

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO

Identificação de Regiões do Cérebro para
Classificação de Severidade do TEA
Utilizando Machine Learning e fMRI

Igor Duarte Rodrigues

Itajubá, 25 de novembro de 2022

**UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA E TECNOLOGIA DA COMPUTAÇÃO**

Igor Duarte Rodrigues

**Identificação de Regiões do Cérebro para
Classificação de Severidade do TEA
Utilizando Machine Learning e fMRI**

submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de em Ciências em Ciência e Tecnologia da Computação.

Área de Concentração:

Orientador: Prof. Dr. Guilherme Sousa Bastos

**25 de novembro de 2022
Itajubá**

Igor Duarte Rodrigues

Identificação de Regiões do Cérebro para Classificação de Severidade do TEA Utilizando Machine Learning e fMRI/ Igor Duarte Rodrigues. – Itajubá, 25 de novembro de 2022-

66 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Guilherme Sousa Bastos

()

Universidade Federal de Itajubá

Programa de Pós-Graduação em Ciência e Tecnologia da Computação, 25 de novembro de 2022.

1. Transtorno do Espectro Autista. 2. Machine Learning. I. Prof. Dr. Guilherme Sousa Bastos. II. Universidade Federal de Itajubá. III. POSCOMP. IV. Identificando Regiões do Cérebro para Classificação de Severidade do ASD Utilizando Machine Learning e fMRI

CDU 07:181:009.3

Igor Duarte Rodrigues

Identificação de Regiões do Cérebro para Classificação de Severidade do TEA Utilizando Machine Learning e fMRI

submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de em Ciências em Ciência e Tecnologia da Computação.

Trabalho aprovado. Itajubá, 26 de Fevereiro de 2021:

Prof. Dr. Guilherme Sousa Bastos
Orientador

Dr Ricardo Zorzetto Nicoliello Vencio

Dr. Rafael de Magalhães Dias
Frinhani

Itajubá
25 de novembro de 2022

Agradecimentos

Agradeço aos meus pais, Rogério e Juliana, e ao meu irmão Victor Hugo, e aos meus avós Rosaly e Delarue por todo apoio, carinho e compreensão. A Deborah Rabello, que muito me incentivou a iniciar o curso de mestrado, sem a qual não teria iniciado esse caminho.

Ao meu orientador Prof. Dr. Guilherme, pela atenção, dedicação, ensinamentos e pela paciência. Foi um período de muito aprendizado, ao qual eu serei eternamente grato. Agradeço aos amigos do laboratório de pesquisa, em especial ao doutorando Emerson Carvalho, tanto pela amizade quanto pelo incentivo e ajuda durante o mestrado, as diversas discussões foram extremamente produtivas e de grande contribuição para minha formação. Também a Caio Pinheiro, Gustavo Leite e Vinícius Paiva, que muito me motivaram a pesquisar, mesmo aos feriados e finais de semana virando noites no laboratório.

Agradeço a Universidade Federal de Itajubá pela oportunidade de cursar o mestrado, que sem dúvida foi uma experiência muito construtiva e engrandecedora. Assim como a Universidade Federal de Ouro Preto, a qual possibilitou a construção do conhecimento necessário para ingressar no mestrado.

Agradeço também aos professores Dr. Ricardo Zorzetto Nicoliello Vencio e Dr. Rafael de Magalhães Dias Frinhani, membros da banca examinadora, pelas considerações feitas sobre o trabalho, que sem dúvida contribuíram.

Agradeço em especial a CAPES, que fomentou a bolsa que permitiu que eu me dedicasse de forma exclusiva ao mestrado, possibilitando que eu aproveitasse ao máximo a oportunidade de aprendizado. Assim como toda a sociedade brasileira que indiretamente permite a existência de agências de fomento como a CAPES assim como as universidades federais. Agradeço à FAPEMIG pelo auxílio financeiro através do projeto APQ-01565-18.

Resumo

O Transtorno do Espectro Autista (TEA) é uma condição neurológica vitalícia relacionada à idade e ao sexo, caracterizada principalmente por disparidades sociais. A prevalência atual do TEA indica que uma em cada 59 crianças estão dentro do espectro. O Autism Diagnostic Observation Schedule (ADOS) é um processo de diagnóstico que classifica o TEA de acordo com a gravidade do transtorno. ADOS classifica sintomas mais graves como casos de “autismo” e os que manifestam sintomas mais leves como casos de “TEA não autista” (TEA-NA). Muitos artigos objetivam criar algoritmos para diagnosticar TEA por meio de aprendizado de máquina (do inglês Machine Learning - ML) e imagens de ressonância magnética funcional (do Inglês Functional Magnetic Resonance Imaging - fMRI). Essas abordagens avaliam o fluxo de oxigênio no cérebro para classificar os indivíduos como TEA ou com desenvolvimento típico. No entanto, em geral, esses trabalhos não fornecem informações sobre a gravidade do transtorno. Esse trabalho tem como objetivo a identificação de regiões do cérebro com diferença funcional entre indivíduos TEA-NA e autistas, como possível biomarcador para a severidade das características TEA. Para isso, o trabalho utilizou dados de fMRI de 202 indivíduos, e suas respectivas pontuações ADOS disponíveis no consórcio ABIDE para determinar a subclasse de TEA correta para cada um. Esses dados foram utilizados para alimentar um algoritmo de ML, de aprendizado supervisionado, o Support Vector Machine (SVC), de forma a selecionar as regiões com maior diferença funcional para os indivíduos da amostra. Os resultados corroboram a hipótese inicial de diferenças funcionais entre as subclasses de TEA, com algumas regiões do cérebro onde a diferença funcional foi suficiente para criar 74% de precisão na classificação. Este trabalho apresenta limitações quanto ao número total de amostras. No entanto, a abordagem mostra-se promissora para o diagnóstico de severidade do TEA.

Palavras-chaves: ABIDE, Transtorno do Espectro Autista, Classificação de Severidade do Transtorno do Espectro Autista, fMRI, Machine Learning.

