10 (Figura 8).

Posteriormente, todos os sinais de EEG foram filtrados com um filtro passa-banda com uma faixa de frequência de 0,5 a 30 Hz para a remoção de ruídos e submetidos a um processo de padronização com base no número de pontos $t = 73.000$ para a base de dados de Sheffield e $T = 9.000$ para a base de dados de King Abdulaziz, para que a quantidade de vértices fosse a mesma para o procedimento de mapeamento em grafos de quantis.

Foram gerados gráficos, representando a relação entre frequência (Hz) e tempo (s), com o propósito de realizar uma comparação entre os grupos TEA e NTEA. Além disso, foram gerados gráficos de espectro de potência com a finalidade de avaliar a eficácia dos filtros aplicados e analisar a distribuição de energia ao longo das séries temporais. Adicionalmente, foram elaborados gráficos de autocorrelação com o intuito de examinar a similaridade de uma série temporal consigo mesma e determinar a adequação dos parâmetros em questão para análise.

Para o estudo das ondas, a quantidade de pontos foi a mesma entretanto foram realizadas quatro etapas de filtragens distintas, utilizando filtros passa-banda nos intervalos correspondentes as frequências listadas na Tabela 2. Também foram gerados gráficos de relação entre frequência (Hz) e tempo (s), potência e autocorrelação.

Filtro	Frequência (Hz)
δ	0,5 a 4
θ	4 a 8
α	8 a 13
β	13 a 30

Tabela 2 – Filtros utilizados para gerar as ondas cerebrais e suas frequências correspondentes.

É importante destacar que todas as etapas dos processos computacionais foram conduzidas integralmente no software MATLAB. Isso inclui desde procedimentos mais simples e básicos, como a extração dos dados de seus respectivos arquivos, até a fase de mapeamento e obtenção dos resultados.

4.2 Extração de características

O mapeamento efetuado em ambas as bases de dados foi proposto por Campanharo *et. al.* (CAMPANHARO *et al.*, 2011; CAMPANHARO; RAMOS, 2016), conforme detalhado no Capítulo 2.2. Na base de dados de King Abdulaziz, foram utilizados $Q = 41$ quantis, enquanto na base de dados de Sheffield, foram empregados $Q = 84$ quantis. Para ambas as bases exploraram-se uma gama de valores $k = 1, \dots, 1000$. Após o mapeamento, aplicaram-se os caracterizadores topológicos descritos no capítulo 2.

4.3 Classificação

Foram utilizadas as medidas obtidas por meio dos caracterizadores topológicos em um processo de classificação, empregando o algoritmo Support Vector Machine (SVM) disponível no Matlab e implementado no programa geral para a classificação dos pacientes. O SVM é uma técnica consagrada em aprendizado de máquina, notória por sua eficácia na resolução de tarefas de classificação e regressão em contextos de alta dimensionalidade. Seu funcionamento pode ser descrito meticulosamente em etapas distintas e fundamentais (GHOLAMI; FAKHARI, 2017).

Inicialmente, realizou-se uma precisa definição do tipo de problema a ser abordado, discriminando entre as categorias de classificação e regressão. Na tarefa de classificação, o objetivo é particionar os dados em múltiplas categorias.

A classificação é dividida em duas etapas fundamentais: a fase de treinamento e a etapa de teste. Na fase de treinamento, um grupo de dados é utilizado para ajustar o modelo preditivo, permitindo que ele aprenda os padrões e relações nos dados de treinamento. Isso é frequentemente

realizado por meio de algoritmos de aprendizado de máquina, como árvores de decisão, redes neurais ou máquinas de vetores de suporte. O objetivo é otimizar os parâmetros do modelo de modo que ele seja capaz de fazer previsões precisas com base nos dados de entrada.

Na segunda etapa, conhecida como fase de teste, um conjunto de dados separado e independente é usado para avaliar o desempenho do modelo. Após o treinamento, se for bem treinado, o modelo SVM se revela apto a classificar novos exemplos (no caso da classificação). Esse desempenho é orientado pela posição dos pontos de dados em relação ao hiperplano de decisão ou à função de regressão.

Neste estudo, a avaliação do desempenho do modelo SVM é realizada mediante a utilização da técnica de validação cruzada K-fold. Essa abordagem envolve a divisão da amostra de dados em K grupos, onde K - 1 grupos são utilizados para treinamento e o grupo restante é reservado para teste (KOHAVI et al., 1995). Esse processo é repetido K vezes, com a permutação dos grupos, garantindo que todos eles tenham sido testados.

Com base nesses valores, foi possível calcular a acurácia (Acu) durante cada repetição, que é definida da seguinte forma:

$$Acu = \frac{(vp + vn)}{(vp + vn + fp + fn)}, \quad (4.1)$$

onde fp são os falsos positivos, fn falsos negativos, vn verdadeiros negativos e vp verdadeiros positivos.

5 Resultados

A aplicação do mapeamento em questão foi utilizada para abordar a distinção entre pacientes neurotípicos e aqueles diagnosticados com o TEA. A seguir apresentamos duas seções trazendo os resultados para cada base de dados utilizada neste trabalho.

5.1 Base de dados de King Abdulaziz

Dado que cada série temporal possui um tamanho de $T = 9.000$ pontos, utilizamos $Q = 2 \times (9.000)^{\frac{1}{3}}$. Consequentemente, foram mapeados os 320 sinais de EEG em um total de 320.000 grafos de quantis ($A_k = 320.000$). Desta forma, obtivemos 320.000 matrizes de transição de Markov, cada uma com $Q^2 = 1.681$ elementos.

As Figuras 11 e 13 exibem exemplos de sinais de EEG selecionados aleatoriamente de indivíduos, representados nas Figuras 23 e 25 após a aplicação do filtro passa-banda (0, 5 a 30 Hz), correspondentes a cada canal de EEG, representando regiões específicas do escalpo (Fp1 = Frontal, C3 = Central, O2 = Occipital, P10 = Parietal). É importante observar que, em todos os casos, os sinais de EEGs de pacientes com o TEA apresentam diferenças em relação aos pacientes NTEA.

Ao examinar atentamente a Figura 14 pode-se destacar não apenas a eficácia, mas também a relevância da aplicação do filtro no contexto de análise. A Figura evidencia de maneira clara como o filtro desempenha um papel crucial na modificação da composição espectral do sinal de EEG em questão.

Inicialmente, é importante notar que o filtro foi aplicado com o propósito de realçar (frequências δ , θ , α , β ou suprimir (artefatos - abaixo de 0, 5 Hz e ruídos - acima de 30 Hz) dentro do sinal. Nesse contexto, a eficácia do filtro é visível por meio das alterações discerníveis nas amplitudes das frequências representadas no espectro. Além disso, ao analisar a figura, também é notável que, à medida que se desloca para frequências mais altas, ocorre uma diminuição na distribuição de energia. Isso reflete a ação do filtro em atenuar ou eliminar componentes de alta frequência que podem conter ruído ou informações menos relevantes para a análise em questão.

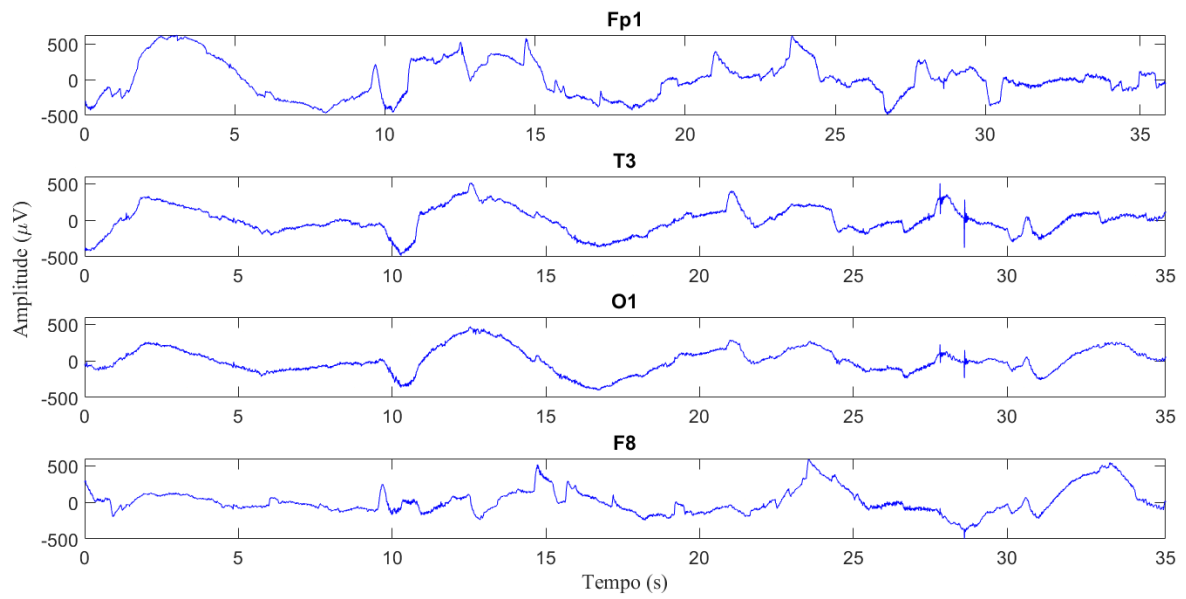


Figura 10 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, T3, O1 e F8.

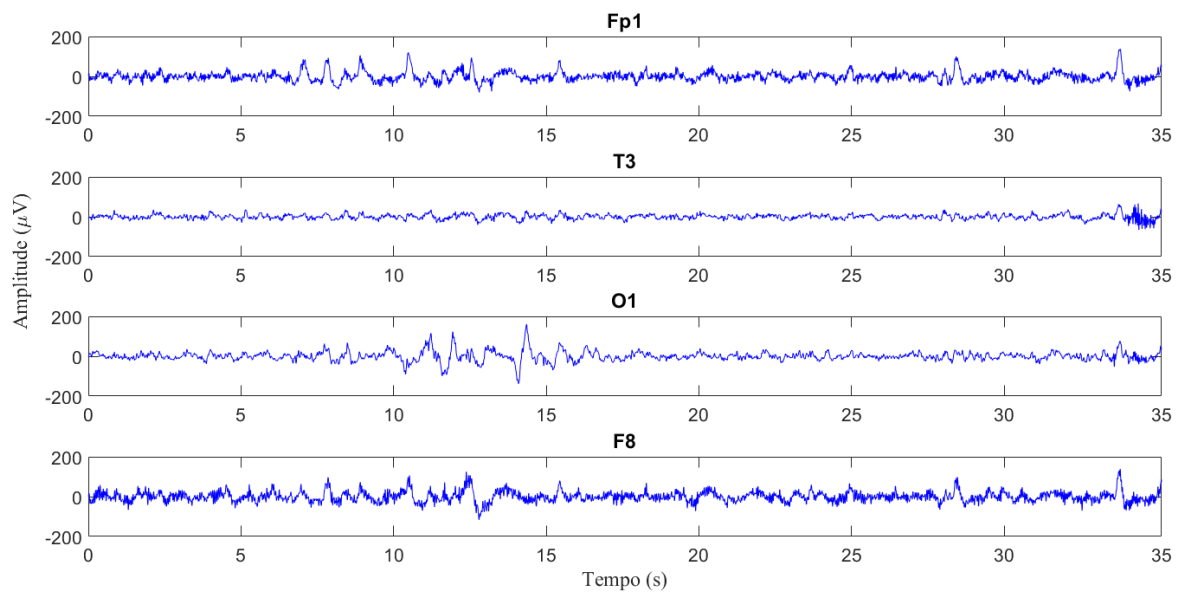


Figura 11 – Exemplificação dos sinais do paciente do grupo TEA representados na Figura 10 após aplicação do filtro passa-banda (0,5 a 30 Hz).

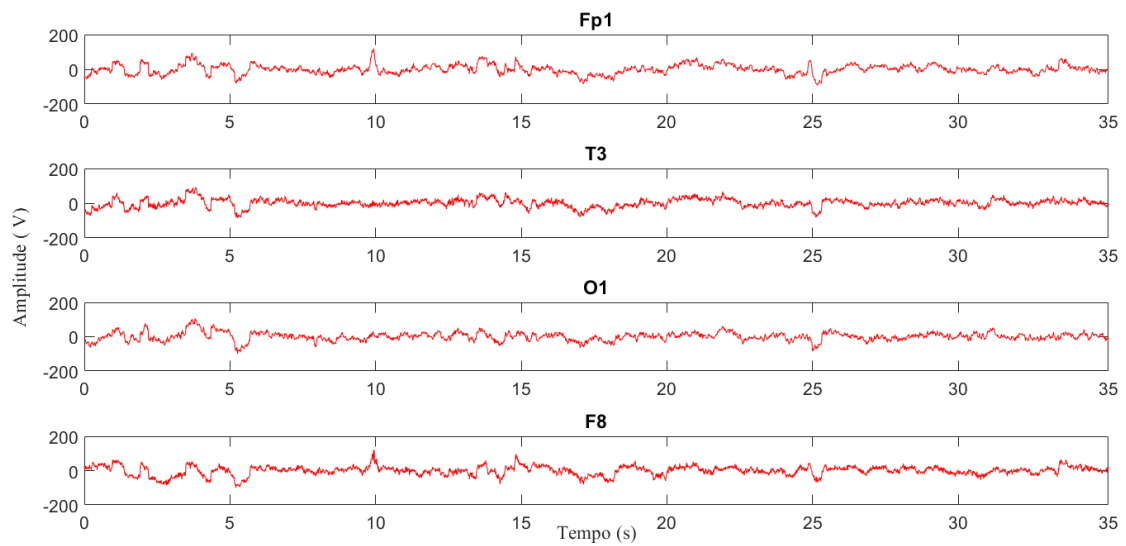


Figura 12 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo NTEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, T3, O1 e F8.

