



Investigação de Sinais Eletrofisiológicos de Processamento Facial como Biomarcadores na Perturbação do Espectro do Autismo: Um Estudo com Arbaclofeno

RICARDO JORGE FERREIRA DA ROCHA

julho de 2025

Investigação de Sinais Eletrofisiológicos de Processamento Facial como Biomarcadores na Perturbação do Espectro do Autismo: Um Estudo com Arbaclofeno

Ricardo Jorge Ferreira da Rocha

**Dissertação para obtenção do Grau de Mestre em
Engenharia Biomédica, Área de Especialização em
Engenharia Biomédica**

Orientador: Luís Coelho

Coorientador: João Castelhana

Coorientador: Andreia Pereira

Júri:

Presidente:

Joaquim Alves, Professor Adjunto, ISEP

Vogais:

Luís Coelho, Professor Adjunto, ISEP

Helena Catarina Pereira, Arguente, Universidade de Coimbra

Porto, julho de 2025

Resumo

A Perturbação do Espectro do Autismo (PEA) é uma condição do neurodesenvolvimento complexa, caracterizada por défices na comunicação social e padrões de comportamento restritos e repetitivos. Uma das hipóteses neurobiológicas centrais no autismo envolve alterações do equilíbrio excitação/inibição (E/I) cerebral, no sentido de maior excitação. Este estudo teve como objetivo avaliar o impacto do arbaclofeno, um agonista seletivo do recetor inibitório GABA-B (ácido gama-aminobutírico) e modulador do equilíbrio E/I em adultos (idades entre 19 e 52 anos) com e sem PEA durante uma tarefa de reconhecimento de faces. Para tal foram utilizados os componentes N170 e P100 de eletroencefalografia (EEG) como biomarcadores do processamento de informação facial.

Foram analisados dados de EEG de 70 participantes (29 com PEA e 41 controlos) num desenho experimental duplamente cego, aleatorizado, cruzado e controlado por placebo. Cada participante foi exposto a administração de três condições farmacológicas (placebo, 15 mg e 30 mg de arbaclofeno) durante uma tarefa de decisão percutual de deteção de faces. O pré-processamento de sinais EEG foi automatizado, e os componentes P100 e N170 foram identificados e quantificados. Os dados foram posteriormente submetidos a análises estatísticas para investigar os efeitos do arbaclofeno nos marcadores eletrofisiológicos, e a modelos de classificação baseados em algoritmos de *machine learning* para avaliar o potencial discriminativo desses marcadores na diferenciação entre grupos diagnósticos e condições experimentais.

As análises estatísticas revelaram que o arbaclofeno modulou a amplitude do P100 e N170 e a latência do N170, com efeitos dose-dependentes e assimetrias hemisféricas. Notavelmente, o fármaco prolongou a latência do N170 no hemisfério esquerdo exclusivamente no grupo de controlo, sugerindo uma resposta diferenciada no grupo com PEA. Contudo, os algoritmos de *machine learning* aplicados demonstraram capacidade discriminativa limitada para classificar participantes por grupo diagnóstico, dose farmacológica ou resposta comportamental.

Os resultados indicam que, embora o arbaclofeno modifique as respostas eletrofisiológicas, a elevada variabilidade interindividual e a sobreposição de dados limitam a utilização dos componentes P100 e N170 como biomarcadores diagnósticos ou de resposta terapêutica isolados na PEA. No entanto, os efeitos observados, nomeadamente as alterações dependentes da dose e assimetrias hemisféricas, evidenciam a sensibilidade destes sinais eletrofisiológicos a modulações farmacológicas e à sua potencial utilização como ferramenta para investigar os mecanismos de ação de fármacos.

Palavras-chave: Perturbação do Espectro do Autismo (PEA), Eletroencefalografia (EEG), N170, Arbaclofeno, Equilíbrio Excitação/Inibição, Processamento Facial.

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterised by deficits in social communication and restricted, repetitive patterns of behaviour. One of the central neurobiological hypotheses in autism involves alterations in the brain's excitation/inhibition (E/I) balance, in the sense of greater excitation. This study aimed to assess the impact of arbaclofen, a selective agonist of the inhibitory GABA-B (gamma-aminobutyric acid) receptor and modulator of the E/I balance in adults with and without ASD during a face recognition task. To this end, the N170 and P100 components of electroencephalography (EEG) were used as biomarkers of facial information processing.

EEG data from 70 participants (29 with PEA and 41 controls) were analysed in a double-blind, randomised, crossover, placebo-controlled experimental design. Each participant was exposed to three pharmacological conditions (placebo, 15 mg and 30 mg of arbaclofen) during a perceptual decision task of face detection. EEG signal preprocessing was automated, and the P100 and N170 components were identified and quantified. The data were subsequently subjected to statistical analyses to investigate the effects of arbaclofen on electrophysiological markers, and to classification models based on machine learning algorithms to assess the discriminatory potential of these markers in differentiating between diagnostic groups and experimental conditions.

Statistical analyses revealed that arbaclofen modulated the amplitude of P100 and N170 and the latency of N170, with dose-dependent effects and hemispheric asymmetries. Notably, the drug prolonged N170 latency in the left hemisphere exclusively in the control group, suggesting a differentiated response in the group with ASD. However, the applied machine learning algorithms demonstrated limited discriminatory ability to classify participants by diagnostic group, pharmacological dose, or behavioural response.

The results indicate that, although arbaclofen modifies electrophysiological responses, high interindividual variability and data overlap limit the use of the P100 and N170 components as isolated diagnostic or therapeutic response biomarkers in ASDICA. However, the observed effects, namely dose-dependent changes and hemispheric asymmetries, demonstrate the sensitivity of these electrophysiological signals to pharmacological modulations and their potential use as a tool to investigate drug mechanisms of action.

Keywords: Autism Spectrum Disorder (ASD), Electroencephalography (EEG), N170, Arbaclofen, Excitation/Inhibition Balance, Facial Processing.

Agradecimentos

A elaboração desta dissertação não poderia ter sido concluída sem o apoio e contribuição de diversas pessoas, às quais quero expressar o meu profundo agradecimento. O seu apoio e contributo foram fundamentais para a minha formação e para o desenvolvimento deste trabalho.

Ao professor Luís Coelho, meu orientador no ISEP, expresso o meu profundo agradecimento pela sua simpatia, disponibilidade e orientação ao longo deste processo. A sua dedicação e apoio foram essenciais para a realização deste trabalho.

Ao engenheiro João Castelhana, meu orientador do CIBIT, agradeço toda a ajuda e suporte fornecidos na compreensão dos métodos utilizados e na elaboração deste documento.

À investigadora Andreia Pereira, minha orientadora do CIBIT, agradeço todo o apoio e orientação prestados no aprofundamento do conhecimento teórico, nas análises realizadas e no aperfeiçoamento da escrita deste trabalho.

À Fundação para a Ciência e Tecnologia, ao *King's College London*, ao CIBIT e ICNAS da Universidade de Coimbra que possibilitaram a realização deste projeto de investigação.

A todos os docentes da Licenciatura de Engenharia Biomédica e colegas de curso que contribuíram para o sucesso do meu percurso académico.

À minha família, expresso o meu profundo agradecimento por terem estado ao meu lado durante toda esta jornada académica. Agradeço por acreditarem em mim e por serem uma fonte constante de motivação.

À minha namorada, por toda a compreensão e paciência demonstradas, e por todo o apoio intelectual e emocional que me deu ao longo da licenciatura e do mestrado.

A todos, o meu obrigado.

Financiamento

Os resultados que conduziram a esta tese receberam financiamento da Parceria Conjunta da Iniciativa de Medicamentos Inovadores 2 (IMI2-JU), ao abrigo do contrato de subvenção n.º 777394, no âmbito do projeto AIMS-2-TRIALS. Esta Parceria Conjunta conta com o apoio do programa de investigação e inovação Horizonte 2020 da União Europeia, bem como da EFPIA, AUTISM SPEAKS, Autistica e SFARI. Financiamento adicional incluiu *Clinical Research Associates*, L.L.C. (CRA), uma entidade afiliada à *Simons Foundation*, assim como apoio do *Institute for Translational Neurodevelopment at King's College London*, do *National Institute for Health Research (NIHR)*, do *Mental Health Biomedical Research Centre (BRC) at South London*, da *Maudsley NHS Foundation Trust*, e do *King's College London*. Este trabalho tem também o apoio da FCT no projeto com o seguinte código: CEECINST/00117/2021/CP2784/CT0001.

AUTISTICA
With knowledge comes understanding

imi | innovative
medicines
initiative

1 2 9 0

INSTITUTO DE
CIÊNCIAS NUCLEARES
APLICADAS À SAÚDE
UNIVERSIDADE DE
COIMBRA

NIHR | Maudsley Biomedical
Research Centre

 **autism
speaks**

 **CIBIT**
Coimbra Institute for Biomedical
Imaging and Translational Research

Clinical Research Associates, LLC
SIMONS FOUNDATION AUTISM RESEARCH INITIATIVE

efpia


AIMS-2-TRIALS
AUTISM RESEARCH FOR EUROPE

KING'S
College
LONDON

fct | Fundação
para a Ciência
e a Tecnologia

Índice

1	Introdução	19
1.1	Contextualização	19
1.2	Objetivos	20
1.3	Estrutura da Tese	21
2	Revisão da Literatura	23
2.1	Perturbação do Espectro do Autismo (PEA)	23
2.1.1	Sintomatologia	24
2.1.2	Diagnóstico	26
2.1.3	Intervenções	29
2.2	Neurobiologia da Perturbação do Espectro do Autismo (PEA)	32
2.2.1	Excitação e inibição na função cerebral	33
2.2.2	Potencial evocado N170 e processamento facial	35
2.2.3	Modulação do equilíbrio E/I com arbaclofeno	39
2.3	Pré-processamento de Eletroencefalografia (EEG)	42
2.3.1	Caraterísticas do Sinal EEG	44
2.3.2	Reamostragem	45
2.3.3	Filtragem	45
2.3.4	Deteção e Remoção de Canais Ruidosos	47
2.3.5	Interpolação dos Canais Removidos	47
2.3.6	Referenciação	48
2.3.7	Extração e Rejeição de Épocas	49
3	Metodologia	51
3.1	Materiais e Métodos	51
3.1.1	Participantes	51
3.1.2	Aquisições de EEG e Paradigma de Decisão Percetual	54
3.2	Pré-Processamento dos Dados	56
3.3	Identificação do N170 e P100	60
3.4	Organização e Análise dos Dados	61
3.5	Métodos de Análise Estatística	64
3.5.1	Análise de Variância (ANOVA)	64
3.5.2	Análise Post Hoc	66
3.6	Métodos de Análise com <i>Machine Learning</i>	66
4	Resultados	71
4.1	Pré-processamento dos Dados e Identificação dos picos	71
4.2	Análise dos ERPs Médios por Condição	74
4.3	Análise Estatística revela efeitos da intervenção	78

4.3.1	Amplitude do P100.....	79
4.3.2	Latência do N170	82
4.3.3	Amplitude do N170	83
4.3.4	Amplitude Pico a Pico.....	85
4.3.5	Diferença hemisférica na amplitude do N170	86
4.4	Análise com <i>Machine Learning</i>	87
4.4.1	Classificação Controlo vs PEA	87
4.4.2	Classificação de Dose.....	91
4.4.3	Classificação face vs não-face	94
5	Discussão	97
5.1	Discussão de Resultados	97
5.2	Limitações do Estudo.....	101
5.3	Trabalhos Futuros.....	101
6	Conclusões	103
	Competências Adquiridas.....	105
	Referências	107
	Anexos	116
	Pré-processamento	116
	Identificação de Picos e Organização de Dados	118
	Algoritmos de <i>Machine Learning</i>	135

Lista de Figuras

Figura 1 - Sequência de etapas de pré-processamento.....	44
Figura 2 - Tipos de filtros mais comuns (adaptado de <i>Peng, 2019</i>).....	46
Figura 3 - Esquema da tarefa realizada pelos participantes durante o ensaio clínico. A ordem de apresentação dos estímulos <i>Mooney</i> era aleatória.	55
Figura 4 - Sequência de etapas de pré-processamento implementada no <i>script</i>	57
Figura 5 - Estrutura dos diretórios após organização manual dos dados por dose.	62
Figura 6 - Segmento de 5 segundos do EEG do sujeito F101.....	71
Figura 7 - ERPs do sujeito F101 após extração de épocas e correção de <i>baseline</i> dos canais PO9 (a) e PO10 (b).	72
Figura 8 - ERPs do sujeito F101 após pré-processamento dos canais PO9 (a) e PO10 (b).	72
Figura 9 - Identificação dos componentes P100 e N170.	73
Figura 10 - ERP médio dos sujeitos do grupo de controlo (azul) e autismo (vermelho) na condição de placebo, quando identificaram faces, do hemisfério esquerdo (a) e direito (b). Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo) e canal PO10 (canal 32, hemisfério direito), com as respetivas áreas sombreadas indicando o erro padrão da média.	75
Figura 11 - ERP médio dos sujeitos do grupo de controlo (a) e autismo (b) na condição de placebo, quando identificaram ou não faces. Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo), com os ensaios em que o sujeito identificou face a azul-escuro e vermelho, e os ensaios que não identificou face a azul-claro e laranja, com as respetivas áreas sombreadas indicando o erro padrão da média.	76
Figura 12 - ERP médio dos sujeitos do grupo de controlo na toma do placebo (traçado verde), 15 mg de arbaclofeno (traçado amarelo) e 30 mg de arbaclofeno (traçado roxo), quando identificaram faces. Os traçados representam a atividade registada no canal 28 (PO9, hemisfério esquerdo, (a)) e no canal 32 (PO10, hemisfério direito, (b)), com as respetivas áreas sombreadas indicando o erro padrão da média.	77
Figura 13 - ERP médio dos sujeitos do grupo de PEA na toma do placebo (traçado verde), 15 mg de arbaclofeno (traçado amarelo) e 30 mg de arbaclofeno (traçado roxo), quando identificaram faces. Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo, (a)) e no canal PO10 (canal 32, hemisfério direito, (b)), com as respetivas áreas sombreadas indicando o erro padrão da média.	78
Figura 14 - Boxplots relativos à amplitude do P100 para a condição de identificação de face, para ambos os grupos e hemisférios.....	80
Figura 15 - Boxplots relativos à amplitude do P100 para a condição de não identificação de face, para ambos os grupos e hemisférios.....	81
Figura 16 - Boxplots relativos à latência do N170 para a condição de identificação de face, para ambos os grupos no hemisfério esquerdo.	83
Figura 17 - Boxplots relativos à amplitude do N170 para a condição de não identificação de face, para ambos os grupos no hemisfério esquerdo.	84

Figura 18 - Boxplots relativos à amplitude pico a pico para a condição de identificação de face, para ambos os grupos no hemisfério esquerdo.	86
Figura 19 - PCA de dois componentes para a predição de controlo vs autismo.	89
Figura 20 - PCA de dois componentes para a predição de controlo vs autismo apenas com os dados da latência do N170.	91
Figura 21 - PCA de dois componentes para a predição de dose farmacológica.	93
Figura 22 - PCA de dois componentes para a classificação comportamental.	95

Lista de Tabelas

Tabela 1 - Perfil demográfico da amostra de voluntários em estudo.....	52
Tabela 2 – Resumo dos algoritmos de <i>machine learning</i> desenvolvidos.....	67
Tabela 3 - Resultados dos modelos de <i>machine learning</i> de predição de grupo vs controlo. ...	87
Tabela 4 - Resultados dos modelos de <i>machine learning</i> de predição de grupo vs controlo apenas com os dados da latência do N170.	90
Tabela 5 - Resultados dos modelos de <i>machine learning</i> de predição de dose farmacológica.	92
Tabela 6 - Resultados dos modelos de <i>machine learning</i> de classificação comportamental. ...	94

Acrónimos e Símbolos

Lista de Acrónimos

ADDM	<i>Autism and Developmental Disabilities Monitoring</i>
ADI-R	<i>Autism Diagnostic Interview-Revised</i>
ADOS	<i>Autism Diagnostic Observation Schedule</i>
CAR	<i>Common Average Reference</i>
CDC	Centro de Controlo de Doenças
CIBIT	Centro de Imagem Biomédica e Investigação Translacional
DALYs	<i>Disability-adjusted life years</i>
DSM-5	Manual Diagnóstico e Estatístico de Perturbações Mentais
EEG	Eletroencefalografia
EMG	Eletromiografia
GABA	Ácido gama-aminobutírico
ICNAS	Instituto de Ciências Nucleares Aplicadas à Saúde
KDD	<i>Knowledge Discovery from Data</i>
KNN	<i>K-Nearest Neighbor</i>
OMS	Organização Mundial de Saúde
PCA	Análise de Componentes Principais
PEA	Perturbação do Espectro do Autismo
QA	Quociente de Autismo
QI	Quociente de Inteligência
SSRIs	<i>Selective serotonin reuptake inhibitors</i>
SVM	<i>Support Vector Machine</i>

Lista de Símbolos

ms	milissegundos
μV	microvolts
α	Nível de significância
F	Valor da estatística
p	valor-p
$\eta^2 p$	Parcial eta quadrado

1 Introdução

1.1 Contextualização

A perturbação do espectro do autismo (PEA) é uma condição do desenvolvimento neuropsicológico caracterizado por défices na comunicação social e pela presença de comportamentos ou interesses restritos e repetitivos. Estes sintomas causam dificuldades significativas nas áreas social, ocupacional e em outras funções importantes, frequentemente levando a dificuldades na vida diária, como na educação e no emprego (Martin, Pepa and Lord, 2022; Hirota and King, 2023). De acordo com a Organização Mundial da Saúde (OMS), aproximadamente 1 em cada 100 crianças apresenta autismo a nível global (World Health Organization, 2023). Já as estimativas do Centro de Controlo de Doenças (CDC) nos Estados Unidos indicam que cerca de 1 em cada 36 crianças é identificada com PEA (U.S. Centers for Disease Control and Prevention, 2023). Face à elevada prevalência da PEA, torna-se essencial compreender as suas bases neurobiológicas. Este conhecimento é fundamental para ajudar no diagnóstico e para desenvolver intervenções eficazes, que possam melhorar significativamente a qualidade de vida dos indivíduos afetados (Kang *et al.*, 2018).

No contexto do funcionamento cerebral, o equilíbrio entre neurotransmissores excitatórios e inibitórios (equilíbrio E-I) é crucial para a regulação das atividades neuronais. Os neurotransmissores excitatórios, como o glutamato, promovem a ativação dos neurónios, enquanto os inibitórios, como o ácido gama-aminobutírico (GABA), moderam e reduzem essa atividade (Canitano and Palumbi, 2021). Este equilíbrio é fundamental para as diferentes funções cerebrais (Zhao *et al.*, 2022).

No entanto, apesar dos avanços na investigação, persiste uma lacuna significativa no conhecimento sobre o equilíbrio E-I em indivíduos com PEA. A compreensão deste equilíbrio é particularmente complexa, pois envolve uma interação de múltiplos fatores, incluindo alterações genéticas, sinápticas e ambientais que se influenciam mutuamente de forma dinâmica. Esta complexidade decorre do facto de o autismo ser de etiologia multifacetada e heterogénea, não podendo ser atribuída a um único mecanismo. A falta de um entendimento claro sobre o equilíbrio E-I, em indivíduos com PEA, também representa um desafio significativo para a translação das descobertas científicas em estratégias de tratamento, pois impede a identificação de alvos terapêuticos específicos e o desenvolvimento de abordagens terapêuticas personalizadas e eficazes (Nisar *et al.*, 2022).

Assim, a identificação de biomarcadores que possam auxiliar na compreensão dos mecanismos moleculares e, potencialmente, na estratificação de pacientes com PEA em subgrupos com perfis biológicos semelhantes é extremamente relevante. Biomarcadores são indicadores biológicos mensuráveis que permitem compreender a presença, gravidade ou progressão de uma doença e/ou de uma resposta terapêutica. A utilização de biomarcadores pode,

potencialmente, facilitar o diagnóstico precoce, a personalização das intervenções e a avaliação terapêutica subsequente (Kala *et al.*, 2021). Neste âmbito, o componente eletrofisiológico cerebral N170 tem sido amplamente estudado como um potencial biomarcador associado ao processamento de estímulos sociais em indivíduos com PEA (Mason *et al.*, 2022; Shan *et al.*, 2023).

A capacidade de reconhecer e interpretar expressões faciais é vital para uma comunicação eficaz e para a formação de relações sociais típicas, aspetos frequentemente comprometidos em indivíduos com PEA (Zhao *et al.*, 2022). Estudos indicam que indivíduos com autismo apresentam um atraso no sinal fisiológico durante o processamento de rostos, sugerindo que estes precisam de mais tempo para reconhecer estímulos sociais, em comparação com indivíduos com desenvolvimento típico (Kang *et al.*, 2018; Mason *et al.*, 2022). Este sinal é, então, conhecido como N170, componente negativa do potencial evocado por eventos (ERP - do inglês: *event-related potentials*), geralmente observado com eletroencefalografia (EEG) cerca de 170 milissegundos após a apresentação de estímulos faciais. Tais alterações no processamento de informação social têm sido associadas a alterações neurobiológicas, nomeadamente ao desequilíbrio entre mecanismos de E-I no cérebro. No entanto, estas observações são indiretas (Parrella *et al.*, 2024).

Neste estudo, procurou-se compreender e avaliar os mecanismos neurobiológicos subjacentes ao equilíbrio E-I na PEA, bem como explorar o seu potencial como biomarcador para a criação de subgrupos com base na sua biologia. Para tal, foi utilizado o agonista dos recetores inibitórios GABA-B, arbaclofeno, com o objetivo de modular este equilíbrio. A resposta ao arbaclofeno foi analisada em adultos com e sem PEA, durante uma tarefa de decisão perceptual de deteção de faces.

1.2 Objetivos

Neste trabalho, serão analisados dados de EEG de participantes com e sem PEA adquiridos no contexto de um estudo farmacológico, isto é, após a toma única de um placebo e de duas doses diferentes de arbaclofeno. O objetivo deste trabalho é realizar o pré-processamento destes dados e identificar os componentes dos potenciais evocados por uma tarefa de deteção de faces, em particular os componentes N170 e P100. Para facilitar este processo, pretende desenvolver-se *scripts* que automatizem o pré-processamento, a extração dos potenciais evocados e a identificação de características relevantes. A automatização destes processos não só facilita e agiliza o trabalho dos investigadores na identificação dos diversos componentes em estudo, como também se revela útil para estudos semelhantes que possam vir a ser realizados no futuro.

Os objetivos do trabalho são os seguintes:

- 1) Elaboração de *scripts* para automatização do pré-processamento de dados de EEG e extração dos componentes P100 e N170.

- 2) Investigar o potencial do N170 como possível biomarcador para a criação de subgrupos de doentes tendo em conta a resposta ao arbaclofeno.

1.3 Estrutura da Tese

Esta dissertação abrange seis capítulos distintos, começando por uma contextualização, os objetivos do trabalho e a estrutura do documento.

No segundo capítulo, é abordada a Revisão da Literatura, onde é fornecida uma visão geral do conhecimento existente, abrangendo a PEA, a sua neurobiologia e o pré-processamento de EEG, elementos essenciais para a compreensão do estudo.

No terceiro capítulo, estão descritos todos os métodos e ferramentas usadas para a elaboração deste estudo, assim como as análises e algoritmos realizados.

No quarto capítulo, são apresentados todos os resultados das análises estatísticas e algoritmos implementados no estudo.

Uma discussão detalhada dos resultados obtidos é feita no quinto capítulo, destacando as principais limitações deste estudo e recomendações para futuras pesquisas.

Por fim, no sexto capítulo, são apresentadas as conclusões do estudo, integrando uma discussão crítica do trabalho realizado.

2 Revisão da Literatura

2.1 Perturbação do Espectro do Autismo (PEA)

De acordo com o Manual Diagnóstico e Estatístico de Perturbações Mentais (DSM-5), a PEA é uma condição do neurodesenvolvimento caracterizada por défices persistentes na comunicação social e na interação, associados a padrões restritos e repetitivos de comportamento, interesses ou atividades. Esses sintomas manifestam-se desde o período inicial de desenvolvimento e causam dificuldades significativas nas áreas social, ocupacional ou em outras funções importantes que afetam o quotidiano (Martin, Pepa and Lord, 2022).

A prevalência global da PEA tem demonstrado variações geográficas substanciais, com estimativas que oscilam entre 1 em 100 crianças segundo a OMS, até 1 em 36 crianças de acordo com dados recentes do CDC dos Estados Unidos (Maenner *et al.*, 2023; World Health Organization, 2023). Regiões com elevado índice socioeconómico apresentam taxas de prevalência e DALYs (*Disability-adjusted life years*) mais elevadas, como o Reino Unido com 112,29 DALYs por 100.000 habitantes, sugerindo uma correlação entre desenvolvimento socioeconómico e capacidade diagnóstica. Em 2019, a PEA foi responsável por 4,3 milhões de DALYs globalmente, com um aumento de 38,66% desde 1990, evidenciando o crescente impacto desta condição na saúde pública mundial (Li *et al.*, 2022; Solmi *et al.*, 2022). As tendências temporais revelam um aumento consistente nas taxas de diagnóstico, com alguns países a reportar aumentos superiores a 250% numa década, muito devido ao aumento da consciencialização profissional (Hodges, Fealko and Soares, 2020).

A PEA é uma condição altamente heterogénea, caracterizada por uma ampla variabilidade na apresentação dos sintomas, na gravidade das manifestações e na presença de comorbilidades. Enquanto alguns indivíduos apresentam défices graves na comunicação e interação social, outros possuem competências cognitivas e linguísticas preservadas, mas enfrentam dificuldades na adaptação a normas sociais (Hirota and King, 2023). Além disso, fatores ambientais, como o acesso a intervenções especializadas e suporte educacional, influenciam significativamente a trajetória de desenvolvimento e a funcionalidade ao longo da vida destes indivíduos (Sauer *et al.*, 2021).

As dificuldades sociais inerentes à PEA impactam significativamente a integração dos indivíduos em contextos educacionais, laborais e sociais, influenciando o seu desenvolvimento e qualidade de vida. As limitações na compreensão e na aplicação de normas sociais podem comprometer a formação de relações interpessoais, resultando em isolamento e redução das oportunidades de socialização. No contexto escolar, os défices na comunicação e na interação dificultam o desempenho académico, enquanto, no mercado de trabalho, as barreiras sociais e a dificuldade em interpretar sinais não verbais reduzem a empregabilidade e a estabilidade financeira. Estes desafios cumulativos contribuem para níveis elevados de stress e angústia entre indivíduos com esta perturbação, afetando o seu bem-estar psicológico e emocional (Salari *et al.*, 2022).

O diagnóstico da PEA é baseado em critérios comportamentais, uma vez que não existe um exame médico específico para a sua identificação. A investigação científica continua a explorar as bases genéticas e ambientais da PEA, bem como a desenvolver estratégias de intervenção que promovam a adaptação e a qualidade de vida dos indivíduos com esta condição (Wang *et al.*, 2023).

A etiologia da PEA caracteriza-se por uma complexa interação entre fatores genéticos e ambientais. A forte componente genética é evidenciada por estudos em gémeos que demonstram taxas de concordância de 70-90% em monozigóticos versus 0-10% em dizigóticos, com 20-25% dos casos apresentando causas genéticas identificáveis, incluindo mutações de novo e variações genéticas raras (Sauer *et al.*, 2021). Os fatores de risco não genéticos abrangem exposições ambientais durante a gestação (toxinas, pesticidas, medicamentos), idade parental avançada, estado nutricional materno, infeções e stress pré-natal, bem como alterações na microbiota intestinal, particularmente a disbiose intestinal associada ao uso precoce de antibióticos (Taniya *et al.*, 2022; Wang *et al.*, 2023).

2.1.1 Sintomatologia

A PEA é caracterizada por dificuldades significativas na comunicação e na interação social, afetando a forma como os indivíduos percebem e respondem ao mundo à sua volta. Um dos principais desafios enfrentados por pessoas com este distúrbio reside na reciprocidade social e emocional, refletindo-se na dificuldade em iniciar ou manter interações, na aparente falta de interesse por trocas sociais e na resposta atípica a estímulos interpessoais (Banker *et al.*, 2021).

Muitos indivíduos apresentam limitações na utilização e interpretação de sinais não-verbais, como expressões faciais, contacto visual e gestos, comprometendo a sua capacidade de compreender o significado implícito nas interações sociais. Estas dificuldades tornam mais desafiante o desenvolvimento e a manutenção de relações interpessoais, uma vez que influenciam a forma como interpretam as intenções, emoções e estados mentais dos outros (Aishworiya *et al.*, 2022).

As dificuldades cognitivas associadas à PEA não se restringem apenas à interação social e comunicação, mas também envolvem falhas em processos como raciocínio espacial e memória episódica. Evidências sugerem que estas dificuldades podem estar relacionadas com alterações no hipocampo, uma estrutura cerebral fundamental para a formação e recuperação de memórias, bem como para a orientação espacial. Anomalias na morfologia e funcionalidade desta região podem comprometer a capacidade dos indivíduos para processar informações espaciais de forma eficiente, influenciando a navegação no ambiente e a organização de experiências passadas. Esta dificuldade em integrar informações sobre as propriedades físicas do ambiente e a sua relação com as próprias capacidades motoras pode prejudicar a autonomia e a adaptação a novas situações (Banker *et al.*, 2021).

Para além das dificuldades referidas, PEA caracteriza-se, também, pela presença de padrões de comportamento restritos, repetitivos e estereotipados. Estes comportamentos podem

manifestar-se através de movimentos repetitivos, como balançar o corpo ou girar sobre si mesmo, bem como pela fixação intensa em interesses específicos, muitas vezes de forma desproporcional à idade ou contexto. A insistência em rotinas rígidas é outra característica marcante, tornando qualquer alteração no ambiente ou nas atividades diárias uma fonte de desconforto significativo (Aishworiya *et al.*, 2022).

As particularidades sensoriais também desempenham um papel relevante no funcionamento dos indivíduos com autismo. A hipersensibilidade ou hipossensibilidade a estímulos externos, como sons, texturas ou padrões luminosos, pode influenciar o comportamento e a adaptação ao meio, levando a reações exacerbadas ou, pelo contrário, a uma procura intensa por determinados estímulos (Buch *et al.*, 2023).

A ansiedade é, também, uma das comorbilidades mais frequentes na PEA, podendo manifestar-se sob diferentes formas, como ansiedade generalizada, fobias específicas ou sintomas obsessivo-compulsivos. A presença de ansiedade pode intensificar as dificuldades já existentes na comunicação e interação social, contribuindo para o aumento do isolamento e da evitação de situações sociais. A literatura sugere uma ampla variabilidade na prevalência de perturbações de ansiedade dentro do espectro, com estudos a reportar taxas que variam significativamente consoante a metodologia utilizada e as características da amostra. Em crianças e adolescentes que apresentam esta perturbação, a ansiedade tende a ser particularmente prevalente, podendo influenciar negativamente o desenvolvimento de competências socioemocionais e a capacidade de adaptação a novas experiências (Bougeard *et al.*, 2021).

Os distúrbios depressivos também são frequentemente associados à PEA, sendo particularmente prevalente em crianças e adolescentes. As dificuldades na comunicação e interação social, combinadas com a perceção das próprias limitações e desafios diários, podem contribuir para o desenvolvimento de sintomas depressivos. A frustração decorrente da dificuldade em estabelecer relações interpessoais, aliada a experiências recorrentes de rejeição ou incompreensão, podem levar ao aumento do isolamento e à anedonia (Aishworiya *et al.*, 2022).

As dificuldades na comunicação verbal são uma característica prevalente nestes indivíduos, manifestando-se de formas variadas, desde atrasos significativos na aquisição da linguagem até a ausência total de fala. Estas dificuldades resultam, em grande parte, de alterações neurológicas que afetam a cognição social e o processamento da linguagem. Estudos indicam que indivíduos com PEA apresentam padrões atípicos de conectividade cerebral, particularmente em regiões envolvidas na compreensão e produção da fala. Alterações estruturais e funcionais no tálamo, responsável pela transmissão de informações sensoriais e motoras, podem afetar a integração dos estímulos necessários para uma comunicação eficaz. Anomalias em áreas visuais e na rede de saliência, que desempenham um papel na atenção e na seleção de estímulos relevantes, podem interferir na perceção e interpretação de sinais não verbais, como expressões faciais e gestos (Buch *et al.*, 2023).

As fobias são bastante comuns entre indivíduos com autismo, o que pode prejudicar significativamente a adaptação destes ao ambiente. Estes medos, muitas vezes intensos e irracionais, podem desencadear respostas de evitamento em relação a determinadas situações, locais ou estímulos sensoriais, limitando a participação em atividades diárias e sociais. A hipersensibilidade a sons, luzes ou movimentos inesperados podem amplificar estas reações. As dificuldades relacionadas à alimentação também são frequentes, refletindo-se em padrões alimentares seletivos. A rejeição de determinados alimentos pode estar associada a uma percepção sensorial alterada, levando a uma forte aversão a certas texturas, sabores, temperaturas ou cores. Esta seletividade pode comprometer a ingestão equilibrada de nutrientes, aumentando o risco de défices nutricionais e impactando o desenvolvimento físico e cognitivo (Aishworiya *et al.*, 2022).

2.1.2 Diagnóstico

O diagnóstico da PEA assume uma importância capital na vida das pessoas afetadas, sendo fundamental não apenas na infância, mas também ao longo de todo o ciclo de vida. A precisão e a precocidade do diagnóstico são cruciais para o desenvolvimento de intervenções personalizadas e eficazes, melhorando significativamente a qualidade de vida dos indivíduos com esta perturbação. No entanto, apesar da existência de diretrizes clínicas que visam garantir avaliações abrangentes e de alta qualidade, a realidade enfrentada por muitas famílias é a de longas listas de espera, resultantes da alta demanda por serviços de diagnóstico e avaliação (Zwaigenbaum and Penner, 2018).

A despeito dos avanços significativos nas evidências científicas sobre os primeiros sinais de autismo, a idade média de diagnóstico permanece elevada, situando-se entre os 4 a 5 anos de idade, sem indícios claros de redução. Esta demora no diagnóstico não é a única preocupação, há também um subdiagnóstico preocupante entre jovens e adultos mais velhos, sugerindo que muitos indivíduos podem não receber o diagnóstico e o apoio de que necessitam (Qin *et al.*, 2024).

O processo de diagnóstico da PEA tem, tradicionalmente, dependido de avaliações minuciosas do comportamento e do histórico de desenvolvimento de cada indivíduo. Essas avaliações são realizadas por profissionais de saúde altamente especializados, como pediatras e psicólogos, que empregam uma abordagem compreensiva para entender as nuances do desenvolvimento da criança. A observação direta do comportamento da criança, aliada a entrevistas aprofundadas com os pais ou cuidadores, constitui um componente essencial desta abordagem. Durante as entrevistas, os profissionais de saúde procuram recolher informações detalhadas sobre as interações sociais, as habilidades de comunicação e os padrões comportamentais da criança (Zwaigenbaum and Penner, 2018; Hirota and King, 2023).

A literatura científica, de facto, sugere que a avaliação ideal para o diagnóstico da PEA deve ser realizada por uma equipa multidisciplinar, composta por pediatras, psiquiatras, terapeutas da fala, entre outros. Esta abordagem permite uma avaliação mais completa, considerando diferentes aspetos do desenvolvimento e do comportamento, essencial para um diagnóstico

preciso. No entanto, na prática clínica, observam-se variações significativas, com muitos clínicos a realizar avaliações de forma isolada, o que pode comprometer a precisão dos diagnósticos (Zwaigenbaum and Penner, 2018).

2.1.2.1 Métodos de *Machine Learning*

Métodos, como os de *machine learning*, têm emergido como mecanismos promissores no diagnóstico de PEA, oferecendo a capacidade de analisar grandes volumes de dados e identificar padrões complexos que podem não ser facilmente detetáveis por métodos tradicionais (Sharma and Tanwar, 2020). Um bom exemplo é o *Support Vector Machine* (SVM), que é um modelo de aprendizagem automática amplamente utilizado em diversas aplicações, incluindo a classificação binária de dados. No contexto do diagnóstico, o SVM permite identificar indivíduos com esta perturbação com base em características específicas extraídas de diferentes conjuntos de dados. O SVM opera através da identificação de uma superfície de decisão num espaço multidimensional, determinando o hiperplano que maximiza a margem entre as classes, ou seja, a distância entre o hiperplano e os pontos de dados das diferentes classes (Erkan and Thanh, 2020; Rasul *et al.*, 2024).

A aplicação do SVM envolve várias etapas metodológicas. Inicialmente, os dados são submetidos a um pré-processamento e seleção de características, onde características relevantes são escolhidas, utilizando métodos estatísticos. Posteriormente, os dados são divididos em conjuntos de treino e teste, e o modelo é treinado no conjunto de treino. Estudos indicam que o SVM pode alcançar elevados níveis de precisão no diagnóstico da PEA, com alguns relatos a apontar para 100% de precisão na deteção em dados de crianças (Rasul *et al.*, 2024).

O método *Random Forest* é outra técnica de *machine learning*, que se destaca pela sua capacidade de classificação através da utilização de múltiplas árvores de decisão. Cada árvore na floresta contribui com uma previsão sobre a classe de um dado da amostra, e a classe final é determinada pela maioria das previsões das árvores individuais. Esta abordagem é considerada um algoritmo de estado da arte em classificação devido à sua eficácia e precisão. A principal característica deste modelo reside na criação de várias árvores de decisão a partir de amostras aleatórias dos dados de treino, o que não só melhora a precisão do modelo, como também reduz o risco de sobreajuste. A combinação das decisões de múltiplas árvores torna o modelo mais robusto e menos suscetível a erros provocados por dados ruidosos ou por variabilidades específicas em amostras individuais (Erkan and Thanh, 2020).

O *Random Forest* é aplicado através de um processo estruturado que começa com a recolha de um conjunto de dados abrangente, incluindo características comportamentais, interações sociais e expressões faciais, entre outros fatores relevantes. Estes dados são então pré-processados para garantir que estejam num formato adequado para análise, o que pode envolver a normalização, a transformação de variáveis categóricas em numéricas e a remoção de dados ausentes (Erkan and Thanh, 2020). Com o conjunto de dados devidamente preparado, o modelo é treinado para identificar padrões. Após o treino, o modelo é capaz de prever, por exemplo, se uma nova observação, como um conjunto de comportamentos observados numa

criança, indica a presença de traços autistas. Cada árvore na floresta contribui com um voto para a classe, e a decisão final é determinada pela maioria dos votos. Estudos têm demonstrado que o *Random Forest* alcança resultados precisos na detecção de traços autistas, com taxas de precisão que variam entre 92,26% e 97,10% em diferentes faixas etárias, incluindo crianças, adolescentes e adultos (Sharma and Tanwar, 2020).

Outra técnica frequentemente utilizada é de *K-Nearest Neighbor* (KNN), que é um algoritmo de classificação que opera com base em instâncias, classificando novos pontos de dados com base na proximidade a pontos de dados já conhecidos num conjunto de treino. O funcionamento do KNN envolve a identificação dos "K" vizinhos mais próximos de um ponto de teste e a atribuição da classe mais comum entre esses vizinhos. Este método não requer a construção de um modelo explícito da relação entre as variáveis, mas sim a manutenção de todos os dados de treino, permitindo que o algoritmo calcule a distância entre o novo ponto e todos os outros pontos de dados no conjunto de treino (Erkan and Thanh, 2020).

O KNN é aplicado utilizando dados que podem incluir, por exemplo, características comportamentais e outros indicadores associados ao autismo. Durante o processo de classificação, o algoritmo determina a classe do novo ponto, que indica a presença ou ausência de traços autistas, com base na classe mais frequente entre os K vizinhos mais próximos. Este método permite ao KNN identificar padrões e relacionar características a classes de forma eficaz. Um aspecto crítico na aplicação do KNN é o ajuste do valor de K, que deve ser cuidadosamente selecionado para otimizar a precisão do modelo. Um valor de K demasiado pequeno pode tornar o modelo sensível ao ruído, enquanto um valor demasiado grande pode suavizar as diferenças entre classes, comprometendo a capacidade do modelo de distinguir entre a presença e a ausência de traços autistas. Na maioria dos estudos, o algoritmo KNN alcançou uma precisão inferior à do RF, ficando pelos 88,5% (Sharma and Tanwar, 2020).

A Regressão Logística, um método que utiliza uma abordagem diferente, é um modelo estatístico utilizado para prever a probabilidade de um resultado binário, estabelecendo uma relação entre uma variável dependente binária e uma ou mais variáveis independentes. Este modelo linear é particularmente útil quando existem duas possíveis saídas, como "autista" ou "não autista". A Regressão Logística gera uma função logística que estima a probabilidade do evento de interesse, por exemplo, a presença de PEA, com base em variáveis explicativas (Panda, 2022).

A Regressão Logística é aplicada para modelar estatisticamente a relação entre características individuais e a probabilidade de diagnóstico de autismo. A escolha deste modelo deve-se à sua capacidade de fornecer uma interpretação clara dos resultados, permitindo que os investigadores compreendam como as variáveis independentes influenciam a probabilidade de ocorrência desta perturbação. O modelo é treinado com dados que contêm características de indivíduos, calculando uma função de probabilidade que classifica os dados em duas categorias, com e sem autismo. As variáveis independentes são as características extraídas dos indivíduos. Em determinados estudos, a Regressão Logística demonstrou uma precisão máxima de 97,14% ao classificar dados de adultos, destacando-se como uma das metodologias que contribuíram para elevados índices de precisão, *recall* e métricas de desempenho geral (Rasul *et al.*, 2024).

Para avaliar o desempenho de modelos de *machine learning*, geralmente são utilizadas quatro métricas complementares. A exatidão (*accuracy*) quantifica a proporção global de classificações corretas, fornecendo uma medida geral de desempenho que, embora intuitiva, pode ser enganadora em conjuntos de dados desequilibrados. A precisão mede a proporção de verdadeiros positivos entre todas as predições positivas, indicando a fiabilidade do modelo quando identifica uma instância como positiva. A sensibilidade (*recall*) avalia a capacidade do modelo em detetar todos os casos positivos existentes, sendo crucial em contextos onde falsos negativos têm consequências significativas. O *F1-score* combina precisão e sensibilidade numa única métrica através da sua média harmónica, proporcionando uma avaliação equilibrada particularmente útil quando ambas as métricas são igualmente importantes. Esta medida compósita permite avaliar o desempenho global do modelo considerando simultaneamente a sua capacidade de deteção e a exatidão das predições positivas, sendo especialmente informativa na presença de classes desequilibradas ou quando existe necessidade de otimizar o compromisso entre falsos positivos e falsos negativos (Schlosser *et al.*, 2024).

2.1.3 Intervenções

A abordagem terapêutica da PEA requer uma perspetiva sistémica e holística que integre uma vasta gama de intervenções comportamentais, educacionais e médicas, a fim de enfrentar as muitas e complexas necessidades dos indivíduos afetados. A colaboração multifacetada entre profissionais, saúde, educação e terapeutas é obrigatória para promover uma resposta eficaz e adaptativa às necessidades únicas de cada indivíduo (Tian, Gao and Yang, 2022).

As intervenções comportamentais englobam uma variedade de abordagens e técnicas baseadas nos princípios da psicologia comportamental, destinadas a modificar comportamentos problemáticos e a facilitar o desenvolvimento de habilidades em indivíduos com PEA para a realização de atividades diárias, melhorar a funcionalidade e promover a autonomia. Estas intervenções são projetadas para aumentar as habilidades adaptativas e reduzir comportamentos indesejados, promovendo maior flexibilidade e ajudando as pessoas a adquirirem habilidades de vida essenciais. São fundamentais no desenvolvimento das competências de comunicação e sociais em crianças com autismo, impactando diretamente na sua qualidade de vida e na capacidade de interação nas atividades diárias (Tian, Gao and Yang, 2022; Qin *et al.*, 2024).

Embora as intervenções comportamentais se demonstrem positivas, a utilização de intervenções farmacológicas pode ser uma opção terapêutica viável para a gestão de sintomas específicos, tais como agressividade, ansiedade e défices de atenção, embora seja fundamental reconhecer que não existe uma cura estabelecida para esta condição. É imperativo que as intervenções farmacológicas sejam objeto de uma monitorização rigorosa, com o objetivo de garantir tanto a eficácia quanto a segurança dos medicamentos utilizados, minimizando os riscos de efeitos adversos e otimizando os resultados terapêuticos (Tian, Gao and Yang, 2022).

Embora determinados medicamentos estejam a ser considerados para o lidar com os comportamentos repetitivos e restritivos, a literatura ainda não reconhece uma intervenção farmacológica como totalmente efetiva. Assim, as intervenções comportamentais continuam a ser a principal abordagem até que opções farmacológicas mais eficazes sejam desenvolvidas (Tian, Gao and Yang, 2022).

As intervenções farmacológicas para a PEA envolve abordagens que visam aliviar sintomas associados e tratar comorbilidades. Embora as intervenções comportamentais sejam fundamentais no apoio a indivíduos com autismo, os medicamentos são frequentemente utilizados para gerir comorbilidades como ansiedade, hiperatividade e comportamentos desafiadores. Muitos indivíduos com autismo são medicados com múltiplos fármacos, especialmente aqueles que apresentam comorbilidades como deficiência intelectual ou problemas de saúde mental, sendo comum o uso de terapias combinadas, dado que um único medicamento pode não ser eficaz para todas as condições. Há um foco crescente em tratamentos direcionados que abordem as alterações neurobiológicas subjacentes ao autismo, particularmente em subgrupos com bases genéticas conhecidas. Existe também a necessidade de considerar a maior sensibilidade dos indivíduos com autismo aos efeitos colaterais dos medicamentos, recomendando-se que estes sejam iniciados em doses mais baixas e ajustados com cautela (Aishworiya *et al.*, 2022).

A utilização de medicamentos farmacológicos entre indivíduos com PEA tem registado um aumento significativo ao longo dos anos. Um estudo revelou que a taxa de uso de intervenções farmacológicas para pessoas com autismo subiu de 57% em 1998 para 64% em 2014. A prescrição de medicamentos psicotrópicos é particularmente comum, começando em crianças a partir dos 2 anos de idade. A taxa de prescrição destes medicamentos aumenta com a idade, sendo de 11% entre crianças de 3 a 5 anos, 46% entre aquelas de 6 a 11 anos e 66% entre adolescentes de 12 a 17 anos. Estes dados traduzem uma tendência crescente na utilização de intervenções farmacológicas como parte do tratamento global para o autismo (Deb, Roy and Limbu, 2023).

Os medicamentos mais comuns utilizados na PEA incluem os antipsicóticos atípicos, que são frequentemente prescritos para gerir comportamentos desafiadores, irritabilidade e agressão, mesmo que muitas vezes sejam utilizados sem uma indicação formal (Aishworiya *et al.*, 2022). Os antipsicóticos atípicos exercem a sua ação terapêutica através da interação com diversos recetores de neurotransmissores, incluindo os de dopamina, serotonina, muscarínicos, adrenérgicos e histaminérgicos (Carli *et al.*, 2021).

Entre os mais utilizados no autismo, a risperidona é frequentemente sugerida, pela sua eficácia na redução de sintomas como agitação, irritabilidade e agressão. O aripiprazol também apresenta alguma evidência preliminar de eficácia para esses comportamentos, embora o suporte para o seu uso seja menos robusto em comparação com a risperidona (Deb, Roy and Limbu, 2023).