Abstract

Autism Spectrum Disorder (ASD) is an age- and sex-related lifelong neurodevelopmental disorder characterized primarily by social impairments. Current ASD prevalence indicates that 1/59 children are diagnosed inside the spectrum. The Autism Diagnostic Observation Schedule (ADOS) classifies ASD according to the disorder severity. ADOS classifies as 'autism' cases that manifest more severe symptoms and as 'ASD non-autism' cases that exhibit milder symptoms. Many papers aimed to create algorithms to diagnose ASD through Machine Learning (ML) and functional Magnetic Resonance Images (fMRI). Such approaches evaluate the oxygen flow in the brain to classify the subjects as ASD or typical development. However, most of these works, do not provided information regarding the disorder severity. This paper aims to use ML and fMRI to classify the disorder severity, aim to find brain regions potentially related to the disorder severity. We used fMRI data of 202 subjects and their ADOS scores available at the ABIDE consortium to determine the correct ASD sub-class for each one. Our results corroborate the initial hypothesis of functional differences within ASD, with some brain regions where the functional difference was enough to create classification accuracy of 74%. This paper has limitations regarding the total number of samples. However, it shows a promising approach to ASD diagnosis.

Key-words: ABIDE, Autism Spectrum Disorders, Autism Spectrum Disorders Severity Classification, fMRI, Machine Learning.

Lista de ilustrações

Figura 1 – Relação de custos entre indivíduos com diagnóstico TEA e DT, conforme idade. Valores em dólares baseados em 2007. Custos de medicamentos, custos médicos e custos de internação respectivamente em azul, vermelho e amarelo para indivíduos com diagnóstico TEA enquanto verde, laranja e rosa para indivíduos DT. FONTE: (SHIMABUKURO; GROSSE; RICE, 2008)	18
Figura 2 – Exemplos de atlas. FONTE: (LONG et al., 2018)	20
Figura 3 – Atlas AAL: (A) lateral esquerdo, (B) medial esquerdo, (C) anterior, (D) superior, (E) inferior, (F) lateral direito, (G) medial direito, e (H) posterior. - FONTE: (LEWIS et al., 2013)	23
Figura 4 – Regiões Cortical e subcortical (58 por hemisfério; totalizando 116). - FONTE: (CAEYENBERGHS et al., 2012)	24
Figura 5 – Exemplo de MRI convencional em corte axial, FONTE: (TANENBAUM et al., 2017)	25
Figura 6 – Exemplo de imagem de ressonância magnética funcional para ativação do córtex durante estímulo visual Fonte: (BETTS et al., 2013), creditado a “Superborsuk”/Wikimedia Commons	26
Figura 7 – Processo de Validação e treinamento utilizando <i>k-fold cross-validation</i> , com valor de $k = 10$, em cada nova interação (linha) é selecionado um <i>fold</i> diferente (coluna) para ser usado como amostra de validação, enquanto os outros 9 ($k - 1$) são utilizados como amostra de treinamento. O processo é repetido 10 vezes (valor de k)	40
Figura 8 – Exemplo de separação usando <i>kernel</i> Linear (a), <i>kernel</i> Poly (b), <i>kernel</i> RBF (c) e <i>kernel</i> Sigmoid (d)	43
Figura 9 – SoRs com 5 melhores resultados de ACC. SoRs 11 - roxo, SoRs 23 - verde, SoRs 4 - azul, SoRs 1 - proto, SoRs 30 amarelo FONTE: (CAEYENBERGHS et al., 2012)	47
Figura 10 – SoRs com ACC, SNS, e SPC superior ou igual a 60%	48
Figura 11 – Resultados de mínimo, máximo, média e desvio mediana respectivamente, para SoRs com ACC, SNS, e SPC superior ou igual a 60%	49

Lista de tabelas

Tabela 1 – Pontuação máxima e nota de corte para TEA-NA e Autismo, segundo ADOS	19
Tabela 2 – Nome e identificação das regiões do atlas AAL	22
Tabela 3 – Opções de estratégia de pré-processamento disponíveis no ABIDE	28
Tabela 4 – Atlas disponíveis no ABIDE	29
Tabela 5 – Tabela Parcial Resultante do processamento DPARSF	34
Tabela 6 – Grupo Indivíduos TEA-NA	35
Tabela 7 – Grupo Indivíduos Autistas	36
Tabela 8 – Distribuição por Fenótipo	36
Tabela 9 – Distribuição por Sexo	36
Tabela 10 – IDs das SoRs e os respectivos IDs das ROI's provenientes do atlas AAL	38
Tabela 11 – Melhores Resultados Por Região	55
Tabela 12 – Melhores Resultados Para Kernel Poly	56
Tabela 13 – Melhores Resultados Para <i>Kernel Sigmoid</i>	57
Tabela 14 – Melhores Resultados Para <i>Kernel RBF</i>	58
Tabela 15 – Comparação dos resultados entre <i>Kernel poly</i> e <i>Sigmoid</i>	59

Lista de abreviaturas e siglas

AAL	Automated Anatomical Labeling
ABIDE	Autism Brain Imaging Data Exchange
ACC	Accuracy (Precisão)
ADIR	Autism Diagnostic Interview-Revised
ADI	Autism Diagnostic Interview
ADOS	Autism Diagnostic Observation Schedule
CCS	Connectome Computation System
CPAC	Configurable Pipeline for the Analysis of Connectomes
DPARSF	Data Processing Assistant for Resting-State fMRI
DT	Desenvolvimento Típico
fMRI	functional Ressonance Magnetic Image
KCV	k-fold Cross-Validation
ML	Machine Learning
MR	Magnetic Resonance
NIAK	Neuroimaging Analysis Kit
ROI	Regions of Interest
SNS	Sensibility (Sensibilidade)
SoRs	Set of Regions
SPC	Specificity (Especificidade)
SRS	Social Communication Questionnaire
TEA	Transtorno do Espectro Autista
TEA-NA	Transtorno do Espectro Autista, porem não autista
T2	Transverse relaxation rate
SVC	Support Vector Classification
SVM	Support Vector Machine