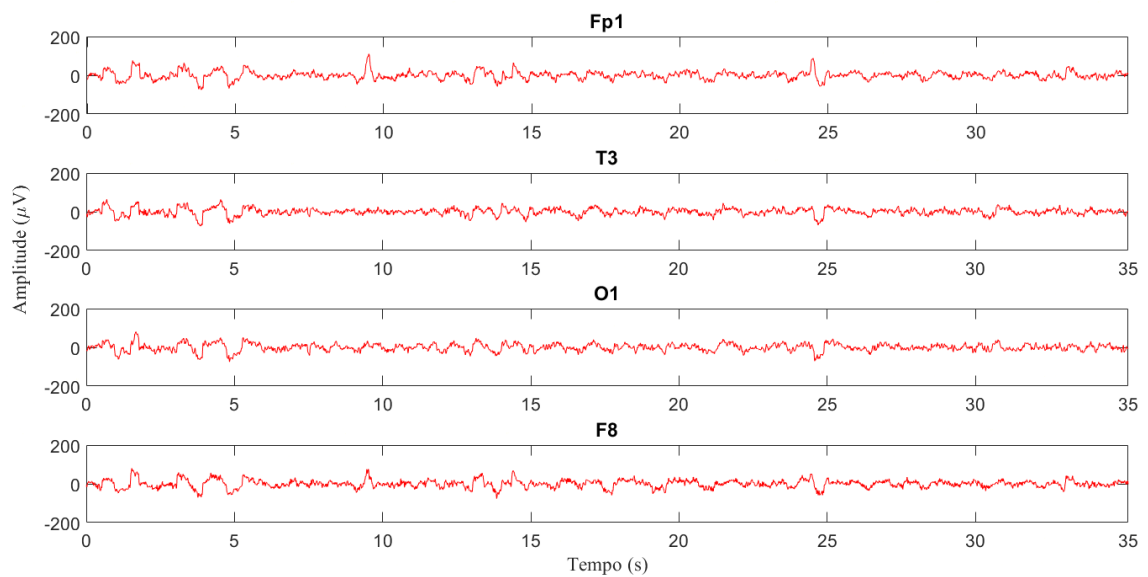


Figura 13 – Exemplificação dos sinais do paciente do grupo NTEA representados na Figura 12 após aplicação do filtro passa-banda (0,5 a 30 Hz).

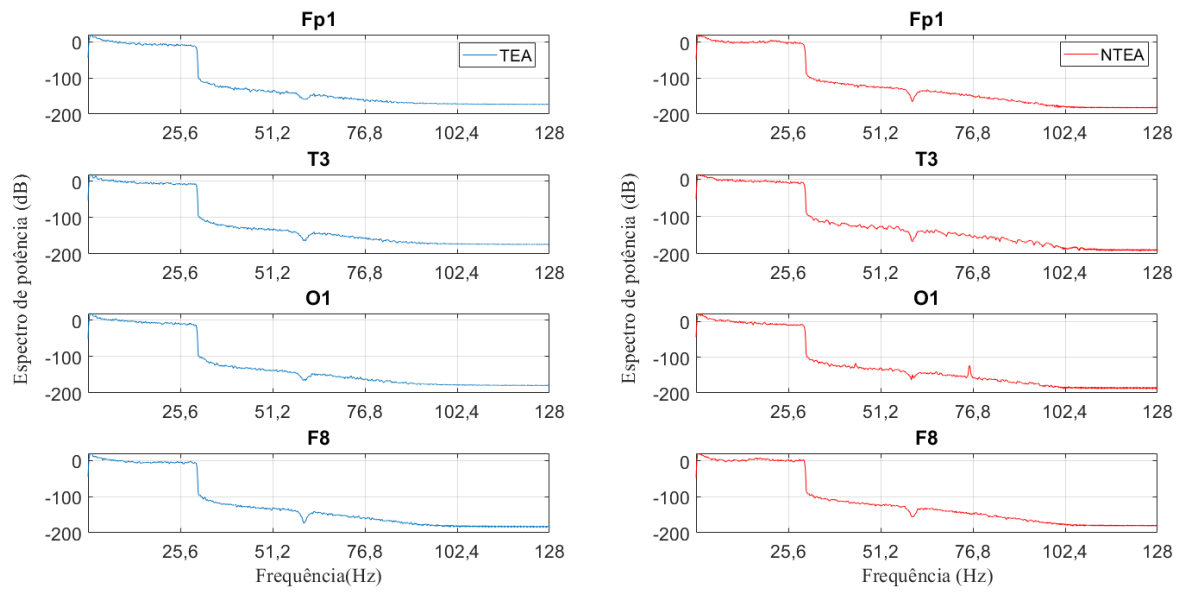


Figura 14 – Espectro de potência dos sinais dos eletrodos Fp1, T3, O1 e F8 dos grupos TEA e NTEA.

A autocorrelação é uma métrica que avalia a relação linear entre os valores de uma série temporal em diferentes momentos no tempo. Em essência, ela proporciona informações sobre a existência de dependências entre os valores anteriores e futuros na série, revelando como os padrões temporais se manifestam. Na Figura 15, no eixo horizontal do gráfico, utilizamos o símbolo τ para indicar o número de intervalos de tempo entre os pontos que estão sendo comparados. No eixo vertical os valores de autocorrelação são apresentados, variando dentro de um intervalo aproximado entre -0,4 e 0,27.

A interpretação desses valores é crucial para entender o comportamento da série temporal. Valores positivos, como 0,27, sugerem uma correlação positiva, indicando que períodos de altos valores anteriores estão associados a altos valores subsequentes, e o oposto para valores negativos, como -0,4, que apontam para uma correlação negativa. Este intervalo de variação não apenas fornece informações sobre as relações temporais nos dados, mas também pode indicar a presença de padrões sazonais ou cíclicos na série que foram observados em $\tau \approx 1.000$. Vale ressaltar que valores dentro desse intervalo são considerados estatisticamente significativos enquanto aqueles fora dele podem ser menos confiáveis.

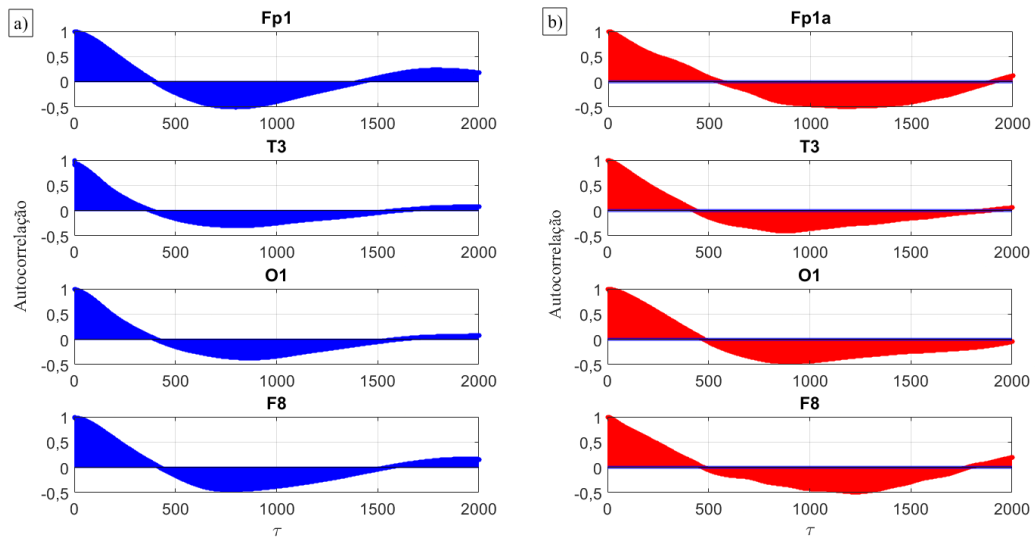


Figura 15 – Autocorrelação da amostra: a) grupo TEA; b) grupo NTEA.

Nesta base de dados, também foi realizada a análise dos seguintes caracterizadores: ec , C , Ca , L , Δ , E e Mo em relação a k para os grupos TEA e NTEA, considerando seus 16 eletrodos, conforme as Equações 2.3, 2.6, 2.7, 2.10, 2.12, 2.14 e 2.15, respectivamente. Novamente, calculado as médias aritméticas e os erros em relação aos valores de k na rede (conforme mostrado na Figura 16).

Observam-se diferenças significativas nas curvas representativas de cada grupo, especialmente nos caracterizadores Ca , Mo , L , C e E . Nos caracterizadores Ca , Mo e E nota-se que as curvas apresentaram valores mais acentuados para o grupo TEA. No entanto, para L e C observamos dois cenários distintos. Para o primeiro, valores mais acentuados na curva do grupo TEA até aproximadamente $k \approx 10^1$, após o qual ocorre uma inversão. Por outro lado, para o outro caracterizador mencionado, valores maiores em toda a variação de k para o grupo NTEA. É importante destacar que, no caso de E , houve uma maior dispersão nas barras de erro.

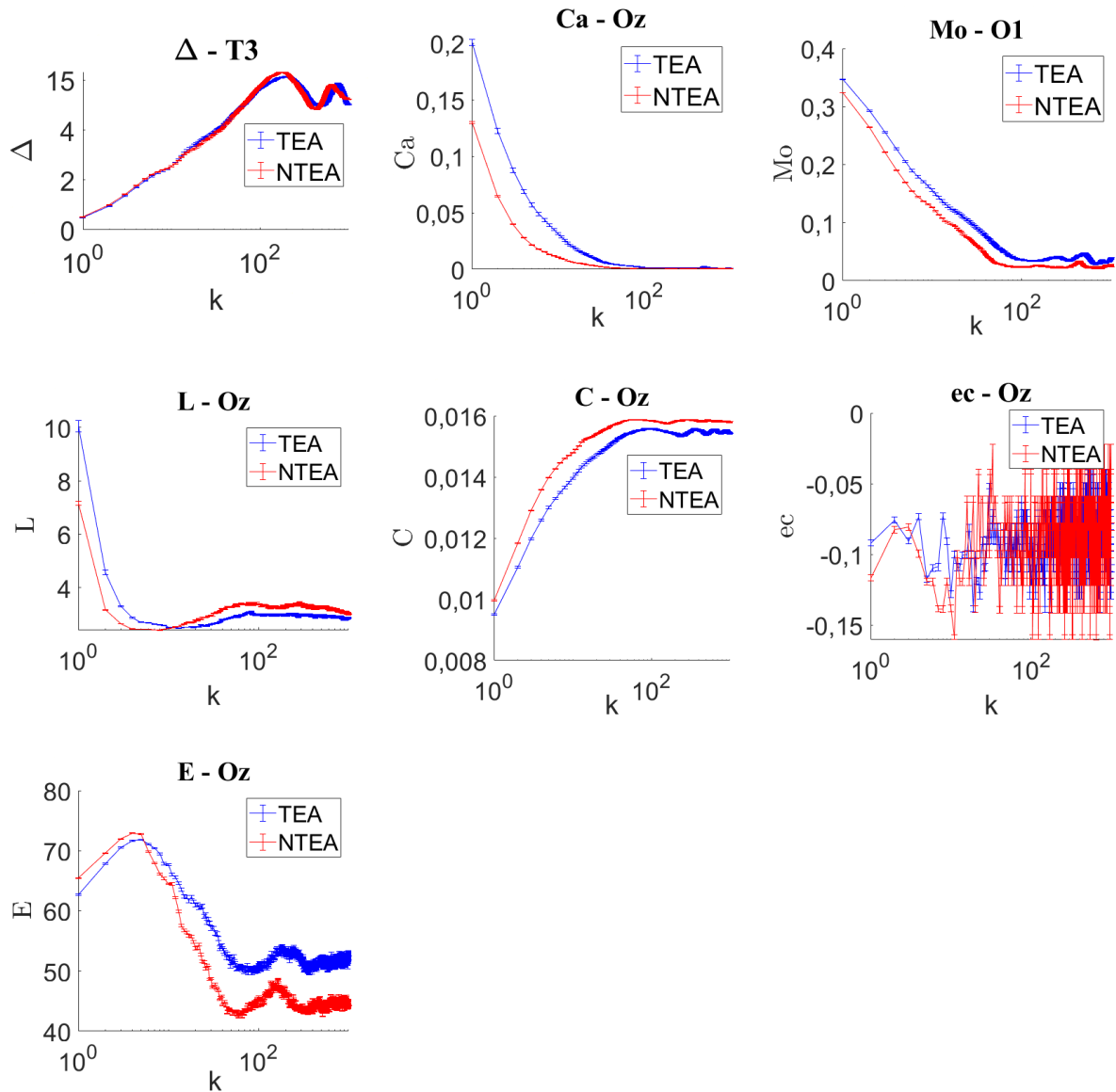


Figura 16 – Gráficos com os valores de variação da média em relação a k para cada caracterizador com cálculo do erro.

As Tabelas 3 e 4 apresentam os pontos de corte (k) que resultaram nos valores máximos nas áreas abaixo da curva ROC para cada um dos caracterizadores analisados, fornecendo uma visão clara do quão bem esses caracterizadores conseguem diferenciar os grupos TEA e NTEA. É importante destacar que o limiar crítico para esta base de dados é de 0,64844, e todos os valores máximos observados estão consistentemente acima desse limite, chegando até mesmo a atingir um limiar ideal de 1,0 para o caracterizador *ec*. Isso evidencia a capacidade excepcional dos caracterizadores em distinguir com sucesso os pacientes com TEA dos pacientes NTEA.

Na Figura 17, representações gráficas sob a forma de boxplots para os caracterizadores Δ , *Ca*, *Mo*, *L*, *C*, *ec*, e *E* em relação aos eletrodos que obtiveram as pontuações mais altas nas áreas abaixo da curva ROC nos grupos TEA e NTEA. Essas pontuações foram associadas a valores de k específicos, sendo eles 602, 387, 763, 319, 948, 759, e 888, respectivamente. É importante notar

Tabela 3 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.