Os inibidores seletivos da recaptção de serotonina (SSRIs - *Selective serotonin reuptake inhibitors*) são também usados na PEA, ao regularem os níveis de serotonina, um

neurotransmissor essencial no sistema nervoso central e em várias funções do organismo. A desregulação da serotonina está frequentemente associada a sintomas comuns em pessoas com autismo, como comportamentos repetitivos e ansiedade. Estudos sugerem que indivíduos com autismo podem ter níveis elevados de serotonina no sangue, embora investigações, como as realizadas através de tomografia por emissão de positrões, indiquem que crianças mais jovens com autismo apresentam níveis mais baixos de serotonina no líquido cefalorraquidiano. A realização de análises de linhas celulares de linfoblastos em pacientes com autismo mostraram que também há uma deficiência de enzimas responsáveis pela conversão de triptofano em serotonina (Aishworiya *et al.*, 2022). Os SSRIs são, por isso, frequentemente utilizados para gerir sintomas associados à PEA, como comportamentos repetitivos e irritabilidade. No entanto, a eficácia destes medicamentos em melhorar os sintomas de comunicação social no autismo tem sido objeto de debate, uma vez que a maioria dos estudos não encontrou benefícios significativos nesta área (Nanjappa *et al.*, 2022).

Os agonistas alfa-2 também são frequentemente utilizados no tratamento de sintomas associados à PEA, particularmente para gerir a hiperatividade e problemas de atenção (Aishworiya *et al.*, 2022). Os agonistas alfa-2 atuam principalmente através da estimulação dos recetores alfa-2 adrenérgicos localizados no sistema nervoso central, particularmente no tronco encefálico. Esta estimulação ativa neurónios inibitórios, resultando numa diminuição do fluxo simpático central. Como consequência, observa-se uma redução da resistência vascular periférica, da frequência cardíaca e da pressão arterial. A ativação dos recetores alfa-2 também inibe a libertação de neurotransmissores excitatórios, como a norepinefrina, o que contribui para os efeitos sedativos e analgésicos associados a estes medicamentos (Nguyen *et al.*, 2017).

Os agonistas alfa-2 são frequentemente prescritos para tratar comportamentos hiperativos e impulsivos, que são comuns em crianças com autismo (Aishworiya *et al.*, 2022). Estes medicamentos podem ser utilizados isoladamente ou em combinação com outros, como medicamentos estimulantes, dependendo das necessidades específicas de cada paciente e da sua resposta à intervenção (Aishworiya *et al.*, 2022).

Os agonistas alfa-2 estão associados a uma série de efeitos secundários que requerem atenção cuidadosa. Entre os mais comuns, destacam-se a sedação e sonolência, especialmente com o uso de clonidina, bem como um aumento potencial da irritabilidade em alguns pacientes. Pode também ocorrer uma diminuição na atividade física e na energia geral, além de distúrbios do sono, que se manifestam como insónia ou sonolência excessiva. A hipotensão é outro efeito secundário relevante, podendo levar a sintomas como tontura ou desmaios, enquanto a bradicardia, caracterizada por uma redução da frequência cardíaca, exige monitorização contínua. Embora raras, reações alérgicas podem surgir e devem ser tratadas de imediato. Dada a possibilidade destes efeitos, é fundamental realizar um acompanhamento regular dos pacientes para avaliar tanto a eficácia do tratamento como a ocorrência de efeitos colaterais (Kaye *et al.*, 2024).

2.2 Neurobiologia da Perturbação do Espectro do Autismo (PEA)

A PEA é uma condição complexa, caracterizada por uma série de disfunções neurobiológicas que afetam o desenvolvimento cerebral e o comportamento. Nos indivíduos com esta perturbação, os processos neurobiológicos são influenciados por uma combinação de fatores genéticos e ambientais, que impactam diversas áreas do desenvolvimento cerebral. Alterações no desenvolvimento cerebral são particularmente marcantes, incluindo um crescimento cerebral atípico durante os primeiros anos de vida, que pode ser seguido por uma fase de estagnação ou até mesmo degeneração em estágios posteriores (Sauer *et al.*, 2021).

A neurogênese, o processo de formação de novos neurónios, desempenha um papel fundamental no desenvolvimento cerebral, e é precisamente durante este processo que o risco genético associado ao autismo parece ter um impacto significativo. Mais especificamente, a migração neuronal no desenvolvimento cortical inicial é uma fase crítica, durante a qual os neurónios são formados e começam a migrar para as áreas específicas do cérebro, o que, por sua vez, influencia a estrutura e função cerebral subsequentes (Eyring and Geschwind, 2021). Nesta fase, o risco genético para a PEA está principalmente associado a uma disfunção em processos de regulação transcricional, o que significa que as proteínas que normalmente controlariam a expressão genética de forma adequada podem estar alteradas, afetando a neurogênese e levando a desenvolvimentos anómalos. A expressão de genes de risco para a PEA está particularmente concentrada em neurónios glutamatérgicos em desenvolvimento durante esta fase, que são importantes para a sinalização excitatória no cérebro e estão envolvidos em processos relacionados com a aprendizagem e a memória (Eyring and Geschwind, 2021; Lamanna and Meldolesi, 2024). Estas células são importantes na formação de circuitos neuronais que suportam a comunicação neuronal e o processamento sensorial, sendo essenciais para o desenvolvimento de funções cognitivas e comportamentais típicas. O desenvolvimento de circuitos neuronais é um processo crítico que ocorre durante períodos específicos do desenvolvimento fetal e infantil. A regulação inadequada de fatores de transcrição e epigenéticos pode, portanto, ter um impacto significativo na criação de redes neuronais apropriadas (Eyring and Geschwind, 2021).

As mutações genéticas associadas ao autismo têm um impacto significativo nos mecanismos sinápticos, pois alteram a expressão de genes que regulam a síntese e a função das proteínas sinápticas. Isto leva a desregulações sinápticas, que podem afetar a comunicação neuronal e contribuir para os sintomas desta perturbação. As espinhas dendríticas, que são projeções das dendrites onde ocorrem a maioria das sinapses, apresentam morfologias alteradas em indivíduos com autismo. Estas alterações morfológicas podem estar relacionadas com flutuações conformacionais que impactam o funcionamento neuronal, afetando a plasticidade sináptica e a capacidade de aprendizagem e memória (Lamanna and Meldolesi, 2024).

As sinapses também apresentam alterações na sua estrutura e na forma como funcionam em indivíduos com PEA, o que pode levar a défices na comunicação neuronal e contribuir para os comportamentos frequentemente associados a este distúrbio (Sauer *et al.*, 2021).

Algumas das áreas cerebrais implicadas na PEA incluem o mesencéfalo, que é crucial para a integração sensorial e a coordenação motora, o hipocampo, fundamental para a formação de memórias e a aprendizagem, e o giro temporal, envolvido no processamento auditivo e na compreensão da linguagem (Lamanna and Meldolesi, 2024). Ademais, o córtex temporal medial, associado à perceção social e à compreensão das interações sociais, o cerebelo, que desempenha funções na coordenação motora e no processamento emocional, e a amígdala, envolvida no processamento emocional e nas respostas ao medo, também são áreas que apresentam alterações em indivíduos com esta condição. Estas alterações podem contribuir para défices nas competências de comunicação, regulação emocional e interação social, características típicas da perturbação (Sauer *et al.*, 2021).

2.2.1 Excitação e inibição na função cerebral

A excitação e inibição neuronal são processos centrais para o desenvolvimento e a função cerebral, desempenhando papéis críticos na modulação da atividade neuronal e integração de informações. A excitação, mediada por neurotransmissores como o glutamato, promove a ativação de neurónios, permitindo a transmissão de sinais no cérebro e o processamento de informações importantes. Por outro lado, a inibição, mediada por neurotransmissores como o GABA, regula e reduz a atividade neuronal, equilibrando a excitação e ajudando a filtrar “ruído” irrelevante (Canitano and Palumbi, 2021). O equilíbrio entre a excitação (E) e a inibição (I), conhecido como E/I, é fulcral para o funcionamento adequado dos circuitos neuronais, permitindo que a atividade cerebral seja organizada e diminuindo a possibilidade de hiperexcitabilidade. Quando esse equilíbrio é perturbado, seja por uma excitação excessiva ou inibição insuficiente, pode resultar em dificuldades cognitivas e comportamentais, como as observadas em pessoas com PEA, incluindo dificuldades de interação social e comportamentos repetitivos (Bjørklund and El-Ansary, 2024).

O equilíbrio E/I é de relevante importância para o funcionamento dos mecanismos subjacentes aos circuitos neuronais, principalmente os corticais. Os circuitos corticais são estruturas complexas compostas por neurónios glutamatérgicos excitatórios, denominados piramidais, e interneurónios inibitórios. Os interneurónios GABAérgicos inibem a atividade dos neurónios excitatórios, regulando a excitabilidade geral do circuito e contribuindo para a organização neuronal. A eficácia destes interneurónios na modulação da excitabilidade é vital para manter o funcionamento normal dos circuitos e prevenir a hiperexcitabilidade, que pode ser prejudicial. Assim, a integração entre excitação e inibição molda a função cerebral sendo fundamental para processos de aprendizagem e interação social (Canitano and Palumbi, 2021).

Para manter o equilíbrio E/I, são importantes dois mecanismos de adaptação sinápticos, a plasticidade sináptica e homeostática. A plasticidade sináptica refere-se à capacidade das sinapses de se adaptarem em resposta a padrões de atividade neuronal, manifestando-se através do fortalecimento ou enfraquecimento das conexões. Este mecanismo é muito relevante para a aprendizagem e a memória, permitindo que os neurónios ajustem a força das suas conexões em resposta a experiências e ao ambiente. A plasticidade homeostática assegura

a estabilidade das redes neuronais, ajustando a excitabilidade intrínseca dos neurónios e as forças sinápticas para manter o equilíbrio entre excitação e inibição (Chen *et al.*, 2022).

A excitação neuronal está intimamente ligada ao papel do glutamato, o principal neurotransmissor excitatório no sistema nervoso central. A função do glutamato na transmissão sináptica ocorre através da ligação aos seus recetores específicos, desencadeando a abertura de canais iónicos que permitem a entrada de sódio e cálcio nas células, promovendo a despolarização neuronal e a propagação dos sinais elétricos. Este neurotransmissor está envolvido na plasticidade sináptica, particularmente na potencialização de longa duração (LTP). Este mecanismo está associado ao fortalecimento das conexões neuronais e a processos cognitivos como a aprendizagem e a memória, sendo mediado, em grande parte, pelos recetores NMDA do glutamato (Nisar *et al.*, 2022). O glutamato também participa na sinaptogénese ao promover a maturação das espinhas dendríticas, influenciando diretamente a formação de novas sinapses e contribuindo para a modulação da atividade neuronal ao longo da vida. Para além do seu impacto nas funções cognitivas, está também associado à regulação do comportamento e da atividade motora, facilitando respostas coordenadas aos estímulos ambientais. A sua influência estende-se ainda a processos metabólicos no cérebro, afetando a disponibilidade de energia e a produção de neuromoduladores (Montanari *et al.*, 2022).

As alterações na neurotransmissão glutamatérgica podem ter consequências graves para a saúde cerebral, incluindo a excitotoxicidade, que é o dano neuronal resultante da ativação excessiva dos recetores de glutamato, levando a uma sobrecarga de cálcio nas células e resultando em danos celulares e morte neuronal. A disfunção na neurotransmissão glutamatérgica pode também contribuir para défices cognitivos e comportamentais, como problemas de aprendizagem, memória e comportamentos disruptivos, observados em condições como a PEA (Nisar *et al.*, 2022). Ademais, o aumento do glutamato pode desencadear stress oxidativo e comprometer a função mitocondrial, resultando numa cascata de danos celulares. Evidências também sugerem que a disfunção glutamatérgica pode estar envolvida em processos de neuroinflamação, uma vez que a ativação excessiva dos recetores NMDA pode induzir a libertação de citocinas pró-inflamatórias. Estas alterações subjacentes são comuns não apenas na PEA, mas também em outras condições como esquizofrenia, depressão e epilepsia, podendo afetar a função neuronal em várias áreas do cérebro, levando a alterações comportamentais e emocionais (Montanari *et al.*, 2022).

Por sua vez, a inibição no sistema nervoso central é essencial para equilibrar a excitação neuronal e manter o equilíbrio entre as atividades excitatórias e inibitórias. A ativação de neurónios excitatórios, que libertam neurotransmissores como o glutamato, deve ser contrabalançada por neurónios inibitórios para evitar a sobrecarga da atividade elétrica e permitir um controlo mais preciso da atividade neuronal (Bjørklund and El-Ansary, 2024). O GABA é o principal neurotransmissor inibitório do sistema nervoso central, sendo fundamental na modulação da atividade neuronal. Ao ligar-se aos seus recetores, provoca a hiperpolarização das células neuronais, diminuindo a probabilidade de ativação e prevenindo a superexcitação. Este neurotransmissor influencia a regulação dos ritmos cerebrais, incluindo a sincronização das oscilações neuronais durante os ciclos de sono e a manutenção de um padrão de atividade

cerebral estável. O GABA também atua durante os períodos críticos de desenvolvimento do sistema nervoso. Nesses momentos, a sinalização GABAérgica influencia a maturação e a organização das redes neuronais, promovendo a formação de circuitos importante para o funcionamento cognitivo adequado. Durante o desenvolvimento embrionário, o GABA inicialmente atua de forma excitatória devido aos elevados níveis de cloro intracelular, conferindo-lhe um papel duplo, sendo fundamental tanto para a ativação inicial como para o estabelecimento de circuitos inibitórios necessários ao desenvolvimento do sistema nervoso central (Zhao *et al.*, 2022).

A desregulação do sistema GABAérgico tem sido amplamente associada a distúrbios do neurodesenvolvimento, incluindo a PEA, devido ao seu papel na modulação da excitabilidade neuronal. A redução da atividade inibitória pode aumentar a excitabilidade neuronal, favorecendo o desenvolvimento de comorbidades como epilepsia, ansiedade e depressão, frequentemente relatadas em indivíduos com PEA (Bjørklund and El-Ansary, 2024).

Diversos estudos têm investigado o papel do sistema GABAérgico, particularmente na supressão de repetições auditivas, uma capacidade fundamental para a percepção e adaptação ao ambiente sonoro. Esta função neurológica, que permite ao cérebro diminuir a resposta a estímulos sonoros repetitivos mantendo simultaneamente a vigilância a sons novos e raros, encontra-se frequentemente comprometida em indivíduos com autismo. Esta disfunção manifesta-se através de uma reatividade neural aumentada a estímulos repetitivos, sugerindo uma falha nos mecanismos de supressão. A investigação atual aponta para uma possível desregulação do sistema GABAérgico como mecanismo subjacente a estas alterações. Esta hipótese é particularmente relevante dado que, embora os indivíduos com PEA mantenham a capacidade de detetar desvios em estímulos auditivos, apresentam dificuldades significativas na supressão de respostas a informações repetitivas e previsíveis, um padrão que sugere um desequilíbrio específico nos mecanismos de regulação auditiva (Huang *et al.*, 2023b).

Evidências também sugerem que níveis alterados de GABA estão associados a um aumento da apoptose em células cerebrais, o que pode levar a alterações estruturais significativas no cérebro durante o desenvolvimento. Esta disfunção pode impactar processos cognitivos fundamentais, incluindo a atenção, a memória e a capacidade de interpretar estímulos sociais, dificultando a adaptação ao ambiente. O impacto da desregulação GABAérgica é ainda mais crítico nas fases iniciais do desenvolvimento, uma vez que a formação adequada das redes neuronais depende de um equilíbrio preciso entre excitação e inibição. Alterações neste sistema podem comprometer a maturação neuronal e predispor o indivíduo a padrões comportamentais repetitivos e dificuldades na interação social (Zhao *et al.*, 2022).

2.2.2 Potencial evocado N170 e processamento facial

O reconhecimento e a interpretação de expressões faciais são fundamentais para a interação social humana, permitindo que os indivíduos infiram os estados emocionais e as intenções dos outros. Esta capacidade é crucial para a adaptação comportamental e a regulação das

interações sociais, facilitando uma comunicação eficaz e a formação de relações interpessoais (Dollion *et al.*, 2022). O processamento facial envolve a identificação de uma pessoa, considerando variáveis como o ângulo de visão e as mudanças na expressão facial, enquanto a interpretação das expressões faciais fornece o contexto emocional necessário para a interação social. Embora estas tarefas sejam processadas por sistemas neuronais distintos, existe uma sobreposição significativa nas áreas cerebrais ativadas. O reconhecimento de rostos é um processo complexo que requer o envolvimento de diferentes mecanismos visuais. Mudanças na orientação do rosto e na expressão podem afetar o reconhecimento de rostos não familiares, mas a capacidade de reconhecer rostos familiares tende a ser resistente a essas variações (Posamentier and Abdi, 2003).

O processamento facial e a percepção visual estão intimamente ligados às dificuldades que indivíduos com PEA enfrentam no reconhecimento e interpretação de expressões faciais. Indivíduos com esta condição frequentemente apresentam dificuldades em descodificar expressões emocionais. Estas dificuldades podem comprometer a capacidade de compreender as emoções dos outros, impactando negativamente a interação social. Estudos indicam que crianças com autismo podem exibir padrões de exploração visual distintos em comparação com indivíduos neurotípicos, como passar menos tempo a observar áreas comuns para a interpretação emocional, como os olhos e a boca, e mais tempo em áreas menos relevantes (Dollion *et al.*, 2022).

A investigação sobre o processamento facial em indivíduos com PEA é fundamental para compreender os mecanismos subjacentes às suas dificuldades na interpretação de expressões emocionais e, conseqüentemente, no desenvolvimento de estratégias de intervenção mais eficazes (Vlamings *et al.*, 2010). Estudos recentes têm explorado diversos métodos avançados, para identificar padrões atípicos na percepção visual e na ativação cerebral durante a análise de rostos (Shu *et al.*, 2024).

O ERP (do inglês, *event-related-potencial*) N170 é um componente bastante relevante no estudo do processamento facial e da percepção visual, sendo amplamente explorado em investigações recentes no autismo. Este componente, geralmente observado num eletroencefalograma (EEG) é caracterizado por uma deflexão negativa que ocorre aproximadamente 170 milissegundos (com um intervalo de variabilidade entre 130 e 200 ms) após a apresentação de um estímulo visual. Esta latência específica é o que lhe confere a designação "N170". O N170 é predominantemente registado nas regiões occipito-temporais do cérebro, particularmente em áreas associadas ao processamento visual e facial, como o giro fusiforme e regiões temporais (Tautvydaitė *et al.*, 2022). Este potencial evocado é altamente sensível ao reconhecimento global de rostos e também a componentes individuais de rostos, como olhos e bocas. Uma característica notável é que a sua resposta ocorre de forma automática, independentemente da atenção ou da familiaridade do rosto apresentado. Esta característica indica que este é um reflexo do processamento neuronal que ocorre rapidamente durante a visualização de rostos (Roberts and Trossman, 2024).

Estudos científicos demonstram que o componente N170 apresenta maior amplitude na resposta cerebral a estímulos faciais comparativamente a objetos não faciais, sendo um

marcador neuronal importante no reconhecimento e processamento facial (Shu *et al.*, 2024). Existe, de facto, uma distinção clara entre a forma de processar estímulos faciais e outros tipos de estímulos visuais, sugerindo que o cérebro reconhece e processa rostos de uma forma singular, possivelmente devido à sua importância social e evolutiva (Kala *et al.*, 2021). O N170 está diretamente relacionado com a capacidade de interpretação de expressões faciais e identificação de indivíduos. A sua amplitude é influenciada por características específicas como a orientação do rosto, sendo menos pronunciada em faces invertidas, e por expressões emocionais, elementos fundamentais para interações sociais eficazes (Tautvydaitė *et al.*, 2022).

A análise do componente N170 revela variações significativas na sua amplitude e latência, correlacionadas com diversos fatores. A idade surge como um determinante importante, observando-se alterações ao longo das diferentes fases do desenvolvimento. Inicialmente, o N170 surge rapidamente na infância, já que as redes neuronais específicas para o processamento de rostos formam-se cedo e continuam a desenvolver-se ao longo dos anos. Durante este período, o N170 começa a exibir características distintivas em resposta a estímulos faciais. À medida que os indivíduos envelhecem, o N170 passa por alterações significativas, com a amplitude a aumentar e a latência a diminuir (Roberts and Trossman, 2024). A capacidade cognitiva geral, incluindo o quociente de inteligência, também influencia as respostas do N170, sendo que indivíduos com maior capacidade cognitiva podem desenvolver estratégias compensatórias no reconhecimento facial, afetando assim as características deste componente neurológico (Kang *et al.*, 2018). Este ERP apresenta maior amplitude perante faces familiares, particularmente de pessoas próximas, em comparação com faces desconhecidas. Esta noção desafia os modelos neurocognitivos tradicionais, que separam as fases perceptuais e mnésicas no reconhecimento facial. O processamento cerebral integra simultaneamente aspetos físicos e informações de familiaridade desde o início do processo perceptivo, indicando que a experiência prévia influencia a percepção facial imediata (Caharel and Rossion, 2021).

O componente N170 manifesta elevada sensibilidade não só à presença de rostos, mas também às emoções e expressões faciais exibidas. A sua latência apresenta variações significativas consoante a expressão facial observada, sendo que faces com emoções negativas, como o medo, provocam respostas distintas comparativamente a expressões neutras ou de felicidade. De facto, o cérebro realiza simultaneamente o reconhecimento facial e a categorização emocional, o que evidencia, de certa forma, a complexidade e importância do processamento emocional nas interações sociais (Kala *et al.*, 2021).

Indivíduos com autismo apresentam diferenças observáveis no componente N170, especialmente em termos de latência, quando comparados a indivíduos com desenvolvimento típico. Estas diferenças manifestam-se principalmente num atraso significativo na resposta a estímulos faciais, indicando uma possível disfunção nos mecanismos neuronais envolvidos no reconhecimento facial. Esta discrepância é particularmente evidente em amostras adultas, em contraste com crianças, o que indica que, enquanto a latência do N170 tende a diminuir com a idade em indivíduos com desenvolvimento típico, em indivíduos com autismo este desenvolvimento pode ser mais lento ou incompleto (Kang *et al.*, 2018). A latência prolongada do N170 em indivíduos com PEA funciona como um indicador neurobiológico de um

processamento facial menos eficiente, estando diretamente associada a dificuldades no reconhecimento e interpretação de expressões faciais. Esta característica tem implicações significativas na capacidade de comunicação não verbal e, conseqüentemente, na qualidade das interações sociais destes indivíduos (Kala *et al.*, 2021).

Determinados estudos afirmam que crianças com PEA apresentam amplitudes do N170 significativamente superiores em resposta a faces que expressam emoções negativas, comparativamente a crianças com desenvolvimento típico. Contudo, não se observam diferenças significativas nas amplitudes do N170 em resposta a outros tipos de estímulos faciais (expressões positivas ou neutras) ou não faciais entre os grupos. Esta reação neuronal diferenciada indica uma possível sensibilidade aumentada a emoções negativas, podendo estar associada às dificuldades características no processamento e reconhecimento de expressões faciais por parte de crianças com autismo (Shu *et al.*, 2024). Outros estudos apresentam resultados diversos, indicando que em análises gerais não existem diferenças significativas na amplitude do N170 entre indivíduos com PEA e indivíduos com desenvolvimento típico. No entanto, em análises de subgrupos específicos, nomeadamente adultos com PEA, pode observa-se uma amplitude do N170 menor (Kang *et al.*, 2018; Márquez *et al.*, 2019).

O denominado efeito de inversão facial encontra-se alterado em indivíduos com um neurodesenvolvimento típico. Este caracteriza-se pela diferença na resposta neuronal durante o processamento de faces em orientação normal comparativamente a faces invertidas, sendo que, a resposta do componente N170 é tipicamente mais pronunciada para faces em orientação normal. Em indivíduos com PEA, este efeito apresenta-se atenuado, evidenciando uma menor diferenciação na resposta do N170 entre faces normais e invertidas (Tavares *et al.*, 2016; Márquez *et al.*, 2019).

A especialização hemisférica do componente N170 também é distinta em indivíduos com autismo. Enquanto os indivíduos com desenvolvimento típico apresentam uma resposta localizada no hemisfério direito e maior ativação no giro fusiforme direito, os indivíduos com PEA demonstram um padrão mais difuso ou equilibrado de resposta eletrofisiológica entre hemisférios, com reduzida ativação do fusiforme. Esta falta de lateralização pode indicar um processamento facial menos especializado e mais global, podendo resultar numa extração menos eficiente de informações emocionais e sociais (Kang *et al.*, 2018).

A utilização do N170 como biomarcador no diagnóstico e terapêutica da PEA oferece uma abordagem mais objetiva, reduzindo a dependência de avaliações subjetivas, que são comuns num contexto onde os diagnósticos, frequentemente, se baseiam em observações comportamentais. Este componente pode ser utilizado para monitorizar a eficácia de intervenções comportamentais. Um estudo indicou que a latência do N170 diminuiu após 16 semanas de tratamento com resposta pivotal, evidenciando que este biomarcador poderá ser utilizado para observar mudanças neuronais associadas a melhorias no comportamento social (Kala *et al.*, 2021).

Um estudo recente demonstrou o potencial do componente N170 como biomarcador da desregulação do equilíbrio E/I, estabelecendo uma correlação significativa entre os níveis de

GABA no córtex visual e as características do N170 (amplitude e latência). Os investigadores verificaram que esta resposta neuronal, intimamente ligada à velocidade de processamento de expressões emocionais, apresentava uma associação com um desempenho reduzido nas competências de comunicação social, que foi avaliado através do questionário de comunicação social. Estes resultados mostram que as assinaturas eletrofisiológicas específicas do N170 podem funcionar como indicadores da desregulação E/I, particularmente em contextos relacionados com capacidades sociais e emocionais (Sousa *et al.*, 2023).

A capacidade de quantificar alterações no processamento facial pode ajudar a avaliar a eficácia de diferentes abordagens terapêuticas e a ajustar intervenções conforme necessário. A identificação de alterações no N170 pode também contribuir para entender quais aspetos do tratamento são mais eficazes em melhorar o processamento social e a comunicação em indivíduos com autismo (Kala *et al.*, 2021).

Um outro componente bastante associado à modulação do N170 é o P100, influenciando diretamente a sua amplitude e topografia através de uma relação temporal e funcional particularmente significativa no contexto do processamento facial. Este componente eletrofisiológico, gerado predominantemente no córtex visual extraestriado, constitui uma resposta precoce da atividade cortical visual, manifestando-se aproximadamente 100 ms após a apresentação de um estímulo visual e demonstrando particular sensibilidade a parâmetros físicos do estímulo, como contraste, luminância, cor, tamanho e frequência espacial (Sousa *et al.*, 2023). Num estudo recente, observou-se que, em indivíduos com PEA, o P100 apresentou latências e amplitudes similares aos grupos de controlo, sugerindo um processamento visual inicial preservado. No entanto, observaram-se alterações significativas no N170, similarmente a outros estudos. Os autores deste estudo concluíram que as alterações no processamento facial observadas não resultavam de deficiências no processamento visual fundamental, mas de uma disfunção específica nos mecanismos de processamento facial, provavelmente relacionada com as regiões neuronais especializadas na perceção facial. Torna-se assim fundamental aprofundar a compreensão da relação entre o P100 e o N170 no processamento facial em indivíduos com PEA, visando elucidar os mecanismos neurobiológicos subjacentes às alterações observadas e o seu potencial como biomarcadores de resposta terapêutica (Parker *et al.*, 2021).

2.2.3 Modulação do equilíbrio E/I com arbaclofeno

Como já referido anteriormente, o equilíbrio E/I neuronal representa um mecanismo essencial para o funcionamento cerebral adequado, desempenhando uma função determinante na regulação de processos cognitivos, emocionais e comportamentais. A sua importância evidencia-se particularmente em diversas condições neurológicas e psiquiátricas, nomeadamente na PEA, esquizofrenia, depressão e perturbações de ansiedade, onde a sua desregulação pode desencadear sintomatologia severa e significativo comprometimento funcional (Chen *et al.*, 2022). A compreensão e modulação deste fenómeno têm-se revelado em áreas de intenso interesse científico e clínico, sendo que as intervenções farmacológicas direcionadas à modulação do equilíbrio E/I visam o seu restabelecimento ou ajuste em

condições onde este se encontra comprometido. Esta abordagem terapêutica demonstra potencial na otimização da sintomatologia e, conseqüentemente, na qualidade de vida dos pacientes, apresentando resultados particularmente promissores quando integrada com terapias comportamentais e psicossociais (Huang *et al.*, 2024).

No contexto da modulação farmacológica do equilíbrio E/I, existem alguns fármacos que estão a ser estudados atualmente, como o arbaclofeno. O arbaclofeno (STX-209) é um fármaco que atua como um agonista seletivo dos recetores inibitórios GABA-B, sendo investigado para potenciais usos terapêuticos, especialmente no tratamento de sintomas relacionados à PEA e outras condições neuropsiquiátricas. Este composto é o enantiómero ativo R do baclofeno, um conhecido agonista do recetor inibitório GABA-B (Roberts *et al.*, 2019). O mecanismo de ação do arbaclofeno envolve a ligação e ativação dos recetores GABA-B, presentes em várias áreas do cérebro, resultando num aumento da atividade inibitória neuronal. Esta ativação potencia os efeitos do GABA, o principal neurotransmissor inibitório do sistema nervoso central, levando a uma maior hiperpolarização das células nervosas e diminuindo a excitabilidade neuronal (Frye, 2014).

O arbaclofeno, administrado por via oral, apresenta características farmacocinéticas superiores comparativamente ao baclofeno racémico, demonstrando uma taxa de absorção mais elevada e uma eliminação mais célere do organismo, devido à ausência de subprodutos metabólicos significativos. A concentração plasmática do arbaclofeno atinge o seu pico aproximadamente 1 a 2 horas após a administração e apresenta uma semivida terminal ($t_{1/2}$) de aproximadamente 5 horas e meia. Estas propriedades farmacocinéticas distintivas não só potenciam a sua eficácia terapêutica, como também contribuem para um perfil de tolerabilidade mais favorável, evidenciando uma redução na prevalência de efeitos secundários adversos tipicamente associados à administração de doses elevadas de baclofeno, nomeadamente sonolência e fadiga (Sanchez-Ponce *et al.*, 2012; Frye, 2014).

Através da ativação da via GABAérgica, o arbaclofeno modula a libertação de outros neurotransmissores importantes, como a dopamina, crucial para o controlo motor, motivação e processamento de recompensa, a norepinefrina, associada à atenção e resposta ao stress, e, também a libertação pré-sináptica do glutamato. A regulação destes neurotransmissores permite a otimização da comunicação entre neurónios nos circuitos envolvidos em funções cognitivas e comportamentais, potenciando assim a eficácia das redes neuronais associadas a estas funções (Gundersen *et al.*, 2023).

O arbaclofeno apresenta potencial terapêutico em diversas condições neurológicas, não só na PEA como já foi referido, mas também na síndrome do X frágil. No contexto da síndrome do X frágil, o arbaclofeno tem sido investigado como potencial tratamento para a evitação social, um sintoma comportamental característico desta condição. A sua ação como agonista GABA-B fundamenta-se em dados pré-clínicos que sugerem a sua capacidade de corrigir perturbações moleculares e comportamentais subjacentes, nomeadamente a síntese proteica excessiva e a diminuição da inibição neuronal. Embora os ensaios clínicos não tenham demonstrado resultados significativos no resultado primário de melhoria da evitação social, observou-se uma potencial eficácia em crianças mais jovens (5-11 anos), particularmente com doses mais

elevadas (Berry-Kravis *et al.*, 2017). Análises secundárias revelaram melhorias significativas em 13% dos participantes em grupos de toma de arbaclofeno, bem como no domínio de socialização, sendo estas respostas mais evidentes em indivíduos com maior funcionalidade (Veenstra-VanderWeele *et al.*, 2017).

Como descrito anteriormente, a disfunção do sistema GABAérgico tem sido associada a alterações significativas no processamento sensorial do ambiente sonoro em indivíduos com autismo. Neste contexto, o arbaclofeno tem sido utilizado como uma ferramenta farmacológica de investigação, permitindo explorar a modulação do sistema GABAérgico e as suas implicações no processamento auditivo. A sua ação como agonista seletivo dos recetores GABA-B tem sido estudada na regulação da excitabilidade neuronal e no restabelecimento da capacidade de supressão auditiva em indivíduos com esta condição, através da administração de diferentes dosagens (Huang *et al.*, 2023b). A administração do arbaclofeno visa avaliar o impacto da ativação dos recetores GABA-B nos mecanismos de supressão de repetição e deteção de desvios em tarefas auditivas, partindo da premissa de que o aumento da atividade GABAérgica poderá contribuir para a normalização das respostas a estímulos auditivos em indivíduos com PEA. Esta abordagem metodológica permite não só elucidar os mecanismos neuronais subjacentes ao processamento auditivo atípico, como também fornece dados sobre a influência das alterações GABAérgicas na perceção sensorial. A análise das respostas individuais ao arbaclofeno possibilita também a identificação de subgrupos específicos dentro da população com este distúrbio, que poderão beneficiar de intervenções terapêuticas direcionadas à modulação do sistema GABAérgico (Huang *et al.*, 2023a).

A investigação das alterações na função GABAérgica e no equilíbrio E/I em indivíduos com PEA tem revelado manifestações significativas ao nível do processamento sensorial retiniano, uma área que permanecia insuficientemente explorada na literatura científica. Estudos recentes têm demonstrado que as disfunções neste equilíbrio podem resultar em padrões de hipersensibilidade ou hiperresponsividade durante o processamento de sinais sensoriais visuais, manifestações frequentemente observadas nesta população (Guimarães-Souza *et al.*, 2019). Neste contexto, a utilização do arbaclofeno como modulador do sistema GABAérgico tem permitido investigar especificamente o impacto da ativação dos recetores GABA-B nas respostas eletrorretinográficas. Os resultados evidenciam que, em condições basais, os indivíduos com autismo apresentam amplitudes da onda-a significativamente mais proeminentes em resposta a estímulos luminosos, sugerindo uma hipersensibilidade retiniana característica. A administração de arbaclofeno demonstrou um efeito diferencial notável, provocando uma redução da amplitude desta onda em indivíduos com PEA, contrastando com um aumento observado em indivíduos com desenvolvimento típico (Huang *et al.*, 2024).

Não obstante as suas vantagens terapêuticas, a administração de arbaclofeno requer uma monitorização cuidadosa, particularmente em populações vulneráveis como adolescentes e crianças com PEA, devido à presença de efeitos secundários significativos. Estudos clínicos têm demonstrado uma correlação positiva entre a dosagem administrada e a incidência de efeitos adversos. Numa investigação com 30 participantes que receberam uma dose de 15 mg de arbaclofeno, 9 indivíduos (5 com PEA e 4 controlos) reportaram efeitos adversos, enquanto

num grupo de 26 participantes que receberam uma dose de 30 mg, 18 indivíduos (9 com PEA e 9 controlos) descreveram efeitos classificados como moderados, incluindo fadiga, tonturas e náuseas (Huang *et al.*, 2022). Também se tem observado que indivíduos com comorbilidades ao nível da saúde mental podem apresentar um risco acrescido de exacerbação de sintomatologia depressiva ou ansiosa, bem como potencial depressão do sistema nervoso central. A heterogeneidade das respostas terapêuticas observadas sugere a existência de perfis distintos de resposta ao fármaco, sublinhando a importância da identificação precoce de "respondedores" e "não-respondedores" durante a fase de investigação clínica. Este facto sublinha a necessidade de um acompanhamento rigoroso e sistemático dos indivíduos sob terapêutica com arbaclofeno, permitindo ajustes posológicos ou alterações na abordagem terapêutica em função da resposta individual observada (Frye, 2014; Roberts *et al.*, 2019).

A integração do arbaclofeno em abordagens terapêuticas multimodais tem surgido como uma estratégia promissora no tratamento de perturbações do neurodesenvolvimento. Os estudos pré-clínicos têm demonstrado o potencial deste fármaco na melhoria de comportamentos e défices cognitivos em ratos, sugerindo benefícios significativos quando associado a intervenções não farmacológicas. Esta abordagem integrada, que combina a terapêutica farmacológica com intervenções comportamentais direcionadas, visa potenciar os resultados terapêuticos através de uma ação sinérgica, particularmente no desenvolvimento de competências sociais e comunicacionais. A implementação desta estratégia requer uma abordagem individualizada e personalizada, considerando as necessidades específicas e as respostas terapêuticas de cada indivíduo. Atualmente, ensaios clínicos em curso procuram avaliar a eficácia desta abordagem multimodal em contextos clínicos, visando estabelecer diretrizes baseadas em evidências para a otimização dos protocolos terapêuticos e a sua integração nas práticas clínicas convencionais (Gundersen *et al.*, 2023).

2.3 Pré-processamento de Eletroencefalografia (EEG)

A crescente complexidade e volume de dados gerados pelos avanços tecnológicos nas diversas áreas científicas têm revolucionado a forma como a informação é adquirida e processada. Os sistemas modernos de aquisição de dados produzem informação em diversos formatos e dimensões, frequentemente acompanhada de variações indesejadas e artefactos que podem comprometer a sua interpretação. Esta realidade tem justificado a necessidade de metodologias complexas de tratamento e análise de dados, sendo o pré-processamento essencial para garantir a precisão e a fidedignidade das análises subsequente (Mishra *et al.*, 2020).

O pré-processamento de dados constitui uma etapa fundamental no processo de Descoberta de Conhecimento em Dados (KDD - *Knowledge Discovery from Data*), sendo imprescindível para a adaptação dos dados às exigências específicas dos algoritmos de mineração. Esta fase preliminar engloba um conjunto diversificado de técnicas, incluindo preparação, redução, limpeza, normalização, integração e seleção de características, visando a otimização da qualidade e utilidade dos dados para análises subsequentes (García *et al.*, 2016; Bhaya and

Alasadi, 2017). A sua relevância reside na capacidade de abordar problemas frequentemente presentes em dados brutos, como inconsistências, redundâncias, valores ausentes e *outliers*, que podem comprometer significativamente a fiabilidade dos resultados obtidos. O pré-processamento permite não só a redução do volume de dados para facilitar o processamento, como também a identificação de relações entre variáveis e a normalização dos dados para garantir a sua comparabilidade. A eficácia desta etapa é determinada pela qualidade do conhecimento extraído, uma vez que dados de baixa qualidade podem resultar em análises enviesadas ou incorretas, comprometendo processos de tomada de decisão subsequentes (García *et al.*, 2016). A aplicação apropriada de métodos de pré-processamento é essencial para garantir um desempenho satisfatório das técnicas de análise de dados, sendo que a sua ausência ou implementação inadequada pode comprometer significativamente a performance dos algoritmos de aprendizagem e mineração (Çetin and Yıldız, 2022).

No contexto específico da aquisição de sinal de EEG, o pré-processamento assume particular relevância, dada a natureza complexa e sensível dos sinais cerebrais registados. Os dados de EEG captados através do couro cabeludo raramente representam exclusivamente a atividade elétrica cerebral, sendo frequentemente contaminados por diversos tipos de ruído e artefactos. Estes elementos interferentes podem ter origem fisiológica, como movimentos oculares e atividade muscular, ou origem externa, como interferências eletromagnéticas, podendo ocultar informações sobre o estado funcional e condição mental do indivíduo (Suveetha Dhanaselvam and Nadia Chellam, 2023). O processo de pré-processamento em EEG engloba uma série de etapas metodológicas específicas, incluindo a remoção de ruídos de linha, ajuste de referências, eliminação de canais comprometidos e remoção de artefactos, visando a extração exclusiva da atividade cerebral relevante. A implementação rigorosa destas técnicas é fundamental para remodelar os dados brutos num formato utilizável, melhorando a separabilidade das características e, conseqüentemente, otimizando o desempenho da classificação em análises posteriores (Kim, 2018).

A aquisição de dados de EEG tem evoluído significativamente, transitando de paradigmas experimentais tradicionais, realizados em ambientes controlados com um número limitado de participantes, para protocolos de recolha em larga escala em condições mais naturalistas. Esta evolução tem sido impulsionada pelos avanços tecnológicos e pelo reconhecimento do potencial do mapeamento cerebral em contextos do mundo real (Bigdely-Shamlo *et al.*, 2015). O processo de aquisição envolve a utilização de múltiplos elétrodos posicionados no couro cabeludo, podendo ser implementado através de duas configurações principais: a montagem bipolar, que analisa o diferencial de voltagem entre elétrodos em áreas eletricamente ativas, e a montagem monopolar/unipolar, que utiliza um elétrodo ativo em conjunto com elétrodos de referência (Shoka *et al.*, 2019). Os elétrodos são frequentemente posicionados em localizações anatómicas específicas, como os mastoides, que funcionam como pontos de referência devido à sua reduzida interferência com os sinais neuronais. Em sistemas de EEG de alta densidade, caracterizados pela utilização de mais de 100 elétrodos, é frequentemente recomendada a utilização de uma média dos dados de referência, assumindo-se que o somatório dos valores elétricos registados nos elétrodos do couro cabeludo apresenta um valor nulo (Peng, 2019). Os sistemas de aquisição modernos, como OpenBCI, EMOTIV, QUASAR e BioSemi, oferecem

diferentes capacidades em termos de amplificação, número de canais e portabilidade, sendo os dados posteriormente processados através de um software especializado, como por exemplo o MATLAB. A preparação adequada do ambiente e do participante, bem como a implementação de protocolos de registo rigorosos, permitem minimizar interferências e garantir a qualidade dos dados recolhidos, particularmente em registos de longa duração necessários para determinadas condições clínicas (Jain *et al.*, 2024).

Como anteriormente referido, o pré-processamento de sinais de EEG compreende diversas etapas metodológicas específicas, cuja descrição detalhada será apresentada nas secções subsequentes. Na Figura 1 pode se observar um diagrama com uma sequência de etapas geralmente realizada durante o pré-processamento de EEG.

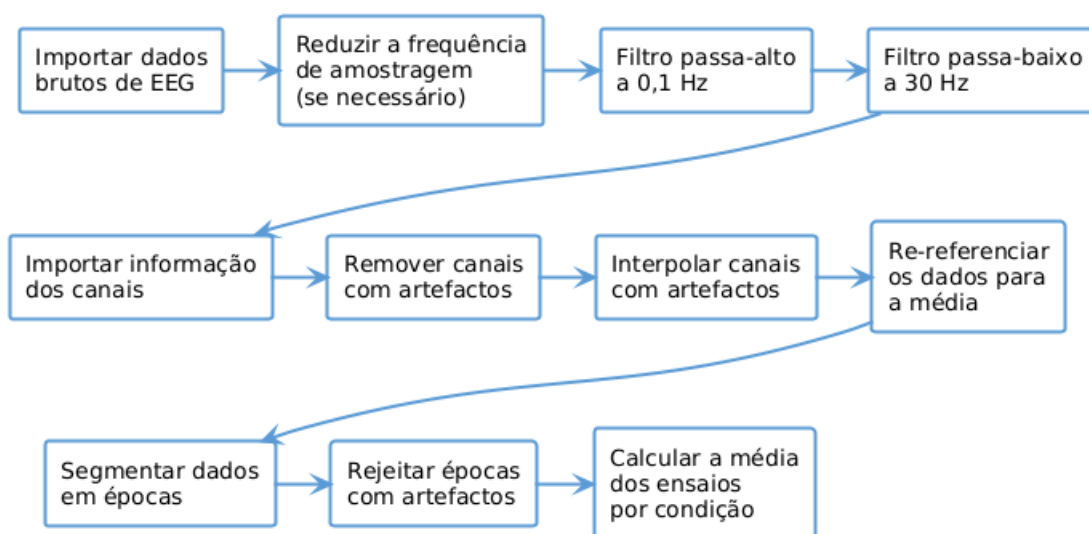


Figura 1 - Sequência de etapas de pré-processamento.

Antes de descrever cada uma das etapas do pré-processamento, é também relevante descrever as características comuns de um sinal EEG.

2.3.1 Características do Sinal EEG

A revisão das características comuns do sinal EEG é fundamental para a compreensão e análise de sinais deste tipo. A atividade elétrica cerebral registada através do EEG encontra-se tipicamente circunscrita a uma faixa de frequência entre 0.01 e 100 Hz, abrangendo as diversas oscilações relevantes para a análise do funcionamento cerebral. Estas oscilações caracterizam-se pela sua natureza não periódica e imprevisível, resultando da atividade coordenada de milhares de milhões de neurónios (Kim, 2018). A amplitude dos sinais, geralmente variando dentro de uma faixa de $\pm 100 \mu\text{V}$, indica a intensidade da atividade elétrica cerebral, apresentando variações significativas em função do estado do indivíduo, como vigília ou sono (Siuly, Li and Zhang, 2016).

Os ritmos cerebrais manifestam-se através de cinco bandas de frequência distintas, cada uma associada a estados mentais e funções cognitivas específicas. A banda delta (δ), compreendida entre 0.5-4 Hz, encontra-se tipicamente associada ao sono profundo; a banda theta (θ), entre 4-8 Hz, relaciona-se com estados de sonolência e relaxamento; a banda alpha (α), entre 8-13 Hz, manifesta-se predominantemente durante o relaxamento e meditação; a banda beta (β), entre 13-30 Hz, está associada a atividade mental intensa e concentração; e a banda gamma (γ), acima de 30 Hz, relaciona-se com processos cognitivos de elevada complexidade (Kim, 2018). A compreensão destas características oscilatórias é importante para a interpretação da dinâmica da atividade cerebral, permitindo a identificação de características morfológicas específicas e a análise de padrões associados a diferentes estados mentais ou de consciência. É importante salientar que a interpretação destes sinais pode ser complexificada pela presença de artefactos de origem não cerebral, cujas amplitudes podem ser comparáveis aos sinais corticais de interesse (Siuly, Li and Zhang, 2016).

2.3.2 Reamostragem

A reamostragem constitui uma etapa fundamental no pré-processamento de sinais de EEG, visando a otimização da resolução dos dados adquiridos e a facilitação das análises subsequentes. Este processo envolve a modificação da taxa de amostragem dos dados de EEG, podendo ser implementado através de duas abordagens principais: o *upsampling*, que visa o aumento da taxa de amostragem através da geração de novos pontos de dados para incrementar a resolução do sinal, e o *downsampling*, que consiste na redução da taxa de amostragem através da recolha de dados em intervalos mais amplos (Suveetha Dhanaselvam and Nadia Chellam, 2023). A reamostragem pode ter múltiplas funções, incluindo a uniformização das taxas de amostragem entre diferentes conjuntos de dados, permitindo a sua comparação e análise conjunta, e a melhoria da qualidade dos dados, particularmente relevante quando estes são adquiridos em diferentes contextos ou com diferentes dispositivos. Este processo facilita também a reconstrução de frequências para análise posterior e permite a otimização dos recursos informáticos relacionados com o armazenamento e velocidade de processamento (Kyriaki, Koukopoulos and Fidas, 2024).