Sumário

1	INTRODUÇÃO	13
1.1	Motivação	14
1.2	Objetivos	14
1.2.1	Objetivo geral	14
1.2.2	Objetivos específicos	14
1.2.3	Hipótese	15
1.2.4	Principais contribuições	15
1.2.5	Organização do trabalho	15
2	TRANSTORNO DO ESPECTRO AUTISTA	17
2.1	Custo de vida	17
2.2	Processo de Diagnóstico	17
2.2.1	ADOS	19
3	MATERIAIS	20
3.1	Atlas	20
3.1.1	Automated Anatomical Labeling	21
3.2	fMRI	21
3.2.1	MRI	21
3.2.2	Imagem de Ressonância Magnética Funcional - fMRI	23
3.2.3	BOLD	26
3.3	ABIDE	27
3.3.1	Material disponibilizado	28
3.3.2	Trabalhos relacionados	28
3.4	Pre-Processamento	30
3.4.1	DPARSF	31
3.4.1.1	Pré-processamento Estrutural	32
3.4.1.2	Pré-processamento Funcional	32
3.4.1.3	Resultado do Pre-Processamento	34
4	MÉTODOS	35
4.1	Conjunto de Dados	35
4.1.1	Seleção de Indivíduos	35
4.1.2	Adequação da Matriz DPARSF ao SVC	36
4.1.3	Seleção de Regiões	37
4.2	Restrições de Precisão, Sensibilidade e Especificidade	38

4.3	Processo de Validação	39
5	APLICAÇÃO	41
5.1	Support Vector Classification	41
5.1.1	Kernel	42
5.1.2	Treinamento	43
5.1.3	Validação	44
5.1.4	Resultado Final	44
6	RESULTADOS E DISCUSSÃO	46
6.1	Classificação	46
6.2	Efeito do Fenótipo	48
6.3	Discussão	49
7	CONCLUSÃO	52
7.1	Trabalhos Futuros	53
	 APÊNDICES	 54
	APÊNDICE A – TABELA MELHORES RESULTADOS POR RE- GIÃO	55
	APÊNDICE B – TABELA MELHORES RESULTADOS PARA <i>KER- NEL POLY</i>	56
	APÊNDICE C – TABELA MELHORES RESULTADOS PARA <i>KER- NEL SIGMOID</i>	57
	APÊNDICE D – TABELA MELHORES RESULTADOS PARA <i>KER- NEL RBF</i>	58
	APÊNDICE E – TABELA COMPARAÇÃO DOS RESULTADOS EN- TRE <i>KERNEL POLY</i> E <i>SIGMOID</i>	59
	 REFERÊNCIAS	 60

1 Introdução

O Transtorno do Espectro Autista (TEA) é uma desordem do desenvolvimento neurológico, caracterizada principalmente por disparidade social. Frequentemente é acompanhada de desafios de comunicação ou comportamentos repetitivos que são definidos como estereotípias motoras (ASSOCIATION et al., 2014). O TEA também é associado diversas vezes a atraso no desenvolvimento motor e cognitivo (PROVOST; LOPEZ; HEIMMERL, 2007; WILSON; ENTICOTT; RINEHART, 2018), e alta sensibilidade a estímulos externos, podendo incluir seletividade alimentar (BANDINI et al., 2010; CHISTOL et al., 2018; PROSPERI et al., 2017). Atualmente o diagnóstico do TEA requer uma combinação multidisciplinar de profissionais qualificados (LEVY et al., 2017; KAMP-BECKER et al., 2018). O diagnóstico está relacionado a um processo de observação do indivíduo sob avaliação ou ainda dependente de relatos de familiares sobre seus hábitos. Dentre as principais ferramentas para auxílio de diagnóstico destacam-se o Autism Diagnostic Observation Schedule (ADOS) e o Autism Diagnostic Interview-Revised (ADI-R) (FALKMER et al., 2013).

Existe dentro do TEA diversos níveis de severidade dos sintomas e tendo em vista essa diferença o ADOS utiliza uma subdivisão do espectro, de tal forma que o resultado do diagnóstico ADOS pode ser: i) não TEA (não diagnosticado); ii) TEA não autista (TEA-NA) (menor severidade); iii) Autista (maior severidade) (LORD et al.,).

Ademais, estudos apontam que intervenções que buscam melhorar as habilidades sociais e motoras de indivíduos diagnosticados com TEA são mais eficazes quando tem um início precoce (antes dos 3 anos) e de um programa adaptado às necessidades específicas de cada indivíduo (DAWSON et al., 2010; BEJARANO-MARTÍN et al., 2020). Isto torna o diagnóstico precoce extremamente importante para uma maior eficiência de técnicas de intervenção, visto que na primeira infância o cérebro possui maior plasticidade (ARANI et al., 2015). Dessa forma, contribuindo para uma melhor qualidade de vida dos indivíduos com diagnóstico positivo para TEA, tanto os TEA-NA quanto os autistas.

Adicionalmente, para resultados mais eficientes das intervenções, além do diagnóstico precoce, é também importante uma classificação que permita segregar o espectro autista em relação a heterogenia e severidade. A exemplo do que é feito em Jiao et al. (2011), onde técnicas de ML são aplicadas à 18 crianças com TEA para tentar diferenciar indivíduos diagnosticados com a síndrome de Asperger (TEA-NA) e autistas, através de análise de imagens de ressonância magnética (do inglês *Magnetic Resonance Image - MRI*) e código genético.

1.1 Motivação

O TEA é uma desordem do desenvolvimento neurológico que afeta uma em cada 59 pessoas no mundo. Apesar do TEA ser uma condição neurológica vitalícia, existem intervenções que visam melhoria do desenvolvimento social e motor dos indivíduos. Tais intervenções mostram-se mais eficientes quando aplicadas precocemente, tornando-se crucial um diagnóstico precoce.