Canais	Δ	k	Ca	k	Mo	k	L	k
Fp1	0,78125	658	0,77344	380	0,78906	638	0,75781	261
F3	0,85938	590	0,85156	764	0,89844	587	0,88281	594
F7	0,71875	604	0,85156	709	0,75781	649	0,70312	637
T3	0,92188	602	0,78125	321	0,75781	647	0,77344	7
T5	0,73438	871	0,78906	603	0,80469	755	0,81250	513
O1	0,78906	257	0,88281	375	0,93750	763	0,91406	313
C4	0,85938	323	0,78125	754	0,85156	633	0,84375	827
Fp2	0,77344	651	0,79688	337	0,75781	211	0,82031	643
Fz	0,77344	639	0,82813	637	0,86719	662	0,82031	587
F4	0,80469	349	0,77344	702	0,74219	752	0,74219	353
F8	0,75000	750	0,71094	116	0,71875	334	0,76563	621
C3	0,78906	300	0,77344	753	0,87500	590	0,77734	828
Cz	0,79688	292	0,78125	786	0,79688	791	0,88281	358
Pz	0,85156	740	0,82813	899	0,87500	594	0,80469	611
Oz	0,84375	733	0,92969	387	0,92188	571	0,92969	319
O2	0,80469	249	0,90625	66	0,86719	532	0,89844	730

Tabela 4 – Maiores valores de áreas abaixo de curva ROC dos caracterizadores para cada eletrodo.

Canais	C	k	ec	k	E	k
Fp1	0,71875	1	0,95313	700	0,88281	3
F3	0,78906	534	0,9375	737	0,85937	73
F7	0,64844	349	0,86719	292	0,79688	102
T3	0,64844	133	0,94531	298	0,82031	93
T5	0,78906	299	0,95313	841	0,875	986
O1	0,88672	322	0,92188	481	0,92188	513
C4	0,83594	357	1,0	759	0,875	514
Fp2	0,74219	109	0,91406	604	0,84375	3
Fz	0,85938	72	0,88281	777	0,89063	959
F4	0,72656	349	0,89063	81	0,75781	108
F8	0,74219	65	0,89063	438	0,78906	527
C3	0,73438	67	0,91406	530	0,82031	67
Cz	0,72656	351	0,85938	483	0,8125	761
Pz	0,78125	811	0,96094	281	0,89844	817
Oz	0,92188	948	0,96094	708	0,9375	888
O2	0,88281	361	0,90625	674	0,90625	773

que os pacientes em diferentes estados de saúde demonstram diferenças notáveis nas medianas, destacadas pelas linhas centrais nos gráficos de caixa. Na Tabela 5, esses valores são apresentados, com ênfase nas linhas correspondentes a Δ , L , ec , e E , onde foram identificadas as maiores discrepâncias nas medianas, indicadas em amarelo.

Com base nos dados obtidos de 16 canais de EEG, foi conduzido um estudo para avaliar o desempenho dos caracterizadores de redes complexas previamente introduzidos no contexto da distinção entre pacientes com TEA e NTEA. Para cada caracterizador específico e em relação a

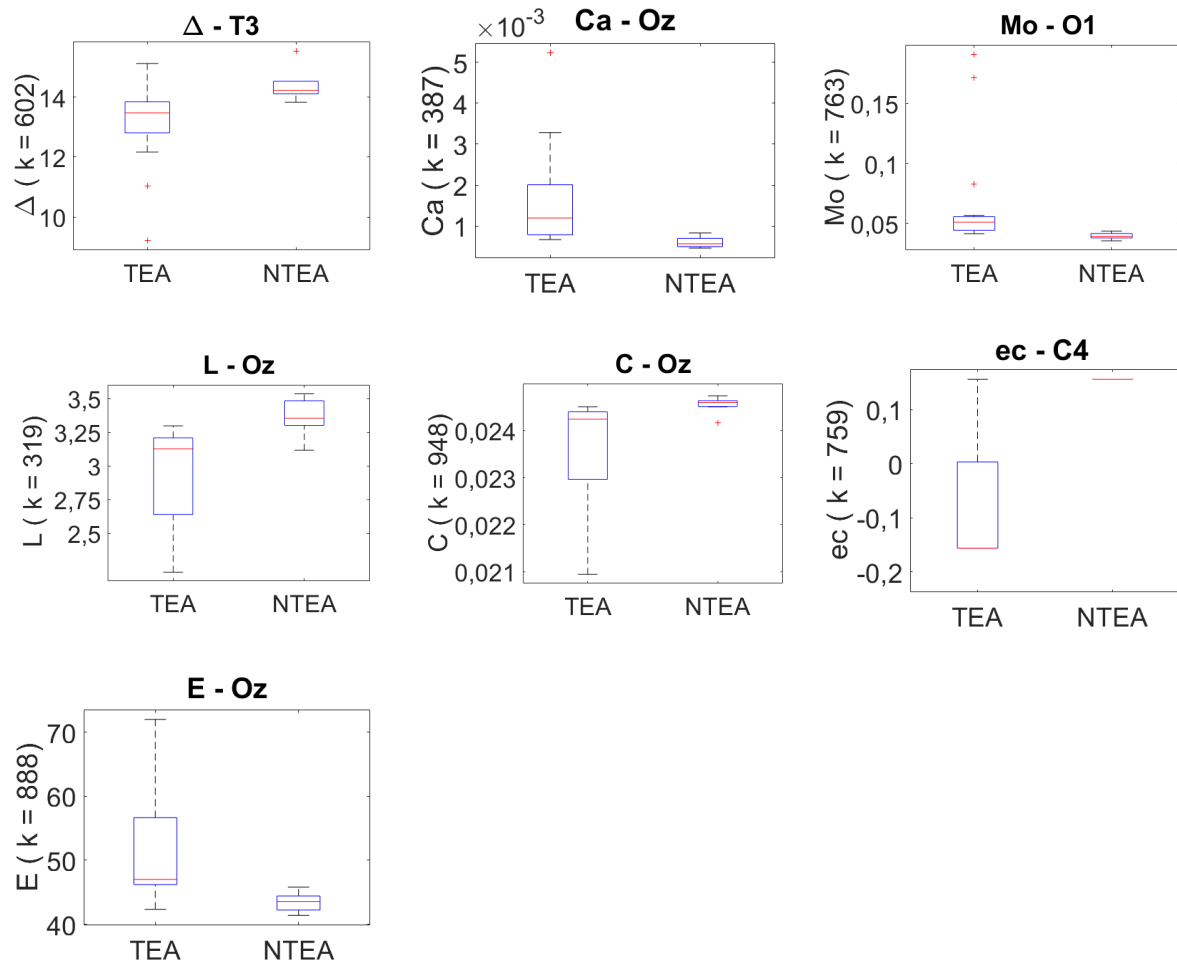


Figura 17 – Boxplots construídos para os maiores valores atingidos de k na área abaixo da curva ROC.

Tabela 5 – Tabela de valores das medianas representadas nos boxplots da Figura 17 para os Grupos TEA e NTEA.

Caracterizador	TEA	NTEA
Δ	13,4745	14,2266
Ca	0,001198	0,00056866
Mo	0,051087	0,039284
L	3,1265	3,3549
c	0,02424	0,024598
ec	-0,15598	0,15614
E	47,007	43,5663

cada canal de EEG, bem como às combinações dos grupos TEA e NTEA, calculamos a média das pontuações das áreas abaixo da curva ROC. Isso permitiu a identificação do canal de EEG que apresentou a maior precisão na diferenciação entre pacientes com o TEA e indivíduos neurotípicos. No geral, o mapeamento realizado e consequentemente os caracterizadores utilizados, demonstraram ser eficazes na distinção dos dois grupos. O caracterizador topológico mais eficaz foi o ec , com valores de curva ROC atingindo 1,0 na análise representada na Figura 18 pelo eletrodo C4. Além disso, observou-se uma proeminência de valores mais elevados nas curvas

ROC dos eletrodos da região occipital.

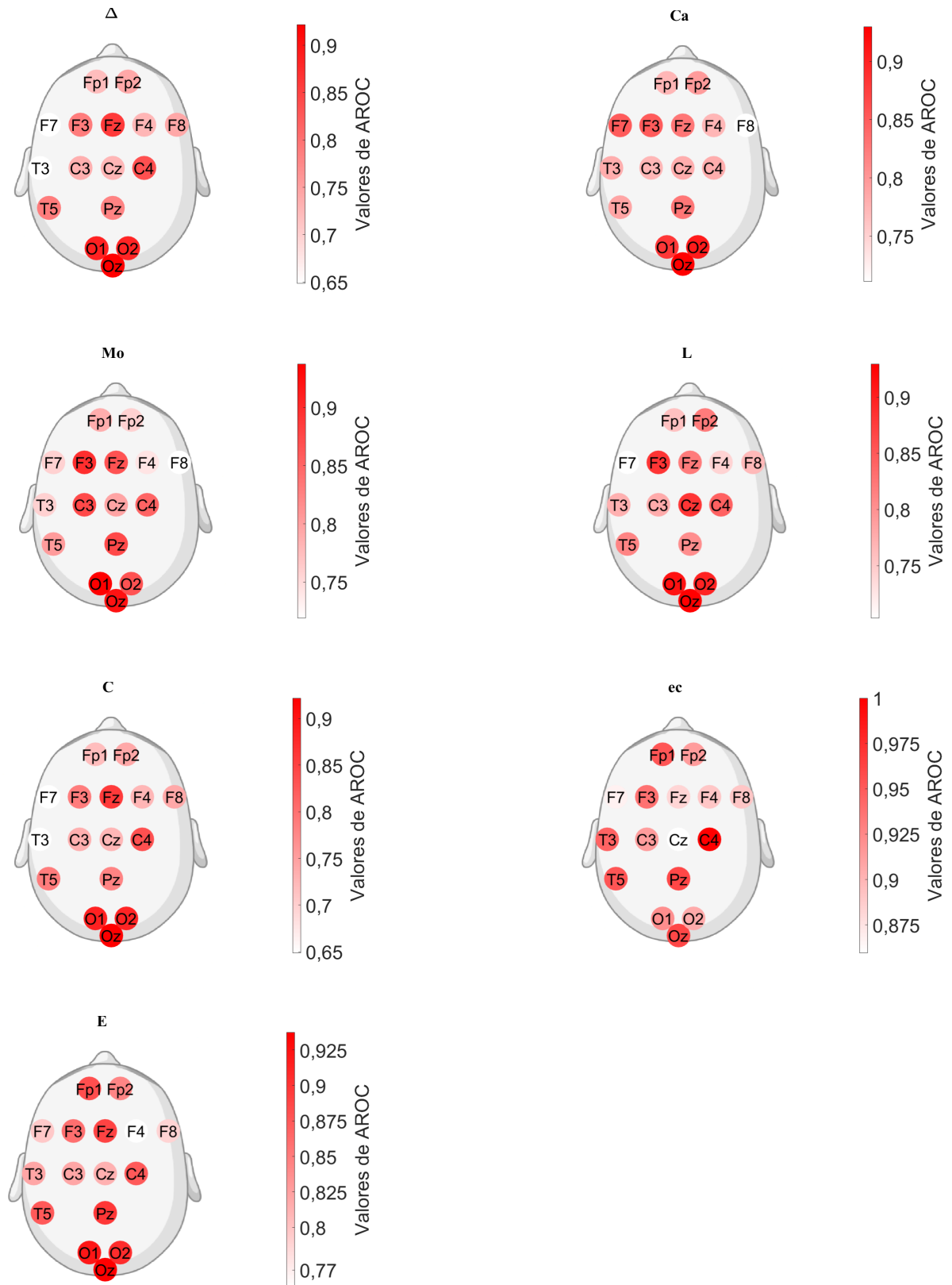


Figura 18 – Ranqueamento dos eletrodos, coloridos de acordo com seu valor abaixo da área de curva ROC para cada um dos caracterizadores utilizados.

Por fim, para otimizar o modelo, os sinais e os eletrodos que apresentaram os resultados mais consistentes em cada base de dados. O algoritmo SVM foi aplicado considerando três caracterizadores: ec , L e Δ . Inicialmente, os pacientes foram selecionados de forma que todas as combinações possíveis fossem consideradas, resultando em um total de 16.384 combinações testadas. Em cada iteração desse processo foi realizada a permutação dos grupos de teste para garantir que todos os grupos fossem avaliados exatamente uma vez.

Ao final desse processo iterativo, os valores de Acu foram calculados com base nos acertos e erros do classificador utilizando a Equação 4.1. O resultado geral foi uma acurácia de 96,86%, como ilustrado na Figura 19. Esse resultado reflete a capacidade do modelo em classificar os pacientes com TEA e NTEA com uma precisão significativa.

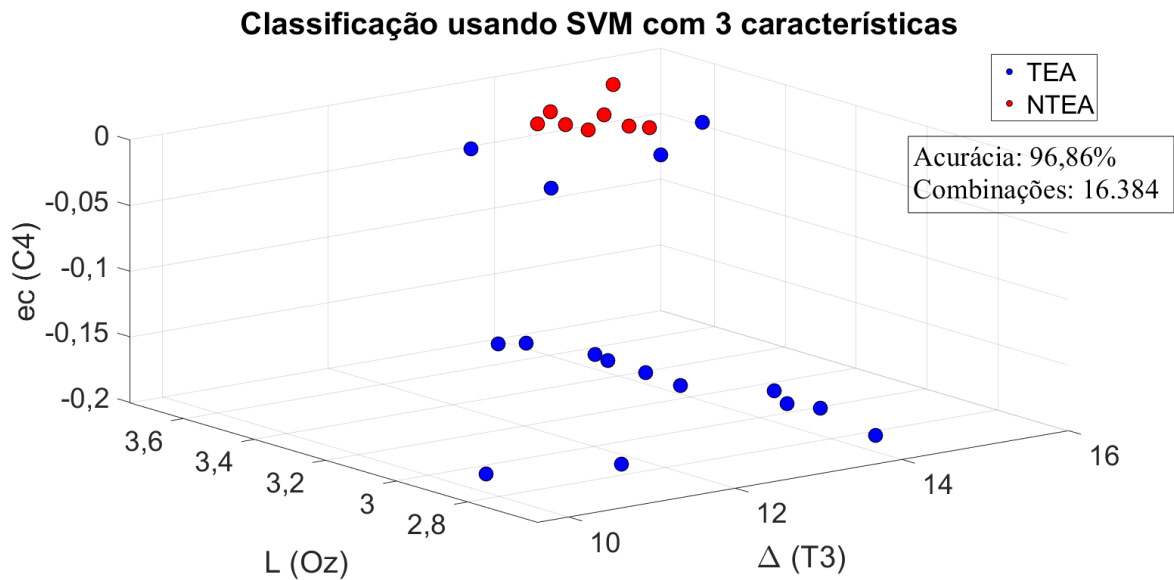


Figura 19 – Classificação usando SVM com 3 caracterizadores testando todas as combinações possíveis, com Acu de 96,86%.