2.3.3 Filtragem

A filtragem constitui uma etapa metodológica fundamental no pré-processamento de sinais de EEG, sendo crucial para a otimização da qualidade dos dados através da minimização de ruídos e artefactos. Este processo baseia-se na aplicação de filtros digitais e técnicas matemáticas específicas que visam atenuar ou eliminar componentes indesejados do sinal, como ruídos de linha (50 Hz ou 60 Hz), ruídos de alta frequência e perturbações em baixas frequências, permitindo a preservação seletiva das frequências relevantes para análise. A implementação desta metodologia pode ser adaptada através de diferentes tipos de filtros, projetados especificamente para responder a necessidades particulares do processamento de sinal. A implementação da filtragem pode ser realizada através de diferentes tipos de filtros digitais: o

filtro passa-baixo, que retém sinais de frequências baixas e atenua frequências altas; o filtro passa-alto, que mantém sinais com frequências acima de um determinado limiar; o filtro passa-banda, que preserva sinais dentro de uma faixa específica de frequências; e o filtro rejeita-banda, que elimina sinais numa faixa específica de frequências (Peng, 2019). Na Figura 2 observa-se uma representação gráfica dos filtros descritos anteriormente.

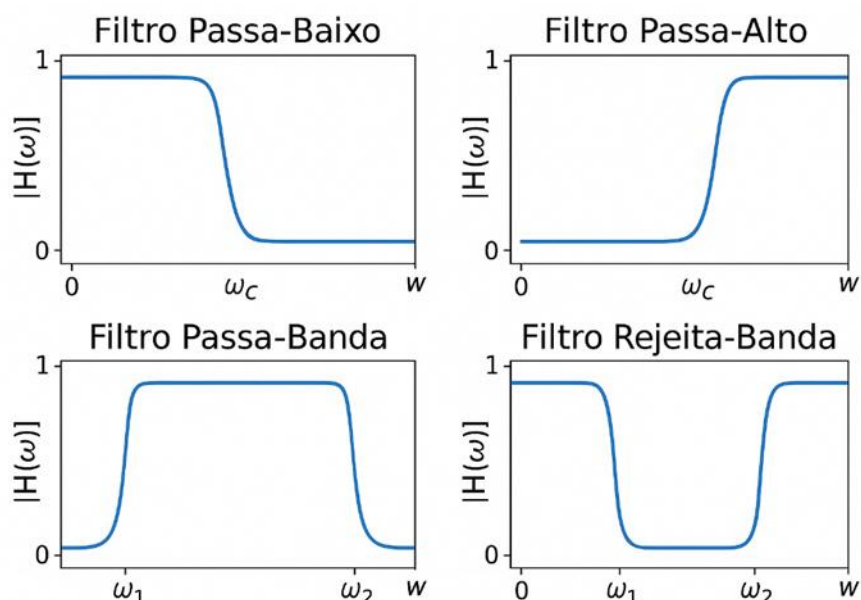


Figura 2 - Tipos de filtros mais comuns (adaptado de Peng, 2019).

A implementação da filtragem visa dois objetivos principais: a remoção de ruídos e o isolamento de frequências específicas de interesse. Os sinais de EEG são frequentemente contaminados por artefactos de origem fisiológica, como movimentos oculares e atividade muscular, e não fisiológica, como interferências eletromagnéticas, cuja eliminação é fulcral para a precisão das análises subsequentes. Esta metodologia permite também o isolamento seletivo de bandas de frequência específicas (theta, alfa, beta), que se encontram associadas a diferentes estados mentais e atividades cerebrais, facilitando assim estudos sobre funções cognitivas e diagnósticos clínicos. Este processo de otimização do sinal é particularmente relevante para análises específicas, como a deteção de ERPs ou a classificação de sinais motores (Suveetha Dhanaselvam and Nadia Chellam, 2023).

A implementação da filtragem requer considerações metodológicas específicas para garantir a sua eficácia e evitar a introdução de artefactos adicionais. É fundamental que o processo de filtragem seja aplicado aos dados contínuos previamente à sua segmentação em épocas, uma vez que a aplicação posterior pode resultar na introdução de artefactos nas extremidades dos segmentos. Relativamente aos parâmetros de filtragem, é preconizada a utilização de um filtro passa-alta de 0.1 Hz para minimização de *drifts* de baixa frequência, em conjugação com um filtro passa-baixo de 30 Hz para eliminação de ruído de alta frequência associado à atividade muscular (Peng, 2019). Adicionalmente, métodos mais sofisticados podem ser implementados para uma discriminação mais precisa de componentes não originários da atividade cerebral (Bigdely-Shamlo *et al.*, 2015; Peng, 2019).

2.3.4 Detecção e Remoção de Canais Ruidosos

A deteção de canais comprometidos constitui uma etapa metodológica fundamental no pré-processamento de sinais de EEG, sendo crucial para garantir a qualidade e fiabilidade dos dados previamente à sua análise. Este processo visa a identificação de canais que apresentam sinais contaminados, ruidosos ou com baixa razão sinal-ruído, manifestando-se através de sinais muito baixos ou inexistentes durante períodos significativos do registo (Pedroni, Bahreini and Langer, 2019). Estas alterações podem resultar de diversos fatores, incluindo má conexão elétrodo-pele, interferências externas ou ruídos impulsivos que distorcem os dados. A implementação desta metodologia é essencial para otimizar a qualidade do sinal de EEG, aumentar a precisão das análises subsequentes, como a classificação de ondas ou estados mentais, e garantir que o processamento de dados não seja comprometido por artefactos provenientes de canais com desempenho deficiente, assegurando assim que apenas os canais com dados fidedignos sejam utilizados nas análises subsequentes (Kim, 2018).

A implementação da deteção de canais comprometidos envolve diversas abordagens metodológicas complementares. A análise de amplitude é um método bastante comum, onde se utiliza o *z-score* para identificar canais que apresentam amplitudes anormalmente elevadas, indicativas de contaminação. A avaliação da correlação entre canais, particularmente nas componentes de baixa frequência, permite identificar canais que não apresentam os padrões de correlação típicos do EEG normal. A análise da energia em bandas de alta frequência baseia-se no princípio de que canais contaminados tendem a exibir maior energia nestas bandas, sendo a razão entre a potência das componentes de alta frequência e a potência geral utilizada como indicador de qualidade (Bigdely-Shamlo *et al.*, 2015; Kim, 2018). O software *Automagic*, por exemplo, implementa rotinas específicas para esta deteção, identificando canais com ruído excessivo, sinais muito baixos ou comportamentos *outlier*. Após a identificação, estes canais podem ser removidos do conjunto de dados original. Alternativamente, quando apenas um número reduzido de canais é afetado, pode-se recorrer à interpolação dos sinais, visando preservar a continuidade dos dados (Pedroni, Bahreini and Langer, 2019). Esta etapa será descrita com mais detalhe de seguida.

2.3.5 Interpolação dos Canais Removidos

A interpolação de canais removidos constitui uma metodologia que se pode utilizar no pré-processamento de sinais de EEG, apresentando-se como uma alternativa à simples eliminação de canais identificados como não funcionais ou comprometidos. Esta técnica visa estimar e preencher os dados dos canais removidos através da utilização de informação proveniente dos canais adjacentes funcionais, permitindo assim preservar a estrutura e continuidade dos dados. A implementação desta metodologia serve múltiplos propósitos: manutenção da dimensionalidade dos dados, aspeto particularmente relevante para análises em grupo; preservação da integridade da informação, evitando situações onde a remoção excessiva de canais poderia comprometer a validade das análises estatísticas; e otimização da qualidade dos dados, facilitando a visualização e extração de resultados significativos (Peng, 2019; Delorme,

2023). Esta abordagem revela-se particularmente útil quando comparada com a simples eliminação de canais, permitindo a realização de análises mais completas, mesmo na presença de canais originalmente comprometidos por ruído ou artefactos (Delorme, 2023).

A implementação da interpolação de canais segue uma metodologia estruturada em três etapas fundamentais. Inicialmente, procede-se à projeção dos elétrodos, tanto funcionais como comprometidos, numa esfera unitária, permitindo a sua representação no espaço tridimensional. Subsequentemente, é desenvolvida uma matriz de mapeamento que estabelece as relações espaciais entre os canais funcionais e os comprometidos. Por fim, realiza-se o cálculo dos dados interpolados através da aplicação desta matriz de mapeamento, estimando os dados dos canais comprometidos com base nos canais funcionais circundantes. O método de *spline* esférica é frequentemente utilizado para esta finalidade, embora seja importante salientar que este processo pode resultar numa redução da resolução espacial dos dados, uma vez que os canais interpolados representam uma combinação ponderada dos canais funcionais adjacentes, não fornecendo dados verdadeiramente únicos (Peng, 2019).

2.3.6 Referenciação

A referenciação é uma etapa metodológica fundamental no pré-processamento de sinais de EEG que influencia significativamente a interpretação dos dados. Este processo envolve o ajuste dos potenciais registados pelos elétrodos em relação a um ou mais elétrodos de referência, visando refletir com precisão as diferenças de potencial entre os elétrodos de interesse e o ponto de referência selecionado (Kim, 2018). A implementação desta metodologia tem alguns objetivos: normalização dos sinais de EEG, minimizando variabilidades não relacionadas com a atividade cerebral; otimização da análise espacial dos dados, facilitando a interpretação das localizações das fontes de atividade elétrica cerebral; redução de ruído e artefactos, incluindo interferências externas e eletromiografia (EMG); e melhoria da razão sinal-ruído (Peng, 2019). Uma seleção adequada do método de referenciação é particularmente relevante para a identificação precisa de ERPs e outros padrões neurofisiológicos significativos, enquanto uma escolha inadequada pode resultar na introdução de artefactos e comprometer a interpretação dos resultados analíticos (Bigdely-Shamlo *et al.*, 2015).

A implementação da referenciação pode ser realizada através de diversas metodologias, cada uma apresentando características e benefícios específicos. A montagem bipolar baseia-se na comparação direta entre pares de elétrodos adjacentes, sendo particularmente eficaz na deteção de potenciais locais e menos sensível a ruídos distantes, embora apresente limitações na captação de potenciais distantes (Peng, 2019). A montagem referencial utiliza um único eléctrodo de referência, como os lóbulos auriculares, vértice ou nariz, permitindo a captação tanto de potenciais locais como distantes, mas apresentando maior suscetibilidade a artefactos, especialmente quando existe uma distância significativa entre o eléctrodo de referência e os elétrodos de interesse. A referência à mastoide, embora comum, apresenta vulnerabilidades significativas quando ocorrem falhas na conexão, podendo resultar num aumento da variabilidade do sinal (Kim, 2018; Peng, 2019). A referência de média comum (CAR - *Common*

Average Reference) utiliza a média de todos os canais como referência, reduzindo o impacto de falhas isoladas, embora permaneça sensível a canais *outliers* (Bigdely-Shamlo *et al.*, 2015; Kim, 2018). Metodologias mais sofisticadas incluem esquemas de referência multi-estágio, que incorporam a identificação prévia de canais problemáticos, e métodos sistemáticos baseados em considerações físicas, eletrodinâmicas ou abordagens estatísticas. É fundamental salientar a importância da remoção de canais comprometidos previamente à aplicação de métodos como o CAR, evitando assim a contaminação dos dados de referência por sinais não fidedignos (Bigdely-Shamlo *et al.*, 2015).

2.3.7 Extração e Rejeição de Épocas

A extração de épocas é uma etapa imprescindível no processamento de sinais de EEG, consistindo na segmentação dos dados contínuos em intervalos temporais específicos, denominadas “épocas”, que correspondem a períodos de interesse experimental. Esta técnica visa a análise detalhada das alterações na atividade elétrica cerebral em resposta a estímulos ou eventos específicos, permitindo um foco preciso em intervalos temporais relevantes. A implementação desta metodologia visa diversos objetivos para a investigação neurofisiológica. Primariamente, permite a identificação e análise de ERPs, facilitando a compreensão das respostas neuronais a estímulos específicos. Através do isolamento de segmentos temporalmente relevantes, otimiza a clareza da análise comparativamente ao processamento de dados contínuos extensos, possibilitando uma investigação mais direcionada (Peng, 2019). Esta segmentação temporal facilita significativamente a comparação entre diferentes condições experimentais, permitindo a identificação de padrões e diferenças significativas na atividade cerebral. Simultaneamente, a extração de épocas contribui para a otimização da remoção de ruídos e artefactos, aspecto crucial para a qualidade dos dados, e potencia a classificação e identificação de padrões nos sinais de EEG (Kyriaki, Koukopoulos and Fidas, 2024).

A implementação da extração de épocas segue uma metodologia estruturada em várias etapas sequenciais. Inicialmente, procede-se à identificação precisa do evento de interesse que servirá como ponto de referência para a extração, podendo este consistir na apresentação de um estímulo específico, como um estímulo visual ou auditivo, ou numa resposta comportamental do participante. Subsequentemente, estabelece-se a janela temporal apropriada em torno do evento identificado, sendo frequentemente definida para capturar a atividade neuronal tanto prévia como posterior ao evento. (Peng, 2019). Por fim, realiza-se a segmentação efetiva dos dados contínuos de EEG, que pode ser executada manualmente ou através de algoritmos automatizados. Este processo transforma a representação bidimensional original dos dados (elétrodos \times tempo) numa estrutura tridimensional (elétrodos \times tempo \times tentativas), onde cada "tentativa" corresponde a uma ocorrência específica do evento em análise (Peng, 2019; Kyriaki, Koukopoulos and Fidas, 2024).

Após a extração das épocas geralmente é necessário fazer uma correção da linha de base. Este passo permite o ajuste dos dados de EEG através da remoção de variações elétricas que precedem os eventos de interesse. Esta técnica é essencial para a normalização dos dados,

permitindo que a análise subsequente se concentre especificamente nas variações provocadas pelos estímulos ou condições experimentais (Peng, 2019). A sua implementação envolve, a seleção de um intervalo de referência pré-estímulo, tipicamente variando entre 200ms e 1s, durante o qual se assume a ausência de resposta ao evento. Subsequentemente, procede-se ao cálculo da amplitude média durante este intervalo, valor que é posteriormente subtraído de cada ponto temporal da época correspondente (Kyriaki, Koukopoulos and Fidas, 2024). Este processo apresenta desafios metodológicos específicos, nomeadamente a seleção apropriada do intervalo de base, que pode influenciar significativamente os resultados, e a necessidade de assumir a equivalência entre condições durante o intervalo selecionado, pressuposto que pode ser complexo de garantir em determinados paradigmas experimentais (Peng, 2019).

A rejeição de épocas é realizada posteriormente, visando a identificação e exclusão de segmentos contaminados por artefactos ou não representativos da atividade cerebral em estudo. Esta contaminação pode resultar de diversos fatores, incluindo movimentos oculares, sinais de EMG ou outros tipos de ruído que interferem com o registo da atividade cerebral e ainda não tenham sido removidos (Peng, 2019; Kyriaki, Koukopoulos and Fidas, 2024). A implementação desta etapa pode seguir duas abordagens principais: a análise visual, que envolve a inspeção manual das épocas, permitindo a identificação de segmentos problemáticos por especialistas, embora apresente limitações relacionadas com a subjetividade e potenciais inconsistências; e a rejeição automática, baseada em critérios objetivos como valores de amplitude pico a pico que excedem limiares pré-definidos, estatísticas baseadas no desvio padrão em relação à média, ou limites temporais específicos. Esta etapa é fundamental para a otimização da relação sinal-ruído e consequente fiabilidade dos dados para análises subsequentes (Peng, 2019). Contudo, apresenta desafios significativos, nomeadamente o risco de perda de dados potencialmente relevantes e a possibilidade de viés do investigador na seleção manual. Para minimizar estas limitações, recomenda-se geralmente a utilização de limiares padronizados baseados em estatísticas, de modo a reduzir a variabilidade interpessoal e assegurar maior consistência nos dados analisados (Kyriaki, Koukopoulos and Fidas, 2024).

3 Metodologia

3.1 Materiais e Métodos

O presente estudo utilizou os dados obtidos no âmbito de um projeto prévio onde se adotou um desenho experimental duplamente cego, aleatorizado, cruzado, controlado por placebo, envolvendo uma amostra de 70 participantes (41 controles, 29 com PEA). A cada participante foi administrada, em 3 dias de estudo distintos, uma dose oral única de 15 mg ou 30 mg de arbaclofeno (STX209) ou placebo, em forma de comprimido. O arbaclofeno atinge os níveis plasmáticos máximos 1h após a ingestão e tem uma meia-vida de 5h (Sanchez-Ponce *et al.*, 2012). A sessão de EEG foi realizada aproximadamente 3 horas após a toma dos comprimidos, dentro da janela de atividade fisiológica do fármaco. A ordem da toma foi aleatorizada. Em conformidade com o protocolo duplamente cego, nem os investigadores nem os participantes tinham conhecimento da natureza da substância administrada em cada visita (placebo ou fármaco ativo).

Durante toda a sessão experimental foi assegurada supervisão médica especializada, sendo solicitado aos participantes que permanecessem nas instalações da unidade de investigação por um período mínimo de 4 horas após a administração do composto em estudo (arbaclofeno ou placebo).

O protocolo de segurança estabeleceu que o profissional médico responsável, embora mantido em condição cega relativamente à sequência de aleatorização, tivesse acesso imediato aos códigos de aleatorização através da farmácia hospitalar e do investigador principal. Embora não tenham ocorrido situações de emergência que requeressem a quebra do código, implementou-se um protocolo de segurança adicional para participantes que apresentaram eventos adversos classificados como moderados a graves segundo os critérios pré-estabelecidos. Nestes casos específicos, após avaliação clínica detalhada e consulta entre a equipa médica e o investigador principal, tomou-se conhecimento da sequência de tratamento do participante afetado. Esta medida preventiva evitou a potencial exposição a doses superiores de arbaclofeno em visitas subsequentes, salvaguardando assim a segurança dos participantes sem comprometer a integridade do desenho duplamente cego para os restantes elementos do estudo.

3.1.1 Participantes

O recrutamento dos participantes foi realizado através do *National Autism and ADHD Adult Service* (NAASA), integrado no *South London and Maudsley NHS Foundation Trust*, que constitui um centro de referência nacional do Serviço Nacional de Saúde britânico especializado na avaliação e intervenção em adultos com PEA. O protocolo de diagnóstico implementado neste serviço segue as diretrizes internacionais e inclui a administração sistemática de instrumentos padronizados de avaliação. Especificamente, utiliza-se a ADI-R (*Autism Diagnostic Interview-*

Revised) quando disponível um informante com conhecimento adequado do desenvolvimento precoce do indivíduo, permitindo a caracterização retrospectiva da história do desenvolvimento, e/ou a ADOS (*Autism Diagnostic Observation Schedule*) para avaliação objetiva da apresentação sintomatológica atual. A ADOS-2 constitui uma avaliação semiestruturada de 45-60 minutos, organizada em módulos específicos adaptados à idade e nível de linguagem do indivíduo, que permite a observação sistemática de comportamentos sociais e comunicativos em contextos de interação controlada. Por sua vez, a ADI-R complementa esta avaliação através de uma entrevista detalhada aos cuidadores, recolhendo informação sobre o desenvolvimento precoce e comportamentos ao longo da vida em domínios-chave do autismo, com ambos os instrumentos a utilizarem sistemas de pontuação estandardizados que permitem determinar o cumprimento dos critérios diagnósticos estabelecidos (Lebersfeld *et al.*, 2021; Hirota and King, 2023). Outros potenciais participantes foram cuidadosamente avaliados por um clínico da NAASA experiente em autismo, que confirmou o diagnóstico de autismo através do relato da avaliação diagnóstica realizada por um serviço de autismo reconhecido no Reino Unido. A caracterização clínica de todos os participantes foi realizada através da aplicação do Quociente do Espetro do Autismo (AQ), um instrumento de rastreio amplamente validado para a identificação de traços de autismo na população geral e clínica. A avaliação cognitiva dos participantes foi complementada através da utilização da Escala de Inteligência Abreviada de Wechsler II (WASI-II) que permitiu a determinação do Quociente de Inteligência Verbal (QI Verbal), Quociente de Inteligência de *Performance* (QI de *Performance*) e Quociente de Inteligência de Escala Completa (*Full-Scale QI*) (Wechsler, 2011). Foi também realizado o registo de toda a medicação que os participantes tomaram durante o mês anterior à participação. Na Tabela 1 está presente a informação do perfil demográfico dos sujeitos pertencentes à amostra.

Tabela 1 - Perfil demográfico da amostra de voluntários em estudo.

Variáveis	Grupo de Controlo (n = 41)	Grupo de Autismo (n = 29)	P value
Idade			0,016
média	28,54	34,06	
desvio padrão	8,02	10,30	
Género			-
masculino	23	21	
feminino	18	8	
QI Verbal			0,290
média	119,03	115,45	
desvio padrão	12,70	11,77	
QI de Performance			0,512
média	117,72	115,82	
desvio padrão	10,38	11,15	
QI Global			0,296
média	120,33	117,36	
desvio padrão	10,39	10,40	
Quociente de Autismo			<0,001
média	16,25	8,15	
desvio padrão	35,15	7,52	
Medicação			-
Sim	16	18	
Não	25	11	

Como referido anteriormente, a amostra final do estudo compreendeu 70 participantes distribuídos por dois grupos experimentais. O grupo de controlo foi constituído por 41 indivíduos (23 homens e 18 mulheres), com idades compreendidas entre os 19 e os 52 anos ($M = 28,5$ anos; $DP = 8,0$ anos). O grupo de autismo incluiu 29 participantes (21 homens e 8 mulheres), com idades entre os 19 e os 51 anos ($M = 34,1$ anos; $DP = 10,3$ anos). A análise comparativa revelou uma diferença significativa entre os grupos para a idade ($p = 0,016$), com o grupo de autismo a apresentar uma média de idade superior em aproximadamente 5,5 anos relativamente ao grupo de controlo.

A avaliação do Quociente de Autismo (QA) revelou diferenças significativas entre os grupos ($p < 0,001$) como esperado. O grupo com diagnóstico de PEA apresentou pontuações mais elevadas ($M = 35,15$; $DP = 7,52$) comparativamente ao grupo de controlo ($M = 16,25$; $DP = 8,15$), confirmando a presença de características fenotípicas distintivas entre as populações em estudo. Estes valores são consistentes com os limiares estabelecidos na literatura, onde pontuações superiores a 32 são indicativas de níveis clinicamente significativos de traços de autismo, enquanto a população neurotípica apresenta tipicamente valores inferiores a 20 (Ruzich *et al.*, 2015).

Relativamente ao funcionamento cognitivo, a análise comparativa entre grupos revelou ausência de diferenças estatisticamente significativas nas três dimensões avaliadas: QI Verbal ($p = 0,290$), QI de Performance ($p = 0,512$) e QI Global ($p = 0,296$). Esta equivalência no funcionamento cognitivo global entre as amostras assegura que as diferenças observadas noutras variáveis não são atribuíveis a disparidades nas capacidades intelectuais.

Relativamente à medicação tomada pelos participantes, 16 participantes reportaram uso de medicação ocasional, no grupo de controlo, incluindo suplementos vitamínicos, anti-histamínicos para uso sintomático, contraceptivos orais e anti-inflamatórios não esteroides para condições pontuais. No grupo com PEA, 18 participantes encontravam-se sob terapêutica farmacológica regular. Destes, 7 participantes utilizavam medicação psicoativa, incluindo antidepressivos (sertralina em 4 casos), ansiolíticos/hipnóticos (zopiclona, solpadeine PRN), antidepressivos com propriedades ansiolíticas (citalopram, mirtazapina), anticonvulsivantes (fexofenadina), estimulantes (metilfenidato) e analgésicos opioides (codeína). A medicação não psicoativa reportada no grupo PEA incluiu anti-inflamatórios (ibuprofeno), medicação para controlo de sintomas gastrointestinais (omeprazol), anti-histamínicos (ciclizina para vertigens), atorvastatina para controlo lipídico, inaladores para asma, e suplementação vitamínica (vitaminas B12, vitamina D, ómega-3 e multivitaminas). É de salientar que todos os participantes medicados com psicofármacos mantiveram a sua terapêutica estável durante o período do estudo, conforme protocolo estabelecido, minimizando assim potenciais interações farmacológicas com o composto em investigação. A única exceção foi um participante cuja dose de sertralina foi ajustada, mas este ajuste ocorreu antes do início da participação no estudo.

A análise dos dados da ADOS revelou heterogeneidade no perfil de apresentação clínica, com pontuações no domínio da comunicação variando entre 1 e 8 (mediana = 4), e no domínio da interação social recíproca entre 3 e 10 (mediana = 7). As pontuações no domínio da

imaginação/criatividade foram consistentemente baixas (0-2), enquanto os comportamentos estereotipados e interesses restritos apresentaram variabilidade mínima (0-3). Os dados obtidos através da ADI-R corroboraram os da ADOS, demonstrando comprometimento substancial nos três domínios nucleares da PEA. No domínio da comunicação, as pontuações variaram entre 8 e 23, no domínio social entre 9 e 28, e no domínio dos comportamentos repetitivos e interesses restritos entre 0 e 8. É de notar que nem todos os participantes dispunham de avaliação completa por ambos os instrumentos, com dois participantes apresentando apenas confirmação diagnóstica através de um dos instrumentos, mantendo, contudo, critérios clínicos inequívocos para inclusão no estudo segundo avaliação por clínicos especializados.

Os critérios de elegibilidade foram estabelecidos de forma rigorosa para assegurar a homogeneidade da amostra e a validade dos resultados. No grupo de autismo, foram excluídos participantes cujo diagnóstico de PEA estivesse associado a síndromes genéticas específicas identificadas, nomeadamente a síndrome do X frágil ou a síndrome de deleção 22q11, garantindo assim o estudo de formas idiopáticas da perturbação. Os critérios de inclusão para toda a amostra foram: funcionamento intelectual dentro dos parâmetros normativos (QI Global > 70), acuidade visual normal ou devidamente corrigida, capacidade preservada para fornecer consentimento informado esclarecido, ausência de comorbilidades psiquiátricas clinicamente significativas (incluindo perturbações psicóticas e perturbações do humor graves), ausência de história pregressa de atividade convulsiva ou diagnóstico estabelecido de epilepsia, e ausência de condições médicas sistémicas relevantes, particularmente patologia cardiovascular, hipertensão arterial não controlada ou compromisso da função renal.

3.1.2 Aquisições de EEG e Paradigma de Decisão Percetual

O protocolo experimental foi conduzido numa sala com pouca luminosidade nas instalações do *Sackler Institute for Translational Neurodevelopment, King's College London*. Os participantes foram posicionados confortavelmente numa cadeira ergonómica, com estabilização cefálica através de um apoio de queixo ajustável, garantindo assim a manutenção de uma distância constante de 60 cm entre os olhos e o monitor de apresentação dos estímulos. Os estímulos visuais utilizados no presente estudo consistiram em imagens *Mooneys* de faces (Castelhano *et al.*, 2018) e versões distorcidas (*scrambled*) destas mesmas imagens, ambas desenvolvidas no Centro de Imagem Biomédica e Investigação Translacional (CIBIT), do Instituto de Ciências Nucleares Aplicadas à Saúde (ICNAS), da Universidade de Coimbra. A apresentação destes estímulos foi implementada através do software MATLAB 9.2.0 versão R2017a, utilizando a biblioteca Psychophysics Toolbox, que permite o controlo preciso de parâmetros temporais e visuais em paradigmas experimentais (Brainard, 1997; The MathWorks Inc., 2017).

O protocolo de apresentação adotou uma estrutura sequencial, iniciando-se com uma fase de *baseline* caracterizada por um ponto de fixação com duração variável (1 a 1,5 segundos), estratégia implementada para minimizar a previsibilidade temporal e, conseqüentemente, potenciais artefactos antecipatórios. Subsequentemente, procedia-se à apresentação do

estímulo (face ou imagem distorcida) durante um período fixo de 1 segundo, seguida do aparecimento de um ponto de interrogação com igual duração, durante o qual os participantes eram instruídos a executar a resposta dicotómica relativa à natureza do estímulo visualizado (face vs. não face). Esta estrutura permitiu a separação temporal entre o processamento visual do estímulo e a resposta motora, contribuindo para a minimização de artefactos musculares durante a fase de interesse. Na Figura 3 está representado um esquema que permite visualizar a tarefa realizada.

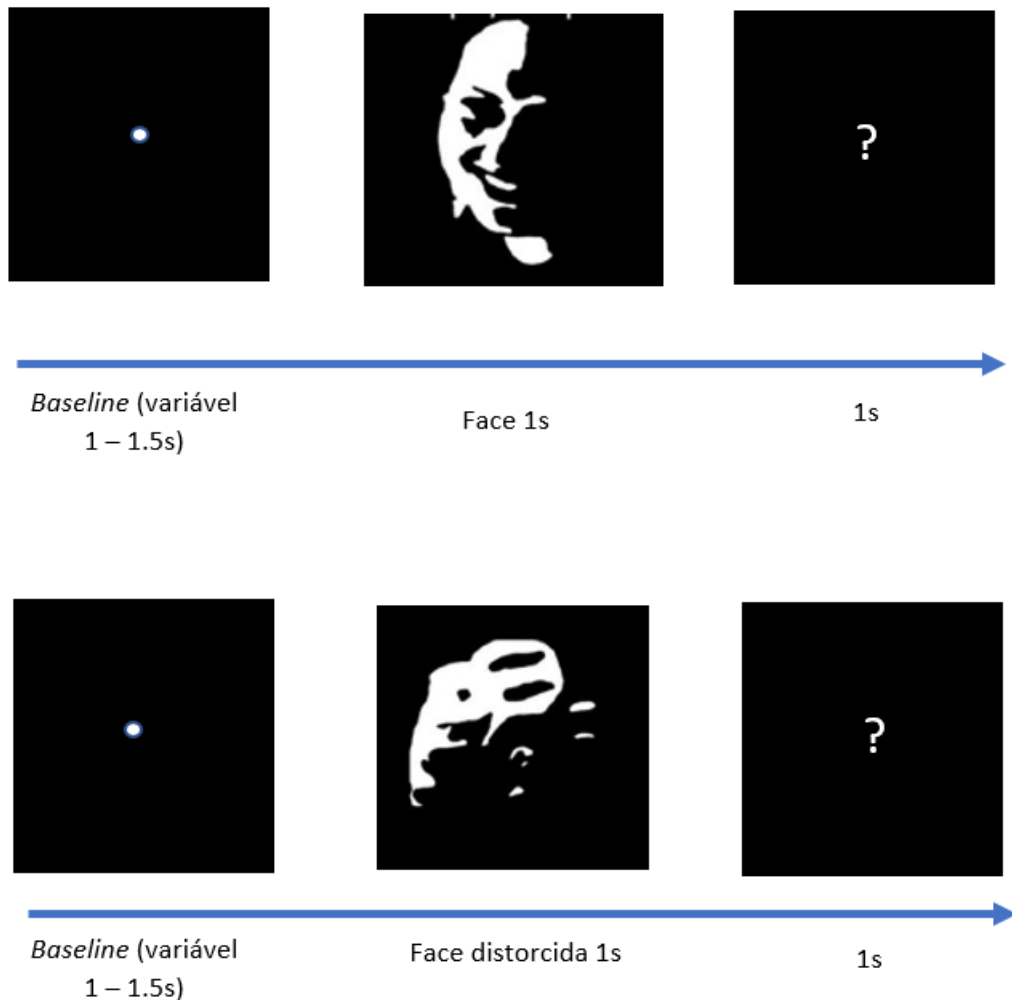


Figura 3 - Esquema da tarefa realizada pelos participantes durante o ensaio clínico. A ordem de apresentação dos estímulos *Mooney* era aleatória.

A aquisição dos dados eletrofisiológicos foi realizada através de um sistema de EEG de alta densidade, utilizando uma touca actiCAP standard de 64 canais (EasyCap Shop, 2025). Os sinais foram amplificados através de um amplificador *BrainAmp* e digitalizados a uma taxa de amostragem de 5 kHz, garantindo resolução temporal adequada para a captura dos potenciais evocados de interesse (Brain Products, 2025). A montagem dos elétrodos seguiu o sistema internacional 10-20, assegurando a padronização e reprodutibilidade do posicionamento espacial (Klem *et al.*, 1999). Durante todo o procedimento de aquisição, as impedâncias foram

55

monitorizadas e mantidas abaixo de 15 k Ω , minimizando assim o ruído e otimizando a qualidade do sinal. O elétrodo FCz foi utilizado como referência online, enquanto o elétrodo localizado na posição fronto-polar central (AFz) serviu como terra.

A aquisição dos dados foi estruturada em dois blocos (*runs*) independentes, cada um compreendendo a apresentação de 40 imagens de faces e 40 imagens distorcidas, numa sequência aleatória, perfazendo um total de 160 estímulos apresentados ao longo da tarefa completa. Esta abordagem metodológica foi implementada com o objetivo de minimizar os efeitos da fadiga nos participantes.

Os dados eletrofisiológicos adquiridos foram sistematicamente organizados numa estrutura hierárquica de diretórios, com segregação inicial por grupo experimental através das pastas principais designadas "ASD" e "CTRL". Dentro de cada grupo, foi criado um diretório individual para cada participante, identificado através de um código único atribuído a cada sujeito. Esta estrutura contemplou ainda um nível adicional de organização temporal, com subdiretórios correspondentes às três sessões experimentais (v1, v2, v3), preservando assim a integridade cronológica das aquisições. Os ficheiros de dados foram armazenados em três formatos complementares gerados pelo sistema de aquisição *Brain Products*: ficheiros de dados brutos (.eeg), ficheiros de cabeçalho contendo metadados da aquisição (.vhdr) e ficheiros de marcadores temporais dos eventos experimentais (.vmrk). Esta arquitetura de armazenamento garantiu a rastreabilidade completa dos dados, facilitando o processamento subsequente e assegurando a manutenção da correspondência entre participantes, condições experimentais e sessões de avaliação ao longo do protocolo longitudinal.

É relevante mencionar que nem todos os participantes completaram o protocolo experimental integral das três visitas previstas. A descontinuação da participação deveu-se a múltiplos fatores, incluindo intercorrências clínicas, incompatibilidade de horários, ou dificuldades logísticas de deslocação. Estas ocorrências resultaram em conjuntos de dados incompletos para vários participantes, com ausência de registos eletrofisiológicos correspondentes a uma ou duas visitas experimentais.

3.2 Pré-Processamento dos Dados

No âmbito da metodologia de pré-processamento dos dados eletrofisiológicos, foi utilizado o software MATLAB 9.8.0 versão R2020a em conjugação com a *toolbox* EEGLAB v2025.0.0, permitindo a implementação de um fluxo de trabalho sistemático e replicável (Delorme and Makeig, 2004; The MathWorks Inc., 2020). Inicialmente, procedeu-se à importação dos ficheiros adquiridos durante as sessões experimentais para o ambiente MATLAB, utilizando o plugin *Brain Vision* v1.73 do EEGLAB para leitura dos dados no formato proprietário do sistema de aquisição. Após a abertura de cada ficheiro de dados, foi possível visualizar o registo contínuo da atividade elétrica cerebral de cada participante, permitindo uma primeira avaliação da qualidade do sinal e da presença de marcadores temporais correspondentes aos eventos experimentais.

Para assegurar a uniformidade e eficiência do processamento, foi desenvolvido um *script* (Anexos, Código 1) que permitiu a automatização do pré-processamento de todos os *datasets* contidos nas pastas designadas "ASD" e "CTRL", correspondentes aos grupos experimentais com PEA e controlo, respetivamente. Esta abordagem metodológica garantiu a aplicação consistente dos parâmetros de processamento a todos os ficheiros, minimizando potenciais variabilidades decorrentes do processamento manual. Na Figura 4 está apresentado um diagrama com a sequência de tarefas realizadas pelo *script* implementado.

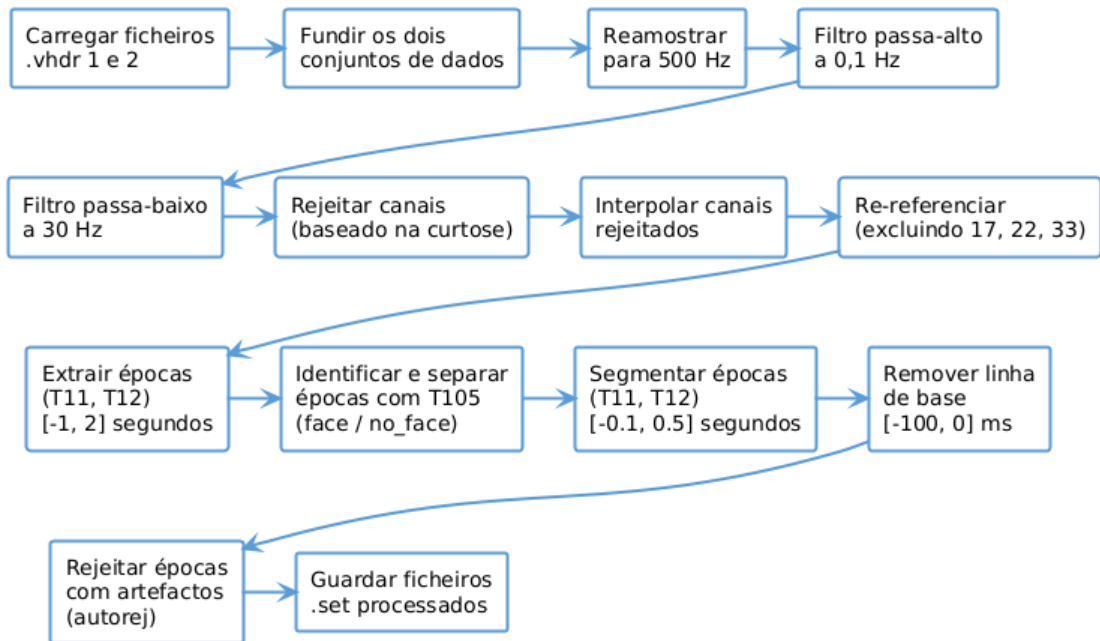


Figura 4 - Sequência de etapas de pré-processamento implementada no *script*.

A primeira etapa do pré-processamento consistiu, na concatenação dos dois *runs* de dados para cada participante, assegurando a continuidade da informação eletrofisiológica para análises subsequentes. Este procedimento foi implementado através de uma sequência estruturada de operações utilizando as funcionalidades do EEGLAB. Inicialmente, procedeu-se ao carregamento dos dois conjuntos de dados através da função *pop_loadbv*, atribuindo a cada um deles um nome específico para identificação. Posteriormente, ambos os conjuntos foram armazenados na estrutura ALLEEG através da função *eeg_store*, que permite a gestão de múltiplos *datasets* dentro do ambiente de trabalho. A junção efetiva dos dados foi realizada utilizando a função *pop_mergeset*, que concatena os dois conjuntos de dados sequencialmente, preservando a ordem temporal dos eventos e mantendo a estrutura dos canais.

A etapa de reamostragem foi implementada através da função *pop_resample* do EEGLAB, estabelecendo uma taxa de amostragem uniforme de 500 Hz para todos os *datasets*. Esta uniformização permitiu não só a otimização da resolução dos dados, como também assegurou a comparabilidade entre os diferentes registos.

Subsequentemente à reamostragem, implementou-se um protocolo de filtragem em duas etapas, utilizando a função *pop_eegfiltnew* do EEGLAB. Primeiramente, aplicou-se um filtro

passa-alto com frequência de corte de 0.1 Hz, visando a minimização de drifts de baixa frequência que poderiam comprometer a qualidade do sinal. De seguida, procedeu-se à aplicação de um filtro passa-baixo com frequência de corte de 30 Hz, permitindo a eliminação do ruído de alta frequência tipicamente associado à atividade muscular (Peng, 2019). Esta estratégia de filtragem, aplicada aos dados contínuos previamente à segmentação em épocas, está em conformidade com as recomendações metodológicas para evitar a introdução de artefactos nas extremidades dos segmentos, garantindo assim a preservação das componentes de frequência relevantes para a análise dos potenciais evocados de interesse.

A deteção de canais comprometidos foi realizada através da função *pop_rejchan* do EEGLAB, utilizando como critério estatístico a curtose (kurtosis), que permite quantificar o grau de afastamento de uma distribuição em relação à distribuição normal. A escolha deste parâmetro baseou-se na sua capacidade de identificar canais com comportamentos atípicos, particularmente aqueles que apresentam picos de atividade irregulares ou ruído impulsivo. A análise foi aplicada aos 64 canais que compunham o registo, estabelecendo-se um limiar de curtose de 5 desvios padrão, com normalização prévia do sinal ('norm', 'on') para garantir a comparabilidade entre diferentes canais e participantes. Os canais cujos valores de curtose excediam este limiar foram automaticamente classificados como maus e removidos do conjunto de dados, resultando numa estrutura temporariamente designada como '*Channel Rejected*'.

Subsequentemente, procedeu-se à interpolação dos canais removidos, visando a manutenção da dimensionalidade dos dados e a preservação da sua estrutura espacial. Este procedimento foi implementado através da função *pop_interp* do EEGLAB, utilizando o método de interpolação esférica. Esta técnica, baseada na representação tridimensional dos elétrodos numa esfera unitária, permite o estabelecimento de relações espaciais precisas entre os canais funcionais e os maus. A interpolação utiliza como referência a localização original dos elétrodos (*ALLEEG(1).chanlocs*) para estimar os valores dos canais removidos com base na atividade dos canais funcionais circundantes. O resultado deste processo foi armazenado numa estrutura designada '*Channel Interpolated*', mantendo assim a integridade topográfica do sinal, aspeto particularmente relevante para análises subsequentes que dependem da distribuição espacial da atividade elétrica.

O método de referenciação implementado neste estudo foi a referência de média comum CAR, aplicada através da função *pop_reref* do EEGLAB. A seleção desta abordagem de referenciação baseou-se na sua capacidade de minimizar o impacto de variabilidades não relacionadas com a atividade cerebral, otimizar a análise espacial dos dados e melhorar substancialmente a razão sinal-ruído. A implementação foi realizada especificando um vetor vazio como segundo parâmetro da função *pop_reref*, indicando que a média de todos os elétrodos deveria ser utilizada como referência. Procedeu-se também à exclusão específica dos elétrodos posicionados sobre os lóbulos oculares e das orelhas do cálculo da referência média. Esta exclusão justifica-se pelo facto de os elétrodos TP9 e TP10 (canais 17 e 22), localizados nos lóbulos auriculares esquerdo e direito, não registarem atividade cortical relevante, podendo introduzir ruído no cálculo da referência. Adicionalmente, o elétrodo AF7 (canal 33),

posicionado na região infraorbital esquerda, foi excluído por registar predominantemente artefactos oculares de elevada amplitude que, se incluídos no cálculo da referência média, contaminariam o sinal dos restantes elétrodos, comprometendo a integridade dos ERPs em estudo.

Posteriormente, procedeu-se à extração de épocas, um processo metodológico estruturado em várias fases sequenciais adaptadas ao desenho experimental específico deste estudo. Esta implementação teve como objetivo a segmentação precisa dos dados contínuos em função dos estímulos de interesse, seguida pela separação dos ensaios em duas categorias experimentais distintas.

Inicialmente, utilizou-se a função *pop_epoch* do EEGLAB para extrair segmentos temporais associados aos estímulos alvo, identificados pelos marcadores 'T 11' (aparição de uma face distorcida) e 'T 12' (aparição de uma face). Nesta primeira segmentação, definiu-se uma janela temporal abrangente, compreendendo o período de -1 a 2 segundos relativamente ao início do estímulo, o que permitiu obter uma representação ampla da atividade elétrica cerebral antes e após a apresentação dos estímulos. Esta abordagem foi implementada para facultar a posterior classificação dos ensaios em função da deteção ou não deteção por participante de estímulos faciais. Subsequentemente, realizou-se uma identificação sistemática dos ensaios que continham o marcador 'T105', indicativo da resposta “não vi face”. Este processo foi implementado através de um algoritmo de pesquisa que examinou os tipos de eventos presentes em cada época, armazenando os índices das épocas que continham o marcador específico. Esta classificação permitiu a criação de dois conjuntos de dados distintos: *EEG_no_face*, contendo exclusivamente as épocas associadas à ausência da perceção de estímulos faciais, e *EEG_face*, englobando as épocas remanescentes, correspondentes à perceção de estímulos faciais por parte dos participantes. Após esta categorização, procedeu-se a uma segunda extração de épocas para ambos os conjuntos de dados, utilizando uma janela temporal mais restrita e otimizada para a análise dos potenciais evocados de interesse. Nesta fase, definiu-se um intervalo de -0.1 a 0.5 segundos relativamente ao início do estímulo, proporcionando uma representação temporal precisa da atividade cerebral imediatamente antes e após a apresentação dos estímulos. Esta janela temporal foi selecionada especificamente para capturar os componentes N170 e P100, objetos centrais da análise subsequente.

Seguidamente, aplicou-se a correção da linha de base utilizando o intervalo pré-estímulo de -100 a 0 ms, através da função *pop_rmbase*. Esta normalização permitiu eliminar variações de potencial não relacionadas com a resposta ao estímulo, otimizando a comparabilidade entre ensaios e participantes. Por fim, implementou-se um procedimento de rejeição automática de épocas contaminadas por artefactos através da função *pop_autorej* do EEGLAB. Este algoritmo realiza uma análise estatística das distribuições de amplitude e aplica critérios adaptativos para identificar épocas que apresentam desvios significativos relativamente ao padrão típico observado na amostra. O método baseia-se na deteção de valores extremos de amplitude, aplicando limiares iterativamente ajustados em função das características estatísticas do conjunto de dados, o que permite uma identificação de artefactos mesmo em presença de

variabilidade interindividual. Esta abordagem automatizada oferece vantagens significativas comparativamente à inspeção visual, nomeadamente a aplicação consistente de critérios objetivos e a eliminação de potenciais enviesamentos introduzidos pelo operador, assegurando que apenas os segmentos com qualidade adequada fossem retidos para as análises subsequentes.

Após a conclusão das etapas de pré-processamento, procedeu-se à organização e armazenamento sistemático dos dados, seguindo uma estrutura hierárquica rigorosa que facilitasse a sua posterior recuperação e análise. Os dados foram organizados segundo uma estrutura em árvore, respeitando a hierarquia original da aquisição.

Dentro de cada diretório de sessão, os conjuntos de dados processados foram armazenados em formato *.set*, padrão do EEGLAB, que preserva não apenas os dados eletrofisiológicos, mas também as informações relevantes sobre os canais, eventos e parâmetros de processamento aplicados. A nomenclatura dos ficheiros seguiu um padrão sistemático, utilizando como base o nome comum extraído do ficheiro original, ao qual se acrescentou o sufixo "*_face*" ou "*_no_face*", dependendo da condição experimental a que os dados se referiam. Esta abordagem garantiu a rastreabilidade completa de cada ficheiro até à sua origem, facilitando a identificação inequívoca da condição experimental e do código do participante correspondente.

Como referido anteriormente, todo o pré-processamento e posterior organização e armazenamento de dados foi implementado através de um *script* automatizado, executado separadamente para as duas populações distintas: indivíduos com PEA, cujos dados foram armazenados na pasta "ASD", e participantes com desenvolvimento típico, armazenados na pasta "CTRL".

Por fim, implementou-se uma reorganização estrutural dos ficheiros para otimização das análises subsequentes. Esta reestruturação consistiu na criação de diretórios específicos, "*ASD_results*" e "*CTRL_results*", dedicados ao armazenamento sistemático dos conjuntos de dados processados para cada grupo experimental. Dentro de cada um destes diretórios principais, estabeleceu-se uma subdivisão adicional em duas categorias funcionais, "*face*" e "*no face*", permitindo a separação dos dados em função da natureza dos estímulos apresentados. Os ficheiros foram automaticamente classificados e copiados para os diretórios correspondentes com base nos sufixos identificativos presentes na sua nomenclatura ("*_face.set*" ou "*_no_face.set*"), mantendo-se os ficheiros originais nas suas localizações iniciais para garantir a integridade dos dados. Todo este processo foi executado, igualmente, através de um *script* codificado no MATLAB (Anexos, Código 2).