Considerando que ainda não existem processos automatizados de diagnóstico, diversos trabalhos buscam identificar um biomarcador - uma característica biológica identificável de forma direta por um exame médico - que permita a criação de um modelo automatizado para identificar indivíduos pertencentes ao espectro. Dentre estes, destacam-se técnicas que utilizam imagens de ressonância magnética funcional (do inglês *functional Resonance Magnetic Image - fMRI*) buscando diferenciar funcionalmente indivíduos TEA e de desenvolvimento típico (DT). [Bi et al. \(2018b\)](#), [Iidaka \(2015\)](#), [Mahanand et al. \(2016\)](#) aplicam ML em fMRI para o diagnóstico do TEA, obtendo precisão entre 81 e 96%, com amostras variando entre 107 e 1035 indivíduos.

1.2 Objetivos

1.2.1 Objetivo geral

Identificar regiões do cérebro como possíveis biomarcadores para a severidade do Transtorno do Espectro Autista.

1.2.2 Objetivos específicos

- Selecionar um atlas de divisão do cérebro que seja bem sucedido em classificação de indivíduos TEA em relação a um grupo controle.
- Selecionar um pipeline de pré-processamento para transformar a imagem medica em uma matriz para o algoritmo de ML
- Criar uma sub-divisão do cérebro, baseado no atlas selecionado, para verificar as regiões com diferenças funcionais.
- Implementar um algoritmo de ML do tipo Support Vector Machine (SVM), para classificar indivíduos dentro do Transtorno do Espectro Autista, conforme o grau de severidade.

1.2.3 Hipótese

Existem regiões do cérebro onde a variação de atividade está relacionada a severidade das características do Transtorno do Espectro Autista.

1.2.4 Principais contribuições

Visto que, os estudos da literatura têm se focado quase primariamente em propor métodos de classificação para indivíduos TEA e DT, a classificação entre indivíduos dentro do próprio espectro é pouco abordada. A principal contribuição deste trabalho é a identificação de regiões com diferente padrão de atividade entre indivíduos com diferentes graus de severidade do TEA, medido pelo nível de oxigênio (mínimo, máximo, media e mediana) na região em um intervalo de tempo. Adicionalmente, o presente trabalho apresenta as seguintes contribuições:

- Desenvolvimento de um método para tratar os dados pré-processados oriundos da base de imagens médicas ABIDE, como insumo para o classificador SVC. Transformando os dados da matriz DPARSF em um vetor;
- Adaptação de um algoritmo SVC para classificação de severidade de indivíduos TEA;
- Verificação do efeito da adição de informações de fenótipo na classificação da severidade TEA.

1.2.5 Organização do trabalho

O presente trabalho divide-se em 7 capítulos, como detalhado a seguir: O Capítulo 2 apresenta uma breve revisão sobre o transtorno do espectro autista. O Capítulo 3 faz uma revisão teórica sobre os materiais utilizados para esse trabalho, e apresenta na Seção 3.2 conceitos sobre ressonância magnética, a Seção 3.3 apresenta o consórcio ABIDE mantenedor dos dados utilizados neste trabalho, além do Atlas de mapeamento do cérebro na Seção 3.1 e do pré-processamento utilizado na Seção 3.4.

O Capítulo 4 apresenta a metodologia utilizada para execução do trabalho, dividida entre a definição dos materiais finais a serem utilizados na Seção 4.1, as especificações de restrição para seleção de regiões com potencial de possuir diferenças funcionais na Seção 4.2 e o processo de validação adotado na Seção 4.3.

O Capítulo 5, apresenta a técnica de ML selecionada para o trabalho, assim como a forma com a qual ela foi utilizada para obter-se os resultados que são apresentados no Capítulo 6, juntamente com a discussão dos mesmos. E por fim o Capítulo 7 apresenta

uma breve conclusão do trabalho, juntamente com algumas possibilidades para trabalhos futuros.

Verificação do efeito da adição de informações de fenótipo na classificação da severidade TEA.

2 Transtorno do Espectro Autista

O Transtorno do Espectro Autista (TEA) é uma desordem com características heterogêneas na qual dois indivíduos diagnosticados podem possuir um conjunto de sintomas completamente diferentes. Alguns estudos estimam uma prevalência no TEA de uma para cada 59 pessoas (BAIO, 2014). Apesar de um possível viés de diagnóstico de gênero, o TEA parece ser um transtorno relacionado ao sexo com uma proporção entre homens e mulheres perto de 3-4 (BAIO, 2014; LOOMES; HULL; MANDY, 2017). Pesquisas atuais apontam o TEA como um transtorno principalmente hereditário. Em torno de 80% dos casos de TEA são atribuídos à herança genética. Cerca de 20% são devidos a fatores de risco ambientais, que incluem problemas durante o período de gestação e a idade dos pais (BAI et al., 2019; SANDIN et al., 2017). Essa carga genética foi corroborada por Carvalho et al. (2020) em um estudo envolvendo técnicas de Inteligência Artificial para analisar a hereditariedade do TEA. A partir de uma amostra de mais de 1,5 milhões de irmãos, estimou-se que pais diagnosticados com TEA têm entre 30% e 80% de probabilidade de gerar filhos diagnosticados com TEA.

2.1 Custo de vida

Os gastos médicos de crianças e adolescentes com diagnóstico TEA chegam a 6 vezes os de uma criança ou adolescente com DT, enquanto gastos gerais podem chegar a 9,5 vezes a média (SHIMABUKURO; GROSSE; RICE, 2008). A Figura 1, apresenta uma comparação feita com dados de 2007, na qual a média de custo baseou-se em um banco de dados contendo informações sobre planos de saúde dos Estados Unidos da América. Utilizando-se de planos privados e públicos, somando dados de milhões de pacientes.

Esses custos são impulsionados principalmente para indivíduos em países de baixa ou média renda e que não recebem suporte adequado dos sistemas de saúde ou de assistência social. O alto custo decorre principalmente de: i) ferramentas proprietárias para diagnóstico; ii) técnicas de intervenções baseadas em evidências e iii) treinamento de pais e profissionais na condução do processo de intervenções para o TEA (DURKIN et al., 2015).