A classificação utilizando o SVM, resultou na seguinte matriz de confusão acumulada:

Tabela 6 – Matriz de confusão acumulada.

		Previsão	
		TEA	NTEA
Verdade	TEA	32.768	2.048
	NTEA	0	30.720

5.1.1 Estudo da divisão das ondas cerebrais

A Figura 20 representa a divisão das ondas após a aplicação do filtro passa-banda nos intervalos correspondentes a δ (0.5 a 4 Hz), θ (4 a 8 Hz), α (8 a 13 Hz) e β (13 a 30 Hz), correspondentes a um canal de EEG (Fp1 = Frontal).

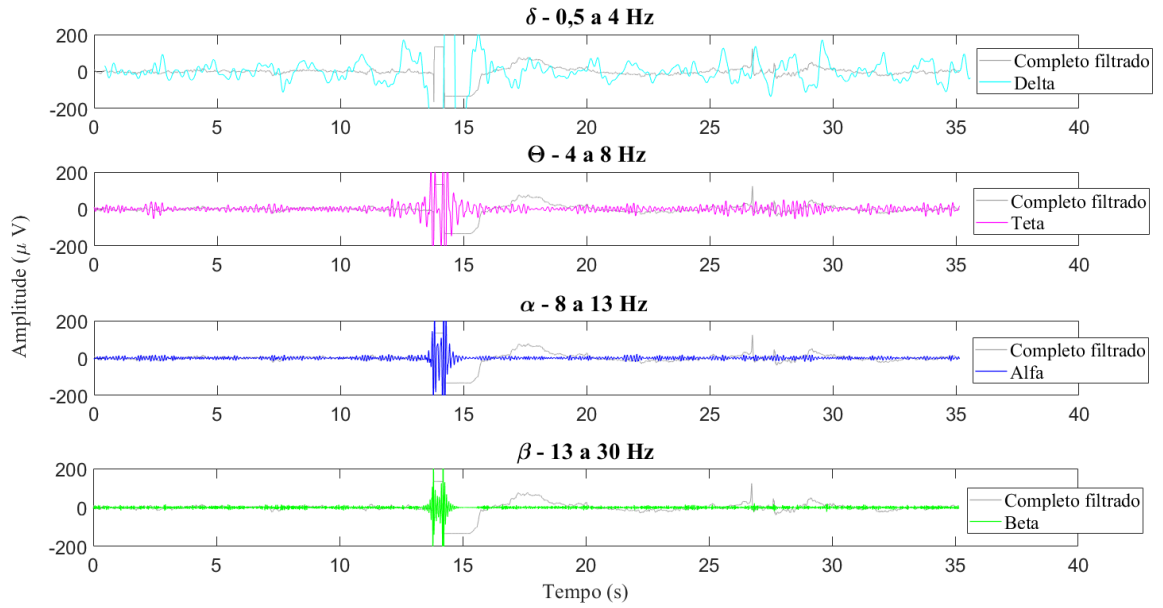


Figura 20 – Exemplos dos sinais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelo eletrodo Fp1 e subdivididos nas ondas δ , θ , α e β .

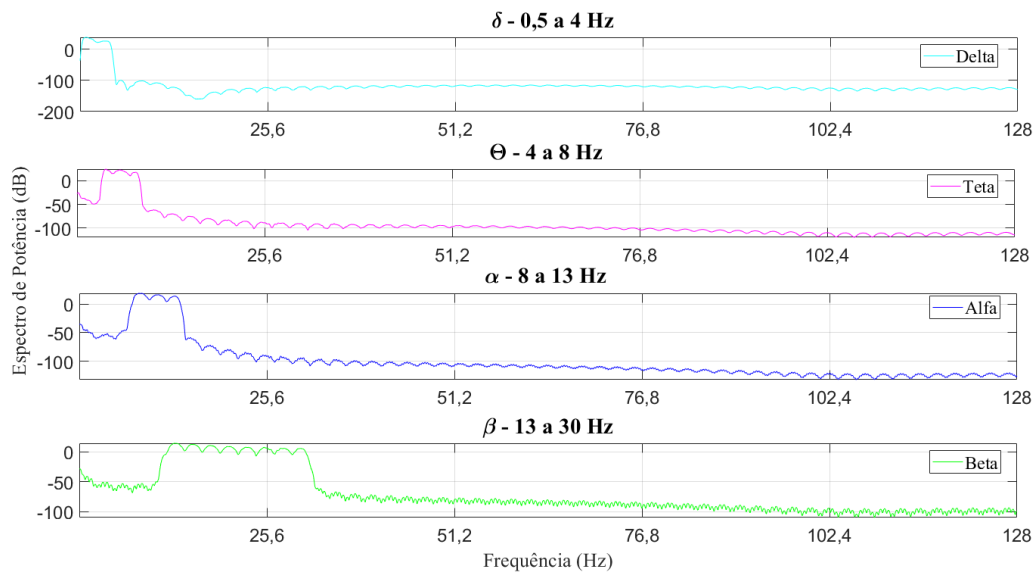


Figura 21 – Espectro de potência dos sinais do eletrodo Fp1 para o grupo TEA nas faixas de frequências correspondentes a δ , θ , α e β .

Na Figura 21, a eficácia do filtro aplicado em cada faixa de frequência correspondente para as divisões δ , θ , α e β torna-se evidente. Isso é perceptível pelas notáveis mudanças nas amplitudes das frequências representadas no espectro. Além disso, ao observar a figura, podemos notar que, à medida que avançamos para frequências fora da faixa especificada, a quantidade de energia diminui. Esse comportamento sugere que o filtro está agindo de forma eficaz, reduzindo ou eliminando componentes fora da faixa de frequência em estudo.

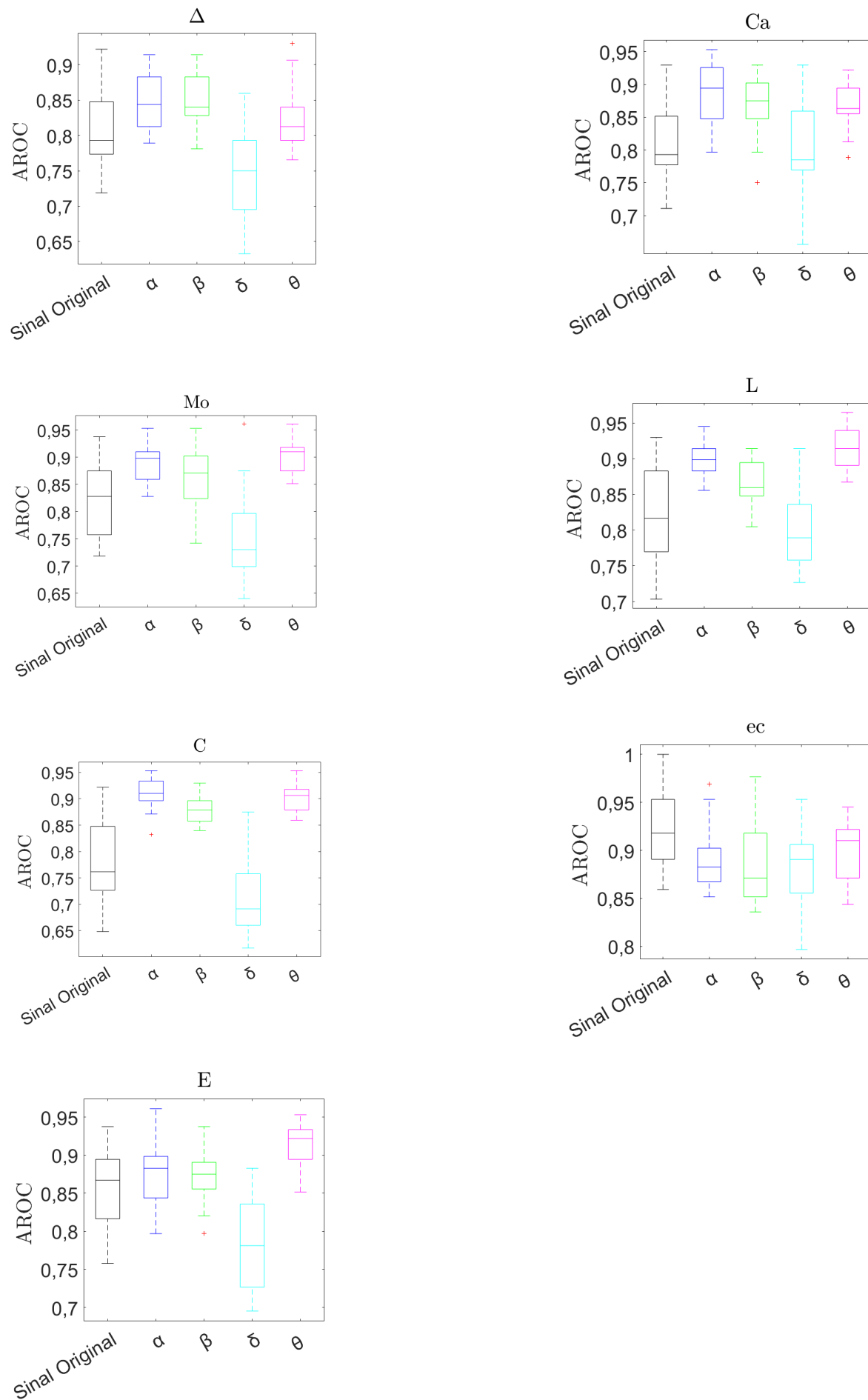


Figura 22 – Boxplots construídos para cada carecterizador, de cada frequência de onda estudada com base nos seus valores de área abaixo da curva ROC.

Tabela 7 – Tabela de valores das medianas representadas nos boxplots da Figura 17 para os Grupos TEA e NTEA.

Caracterizador	Completa	α	β	δ	θ
Δ	0,7930	0,8437	0,8398	0,7500	0,8125
Ca	0,7930	0,8945	0,87500	0,78515	0,8633
M_o	0,8281	0,8984	0,8711	0,7305	0,9101
L	0,8164	0,8984	0,8594	0,7890	0,9141
c	0,7617	0,9102	0,8789	0,6914	0,9062
ec	0,9197	0,8828	0,8711	0,8906	0,9101
E	0,8671	0,8828	0,87500	0,7812	0,9219

Na Figura 22, são apresentadas representações gráficas na forma de boxplots para os caracterizadores Δ , Ca , M_o , L , C , ec , e E em relação aos eletrodos que alcançaram as pontuações mais elevadas nas áreas abaixo da curva ROC nos grupos TEA e NTEA. Vale ressaltar que a faixa de frequência α e θ exibem diferenças notáveis nas medianas, destacadas pelas linhas centrais nos gráficos de caixa e destacadas numericamente na Tabela 7.

5.2 Base de dados de Sheffield

Dado que cada série temporal possui um tamanho de $T = 73.000$ pontos, utilizamos $Q = 2 \times (73.000)^{\frac{1}{3}}$. Consequentemente, mapeamos os 384 sinais de EEG em um total de 384.000 grafos de quantis ($A_k = 384.000$). Desta forma, obtivemos 384.000 matrizes de transição de Markov, cada uma com $Q^2 = 7.056$ elementos.

As Figuras 24 e 26 exibem exemplos de sinais de EEG selecionados aleatoriamente de indivíduos, representados nas Figuras 23 e 25 após a aplicação do filtro passa-banda (0, 5 a 30 Hz), correspondentes a cada canal de EEG (Fp1 = Frontal, C3 = Central, O2 = Occipital, P10 = Parietal). É importante observar que, em todos os casos, os sinais de EEG de pacientes com o TEA apresentam diferenças em relação aos pacientes NTEA.

Novamente, observa-se a relevância da aplicação do filtro na segunda base de dados. Na Figura 27 a eficácia do filtro se torna evidente ao observarmos as mudanças notáveis nas amplitudes das frequências representadas no espectro. Além disso, ao examinar a figura é possível perceber que à medida que se avança para frequências mais elevadas, a quantidade de energia diminui, indicando que o filtro está agindo para reduzir ou eliminar componentes de alta frequência que possam conter ruído ou informações menos relevantes para a análise em questão.