3.3 Identificação do N170 e P100

Para a identificação e quantificação dos componentes eletrofisiológicos de interesse, nomeadamente o P100 e o N170, implementou-se uma metodologia baseada na análise sistemática dos ERPs, derivados da média dos ensaios para cada participante e condição experimental. Esta análise foi realizada especificamente nos elétrodos 32 (PO10) e 28 (PO9),

posicionados nas regiões occipito-temporais direita e esquerda respetivamente, áreas onde os componentes associados ao processamento facial tipicamente apresentam maior proeminência (Tautvydaitė *et al.*, 2022).

A extração dos parâmetros do componente P100 foi efetuada através da delimitação de uma janela temporal específica, compreendida entre 80 e 200 milissegundos após a apresentação do estímulo. Dentro deste intervalo, procedeu-se à identificação do valor máximo de amplitude, correspondente ao pico positivo característico deste componente, registando-se tanto o seu valor em microvolts (μV) como a sua latência em milissegundos (ms). Esta abordagem permitiu capturar o componente positivo inicial associado ao processamento visual primário.

Para o componente N170, implementou-se uma metodologia análoga, utilizando uma janela temporal de 130 a 240 milissegundos pós-estímulo, intervalo que abrange tipicamente a manifestação deste componente associado ao processamento facial estrutural (Tautvydaitė *et al.*, 2022). Neste intervalo, identificou-se o valor mínimo de amplitude, correspondente ao pico negativo característico do N170, registando-se igualmente a sua amplitude e latência.

O processamento foi implementado através de um *script* automatizado (Anexos, Código 3) que percorreu sistematicamente todos os ficheiros de dados (.set) contidos em cada pasta de interesse, identificando automaticamente os dados do sujeito e da sessão a partir da nomenclatura dos ficheiros. Esta abordagem permitiu o processamento eficiente dos dados, minimizando potenciais erros associados à manipulação manual. Os parâmetros extraídos foram organizados numa estrutura de dados hierárquica, discriminando os valores por sujeito e por sessão experimental (v1, v2, v3), e posteriormente exportados para ficheiros do tipo CSV (*Comma-Separated Values*). Este procedimento foi executado de forma independente para cada uma das quatro pastas: "ASD_results - face", "ASD_results - no_face", "CTRL_results - face" e "CTRL_results - no_face", permitindo assim a comparação sistemática dos parâmetros entre grupos clínicos (PEA vs. controlos) e entre respostas ao estímulo (face vs. não face).

3.4 Organização e Análise dos Dados

De modo a otimizar a subsequente análise, foi desenvolvido um *script* em Python (Anexos, Código 4) que converteu os dados para o formato Excel (.xlsx), implementando simultaneamente transformações estruturais que maximizavam a sua usabilidade analítica. O *script* implementado foi desenvolvido utilizando as bibliotecas *Pandas* e *NumPy*. Inicialmente, procedeu-se à leitura dos ficheiros CSV, preservando a estrutura hierárquica dos cabeçalhos. Subsequentemente, implementou-se a substituição sistemática dos valores ausentes, representados originalmente por hífens ("-"), por valores NaN (*Not a Number*), facilitando o seu reconhecimento como dados omissos nas análises estatísticas posteriores. Paralelamente, realizou-se a conversão dos dados para formato numérico em todas as colunas exceto a identificativa do sujeito, assegurando a correta tipologia dos dados para operações matemáticas e estatísticas. Por fim, procedeu-se à estruturação adequada do índice multinível das colunas, garantindo a preservação da hierarquia informacional presente nos dados originais.

Esta abordagem resultou em quatro ficheiros Excel, facilitando significativamente a exploração, visualização e interpretação dos resultados. Todo este processo foi executado duas vezes (uma para cada elétrodo), resultando em 8 ficheiros Excel, que posteriormente foram organizados e incorporados apenas num só ficheiro.

Seguidamente, procedeu-se à reorganização sistemática dos dados, transitando de uma estrutura baseada em visitas cronológicas para uma arquitetura orientada pela dose farmacológica administrada. Esta reestruturação foi realizada manualmente, tanto nos dados presentes no ficheiro Excel como nos *datasets* resultantes do pré-processamento, com base nas informações de aleatorização previamente desmascaradas e consolidadas em ficheiro Word. Os *datasets* processados foram redistribuídos na seguinte hierarquia de diretórios: no nível superior, os dados foram segregados por dose (placebo, 15 mg Arbaclofeno, 30 mg Arbaclofeno), seguindo-se uma subdivisão por grupo experimental (ASD e CTRL) e, no nível mais específico, uma separação por resposta ao estímulo (face e no face). Esta arquitetura de armazenamento, ilustrada esquematicamente na Figura 5 através de um diagrama de estrutura de diretórios, permitiu o acesso sistemático aos dados segundo os fatores experimentais relevantes, facilitando as análises subsequentes e garantindo a rastreabilidade completa dos resultados em função das variáveis independentes do estudo.

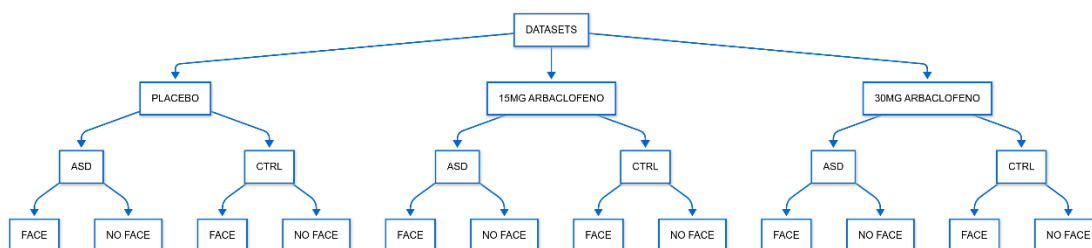


Figura 5 - Estrutura dos diretórios após organização manual dos dados por dose.

A partir dos dados consolidados no ficheiro Excel contendo as amplitudes e latências dos componentes P100 e N170 previamente extraídos, procedeu-se ao cálculo de métricas derivadas através de um *script* automatizado (Anexos, Código 5) desenvolvido em Python. Estas métricas incluem a amplitude pico-a-pico e a latência interpicos, parâmetros que fornecem informação complementar sobre a dinâmica temporal e a magnitude da resposta neuronal ao processamento de faces.

O algoritmo implementado realiza uma análise sistemática da estrutura de dados, identificando automaticamente as colunas correspondentes às amplitudes e latências dos componentes P100 e N170 através de reconhecimento de padrões nos cabeçalhos. Para cada condição experimental (hemisfério, dose farmacológica e tipo de estímulo), o *script* calcula a amplitude pico-a-pico através da diferença absoluta entre as amplitudes dos componentes P100 e N170. Paralelamente, a latência interpicos é calculada como a diferença temporal entre os picos N170 e P100. O processamento gera um novo ficheiro Excel contendo exclusivamente as colunas com os valores calculados de amplitude pico-a-pico e latência interpicos, mantendo a correspondência com os identificadores dos participantes. Estes dados foram posteriormente integrados no ficheiro principal de resultados através de cópia manual das colunas,

consolidando assim toda a informação necessária para as análises estatísticas subsequentes num único repositório de dados, facilitando a gestão e análise integrada de todas as variáveis dependentes do estudo.

Os dados de caracterização demográfica e clínica dos participantes foram, nesta fase, integrados no ficheiro Excel, consolidando assim toda a informação relevante numa base de dados unificada. Esta integração incluiu os valores do Quociente de Inteligência, as pontuações do Quociente do Espectro do Autismo, bem como os resultados dos instrumentos diagnósticos padronizados, nomeadamente a ADI-R e a ADOS.

Subsequentemente, procedeu-se à reestruturação dos dados para facilitar as análises estatísticas, convertendo o formato de organização dos dados de *wide* (formato largo) para *long* (formato longo) através de outro *script* desenvolvido em Python (Anexos, Código 6). O algoritmo implementado realiza uma análise sistemática da estrutura do ficheiro Excel original, identificando e categorizando automaticamente três tipos de colunas: medidas eletrofisiológicas regulares (amplitudes e latências dos componentes P100 e N170), métricas derivadas (amplitude pico-a-pico e latência interpicos) e variáveis demográficas. Para cada coluna de medição, o *script* extrai informação contextual através do reconhecimento de padrões nos cabeçalhos, identificando o hemisfério (direito/esquerdo), a dose farmacológica administrada (placebo, 15 mg ou 30 mg de arbaclofeno), a condição experimental (face/no face), o componente eletrofisiológico (P100/N170) e o tipo de medida (amplitude/latência). O processo de transformação preserva toda a informação demográfica dos participantes, incluindo o identificador único, o grupo experimental e outras variáveis relevantes, replicando estas informações para cada medição individual. Esta abordagem garante que cada linha do novo formato contém não apenas o valor da medição específica, mas também todo o contexto experimental e demográfico necessário para as análises subsequentes. O *script* gera um novo ficheiro Excel com os dados reorganizados no formato *long*, onde cada linha representa uma observação única caracterizada por múltiplas variáveis categóricas (participante, hemisfério, dose, condição, componente, tipo de medida) e o respetivo valor numérico. A manutenção dos dados em ambos os formatos, *wide* e *long*, permite flexibilidade analítica, possibilitando a seleção do formato mais adequado consoante os requisitos específicos de cada procedimento estatístico, análise, ou técnica de visualização a implementar.

Desenvolveu-se, adicionalmente, um *script* automatizado em MATLAB (Anexos, Código 7) para o cálculo e visualização dos ERPs para todas as condições experimentais. O algoritmo implementado percorre sistematicamente a estrutura hierárquica de diretórios previamente estabelecida, processando os *datasets* de forma iterativa através das diferentes doses farmacológicas, grupos experimentais e condições de estímulo. Para cada combinação de fatores, o *script* identifica e carrega automaticamente todos os ficheiros de dados (.set) disponíveis, extraíndo os sinais dos canais de interesse (elétrodos 32 e 28).

O processamento estatístico implementado calcula o ERP individual de cada participante através da média das épocas válidas, seguido do cálculo da média das médias (*grand average*) para cada condição experimental. A variabilidade interindividual é quantificada através do erro

padrão da média, permitindo a visualização da dispersão dos dados em torno da tendência central. As formas de onda resultantes são apresentadas graficamente com sombreamento representativo do erro padrão, facilitando a interpretação visual da significância estatística das diferenças observadas. O *script* desenvolvido gera três conjuntos distintos de visualizações comparativas, estrategicamente selecionados para evidenciar os efeitos principais e interações do paradigma experimental. O primeiro conjunto apresenta a comparação direta entre os grupos PEA e controlo em condições basais (placebo, estímulo face), permitindo identificar diferenças eletrofisiológicas intrínsecas entre as populações. O segundo conjunto contrasta as respostas face versus não-face dentro de cada grupo na condição placebo, isolando o efeito do processamento específico de faces sem a influência farmacológica, apenas de um hemisfério. Por fim, o terceiro conjunto examina o efeito dose-resposta do arbaclofeno (placebo, 15 mg, 30 mg de arbaclofeno) na condição de apresentação de faces, possibilitando a avaliação do impacto modulatório do fármaco sobre os componentes ERP relacionados com o processamento social. Esta abordagem sistemática de visualização facilita a identificação de padrões e tendências nos dados, mantendo a clareza interpretativa ao isolar variáveis específicas em cada comparação.

3.5 Métodos de Análise Estatística

A análise estatística dos dados foi conduzida através de duas abordagens complementares, implementadas com recurso ao software *IBM SPSS Statistics* (Versão 29) (IBM Corp., 2024). Inicialmente, procedeu-se à aplicação de modelos de Análise de Variância (ANOVA) para avaliar os efeitos principais e de interação entre os fatores experimentais. Subsequentemente, foram realizadas análises post hoc para identificar e caracterizar diferenças específicas entre condições experimentais quando os efeitos principais ou de interação se revelaram estatisticamente significativos, permitindo assim uma compreensão mais detalhada dos padrões observados nos dados. As duas análises são especificadas, em detalhe, nas secções seguintes.

3.5.1 Análise de Variância (ANOVA)

A ANOVA constitui uma técnica estatística fundamental para a investigação de efeitos experimentais em desenhos complexos, permitindo a partição sistemática da variabilidade total observada numa variável dependente em componentes atribuíveis a diferentes fontes de variação. Esta metodologia possibilita a decomposição da variância em efeitos aditivos correspondentes aos fatores experimentais principais, às suas interações e ao componente residual, distinguindo assim as variações sistemáticas explicadas pelos fatores manipulados das flutuações aleatórias não explicadas pelo modelo (Bertinetto, Engel and Jansen, 2020).

No contexto de desenhos experimentais multifatoriais, a ANOVA permite avaliar simultaneamente o impacto de múltiplas variáveis independentes e as suas interações sobre a variável de resposta. A ANOVA unifatorial examina o efeito de um único fator com múltiplos

níveis, enquanto a ANOVA multifatorial estende esta análise para dois ou mais fatores, testando não apenas os efeitos principais de cada variável independente, mas também os efeitos de interação entre elas. A ANOVA de medidas repetidas constitui uma extensão particularmente relevante desta metodologia, aplicável quando as mesmas unidades experimentais são avaliadas em múltiplas condições ou momentos temporais, controlando assim a variabilidade interindividual e aumentando o poder estatístico para detetar efeitos intra-sujeitos. Esta capacidade de detetar interações e controlar a variabilidade individual é particularmente relevante em investigações psicológicas e neurocientíficas, onde os fenómenos estudados frequentemente resultam da interação complexa entre múltiplos fatores e onde o desenho longitudinal ou de medidas repetidas permite capturar a dinâmica temporal dos processos cognitivos e neurais. (Cramer *et al.*, 2016).

Neste estudo, a análise foi implementada através do módulo de medidas repetidas (*Repeated Measures*) do SPSS, configurando-se um desenho misto com um fator intra-sujeitos (*within-subjects*) correspondente à dose farmacológica com três níveis (placebo, 15mg de arbaclofeno e 30mg de arbaclofeno) e um fator inter-sujeitos (*between-subjects*) representado pelo grupo diagnóstico (PEA ou controlo). Esta configuração permite avaliar simultaneamente o efeito da manipulação farmacológica dentro de cada participante, controlando a variabilidade individual, e as diferenças entre os grupos experimentais, assim como a sua interação.

Os parâmetros analíticos foram configurados para incluir estatísticas descritivas completas, estimativas do tamanho do efeito (eta quadrado parcial), poder observado e testes de homogeneidade de variâncias, garantindo uma avaliação abrangente dos pressupostos estatísticos e da magnitude dos efeitos observados. As médias marginais estimadas foram calculadas para os efeitos principais da dose e grupo, bem como para a interação dose × grupo, com comparação dos efeitos principais e ajustamento dos intervalos de confiança através do método *Least Significant Difference* (LSD).

O procedimento analítico foi aplicado sistematicamente a cada variável dependente em estudo, abrangendo as amplitudes e latências dos componentes P100 e N170, bem como as métricas derivadas de amplitude pico-a-pico e latência interpicos. As análises foram conduzidas separadamente para cada condição de estímulo (face e não face) e cada hemisfério cerebral (direito e esquerdo). Foi realizada adicionalmente uma ANOVA de medidas repetidas com dois fatores intra-sujeitos, correspondentes à dose e ao hemisfério, resultando num total de 25 modelos estatísticos independentes. O nível de significância foi estabelecido em $\alpha = 0,05$ para todos os testes estatísticos. Dado o carácter exploratório deste estudo e o elevado número de comparações realizadas, optou-se por não aplicar correções para comparações múltiplas, reconhecendo-se o conseqüente aumento do risco de erro tipo I. Esta decisão metodológica privilegiou a sensibilidade para detetar potenciais efeitos do arbaclofeno, sendo os resultados interpretados como preliminares e requerendo replicação em estudos confirmatórios futuros.

No que respeita ao tratamento dos dados em falta, o SPSS utilizou a abordagem de exclusão por lista, por defeito, nas análises de medidas repetidas. Esta metodologia exclui automaticamente da análise qualquer participante que apresente dados em falta em pelo

menos uma das condições experimentais, garantindo que apenas participantes com dados completos nas três visitas sejam incluídos em cada modelo ANOVA. Embora esta abordagem resulte numa redução do tamanho da amostra efetivo e potencial perda de poder estatístico, assegura a comparabilidade das condições experimentais. Consequentemente, o número de participantes incluídos em cada análise variou em função da disponibilidade de dados completos para a variável dependente específica em análise.

3.5.2 Análise Post Hoc

Neste estudo, as análises post hoc foram conduzidas para examinar as diferenças específicas entre doses farmacológicas nas variáveis que demonstraram efeitos significativos na ANOVA de medidas repetidas. Para este efeito, procedeu-se inicialmente à segmentação do ficheiro de dados por grupo diagnóstico através da função *Split File* do SPSS, permitindo a análise independente dos efeitos farmacológicos dentro de cada grupo experimental. Subsequentemente, foram realizados testes t para amostras emparelhadas (*paired-samples t-test*) para avaliar as diferenças entre pares de doses específicas. Este procedimento envolveu três comparações sistemáticas: placebo versus 15 mg de arbaclofeno, placebo versus 30 mg de arbaclofeno, e 15 mg versus 30 mg de arbaclofeno. A utilização de testes para amostras emparelhadas é apropriada neste contexto, dado o desenho intra-sujeitos do estudo, onde cada participante foi avaliado nas três condições farmacológicas. A configuração dos testes incluiu o cálculo de tamanhos de efeito padronizados, utilizando o desvio-padrão da diferença como denominador, permitindo assim a quantificação da magnitude das diferenças observadas independentemente da escala de medida. Esta abordagem analítica estratificada por grupo permitiu identificar padrões diferenciais de resposta farmacológica entre participantes com PEA e controlos, fornecendo informações sobre a especificidade dos efeitos do arbaclofeno em função do perfil diagnóstico dos participantes.

3.6 Métodos de Análise com *Machine Learning*

A aplicação de técnicas de *machine learning* foi implementada como abordagem complementar à análise estatística tradicional, com o objetivo de explorar padrões subjacentes nos dados eletrofisiológicos e avaliar a capacidade discriminativa dos componentes P100 e N170 em diferentes contextos experimentais. Esta metodologia permitiu investigar se as características extraídas dos potenciais evocados apresentavam estruturas de agrupamento natural que pudessem não ser totalmente capturadas pelas análises estatísticas convencionais. Para este efeito, foram implementados três algoritmos independentes de classificação, cada um direcionado para uma questão experimental específica. O primeiro modelo foi treinado para discriminar entre participantes do grupo de controlo e participantes com PEA, utilizando as características eletrofisiológicas como preditores. O segundo modelo focou-se na distinção entre as três condições farmacológicas (placebo, 15 mg e 30 mg de arbaclofeno). Por fim, o terceiro modelo foi desenvolvido para classificar os dados eletrofisiológicos com base na resposta comportamental dos participantes, distinguindo entre condições em que foram

identificadas faces versus condições sem identificação facial. Na Tabela 2 está presente um resumo dos dados de entrada, saída e os modelos de classificação usados em cada um dos algoritmos de *machine learning* desenvolvido.

Tabela 2 – Resumo dos algoritmos de *machine learning* desenvolvidos.

Algoritmo	Dados de Entrada/Saída	Modelos de Classificação
Classificação PEA vs. Controlo	Entrada: Amplitudes e latências dos componentes P100 e N170 e métricas associadas (placebo, face). Saída: Grupo (0=Controlo, 1=PEA)	<i>Regressão Logística</i>
Predição de Dose	Entrada: Amplitudes e latências dos componentes P100 e N170 e métricas associadas (todas as doses, face). Saída: Dose (0=Placebo, 1=15 mg de arbaclofeno, 2=30 mg de arbaclofeno)	<i>Support Vector Machine</i> <i>Random Forest</i> <i>K-Nearest Neighbors</i>
Classificação Face vs. Não-Face	Entrada: Amplitudes do componente N170 (todas as doses, ambos os hemisférios). Saída: Condição (0=Não-face, 1=Face)	

O primeiro algoritmo de classificação (Anexos, Código 8) foi, então, desenvolvido para discriminar entre participantes do grupo de controlo e participantes com PEA, utilizando exclusivamente as características eletrofisiológicas extraídas dos componentes P100 e N170 na condição placebo e apenas para os ensaios em que os participantes identificaram faces. Esta decisão metodológica fundamentou-se em duas considerações principais: primeiro, a utilização dos dados da condição placebo representa o estado eletrofisiológico basal dos participantes sem interferência farmacológica; segundo, permite a comparabilidade com estudos prévios na literatura que tipicamente avaliam diferenças basais entre populações. A restrição aos ensaios com identificação de faces assegura que os componentes analisados retratam efetivamente o processamento neuronal específico de estímulos faciais, maximizando assim a relevância dos marcadores eletrofisiológicos para a distinção entre grupos. O ficheiro Excel preparado para este algoritmo continha os dados organizados em formato *wide*, incluindo exclusivamente as colunas referentes às medidas de amplitude e latência dos componentes P100 e N170 obtidas na condição placebo com resposta face, em ambos os hemisférios, juntamente com uma coluna identificadora do grupo (0 = controlo, 1 = autismo). Adicionalmente, foi criado um ficheiro Excel apenas com as colunas relativas à latência do componente N170 em ambos os hemisférios, nas mesmas condições, para realizar um teste adicional com os dados que reportaram maior diferença entre grupos na literatura.

O pré-processamento dos dados iniciou-se com a identificação e remoção de participantes que apresentavam mais de 50% de valores em falta, assegurando assim a qualidade dos dados utilizados para treino dos modelos. Para os valores em falta restantes, foi aplicado o algoritmo *K-Nearest Neighbors (KNN) Imputer* com $k=5$, que estima os valores ausentes com base na similaridade com os vizinhos mais próximos no espaço de características. Subsequentemente, todos os dados foram normalizados através do método *StandardScaler*, garantindo que todas

as variáveis apresentassem média zero e desvio padrão unitário, prevenindo assim que características com maiores escalas dominassem o processo de aprendizagem.

A estratégia de validação adotada incluiu a divisão dos dados em conjuntos de treino (70%) e teste (30%), mantendo a estratificação para preservar a proporção original entre grupos em ambos os conjuntos. Foram implementados quatro algoritmos de classificação distintos: *Regressão Logística*, *Support Vector Machine (SVM)* com *kernel* RBF, *Random Forest* com 100 árvores de decisão, e *K-Nearest Neighbors (KNN)* com $k=5$ (Erkan and Thanh, 2020; Sharma and Tanwar, 2020; Panda, 2022). Esta diversidade de abordagens permitiu avaliar qual metodologia seria mais adequada para capturar os padrões discriminativos nos dados eletrofisiológicos. A avaliação do desempenho foi realizada através de múltiplas métricas, incluindo exatidão (*accuracy*), precisão, sensibilidade (*recall*) e *F1-score*, complementadas por matrizes de confusão para análise detalhada dos padrões de classificação. Adicionalmente, foi implementada validação cruzada estratificada com 5 partições (*5-fold*) para obter estimativas do desempenho de generalização de cada modelo. Para complementar a análise de classificação, o algoritmo RF foi utilizado para extrair a importância relativa de cada característica, permitindo identificar quais componentes eletrofisiológicos contribuíam mais significativamente para a discriminação entre grupos. Por fim, foi realizada uma Análise de Componentes Principais (PCA) para visualização bidimensional da distribuição dos dados e avaliação da separabilidade natural entre os grupos no espaço reduzido de características. Dado o design de classificação multiclasse com duas categorias equiprováveis, o nível de acaso (*chance level*) para este algoritmo é de 50%, representando o desempenho esperado de um classificador aleatório.

O segundo algoritmo de classificação (Anexos, Código 9) foi desenvolvido para discriminar entre as três condições farmacológicas testadas (placebo, 15 mg e 30 mg de arbaclofeno), independentemente do diagnóstico, utilizando as características eletrofisiológicas como preditores. Contrariamente ao algoritmo anterior, este modelo trabalhou com dados em formato *long*, requerendo uma etapa adicional de transformação para formato *wide* antes do processamento. O ficheiro Excel de entrada continha os dados organizados com cada linha representando uma medida individual (amplitude ou latência) de um componente específico (P100 ou N170) num determinado hemisfério, incluindo informação sobre o sujeito, grupo, sexo, dose administrada e condição experimental (face ou não-face). Para assegurar a especificidade da análise ao processamento de faces, os dados foram filtrados para incluir exclusivamente os ensaios em que os participantes identificaram corretamente estímulos faciais. A transformação dos dados de formato *long* para *wide* criou uma matriz onde cada participante em cada condição de dose era representado por uma linha única, com as colunas correspondendo às diferentes medidas eletrofisiológicas (amplitude e latência de P100 e N170 em ambos os hemisférios). A variável alvo foi codificada numericamente, atribuindo 0 ao placebo, 1 à dose de 15 mg e 2 à dose de 30 mg de arbaclofeno, permitindo assim a implementação de algoritmos de classificação multiclasse. Esta abordagem possibilitou investigar se os padrões eletrofisiológicos continham informação suficiente para distinguir não apenas entre presença ou ausência do fármaco, mas também entre diferentes doses.

O pré-processamento seguiu a mesma metodologia do algoritmo anterior, incluindo remoção de observações com elevada percentagem de valores em falta, imputação através de KNN e normalização dos dados. A informação sobre o grupo diagnóstico (controlo vs PEA) foi codificada e incluída como característica adicional, permitindo ao modelo considerar potenciais interações entre o diagnóstico e a resposta farmacológica. A implementação incluiu os mesmos quatro algoritmos de classificação, adaptados para o contexto multiclasse. Para a regressão logística, foi utilizada a abordagem multinominal, enquanto o SVM implementou a estratégia *one-vs-rest* para lidar com as três classes. A análise de importância das características, extraída do modelo *Random Forest*, foi expandida para incluir agregações por componente (P100 vs N170) e por hemisfério (esquerdo vs direito), de modo a fornecer informações sobre quais aspetos do processamento neuronal eram mais informativos para a discriminação entre doses. A análise PCA foi adaptada para visualizar a separabilidade entre as três condições farmacológicas, incluindo cálculos de distâncias entre centroides e análises de dispersão para quantificar o grau de sobreposição entre grupos no espaço bidimensional reduzido. Dado o design de classificação multiclasse com três categorias equiprováveis, o *chance level* para este algoritmo é de 33,33%.

O terceiro algoritmo de classificação (Anexos, Código 10) foi desenvolvido para discriminar entre respostas comportamentais, distinguindo dados eletrofisiológicos em que os participantes identificaram faces daqueles sem identificação facial, focando-se especificamente nas características da amplitude do componente N170. Esta abordagem metodológica diferenciou-se dos algoritmos anteriores ao examinar a capacidade discriminativa de um único componente eletrofisiológico, reconhecido na literatura como particularmente sensível ao processamento facial. O modelo utilizou o mesmo ficheiro Excel em formato long do algoritmo anterior, mas implementou critérios de filtragem distintos, selecionando exclusivamente dados referentes à amplitude do N170 em ambos os hemisférios, independentemente do grupo diagnóstico. A transformação dos dados manteve a estrutura por dose, criando características que representavam a amplitude do N170 em cada hemisfério para cada condição farmacológica (placebo, 15 mg e 30 mg). Esta organização permitiu averiguar se os padrões de ativação neuronal associados ao reconhecimento facial eram suficientemente distintos para possibilitar classificação automática, e se esta distinção era modulada pela dose de arbaclofeno. A variável alvo foi codificada binariamente, atribuindo 1 aos ensaios com identificação facial e 0 aos ensaios sem identificação. A análise de importância das características foi expandida para incluir agregações por dose farmacológica, de modo a observar como diferentes níveis de modulação GABAérgica influenciavam a discriminação entre condições.

A análise PCA resultou em duas figuras distintas, uma focada na separação entre condições comportamentais e outra estratificada por grupo diagnóstico, permitindo examinar se os padrões de separação eram consistentes entre populações. Adicionalmente, foram implementadas análises post hoc que incluíram a identificação de *outliers* no espaço PCA e a avaliação da performance de classificação estratificada por grupo diagnóstico. Esta última análise, permitiu verificar se a capacidade discriminativa do modelo era equivalente entre participantes com PEA e controlos. A análise dos padrões no espaço PCA incluiu também a quantificação de observações extremas, de modo a observar se estas se distribuíam

uniformemente entre condições e grupos ou se representavam subpopulações específicas com padrões atípicos de processamento facial. Para este algoritmo, devido ao design com duas classes equiprováveis, o *chance level* é também de 50%.

4 Resultados

4.1 Pré-processamento dos Dados e Identificação dos picos

Os ficheiros EEG adquiridos foram importados com sucesso para o ambiente computacional do EEGLAB, mantendo a integridade dos registos e a sincronização temporal dos marcadores de eventos. A visualização inicial dos dados permitiu confirmar a adequada aquisição dos sinais em todos os canais e a correta codificação dos eventos experimentais ao longo das sessões. A Figura 6 ilustra um exemplo representativo desta fase inicial de análise, apresentando um segmento de 5 segundos do registo eletroencefalográfico contínuo do participante F101 na sua primeira sessão (grupo PEA).

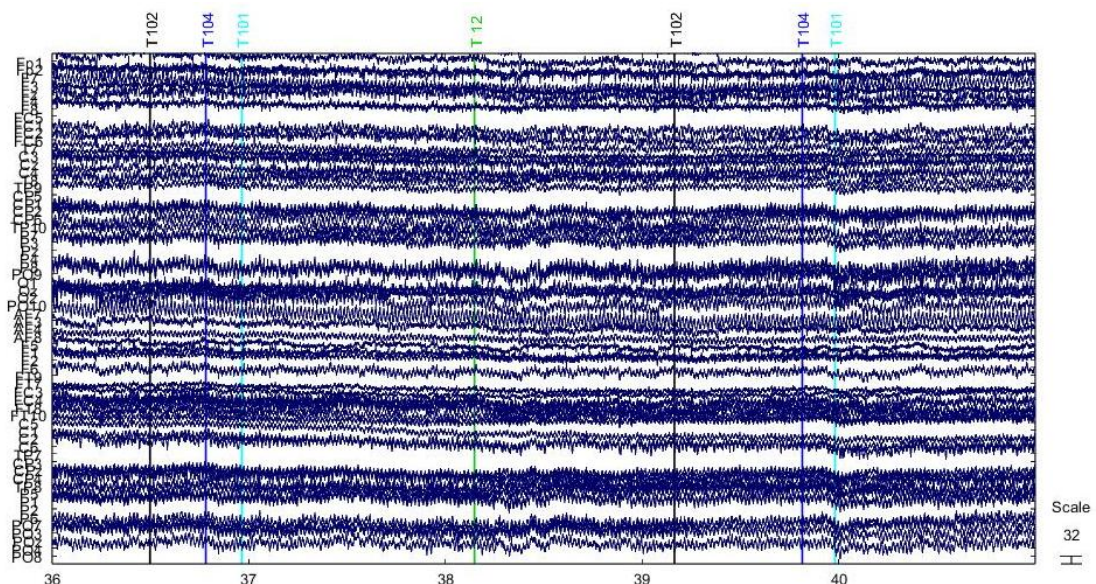


Figura 6 - Segmento de 5 segundos do EEG do sujeito F101.

O registo apresenta os diferentes canais de EEG dispostos verticalmente segundo a montagem do sistema internacional 10-20, com os marcadores de eventos sobrepostos temporalmente. Neste segmento específico são visíveis quatro marcadores: um marcador de *baseline* (T 101), dois marcadores de pergunta (T 102), um marcador de resposta afirmativa (T 104) indicando que o participante identificou ter visto uma face, e um marcador de apresentação de estímulo facial (T 12), que fazem parte do protocolo de apresentação de estímulos. A atividade eletroencefalográfica apresenta as características típicas de um registo em estado de vigília, com oscilações de amplitude variável entre canais (escala de 32 μ V), sendo possível observar tanto a atividade neuronal espontânea como as respostas aos estímulos apresentados.

De seguida, procedeu-se à segmentação temporal dos registos contínuos em épocas centradas nos eventos de interesse, seguida da correção de *baseline*. Este processamento inicial permitiu a visualização dos ERPs para cada canal, sem o pré-processamento efetuado. A

Figura 7 apresenta os ERPs obtidos nos canais PO9 e PO10 (hemisfério esquerdo e direito, respectivamente), para o mesmo participante F101 após a extração de épocas e correção de baseline.

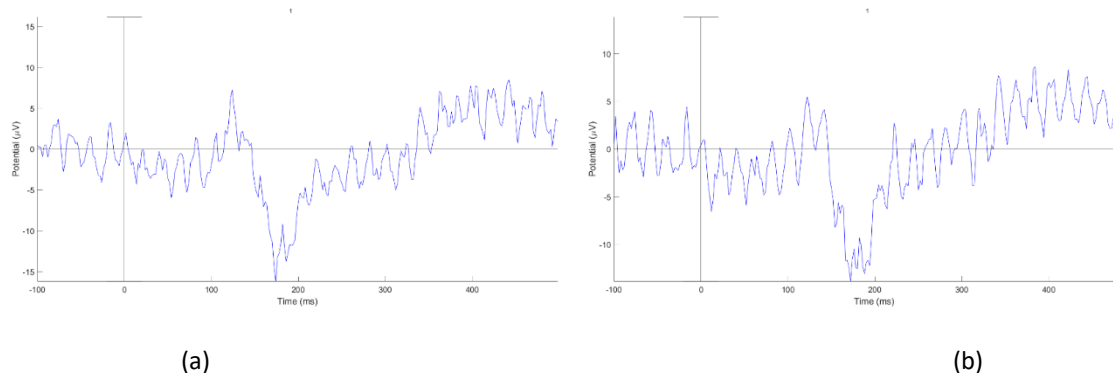


Figura 7 - ERPs do sujeito F101 após extração de épocas e correção de *baseline* dos canais PO9 (a) e PO10 (b).

A análise visual dos traçados revela a presença de uma deflexão negativa proeminente em ambos os canais, com latência aproximada de 170 ms após a apresentação do estímulo, consistente com o componente N170 descrito na literatura. No entanto, é evidente a presença de considerável variabilidade no sinal, com flutuações de amplitude que se estendem ao longo de toda a janela temporal analisada. Esta variabilidade, sugere a necessidade de procedimentos adicionais de pré-processamento para otimização da relação sinal-ruído.

Após a implementação do pipeline completo de pré-processamento através do *script* automatizado desenvolvido, obteve-se uma melhoria substancial na qualidade dos potenciais evocados. A Figura 8 apresenta os ERPs resultantes para os mesmos canais PO9 e PO10 do participante F101, quando este respondeu 'face', após a aplicação de todos os procedimentos de limpeza do sinal.

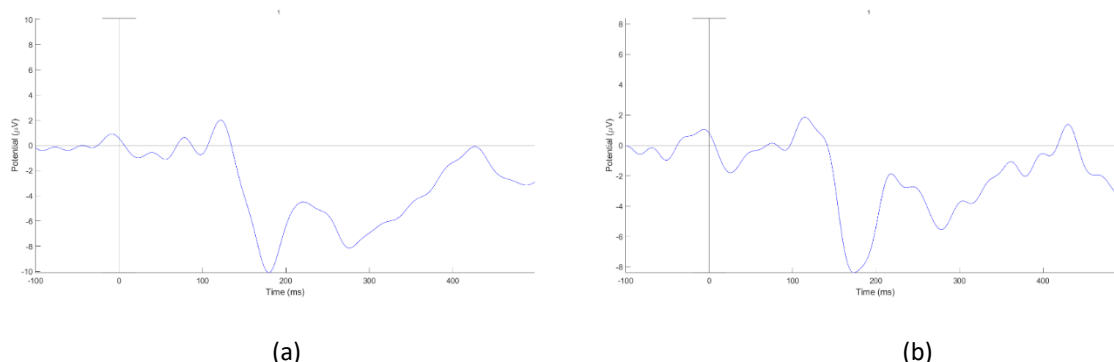


Figura 8 - ERPs do sujeito F101 após pré-processamento dos canais PO9 (a) e PO10 (b).

A comparação entre os traçados pré e pós-processamento demonstra o impacto significativo das técnicas de otimização do sinal implementadas. Os ERPs processados apresentam uma morfologia consideravelmente mais definida, com redução das oscilações do ruído de fundo que anteriormente dificultavam a visualização dos componentes de interesse. A redução destas oscilações sugere uma remoção eficaz de artefactos transitórios e uma melhoria substancial da relação sinal-ruído.

De notar que, conforme mencionado anteriormente, o algoritmo procedeu à separação dos ensaios com base nas respostas comportamentais dos participantes, classificando as épocas em função da identificação subjetiva de faces versus não-faces. É importante salientar que, embora o paradigma experimental tenha apresentado um número equilibrado de estímulos (80 imagens de faces e 80 imagens de não-faces), a distribuição final das épocas nos conjuntos de dados "face" e "não-face" variou entre participantes devido a erros de identificação. Esta variabilidade comportamental resultou em conjuntos de dados assimétricos, nos quais alguns participantes apresentaram um número superior de épocas classificadas como "face" relativamente às classificadas como "não-face", ou vice-versa. De notar que foram removidos em média, e posteriormente interpolados, 3 a 5 canais por *dataset*. Também é importante salientar que, após a aplicação dos critérios de rejeição de artefactos, foram excluídas entre 4 e 8 épocas por *dataset*, independentemente da condição, resultando em *datasets* com 70 a 80 épocas.

A otimização do sinal eletrofisiológico possibilitou uma identificação mais precisa dos componentes P100 e N170. Na Figura 9 é possível observar os picos característicos dos componentes P100 e N170, identificados com pontos coloridos.

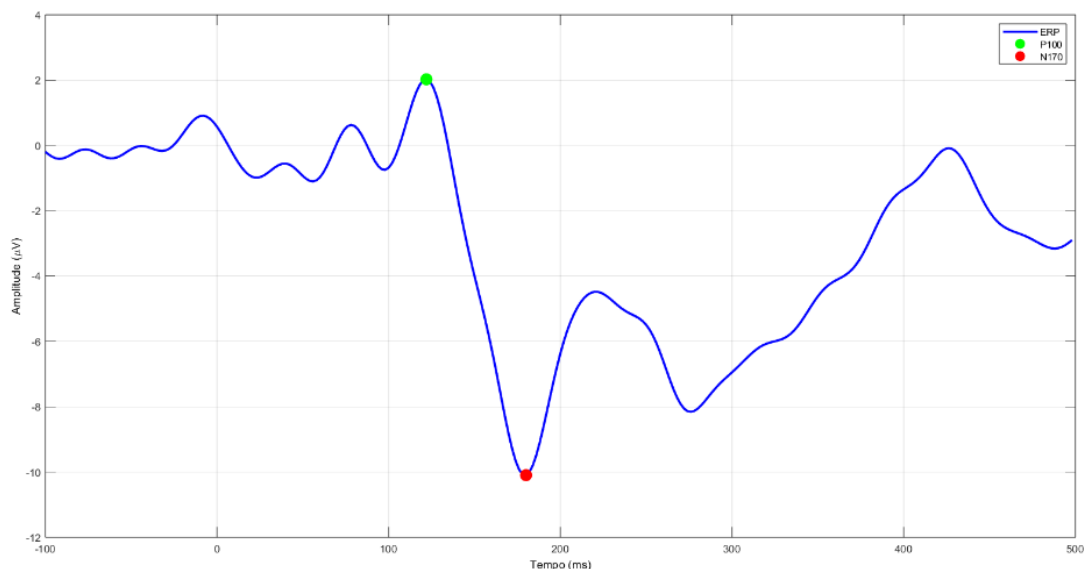


Figura 9 - Identificação dos componentes P100 e N170.

A implementação do algoritmo automatizado de detecção de picos demonstrou elevada eficácia na identificação e quantificação dos componentes eletrofisiológicos de interesse. A figura ilustra um exemplo representativo do desempenho do *script*, onde os marcadores identificados automaticamente (ponto verde para o P100 e ponto vermelho para o N170) correspondem com precisão aos picos máximo e mínimo dos respetivos componentes no traçado do ERP. O algoritmo permitiu extrair de forma sistemática os valores de amplitude e latência para a grande maioria dos potenciais evocados processados, assegurando consistência metodológica e reduzindo substancialmente o tempo de análise comparativamente a métodos manuais. Contudo, a heterogeneidade morfológica dos ERPs entre participantes, resultou em situações pontuais onde a detecção automática não capturou adequadamente os componentes de interesse. Nestas circunstâncias, que ocorreram em ambos os canais analisados (PO9 e PO10), foi necessária a intervenção manual para corrigir os valores de amplitude e latência, garantindo assim a validade das medições.

Após a extração bem-sucedida dos parâmetros eletrofisiológicos, procedeu-se à exportação e organização dos dados num formato estruturado para análise estatística. Previamente à finalização desta base de dados, realizou-se uma inspeção visual sistemática de todos ERPs resultantes do processamento, durante a qual se identificaram traçados que não apresentavam a morfologia característica esperada dos componentes P100 e N170, potencialmente devido a artefactos residuais ou baixa qualidade do sinal. Para assegurar a integridade das análises estatísticas e evitar enviesamento dos resultados, foram excluídos os *datasets* que apresentavam estas anomalias morfológicas, abrangendo registos de ambos os hemisférios e todas as condições experimentais. Esta triagem resultou na exclusão de 13 visitas de participantes com PEA e 14 do grupo de controlo, perfazendo um total de 27 conjuntos de dados removidos. Estes *datasets* excluídos não foram considerados em nenhuma das análises subsequentes, garantindo assim que apenas dados de qualidade adequada contribuíssem para os resultados do estudo.

4.2 Análise dos ERPs Médios por Condição

A implementação do *script* de análise dos ERPs médios permitiu a geração sistemática de 8 figuras, correspondentes às diferentes combinações de grupos experimentais (controlo vs. PEA), condições farmacológicas (placebo, 15mg e 30mg de arbaclofeno), tipos de resposta comportamental (identificação de face vs. não face) e hemisférios cerebrais (esquerdo vs. direito). Esta visualização dos dados eletrofisiológicos possibilitou uma análise comparativa dos padrões de ativação neuronal em resposta aos estímulos faciais. A Figura 10 apresenta os potenciais evocados médios do grupo de controlo e PEA na toma de placebo, especificamente para os ensaios em que os participantes identificaram a presença de faces.

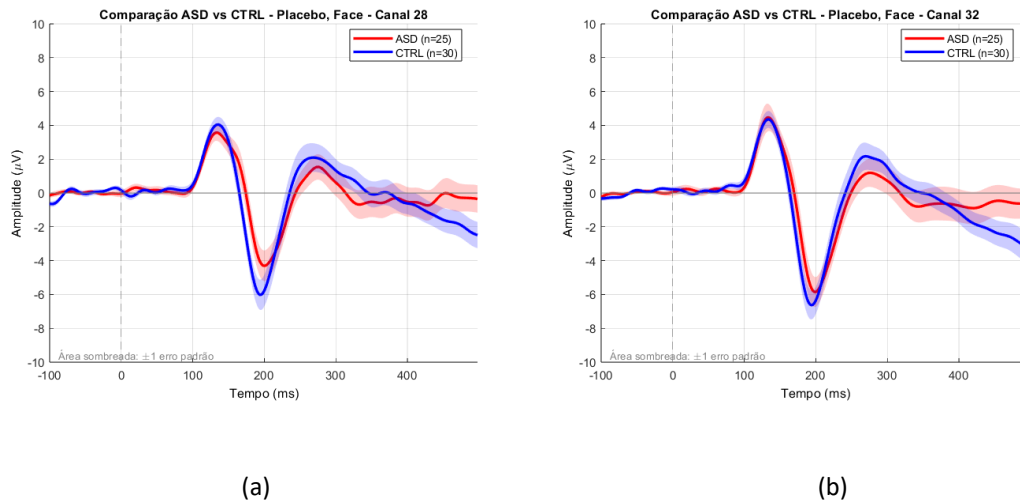


Figura 10 - ERP médio dos sujeitos do grupo de controlo (azul) e autismo (vermelho) na condição de placebo, quando identificaram faces, do hemisfério esquerdo (a) e direito (b). Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo) e canal PO10 (canal 32, hemisfério direito), com as respetivas áreas sombreadas indicando o erro padrão da média.

A análise dos traçados revela padrões eletrofisiológicos característicos do processamento facial em ambos os hemisférios, com clara identificação dos componentes P100 e N170. O componente P100 manifesta-se como uma deflexão positiva com pico aproximadamente aos 150 ms, apresentando amplitudes comparáveis entre hemisférios e sem diferenças observáveis entre os grupos diagnósticos. O N170 observa-se na proeminente deflexão negativa, atingindo o seu valor máximo entre os 170-200 ms, demonstrando morfologia típica do processamento de estímulos faciais. São, também, notórias uma ligeira latência prolongada e uma menor amplitude do componente N170 no grupo PEA comparativamente ao grupo de controlo. Estas diferenças nas médias, embora subtis, são consistentes com alguns dos resultados reportados na literatura que documentam um processamento facial mais lento e menos acentuado em indivíduos com PEA (Neath-Tavares and Itier, 2016; Márquez *et al.*, 2019; Kala *et al.*, 2021). Destaca-se ainda uma assimetria hemisférica pronunciada no grupo PEA, com o hemisfério direito a demonstrar uma amplitude N170 visivelmente superior à do hemisfério esquerdo. Esta diferença inter-hemisférica contrasta com as evidências da literatura, que geralmente reportam uma lateralização atenuada ou mais difusa do processamento facial em indivíduos com PEA (Kang *et al.*, 2018). Na Figura 11 encontram-se apresentados os potenciais evocados médios do grupo de controlo e PEA na toma de placebo, comparando os ensaios que os participantes identificaram a presença de faces com os que não identificaram.

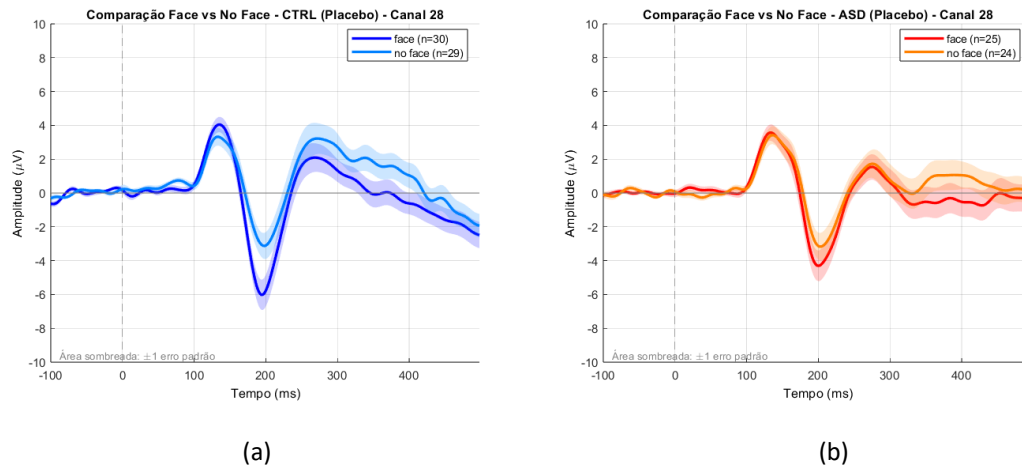


Figura 11 - ERP médio dos sujeitos do grupo de controlo (a) e autismo (b) na condição de placebo, quando identificaram ou não faces. Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo), com os ensaios em que o sujeito identificou face a azul-escuro e vermelho, e os ensaios que não identificou face a azul-claro e laranja, com as respetivas áreas sombreadas indicando o erro padrão da média.