2.2 Processo de Diagnóstico

Atualmente existem muitos protocolos para diagnosticar e medir a severidade do TEA, como ADOS, ADIR, Social Communication Questionnaire (SRS). O ADOS (GOTHAM et al., 2007) em sua segunda versão (ADOS-2) é atualmente um dos proto-

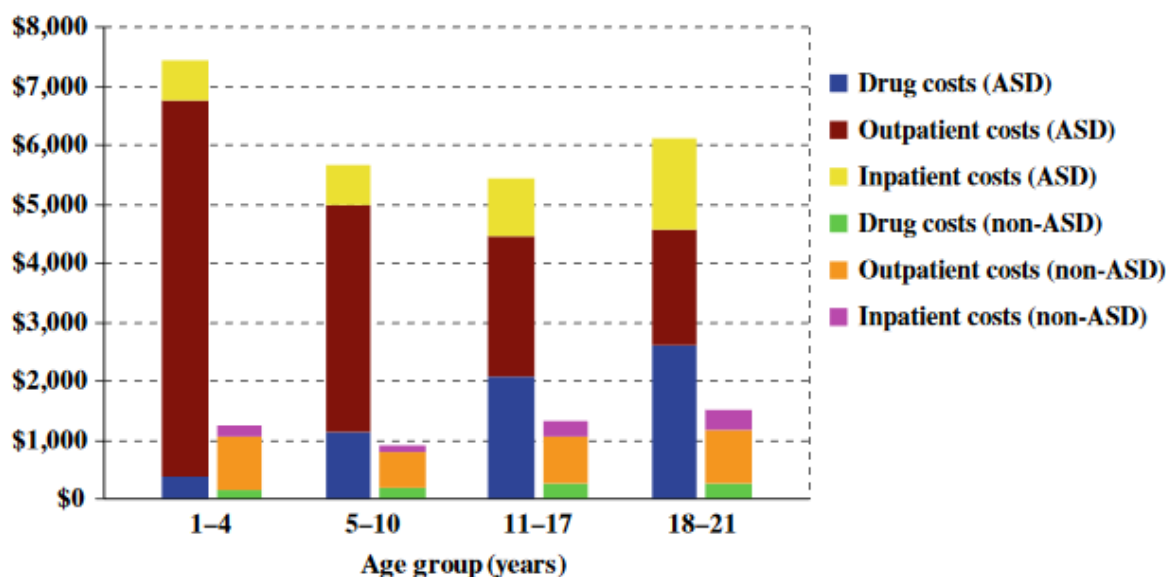


Figura 1 – Relação de custos entre indivíduos com diagnóstico TEA e DT, conforme idade. Valores em dólares baseados em 2007. Custos de medicamentos, custos médicos e custos de internação respectivamente em azul, vermelho e amarelo para indivíduos com diagnóstico TEA enquanto verde, laranja e rosa para indivíduos DT. FONTE: (SHIMABUKURO; GROSSE; RICE, 2008)

colos mais usados no mundo (DUDA; KOSMICKI; WALL, 2014). O ADOS divide sua classificação entre “autismo” (indivíduos com sintomas mais graves), “TEA-NA” (indivíduos com sintomas menos graves), e “fora do espectro” (indivíduos não diagnosticados como dentro do espectro) (LORD et al., ; GOTHAM et al., 2007).

O diagnóstico antes dos três anos de idade, e as intervenções adequadas são fatores críticos na reversão do comportamento gerado pelo TEA em crianças. As intervenções iniciais podem resultar em melhoria do funcionamento cognitivo, comportamental e social, permitindo para um subconjunto de pessoas, uma evolução que pode levar a uma vida adulta saudável e com reduções significativas de impacto social a longo prazo (HAZLETT et al., 2017).

Atualmente não existem testes automatizados de baixo custo para identificar o transtorno. O diagnóstico do TEA é realizado por meio de observação clínica, o que é um desafio de se realizar em crianças pequenas, principalmente, nos primeiros anos de vida (Brazil’s Ministry of Health, 2014). Além disso, atualmente o diagnóstico é realizado por profissionais treinados utilizando diversas ferramentas como ADOS, que possui sensibilidade e especificidade acima de 80% (FALKMER et al., 2013).

Cabe ressaltar que devido a ausência de ferramentas automatizadas para o diagnóstico, o treinamento e habilidade do examinador, assim como a sequência como os testes são aplicados, são fatores que implicam diretamente na precisão do diagnóstico (GOTHAM et al., 2007). Devido a esses fatores, o ideal é que sejam utilizados mais de

uma ferramenta de avaliação, de forma a munir o examinador com mais informações, possibilitando um diagnóstico mais preciso [Gotham et al. \(2007\)](#).

2.2.1 ADOS

Consiste em um processo de avaliação do TEA, padronizado de forma a avaliar 3 grandes áreas: comunicação, relações sociais e comportamento ([LORD et al.](#), ; [GOTHAM et al., 2007](#)). Cada área possui um conjunto de tarefas a serem avaliadas, com pontuações totais diferentes. Dado o diagnóstico com base nas pontuações obtidas em cada área e o ponto de corte estipulado. Os pontos de cortes utilizados são apresentados na Tabela 1. ADOS possui 4 módulos, cada um para uma faixa específica de idade. Para cada módulo uma nota de corte diferente é aplicada. A divisão de módulos visa avaliações com base em diferentes níveis de comunicação, uma vez que a comunicação de cada indivíduo evolui ao longo do tempo ([GOTHAM et al., 2007](#)).