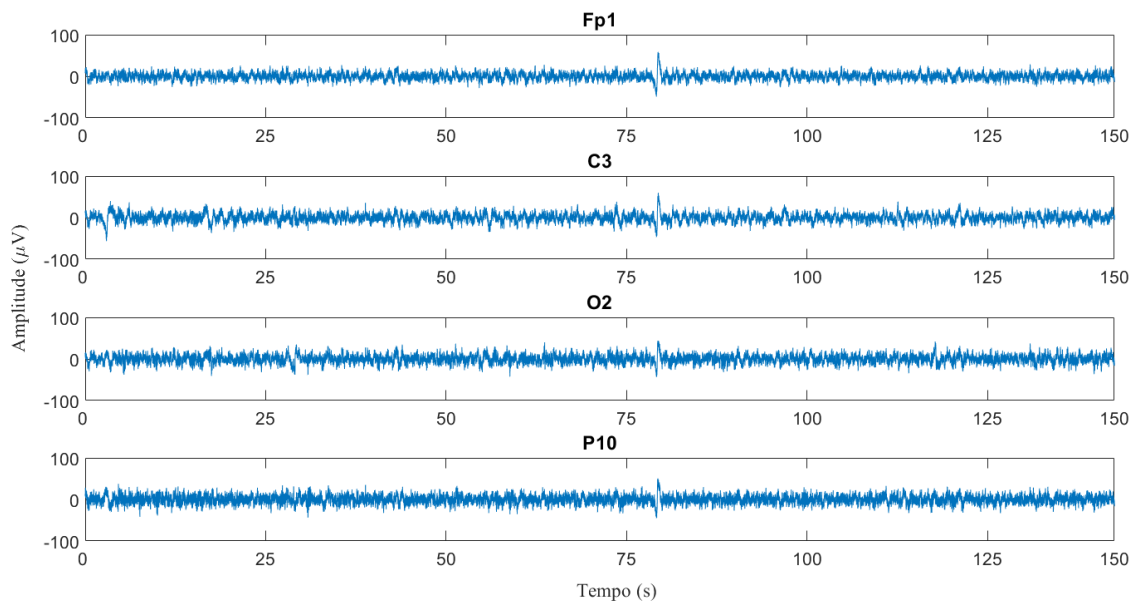


Figura 23 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, C3, O2 e P10.

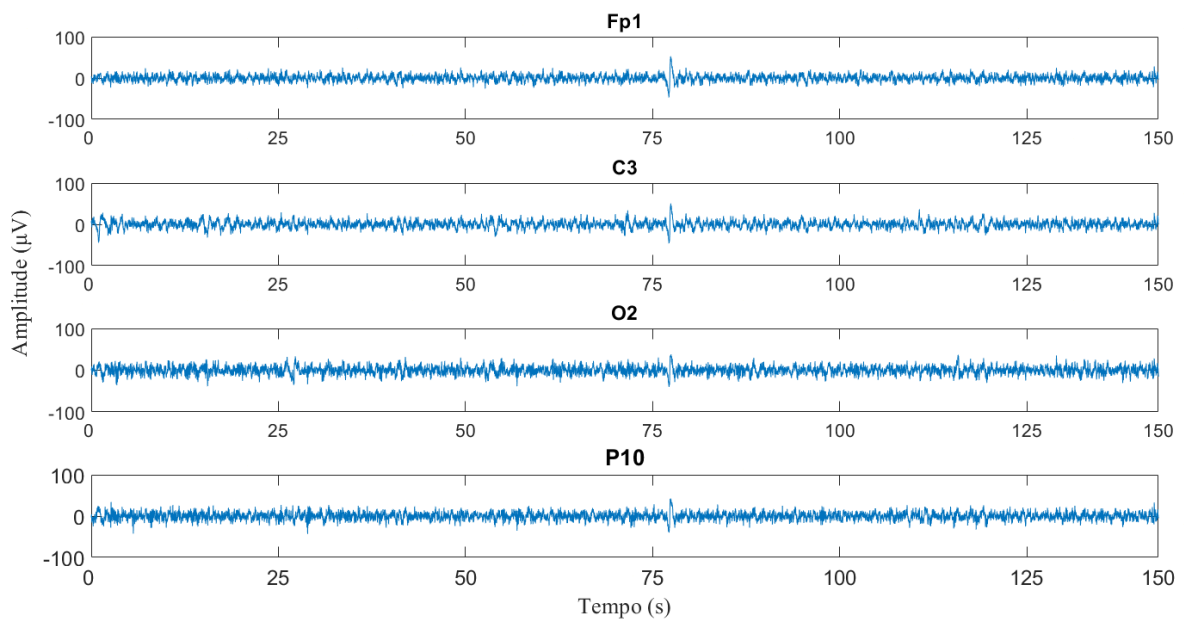


Figura 24 – Exemplificação dos sinais do paciente do grupo TEA representados na Figura 23 após aplicação do filtro passa-banda (0,5 a 30 Hz).

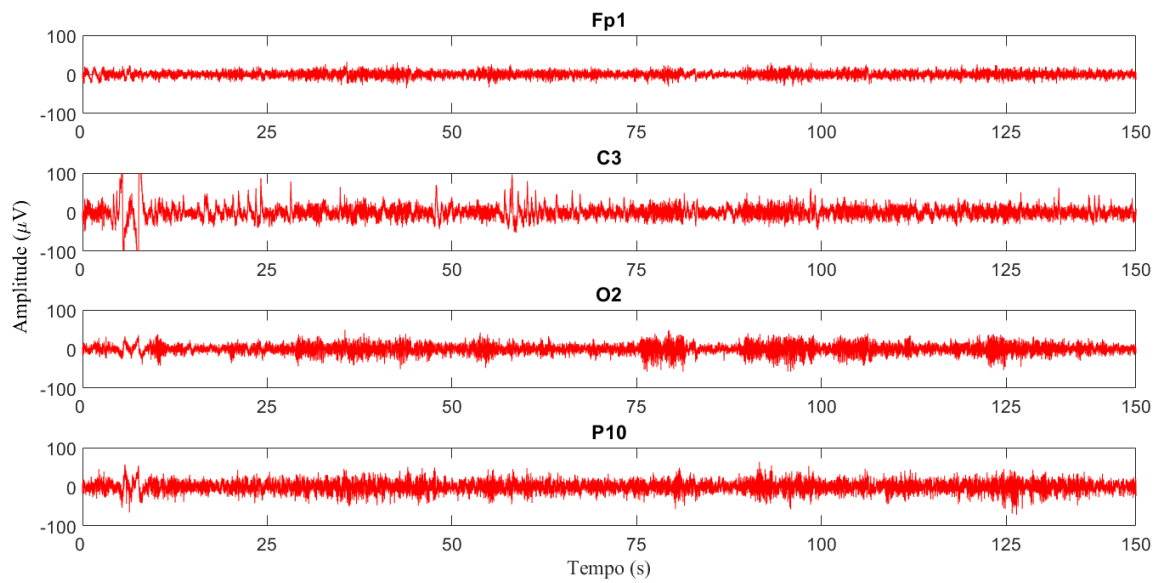


Figura 25 – Exemplos dos sinais originais de EEG de um paciente escolhido de maneira aleatória do grupo NTEA. Os sinais exemplificados foram obtidos pelos eletrodos Fp1, C3, O2 e P10.

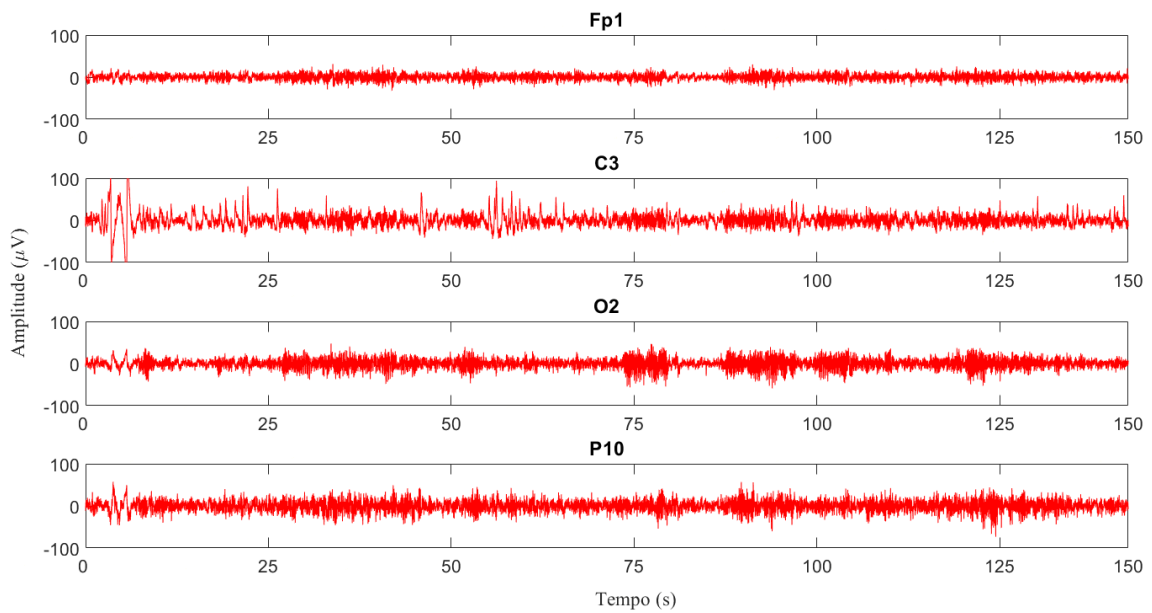


Figura 26 – Exemplificação dos sinais do paciente do grupo NTEA representados na Figura 25 após aplicação do filtro passa-banda (0,5 a 30 Hz).

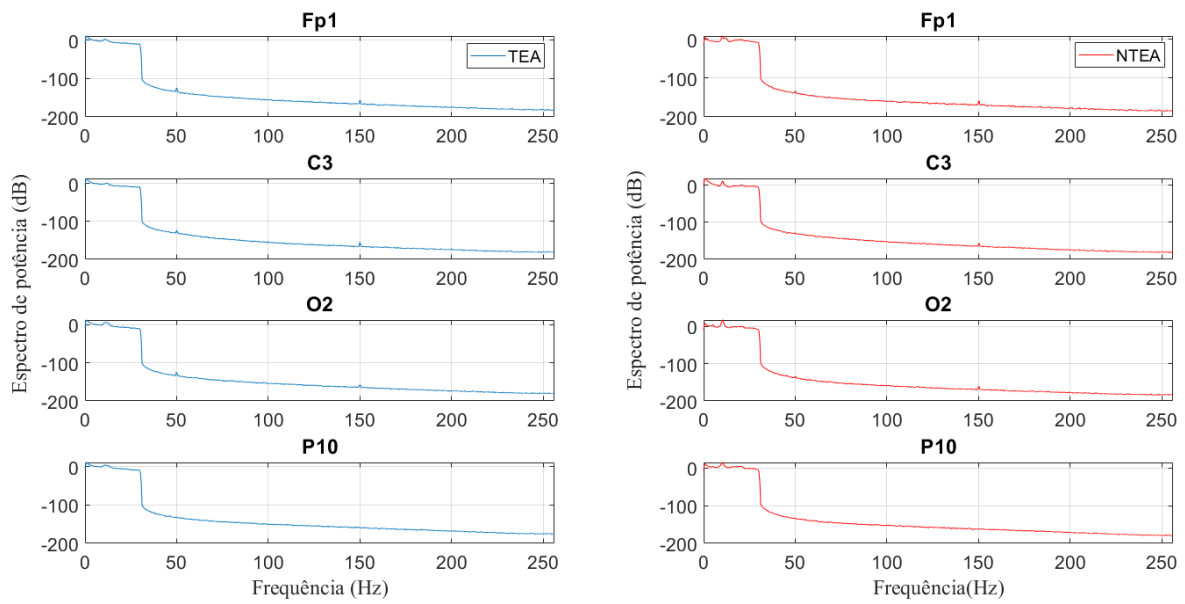


Figura 27 – Espectro de potência dos sinais dos eletrodos Fp1, C3, O2 e P10 dos grupos TEA e NTEA.

Para a autocorrelação, no eixo vertical da Figura 28 observa-se uma variação de valores aproximados a -0,3 e 0,2, valores positivos nesse intervalo indicam uma correlação positiva entre os dados em atrasos temporais específicos, enquanto valores negativos apontam para uma correlação negativa. Valores próximos aos extremos indicam correlações mais fortes. No eixo horizontal do gráfico, o símbolo τ para indicar o número de intervalos de tempo entre os pontos que estão sendo comparados. A presença de padrões sazonais ou cíclicos na série foram observados em $\tau \approx 1.250$.

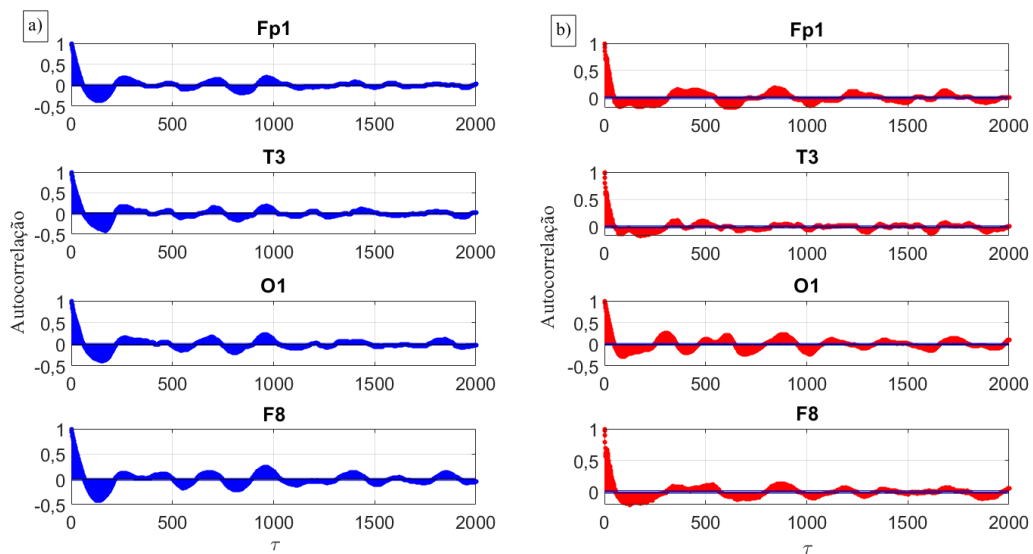


Figura 28 – Autocorrelação da amostra: a) grupo TEA; b) grupo NTEA.