A comparação entre identificação ou não de face revela diferenças eletrofisiológicas marcantes em ambos os componentes analisados. O componente N170 demonstra uma modulação substancial pela perceção consciente de faces, com uma redução da sua amplitude quando os participantes não identificam faces nos estímulos apresentados. Esta atenuação da amplitude do N170 na ausência de perceção consciente de faces é consistente com a literatura que estabelece este componente como um marcador eletrofisiológico específico do processamento facial (Shu *et al.*, 2024). Esta diferenciação entre condições é evidente em ambos os grupos diagnósticos, manifestando-se de forma mais pronunciada no grupo de controlo. No grupo PEA, embora se mantenha o padrão de modulação, a diferença entre condições apresenta-se visivelmente atenuada, sugerindo uma resposta neuronal menos diferenciada ao processamento de estímulos faciais. O componente P100 também evidencia sensibilidade à categoria do estímulo percebido, apresentando uma redução de amplitude na condição não-face, particularmente visível no grupo de controlo. Na Figura 12 encontram-se apresentados os potenciais evocados médios do grupo de controlo, nos ensaios em que os participantes identificaram faces, comparando as doses administradas e os hemisférios.

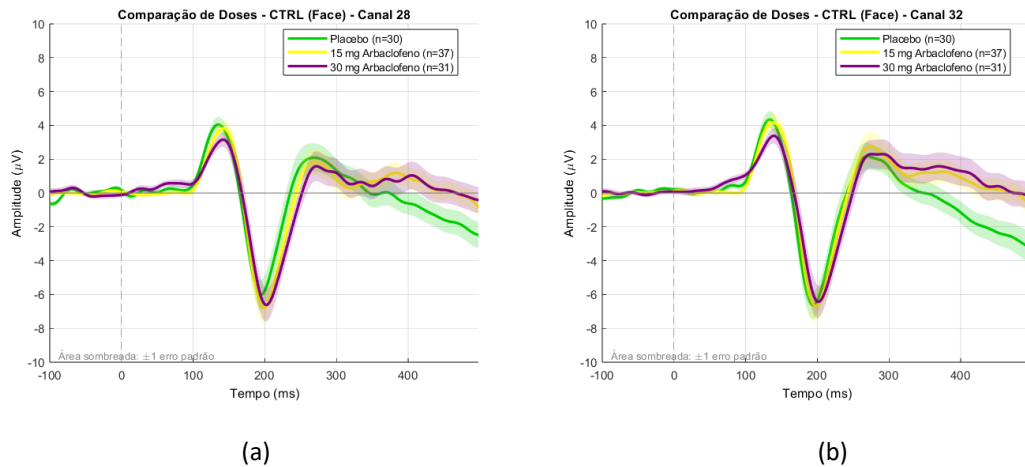


Figura 12 - ERP médio dos sujeitos do grupo de controlo na toma do placebo (traçado verde), 15 mg de arbaclofeno (traçado amarelo) e 30 mg de arbaclofeno (traçado roxo), quando identificaram faces. Os traçados representam a atividade registada no canal 28 (PO9, hemisfério esquerdo, (a)) e no canal 32 (PO10, hemisfério direito, (b)), com as respetivas áreas sombreadas indicando o erro padrão da média.

A análise comparativa dos efeitos das diferentes doses de arbaclofeno (placebo, 15 mg e 30 mg de arbaclofeno) no grupo de controlo revela uma notável sobreposição dos traçados eletrofisiológicos, sugerindo que a manipulação farmacológica não induz alterações substanciais nos padrões médios dos potenciais evocados. Os componentes P100 e N170 mantêm morfologia e temporalidade consistentes através das três condições experimentais, sem evidência de modulação dose-dependente sistemática. Contudo, observam-se diferenças subtis na lateralização dos efeitos farmacológicos. No hemisfério esquerdo, observa-se uma maior variabilidade entre condições, particularmente na janela temporal pós-N170, onde os traçados apresentam trajetórias ligeiramente divergentes entre doses. Em contraste, o hemisfério direito demonstra uma sobreposição quase completa dos traçados nas três condições farmacológicas, com mínima variação nas amplitudes e latências dos componentes principais. Esta assimetria na sensibilidade hemisférica ao arbaclofeno sugere que eventuais efeitos farmacológicos possam manifestar-se de forma lateralizada. Na Figura 13, encontram-se apresentados os potenciais evocados médios do grupo de PEA, nos ensaios em que os participantes identificaram faces, comparando as doses administradas e os hemisférios.

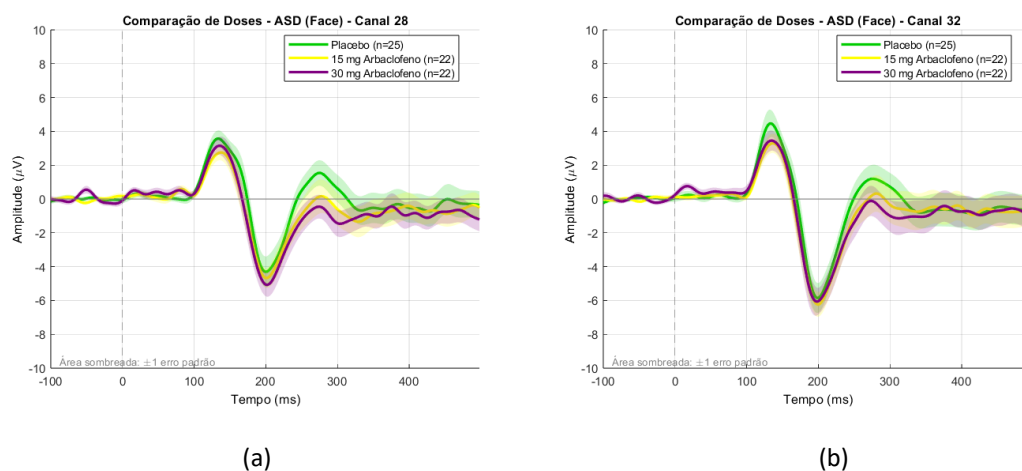


Figura 13 - ERP médio dos sujeitos do grupo de PEA na toma do placebo (traçado verde), 15 mg de arbaclofeno (traçado amarelo) e 30 mg de arbaclofeno (traçado roxo), quando identificaram faces. Os traçados representam a atividade registada no canal PO9 (canal 28, hemisfério esquerdo, (a)) e no canal PO10 (canal 32, hemisfério direito, (b)), com as respetivas áreas sombreadas indicando o erro padrão da média.

A análise dos efeitos farmacológicos no grupo PEA replica o padrão geral observado no grupo de controlo, com ausência de alterações marcadas nos potenciais evocados médios através das diferentes doses de arbaclofeno. Tal como anteriormente, verifica-se uma assimetria hemisférica na sensibilidade aos efeitos farmacológicos, com o hemisfério esquerdo a demonstrar maior variabilidade entre condições. Neste hemisfério, distinguem-se modulações ligeiras, mas discerníveis nas amplitudes de ambos os componentes principais. O P100 apresenta variações subtis entre doses, com a toma de 15 mg de arbaclofeno a demonstrar a amplitude mais reduzida. De forma mais evidente, o componente N170 exhibe diferenças na amplitude máxima entre condições, sugerindo uma possível modulação dose-dependente neste grupo clínico. O hemisfério direito mantém o padrão de elevada sobreposição entre condições farmacológicas, com traçados virtualmente indistinguíveis para as três doses testadas.

4.3 Análise Estatística revela efeitos da intervenção

A análise estatística dos dados eletrofisiológicos através de ANOVA de medidas repetidas revelou padrões diferenciados de modulação dos componentes P100 e N170 em função da dose de arbaclofeno administrada e do grupo experimental. Nesta secção serão apresentados os resultados dos modelos que demonstraram efeitos principais ou que se aproximaram do nível de significância ($p < .05$), organizados por componente, hemisfério e tipo de resposta comportamental. A análise identificou efeitos significativos em múltiplas condições experimentais, sugerindo que o arbaclofeno exerce uma influência seletiva sobre determinados aspetos do processamento visual, com padrões de resposta que variam entre os grupos de controlo e PEA. Os resultados serão apresentados seguindo uma estrutura sistemática,

reportando para cada análise os valores da estatística F, graus de liberdade, significância estatística e tamanho do efeito (eta quadrado parcial), complementados por análises post hoc quando apropriado.

4.3.1 Amplitude do P100

A análise da amplitude do componente P100 no hemisfério direito revelou um efeito principal significativo da dose ($F(2, 76) = 3.853, p = .025, \eta^2p = .092$), indicando que a administração de arbaclofeno modulou a amplitude deste componente. A análise de contrastes revelou um efeito linear significativo da dose ($F(1, 38) = 5.656, p = .023, \eta^2p = .130$), indicando uma modulação progressiva da amplitude do P100 com o aumento da dose de arbaclofeno. O contraste quadrático não foi significativo ($F(1, 38) = 1.537, p = .223, \eta^2p = .039$). Não se verificou um efeito significativo do grupo ($F(1, 38) = .263, p = .611, \eta^2p = .007$), sugerindo que não existem diferenças na amplitude do P100 entre participantes com PEA e controlos neste hemisfério. A interação dose \times grupo não atingiu significância estatística ($F(2, 76) = 1.532, p = .223, \eta^2p = .039$). As comparações post hoc demonstraram diferenças significativas entre o placebo e a dose de 30 mg (diferença média = $.923 \mu V, p = .023$), com a dose mais elevada a produzir, em média, amplitudes inferiores. A comparação entre placebo e 15 mg aproximou-se da significância (diferença média = $.830 \mu V, p = .050$), enquanto a diferença entre 15 mg e 30 mg não foi estatisticamente significativa ($p = .747$).

A análise da amplitude do componente P100 no hemisfério esquerdo revelou um efeito principal significativo da dose ($F(2, 84) = 4.194, p = .018, \eta^2p = .091$), confirmando que o arbaclofeno modulou bilateralmente a amplitude deste componente. A análise de contrastes não revelou efeitos lineares significativos ($p = .066$), mas identificou um efeito quadrático significativo da dose $F(1, 42) = 4.623, p = .037, \eta^2p = .099$, indicando um padrão não-linear de modulação. Similarmente ao hemisfério direito, não se verificou efeito significativo do grupo ($F(1, 42) = .186, p = .668, \eta^2p = .004$), nem da interação dose \times grupo ($F(2, 84) = 1.811, p = .170, \eta^2p = .041$). As comparações post hoc demonstraram que a diferença entre placebo e 15 mg atingiu significância estatística (diferença média = $.898 \mu V, p = .006$), assim como a diferença entre placebo e 30 mg (diferença média = $.534 \mu V, p = .066$), embora esta última se aproxime apenas do limiar de significância. Não se observaram diferenças significativas entre as doses de 15 mg e 30 mg ($p = .288$). Este padrão sugere que a dose de 15 mg produziu o efeito máximo na amplitude do P100 no hemisfério esquerdo, sem benefício adicional com o aumento para 30 mg. A visualização através de boxplots (Figura 14) permite uma análise dos padrões de modulação da amplitude do P100 em ambos os hemisférios e grupos experimentais.

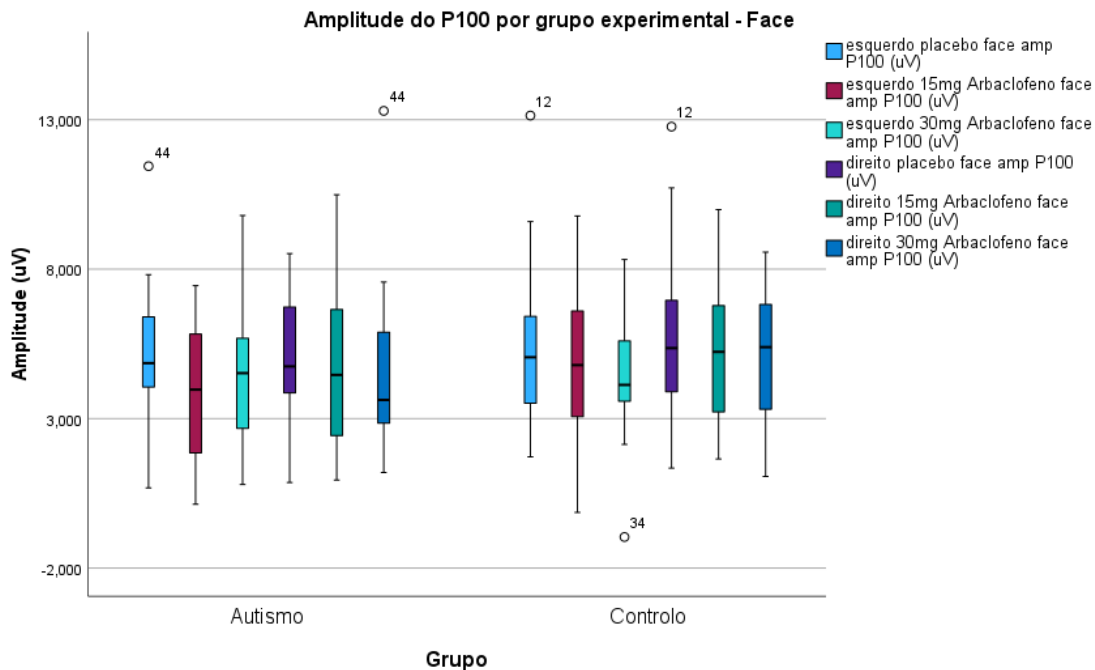


Figura 14 - Boxplots relativos à amplitude do P100 para a condição de identificação de face, para ambos os grupos e hemisférios.

Os dados revelam uma notável consistência nas amplitudes basais (placebo) entre grupos, com medianas aproximadamente entre 3-5 µV em ambos os hemisférios, corroborando a ausência de diferenças significativas entre controlos e participantes com PEA reportada nas análises estatísticas.

A análise da amplitude do P100 no hemisfério direito para a condição de não identificação de faces revelou um efeito principal significativo da dose ($F(2, 80) = 5.669, p = .005, \eta^2p = .124$), demonstrando que o arbaclofeno modula este componente mesmo na ausência de percepção consciente de faces. A análise de contrastes revelou um efeito linear significativo da dose ($F(1, 40) = 8.339, p = .006, \eta^2p = .173$), indicando uma diminuição progressiva da amplitude do P100 com o incremento da dose. O contraste quadrático não atingiu significância ($p = .290$). Não se verificaram efeitos significativos do grupo ($F(1, 40) = .076, p = .784, \eta^2p = .002$), nem da interação dose \times grupo ($F(2, 80) = .465, p = .630, \eta^2p = .011$). As comparações post hoc demonstraram diferenças significativas entre placebo e 15 mg (diferença média = .733 µV, $p = .021$) e entre placebo e 30 mg (diferença média = .981 µV, $p = .006$). A comparação entre 15 mg e 30 mg não foi significativa ($p = .345$), sugerindo que ambas as doses produzem diminuições substanciais na amplitude do P100, com um efeito ligeiramente mais pronunciado, em média, na dose de 30 mg.

A análise da amplitude do P100 no hemisfério esquerdo para a condição de não identificação de faces revelou um efeito principal da dose que se aproximou da significância estatística ($F(2, 84) = 3.213, p = .045, \eta^2p = .071$). A análise de contrastes identificou um efeito linear significativo da dose ($F(1, 42) = 5.384, p = .025, \eta^2p = .114$), sugerindo uma diminuição progressiva da amplitude com o incremento da dose. O contraste quadrático não foi significativo ($p = .964$).

Não se verificaram efeitos significativos do grupo ($F(1, 42) = .006, p = .939, \eta^2p = .000$), nem da interação dose \times grupo ($F(2, 84) = .345, p = .709, \eta^2p = .008$), indicando, igualmente, padrões de resposta similares entre grupos. As comparações post hoc revelaram que apenas a diferença entre placebo e 30 mg atingiu significância estatística (diferença média = $.808 \mu\text{V}$, $p = .025$), enquanto as comparações entre placebo e 15 mg ($p = .137$) e entre 15 mg e 30 mg ($p = .240$) não foram significativas. É importante notar que o teste de Levene indicou violação do pressuposto de homogeneidade de variâncias para a condição 30 mg ($p = .043$), sugerindo cautela na interpretação destes resultados. O padrão observado indica que o hemisfério esquerdo apresenta uma resposta mais gradual ao arbaclofeno na condição de não-face, requerendo a dose máxima para produzir efeitos significativos sobre a amplitude do P100. Na Figura 15 observam-se os boxplots para a condição de não identificação de faces, da amplitude do P100 em ambos os hemisférios e grupos experimentais.

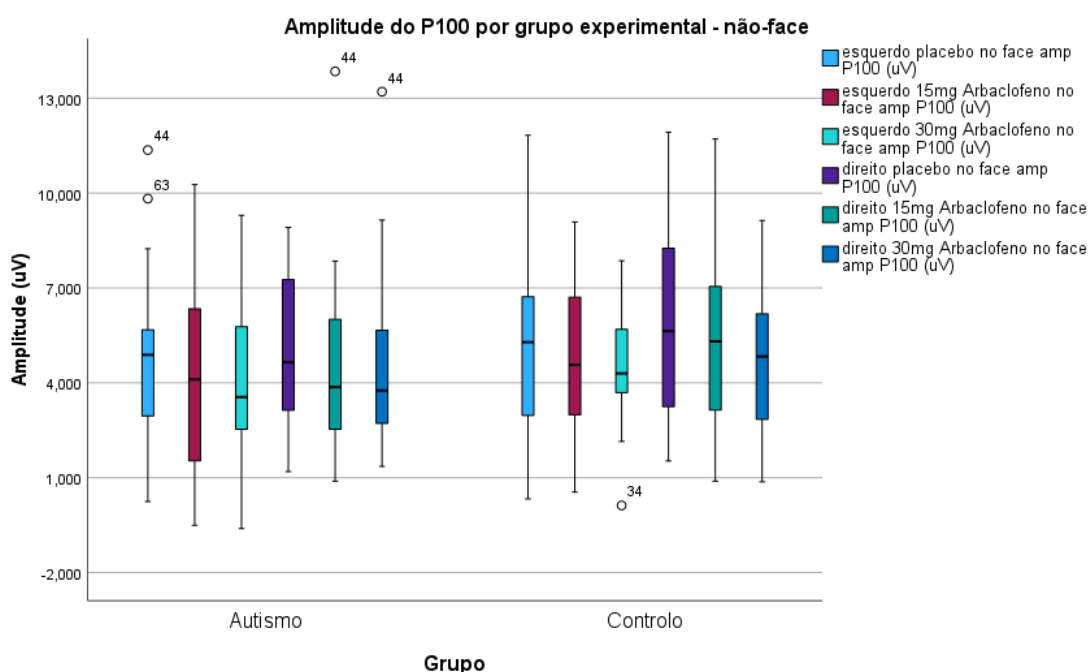


Figura 15 - Boxplots relativos à amplitude do P100 para a condição de não identificação de face, para ambos os grupos e hemisférios.

A análise visual dos boxplots revela padrões de amplitude do P100 comparáveis aos observados na condição de identificação facial, com medianas tipicamente entre 3-5 μV . Um aspeto distintivo desta condição é a clara demonstração do efeito dose-dependente do arbaclofeno em ambos os hemisférios e grupos. Contrariamente ao padrão nem sempre linear observado na condição de identificação de faces, aqui observa-se uma progressão linear consistente, com as amplitudes diminuindo progressivamente do placebo para 15 mg e desta para 30 mg. Este padrão bilateral e consistente entre grupos sugere que o arbaclofeno modula de forma mais uniforme os mecanismos visuais básicos quando o processamento de faces não está ativo. A maior variabilidade interindividual observada no grupo com PEA mantém-se evidente, com múltiplos *outliers* extremos (valores superiores a 10 μV). A presença de *outliers* negativos no

grupo de controlo (valor próximo de zero na condição 30 mg) indica participantes com respostas atípicas que podem ter influenciado a heterogeneidade de variâncias detetada pelo teste de Levene.

4.3.2 Latência do N170

A análise da latência do componente N170 no hemisfério esquerdo para a condição de identificação de faces não revelou efeito principal significativo da dose ($F(2, 84) = .663, p = .518, \eta^2p = .016$), indicando que o arbaclofeno não alterou o timing do processamento facial neste hemisfério. Similarmente, não se verificou efeito principal do grupo ($F(1, 42) = .082, p = .776, \eta^2p = .002$). Contudo, observou-se uma interação dose \times grupo significativa ($F(2, 84) = 3.354, p = .040, \eta^2p = .074$), sugerindo que os grupos responderam diferencialmente à administração do fármaco. A análise de contrastes para a interação revelou um efeito linear significativo ($F(1, 42) = 4.795, p = .034, \eta^2p = .102$), indicando que a modulação da latência do N170 pelo arbaclofeno seguiu padrões opostos nos dois grupos. O contraste quadrático não foi significativo ($p = .283$). Estes resultados sugerem que o arbaclofeno pode ter efeitos diferenciais sobre a velocidade do processamento facial dependendo da presença do diagnóstico de autismo.

A análise dos testes t emparelhados revelou padrões distintos de modulação da latência do N170 pelo arbaclofeno em cada grupo. No grupo com PEA, não se observaram diferenças significativas entre nenhuma das comparações de doses: placebo vs. 15 mg ($t(21) = -.130, p = .897$), placebo vs. 30 mg ($t(20) = .667, p = .513$), e 15 mg vs. 30 mg ($t(21) = 1.810, p = .085$). As latências médias mantiveram-se relativamente estáveis ao longo das doses (placebo: 202.91 ms; 15 mg: 203.27 ms; 30 mg: 200.95 ms), com tamanhos de efeito negligíveis (d de Cohen a variar entre $-.028$ e $.386$). Em contraste, o grupo de controlo demonstrou alterações significativas e progressivas na latência do N170 com o aumento da dose de arbaclofeno. Enquanto a comparação placebo vs. 15 mg não atingiu significância ($t(27) = -1.045, p = .304$), observaram-se aumentos significativos da latência tanto na comparação placebo vs. 30 mg ($t(24) = -2.602, p = .016, d = -.520$) como na comparação 15 mg vs. 30 mg ($t(27) = -2.574, p = .016, d = -.486$). As latências médias aumentaram progressivamente de 196.07 ms (placebo) para 198.43 ms (15 mg) e 201.20 ms (30 mg), representando tamanhos de efeito médios.

Em contraste com os efeitos observados no hemisfério esquerdo, a análise da latência do N170 no hemisfério direito não revelou quaisquer diferenças estatisticamente significativas. Não se verificaram efeitos principais da dose ou do grupo, nem interação entre estes fatores, sugerindo que o hemisfério direito mantém uma estabilidade temporal no processamento facial que não é afetada pela modulação GABAérgica. Esta assimetria hemisférica na resposta ao arbaclofeno reforça a noção de que os mecanismos de processamento facial podem apresentar diferentes sensibilidades farmacológicas entre hemisférios. Na Figura 16 observam-se os boxplots para a condição de identificação de faces, da latência do N170 em ambos os grupos no hemisfério esquerdo.

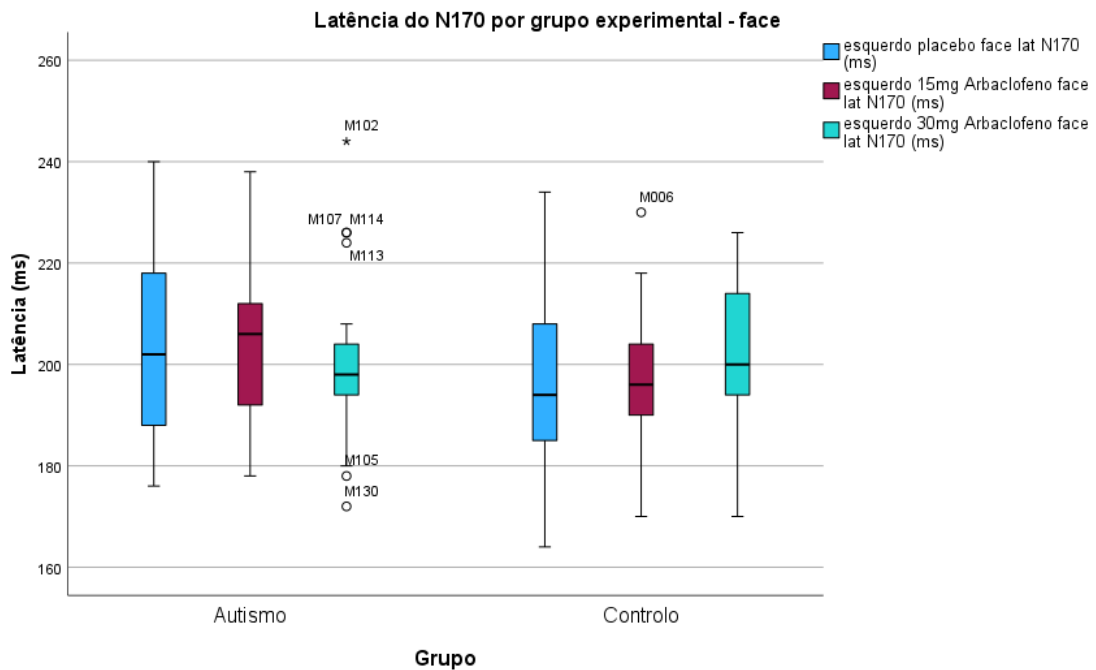


Figura 16 - Boxplots relativos à latência do N170 para a condição de identificação de face, para ambos os grupos no hemisfério esquerdo.

A visualização através de boxplots da latência do N170 no hemisfério esquerdo permite identificar a interação significativa entre dose e grupo identificada na análise estatística. No grupo com PEA, observa-se um padrão não-linear de resposta ao arbaclofeno, com as latências a aumentarem inicialmente da condição placebo (≈ 202 ms) para 15 mg (≈ 210 ms), seguido de uma redução acentuada na dose de 30 mg para valores ligeiramente inferiores a 200 ms. Este padrão em forma de U invertido contrasta com o grupo de controlo, que demonstra um aumento progressivo e consistente das latências com o incremento da dose, partindo de aproximadamente 195 ms no placebo para 200 ms na dose máxima. Na análise da distribuição dos dados, identifica-se maior variabilidade no grupo PEA na condição placebo, com um intervalo interquartil substancialmente maior comparado ao grupo de controlo.

4.3.3 Amplitude do N170

A análise da amplitude do componente N170 no hemisfério esquerdo para a condição de não identificação de faces revelou um efeito principal da dose significativo ($F(2, 84) = 3.952, p = .023, \eta^2 p = .086$). A análise de contrastes identificou um efeito linear significativo da dose ($F(1, 42) = 4.793, p = .034, \eta^2 p = .102$), indicando um aumento progressivo da amplitude com o aumento da dose. Não se verificaram efeitos significativos do grupo ($F(1, 42) = .002, p = .964, \eta^2 p = .000$) nem da interação dose \times grupo ($F(2, 84) = .379, p = .685, \eta^2 p = .009$). As comparações post hoc revelaram diferenças significativas entre placebo e 30 mg (diferença média = $1.385 \mu V, p = .034$) e entre 15 mg e 30 mg (diferença média = $1.147 \mu V, p = .037$). A comparação entre placebo e 15 mg não atingiu significância ($p = .543$). É importante notar que o teste de Levene indicou

violação do pressuposto de homogeneidade de variâncias para a condição 15 mg ($p = .003$), sugerindo cautela na interpretação. O padrão de resultados indica que a dose de 30 mg produz um aumento significativo da amplitude do N170, sugerindo um efeito do arbaclofeno sobre este componente na ausência de processamento facial específico.

De forma contrária, a análise da amplitude do N170 no hemisfério direito para a mesma condição de não identificação de faces não revelou quaisquer efeitos estatisticamente significativos. Não se observaram efeitos principais da dose, do grupo, nem interação entre estes fatores, indicando que o hemisfério direito mantém uma estabilidade na amplitude do N170 durante o processamento de estímulos não-faciais que não é afetada pela modulação GABAérgica. Esta assimetria hemisférica na resposta ao arbaclofeno sugere mecanismos diferenciais de processamento visual entre hemisférios, mesmo quando os estímulos não são categorizados como faces. Na Figura 17 observam-se os boxplots para a condição de não identificação de faces, da amplitude do N170 em ambos os grupos no hemisfério esquerdo.

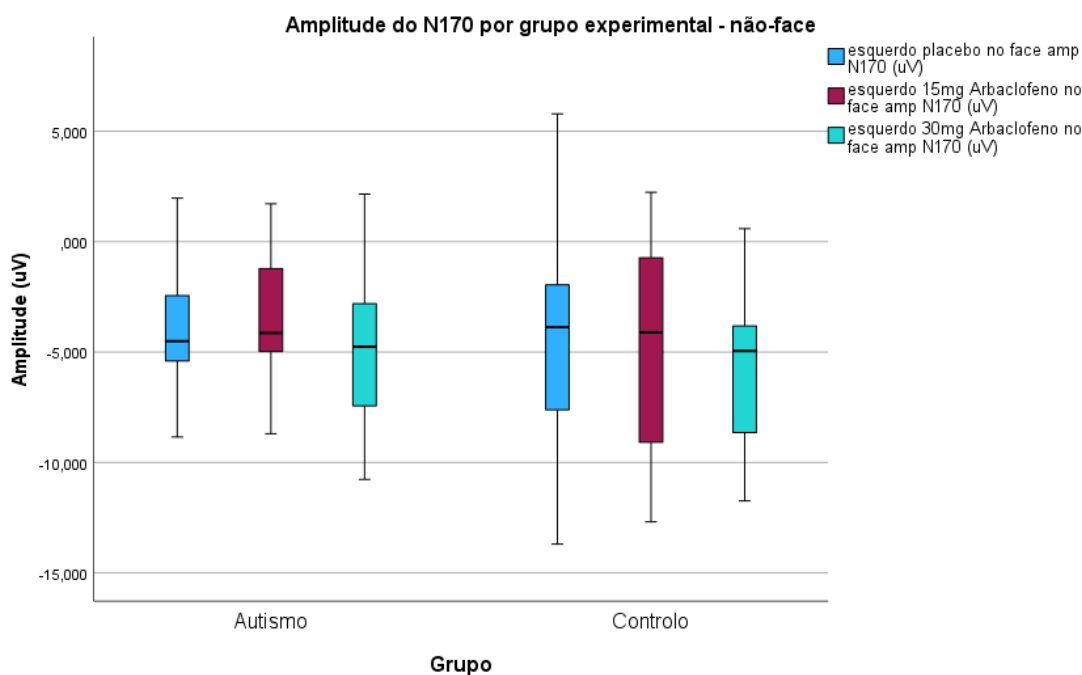


Figura 17 - Boxplots relativos à amplitude do N170 para a condição de não identificação de face, para ambos os grupos no hemisfério esquerdo.

A visualização através de boxplots da amplitude do N170 no hemisfério esquerdo para a condição de não identificação de faces clarifica o padrão de modulação identificado na análise estatística. É possível observar que o arbaclofeno produz um aumento progressivo da amplitude do N170 (valores mais negativos) com o incremento da dose, particularmente evidente na transição do placebo (medianas aproximadamente $-4 \mu\text{V}$) para a dose de 30 mg (medianas aproximadamente $-5 \mu\text{V}$) em ambos os grupos. O padrão de resposta é notavelmente consistente entre grupos, com trajetórias paralelas de aumento da amplitude, sugerindo que este efeito do arbaclofeno sobre o N170 na condição não-face é independente do diagnóstico de PEA. Como indicado pelos testes post hoc, a dose de 15 mg ocupa uma posição intermédia,

com amplitudes que não diferem significativamente do placebo, confirmando que o efeito requer a dose máxima de 30 mg para se manifestar plenamente.

4.3.4 Amplitude Pico a Pico

A análise da amplitude pico a pico no hemisfério esquerdo para a condição de identificação de faces revelou um efeito principal significativo da dose ($F(2, 84) = 197.579, p < .001, \eta^2p = .825$), indicando uma modulação desta medida pelo arbaclofeno. A análise de contrastes revelou um efeito quadrático significativo ($F(1, 42) = 221.431, p < .001, \eta^2p = .841$), caracterizando um padrão não-linear de resposta à dose. Observou-se também um efeito principal significativo do grupo ($F(1, 42) = 4.248, p = .046, \eta^2p = .092$), sugerindo diferenças basais entre controlos e participantes com PEA. A interação dose \times grupo não atingiu significância estatística ($F(2, 84) = 2.534, p = .085, \eta^2p = .057$), indicando padrões de resposta similares ao fármaco em ambos os grupos. As comparações post hoc demonstraram reduções significativas da amplitude interpicos entre placebo e 15 mg (diferença média = $-16.242 \mu V, p < .001$) e entre 15 mg e 30 mg (diferença média = $-16.488 \mu V, p < .001$). A comparação entre placebo e 30 mg não foi significativa ($p = .581$), sugerindo um efeito em forma de U invertido, onde a dose de 15 mg produz a redução máxima na diferença entre os picos P100 e N170. É importante notar que o teste de Levene indicou violação do pressuposto de homogeneidade de variâncias para a condição 15 mg ($p = .003$), requerendo cautela interpretativa.

Contrariamente, a análise da amplitude pico a pico no hemisfério direito para a mesma condição de identificação de faces não revelou, novamente, quaisquer efeitos estatisticamente significativos. Na Figura 18 observam-se os boxplots para a condição de identificação de faces, da amplitude pico a pico em ambos os grupos no hemisfério esquerdo.

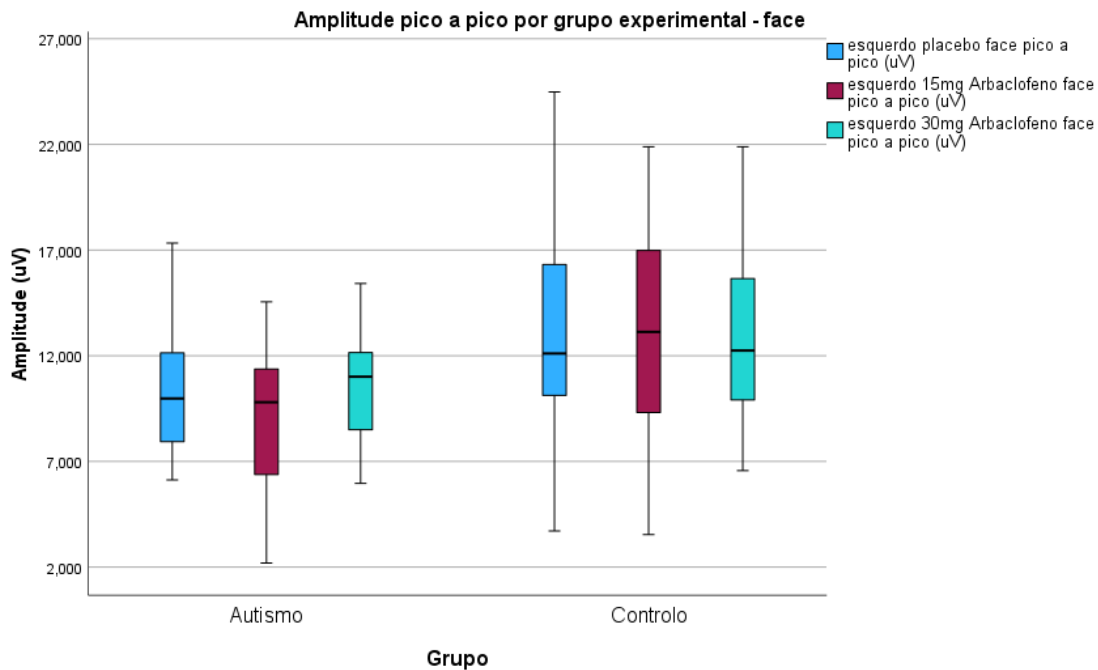


Figura 18 - Boxplots relativos à amplitude pico a pico para a condição de identificação de face, para ambos os grupos no hemisfério esquerdo.

A visualização através de boxplots da amplitude pico a pico no hemisfério esquerdo ilustra os padrões diferenciais de resposta ao arbaclofeno entre grupos. No grupo de controlo, observa-se o característico padrão em U invertido, com amplitudes pico a pico que aumentam da condição placebo (mediana $\approx 12 \mu\text{V}$) para um pico na dose de 15 mg (mediana $\approx 14 \mu\text{V}$), seguido de uma redução na dose de 30 mg (mediana $\approx 12 \mu\text{V}$). Este padrão não-linear confirma o efeito quadrático significativo identificado na análise estatística. A maior variabilidade observada no grupo de controlo, particularmente evidente na toma de 15 mg de arbaclofeno com um intervalo interquartil substancialmente maior, contrasta com a relativa homogeneidade do grupo PEA. Esta heterogeneidade no grupo neurotípico pode indicar diferenças individuais nas estratégias de processamento facial que são moduladas de forma mais pronunciada pelo arbaclofeno, enquanto o grupo PEA apresenta um padrão mais rígido e menos responsivo à manipulação farmacológica.

4.3.5 Diferença hemisférica na amplitude do N170

Adicionalmente, tendo em consideração as diferenças de amplitude do N170 observadas entre os hemisférios esquerdo e direito na análise dos ERPs médios, procedeu-se à realização de uma ANOVA de medidas repetidas com design 3×2 , incluindo dois fatores intra-sujeitos: dose (placebo, 15 mg, 30 mg) e hemisfério (esquerdo, direito). Esta análise permitiu avaliar formalmente se as assimetrias hemisféricas observadas visualmente atingiam significância estatística e se eram moduladas pela administração de arbaclofeno. A inclusão do fator hemisfério como variável intra-sujeitos possibilitou ainda a investigação de potenciais

interações dose × hemisfério, que poderiam revelar efeitos diferenciais do fármaco sobre os mecanismos de lateralização do processamento facial em cada grupo experimental.

A análise da lateralização hemisférica através de ANOVA de medidas repetidas não revelou um efeito principal do hemisfério ($F(1, 38) = 3.591, p = .066, \eta^2p = .086$). Não se verificaram efeitos principais significativos da dose ($p = .660$) ou do grupo ($F(1, 38) = 2.323, p = .136$), nem interação dose × grupo ($p = .681$). Observou-se, ainda, uma interação dose × hemisfério significativa ($F(2, 76) = 2.539, p = .039, \eta^2p = .160$), com um tamanho de efeito grande. A análise de contrastes revelou que esta interação segue um padrão linear ($F(1, 38) = 7.235, p = .011, \eta^2p = .160$), indicando que o efeito da dose sobre a amplitude do N170 difere sistematicamente entre hemisférios. A análise das médias marginais revela um padrão distintivo de lateralização do efeito farmacológico. No grupo de controlo, o hemisfério esquerdo demonstra uma modulação dose-dependente pronunciada, com amplitudes que progridem de $-7.768 \mu V$ (placebo) para $-8.723 \mu V$ (15 mg) e $-9.190 \mu V$ (30 mg), enquanto o hemisfério direito mantém-se relativamente estável ($-8.809 \mu V, -8.803 \mu V$ e $-8.711 \mu V$, respetivamente). O grupo PEA apresenta um padrão similar, mas atenuado, com o hemisfério esquerdo a variar de $-5.925 \mu V$ a $-6.365 \mu V$, enquanto o hemisfério direito permanece estável entre $-7.329 \mu V$ e $-7.736 \mu V$. Este padrão de resultados sugere que o arbaclofeno afeta preferencialmente o hemisfério esquerdo, independentemente do grupo diagnóstico, enquanto o hemisfério direito demonstra resistência aos efeitos moduladores do fármaco.

4.4 Análise com *Machine Learning*

4.4.1 Classificação Controlo vs PEA

Os resultados do primeiro algoritmo de classificação, desenvolvido para discriminar entre participantes do grupo de controlo e participantes com PEA utilizando características eletrofisiológicas basais, revelaram desempenhos modestos, mas informativos sobre a capacidade discriminativa dos componentes P100 e N170. Após a remoção de 13 sujeitos com dados totalmente ausentes (10 do grupo controlo e 3 do grupo PEA), a análise prosseguiu com 57 participantes, mantendo-se uma distribuição relativamente equilibrada entre grupos (31 controlos e 26 PEA). Na Tabela 3 encontram-se apresentados os principais resultados dos modelos de *machine learning* obtidos com este algoritmo.

Tabela 3 - Resultados dos modelos de *machine learning* de predição de grupo vs controlo.

	Exatidão	Precisão	Recall	F1-score
Regressão Logística	0,611	0,600	0,375	0,462
SVM	0,556	0,500	0,375	0,429
Random Forest	0,611	0,600	0,375	0,462
KNN	0,500	0,429	0,375	0,400

A análise comparativa dos modelos demonstrou que a Regressão Logística e o *Random Forest* alcançaram o melhor desempenho no conjunto de teste, ambos com uma exatidão de 61.1%, seguidos pelo SVM (55.6%) e KNN (50.0%). Contudo, é notável que todos os modelos apresentaram valores de sensibilidade consistentemente baixos (37.5%), indicando dificuldade sistemática na identificação correta de participantes com PEA. Esta limitação refletiu-se nos valores de F1-Score, que variaram entre 0.400 e 0.462, o que indica um desequilíbrio entre precisão e sensibilidade na classificação.

A validação cruzada estratificada com 5 partições revelou que o SVM apresentou o melhor desempenho médio ($54.4\% \pm 6.3\%$), seguido pela Regressão Logística ($54.2\% \pm 9.1\%$). Importa salientar que todos os modelos demonstraram desempenhos próximos ao acaso, com médias de validação cruzada inferiores a 55%, indicando limitada capacidade de generalização.

A análise de importância das características, extraída do modelo *Random Forest*, identificou a amplitude do P100 no hemisfério esquerdo como a característica mais informativa (12.5%), seguida pela latência do N170 esquerdo (10.6%) e latência do N170 direito (9.7%). Notavelmente, observou-se uma ligeira predominância do hemisfério esquerdo na contribuição total para a classificação (53.8% vs 46.2% para o hemisfério direito).

As matrizes de confusão revelaram um padrão consistente de enviesamento dos modelos para a classe majoritária (controlo), com os melhores modelos a classificar corretamente apenas 3 dos 8 participantes com PEA no conjunto de teste. Este padrão sugere que as diferenças eletrofisiológicas basais entre grupos, embora presentes, apresentam considerável sobreposição, dificultando a discriminação automática baseada exclusivamente nestes marcadores.

A análise PCA permitiu complementar os resultados de classificação. Os dois primeiros componentes principais capturaram 58.4% da variância total dos dados, com o PC1 a contribuir 35.5% e o PC2 com os 22.9% adicionais. Esta percentagem moderada de variância explicada indica que a informação eletrofisiológica se distribui por múltiplas dimensões, não sendo facilmente redutível a um espaço de baixa dimensionalidade. Na Figura 19 observa-se a distribuição dos dados de ambos os grupos na análise PCA realizada.

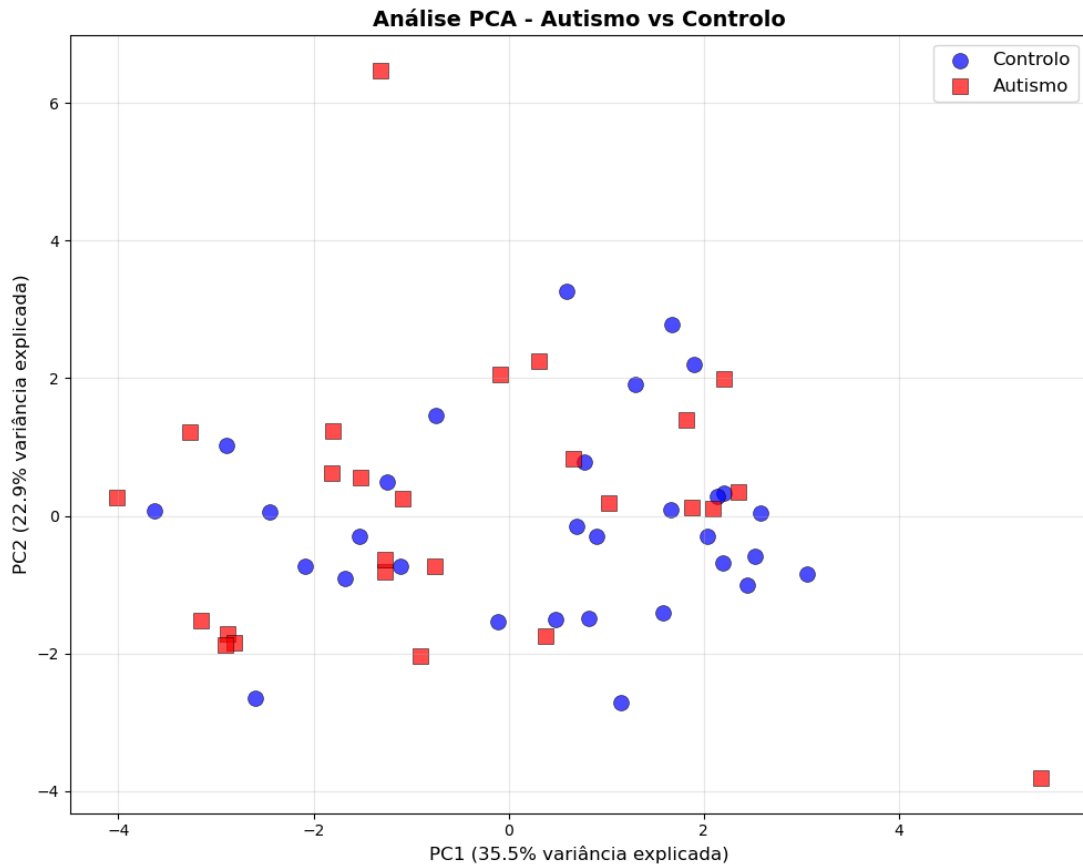


Figura 19 - PCA de dois componentes para a predição de controlo vs autismo.

A visualização dos dados representados pela análise PCA mostrou uma extensa sobreposição entre os grupos de controlo e PEA, corroborando os modestos resultados de classificação obtidos. A distância entre os centroides dos grupos foi de apenas 0.923 unidades no espaço normalizado, indicando proximidade substancial entre os perfis eletrofisiológicos médios de ambas as populações. Esta sobreposição é particularmente evidente na região central do gráfico, onde participantes de ambos os grupos se misturam indistintamente. Observou-se, contudo, alguma tendência para separação parcial ao longo do PC1, com uma ligeira concentração de participantes com PEA em valores mais negativos e participantes controlo distribuindo-se mais uniformemente ao longo deste eixo. A dispersão ao longo do PC2 mostrou-se menos informativa para a distinção entre grupos, com ambas as populações apresentando distribuições similares nesta dimensão.

Estes resultados indicam que, apesar dos componentes P100 e N170 apresentarem algumas diferenças entre populações, estas não são suficientemente robustas para permitir classificação fiável ao nível individual. A modesta performance dos algoritmos sugere que a heterogeneidade fenotípica característica da PEA se reflete também ao nível eletrofisiológico, requerendo possivelmente abordagens multidimensionais ou a inclusão de marcadores adicionais para melhorar a capacidade discriminativa.

A análise focada exclusivamente na latência do N170 revelou resultados comparáveis ao modelo completo, embora com algumas diferenças na distribuição do desempenho entre algoritmos. Utilizando apenas duas características (latência N170 bilateral), os modelos mantiveram capacidade discriminativa similar, sugerindo que esta medida específica contém grande parte da informação relevante para a distinção entre grupos. Na Tabela 4 encontram-se apresentados os principais resultados dos modelos de *machine learning* obtidos com este algoritmo.

Tabela 4 - Resultados dos modelos de *machine learning* de predição de grupo vs controlo apenas com os dados da latência do N170.