Tabela 1 – Pontuação máxima e nota de corte para TEA-NA e Autismo, segundo ADOS

		Comunicação	Interação Social	Total	Comportamento Repetitivo
Módulo 1	Máximo	10	14	24	6
	corte TEA-NA	2	4	7	-
	corte Autista	4	7	12	-
Módulo 2	Máximo	10	14	24	6
	corte TEA-NA	3	4	8	-
	corte Autista	5	6	12	-
Módulo 3	Máximo	8	14	22	8
	corte TEA-NA	2	4	7	-
	corte Autista	3	6	10	-
Módulo 4	Máximo	8	14	22	8
	corte TEA-NA	2	4	7	-
	corte Autista	3	6	10	-

FONTE: ([LORD et al.](#),)

3 Materiais

Esse capítulo faz uma revisão teórica sobre os materiais utilizados neste trabalho. Seção 3.1 faz uma breve revisão sobre o que é um atlas do cérebro, além de o atlas utilizado neste trabalho. Seção 3.2 aborda os conceitos de ressonância magnética e ressonância magnética funcional, enquanto a Seção 3.3 apresenta o consórcio que mantém os dados utilizados neste trabalho. E por fim a Seção 3.4 apresenta o pipeline de pré processamento, disponível no ABIDE, utilizado para processar a imagem de ressonância magnética de forma a transformá-la nos dados utilizados neste estudo.

3.1 Atlas

O atlas consiste em uma padronização de mapeamento das regiões do cérebro. Esse mapeamento acontece na separação e nomeação de regiões anatômicas e estruturas funcionais (corticais e subcorticais) do cérebro [Fan et al. \(2016\)](#). Essas regiões são chamadas Regiões de Interesse (*Regions of Interest - ROI*). A Figura 2 apresenta 3 exemplos de atlas, AAL-90, BN-246 e AAL-1024, nos quais cada cor representa uma ROI diferente.

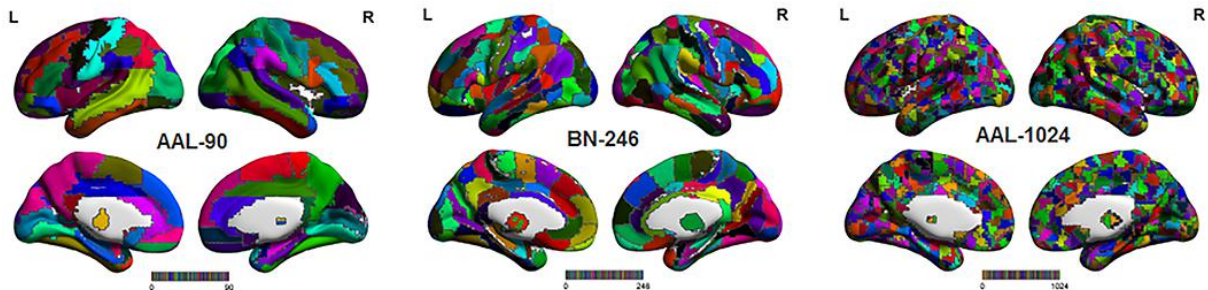


Figura 2 – Exemplos de atlas. FONTE: ([LONG et al., 2018](#))

O uso de um atlas facilita a comparação entre cérebros de indivíduos diferentes, uma vez que esse mapeamento possibilita a comparação entre ROI's, permitindo assim a comparação entre dois indivíduos tanto em termos da estrutura que compõe a região (MRI) quanto no nível de oxigênio (fMRI) entre dois indivíduos [Greitz et al. \(1991\)](#). No caso da fMRI, permite ainda uma comparação sobre um mesmo indivíduo e as atividades de diferentes regiões do cérebro dentro de um determinado intervalo de tempo.

Uma vez que as fMRI são um conjunto de imagens dentro de um intervalo de tempo ([HEEGER; RESS, 2002](#)), o uso de um atlas é extremamente importante, tanto para facilitar os experimentos e análises, quanto para explicar os resultados encontrados, uma vez que facilita a apresentação das regiões relevantes ([GREITZ et al., 1991](#)).

A existência dessa correlação permite duas abordagens diferentes, a primeira considera regiões do cérebro geralmente ligadas à diferença entre dois grupos, além de investigar a existência de uma diferença funcional nas fMRI de indivíduos de dois grupos. Já a segunda permite buscar os pontos relevantes e, após encontrá-los, descobrir qual a atividade referente às ROI's, para em seguida descobrir quais são as atividades geralmente relacionadas a cada uma dessas regiões, e conjecturar a possível razão para essa descoberta.

3.1.1 Automated Anatomical Labeling

O ABIDE disponibiliza diversas opções de atlas dentro de seu acervo de fMRI pré-processadas, dentre eles está o atlas AAL (ROLLS et al., 2019). Adicionalmente, dentre os trabalhos que objetivam a classificação de indivíduos entre neurotípico e TEA, a maioria utiliza o atlas AAL. Dessa forma, ele foi o atlas selecionado para este trabalho. Este atlas consiste em uma divisão do cérebro em 116 ROI's, as quais estão listadas na Tabela 2, com os limites representados graficamente na Figura 3, e com cada região apontada na Figura 4.

3.2 fMRI

Essa seção sumariza a imagem de ressonância magnética, uma técnica amplamente utilizada para análise do funcionamento do cérebro. O restante apresenta na Subseção 3.2.1 a imagem de ressonância magnética (do inglês Magnetic Resonance Imaging - MRI), em seguida na Subseção 3.2.2 são apresentados os conceitos de imagem de ressonância magnética funcional (do inglês functional Magnetic Resonance Image - fMRI), por fim 3.2.3 apresenta a técnica de realce mais utilizada na aquisição de fMRI.

3.2.1 MRI

A MRI é uma técnica de imagem médica não invasiva, sendo uma dos maiores avanços na área desde a introdução do raio x em 1895 (DEYOE et al., 1994; YAN; ZANG, 2010; BETTS et al., 2013; LOGOTHETIS, 2008, 2008). Ela se baseia em fenômenos da física nuclear, no qual matérias expostas a ondas de rádio e campos magnéticos emitem sinais de rádio Betts et al. (2013). A captação desses sinais de rádio são então convertidas em uma imagem, conforme apresentado na Figura 5.

A primeira máquina de MRI foi concebida pelo pesquisador e médico Raymond Damadian, em 1970, a partir de observações sobre a reação de tecido de células cancerígenas e o diferente sinal emitido por elas em relação ao tecido corporal saudável (BETTS et al., 2013; LOGOTHETIS, 2008).