Para os grupos TEA e NTEA, bem como para os 8 canais de EEG, foram calculados

diversos caracterizadores, a saber: ec , C , Ca , L , Δ , E e Mo em relação a k , utilizando as Equações 2.3, 2.6, 2.7, 2.10, 2.12, 2.14 e 2.15, respectivamente. Em seguida, para cada paciente em ambos os grupos, calculamos a média aritmética e o erro em função dos valores de k na rede (Figura 29).

Observaram-se diferenças significativas entre as curvas representativas de cada grupo, especialmente nos caracterizadores Δ , L e ec . Notavelmente, para os grupos NTEA, essas curvas apresentaram um pico mais acentuado em comparação com o grupo TEA, ocorrendo em valores de k superiores a 10 para os caracterizadores Δ e L . Para o caracterizador ec , as curvas começaram a divergir já a partir de 1. Vale ressaltar que no caso de ec houve uma maior dispersão nas barras de erro.

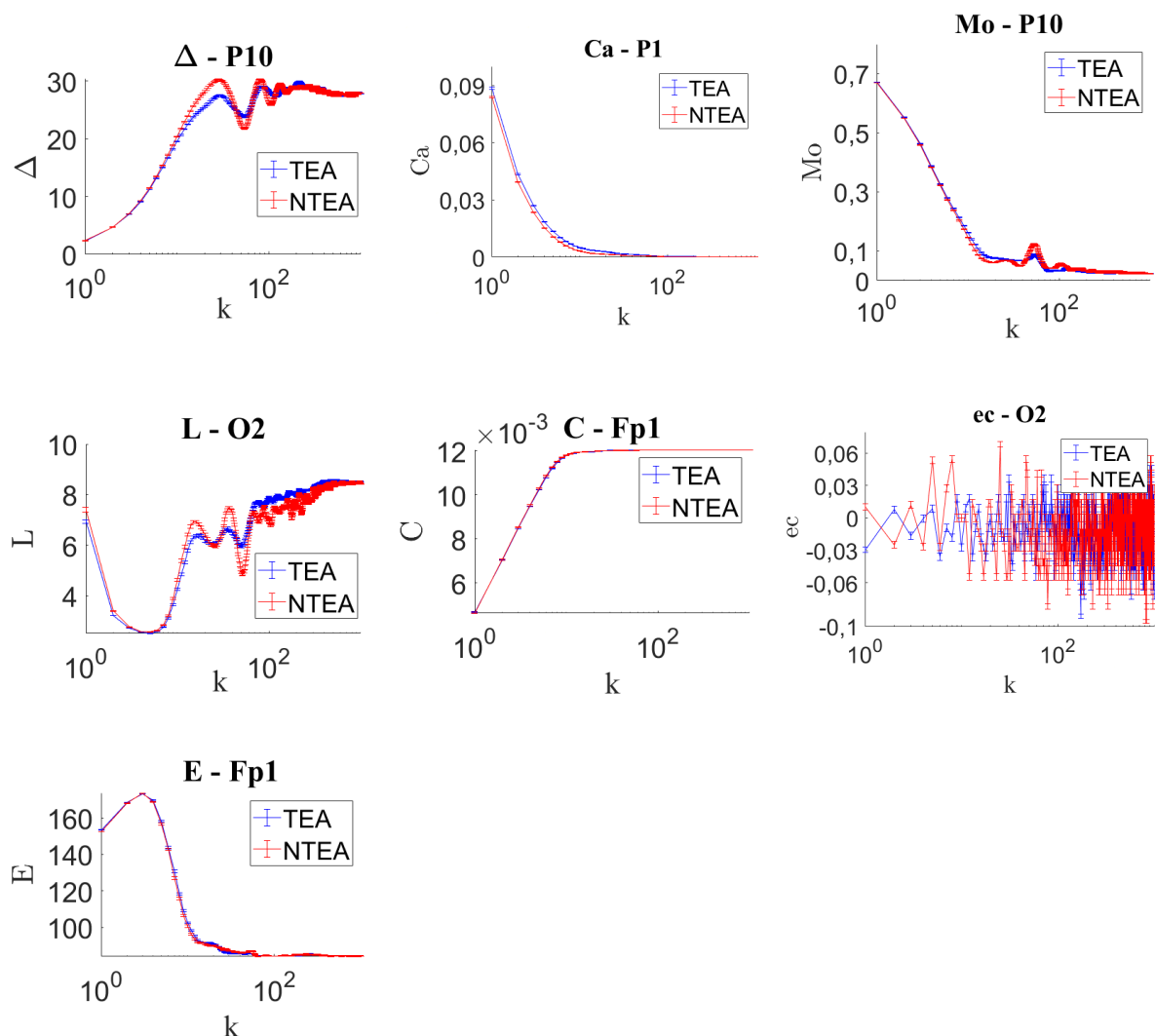


Figura 29 – Gráficos com os valores de variação da média em relação a k para cada caracterizador com cálculo do erro.

As áreas abaixo da curva ROC foram meticulosamente calculadas para cada um dos grupos de estudo. As Tabelas 8 e 9 destacam os pontos de corte (k) que resultaram nos maiores valores de área da curva ROC para cada um dos caracterizadores analisados, oferecendo uma

Tabela 8 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.

Eletrodos	Δ	k	Ca	k	Mo	k	L	k
C3	0,755208	633	0,731771	554	0,692708	542	0,638021	157
F7	0,75	823	0,710938	278	0,757812	604	0,71875	604
Fp1	0,752604	906	0,783854	906	0,794271	899	0,755208	826
Fz	0,747396	354	0,710938	430	0,747396	222	0,708333	36
O2	0,708333	354	0,791667	941	0,78125	219	0,760417	466
P1	0,739583	985	0,794271	471	0,731771	979	0,713542	946
P10	0,773438	599	0,747396	734	0,822917	753	0,716146	873
Poz	0,757812	361	0,716146	950	0,763021	344	0,747396	102

Tabela 9 – Maiores valores de áreas abaixo da curva ROC dos caracterizadores para cada eletrodo.

Eletrodos	C	k	ec	k	E	k
C3	0,695312	522	0,81901	173	0,740885	531
F7	0,742188	557	0,825521	311	0,761719	135
Fp1	0,820312	735	0,768229	199	0,790365	735
Fz	0,779948	721	0,815104	109	0,766927	721
O2	0,777344	368	0,833333	810	0,768229	396
P1	0,798177	583	0,789062	355	0,773438	583
P10	0,768229	798	0,815104	456	0,776042	585
Poz	0,792969	104	0,778646	244	0,78125	301

clara perspectiva do desempenho discriminatório desses caracterizadores em relação aos grupos TEA e NTEA.

O que merece destaque e é altamente significativo é que todos esses valores máximos da área abaixo da curva ROC para os grupos de estudo são consistentemente superiores a um limiar crítico de 0.638021. Essa constância de resultados acima desse limiar é notável e revela a eficácia robusta dos caracterizadores e do modelo de classificação na tarefa de diferenciação entre os grupos TEA e NTEA. Essa consistência é um indicativo poderoso de que o modelo tem uma capacidade sólida de distinguir com confiabilidade os pacientes com o TEA dos pacientes NTEA, o que é de suma importância em contextos clínicos e de pesquisa.

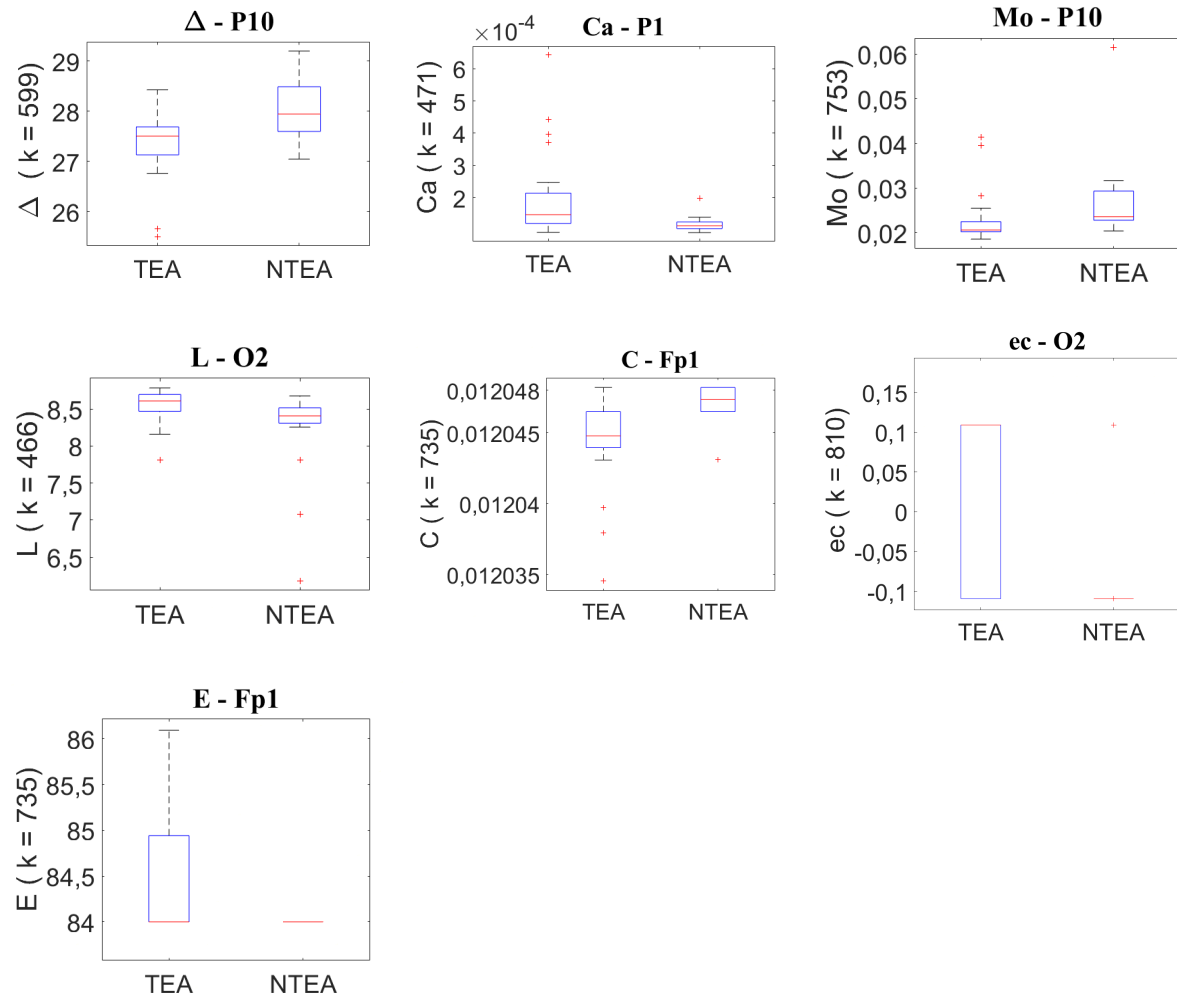


Figura 30 – Boxplots construídos para os maiores valores atingidos de k na área abaixo da curva ROC.

Na Figura 30, apresento representações gráficas em forma de boxplots para os caracterizadores Δ , Ca , Mo , L , C , ec , e E em relação aos eletrodos que exibiram os maiores valores na área abaixo da curva ROC nos grupos TEA e NTEA, com valores de k correspondentes a 599, 471, 753, 466, 735, 810, e 735, respectivamente. É importante destacar que os pacientes em diferentes estados de saúde mostram diferenças notáveis nas medianas, evidenciadas pelas linhas centrais nos gráficos de caixa. Na Tabela 10 apresentamos esses valores, enfatizando em amarelo as linhas correspondentes a Δ e a L , onde foram identificadas as maiores discrepâncias nas medianas.

Com base nos dados provenientes de 8 canais de EEG, foi conduzido um estudo para avaliar o desempenho dos caracterizadores de redes complexas já previamente apresentados no contexto da distinção entre pacientes com TEA e NTEA. Para cada caracterizador específico, em relação a cada canal de EEG e às combinações dos conjuntos de TEA e NTEA, calculamos a média das áreas abaixo da curva ROC. Isso permitiu a identificação do canal de EEG que demonstrou a maior precisão na diferenciação entre os pacientes. No geral, o mapeamento realizado e por conseguinte os caracterizadores utilizados mostraram-se eficazes na distinção dos

Tabela 10 – Tabela de valores das medianas representadas nos boxplots da Figura 30 para os grupos TEA e NTEA.

Caracterizador	TEA	NTEA
Δ	27,5011	27,9418
Ca	0,012045	0,012046
M_o	0,020621	0,023596
L	8,6076	8,4067
c	0,012045	0,012047
ec	0,10911	0,10911
E	84	84

dois grupos. O Δ mostrou-se o caracterizador topológico mais eficaz, possuindo valor de curva ROC igual a 0,84 na análise representada na Figura 31 pelo eletrodo P1.