	Exatidão	Precisão	Recall	F1-score
Regressão Logística	0,556	0,500	0,375	0,429
SVM	0,611	0,571	0,500	0,533
Random Forest	0,444	0,375	0,375	0,375
KNN	0,611	0,556	0,625	0,588

A análise comparativa dos modelos demonstrou que o SVM e o KNN alcançaram o melhor desempenho no conjunto de teste (61.1%), igualando a performance dos melhores modelos do algoritmo completo. Destaca-se o KNN, que apresentou o maior F1-Score (0.588) e a melhor sensibilidade (62.5%), conseguindo identificar corretamente 5 dos 8 participantes com PEA - uma melhoria substancial relativamente aos 3 identificados no modelo completo. Esta melhoria na sensibilidade sugere que a simplificação do espaço de características pode ter reduzido o ruído e melhorado a capacidade de deteção do grupo PEA.

A validação cruzada revelou um padrão similar, com a Regressão Logística a apresentar o melhor desempenho médio ($59.2\% \pm 10.4\%$), superior ao obtido com o conjunto completo de características. Observou-se também uma elevada variabilidade em todos os modelos (desvios-padrão entre 10.4% e 14.0%), característica esperada dado o reduzido espaço de características. A análise de importância confirmou uma ligeira predominância do hemisfério esquerdo (52.4% vs 47.6%), consistente com os resultados do modelo completo.

A análise PCA com apenas duas variáveis capturou, naturalmente, 100% da variância, com o PC1 a apresentar 90.6% de variância explicada. Na Figura 20 observa-se a distribuição dos dados de ambos os grupos na análise PCA realizada.

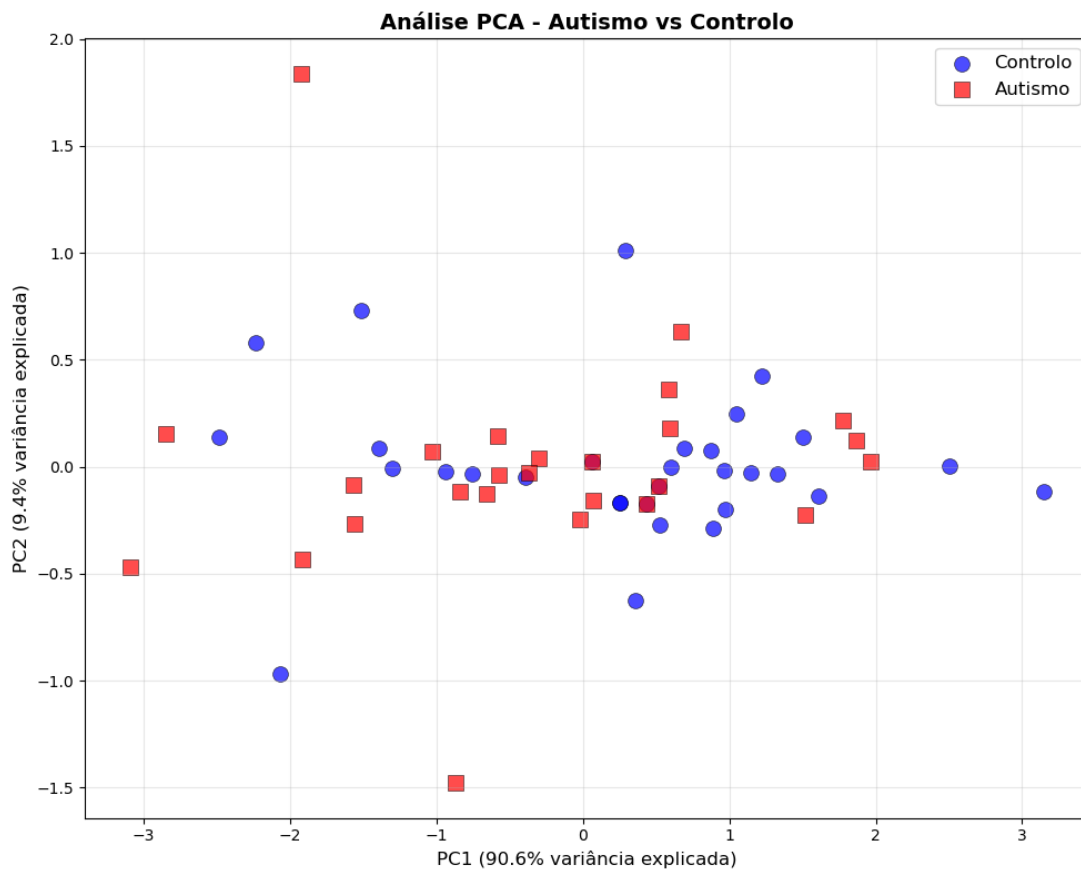


Figura 20 - PCA de dois componentes para a predição de controlo vs autismo apenas com os dados da latência do N170.

A análise PCA revelou uma distribuição essencialmente unidimensional dos dados ao longo do PC1, refletindo a alta correlação entre as latências bilaterais do N170. A distância entre centroides reduziu-se para 0.574, inferior aos 0.923 observados no modelo completo, indicando maior sobreposição entre grupos quando consideradas apenas estas características.

Estes resultados sugerem que, embora a latência do N170 contenha informação discriminativa relevante, a sua utilização isolada não melhora substancialmente a capacidade de classificação. A manutenção de desempenhos próximos ao acaso reforça a conclusão de que as diferenças eletrofisiológicas entre grupos, mesmo quando focadas em componentes específicos reconhecidamente sensíveis ao processamento facial, apresentam sobreposição considerável ao nível individual, limitando sua aplicabilidade como biomarcadores diagnósticos únicos.

4.4.2 Classificação de Dose

Os resultados do algoritmo de classificação de doses farmacológicas revelaram capacidade discriminativa limitada, com todos os modelos a apresentar desempenhos próximos ou inferiores ao esperado por acaso para classificação triclasse. A análise processou 173 observações correspondentes a ensaios com identificação facial, distribuídas de forma

relativamente equilibrada entre as três condições: placebo (57), 15 mg (63) e 30 mg (53) de arbaclofeno. Na Tabela 5 encontram-se apresentados os principais resultados dos modelos de *machine learning* obtidos com este algoritmo.

Tabela 5 - Resultados dos modelos de *machine learning* de predição de dose farmacológica.

	Exatidão	Precisão	Recall	F1-score
Regressão Logística	0,423	0,427	0,423	0,414
SVM	0,269	0,232	0,269	0,223
Random Forest	0,346	0,324	0,346	0,327
KNN	0,327	0,239	0,327	0,276

A comparação dos modelos determinou que a Regressão Logística alcançou o melhor desempenho com uma exatidão de 42.3%, valor marginalmente superior ao acaso (33.3%). O *Random Forest* e KNN apresentaram desempenhos intermédios (34.6% e 32.7%, respetivamente), enquanto o SVM obteve o pior resultado (26.9%), inferior ao esperado aleatoriamente. As métricas ponderadas de precisão, sensibilidade e F1-Score acompanharam este padrão, confirmando a dificuldade generalizada na discriminação entre doses.

A análise das matrizes de confusão apresentou padrões interessantes de classificação. O modelo com melhor desempenho (Regressão Logística) demonstrou maior capacidade para identificar corretamente a dose de 15 mg (52.6% de sensibilidade), mas apresentou dificuldades consideráveis com a dose de 30 mg (25.0% de sensibilidade). Notavelmente, o KNN foi incapaz de classificar corretamente qualquer observação da dose mais elevada, sugerindo que esta condição apresenta características eletrofisiológicas particularmente ambíguas.

A validação cruzada estratificada revelou desempenhos ainda mais modestos, com o *Random Forest* a apresentar a melhor média de apenas 30.6% ($\pm 4.7\%$). Todos os modelos demonstraram médias de validação cruzada inferiores ao acaso, indicando sobreajustamento aos dados de treino e incapacidade de generalização. O SVM apresentou o menor desvio padrão ($\pm 1.9\%$), enquanto a Regressão Logística apresentou o maior ($\pm 9.7\%$).

A análise de importância das características identificou as medidas derivadas (pico a pico e latência interpicos) como as mais informativas, representando 35.5% da importância total nas 10 características principais. O componente N170 contribuiu ligeiramente mais que o P100 (32.2% vs 30.9%), com uma distribuição equilibrada entre hemisférios (50.6% esquerdo vs 47.9% direito). Esta distribuição relativamente uniforme sugere ausência de marcadores eletrofisiológicos dominantes para a discriminação entre doses.

A análise PCA capturou 54.4% da variância total com dois componentes principais, indicando considerável complexidade multidimensional nos dados. Na Figura 21 observa-se a distribuição dos dados das três doses na análise PCA realizada.

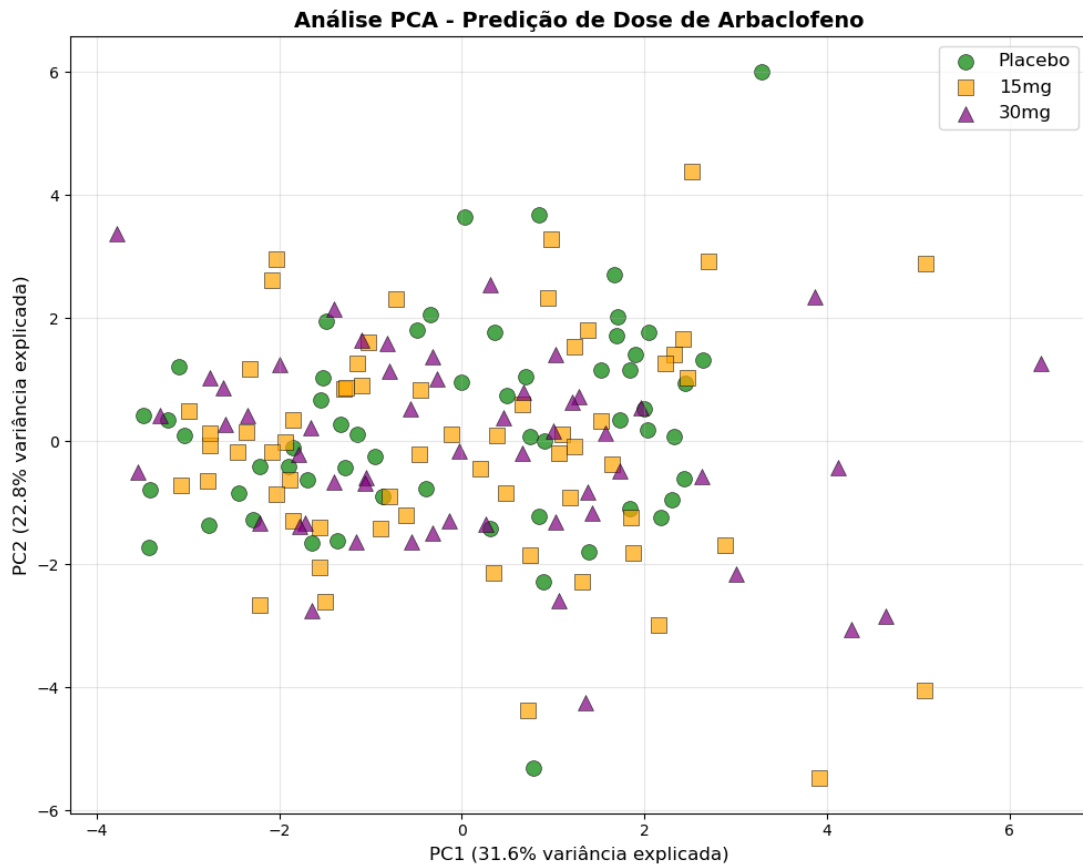


Figura 21 - PCA de dois componentes para a predição de dose farmacológica.

A visualização no espaço PCA revelou extensa sobreposição entre as três condições farmacológicas, formando essencialmente uma nuvem contínua sem separação clara. Embora as distâncias entre centroides sugerissem alguma estrutura (Placebo-15 mg: 0.343; Placebo-30 mg: 0.425), a mínima separação entre 15 mg e 30 mg de arbaclofeno (0.084) indica ausência de efeito dose-resposta linear nos padrões eletrofisiológicos. A análise de dispersão revelou desvios-padrão elevados e similares entre condições, com ligeiro aumento da variabilidade na dose mais elevada (PC1: 2.197), indicando uma maior heterogeneidade na resposta individual ao arbaclofeno em doses superiores.

Estes resultados indicam que as modulações eletrofisiológicas induzidas pelo arbaclofeno, embora detetáveis através de análises estatísticas convencionais, não produzem assinaturas suficientemente robustas ou consistentes para permitir classificação fiável da dose administrada. A elevada variabilidade interindividual e a aparente não-linearidade dos efeitos farmacológicos limitam significativamente a utilidade dos componentes P100 e N170 como marcadores de resposta ao tratamento com arbaclofeno em contextos de processamento facial.

4.4.3 Classificação face vs não-face

Os resultados do algoritmo de classificação comportamental, desenvolvido para discriminar entre ensaios com e sem identificação facial baseando-se exclusivamente na amplitude do N170, revelaram desempenhos modestos, mas consistentes entre modelos. A análise processou 116 observações equilibradas entre condições após remoção de 24 observações com dados insuficientes, utilizando seis características correspondentes à amplitude bilateral do N170 nas três condições farmacológicas. Na Tabela 6 encontram-se apresentados os principais resultados dos modelos de *machine learning* obtidos com este algoritmo.

Tabela 6 - Resultados dos modelos de *machine learning* de classificação comportamental.

	Exatidão	Precisão	Recall	F1-score
Regressão Logística	0,543	0,529	0,529	0,529
SVM	0,486	0,462	0,353	0,400
Random Forest	0,543	0,538	0,412	0,467
KNN	0,543	0,529	0,529	0,529

A comparação dos modelos demonstrou notável uniformidade nos desempenhos, com três algoritmos (Regressão Logística, *Random Forest* e KNN) alcançando exatidão idêntica de 54.3% no conjunto de teste. Apenas o SVM apresentou desempenho inferior (48.6%), aproximando-se do acaso. Embora modestos, estes resultados representam capacidade discriminativa marginalmente superior ao esperado aleatoriamente, sugerindo que a amplitude do N170 contém informação relevante, ainda que limitada, sobre o processamento facial.

A análise das matrizes de confusão revelou padrões equilibrados de classificação para a Regressão Logística e KNN, com sensibilidade e especificidade idênticas (52.9%), indicando ausência de enviesamento sistemático. O *Random Forest* demonstrou padrão distinto, com maior especificidade (66.7%) mas sensibilidade reduzida (41.2%), sugerindo critério mais conservador na identificação de processamento facial.

A validação cruzada estratificada revelou desempenhos substancialmente superiores, com o KNN a alcançar 62.1% ($\pm 4.7\%$) e a Regressão Logística 60.4% ($\pm 7.3\%$). Esta discrepância entre validação cruzada e conjunto de teste sugere que o desempenho real dos modelos pode ser superior ao observado, sendo limitado pelo tamanho reduzido da amostra de teste e pelo elevado número de valores imputados (66 em 696 pontos de dados).

A análise de importância das características revelou que a condição de 15 mg de arbaclofeno apresentou a maior contribuição agregada (35.2%), com destaque para a amplitude no hemisfério direito (19.2%). A distribuição hemisférica mostrou-se notavelmente equilibrada (50.4% esquerdo vs 49.6% direito).

A análise PCA capturou 84.0% da variância total, com predominância do PC1 (74.0%), indicando que a informação discriminativa se concentra numa dimensão principal. A distância entre centroides (1.367) foi substancialmente superior às observadas nos algoritmos de classificação

de grupo ou dose, confirmando que as diferenças na amplitude do N170 entre condições comportamentais são mais pronunciadas que entre grupos diagnósticos ou condições farmacológicas. Na Figura 22 observa-se a distribuição dos dados de ambos os grupos e condições na análise PCA realizada.

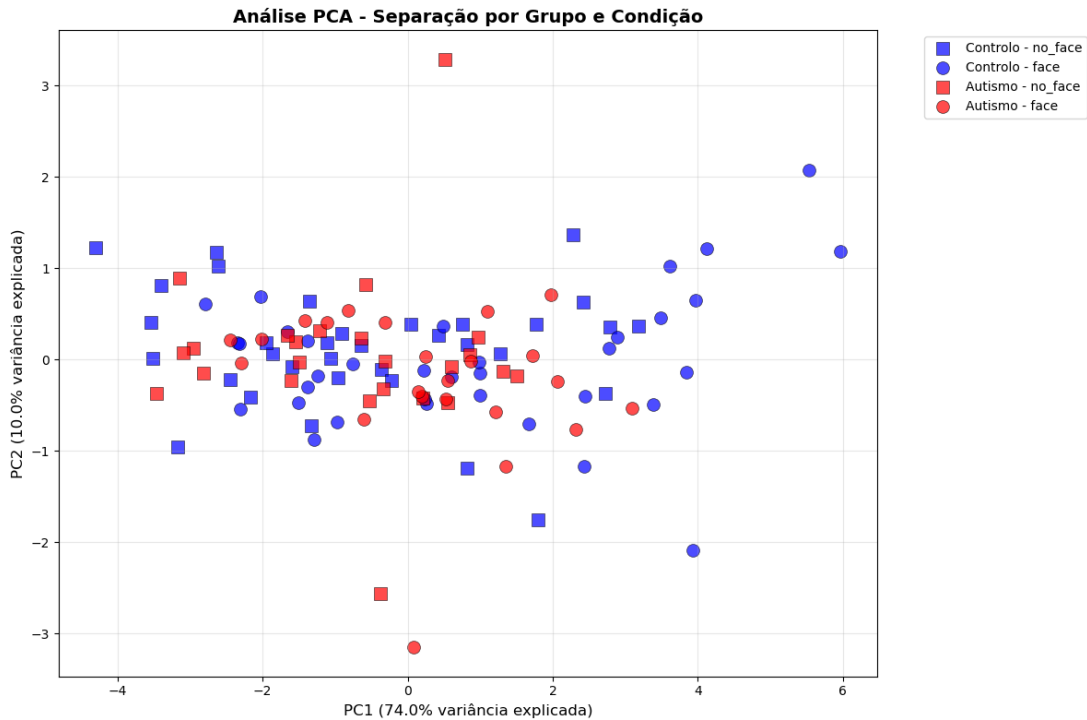


Figura 22 - PCA de dois componentes para a classificação comportamental.

A análise PCA revelou uma distribuição complexa sem separação clara entre condições ou grupos. Notavelmente, a análise de padrões extremos identificou 12 observações acima do percentil 90 do PC1, das quais 11 correspondiam a ensaios com identificação facial e 11 pertenciam ao grupo controlo. A análise post hoc por grupo indicou que o melhor modelo, neste caso o KNN, apresentou fraco desempenho no grupo controlo (47.6%). No entanto, demonstrou capacidade discriminativa superior no grupo PEA (64.3%). Esta diferença de desempenho entre os grupos, sugere que os padrões de amplitude do N170 associados ao processamento facial foram mais consistentes ou pronunciados nos indivíduos que pertenciam ao grupo com PEA. Estes resultados indicam que, embora a amplitude do N170 contenha informação discriminativa sobre o processamento facial, esta capacidade é modesta e varia significativamente entre populações.

5 Discussão

5.1 Discussão de Resultados

Neste estudo foram estudados os efeitos de uma intervenção farmacológica de dose única nos componentes eletrofisiológicos P100 e N170 em indivíduos com e sem PEA, durante uma tarefa de identificação de faces. Os resultados revelaram padrões complexos de resposta eletrofisiológica, diferenças subtis entre grupos diagnósticos e limitações importantes na aplicação de técnicas de *machine learning* para classificação automática baseada em marcadores eletrofisiológicos.

O pipeline de pré-processamento implementado demonstrou eficácia na melhoria da qualidade dos sinais eletrofisiológicos, conforme evidenciado pela substancial redução do ruído e pela clara definição dos componentes P100 e N170 nos ERPs processados. A aplicação sequencial de filtros, detecção automática de canais ruidosos, e remoção de épocas contaminadas criou um protocolo reprodutível que preservou a integridade dos sinais neuronais, eliminando fontes de variabilidade não relacionadas com o processamento cognitivo de interesse.

O algoritmo de detecção automática de picos demonstrou desempenho geralmente satisfatório, identificando corretamente os componentes P100 e N170 na maioria dos participantes. Contudo, a necessidade de correção manual em alguns dos casos indica limitações importantes que merecem reflexão. Estas situações de falha ocorreram predominantemente em participantes que apresentavam morfologias atípicas dos componentes ou latências fora do intervalo esperado, sugerindo que a variabilidade interindividual nos padrões eletrofisiológicos excede ocasionalmente os parâmetros definidos no algoritmo. Esta observação é particularmente relevante no contexto da PEA, onde a heterogeneidade neurobiológica é uma característica predominante da condição (Hirota and King, 2023). A necessidade de intervenção manual, embora introduza um elemento de subjetividade, foi essencial para assegurar a validade das medições e realça a importância da supervisão especializada mesmo em *pipelines* altamente automatizados.

O componente N170 revelou-se como marcador do processamento facial em ambos os grupos diagnósticos e através de todas as condições experimentais. A morfologia típica do componente, com latência entre 140-200 ms e distribuição occipito-temporal, foi observada de forma fiável, embora com variações individuais na amplitude e latência específicas. A presença sistemática deste componente, observável em todas as condições, valida a sua utilização como índice eletrofisiológico do processamento facial, confirmando a adequação do paradigma experimental implementado. A sua presença também demonstra que, independentemente do diagnóstico de PEA ou da administração de arbaclofeno, os mecanismos neuronais fundamentais subjacentes ao processamento facial permanecem suficientemente preservados para gerar esta resposta característica.

A comparação entre presença ou ausência de estímulo facial revelou diferenças na amplitude do N170 visualmente aparentes, com reduções nos ensaios em que os participantes não identificaram faces comparativamente aos ensaios em que identificaram. No entanto, quando estas diferenças foram submetidas a análise através de algoritmos de *machine learning*, os resultados revelaram-se modestos, com os melhores modelos a alcançar apenas 54.3% de exatidão, marginalmente superior ao esperado por acaso. Esta discrepância entre a observação visual de diferenças e a incapacidade dos algoritmos em classificar eficazmente as condições sugere que, embora existam diferenças médias entre grupos, os dados apresentam elevada variabilidade e sobreposição considerável. Esta modulação seletiva da amplitude, não produz *clusters* suficientemente distintos para permitir classificação fiável ao nível individual. No entanto, estudos anteriores demonstraram que a resposta no domínio das frequências poderia ser informativa como biomarcador na PEA, usando a mesma tarefa experimental (Castelhana *et al.*, 2018).

No geral, estes resultados alinham-se com a literatura sobre o N170, que estabelece este componente como um marcador específico e sensível do processamento facial. Estudos prévios demonstraram que o N170 apresenta maior amplitude para faces comparativamente a outros estímulos visuais complexos, e que esta resposta é modulada por fatores como a orientação facial, familiaridade e conteúdo emocional (Neath-Tavares and Itier, 2016; Caharel and Rossion, 2021; Shu *et al.*, 2024).

Os efeitos do arbaclofeno sobre os componentes eletrofisiológicos revelaram padrões complexos e dependentes da dose que diferem entre componentes, hemisférios e condições experimentais. A amplitude do P100 demonstrou modulação significativa pelo arbaclofeno em ambos os hemisférios, com efeitos particularmente robustos no hemisfério direito. Neste hemisfério, observou-se uma diminuição linear progressiva da amplitude com o aumento da dose, atingindo reduções significativas tanto na dose de 15 mg como na de 30 mg. O hemisfério esquerdo apresentou um padrão quadrático, com a dose de 15 mg a produzir o efeito máximo, sem benefício adicional na dose superior. Notavelmente, estes efeitos foram consistentes entre grupos diagnósticos, sugerindo que o arbaclofeno modula mecanismos visuais básicos de forma similar em indivíduos neurotípicos e com PEA. Na condição de não identificação de faces, os efeitos foram ainda mais pronunciados, com padrões lineares bilaterais indicando que a modulação GABAérgica do P100 é independente do processamento de faces.

O componente N170 demonstrou um padrão de resposta distinto e altamente lateralizado. Enquanto o hemisfério direito permaneceu notavelmente resistente aos efeitos do arbaclofeno em todas as condições, o hemisfério esquerdo apresentou modulações significativas, particularmente na condição de não identificação de faces. Nesta condição, observou-se um aumento progressivo da amplitude do N170 com o incremento da dose, atingindo significância apenas na dose de 30 mg. Este padrão sugere que o arbaclofeno pode intensificar o processamento visual não-específico no hemisfério esquerdo. A medida de amplitude pico a pico no hemisfério esquerdo revelou o efeito mais robusto do estudo, com um padrão quadrático altamente significativo e tamanho de efeito elevado. O grupo de controlo demonstrou o característico padrão em U invertido, com amplitudes máximas na dose de 15

mg, enquanto o grupo PEA apresentou uma resposta atenuada, mantendo amplitudes relativamente estáveis através das doses.

A latência do componente N170 no hemisfério esquerdo demonstrou diferenças significativas entre grupos, existindo uma interação dose × grupo significativa, indicando que os grupos responderam de forma diferencial à administração de arbaclofeno. Enquanto o grupo de controlo demonstrou um prolongamento progressivo e linear da latência do N170 com o aumento da dose, o grupo PEA apresentou um padrão não-linear e inconsistente, com latências que permaneceram relativamente estáveis através das condições farmacológicas. Esta divergência foi confirmada pelos testes t emparelhados, que revelaram aumentos significativos da latência apenas no grupo de controlo nas doses mais elevadas, enquanto o grupo PEA não apresentou alterações significativas em nenhuma comparação.

Apesar destas diferenças estatisticamente significativas ao nível de grupo, os algoritmos de *machine learning* apresentaram desempenhos modestos na classificação entre controlos e PEA, com exatidões máximas de 61.1%. Esta discrepância entre significância estatística e capacidade classificatória reforça que as diferenças eletrofisiológicas entre grupos, embora reais, apresentam elevada sobreposição interindividual.

Adicionalmente, apesar dos efeitos estatisticamente significativos e dos padrões dose-resposta claramente identificáveis, os algoritmos de *machine learning* não foram eficazes na classificação das doses administradas. O melhor desempenho foi alcançado pela Regressão Logística com apenas 42.3% de exatidão, valor não muito superior ao esperado por acaso (33.3%). A validação cruzada revelou desempenhos ainda mais modestos, com médias inferiores a 31% para todos os modelos testados. Esta discrepância entre significância estatística e incapacidade classificatória indica que, embora o arbaclofeno produza alterações mensuráveis nos componentes eletrofisiológicos, estas mudanças são caracterizadas por elevada variabilidade interindividual e considerável sobreposição entre doses. A análise PCA confirmou esta interpretação, revelando extensa sobreposição entre as três condições farmacológicas no espaço multidimensional, com distâncias mínimas entre centroides, particularmente entre as doses de 15 mg e 30 mg de arbaclofeno.

Um resultado particularmente intrigante foi encontrado na análise de lateralização hemisférica, revelando padrões que contrastam com parte da literatura sobre processamento facial na PEA. A observação dos ERPs médios e a análise ANOVA revelaram que o grupo PEA manteve amplitudes consistentemente elevadas no hemisfério direito em todas as condições farmacológicas, enquanto demonstrava amplitudes reduzidas no hemisfério esquerdo. Este padrão de lateralização acentuada para o hemisfério direito no grupo PEA contrasta com estudos prévios que frequentemente reportam lateralização reduzida ou padrões de ativação mais difusos nesta população (Kang *et al.*, 2018). A interação dose × hemisfério significativa revelou ainda que o arbaclofeno afetou preferencialmente o hemisfério esquerdo em ambos os grupos, enquanto o hemisfério direito permaneceu notavelmente resistente à modulação farmacológica. A heterogeneidade na distribuição ou funcionalidade dos recetores GABA-B entre hemisférios, evidenciada pela resistência diferencial à modulação farmacológica, tem

implicações cruciais para o desenvolvimento de intervenções terapêuticas direcionadas. Os resultados sugerem que o hemisfério esquerdo, demonstrando maior plasticidade em resposta ao arbaclofeno, pode constituir um alvo terapêutico mais promissor para a modulação do sistema GABAérgico em contextos clínicos.

Os algoritmos de *machine learning* desenvolvidos, revelaram limitações importantes que refletem a natureza dos dados eletrofisiológicos e a complexidade dos fenômenos estudados. A incapacidade de classificar corretamente grupos diagnósticos, doses farmacológicas ou respostas comportamentais com precisão substancialmente superior ao acaso demonstra que os biomarcadores eletrofisiológicos isolados são insuficientes para capturar a variabilidade fenotípica e as respostas individuais ao tratamento. A necessidade de imputação de valores ausentes em parte dos dados de entrada pode, também, ter introduzido ruído artificial nos modelos, comprometendo potencialmente a identificação de padrões. Embora a imputação através de *K-Nearest Neighbors* tenha permitido preservar a completude dos dados, esta estratégia pode distorcer relações sutis entre variáveis, particularmente considerando o tamanho da amostra limitado. Apesar destas limitações, as ferramentas computacionais revelaram-se úteis como método complementar de análise, permitindo identificar padrões de variabilidade e confirmar a ausência de *clusters* claramente definidos.

As implicações clínicas e terapêuticas dos resultados obtidos apontam para um cenário complexo quanto ao potencial do arbaclofeno como intervenção farmacológica na PEA. Os dados não revelaram um efeito normalizador claro do arbaclofeno sobre os marcadores eletrofisiológicos do processamento facial em indivíduos com PEA. Enquanto o grupo de controlo demonstrou respostas dose-dependentes consistentes e previsíveis, o grupo PEA apresentou padrões de resposta atenuados ou inconsistentes em alguns componentes. Esta ausência de modulação significativa sugere que o arbaclofeno, nas doses testadas, pode ter eficácia limitada na normalização dos défices de processamento facial geralmente associados à PEA.

Várias hipóteses podem explicar estes resultados do ponto de vista biológico. A heterogeneidade fenotípica característica da PEA representa provavelmente o fator mais significativo, onde diferentes subgrupos etiológicos podem apresentar perfis distintos de disfunção GABAérgica que respondem diferencialmente ao tratamento. É plausível que existam subtipos de resposta ao arbaclofeno dentro do espectro do autismo, com alguns indivíduos potencialmente a beneficiar da intervenção enquanto outros permanecem refratários ou até apresentam respostas adversas. As doses utilizadas (15 mg e 30 mg) podem também ter sido subótimas para produzir efeitos significativos nesta população, considerando possíveis alterações na farmacocinética ou na sensibilidade dos recetores GABA-B em indivíduos com PEA. A modulação isolada do sistema GABAérgico pode, também, ser insuficiente para corrigir as múltiplas alterações neurobiológicas subjacentes aos défices sociais no autismo.

5.2 Limitações do Estudo

O presente estudo apresenta limitações metodológicas que devem ser consideradas na interpretação dos resultados. Primeiramente, o tamanho da amostra foi afetado pela ausência de dados completos em parte dos participantes, onde nem todos completaram as três visitas experimentais correspondentes às diferentes doses de arbaclofeno. Isto resultou numa redução efetiva da dimensão da amostra, limitando o poder estatístico para detetar efeitos subtis e comprometendo particularmente as análises que requerem maior número de observações, como os algoritmos de *machine learning* e as interações complexas nas ANOVA's de medidas repetidas.

A elevada variabilidade interindividual observada nos marcadores eletrofisiológicos, embora esperada nesta população, constituiu um desafio adicional para a identificação de padrões consistentes e a obtenção de resultados estatisticamente robustos, contribuindo para a sobreposição substancial entre condições experimentais.

A natureza transversal do desenho experimental representa outra limitação significativa, impedindo a avaliação da dinâmica temporal dos efeitos do arbaclofeno. A ausência de medidas longitudinais impossibilita determinar se as modulações eletrofisiológicas observadas são mantidas, potenciadas ou atenuadas com administrações repetidas.

O elevado número de comparações estatísticas realizadas (25 modelos ANOVA independentes e múltiplas análises post hoc) sem aplicação de correções para comparações múltiplas aumenta substancialmente o risco de erro tipo I, tornando provável a ocorrência de falsos positivos. Embora esta abordagem exploratória seja justificável numa fase inicial de investigação farmacológica, implica que os resultados devem ser interpretados com cautela e considerados preliminares até serem replicados em estudos confirmatórios com correções estatísticas apropriadas e amostras maiores.

A restrição da análise aos canais occipito-temporais PO9 e PO10, embora justificada pela literatura sobre o processamento facial, limitou a compreensão abrangente dos efeitos do arbaclofeno sobre a atividade cortical. A análise dos restantes 62 canais poderia revelar modulações em outras regiões cerebrais relevantes para o processamento social ou identificar padrões de conectividade funcional alterados pelo fármaco, melhorando a caracterização dos seus efeitos neurobiológicos.

Finalmente, os participantes incluídos neste estudo tinham todos um coeficiente de inteligência global acima de 70 (sem défice intelectual) e eram todos adultos, o que limita a generalização dos resultados para a população mais alargada com PEA.

5.3 Trabalhos Futuros

Os resultados obtidos neste estudo apontam para várias direções promissoras para investigação futura. O aumento do tamanho e diversidade da amostra constitui uma prioridade,

requerendo estudos multicêntricos com amostras substancialmente maiores e representativas da heterogeneidade do espectro do autismo. A inclusão de participantes com diferentes subtipos etiológicos de PEA, múltiplas faixas etárias e diversos níveis de funcionamento cognitivo permitirá identificar potenciais moderadores da resposta ao arbaclofeno e desenvolver modelos preditivos mais eficazes.

A implementação de desenhos longitudinais será também importante para compreender a dinâmica temporal dos efeitos do arbaclofeno. Estudos futuros deverão avaliar participantes em múltiplos pontos temporais durante regimes terapêuticos prolongados, permitindo caracterizar padrões de resposta inicial, manutenção de efeitos, desenvolvimento de tolerância e potenciais benefícios cumulativos. Esta abordagem longitudinal possibilitará também a identificação de biomarcadores preditivos de resposta sustentada ao tratamento.

A inclusão de grupos clínicos adicionais, particularmente indivíduos com outras perturbações do neurodesenvolvimento ou perturbações da linguagem, permitirá determinar a especificidade dos resultados encontrados para a PEA. Esta abordagem poderá revelar fenótipos eletrofisiológicos comuns que transcendem categorias diagnósticas tradicionais.

A integração de marcadores neurofisiológicos complementares expandirá significativamente a compreensão dos efeitos do arbaclofeno. A análise de componentes como o P300, indexando processos de atenção, poderá fornecer uma caracterização mais abrangente da modulação GABAérgica sobre múltiplos domínios cognitivos relevantes para o fenótipo da PEA (Cui et al., 2017).

O desenvolvimento e aplicação de técnicas analíticas avançadas, incluindo algoritmos de *deep learning* e redes neurais convolucionais, poderá revelar padrões complexos não-lineares nos dados eletrofisiológicos. Estas abordagens computacionais mais avançadas, treinadas em datasets de maior dimensão, poderão identificar assinaturas multivariadas subtis que escapam às análises tradicionais, potencialmente alcançando capacidade classificatória clinicamente útil.

A exploração da lateralização funcional através de paradigmas específicos e análises de conectividade inter-hemisférica poderá elucidar os mecanismos subjacentes às assimetrias observadas. Esta linha de investigação poderá revelar se as diferenças de lateralização representam um biomarcador relevante no contexto da PEA, auxiliando no desenvolvimento de intervenções direcionadas para normalizar padrões de ativação hemisférica.

6 Conclusões

O presente estudo teve como principal objetivo explorar os efeitos do arbaclofeno sobre os componentes eletrofisiológicos N170 e P100, associados ao processamento facial, em indivíduos com e sem PEA, utilizando dados de EEG recolhidos no contexto de uma tarefa de detecção de faces. Foi implementado um pipeline de pré-processamento automatizado e desenvolvidos *scripts* para a extração sistemática dos componentes eletrofisiológicos, complementados por análises estatísticas e algoritmos de *machine learning*.

Os resultados obtidos demonstraram efeitos diferenciais do arbaclofeno sobre os componentes eletrofisiológicos estudados. O componente P100 revelou-se consistentemente modulado pelo arbaclofeno em ambos os grupos diagnósticos, apresentando reduções dose-dependentes da amplitude tanto na condição de processamento facial como não-facial. Esta modulação sugere que o arbaclofeno atua eficazmente sobre os mecanismos de processamento visual precoce, independentemente do diagnóstico ou do tipo de estímulo processado. Em contraste, o componente N170, embora se tenha revelado um marcador consistente do processamento facial, demonstrou padrões de resposta farmacológica mais complexos e lateralizados, com modulação preferencial do hemisfério esquerdo e resistência notável do hemisfério direito.

No grupo PEA, o arbaclofeno não demonstrou efeitos normalizadores consistentes sobre o N170, componente associado ao processamento facial de alto nível. Enquanto o grupo de controlo apresentou respostas dose-dependentes, o grupo PEA exibiu padrões atenuados ou inconsistentes, particularmente nas medidas relacionadas com o processamento social. Esta dissociação entre os efeitos sobre o processamento visual básico (P100) e o processamento facial especializado (N170) sugere que o arbaclofeno pode atuar preferencialmente como modulador dos circuitos sensoriais primários, com impacto limitado sobre os mecanismos neuronais de processamento social mais complexos que se encontram alterados na PEA.

A aplicação de metodologias de classificação baseadas em *machine learning* produziu resultados de desempenho limitado, reforçando a ideia de que os marcadores eletrofisiológicos isolados, como o N170, embora informativos, não são suficientes para discriminar de forma fiável indivíduos com PEA ao nível individual. Contudo, observou-se que a amplitude do N170 apresentou alguma capacidade discriminativa em subgrupos específicos, apontando para a possível utilidade deste marcador na identificação de perfis de resposta diferenciados dentro do espectro autista.

Do ponto de vista metodológico, este trabalho demonstrou a viabilidade da automatização do processamento de dados de EEG em contextos de investigação clínica, permitindo a replicação de análises com maior eficiência e fiabilidade. A implementação de ferramentas computacionais demonstrou benefícios consideráveis para a sistematização e exploração de grandes volumes de dados neurofisiológicos, permitindo a identificação de padrões complexos, mesmo que subtis.

Em síntese, os resultados obtidos revelam que o arbaclofeno exerce efeitos diferenciais sobre os componentes eletrofisiológicos, com modulação consistente do processamento visual básico (P100) mas efeitos limitados e inconsistentes sobre os marcadores de processamento facial (N170) em indivíduos com PEA. Esta especificidade sugere que, embora o arbaclofeno possa ter utilidade na modulação de aspetos sensoriais básicos, a sua eficácia como agente terapêutico para os défices sociais nucleares da PEA permanece incerta. A metodologia aplicada fornece uma base sólida para futuras investigações, que deverão contemplar amostras mais alargadas, delineamentos longitudinais e abordagens multivariadas que incorporem múltiplos domínios neurobiológicos e comportamentais. Este trabalho contribui assim para a crescente compreensão dos mecanismos neuronais subjacentes ao autismo e para o desenvolvimento de intervenções mais direccionadas e baseadas em evidência.

Competências Adquiridas

A realização deste trabalho permitiu o desenvolvimento de um conjunto alargado de competências científicas, técnicas e transversais. No domínio técnico, foi adquirida experiência aprofundada em eletroencefalografia (EEG), nomeadamente no pré-processamento de sinais, deteção de artefactos, interpolação de canais e extração de componentes de potenciais evocados. Paralelamente, foram consolidadas competências na utilização de ferramentas de análise estatística, como a ANOVA de medidas repetidas e testes post hoc, essenciais para a interpretação dos dados neurofisiológicos.

Foram ainda desenvolvidas competências em programação e automação de processos analíticos em MATLAB e Python, permitindo a criação de *pipelines* complexos e reprodutíveis. A aplicação de algoritmos de *machine learning* para classificação e análise de dados multidimensionais contribuiu para o reforço das competências em ciência de dados aplicada à neuroengenharia.

Ao nível científico, este trabalho promoveu o desenvolvimento da capacidade crítica na análise e interpretação de resultados, bem como na revisão e integração da literatura relevante. Finalmente, foram igualmente fortalecidas competências de comunicação científica, redação técnica e gestão de projeto, fundamentais para a futura prática profissional e académica.

Referências

Aishworiya, R. *et al.* (2022) 'An Update on Psychopharmacological Treatment of Autism Spectrum Disorder', *Neurotherapeutics*, 19(1), pp. 248–262. Available at: <https://doi.org/10.1007/s13311-022-01183-1>.

Banker, S.M. *et al.* (2021) 'Hippocampal contributions to social and cognitive deficits in autism spectrum disorder', *Trends in Neurosciences*, 44(10), pp. 793–807. Available at: <https://doi.org/10.1016/j.tins.2021.08.005>.

Berry-Kravis, E. *et al.* (2017) 'Arbaclofen in fragile X syndrome: results of phase 3 trials', *Journal of Neurodevelopmental Disorders*, 9(1), p. 3. Available at: <https://doi.org/10.1186/s11689-016-9181-6>.

Bertinetto, C., Engel, J. and Jansen, J. (2020) 'ANOVA simultaneous component analysis: A tutorial review', *Analytica Chimica Acta: X*, 6, p. 100061. Available at: <https://doi.org/10.1016/j.acax.2020.100061>.

Bhaya, W. and Alasadi, S.A. (2017) 'Review of Data Preprocessing Techniques in Data Mining', *Journal of Engineering and Applied Sciences*, 12, pp. 4102–4107. Available at: <https://doi.org/10.3923/jeasci.2017.4102.4107>.

Bigdely-Shamlo, N. *et al.* (2015) 'The PREP pipeline: standardized preprocessing for large-scale EEG analysis', *Frontiers in Neuroinformatics*, 9. Available at: <https://doi.org/10.3389/fninf.2015.00016>.

Bjørklund, G. and El-Ansary, A. (2024) 'GABAergic Dysregulation in Autism Spectrum Disorder: Collaborative Insights on Neurobiological Complexity', *International Journal for Autism Challenges & Solution*, 1(1), pp. 72–78. Available at: <https://doi.org/10.54878/4v2w1f14>.

Bougeard, C. *et al.* (2021) 'Prevalence of Autism Spectrum Disorder and Co-morbidities in Children and Adolescents: A Systematic Literature Review', *Frontiers in Psychiatry*, 12. Available at: <https://doi.org/10.3389/fpsy.2021.744709>.

Brain Products (2025) 'BrainAmp series', *brainproducts.com* [Preprint]. Gilching, Alemanha. Available at: <https://www.brainproducts.com/solutions/brainamp/> (Accessed: 3 June 2025).

Brainard, D.H. (1997) 'The Psychophysics Toolbox.', *Spatial vision*, 10(4), pp. 433–6.

Buch, A.M. *et al.* (2023) 'Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder', *Nature Neuroscience*, 26(4), pp. 650–663. Available at: <https://doi.org/10.1038/s41593-023-01259-x>.

Caharel, S. and Rossion, B. (2021) 'The N170 is Sensitive to Long-term (Personal) Familiarity of a Face Identity', *Neuroscience*, 458, pp. 244–255. Available at: <https://doi.org/10.1016/j.neuroscience.2020.12.036>.

Canitano, R. and Palumbi, R. (2021) 'Excitation/Inhibition Modulators in Autism Spectrum Disorder: Current Clinical Research', *Frontiers in Neuroscience*, 15. Available at: <https://doi.org/10.3389/fnins.2021.753274>.

Carli, M. *et al.* (2021) 'Atypical Antipsychotics and Metabolic Syndrome: From Molecular Mechanisms to Clinical Differences', *Pharmaceuticals*, 14(3), p. 238. Available at: <https://doi.org/10.3390/ph14030238>.

Castelhano, J. *et al.* (2018) 'Stimulus dependent neural oscillatory patterns show reliable statistical identification of autism spectrum disorder in a face perceptual decision task', *Clinical Neurophysiology*, 129(5), pp. 981–989. Available at: <https://doi.org/10.1016/j.clinph.2018.01.072>.

Çetin, V. and Yıldız, O. (2022) 'A comprehensive review on data preprocessing techniques in data analysis', *Pamukkale University Journal of Engineering Sciences*, 28(2), pp. 299–312. Available at: <https://doi.org/10.5505/pajes.2021.62687>.

Chen, L. *et al.* (2022) 'Homeostatic plasticity and excitation-inhibition balance: The good, the bad, and the ugly', *Current Opinion in Neurobiology*, 75, p. 102553. Available at: <https://doi.org/10.1016/j.conb.2022.102553>.

Cramer, A.O.J. *et al.* (2016) 'Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies', *Psychonomic Bulletin & Review*, 23(2), pp. 640–647. Available at: <https://doi.org/10.3758/s13423-015-0913-5>.

Deb, S., Roy, M. and Limbu, B. (2023) 'Pharmacological management of psychopathology in people with intellectual disabilities and/or autism spectrum disorder', *BJPsyche Advances*, 29(5), pp. 322–333. Available at: <https://doi.org/10.1192/bja.2022.61>.

Delorme, A. (2023) 'EEG is better left alone', *Scientific Reports*, 13(1), p. 2372. Available at: <https://doi.org/10.1038/s41598-023-27528-0>.

Delorme, A. and Makeig, S. (2004) 'EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis', *Journal of Neuroscience Methods*, 134(1), pp. 9–21. Available at: <https://doi.org/10.1016/j.jneumeth.2003.10.009>.

Dollion, N. *et al.* (2022) 'Emotion Facial Processing in Children With Autism Spectrum Disorder: A Pilot Study of the Impact of Service Dogs', *Frontiers in Psychology*, 13. Available at: <https://doi.org/10.3389/fpsyg.2022.869452>.

EasyCap Shop (2025) '64Ch Empty Cap with actiCAP snap holders', *shop.easycap* [Preprint]. Gilching, Alemanha. Available at: <https://shop.easycap.de/products/64ch-empty-cap-with-acticap-snap-holders> (Accessed: 3 June 2025).

Erkan, U. and Thanh, D.N.H. (2020) 'Autism Spectrum Disorder Detection with Machine Learning Methods', *Current Psychiatry Research and Reviews*, 15(4), pp. 297–308. Available at: <https://doi.org/10.2174/2666082215666191111121115>.

Eyring, K.W. and Geschwind, D.H. (2021) 'Three decades of ASD genetics: building a foundation for neurobiological understanding and treatment', *Human Molecular Genetics*, 30(20), pp. R236–R244. Available at: <https://doi.org/10.1093/hmg/ddab176>.

Frye, R. (2014) 'Clinical potential, safety, and tolerability of arbaclofen in the treatment of autism spectrum disorder', *Drug, Healthcare and Patient Safety*, p. 69. Available at: <https://doi.org/10.2147/DHPS.S39595>.

García, S. *et al.* (2016) 'Big data preprocessing: methods and prospects', *Big Data Analytics*, 1(1), p. 9. Available at: <https://doi.org/10.1186/s41044-016-0014-0>.

Guimarães-Souza, E.M. *et al.* (2019) 'Retinal alterations in a pre-clinical model of an autism spectrum disorder', *Molecular Autism*, 10(1), p. 19. Available at: <https://doi.org/10.1186/s13229-019-0270-8>.

Gundersen, B.B. *et al.* (2023) 'Towards Preclinical Validation of Arbaclofen (R-baclofen) Treatment for 16p11.2 Deletion Syndrome'. Available at: <https://doi.org/10.1101/2023.05.01.538987>.