Tabela 2 – Nome e identificação das regiões do atlas AAL

ID	Nome	ID	Nome	ID	Nome
0	Precentral L	46	Lingual L	92	Cerebelum C2 L
1	Precentral R	47	Lingual R	93	Cerebelum C2 R
2	Frontal S L	48	Occipital S L	94	Cerebelum 3 L
3	Frontal S R	49	Occipital S R	95	Cerebelum 3 R
4	Frontal S Orb L	50	Occipital Mid L	96	Cerebelum 4 5 L
5	Frontal S Orb R	51	Occipital Mid R	97	Cerebelum 4 5 R
6	Frontal Mid L	52	Occipital Inf L	98	Cerebelum 6 L
7	Frontal Mid R	53	Occipital Inf R	99	Cerebelum 6 R
8	Frontal Mid Orb L	54	Fusiform L	100	Cerebelum 7b L
9	Frontal Mid Orb R	55	Fusiform R	101	Cerebelum 7b R
10	Frontal Inf Oper L	56	Postcentral L	102	Cerebelum 8 L
11	Frontal Inf Oper R	57	Postcentral R	103	Cerebelum 8 R
12	Frontal Inf Tri L	58	Parietal Sup L	104	Cerebelum 9 L
13	Frontal Inf Tri R	59	Parietal Sup R	105	Cerebelum 9 R
14	Frontal Inf Orb L	60	Parietal Inf L	106	Cerebelum 10 L
15	Frontal Inf Orb R	61	Parietal Inf R	107	Cerebelum 10 R
16	Rolandic Oper L	62	SupraMarginal L	108	Vermis 1 2
17	Rolandic Oper R	63	SupraMarginal R	109	Vermis 3
18	Supp Motor Area L	64	Angular L	110	Vermis 4 5
19	Supp Motor Area R	65	Angular R	111	Vermis 6
20	Olfactory L	66	Precuneus L	112	Vermis 7
21	Olfactory R	67	Precuneus R	113	Vermis 8
22	Frontal Sup Medial L	68	Paracentral Lobule L	114	Vermis 9
23	Frontal Sup Medial R	69	Paracentral Lobule R	115	Vermis 10
24	Frontal Med Orb L	70	Caudate L		
25	Frontal Med Orb R	71	Caudate R		
26	Rectus L	72	Putamen L		
27	Rectus R	73	Putamen R		
28	Insula L	74	Pallidum L		
29	Insula R	75	Pallidum R		
30	Cingulum Ant L	76	Thalamus L		
31	Cingulum Ant R	77	Thalamus R		
32	Cingulum Mid L	78	Heschl L		
33	Cingulum Mid R	79	Heschl R		
34	Cingulum Post L	80	Temporal Sup L		
35	Cingulum Post R	81	Temporal Sup R		
36	Hippocampus L	82	Temporal Pole Sup L		
37	Hippocampus R	83	Temporal Pole Sup R		
38	ParaHippocampal L	84	Temporal Mid L		
39	ParaHippocampal R	85	Temporal Mid R		
40	Amygdala L	86	Temporal Pole Mid L		
41	Amygdala R	87	Temporal Pole Mid R		
42	Calcarine L	88	Temporal Inf L		
43	Calcarine R	89	Temporal Inf R		
44	Cuneus L	90	Cerebelum Crus1 L		
45	Cuneus R	91	Cerebelum Crus1 R		