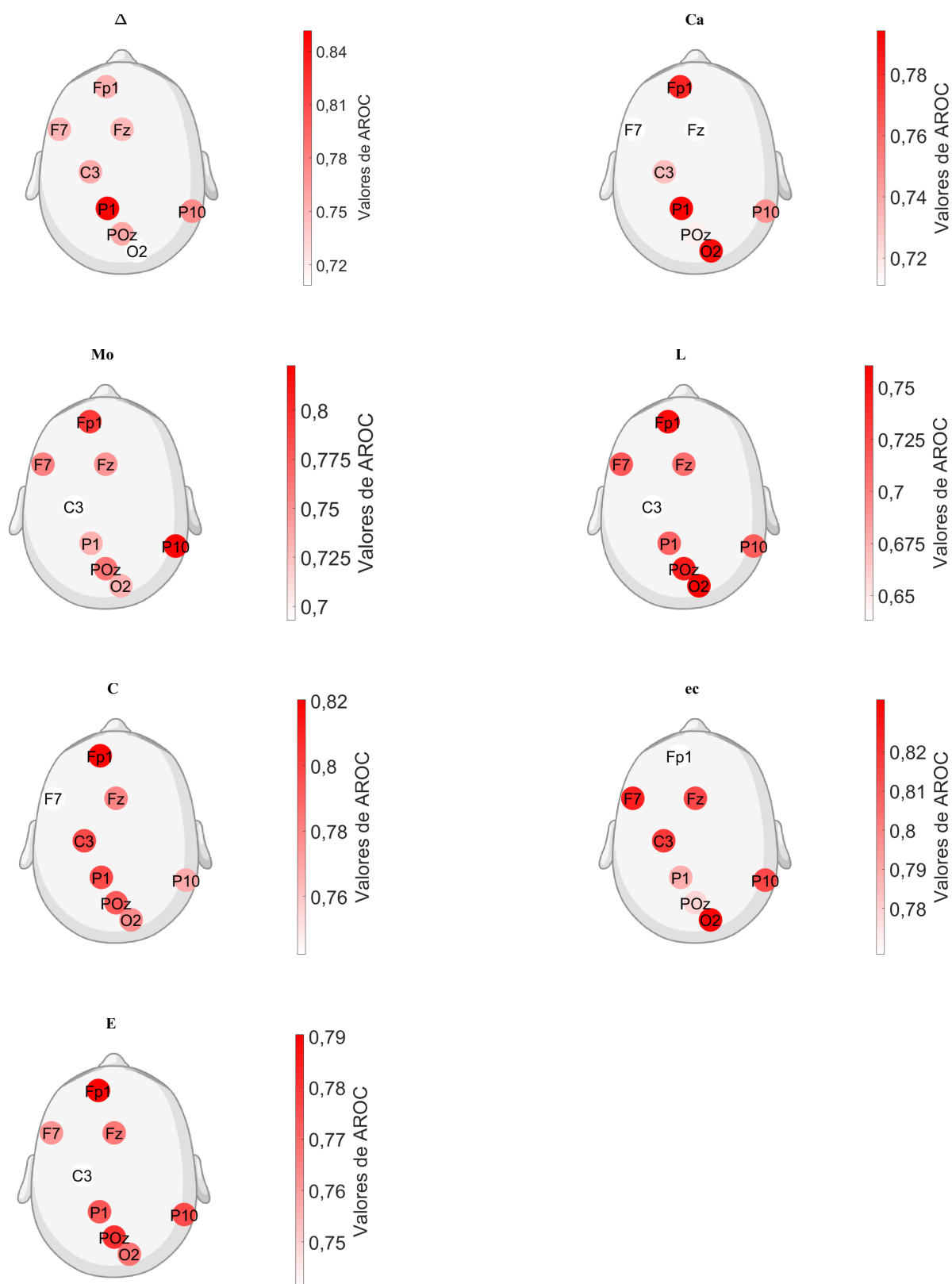


Figura 31 – Ranqueamento dos eletrodos, coloridos de acordo com seu valor de *Acu* para cada um dos caracterizadores utilizados.

A seguir, utilizando os sinais e os eletrodos que apresentaram os resultados mais consis-

tentes em cada técnica, foi aplicado o SVM com três caracterizadores: C , ec e Mo . Inicialmente, os pacientes foram selecionados de tal forma que todas as combinações possíveis fossem consideradas, resultando em um total de 147.456 combinações testadas. A cada iteração foi realizada a permutação dos grupos de teste, assegurando que todos os grupos fossem testados exatamente uma vez.

Ao final desse processo iterativo, os valores de Acu com base nos acertos e erros do classificador foram calculados utilizando a Equação 4.1. Isso resultou em uma acurácia geral de 70,83%, conforme ilustrado na Figura 32.

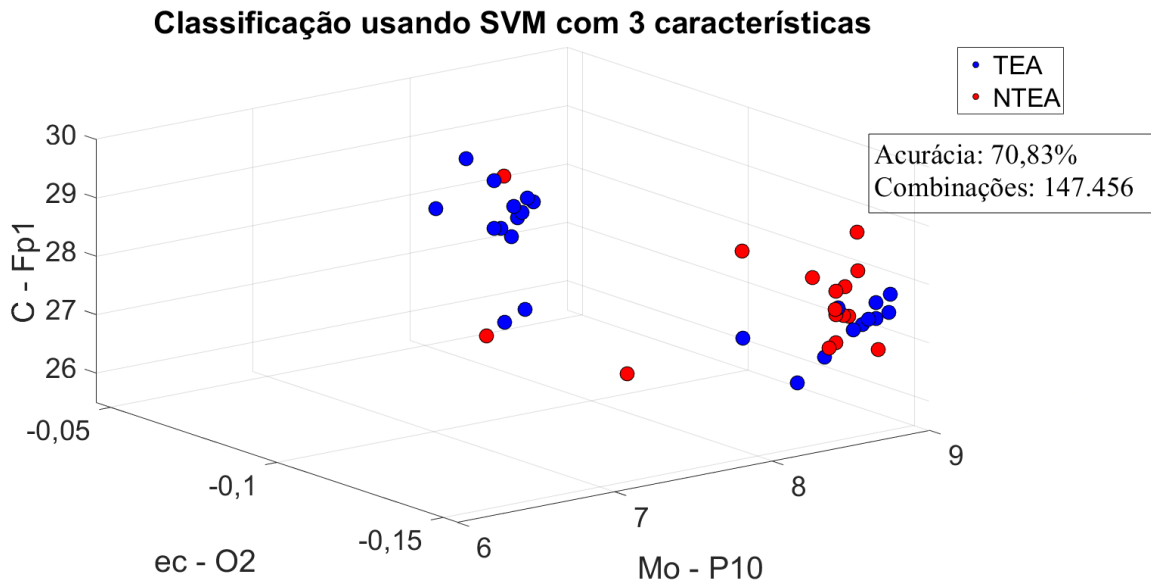


Figura 32 – Classificação usando SVM com 3 caracterizadores testando todas as combinações possíveis, com Acu de 70,83%.

A classificação utilizando o SVM resultou na seguinte matriz de confusão acumulada:

Tabela 11 – Matriz de confusão acumulada.

		Previsão	
		TEA	NTEA
Verdade	TEA	147.456	24.576
	NTEA	147.456	270.336

5.2.1 Estudo da divisão das ondas cerebrais

A Figura 33 representa a divisão das ondas após a aplicação do filtro passa-banda nos intervalos correspondentes a δ (0,5 a 4 Hz), θ (4 a 8 Hz), α (8 a 13 Hz) e β (13 a 30 Hz), correspondentes a um canal de EEG (Fp1 = Frontal).

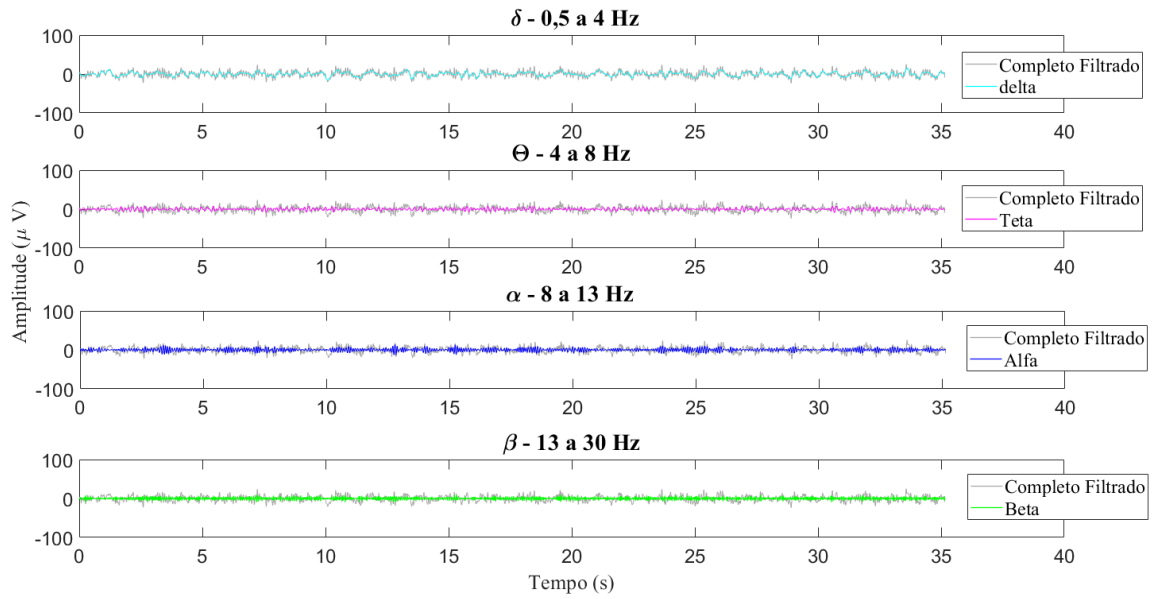


Figura 33 – Exemplos dos sinais de EEG de um paciente escolhido de maneira aleatória do grupo TEA. Os sinais exemplificados foram obtidos pelo eletrodo Fp1 e subdivididos nas ondas δ , θ , α e β .

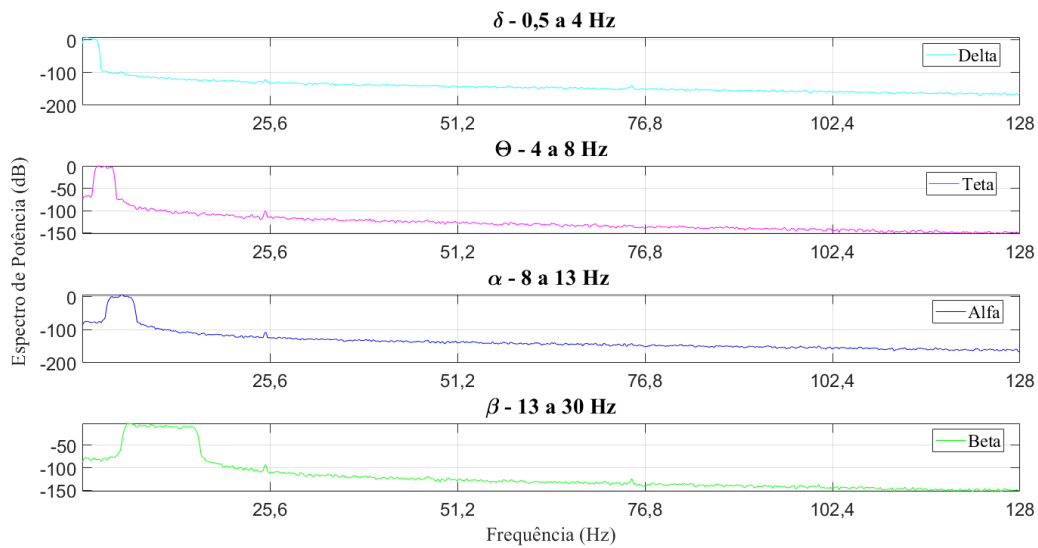


Figura 34 – Espectro de potência dos sinais do eletrodo Fp1 para o grupo TEA nas faixas de frequências correspondentes a δ , θ , α e β .

Na Figura 34, reiteramos a evidência da eficácia do filtro aplicado em cada faixa de frequência correspondente para as divisões δ , θ , α e β .

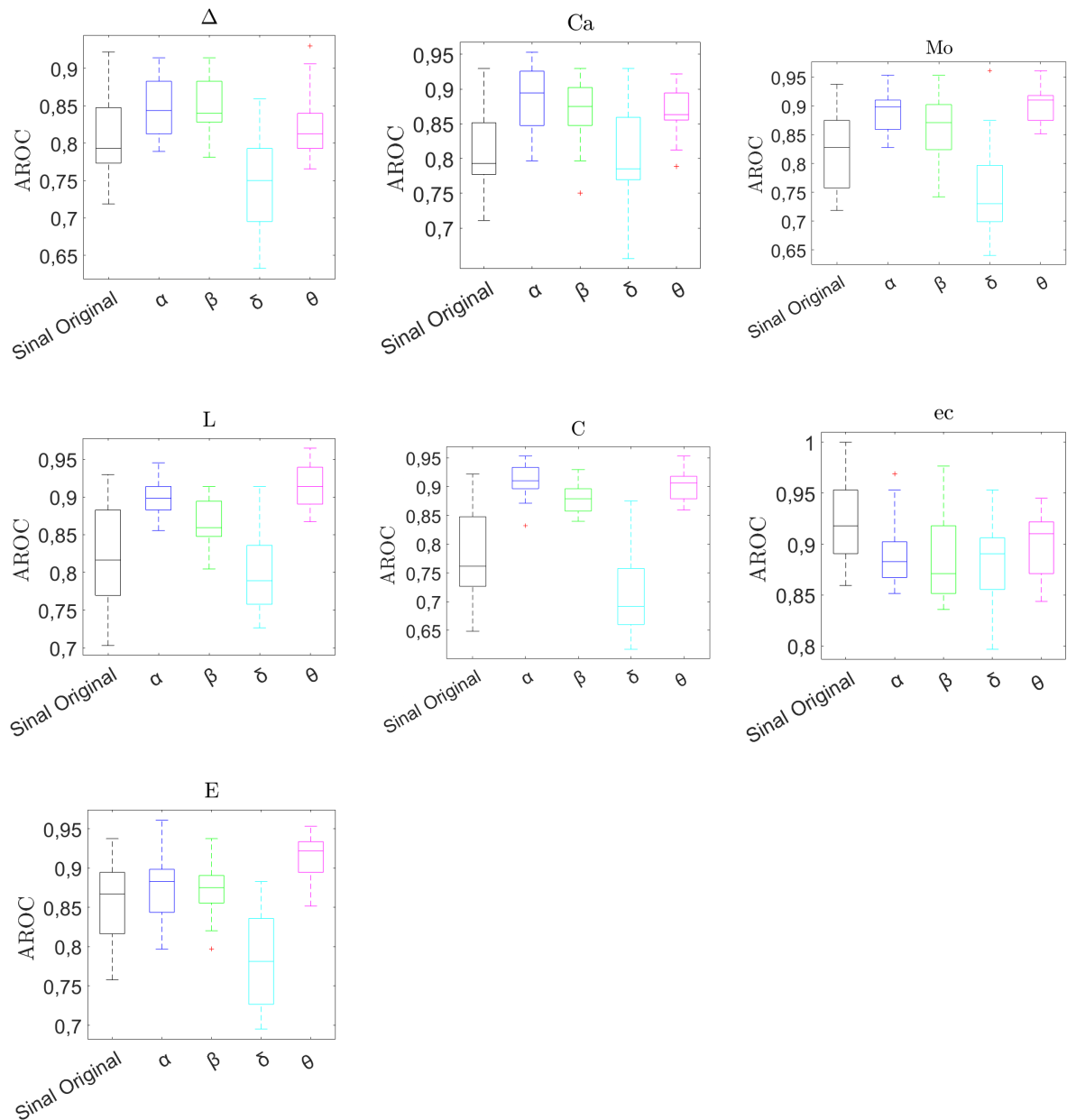


Figura 35 – Boxplots construídos para cada carecterizizador com base nos seus valores de área abaixo da curva ROC.