Hirota, T. and King, B.H. (2023) 'Autism Spectrum Disorder', *JAMA*, 329(2), p. 157. Available at: <https://doi.org/10.1001/jama.2022.23661>.

Hodges, H., Fealko, C. and Soares, N. (2020) 'Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation', *Translational Pediatrics*, 9(S1), pp. S55–S65. Available at: <https://doi.org/10.21037/tp.2019.09.09>.

Huang, Q. *et al.* (2022) 'GABA-B receptor modulation of visual sensory processing in adults with and without autism spectrum disorder', *Science Translational Medicine*, 14(626). Available at: <https://doi.org/10.1126/scitranslmed.abg7859>.

Huang, Q. *et al.* (2023a) 'Exploratory evidence for differences in GABAergic regulation of auditory processing in autism spectrum disorder', *Translational Psychiatry*, 13(1), p. 320. Available at: <https://doi.org/10.1038/s41398-023-02619-8>.

Huang, Q. *et al.* (2023b) 'GABAergic regulation of auditory repetition suppression in adults with and without Autism Spectrum Disorder'. Available at: <https://doi.org/10.1101/2023.02.15.23285928>.

Huang, Q. *et al.* (2024) 'Retinal GABAergic Alterations in Adults with Autism Spectrum Disorder', *The Journal of Neuroscience*, 44(14), p. e1218232024. Available at: <https://doi.org/10.1523/JNEUROSCI.1218-23.2024>.

IBM Corp. (2024) 'IBM SPSS Statistics for Windows (Version 29.0)'. Available at: <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-29> (Accessed: 9 June 2025).

Jain, A. *et al.* (2024) 'Analysis of EEG signals and data acquisition methods: a review', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 12(1). Available at: <https://doi.org/10.1080/21681163.2024.2304574>.

Kala, S. *et al.* (2021) 'Brief Report: Preliminary Evidence of the N170 as a Biomarker of Response to Treatment in Autism Spectrum Disorder', *Frontiers in Psychiatry*, 12. Available at: <https://doi.org/10.3389/fpsy.2021.709382>.

Kang, E. *et al.* (2018) 'Atypicality of the N170 Event-Related Potential in Autism Spectrum Disorder: A Meta-analysis', *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(8), pp. 657–666. Available at: <https://doi.org/10.1016/j.bpsc.2017.11.003>.

Kaye, A.D. *et al.* (2024) 'Efficacy and Safety of Alpha-2 Agonists in Autism Spectrum Disorder: A Systematic Review', *Advances in Therapy*, 41(11), pp. 4299–4311. Available at: <https://doi.org/10.1007/s12325-024-02980-0>.

Kim, S.-P. (2018) 'Preprocessing of EEG', in C.-H. Im (ed.) *Computational EEG Analysis: Methods and Applications*. Singapore: Springer Singapore, pp. 15–33. Available at: https://doi.org/10.1007/978-981-13-0908-3_2.

Klem, G.H. *et al.* (1999) 'The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology', *Electroencephalography and Clinical Neurophysiology. Supplement*, 52, pp. 3–6.

Kyriaki, K., Koukopoulos, D. and Fidas, C.A. (2024) 'A Comprehensive Survey of EEG Preprocessing Methods for Cognitive Load Assessment', *IEEE Access*, 12, pp. 23466–23489. Available at: <https://doi.org/10.1109/ACCESS.2024.3360328>.

Lamanna, J. and Meldolesi, J. (2024) 'Autism Spectrum Disorder: Brain Areas Involved, Neurobiological Mechanisms, Diagnoses and Therapies', *International Journal of Molecular Sciences*, 25(4), p. 2423. Available at: <https://doi.org/10.3390/ijms25042423>.

Lebersfeld, J.B. *et al.* (2021) 'Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children', *Journal of Autism and Developmental Disorders*, 51(11), pp. 4101–4114. Available at: <https://doi.org/10.1007/s10803-020-04839-z>.

Li, Y.-A. *et al.* (2022) 'Epidemiology of autism spectrum disorders: Global burden of disease 2019 and bibliometric analysis of risk factors', *Frontiers in Pediatrics*, 10. Available at: <https://doi.org/10.3389/fped.2022.972809>.

Maenner, M.J. *et al.* (2023) 'Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites,

United States, 2020', *MMWR. Surveillance Summaries*, 72(2), pp. 1–14. Available at: <https://doi.org/10.15585/mmwr.ss7202a1>.

Márquez, C. *et al.* (2019) 'Early processing (N170) of infant faces in mothers of children with autism spectrum disorder and its association with maternal sensitivity', *Autism Research*, 12(5), pp. 744–758. Available at: <https://doi.org/10.1002/aur.2102>.

Martin, C., Pepa, L. and Lord, C. (2022) 'DSM-5 diagnosis of autism spectrum disorder', in *Assessment of autism spectrum disorder, 2nd ed.* New York, NY, US: The Guilford Press, pp. 72–95.

Mason, L. *et al.* (2022) 'Stratifying the autistic phenotype using electrophysiological indices of social perception', *Science Translational Medicine*, 14(658). Available at: <https://doi.org/10.1126/scitranslmed.abf8987>.

Mishra, P. *et al.* (2020) 'New data preprocessing trends based on ensemble of multiple preprocessing techniques', *TrAC Trends in Analytical Chemistry*, 132, p. 116045. Available at: <https://doi.org/10.1016/j.trac.2020.116045>.

Montanari, M. *et al.* (2022) 'Autism Spectrum Disorder: Focus on Glutamatergic Neurotransmission', *International Journal of Molecular Sciences*, 23(7), p. 3861. Available at: <https://doi.org/10.3390/ijms23073861>.

Nanjappa, M.S. *et al.* (2022) 'Use of selective serotonin and norepinephrine reuptake inhibitors (SNRIs) in the treatment of autism spectrum disorder (ASD), comorbid psychiatric disorders and ASD-associated symptoms: a clinical review', *CNS Spectrums*, 27(3), pp. 290–297. Available at: <https://doi.org/10.1017/S109285292000214X>.

Neath-Tavares, K.N. and Itier, R.J. (2016) 'Neural processing of fearful and happy facial expressions during emotion-relevant and emotion-irrelevant tasks: A fixation-to-feature approach', *Biological Psychology*, 119, pp. 122–140. Available at: <https://doi.org/10.1016/j.biopsycho.2016.07.013>.

Nguyen, V. *et al.* (2017) 'Alpha-2 Agonists', *Anesthesiology Clinics*, 35(2), pp. 233–245. Available at: <https://doi.org/10.1016/j.anclin.2017.01.009>.

Nisar, S. *et al.* (2022) 'Genetics of glutamate and its receptors in autism spectrum disorder', *Molecular Psychiatry*, 27(5), pp. 2380–2392. Available at: <https://doi.org/10.1038/s41380-022-01506-w>.

Panda, N.R. (2022) 'A Review on Logistic Regression in Medical Research', *National Journal of Community Medicine*, 13(4), pp. 265–270. Available at: <https://doi.org/10.55489/njcm.134202222>.

Parker, T.C. *et al.* (2021) 'The N170 event-related potential reflects delayed neural response to faces when visual attention is directed to the eyes in youths with ASD', *Autism Research*, 14(7), pp. 1347–1356. Available at: <https://doi.org/10.1002/aur.2505>.

Parrella, N.-F. *et al.* (2024) 'Inhibitory dysfunction and social processing difficulties in autism: A comprehensive narrative review', *Journal of Psychiatric Research*, 169, pp. 113–125. Available at: <https://doi.org/10.1016/j.jpsychires.2023.11.014>.

Pedroni, A., Bahreini, A. and Langer, N. (2019) 'Automagic: Standardized preprocessing of big EEG data', *NeuroImage*, 200, pp. 460–473. Available at: <https://doi.org/10.1016/j.neuroimage.2019.06.046>.

Peng, W. (2019) 'EEG Preprocessing and Denoising', in *EEG Signal Processing and Feature Extraction*. Singapore: Springer Singapore, pp. 71–87. Available at: https://doi.org/10.1007/978-981-13-91113-2_5.

Posamentier, M.T. and Abdi, H. (2003) 'Processing Faces and Facial Expressions', *Neuropsychology Review*, 13(3), pp. 113–143. Available at: <https://doi.org/10.1023/A:1025519712569>.

Qin, L. *et al.* (2024) 'New advances in the diagnosis and treatment of autism spectrum disorders', *European Journal of Medical Research*, 29(1), p. 322. Available at: <https://doi.org/10.1186/s40001-024-01916-2>.

Rasul, R.A. *et al.* (2024) 'An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder', *Healthcare Analytics*, 5, p. 100293. Available at: <https://doi.org/10.1016/j.health.2023.100293>.

Roberts, B.R.T. and Trossman, R. (2024) 'Face processing in ADHD: A review of the N170 event-related potential.', *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, 78(1), pp. 36–49. Available at: <https://doi.org/10.1037/cep0000321>.

Roberts, T.P.L. *et al.* (2019) 'A MEG Study of Acute Arbaclofen (STX-209) Administration', *Frontiers in Integrative Neuroscience*, 13. Available at: <https://doi.org/10.3389/fnint.2019.00069>.

Ruzich, E. *et al.* (2015) 'Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females', *Molecular Autism*, 6(1), p. 2. Available at: <https://doi.org/10.1186/2040-2392-6-2>.

Salari, N. *et al.* (2022) 'The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis', *Italian Journal of Pediatrics*, 48(1), p. 112. Available at: <https://doi.org/10.1186/s13052-022-01310-w>.

Sanchez-Ponce, R. *et al.* (2012) 'Metabolic and Pharmacokinetic Differentiation of STX209 and Racemic Baclofen in Humans', *Metabolites*, 2(3), pp. 596–613. Available at: <https://doi.org/10.3390/metabo2030596>.

Sauer, A.K. *et al.* (2021) 'Autism Spectrum Disorders: Etiology and Pathology', in *Autism Spectrum Disorders*. Exon Publications, pp. 1–16. Available at: <https://doi.org/10.36255/exonpublications.autismspectrumdisorders.2021.etiology>.

Schlosser, T. *et al.* (2024) *A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision*. Available at: <https://doi.org/10.13140/RG.2.2.14331.69928>.

Shan, J. *et al.* (2023) 'A scoping review of physiological biomarkers in autism', *Frontiers in Neuroscience*, 17. Available at: <https://doi.org/10.3389/fnins.2023.1269880>.

Sharma, A. and Tanwar, P. (2020) 'Deep Analysis of Autism Spectrum Disorder Detection Techniques', in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE, pp. 455–459. Available at: <https://doi.org/10.1109/ICIEM48762.2020.9160123>.

Shoka, A. *et al.* (2019) 'Literature Review on EEG Preprocessing, Feature Extraction, and Classifications Techniques', *Menoufia Journal of Electronic Engineering Research*, 28(1), pp. 292–299. Available at: <https://doi.org/10.21608/mjeer.2019.64927>.

Shu, D. *et al.* (2024) 'Intervention Effect of Group Sensory Integration Training on Social Responsiveness and N170 Event-Related Potential of Children with Autism', *Behavioral Sciences*, 14(3), p. 202. Available at: <https://doi.org/10.3390/bs14030202>.

Siuly, S., Li, Y. and Zhang, Y. (2016) 'Electroencephalogram (EEG) and Its Background', in S. Siuly, Y. Li, and Y. Zhang (eds) *EEG Signal Analysis and Classification: Techniques and Applications*. Cham: Springer International Publishing, pp. 3–21. Available at: https://doi.org/10.1007/978-3-319-47653-7_1.

Solmi, M. *et al.* (2022) 'Incidence, prevalence, and global burden of autism spectrum disorder from 1990 to 2019 across 204 countries', *Molecular Psychiatry*, 27(10), pp. 4172–4180. Available at: <https://doi.org/10.1038/s41380-022-01630-7>.

Sousa, D. *et al.* (2023) 'A neurophysiological signature of dynamic emotion recognition associated with social communication skills and cortical gamma-aminobutyric acid levels in children', *Frontiers in Neuroscience*, 17. Available at: <https://doi.org/10.3389/fnins.2023.1295608>.

Suveetha Dhanaselvam, P. and Nadia Chellam, C. (2023) 'A Review on Preprocessing of EEG Signal', in *2023 International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*. IEEE, pp. 1–7. Available at: <https://doi.org/10.1109/ICBSII58188.2023.10181071>.

Taniya, M.A. *et al.* (2022) 'Role of Gut Microbiome in Autism Spectrum Disorder and Its Therapeutic Regulation', *Frontiers in Cellular and Infection Microbiology*, 12. Available at: <https://doi.org/10.3389/fcimb.2022.915701>.

Tautvydaitė, D. *et al.* (2022) 'Effect of perceived eye gaze on the N170 component – A systematic review', *Neuroscience & Biobehavioral Reviews*, 143, p. 104913. Available at: <https://doi.org/10.1016/j.neubiorev.2022.104913>.

Tavares, P.P. *et al.* (2016) 'Preserved face inversion effects in adults with autism spectrum disorder', *NeuroReport*, 27(8), pp. 587–592. Available at: <https://doi.org/10.1097/WNR.0000000000000576>.

The MathWorks Inc. (2017) 'MATLAB version: 9.2.0 (R2017a)'. Natick, Massachusetts, United States: The MathWorks Inc. Available at: <https://www.mathworks.com> (Accessed: 30 June 2025).

The MathWorks Inc. (2020) 'MATLAB version: 9.8.0 (R2020a)'. Natick, Massachusetts, United States: The MathWorks Inc. Available at: <https://www.mathworks.com>.

Tian, J., Gao, X. and Yang, L. (2022) 'Repetitive Restricted Behaviors in Autism Spectrum Disorder: From Mechanism to Development of Therapeutics', *Frontiers in Neuroscience*, 16. Available at: <https://doi.org/10.3389/fnins.2022.780407>.

U.S. Centers for Disease Control and Prevention (2023) *Data and Statistics on Autism Spectrum Disorder*. Available at: <https://www.cdc.gov/autism/data-research/index.html> (Accessed: 19 March 2025).

Veenstra-VanderWeele, J. *et al.* (2017) 'Arbaclofen in Children and Adolescents with Autism Spectrum Disorder: A Randomized, Controlled, Phase 2 Trial', *Neuropsychopharmacology*, 42(7), pp. 1390–1398. Available at: <https://doi.org/10.1038/npp.2016.237>.

Vlamings, P.H.J.M. *et al.* (2010) 'Basic Abnormalities in Visual Processing Affect Face Processing at an Early Age in Autism Spectrum Disorder', *Biological Psychiatry*, 68(12), pp. 1107–1113. Available at: <https://doi.org/10.1016/j.biopsych.2010.06.024>.

Wang, L. *et al.* (2023) 'Autism Spectrum Disorder: Neurodevelopmental Risk Factors, Biological Mechanism, and Precision Therapy', *International Journal of Molecular Sciences*, 24(3), p. 1819. Available at: <https://doi.org/10.3390/ijms24031819>.

Wechsler, D. (2011) 'Wechsler Abbreviated Scale of Intelligence--Second Edition (WASI-II)', *PsycTESTS Dataset* [Preprint]. Available at: <https://doi.org/10.1037/t15171-000>.

World Health Organization (2023) *Autism*. Available at: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders> (Accessed: 11 March 2025).

Zhao, H. *et al.* (2022) 'GABAergic System Dysfunction in Autism Spectrum Disorders', *Frontiers in Cell and Developmental Biology*, 9. Available at: <https://doi.org/10.3389/fcell.2021.781327>.

Zwaigenbaum, L. and Penner, M. (2018) 'Autism spectrum disorder: advances in diagnosis and evaluation', *BMJ*, p. k1674. Available at: <https://doi.org/10.1136/bmj.k1674>.

Anexos

Pré-processamento

O seguinte *script* permitiu fazer o pré-processamento automático de todos os datasets em estudo.

```
% Define o diretório principal
mainDir = pwd;

% Lista todas as pastas de sujeitos
subjectDirs = dir(mainDir);

% Filtra apenas diretórios, excluindo '.' e '..'
subjectDirs = subjectDirs([subjectDirs.isdir] & ~ismember({subjectDirs.name},
{'.' , '..'}));

% Itera sobre cada pasta de sujeito
for subj = 1:length(subjectDirs)
    subjPath = fullfile(mainDir, subjectDirs(subj).name);
    disp(['Processando sujeito: ', subjectDirs(subj).name]);

    % Lista todas as pastas de sessões (v1, v2, v3)
    sessionDirs = dir(fullfile(subjPath, 'v*'));

    % Filtra apenas diretórios
    sessionDirs = sessionDirs([sessionDirs.isdir]);

    % Itera sobre cada pasta de sessão
    for sess = 1:length(sessionDirs)
        sessPath = fullfile(subjPath, sessionDirs(sess).name);
        disp(['Processando sessão: ', sessionDirs(sess).name]);

        % Lista todos os arquivos .vhdr na pasta de sessão
        files = dir(fullfile(sessPath, '*.vhdr'));

        % Verifica se há exatamente dois arquivos
        if length(files) ~= 2
            warning(['A pasta ', sessPath, ' não contém exatamente dois
arquivos .vhdr.']);
            continue;
        end

        % Carrega o primeiro conjunto de dados
        EEG1 = pop_loadbv(sessPath, files(1).name);
        EEG1.setname = files(1).name;

        % Carrega o segundo conjunto de dados
        EEG2 = pop_loadbv(sessPath, files(2).name);
        EEG2.setname = files(2).name;

        % Armazena os conjuntos de dados em ALLEEG
        ALLEEG = [];
        [ALLEEG, EEG1, CURRENTSET] = eeg_store(ALLEEG, EEG1, 0);
        [ALLEEG, EEG2, CURRENTSET] = eeg_store(ALLEEG, EEG2, 0);
    end
end
```

```

% Funde os dois conjuntos de dados
EEG = pop_mergeset(ALLEEG, [1 2], 0);

% Define o nome do conjunto de dados fundido
EEG.setname = 'Merged Data';

% Reamostra os dados para 500 Hz
EEG = pop_resample(EEG, 500);

% Aplica filtro passa-alto e passa-baixo
EEG = pop_eegfiltnew(EEG, 'locutoff', 0.1, 'plotfreqz', 0);
EEG = pop_eegfiltnew(EEG, 'hicutoff', 30, 'plotfreqz', 0);

% Armazena o conjunto de dados original em ALLEEG
[ALLEEG, EEG, CURRENTSET] = eeg_store(ALLEEG, EEG, 0);

% Rejeita canais com base na kurtosis sem exibir gráficos
EEG = pop_rejchan(EEG, 'elec', [1:64], 'threshold', 5, 'norm', 'on',
'measure', 'kurt');
EEG.setname = 'Channel Rejected';

% Interpola canais rejeitados usando o conjunto de dados original
EEG = pop_interp(EEG, ALLEEG(1).chanlocs, 'spherical');
EEG.setname = 'Channel Interpolated';

% Re-referencia os dados, excluindo alguns canais
EEG = pop_reref(EEG, [], 'exclude', [17 22 33]);

% Extrai épocas de -1 a 2 segundos
EEG = pop_epoch(EEG, {'T 11', 'T 12'}, [-1 2], 'newname', 'Epoch
Data1', 'epochinfo', 'yes');

% Identifica as épocas com 'T105'
epochsWithT105 = [];
for i = 1:length(EEG.epoch)
    eventTypes = EEG.epoch(i).eventtype;
    for j = 1:length(eventTypes)
        if strcmp(strtrim(eventTypes{j}), 'T105')
            epochsWithT105 = [epochsWithT105, i];
            break;
        end
    end
end

% Cria um dataset com épocas 'T105'
EEG_no_face = pop_select(EEG, 'trial', epochsWithT105);

% Cria um dataset sem épocas 'T105'
EEG_face = pop_select(EEG, 'notrial', epochsWithT105);

% Aplica o pré-processamento a ambos os datasets
datasets = {EEG_face, EEG_no_face};
names = {'face', 'no_face'};

for d = 1:2
    % Segmenta os dados em épocas novamente
    datasets{d} = pop_epoch(datasets{d}, {'T 11', 'T 12'}, [-0.1 0.5],
'newname', 'Epoch Data', 'epochinfo', 'yes');
end

```

```

        % Remove a linha de base
        datasets{d} = pop_rmbase(datasets{d}, [-100 0], []);

        % Rejeita automaticamente épocas com artefactos
        datasets{d} = pop_autorej(datasets{d}, 'nogui', 'on', 'eegplot',
'off');

        % Plota mapas topográficos
        figure; pop_plottopo(datasets{d}, 32, 'Epoch Data', 0, 'ydir',
1);
        %figure; pop_plottopo(datasets{d}, [1:64], 'Epoch Data', 0,
'ydir', 1);
        end

        % Extrai a parte comum dos nomes dos arquivos
        commonName = extractBefore(files(1).name, '_r1');

        % Guarda os conjuntos de dados processados
        EEG_face = datasets{1};
        EEG_no_face = datasets{2};

        EEG_face = pop_saveset(EEG_face, 'filename', [commonName
'_face.set'], 'filepath', sessPath);
        EEG_no_face = pop_saveset(EEG_no_face, 'filename', [commonName
'_no_face.set'], 'filepath', sessPath);

        disp(['Concluído: ', sessPath]);
    end
end

disp('Processamento completo para todos os sujeitos e sessões.');
```

Código 1 - *Script* em MATLAB de pré-processamento dos ficheiros.

Identificação de Picos e Organização de Dados

O seguinte *script* organiza automaticamente todos os datasets processados (ficheiros .set) nas pastas *CTRL_results* e *ASD_results*, posteriormente nas pastas *face* e *no face* tendo em conta as suas designações.

```

% Define o diretório principal
mainDir = pwd;

% Define o diretório de resultados
resultsDir = fullfile(mainDir, 'CTRL results');

% Cria as pastas de resultados, se não existirem
if ~exist(resultsDir, 'dir')
    mkdir(resultsDir);
end
faceDir = fullfile(resultsDir, 'face');
if ~exist(faceDir, 'dir')
    mkdir(faceDir);
end
```

```

noFaceDir = fullfile(resultsDir, 'no face');
if ~exist(noFaceDir, 'dir')
    mkdir(noFaceDir);
end

% Lista todas as pastas de sujeitos
subjectDirs = dir(mainDir);

% Filtra apenas diretórios, excluindo '.' e '..'
subjectDirs = subjectDirs([subjectDirs.isdir] & ~ismember({subjectDirs.name},
{'.', '..'}));

% Itera sobre cada pasta de sujeito
for subj = 1:length(subjectDirs)
    subjPath = fullfile(mainDir, subjectDirs(subj).name);

    % Lista todas as pastas de sessões (v1, v2, v3)
    sessionDirs = dir(fullfile(subjPath, 'v*'));

    % Filtra apenas diretórios
    sessionDirs = sessionDirs([sessionDirs.isdir]);

    % Itera sobre cada pasta de sessão
    for sess = 1:length(sessionDirs)
        sessPath = fullfile(subjPath, sessionDirs(sess).name);

        % Lista todos os arquivos .set na pasta de sessão
        setFiles = dir(fullfile(sessPath, '*.set'));

        % Itera sobre cada arquivo .set
        for file = 1:length(setFiles)
            srcFile = fullfile(sessPath, setFiles(file).name);

            % Determina o destino com base no nome do arquivo
            if endsWith(setFiles(file).name, '_no_face.set')
                destFile = fullfile(noFaceDir, setFiles(file).name);
            elseif endsWith(setFiles(file).name, '_face.set')
                destFile = fullfile(faceDir, setFiles(file).name);
            else
                warning(['Arquivo não reconhecido: ', setFiles(file).name]);
                continue;
            end

            % Copia o arquivo para o destino apropriado
            copyfile(srcFile, destFile);
            disp(['Copiado: ', srcFile, ' para ', destFile]);
        end
    end
end

disp('Cópia completa de todos os arquivos .set.');
```

Código 2 - *Script* em MATLAB de organização de resultados.

O seguinte *script* identifica os picos dos componentes P100 e N170 através de janelas temporais específica, calcula as suas amplitudes e latências e exporta os resultados em formato CSV.

```

% Define sessões
sess_names = {'v1', 'v2', 'v3'};
```

```

col_labels = {'lat P100 (ms)', 'amp P100 (uV)', 'lat N170 (ms)', 'amp N170
(uV)'};

% Listar os ficheiros .set na pasta atual
lista = dir('*.set');
canal = 28;
dados = struct;

for f = 1:length(lista)
    filename = lista(f).name;
    fprintf('A processar: %s\n', filename);

    % Extrair sujeito (até ao primeiro _) e sessão (logo a seguir ao sujeito)
    tokens = regexp(filename, '([^\_]+)_([^\_]+)', 'tokens');
    if isempty(tokens)
        warning('Nome de ficheiro inesperado: %s', filename);
        continue;
    end
    sujeito = tokens{1}{1};
    sujeito = matlab.lang.makeValidName(sujeito); %converte "R-M021" em
"R_M021"
    sessao = tokens{1}{2}; % Ex: v1, v2, v3

    EEG = pop_loadset('filename', filename);

    if ndims(EEG.data) ~= 3
        warning('Dataset %s não está epocado! Ignorado.\n', filename);
        continue
    end

    erp = mean(squeeze(EEG.data(canal, :, :)), 2);
    times = EEG.times;

    % --- N170 ---
    janela_n170 = [130 240];
    [~, idx_ini_n170] = min(abs(times - janela_n170(1)));
    [~, idx_fim_n170] = min(abs(times - janela_n170(2)));
    [amp_n170, idx_pico_n170] = min(erp(idx_ini_n170:idx_fim_n170));
    idx_n170_global = idx_ini_n170 + idx_pico_n170 - 1;
    lat_n170 = times(idx_n170_global);

    % --- P100 ---
    janela_p100 = [80 200];
    [~, idx_ini_p100] = min(abs(times - janela_p100(1)));
    [~, idx_fim_p100] = min(abs(times - janela_p100(2)));
    [amp_p100, idx_pico_p100] = max(erp(idx_ini_p100:idx_fim_p100));
    idx_p100_global = idx_ini_p100 + idx_pico_p100 - 1;
    lat_p100 = times(idx_p100_global);

    % Guardar valores na estrutura
    dados.(sujeito).(sessao).lat_p100 = lat_p100;
    dados.(sujeito).(sessao).amp_p100 = amp_p100;
    dados.(sujeito).(sessao).lat_n170 = lat_n170;
    dados.(sujeito).(sessao).amp_n170 = amp_n170;
end

% Gerar lista única de sujeitos
sujeitos = fieldnames(dados);

% Preparação dos cabeçalhos

```

```

nSesoes = numel(sess_names);
nCol = numel(col_labels)*nSesoes + 1; % +1 para sujeito

% Linha 1 do header (nome das sessões)
header1 = cell(1,nCol);
header1{1} = '';
for i=1:nSesoes
    for j=1: numel(col_labels)
        header1{1+(i-1)*numel(col_labels)+j} = sess_names{i};
    end
end

% Linha 2 do header (lat/amp por sessão)
header2 = cell(1,nCol);
header2{1} = 'Sujeito';
for i=1:nSesoes
    for j=1: numel(col_labels)
        header2{1+(i-1)*numel(col_labels)+j} = col_labels{j};
    end
end

% Preencher tabela de dados
data = cell(numel(sujeitos), nCol);
for i=1: numel(sujeitos)
    data{i,1} = sujeitos{i};
    for s=1:nSesoes
        sess = sess_names{s};
        colBase = (s-1)*numel(col_labels)+2;
        if isfield(dados.(sujeitos{i}), sess)
            d = dados.(sujeitos{i}).(sess);
            data{i,colBase} = d.lat_p100;
            data{i,colBase+1} = d.amp_p100;
            data{i,colBase+2} = d.lat_n170;
            data{i,colBase+3} = d.amp_n170;
        else
            data{i,colBase} = '-';
            data{i,colBase+1} = '-';
            data{i,colBase+2} = '-';
            data{i,colBase+3} = '-';
        end
    end
end

% Exportar CSV de duas linhas de cabeçalho
fid = fopen('resultados_ASD_face_28.csv','w');
% Escrever cabeçalho 1
fprintf(fid, '%s', header1{1});
for h = 2: numel(header1)
    fprintf(fid, ',%s', header1{h});
end
fprintf(fid, '\n');
% Escrever cabeçalho 2
fprintf(fid, '%s', header2{1});
for h = 2: numel(header2)
    fprintf(fid, ',%s', header2{h});
end
fprintf(fid, '\n');
% Escrever dados
for i=1: size(data,1)

```

```

    for j=1:size(data,2)
        if isnumeric(data{i,j})
            fprintf(fid, '%g', data{i,j});
        else
            fprintf(fid, '%s', string(data{i,j}));
        end
        if j<size(data,2)
            fprintf(fid, ',');
        end
    end
    fprintf(fid, '\n');
end
fclose(fid);

disp('CSV exportado!');

```

Código 3 - *Script* em MATLAB de identificação de picos e exportação para ficheiro CSV.

O seguinte *script* converte os ficheiros no formato CSV com os dados pretendidos, para ficheiros no formato *.xlsx* (Excel).

```

import pandas as pd
import numpy as np

csv_file = 'resultados_ASD_face_28.csv'
excel_file = 'resultados_ASD_face_28.xlsx'

# Lê CSV como string para poder substituir '-'
df = pd.read_csv(csv_file, header=[0,1], dtype=str)

# Substituir '-' por NaN
df = df.replace('-', np.nan)

# Tentar converter para número sempre que possível (excepto a coluna 'Sujeito')
for col in df.columns:
    if 'Sujeito' not in col:
        df[col] = pd.to_numeric(df[col], errors='coerce')

# Corrigir nome da coluna do sujeito para multinível correto
columns = df.columns.tolist()
new_columns = []
for col in columns:
    if 'Sujeito' in col:
        new_columns.append((' ', 'Sujeito'))
    else:
        new_columns.append(col)
df.columns = pd.MultiIndex.from_tuples(new_columns)

# Exporta para Excel
df.to_excel(excel_file, index=True)
print(f'Exportado para {excel_file}!')

```

Código 4 - *Script* em Python para conversão de ficheiros *.csv* em ficheiros *.xlsx*.

O seguinte *script* calcula as métricas amplitude pico a pico e latência interpicos dos componentes P100 e N170 através dos seus valores de amplitude e latência. O *script* origina um novo ficheiro Excel apenas com as métricas calculadas de cada sujeito.

```

import pandas as pd
import numpy as np
import re

# Carregar o arquivo Excel com os dados
input_file = 'resultados.xlsx'
output_file = 'resultados_pico_a_pico_e_interpicos.xlsx'

# Ler os dados
df = pd.read_excel(input_file)

# Criar DataFrame para resultados apenas com ID do sujeito
result_df = pd.DataFrame()
result_df[df.columns[0]] = df[df.columns[0]] # Coluna de ID do sujeito

# Adicionar coluna de Grupo se existir
if 'Grupo' in df.columns:
    result_df['Grupo'] = df['Grupo']

print("=== PROCESSANDO AMPLITUDE PICO A PICO ===")

# Identificar colunas de amplitude
amp_p100_cols = [col for col in df.columns if 'amp' in col.lower() and 'p100'
in col.lower()]
amp_n170_cols = [col for col in df.columns if 'amp' in col.lower() and 'n170'
in col.lower()]

print(f"Encontradas {len(amp_p100_cols)} colunas de amplitude P100")
print(f"Encontradas {len(amp_n170_cols)} colunas de amplitude N170")

# Verificar se temos o mesmo número de colunas para ambos os componentes
if len(amp_p100_cols) != len(amp_n170_cols):
    print("AVISO: Número diferente de colunas de amplitude para P100 e N170!")

# Criar colunas para os valores pico a pico (amplitude)
for i in range(min(len(amp_p100_cols), len(amp_n170_cols))):
    # Obter nomes das colunas
    col_p100 = amp_p100_cols[i]
    col_n170 = amp_n170_cols[i]

    # Extrair contexto (hemisfério, dose, condição)
    context_match = re.search(r'(direito|esquerdo)\s+(placebo|15mg|30mg).*(face|no face)',
col_p100, re.IGNORECASE)
    if context_match:
        context = context_match.group(0)

        # Calcular amplitude pico a pico (valor absoluto da diferença)
        result_df[f'{context} pico a pico (uV)'] = np.abs(df[col_p100] -
df[col_n170])

        print(f"Calculado: {context} pico a pico (uV)")
    else:
        # Tentativa alternativa para extrair contexto
        # Remover "amp" e "P100" para obter o contexto
        context = col_p100.replace('amp', '').replace('P100',
'').replace('(uV)', '').strip()
        new_col = f"{context} pico a pico (uV)"

```

```

        result_df[new_col] = np.abs(df[col_p100] - df[col_n170])
        print(f"Calculado: {new_col} (usando método alternativo)")

print("\n=== PROCESSANDO LATÊNCIA INTERPICOS ===")

# Identificar colunas de latência
lat_p100_cols = [col for col in df.columns if 'lat' in col.lower() and 'p100'
in col.lower()]
lat_n170_cols = [col for col in df.columns if 'lat' in col.lower() and 'n170'
in col.lower()]

print(f"Encontradas {len(lat_p100_cols)} colunas de latência P100")
print(f"Encontradas {len(lat_n170_cols)} colunas de latência N170")

# Verificar se temos o mesmo número de colunas para ambos os componentes
if len(lat_p100_cols) != len(lat_n170_cols):
    print("AVISO: Número diferente de colunas de latência para P100 e N170!")

# Criar colunas para os valores de latência interpicos
for i in range(min(len(lat_p100_cols), len(lat_n170_cols))):
    # Obter nomes das colunas
    col_p100 = lat_p100_cols[i]
    col_n170 = lat_n170_cols[i]

    # Extrair contexto (hemisfério, dose, condição)
    context_match = re.search(r'(direito|esquerdo)\s+(placebo|15mg|30mg).*(face|no face)',
col_p100, re.IGNORECASE)
    if context_match:
        context = context_match.group(0)

        # Calcular latência interpicos (N170 - P100)
        result_df[f'{context} latência interpicos (ms)'] = df[col_n170] -
df[col_p100]

        print(f"Calculado: {context} latência interpicos (ms)")
    else:
        # Tentativa alternativa para extrair contexto
        context = col_p100.replace('lat', '').replace('P100',
'').replace('(ms)', '').strip()
        new_col = f"{context} latência interpicos (ms)"

        result_df[new_col] = df[col_n170] - df[col_p100]
        print(f"Calculado: {new_col} (usando método alternativo)")

# Guardar resultados finais
result_df.to_excel(output_file, index=False)
print(f"\n=== PROCESSAMENTO CONCLUÍDO ===")
print(f"Resultados salvos em: {output_file}")
print(f"Total de colunas no arquivo final: {len(result_df.columns)}")
print(f"- ID do sujeito: 1")
if 'Grupo' in result_df.columns:
    print(f"- Grupo: 1")
print(f"- Amplitude pico a pico: {len([col for col in result_df.columns if
'pico a pico' in col])}")
print(f"- Latência interpicos: {len([col for col in result_df.columns if
'latência interpicos' in col])}")

```

Código 5 - Script em Python para o cálculo dos valores de amplitude pico a pico e latência interpicos.

O seguinte *script* converte o Excel dos dados obtidos, de um formato *wide* para um formato *long*. O *script* dá origem a um novo ficheiro Excel com o formato pretendido.

```
import pandas as pd

# Carregar o arquivo Excel
caminho_arquivo = "resultados.xlsx"
df_wide = pd.read_excel(caminho_arquivo)

# Passo 1: Identificar colunas de medição vs. colunas demográficas
cols_medidas = []
cols_demograficas = []
cols_pico_a_pico = [] # Lista para colunas de pico a pico
cols_latencia_interpicos = [] # Nova lista para colunas de latência
interpicos

for col in df_wide.columns:
    col_lower = str(col).lower()
    # Verificar se é coluna de latência interpicos
    if "latência interpicos" in col_lower or "latencia interpicos" in
col_lower:
        cols_latencia_interpicos.append(col)
    # Verificar se é coluna de pico a pico
    elif "pico a pico" in col_lower:
        cols_pico_a_pico.append(col)
    # Verificar se é uma coluna de medição de EEG regular
    elif any(term in col_lower for term in ["direito", "esquerdo", "face",
"placebo", "15mg", "30mg", "p100", "n170", "amp", "lat"]):
        cols_medidas.append(col)
    # Verificar se é uma coluna demográfica/metadata
    elif col != df_wide.columns[0]: # Não incluir a coluna de ID do sujeito
        cols_demograficas.append(col)

print(f"Encontradas {len(cols_medidas)} colunas de medidas regulares")
print(f"Encontradas {len(cols_pico_a_pico)} colunas de pico a pico")
print(f"Encontradas {len(cols_latencia_interpicos)} colunas de latência
interpicos")
print(f"Encontradas {len(cols_demograficas)} colunas demográficas")

# Passo 2: Extrair a coluna de ID do sujeito
id_col = df_wide.columns[0]

# Passo 3: Criar um DataFrame apenas com os dados demográficos
demo_df = df_wide[[id_col] + cols_demograficas].copy()

# Função para extrair informações dos cabeçalhos
def parse_column_info(col_name, tipo_especial=None):
    col_str = str(col_name).lower()

    # Determinar hemisfério
    if "direito" in col_str:
        hemisferio = "Direito"
    elif "esquerdo" in col_str:
        hemisferio = "Esquerdo"
    else:
        hemisferio = "Desconhecido"

    # Determinar dose
    if "placebo" in col_str:
```

```

        dose = "Placebo"
    elif "15mg" in col_str or "15 mg" in col_str:
        dose = "15 mg Arbaclofeno"
    elif "30mg" in col_str or "30 mg" in col_str:
        dose = "30 mg Arbaclofeno"
    else:
        dose = "Desconhecido"

    # Determinar condição
    if " no face " in f" {col_str} " or " no_face " in f" {col_str} ":
        condicao = "no_face"
    elif " face " in f" {col_str} " and " no face " not in f" {col_str} "
and " no_face " not in f" {col_str} ":
        condicao = "face"
    elif "noface" in col_str:
        condicao = "no_face"
    elif any(pattern in col_str for pattern in ["placebo face", "15mg face",
"30mg face", "15 mg face", "30 mg face"]):
        condicao = "face"
    elif any(pattern in col_str for pattern in ["placebo no", "15mg no",
"30mg no", "15 mg no", "30 mg no"]):
        condicao = "no_face"
    else:
        parts = col_str.split()
        if "face" in parts:
            idx_face = parts.index("face")
            if idx_face > 0 and parts[idx_face-1] == "no":
                condicao = "no_face"
            else:
                condicao = "face"
        else:
            condicao = "Desconhecido"

    # Para colunas pico a pico e latência interpicos
    if tipo_especial == 'pico_a_pico':
        componente = "-" # Sem componente específico
        medida = "Pico a Pico"
        unidade = "uV" if "(uv)" in col_str or "(μv)" in col_str else "ms"
    elif tipo_especial == 'latencia_interpicos':
        componente = "-" # Sem componente específico
        medida = "Latência Interpicos"
        unidade = "ms" # Latência sempre em ms
    else:
        # Determinar componente para colunas regulares
        if "p100" in col_str:
            componente = "P100"
        elif "n170" in col_str:
            componente = "N170"
        else:
            componente = "Desconhecido"

    # Determinar tipo de medida e unidade
    if "amp" in col_str:
        medida = "Amplitude"
        unidade = "uV"
    elif "lat" in col_str:
        medida = "Latency"
        unidade = "ms"
    else:
        medida = "Desconhecido"

```

```

        unidade = "Desconhecido"

    return {
        'Hemisferio': hemisferio,
        'Dose': dose,
        'Condicao': condicao,
        'Componente': componente,
        'Medida': medida,
        'Unidade': unidade
    }

# Processar cada coluna de medição separadamente
long_results = []

# Primeiro processar as colunas de medição regulares
for col in cols_medidas:
    # Extrair informações da coluna
    info = parse_column_info(col, tipo_especial=None)

    # Criar um DataFrame temporário para esta coluna
    temp_df = df_wide[[id_col, col]].copy()
    temp_df.columns = [id_col, 'Valor']

    # Adicionar as informações extraídas
    for key, value in info.items():
        temp_df[key] = value

    # Adicionar ao resultado acumulado
    long_results.append(temp_df)

# Depois processar as colunas de pico a pico
for col in cols_pico_a_pico:
    # Extrair informações da coluna
    info = parse_column_info(col, tipo_especial='pico_a_pico')

    # Criar um DataFrame temporário para esta coluna
    temp_df = df_wide[[id_col, col]].copy()
    temp_df.columns = [id_col, 'Valor']

    # Adicionar as informações extraídas
    for key, value in info.items():
        temp_df[key] = value

    # Adicionar ao resultado acumulado
    long_results.append(temp_df)

# Processar as colunas de latência interpicos
for col in cols_latencia_interpicos:
    # Extrair informações da coluna
    info = parse_column_info(col, tipo_especial='latencia_interpicos')

    # Criar um DataFrame temporário para esta coluna
    temp_df = df_wide[[id_col, col]].copy()
    temp_df.columns = [id_col, 'Valor']

    # Adicionar as informações extraídas
    for key, value in info.items():
        temp_df[key] = value

```

```

# Adicionar ao resultado acumulado
long_results.append(temp_df)

# Concatenar todos os resultados
long_df = pd.concat(long_results, ignore_index=True)

# Adicionar informações de sexo
long_df['Sexo'] = long_df[id_col].apply(lambda x: "Feminino" if
str(x).startswith('F') else
                                     "Masculino" if
str(x).startswith('M') or str(x).startswith('R_M') else
                                     "Desconhecido")

# Mesclar com os dados demográficos
result_df = pd.merge(long_df, demo_df, on=id_col, how='left')

# Ordenar o DataFrame
if 'Grupo' in result_df.columns:
    # Definir a ordem personalizada de classificação para o grupo
    ordem_grupo = {'Controlo': 0, 'Autismo': 1, 'ASD': 1}
    result_df['Ordem_Grupo'] = result_df['Grupo'].map(lambda g:
ordem_grupo.get(str(g), 2))

    # Ordenar os dados
    result_df = result_df.sort_values(['Ordem_Grupo', id_col, 'Hemisferio',
'Dose', 'Condicao', 'Componente', 'Medida'])

    # Remover a coluna auxiliar
    result_df = result_df.drop(columns=['Ordem_Grupo'])
else:
    # Ordenação simples sem a coluna Grupo
    result_df = result_df.sort_values([id_col, 'Hemisferio', 'Dose',
'Condicao', 'Componente', 'Medida'])

# Remover valores NaN
result_df = result_df.dropna(subset=['Valor'])

# Reorganizar colunas para o formato desejado
colunas_fundamentais = [id_col, 'Sexo', 'Hemisferio', 'Dose', 'Condicao',
'Componente', 'Medida', 'Valor', 'Unidade']
if 'Grupo' in result_df.columns:
    colunas_fundamentais.insert(1, 'Grupo')

# Depois todas as outras colunas (demográficas)
outras_colunas = [col for col in result_df.columns if col not in
colunas_fundamentais]
ordem_final = colunas_fundamentais + outras_colunas

# Aplicar a ordem final
result_df = result_df[ordem_final]

# Guardar o resultado
result_df.to_excel("resultados_long_completos_com_todas_medidas.xlsx",
index=False)

print("Conversão concluída! Dados salvos em
'resultados_long_completos_com_todas_medidas.xlsx'")

```

Código 6 - Script em Python de conversão para formato *long*.