As representações gráficas na forma de boxplots para os caracterizadores Δ , Ca , Mo , L , C , ec , e E em relação aos eletrodos que obtiveram as pontuações mais elevadas nas áreas abaixo da curva ROC, nos grupos TEA e NTEA, reafirmam uma diferença notável para a onda α .

6 Conclusões

A técnica explorada neste estudo oferece uma abordagem promissora e não invasiva para o diagnóstico do TEA com base na análise de dados de EEG. Os resultados obtidos demonstram que essa técnica é estatisticamente significativa e eficaz na diferenciação de pacientes com TEA e NTEA.

A habilidade de discernir efetivamente entre esses grupos foi alcançada por meio da aplicação de todos os caracterizadores. No entanto, é notável que tanto na primeira quanto na segunda base de dados, os caracterizadores Δ e L se destacaram, indicando que essa metodologia apresenta um potencial significativo. Isso sugere que a técnica pode ser valiosa para facilitar o diagnóstico precoce e a implementação de intervenções oportunas no TEA futuramente.

Além disso, a técnica pode contribuir para uma melhor compreensão das características neurofisiológicas associadas ao TEA. Dada a complexidade do TEA, é desafiador afirmar com certeza qual região específica do cérebro é afetada, no entanto, os resultados obtidos na base de dados de King Abdulaziz, com maiores valores na região occipital, e na base de dados de Sheffield, na área frontal, estão em consonância com estudos previamente observados na literatura (CATARINO et al., 2011; LÉVEILLÉ et al., 2010; BARTTFELD et al., 2013). Isso sugere que a técnica pode proporcionar insights valiosos para a pesquisa e o tratamento desse transtorno complexo.

A análise das ondas cerebrais revelou, em ambas as bases de dados, uma marcante inclinação em direção às frequências α . Estudos indicam que, em comparação com indivíduos neurotípicos, aqueles com TEA podem apresentar alterações nas atividades das ondas α (SANTANA et al., 2021; DICKINSON et al., 2018).

Portanto, os resultados deste estudo ressaltam o potencial desta abordagem como uma ferramenta útil na análise de sinais cerebrais complexos e não lineares, como os obtidos por meio de EEGs em pacientes com TEA. Essa compreensão mais profunda das características neurofisiológicas do TEA pode, por sua vez, abrir caminho para abordagens mais eficazes no diagnóstico e no tratamento, representando um avanço importante no campo da neurociência clínica e na melhoria da qualidade de vida das pessoas afetadas pelo TEA.

Adicionalmente, a diferença na capacidade de classificação abre oportunidades significativas para aprimorar ainda mais a abordagem nos processos de pré-processamento de dados

e mapeamento, abrindo novas perspectivas para pesquisas futuras. Isso não apenas aprofunda nosso entendimento da aplicação da técnica atual, mas também estimula a exploração de como a metodologia pode ser otimizada e adaptada para desafios mais complexos e diversos na área da classificação e diagnóstico de transtornos como o TEA. Em última análise, essa distinção de resultados representa um passo fundamental em direção ao aprimoramento contínuo das técnicas de classificação.

Referências

- ABDULBAQI, A. S. et al. A hybrid technique for EEG signals evaluation and classification as a step towards to neurological and cerebral disorders diagnosis. *International Journal of Nonlinear Analysis and Applications*, Semnan University, v. 13, n. 1, p. 773–781, 2022. [4](#)
- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, APS, v. 74, n. 1, p. 47, 2002. [18](#)
- ALHADDAD, M. J. et al. Diagnosis autism by fisher linear discriminant analysis flda via EEG. *International Journal of Bio-Science and Bio-Technology*, Science and Engineering Research Support Society, v. 4, n. 2, p. 45–54, 2012. [15](#)
- AMARAL, L. A.; OTTINO, J. M. Complex networks: Augmenting the framework for the study of complex systems. *The European physical journal B*, Springer, v. 38, p. 147–162, 2004. [6](#), [7](#)
- BARTTFELD, P. et al. Organization of brain networks governed by long-range connections index autistic traits in the general population. *Journal of neurodevelopmental disorders*, BioMed Central, v. 5, n. 1, p. 1–9, 2013. [46](#)
- BONACICH, P. Some unique properties of eigenvector centrality. *Social networks*, Elsevier, v. 29, n. 4, p. 555–564, 2007. [11](#)
- BRITANNICA, E. Ilustração problema das sete pontes de euler. Universal Images Group Editorial, 2003. [vii](#), [6](#)
- CAMPANHARO, A. S.; DOESCHER, E.; RAMOS, F. M. Application of quantile graphs to the automated analysis of EEG signals. *Neural Processing Letters*, Springer, v. 52, p. 5–20, 2020. [4](#)
- CAMPANHARO, A. S.; RAMOS, F. M. Hurst exponent estimation of self-affine time series using quantile graphs. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 444, p. 43–48, 2016. [10](#), [13](#), [19](#)
- CAMPANHARO, A. S. et al. Duality between time series and networks. *PloS one*, Public Library of Science San Francisco, USA, v. 6, n. 8, p. e23378, 2011. [9](#), [19](#)
- CATARINO, A. et al. Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. *Clinical neurophysiology*, Elsevier, v. 122, n. 12, p. 2375–2383, 2011. [46](#)
- COSTA, D. R. T. d.; BARRETO, J. O. M.; SAMPAIO, R. B. Modelo teórico-metodológico baseado na teoria de redes complexas para análise da oferta potencial dos serviços de saúde. *Ciência & Saúde Coletiva*, SciELO Brasil, v. 26, p. 3791–3804, 2021. [9](#)
- 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. [13](#)
- COUTINHO, J. V. S. C.; BOSSO, R. M. do V. Gêmeos dizigóticos concordantes para autismo: relato de caso. *Revista Brasileira de Neurologia e Psiquiatria*, v. 20, n. 2, 2016. [2](#)
- DICKINSON, A. et al. Peak alpha frequency is a neural marker of cognitive function across the autism spectrum. *European Journal of Neuroscience*, Wiley Online Library, v. 47, n. 6, p. 643–651, 2018. [46](#)

- EULER, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, p. 128–140, 1741. 6
- FAGIOLO, G. Clustering in complex directed networks. *Physical Review E*, APS, v. 76, n. 2, p. 026107, 2007. 12
- FAYAZ, M. The bibliometric analysis of eeglab software in the web of science indexed articles. *Neuroscience Informatics*, Elsevier, p. 100154, 2023. 18
- FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. Uma introdução sucinta à teoria dos grafos. 2011. 9
- FILIPEK, P. A. et al. Practice parameter: screening and diagnosis of autism: report of the quality standards subcommittee of the american academy of neurology and the child neurology society. *Neurology*, AAN Enterprises, v. 55, n. 4, p. 468–479, 2000. 1
- FRAGA, S. F.; MONDRAGÓN, J. R. Comparativo dos algoritmos de dimensão fractal higuchi, katz e multi-resolução de contar as caixas em sinais EEG baseadas em potenciais relacionados por eventos. *Revista EIA*, n. 27, p. 73–83, 2017. 4
- GHOLAMI, R.; FAKHARI, N. Support vector machine: principles, parameters, and applications. In: *Handbook of neural computation*. [S.l.]: Elsevier, 2017. p. 515–535. 19
- GROSSI, E.; VALBUSA, G.; BUSCEMA, M. Detection of an autism EEG signature from only two eeg channels through features extraction and advanced machine learning analysis. *Clinical EEG and Neuroscience*, SAGE Publications Sage CA: Los Angeles, CA, v. 52, n. 5, p. 330–337, 2021. 4
- GUIMERA, R.; SALES-PARDO, M.; AMARAL, L. A. N. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, APS, v. 70, n. 2, p. 025101, 2004. 14
- GUTMAN, I.; LI, X.; ZHANG, J. Graph energy. *Analysis of Complex Networks: From Biology to Linguistics*, Wiley Online Library, p. 145–174, 2009. 14
- JIN, E. M.; GIRVAN, M.; NEWMAN, M. E. Structure of growing social networks. *Physical review E*, APS, v. 64, n. 4, p. 046132, 2001. 9, 10, 14
- KIM, Y. et al. Structural investigation of supply networks: A social network analysis approach. *Journal of operations management*, Elsevier, v. 29, n. 3, p. 194–211, 2011. 11, 12
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145. 20
- 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. 4
- LACASA, L. et al. The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion. *Europhysics Letters*, IOP Publishing, v. 86, n. 3, p. 30001, 2009. 4
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. *Physical review letters*, APS, v. 87, n. 19, p. 198701, 2001. 13
- LATORA, V.; NICOSIA, V.; RUSSO, G. *Complex networks: principles, methods and applications*. [S.l.]: Cambridge University Press, 2017. 7, 8

- LAVOR, M. D. L. S. S. et al. O autismo: aspectos genéticos e seus biomarcadores: uma revisão integrativa. *Brazilian Journal of Health Review*, v. 4, n. 1, p. 3274–3289, 2021. [2](#)
- LÉVEILLÉ, C. et al. Enhanced connectivity between visual cortex and other regions of the brain in autism: a rem sleep EEG coherence study. *Autism Research*, Wiley Online Library, v. 3, n. 5, p. 280–285, 2010. [46](#)
- LORD, C. et al. Autism spectrum disorder. *The lancet*, Elsevier, v. 392, n. 10146, p. 508–520, 2018. [1](#)
- MILNE, E. et al. Atypical EEG in autism spectrum disorder: Comparing a dimensional and a categorical approach. *Journal of abnormal psychology*, American Psychological Association, v. 128, n. 5, p. 442, 2019. [17](#), [18](#)
- MONTENEGRO, M. A. et al. *EEG na prática clínica*. [S.l.]: Thieme Revinter Publicações LTDA, 2018. [3](#)
- NEWMAN, M. *Networks*. [S.l.]: Oxford university press, 2018. [7](#), [8](#)
- PINEDA, A. M. et al. Quantile graphs for EEG-based diagnosis of alzheimer’s disease. *Plos one*, Public Library of Science San Francisco, CA USA, v. 15, n. 6, p. e0231169, 2020. [4](#)
- PSIQUIATRIA, A. A. D. Manual diagnóstico e estatístico de transtornos mentais. *Editora Artmed*, v. 5, 2014. [2](#), [3](#)
- RUHNAU, B. Eigenvector-centrality—a node-centrality? *Social networks*, Elsevier, v. 22, n. 4, p. 357–365, 2000. [11](#)
- SACREY, L.-A. R.; BENNETT, J. A.; ZWAIGENBAUM, L. Early infant development and intervention for autism spectrum disorder. *Journal of child neurology*, Sage Publications Sage CA: Los Angeles, CA, v. 30, n. 14, p. 1921–1929, 2015. [3](#), [4](#)
- SANTANA, J. P. M. de et al. Potência do ritmo alfa em pacientes com transtorno do espectro autista e indivíduos controle durante realização de tarefas cognitivas. *Revista Eletrônica Acervo Saúde*, v. 13, n. 3, p. e6909–e6909, 2021. [46](#)
- SARMUKADAM, K. et al. A review of the use of EEG connectivity to measure the neurological characteristics of the sensory features in young people with autism. *Reviews in the Neurosciences*, De Gruyter, v. 30, n. 5, p. 497–510, 2019. [4](#)
- TOMANIK, G.; CAMPANHARO, A. Estudo da dinâmica de epidemias em redes complexas. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 5, n. 1, 2017. [4](#)
- TUCHMAN, R.; RAPIN, I. Epilepsy in autism. *The Lancet Neurology*, Elsevier, v. 1, n. 6, p. 352–358, 2002. [2](#)
- VARANIS, M. V.; PEDERIVA, R. Detecção de correlação entre séries temporais utilizando espectro cruzado wavelet e coerência wavelet. In: SBMAC. *Conferência Brasileira de Dinâmica, Controle e Aplicações*. [S.l.], 2011. v. 10, p. 711–714. [4](#)
- VICCHIETTI, M. L. et al. Computational methods of EEG signals analysis for alzheimer’s disease classification. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 8184, 2023. [4](#)

WANG, Y. et al. Hypermethylation of the enolase gene (eno2) in autism. *European journal of pediatrics*, Springer, v. 173, p. 1233–1244, 2014. [2](#)

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. [13](#)

ZEIDAN, J. et al. Global prevalence of autism: A systematic review update. *Autism research*, Wiley Online Library, v. 15, n. 5, p. 778–790, 2022. [3](#)

ZHANG, J.; LUO, Y. Degree centrality, betweenness centrality, and closeness centrality in social network. In: ATLANTIS PRESS. *2017 2nd international conference on modelling, simulation and applied mathematics (MSAM2017)*. [S.l.], 2017. p. 300–303. [11](#)