O seguinte *script* automatiza a geração de visualizações comparativas dos ERPs, processando iterativamente os datasets EEG para criar três conjuntos de gráficos que ilustram as diferenças eletrofisiológicas entre grupos (PEA e controlo), condições de estímulo (face e não-face) e doses farmacológicas, com representação do erro padrão através de áreas sombreadas transparentes.

```

% Definir canais de interesse
canais = [32, 28];
canal_labels = {'Canal 32', 'Canal 28'};

% Diretório raiz
dir_raiz = pwd;

% Estrutura das pastas
doses = {'Placebo', '15 mg Arbaclofeno', '30 mg Arbaclofeno'};
grupos = {'ASD', 'CTRL'};
condicoes = {'face', 'no face'};

%% GRÁFICO 1: Comparação ASD vs CTRL (Placebo, Face) para ambos os canais
fprintf('\n=== GRÁFICO 1: Comparação ASD vs CTRL (Placebo, Face) ===\n');

% Cores para os grupos
cores_grupos = {[1 0 0], [0 0 1]}; % Vermelho para ASD, Azul para CTRL

% Loop para cada canal
for c = 1:length(canais)
    canal = canais(c);

    figure;
    hold on;

    % Variável para armazenar handles das linhas principais para a legenda
    h_lines = [];
    legend_labels = {};

    times = []; % Inicializar times

    % Para cada grupo (ASD e CTRL)
    for g = 1:length(grupos)
        % Construir o caminho para a pasta
        caminho_pasta = fullfile(dir_raiz, 'Placebo', grupos{g}, 'face');

        if ~exist(caminho_pasta, 'dir')
            fprintf('Pasta não encontrada: %s\n', caminho_pasta);
            continue;
        end

        cd(caminho_pasta);
        fprintf('Processando: Placebo - %s - face\n', grupos{g});

        % Listar os arquivos .set
        lista = dir('*.set');
        todos_ERPs = [];

        % Processar cada arquivo
        for f = 1:length(lista)
            try
                EEG = pop_loadset(lista(f).name);
            end
        end
    end
end

```

```

        if ndims(EEG.data) ~= 3 || size(EEG.data, 1) < canal
            continue;
        end

        erp = mean(squeeze(EEG.data(canal, :, :)), 2);
        todos_ERPs(:, end+1) = erp;

        if isempty(times), times = EEG.times; end
    catch
        continue;
    end
end

if ~isempty(todos_ERPs)
    % Calcular média e erro padrão
    erp_mean = mean(todos_ERPs, 2, 'omitnan');
    erp_std = std(todos_ERPs, 0, 2, 'omitnan');
    erp_sem = erp_std / sqrt(size(todos_ERPs, 2));

    % Criar área sombreada com transparência muito suave
    x_fill = [times, fliplr(times)];
    y_fill = [erp_mean + erp_sem; flipud(erp_mean - erp_sem)];

    % Plot da área sombreada primeiro (para ficar atrás)
    fill(x_fill, y_fill, cores_grupos{g}, 'FaceAlpha', 0.2,
'EdgeColor', 'none', ...
        'HandleVisibility', 'off');

    % Plot da linha média por cima
    h = plot(times, erp_mean, 'LineWidth', 2, 'Color',
cores_grupos{g});

    % Guardar handle e label para a legenda
    h_lines(end+1) = h;
    legend_labels{end+1} = sprintf('%s (n=%d)', grupos{g},
size(todos_ERPs, 2));
end
end

% Verificar se temos dados para formatar o gráfico
if ~isempty(times)
    % Formatar o gráfico
    xlabel('Tempo (ms)');
    ylabel('Amplitude (\muV)');
    title(sprintf('Comparação ASD vs CTRL - Placebo, Face - %s',
canal_labels{c}));
    grid on;
    xlim([min(times) max(times)]);
    ylim([-10 10]);

    % Linhas de referência
    line([0 0], ylim, 'Color', [0.5 0.5 0.5], 'LineStyle', '--');
    line(xlim, [0 0], 'Color', [0.5 0.5 0.5], 'LineStyle', '-');

    % Legenda apenas com as linhas principais
    if ~isempty(h_lines)
        legend(h_lines, legend_labels, 'Location', 'best');
    end

    % Adicionar texto explicativo sobre a área sombreada

```

```

        text(0.02, 0.02, 'Área sombreada: \pm1 erro padrão', ...
            'Units', 'normalized', 'FontSize', 8, 'Color', [0.5 0.5 0.5]);

        hold off;

        % Guardar figura
        nome_arquivo = sprintf('Comparacao_ASD_vs_CTRL_Placebo_Face_%s.fig',
canal_labels{c});
        nome_arquivo = strrep(nome_arquivo, 'Canal ', 'ch');
        savefig(nome_arquivo);
        saveas(gcf, strrep(nome_arquivo, '.fig', '.png'));
        fprintf('Figura guardada: %s\n', nome_arquivo);
    else
        close(gcf);
        warning('Não foi possível criar o gráfico para o canal %d', canal);
    end

    cd(dir_raiz);
end

%% GRÁFICO 2: Comparação Face vs No Face (Placebo, Canal 28)
fprintf('\n=== GRÁFICO 2: Comparação Face vs No Face (Placebo, Canal 28)
===\n');

canal = 28;

% Cores para face/no face
cores_face_ASD = {[1 0 0], [1 0.5 0]}; % Vermelho e laranja para ASD
cores_face_CTRL = {[0 0 1], [0 0.5 1]}; % Azul escuro e claro para CTRL

% Para cada grupo
for g = 1:length(grupos)
    figure;
    hold on;

    % Variável para armazenar handles das linhas principais para a legenda
    h_lines = [];
    legend_labels = {};

    times = []; % Inicializar times

    if strcmp(grupos{g}, 'ASD')
        cores_linha = cores_face_ASD;
    else
        cores_linha = cores_face_CTRL;
    end

    % Para cada condição (face e no face)
    for cond = 1:length(condicoes)
        % Construir o caminho
        caminho_pasta = fullfile(dir_raiz, 'Placebo', grupos{g},
condicoes{cond});

        if ~exist(caminho_pasta, 'dir')
            fprintf('Pasta não encontrada: %s\n', caminho_pasta);
            continue;
        end

        cd(caminho_pasta);

```

```

        fprintf('Processando:  Placebo  -  %s  -  %s\n',  grupos{g},
condicoes{cond});

% Processar arquivos
lista = dir('*.set');
todos_ERPs = [];

for f = 1:length(lista)
    try
        EEG = pop_loadset(lista(f).name);

        if ndims(EEG.data) ~= 3 || size(EEG.data, 1) < canal
            continue;
        end

        erp = mean(squeeze(EEG.data(canal,:,:), 2), 2);
        todos_ERPs(:,end+1) = erp;

        if isempty(times), times = EEG.times; end
    catch
        continue;
    end
end

if ~isempty(todos_ERPs)
    % Calcular média e erro padrão
    erp_mean = mean(todos_ERPs, 2, 'omitnan');
    erp_std = std(todos_ERPs, 0, 2, 'omitnan');
    erp_sem = erp_std / sqrt(size(todos_ERPs, 2));

    % Criar área sombreada com transparência muito suave
    x_fill = [times, fliplr(times)];
    y_fill = [erp_mean + erp_sem; flipud(erp_mean - erp_sem)];

    % Plot da área sombreada primeiro
    fill(x_fill, y_fill, cores_linha{cond}, 'FaceAlpha', 0.2,
'EdgeColor', 'none', ...
        'HandleVisibility', 'off');

    % Plot da linha média
    h = plot(times, erp_mean, 'LineWidth', 2, 'Color',
cores_linha{cond});

    % Guardar handle e label para a legenda
    h_lines(end+1) = h;
    legend_labels{end+1} = sprintf('%s (n=%d)', condicoes{cond},
size(todos_ERPs, 2));
end
end

% Formatar gráfico
if ~isempty(times)
    xlabel('Tempo (ms)');
    ylabel('Amplitude (\muV)');
    title(sprintf('Comparação Face vs No Face - %s (Placebo) - Canal 28',
grupos{g}));
    grid on;
    xlim([min(times) max(times)]);
    ylim([-10 10]);
end

```

```

% Linhas de referência
line([0 0], ylim, 'Color', [0.5 0.5 0.5], 'LineStyle', '--');
line(xlim, [0 0], 'Color', [0.5 0.5 0.5], 'LineStyle', '-');

% Legenda
if ~isempty(h_lines)
    legend(h_lines, legend_labels, 'Location', 'best');
end

% Adicionar texto explicativo
text(0.02, 0.02, 'Área sombreada: \pm1 erro padrão', ...
     'Units', 'normalized', 'FontSize', 8, 'Color', [0.5 0.5 0.5]);

hold off;

% Guardar figura
nome_arquivo
sprintf('Comparacao_Face_vs_NoFace_%s_Placebo_ch28.fig', grupos{g});
savefig(nome_arquivo);
saveas(gcf, strrep(nome_arquivo, '.fig', '.png'));
fprintf('Figura guardada: %s\n', nome_arquivo);
else
    close(gcf);
    warning('Não foi possível criar o gráfico para o grupo %s',
grupos{g});
end

cd(dir_raiz);
end

%% GRÁFICO 3: Comparação das três doses (Face) para ambos os canais
fprintf('\n=== GRÁFICO 3: Comparação das três doses (Face) ===\n');

% Cores para as doses
cores_doses = {[0 0.8 0], [1 1 0], [0.5 0 0.5]}; % Verde, Amarelo, Roxo

% Para cada grupo e canal
for g = 1:length(grupos)
    for c = 1:length(canais)
        canal = canais(c);

        figure;
        hold on;

% Variável para armazenar handles das linhas principais para a
legenda
        h_lines = [];
        legend_labels = {};

        times = []; % Inicializar times

% Para cada dose
        for d = 1:length(doses)
            % Construir o caminho
            caminho_pasta = fullfile(dir_raiz, doses{d}, grupos{g}, 'face');

            if ~exist(caminho_pasta, 'dir')
                fprintf('Pasta não encontrada: %s\n', caminho_pasta);
                continue;
            end
        end
    end
end

```

```

end

cd(caminho_pasta);
fprintf('Processando: %s - %s - face\n', doses{d}, grupos{g});

% Processar arquivos
lista = dir('*.set');
todos_ERPs = [];

for f = 1:length(lista)
    try
        EEG = pop_loadset(lista(f).name);

        if ndims(EEG.data) ~= 3 || size(EEG.data, 1) < canal
            continue;
        end

        erp = mean(squeeze(EEG.data(canal,:,:)), 2);
        todos_ERPs(:,end+1) = erp;

        if isempty(times), times = EEG.times; end
    catch
        continue;
    end
end

if ~isempty(todos_ERPs)
    % Calcular média e erro padrão
    erp_mean = mean(todos_ERPs, 2, 'omitnan');
    erp_std = std(todos_ERPs, 0, 2, 'omitnan');
    erp_sem = erp_std / sqrt(size(todos_ERPs, 2));

    % Criar área sombreada com transparência muito suave
    x_fill = [times, fliplr(times)];
    y_fill = [erp_mean + erp_sem; flipud(erp_mean - erp_sem)];

    % Plot da área sombreada primeiro
    fill(x_fill, y_fill, cores_doses{d}, 'FaceAlpha', 0.2,
'EdgeColor', 'none', ...
        'HandleVisibility', 'off');

    % Plot da linha média
    h = plot(times, erp_mean, 'LineWidth', 2, 'Color',
cores_doses{d});

    % Guardar handle e label para a legenda
    h_lines(end+1) = h;
    legend_labels{end+1} = sprintf('%s (n=%d)', doses{d},
size(todos_ERPs, 2));
end
end

% Formatar gráfico
if ~isempty(times)
    xlabel('Tempo (ms)');
    ylabel('Amplitude (\muV)');
    title(sprintf('Comparação de Doses - %s (Face) - %s', grupos{g},
canal_labels{c}));
    grid on;
    xlim([min(times) max(times)]);
end

```

```

        ylim([-10 10]);

        % Linhas de referência
        line([0 0], ylim, 'Color', [0.5 0.5 0.5], 'LineStyle', '--');
        line(xlim, [0 0], 'Color', [0.5 0.5 0.5], 'LineStyle', '--');

        % Legenda
        if ~isempty(h_lines)
            legend(h_lines, legend_labels, 'Location', 'best');
        end

        % Adicionar texto explicativo
        text(0.02, 0.02, 'Área sombreada: \pm1 erro padrão', ...
            'Units', 'normalized', 'FontSize', 8, 'Color', [0.5 0.5 0.5]);

        hold off;

        % Guardar figura
        nome_arquivo = sprintf('Comparacao_Doses_%s_Face_%s.fig',
grupos{g}, canal_labels{c});
        nome_arquivo = strrep(nome_arquivo, 'Canal ', 'ch');
        savefig(nome_arquivo);
        saveas(gcf, strrep(nome_arquivo, '.fig', '.png'));
        fprintf('Figura guardada: %s\n', nome_arquivo);
    else
        close(gcf);
        warning('Não foi possível criar o gráfico para %s - %s',
grupos{g}, canal_labels{c});
    end

    cd(dir_raiz);
end
end
end

```

Código 7 - *Script* em MATLAB para o cálculo e visualização dos ERPs médios por condição.

Algoritmos de *Machine Learning*

O seguinte algoritmo implementa um sistema de classificação baseado em aprendizagem automática que utiliza características eletrofisiológicas dos componentes P100 e N170 para discriminar automaticamente entre participantes com PEA e controlos.

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score,
StratifiedKFold
from sklearn.preprocessing import StandardScaler
from sklearn.impute import KNNImputer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, classification_report, confusion_matrix
import warnings
warnings.filterwarnings('ignore')

```

```

# --- 1. Configuração ---
EXCEL_FILE_PATH = 'ML_grupo_N170.xlsx'
TARGET_COLUMN = 'Grupo' # Coluna alvo (0=Controlo, 1=Autismo)

# --- 2. Carregar os dados ---
print("=== CLASSIFICAÇÃO AUTISMO vs CONTROLO ===\n")
try:
    df = pd.read_excel(EXCEL_FILE_PATH)
    print(f"Dados carregados com sucesso de '{EXCEL_FILE_PATH}'")
    print(f"Dimensões iniciais: {df.shape[0]} sujeitos, {df.shape[1]}
colunas\n")

    # Mostrar distribuição dos grupos inicial
    print("Distribuição dos grupos (inicial):")
    print(df['Grupo'].value_counts())
    print(f"  Controlo (0): {(df['Grupo'] == 0).sum()} sujeitos")
    print(f"  Autismo (1): {(df['Grupo'] == 1).sum()} sujeitos\n")

except FileNotFoundError:
    print(f"Erro: O arquivo '{EXCEL_FILE_PATH}' não foi encontrado.")
    exit()
except Exception as e:
    print(f"Erro ao carregar o arquivo: {e}")
    exit()

# --- 3. Remover sujeitos com muitos valores em falta ---
# Calcular percentagem de valores em falta por sujeito (excluindo coluna
Grupo)
feature_columns = [col for col in df.columns if col != TARGET_COLUMN]
missing_per_subject = df[feature_columns].isnull().sum(axis=1)
total_features = len(feature_columns)
missing_percentage = (missing_per_subject / total_features) * 100

# Identificar sujeitos com >50% dados em falta
subjects_to_remove = missing_percentage > 50
n_subjects_removed = subjects_to_remove.sum()

if n_subjects_removed > 0:
    print(f"=== LIMPEZA DE DADOS ===")
    print(f"Removendo {n_subjects_removed} sujeitos com >50% dados em falta:")

    # Mostrar detalhes dos sujeitos removidos
    removed_subjects = df[subjects_to_remove]
    for idx in removed_subjects.index:
        grupo = "Autismo" if removed_subjects.loc[idx, 'Grupo'] == 1 else
"Controlo"
        pct = missing_percentage[idx]
        print(f" Sujeito índice {idx} ({grupo}): {pct:.1f}% dados em falta")

    # Remover sujeitos
    df_clean = df[~subjects_to_remove].copy()
    print(f"\nDimensões após limpeza: {df_clean.shape[0]} sujeitos,
{df_clean.shape[1]} colunas")

    # Mostrar nova distribuição
    print("\nDistribuição dos grupos (após limpeza):")
    print(df_clean['Grupo'].value_counts())
    print(f"  Controlo (0): {(df_clean['Grupo'] == 0).sum()} sujeitos")
    print(f"  Autismo (1): {(df_clean['Grupo'] == 1).sum()} sujeitos\n")

```

```

df = df_clean

# --- 4. Preparação dos dados ---
# Separar features (X) e target (y)
X = df.drop(columns=[TARGET_COLUMN])
y = df[TARGET_COLUMN]

print(f"Features utilizadas ({X.shape[1]} total):")
for i, col in enumerate(X.columns, 1):
    print(f" {i}. {col}")
print()

# --- 5. Imputação de valores em falta restantes ---
# Verificar valores em falta após remoção de sujeitos
missing_before = X.isnull().sum().sum()
if missing_before > 0:
    print(f"Valores em falta restantes: {missing_before}")
    print("\nValores em falta por coluna:")
    missing_counts = X.isnull().sum()
    for col in missing_counts[missing_counts > 0].index:
        print(f" {col}: {missing_counts[col]} valores em falta")

    # Aplicar KNN Imputer apenas nos valores restantes
    print("\nAplicando KNN Imputer (k=5) para valores em falta restantes...")
    imputer = KNNImputer(n_neighbors=5)
    X_imputed = imputer.fit_transform(X)
    X = pd.DataFrame(X_imputed, columns=X.columns)

    print("Imputação concluída!\n")
else:
    print("Nenhum valor em falta restante.\n")

# --- 6. Normalização dos dados ---
print("Normalizando os dados...")
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled = pd.DataFrame(X_scaled, columns=X.columns)

# --- 7. Divisão treino/teste ---
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.3, random_state=42, stratify=y
)
print(f"\nDivisão treino/teste:")
print(f" Treino: {X_train.shape[0]} sujeitos ({len(y_train[y_train==0])}
controle, {len(y_train[y_train==1])} autismo)")
print(f" Teste: {X_test.shape[0]} sujeitos ({len(y_test[y_test==0])}
controle, {len(y_test[y_test==1])} autismo)\n")

# --- 8. Definir os 4 classificadores ---
classifiers = {
    "Regressão Logística": LogisticRegression(random_state=42,
max_iter=1000),
    "SVM": SVC(random_state=42, kernel='rbf', probability=True),
    "Random Forest": RandomForestClassifier(random_state=42,
n_estimators=100),
    "KNN": KNeighborsClassifier(n_neighbors=5)
}

```

```

# --- 9. Treinar e avaliar cada modelo ---
print("=== RESULTADOS DOS MODELOS ===\n")
results = {}

for name, clf in classifiers.items():
    print(f"--- {name} ---")

    # Treinar
    clf.fit(X_train, y_train)

    # Prever
    y_pred = clf.predict(X_test)

    # Calcular métricas
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred, zero_division=0)
    recall = recall_score(y_test, y_pred, zero_division=0)
    f1 = f1_score(y_test, y_pred, zero_division=0)

    results[name] = {
        "Acurácia": accuracy,
        "Precisão": precision,
        "Recall": recall,
        "F1-Score": f1
    }

    print(f"Acurácia: {accuracy:.3f}")
    print(f"Precisão: {precision:.3f}")
    print(f"Recall: {recall:.3f}")
    print(f"F1-Score: {f1:.3f}")

    # Matriz de confusão
    cm = confusion_matrix(y_test, y_pred)
    print(f"\nMatriz de Confusão:")
    print(f"
          Pred_Controlo  Pred_Autismo")
    print(f"Real_Controlo    {cm[0,0]:3d}      {cm[0,1]:3d}")
    print(f"Real_Autismo       {cm[1,0]:3d}      {cm[1,1]:3d}")
    print()

# --- 10. Comparação dos resultados ---
print("\n=== COMPARAÇÃO DOS MODELOS ===")
results_df = pd.DataFrame(results).T
print(results_df.round(3))
print(f"\nMelhor modelo por Acurácia: {results_df['Acurácia'].idxmax()}")
print(f"\nMelhor modelo por F1-Score: {results_df['F1-Score'].idxmax()}")

# --- 11. Validação cruzada ---
print("\n=== VALIDAÇÃO CRUZADA (5-fold) ===")
cv_results = {}

# Usar StratifiedKFold para manter proporção das classes
skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

for name, clf in classifiers.items():
    scores = cross_val_score(clf, X_scaled, y, cv=skf, scoring='accuracy')
    cv_results[name] = {
        "Média": scores.mean(),
        "Desvio Padrão": scores.std(),
        "Min": scores.min(),
        "Max": scores.max()
    }

```

```

    }
    print(f"{name}:")
    print(f"  Acurácia média: {scores.mean():.3f} (+/- {scores.std():.3f})")
    print(f"  Min: {scores.min():.3f}, Max: {scores.max():.3f}")

# --- 12. Feature Importance (para Random Forest) ---
print("\n=== IMPORTÂNCIA DAS FEATURES (Random Forest) ===")
rf_model = classifiers["Random Forest"]
rf_model.fit(X_train, y_train)
feature_importance = pd.DataFrame({
    'feature': X.columns,
    'importance': rf_model.feature_importances_
}).sort_values('importance', ascending=False)

print("Top 6 features mais importantes:")
for idx, row in feature_importance.head(6).iterrows():
    print(f"  {row['feature']}: {row['importance']:.3f}")

# --- 13. Análise adicional: Verificar se há diferença HD vs HE ---
print("\n=== ANÁLISE HEMISFÉRICA ===")
hd_features = [col for col in X.columns if 'direito' in col.lower()]
he_features = [col for col in X.columns if 'esquerdo' in col.lower()]

hd_importance = feature_importance[feature_importance['feature'].isin(hd_features)]['importance'].sum()
he_importance = feature_importance[feature_importance['feature'].isin(he_features)]['importance'].sum()

print(f"Importância total Hemisfério Direito: {hd_importance:.3f}")
print(f"Importância total Hemisfério Esquerdo: {he_importance:.3f}")

# --- 14. Resumo final ---
print("\n=== RESUMO FINAL ===")
cv_df = pd.DataFrame(cv_results).T
print("\nResultados de Validação Cruzada:")
print(cv_df[['Média', 'Desvio Padrão']].round(3))
print(f"\nMelhor modelo por CV: {cv_df['Média'].idxmax()} (Acurácia: {cv_df['Média'].max():.3f})")

# --- 15. Análise PCA ---
print("\n=== ANÁLISE PCA ===")
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

# Aplicar PCA com 2 componentes
pca = PCA(n_components=2, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Criar dataframe com componentes PCA e grupos
pca_df = pd.DataFrame(data=X_pca, columns=['PC1', 'PC2'])
pca_df['Grupo'] = y.values

# Variância explicada
var_explained = pca.explained_variance_ratio_
print(f"Variância explicada por PC1: {var_explained[0]:.3f}")
print(f"Variância explicada por PC2: {var_explained[1]:.3f}")
print(f"Variância total explicada: {sum(var_explained):.3f}")

```

```

# Criar o gráfico
plt.figure(figsize=(10, 8))

# Plot para cada grupo
colors = ['blue', 'red']
markers = ['o', 's']
labels = ['Controlo', 'Autismo']

for i, grupo in enumerate([0, 1]):
    grupo_data = pca_df[pca_df['Grupo'] == grupo]
    plt.scatter(grupo_data['PC1'], grupo_data['PC2'],
                c=colors[i], marker=markers[i], label=labels[i],
                alpha=0.7, s=100, edgecolors='black', linewidth=0.5)

plt.xlabel(f'PC1 ({var_explained[0]:.1%} variância explicada)', fontsize=12)
plt.ylabel(f'PC2 ({var_explained[1]:.1%} variância explicada)', fontsize=12)
plt.title('Análise PCA - Autismo vs Controlo', fontsize=14, fontweight='bold')
plt.legend(fontsize=12)
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

# Estatísticas de separação
controlo_centroid = pca_df[pca_df['Grupo'] == 0][['PC1', 'PC2']].mean()
autismo_centroid = pca_df[pca_df['Grupo'] == 1][['PC1', 'PC2']].mean()
distance = np.sqrt(((controlo_centroid - autismo_centroid) ** 2).sum())
print(f"\nDistância entre centróides: {distance:.3f}")

```

Código 8 – Algoritmo em Python de *Machine Learning* para a classificação de dados por grupos.

O seguinte algoritmo implementa um sistema de classificação multiclasse baseado em aprendizagem automática que utiliza características eletrofisiológicas dos componentes P100 e N170 para prever automaticamente a dose de arbaclofeno administrada (placebo, 15 mg ou 30 mg de arbaclofeno), avaliando o impacto farmacológico nos padrões neuronais através de múltiplos classificadores e análise de componentes principais.

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score, StratifiedKFold
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.impute import KNNImputer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, classification_report, confusion_matrix
from sklearn.multiclass import OneVsRestClassifier
import warnings
warnings.filterwarnings('ignore')

# --- 1. Configuração ---
df = pd.read_excel('ML_dose.xlsx')

# --- 2. Carregar e preparar os dados ---
print("=== PREDIÇÃO DE DOSE DE ARBACLOFENO ===\n")

```

```

print("Classes: Placebo (0) vs 15mg (1) vs 30mg (2)\n")

# --- 3. Filtrar apenas condição 'face' ---
print("Filtrando dados para condição 'face'...")
df_face = df[df['Condicao'] == 'face'].copy()
print(f"Linhas após filtro: {df_face.shape[0]}\n")

# --- 4. Transformar formato long para wide ---
print("Transformando formato long para wide...")

# Criar identificador único para cada medida
df_face['Medida_Completa'] = df_face['Componente'] + '_' + df_face['Medida']
+ '_' + df_face['Hemisferio']

# Limpar valores da coluna Valor (remover unidades se houver)
df_face['Valor_Numerico'] =
pd.to_numeric(df_face['Valor'].astype(str).str.replace(r'^\d.-', ''),
              regex=True, errors='coerce')

# Pivot para formato wide
df_wide = df_face.pivot_table(
    index=['Sujeito', 'Grupo', 'Sexo', 'Dose'],
    columns='Medida_Completa',
    values='Valor_Numerico',
    aggfunc='first'
).reset_index()

print(f"Dimensões após transformação: {df_wide.shape[0]} sujeitos,
{df_wide.shape[1]} colunas")
print(f"Features criadas: {len(df_wide.columns) - 4}\n") # -4 por Sujeito,
Grupo, Sexo, Dose

# --- 5. Criar variável target (dose) ---
print("Distribuição das doses:")
print(df_wide['Dose'].value_counts().sort_index())

# Codificar doses
dose_mapping = {
    'Placebo': 0,
    '15 mg Arbaclofeno': 1,
    '30 mg Arbaclofeno': 2
}

df_wide['Dose_Coded'] = df_wide['Dose'].map(dose_mapping)

# Verificar se todas as doses foram mapeadas
if df_wide['Dose_Coded'].isnull().any():
    print("\nATENÇÃO: Algumas doses não foram mapeadas. Valores únicos de
dose:")
    print(df_wide['Dose'].unique())
    # Tentar mapear automaticamente
    unique_doses = df_wide['Dose'].unique()
    for dose in unique_doses:
        if pd.isna(dose_mapping.get(dose)):
            if 'placebo' in str(dose).lower():
                dose_mapping[dose] = 0
            elif '15' in str(dose):
                dose_mapping[dose] = 1
            elif '30' in str(dose):

```

```

        dose_mapping[dose] = 2
    df_wide['Dose_Coded'] = df_wide['Dose'].map(dose_mapping)

print("\nDistribuição das doses codificadas:")
print(f" Placebo (0): {(df_wide['Dose_Coded'] == 0).sum()} observações")
print(f" 15mg (1): {(df_wide['Dose_Coded'] == 1).sum()} observações")
print(f" 30mg (2): {(df_wide['Dose_Coded'] == 2).sum()} observações\n")

# --- 6. Preparar features e target ---
# Remover colunas não numéricas e identificadores
feature_columns = [col for col in df_wide.columns if col not in ['Sujeito',
'Grupo', 'Sexo', 'Dose', 'Dose_Coded']]
X = df_wide[feature_columns]
y = df_wide['Dose_Coded']

# Adicionar informação sobre grupo como feature
if 'Grupo' in df_wide.columns:
    grupo_encoded = LabelEncoder().fit_transform(df_wide['Grupo'])
    X['Grupo_Encoded'] = grupo_encoded

print(f"Total de features: {X.shape[1]}")
print(f"Exemplos de features:")
for i, col in enumerate(X.columns[:10], 1):
    print(f" {i}. {col}")
if len(X.columns) > 10:
    print(f" ... e mais {len(X.columns) - 10} features\n")

# --- 7. Remover sujeitos com muitos valores em falta ---
missing_per_subject = X.isnull().sum(axis=1)
missing_percentage = (missing_per_subject / X.shape[1]) * 100
subjects_to_remove = missing_percentage > 50

if subjects_to_remove.sum() > 0:
    print(f"Removendo {subjects_to_remove.sum()} observações com >50% dados
em falta...")
    X = X[~subjects_to_remove]
    y = y[~subjects_to_remove]
    print(f"Dimensões após limpeza: {X.shape[0]} observações\n")

# --- 8. Imputação de valores em falta ---
missing_before = X.isnull().sum().sum()
if missing_before > 0:
    print(f"Valores em falta restantes: {missing_before}")
    print("Aplicando KNN Imputer (k=5)...")
    imputer = KNNImputer(n_neighbors=5)
    X_imputed = imputer.fit_transform(X)
    X = pd.DataFrame(X_imputed, columns=X.columns)
    print("Imputação concluída!\n")

# --- 9. Normalização ---
print("Normalizando os dados...")
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled = pd.DataFrame(X_scaled, columns=X.columns)

# --- 10. Divisão treino/teste ---
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.3, random_state=42, stratify=y
)

```

```

print(f"\nDivisão treino/teste:")
print(f" Treino: {X_train.shape[0]} observações")
for dose in [0, 1, 2]:
    count = (y_train == dose).sum()
    print(f" Dose {dose}: {count} observações")
print(f" Teste: {X_test.shape[0]} observações")
for dose in [0, 1, 2]:
    count = (y_test == dose).sum()
    print(f" Dose {dose}: {count} observações")
print()

# --- 11. Definir os 4 classificadores ---
classifiers = {
    "Regressão Logística": LogisticRegression(random_state=42, max_iter=1000,
multi_class='multinomial'),
    "SVM": SVC(random_state=42, kernel='rbf', probability=True,
decision_function_shape='ovr'),
    "Random Forest": RandomForestClassifier(random_state=42,
n_estimators=100),
    "KNN": KNeighborsClassifier(n_neighbors=5)
}

# --- 12. Treinar e avaliar cada modelo ---
print("=== RESULTADOS DOS MODELOS ===\n")
results = {}

for name, clf in classifiers.items():
    print(f"--- {name} ---")

    # Treinar
    clf.fit(X_train, y_train)

    # Prever
    y_pred = clf.predict(X_test)

    # Calcular métricas
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred, average='weighted',
zero_division=0)
    recall = recall_score(y_test, y_pred, average='weighted',
zero_division=0)
    f1 = f1_score(y_test, y_pred, average='weighted', zero_division=0)

    results[name] = {
        "Acurácia": accuracy,
        "Precisão": precision,
        "Recall": recall,
        "F1-Score": f1
    }

    print(f"Acurácia: {accuracy:.3f}")
    print(f"Precisão (weighted): {precision:.3f}")
    print(f"Recall (weighted): {recall:.3f}")
    print(f"F1-Score (weighted): {f1:.3f}")

    # Matriz de confusão
    cm = confusion_matrix(y_test, y_pred)
    print(f"\nMatriz de Confusão:")
    print(f"          Pred_0  Pred_1  Pred_2")

```

```

for i, dose in enumerate(['Placebo', '15mg', '30mg']):
    print(f"{dose:8s}    {cm[i,0]:3d}    {cm[i,1]:3d}    {cm[i,2]:3d}")

# Classification report detalhado
print("\nRelatório de Classificação:")
print(classification_report(y_test, y_pred,
                            target_names=['Placebo', '15mg', '30mg'],
                            digits=3))

print()

# --- 13. Comparação dos resultados ---
print("\n=== COMPARAÇÃO DOS MODELOS ===")
results_df = pd.DataFrame(results).T
print(results_df.round(3))
print(f"\nMelhor modelo por Acurácia: {results_df['Acurácia'].idxmax()}")
print(f"Melhor modelo por F1-Score: {results_df['F1-Score'].idxmax()}")

# --- 14. Validação cruzada ---
print("\n=== VALIDAÇÃO CRUZADA (5-fold) ===")
cv_results = {}

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

for name, clf in classifiers.items():
    scores = cross_val_score(clf, X_scaled, y, cv=skf, scoring='accuracy')
    cv_results[name] = {
        "Média": scores.mean(),
        "Desvio Padrão": scores.std(),
        "Min": scores.min(),
        "Max": scores.max()
    }
    print(f"{name}:")
    print(f"  Acurácia média: {scores.mean():.3f} (+/- {scores.std():.3f})")
    print(f"  Min: {scores.min():.3f}, Max: {scores.max():.3f}")

# --- 15. Feature Importance (para Random Forest) ---
print("\n=== IMPORTÂNCIA DAS FEATURES (Random Forest) ===")
rf_model = classifiers["Random Forest"]
rf_model.fit(X_train, y_train)
feature_importance = pd.DataFrame({
    'feature': X.columns,
    'importance': rf_model.feature_importances_
}).sort_values('importance', ascending=False)

print("Top 10 features mais importantes:")
for idx, row in feature_importance.head(10).iterrows():
    print(f"  {row['feature']}: {row['importance']:.3f}")

# --- 16. Análise por componente e hemisfério ---
print("\n=== ANÁLISE POR COMPONENTE E HEMISFÉRIO ===")

# Agrupar importância por componente
for componente in ['N170', 'P100']:
    comp_features = [col for col in X.columns if componente in col]
    comp_importance = feature_importance[feature_importance['feature'].isin(comp_features)]['importance'].sum()
    print(f"Importância total {componente}: {comp_importance:.3f}")

print()

```

```

# Agrupar importância por hemisfério
for hemisferio in ['Direito', 'Esquerdo']:
    hem_features = [col for col in X.columns if hemisferio in col]
    hem_importance = feature_importance[feature_importance['feature'].isin(hem_features)]['importance'].sum()
    print(f"Importância total Hemisfério {hemisferio}: {hem_importance:.3f}")

# --- 17. Análise de confusão entre doses ---
print("\n=== ANÁLISE DE CONFUSÃO ENTRE DOSES ===")
# Usar o melhor modelo
best_model_name = results_df['Acurácia'].idxmax()
best_model = classifiers[best_model_name]

# --- 18. Análise PCA ---
print("\n=== ANÁLISE PCA - PREDIÇÃO DE DOSE ===")
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

# Aplicar PCA com 2 componentes
pca = PCA(n_components=2, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Criar dataframe com componentes PCA e doses
pca_df = pd.DataFrame(data=X_pca, columns=['PC1', 'PC2'])
pca_df['Dose'] = y.values

# Variância explicada
var_explained = pca.explained_variance_ratio_
print(f"Variância explicada por PC1: {var_explained[0]:.3f}")
print(f"Variância explicada por PC2: {var_explained[1]:.3f}")
print(f"Variância total explicada: {sum(var_explained):.3f}")

# Criar o gráfico
plt.figure(figsize=(10, 8))

# Plot para cada dose
colors = ['green', 'orange', 'purple']
markers = ['o', 's', '^']
labels = ['Placebo', '15mg', '30mg']

for i, dose in enumerate([0, 1, 2]):
    dose_data = pca_df[pca_df['Dose'] == dose]
    plt.scatter(dose_data['PC1'], dose_data['PC2'],
                c=colors[i], marker=markers[i], label=labels[i],
                alpha=0.7, s=100, edgecolors='black', linewidth=0.5)

plt.xlabel(f'PC1 ({var_explained[0]:.1%} variância explicada)', fontsize=12)
plt.ylabel(f'PC2 ({var_explained[1]:.1%} variância explicada)', fontsize=12)
plt.title('Análise PCA - Predição de Dose de Arbaclofeno', fontsize=14,
fontweight='bold')
plt.legend(fontsize=12)
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

# Estatísticas de separação entre doses
print("\nDistâncias entre centróides das doses:")

```

```

centroids = {}
for dose in [0, 1, 2]:
    centroids[dose] = pca_df[pca_df['Dose'] == dose][['PC1', 'PC2']].mean()

# Calcular distâncias entre pares
dose_names = ['Placebo', '15mg', '30mg']
for i in range(3):
    for j in range(i+1, 3):
        dist = np.sqrt(((centroids[i] - centroids[j]) ** 2).sum())
        print(f" {dose_names[i]} - {dose_names[j]}: {dist:.3f}")

# Análise de sobreposição
print("\nAnálise de dispersão por dose:")
for i, dose in enumerate([0, 1, 2]):
    dose_data = pca_df[pca_df['Dose'] == dose][['PC1', 'PC2']]
    std_pc1 = dose_data['PC1'].std()
    std_pc2 = dose_data['PC2'].std()
    print(f"          {dose_names[i]}:          std(PC1)={std_pc1:.3f},
std(PC2)={std_pc2:.3f}")

```

Código 9 - Algoritmo em Python de *Machine Learning* para a classificação de dados por dose.

O seguinte algoritmo implementa um sistema de classificação binária baseado em aprendizagem automática que utiliza exclusivamente as amplitudes do componente N170 em diferentes condições experimentais (doses e hemisférios) para discriminar automaticamente entre respostas a estímulos de face versus não-face, incluindo análise de componentes principais para visualização da separabilidade entre condições e avaliação da importância relativa de cada característica eletrofisiológica.

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score,
StratifiedKFold
from sklearn.preprocessing import StandardScaler
from sklearn.impute import KNNImputer
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, classification_report, confusion_matrix
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

# --- 1. Configuração ---
EXCEL_FILE_PATH = 'ML_dose.xlsx' # Usar o arquivo em formato long
TARGET_COLUMN = 'Condicao' # Coluna alvo (face vs no_face)

# --- 2. Carregar os dados ---
print("=== CLASSIFICAÇÃO FACE vs NO_FACE (AMPLITUDE N170) ===\n")
try:
    df = pd.read_excel(EXCEL_FILE_PATH)
    print(f"Dados carregados com sucesso de '{EXCEL_FILE_PATH}'")
    print(f"Dimensões iniciais: {df.shape[0]} linhas, {df.shape[1]}
colunas\n")
except Exception as e:

```

```

    print(f"Erro ao carregar o arquivo: {e}")
    exit()

# --- 3. Filtrar apenas amplitudes N170 ---
print("Filtrando dados para amplitude N170...")
df_n170 = df[(df['Componente'] == 'N170') & (df['Medida'] ==
'Amplitude')].copy()
print(f"Linhas após filtro: {df_n170.shape[0]}\n")

# --- 4. Transformar formato long para wide ---
print("Transformando formato long para wide...")

# Criar identificador único para cada medida
df_n170['Medida_Completa'] = 'N170_Amplitude_' + df_n170['Hemisferio'] + '_'
+ df_n170['Dose']

# Limpar valores da coluna Valor
df_n170['Valor_Numerico']
pd.to_numeric(df_n170['Valor'].astype(str).str.replace(r'[^\d.-]',
'',
regex=True), errors='coerce')

# Pivot para formato wide
df_wide = df_n170.pivot_table(
    index=['Sujeito', 'Grupo', 'Sexo', 'Condicao'],
    columns='Medida_Completa',
    values='Valor_Numerico',
    aggfunc='first'
).reset_index()

print(f"Dimensões após transformação: {df_wide.shape[0]} observações,
{df_wide.shape[1]} colunas")

# Guardar informações de grupo e sexo para análise posterior
grupos_info = df_wide['Grupo'].copy()
sexo_info = df_wide['Sexo'].copy()

# Mostrar distribuição das condições
print("\nDistribuição das condições (inicial):")
print(df_wide['Condicao'].value_counts())
face_count = (df_wide['Condicao'] == 'face').sum()
no_face_count = (df_wide['Condicao'] == 'no_face').sum()
print(f" face: {face_count} observações")
print(f" no_face: {no_face_count} observações")

# Mostrar distribuição por grupo
print("\nDistribuição por grupo:")
for grupo in df_wide['Grupo'].unique():
    grupo_data = df_wide[df_wide['Grupo'] == grupo]
    face_grupo = (grupo_data['Condicao'] == 'face').sum()
    no_face_grupo = (grupo_data['Condicao'] == 'no_face').sum()
    print(f" {grupo}: {face_grupo} face, {no_face_grupo} no_face")
print()

# --- 5. Codificar variável target ---
# Codificar condições: face=1, no_face=0
df_wide['Condicao_Coded'] = (df_wide['Condicao'] == 'face').astype(int)

# --- 6. Preparação dos dados ---
# Separar features e target

```

```

feature_columns = [col for col in df_wide.columns if 'N170_Amplitude' in col]
X = df_wide[feature_columns]
y = df_wide['Condicao_Coded']

print(f"Features utilizadas ({X.shape[1]} total):")
for i, col in enumerate(X.columns, 1):
    print(f" {i}. {col}")
print()

# --- 7. Remover observações com muitos valores em falta ---
missing_per_obs = X.isnull().sum(axis=1)
missing_percentage = (missing_per_obs / X.shape[1]) * 100
obs_to_remove = missing_percentage > 50
n_obs_removed = obs_to_remove.sum()

if n_obs_removed > 0:
    print(f"=== LIMPEZA DE DADOS ===")
    print(f"Removendo {n_obs_removed} observações com >50% dados em falta")

    X = X[~obs_to_remove]
    y = y[~obs_to_remove]
    grupos_info = grupos_info[~obs_to_remove]
    sexo_info = sexo_info[~obs_to_remove]

    print(f"\nDimensões após limpeza: {X.shape[0]} observações, {X.shape[1]}
colunas")

    # Mostrar nova distribuição
    print("\nDistribuição das condições (após limpeza):")
    print(f" no_face (0): {(y == 0).sum()} observações")
    print(f" face (1): {(y == 1).sum()} observações\n")

# --- 8. Imputação de valores em falta ---
missing_before = X.isnull().sum().sum()
if missing_before > 0:
    print(f"Valores em falta restantes: {missing_before}")
    print("Aplicando KNN Imputer (k=5)...")
    imputer = KNNImputer(n_neighbors=5)
    X_imputed = imputer.fit_transform(X)
    X = pd.DataFrame(X_imputed, columns=X.columns)
    print("Imputação concluída!\n")
else:
    print("Nenhum valor em falta restante.\n")

# --- 9. Normalização dos dados ---
print("Normalizando os dados...")
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled = pd.DataFrame(X_scaled, columns=X.columns)

# --- 10. Divisão treino/teste ---
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.3, random_state=42, stratify=y
)
print(f"\nDivisão treino/teste:")
print(f" Treino: {X_train.shape[0]} observações ({len(y_train[y_train==0])}
no_face, {len(y_train[y_train==1])} face)")
print(f" Teste: {X_test.shape[0]} observações ({len(y_test[y_test==0])}
no_face, {len(y_test[y_test==1])} face)\n")

```

```

# --- 11. Definir os 4 classificadores ---
classifiers = {
    "Regressão Logística": LogisticRegression(random_state=42,
max_iter=1000),
    "SVM": SVC(random_state=42, kernel='rbf', probability=True),
    "Random Forest": RandomForestClassifier(random_state=42,
n_estimators=100),
    "KNN": KNeighborsClassifier(n_neighbors=5)
}

# --- 12. Treinar e avaliar cada modelo ---
print("=== RESULTADOS DOS MODELOS ===\n")
results = {}

for name, clf in classifiers.items():
    print(f"--- {name} ---")

    # Treinar
    clf.fit(X_train, y_train)

    # Prever
    y_pred = clf.predict(X_test)

    # Calcular métricas
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred, zero_division=0)
    recall = recall_score(y_test, y_pred, zero_division=0)
    f1 = f1_score(y_test, y_pred, zero_division=0)

    results[name] = {
        "Acurácia": accuracy,
        "Precisão": precision,
        "Recall": recall,
        "F1-Score": f1
    }

    print(f"Acurácia: {accuracy:.3f}")
    print(f"Precisão: {precision:.3f}")
    print(f"Recall: {recall:.3f}")
    print(f"F1-Score: {f1:.3f}")

    # Matriz de confusão
    cm = confusion_matrix(y_test, y_pred)
    print(f"\nMatriz de Confusão:")
    print(f"
        Pred_no_face  Pred_face")
    print(f"Real_no_face    {cm[0,0]:3d}    {cm[0,1]:3d}")
    print(f"Real_face        {cm[1,0]:3d}    {cm[1,1]:3d}")
    print()

# --- 13. Comparação dos resultados ---
print("\n=== COMPARAÇÃO DOS MODELOS ===")
results_df = pd.DataFrame(results).T
print(results_df.round(3))
print(f"\nMelhor modelo por Acurácia: {results_df['Acurácia'].idxmax()}")
print(f"Melhor modelo por F1-Score: {results_df['F1-Score'].idxmax()}")

# --- 14. Validação cruzada ---
print("\n=== VALIDAÇÃO CRUZADA (5-fold) ===")
cv_results = {}

```

```

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

for name, clf in classifiers.items():
    scores = cross_val_score(clf, X_scaled, y, cv=skf, scoring='accuracy')
    cv_results[name] = {
        "Média": scores.mean(),
        "Desvio Padrão": scores.std(),
        "Min": scores.min(),
        "Max": scores.max()
    }
    print(f"{name}:")
    print(f"  Acurácia média: {scores.mean():.3f} (+/- {scores.std():.3f})")
    print(f"  Min: {scores.min():.3f}, Max: {scores.max():.3f}")

# --- 15. Feature Importance (para Random Forest) ---
print("\n=== IMPORTÂNCIA DAS FEATURES (Random Forest) ===")
rf_model = classifiers["Random Forest"]
rf_model.fit(X_train, y_train)
feature_importance = pd.DataFrame({
    'feature': X.columns,
    'importance': rf_model.feature_importances_
}).sort_values('importance', ascending=False)

print("Features mais importantes:")
for idx, row in feature_importance.iterrows():
    print(f"  {row['feature']}: {row['importance']:.3f}")

# --- 16. Análise por hemisfério ---
print("\n=== ANÁLISE POR HEMISFÉRIO ===")

# Agrupar importância por hemisfério
for hemisferio in ['Direito', 'Esquerdo']:
    hem_features = [col for col in X.columns if hemisferio in col]
    hem_importance = feature_importance[feature_importance['feature'].isin(hem_features)]['importance'].sum()
    print(f"Importância total Hemisfério {hemisferio}: {hem_importance:.3f}")

# --- 17. Análise PCA ---
print("\n=== ANÁLISE PCA ===")

# Aplicar PCA com 2 componentes
pca = PCA(n_components=2, random_state=42)
X_pca = pca.fit_transform(X_scaled)

# Criar dataframe com componentes PCA e condições
pca_df = pd.DataFrame(data=X_pca, columns=['PC1', 'PC2'])
pca_df['Condição'] = y.values
pca_df['Grupo'] = grupos_info.values # Adicionar info de grupo para visualização

# Variância explicada
var_explained = pca.explained_variance_ratio_
print(f"Variância explicada por PC1: {var_explained[0]:.3f}")
print(f"Variância explicada por PC2: {var_explained[1]:.3f}")
print(f"Variância total explicada: {sum(var_explained):.3f}")

# Criar o gráfico PCA - Visualização 1: Por condição
plt.figure(figsize=(10, 8))

```

```

colors = ['purple', 'green']
markers = ['s', 'o']
labels = ['no_face', 'face']

for i, cond in enumerate([0, 1]):
    cond_data = pca_df[pca_df['Condicao'] == cond]
    plt.scatter(cond_data['PC1'], cond_data['PC2'],
               c=colors[i], marker=markers[i], label=labels[i],
               alpha=0.7, s=100, edgecolors='black', linewidth=0.5)

plt.xlabel(f'PC1 ({var_explained[0]:.1%} variância explicada)', fontsize=12)
plt.ylabel(f'PC2 ({var_explained[1]:.1%} variância explicada)', fontsize=12)
plt.title('Análise PCA - Face vs No_Face (Amplitude N170)', fontsize=14,
          fontweight='bold')
plt.legend(fontsize=12)
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

# Estatísticas de separação
no_face_centroid = pca_df[pca_df['Condicao'] == 0][['PC1', 'PC2']].mean()
face_centroid = pca_df[pca_df['Condicao'] == 1][['PC1', 'PC2']].mean()
distance = np.sqrt(((face_centroid - no_face_centroid) ** 2).sum())
print(f"\nDistância entre centróides: {distance:.3f}")

# Criar segunda visualização: Por grupo e condição
plt.figure(figsize=(12, 8))

# Plot separado por grupo para ver se há padrões
grupos_unicos = pca_df['Grupo'].unique()
colors_grupo = ['blue', 'red']
markers_cond = {0: 's', 1: 'o'} # 0=no_face, 1=face

for i, grupo in enumerate(grupos_unicos):
    for cond in [0, 1]:
        mask = (pca_df['Grupo'] == grupo) & (pca_df['Condicao'] == cond)
        data = pca_df[mask]
        label = f'{grupo} - {"face" if cond == 1 else "no_face"}'

        plt.scatter(data['PC1'], data['PC2'],
                   c=colors_grupo[i], marker=markers_cond[cond],
                   label=label, alpha=0.7, s=100,
                   edgecolors='black', linewidth=0.5)

plt.xlabel(f'PC1 ({var_explained[0]:.1%} variância explicada)', fontsize=12)
plt.ylabel(f'PC2 ({var_explained[1]:.1%} variância explicada)', fontsize=12)
plt.title('Análise PCA - Separação por Grupo e Condição', fontsize=14,
          fontweight='bold')
plt.legend(fontsize=10, bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

# --- 18. Análise por dose ---
print("\n=== ANÁLISE POR DOSE ===")
dose_features = {
    'Placebo': [col for col in X.columns if 'Placebo' in col],
    '15mg': [col for col in X.columns if '15 mg' in col],

```

```

    '30mg': [col for col in X.columns if '30 mg' in col]
}

for dose, features in dose_features.items():
    dose_importance =
feature_importance[feature_importance['feature'].isin(features)]['importance'].sum()
    print(f"Importância total {dose}: {dose_importance:.3f}")

# Análise adicional dos clusters PCA
print("\n=== ANÁLISE DOS PADRÕES NO PCA ===")
# Identificar possíveis outliers ou padrões
pc1_threshold = pca_df['PC1'].quantile(0.9)
outliers = pca_df[pca_df['PC1'] > pc1_threshold]
print(f"Observações com PC1 > percentil 90 ({pc1_threshold:.2f}):")
print(f" Total: {len(outliers)}")
print(f"
          Distribuição          por          grupo:
{outliers['Grupo'].value_counts().to_dict()}")
print(f"
          Distribuição          por          condição:
face={len(outliers[outliers['Condicao']==1])},
no_face={len(outliers[outliers['Condicao']==0])}")

# Análise da performance por grupo (post hoc)
print("\n=== ANÁLISE POST HOC: PERFORMANCE POR GRUPO ===")
# Identificar índices de teste que pertencem a cada grupo
test_indices = y_test.index
grupos_test = grupos_info.loc[test_indices]

best_model = classifiers[results_df['Acurácia'].idxmax()]
y_pred_best = best_model.predict(X_test)

for grupo in grupos_test.unique():
    mask = grupos_test == grupo
    acc_grupo = accuracy_score(y_test[mask], y_pred_best[mask])
    print(f" Acurácia no grupo {grupo}: {acc_grupo:.3f}")

```

Código 10 - Algoritmo em Python de *Machine Learning* para a classificação de dados por resposta comportamental.