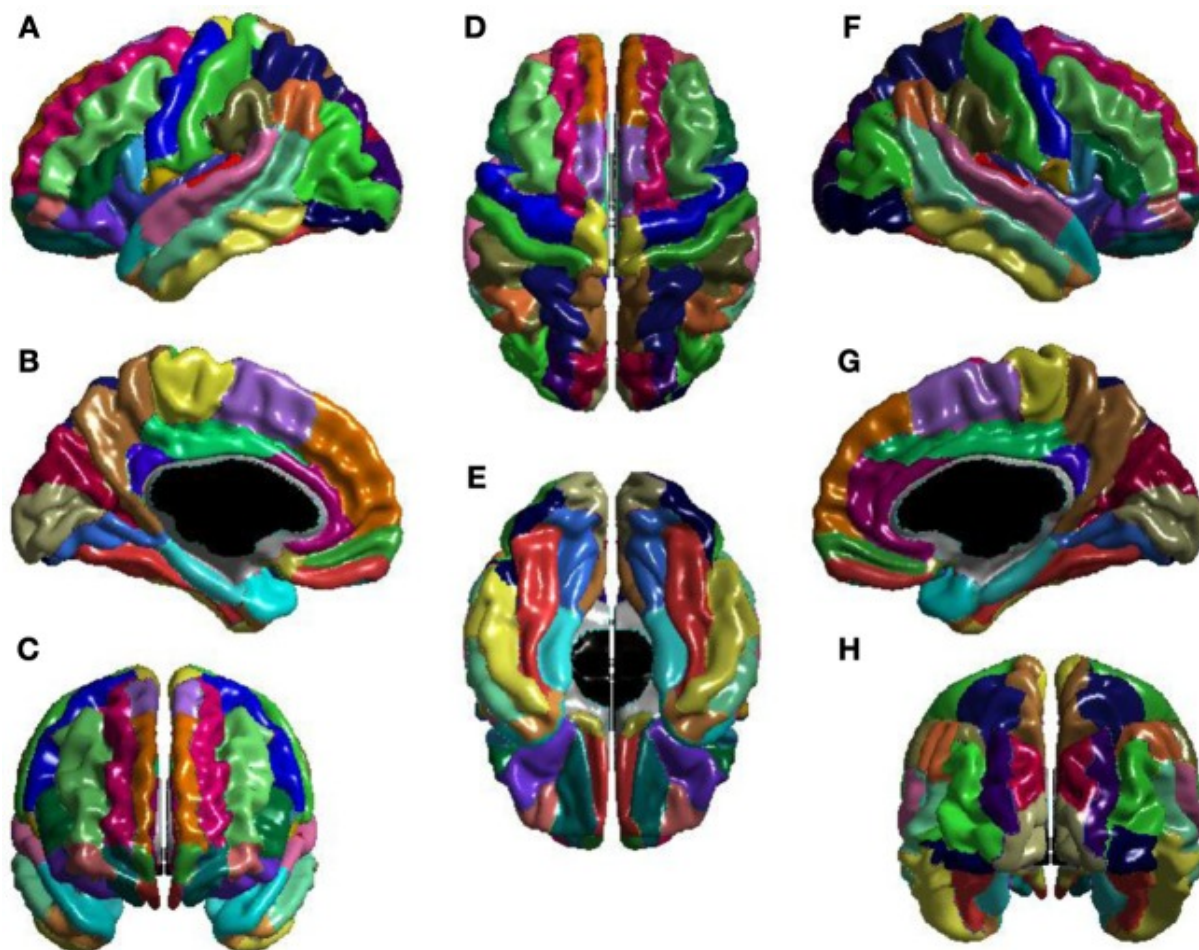


Figura 3 – Atlas AAL: (A) lateral esquerdo, (B) medial esquerdo, (C) anterior, (D) superior, (E) inferior, (F) lateral direito, (G) medial direito, e (H) posterior. - FONTE: (LEWIS et al., 2013)

Seu principal uso é para aquisição de imagens de órgãos, incluindo o cérebro e o restante do sistema nervoso central, mas também pode ser utilizada para a aquisição de imagens de outros órgãos Logothetis (2008).

Com o surgimento da fMRI vieram maiores avanços, permitindo a medição da atividade cerebral de forma indireta, através da medição de alterações do nível de oxigênio no fluxo sanguíneo. Levando no começo da década de 80 a um enorme impacto na neurociência Logothetis (2008). A próxima subseção apresenta maiores detalhes sobre a fMRI.

3.2.2 Imagem de Ressonância Magnética Funcional - fMRI

A fMRI, detecta a concentração do fluxo sanguíneo em certas partes do corpo, e é muito utilizada para estudo do cérebro (BETTS et al., 2013; LOGOTHETIS, 2008). Essa técnica tem contribuído para a compreensão do cérebro assim como anomalias, doenças e transtornos relacionados a ele Betts et al. (2013).

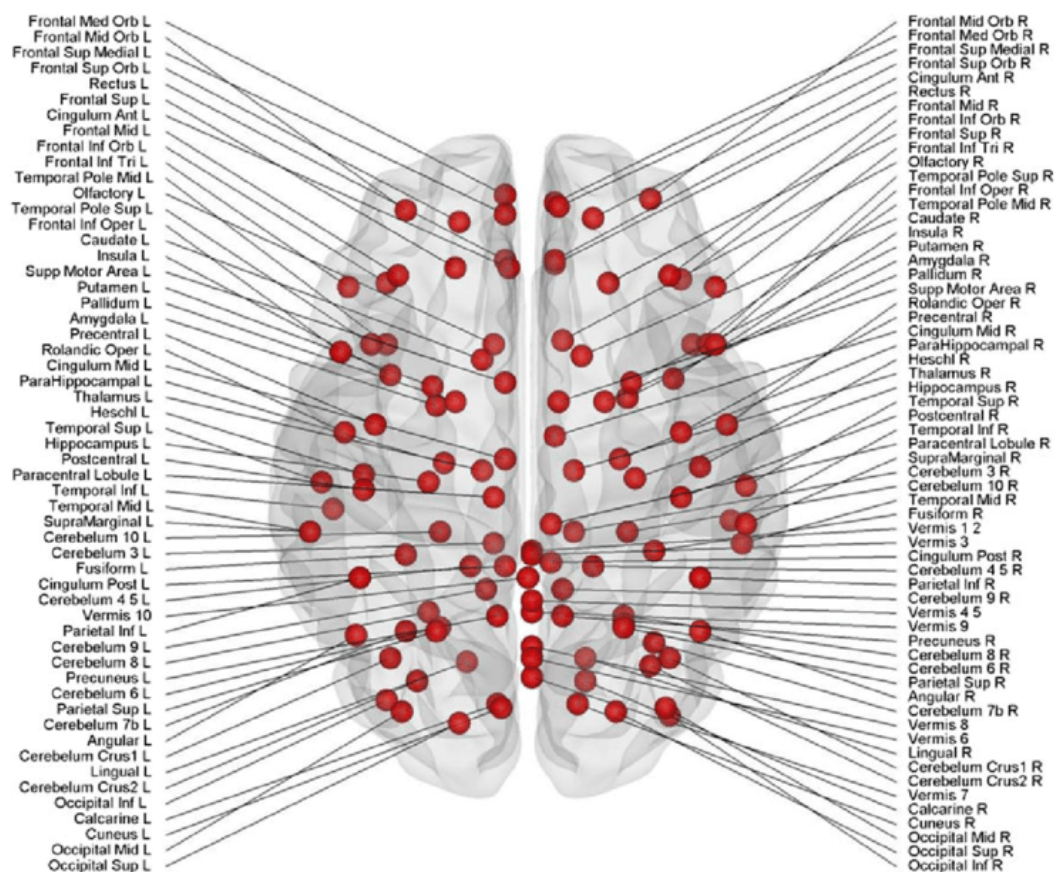


Figura 4 – Regiões Cortical e subcortical (58 por hemisfério; totalizando 116). - FONTE: (CAEYENBERGHS et al., 2012)

A fMRI consiste na medição da diferença do campo magnético DeYoe et al. (1994), geralmente utilizando-se de uma substância para realçar o contraste do oxigênio no sangue. Essa substância possui propriedades que afetam a saturação da hemoglobina no sangue Bandettini et al. (1992a). Isso ocorre devido a uma interação entre as hemácias (eritrócitos) e o plasma ao redor, gerando um microscópico campo magnético, devido a falta de homogeneidade entre ambas Bandettini et al. (1992a). Essa reação é devido a uma troca de prótons de hidrogênio feita pela membrana das hemácias (DEYOE et al., 1994; BANDETTINI et al., 1992a). Esse fenômeno é gerado pela queda de concentração da desoxihemoglobina (estado da hemácia quando não está oxigenada) deixando o tecido da região mais suscetível ao efeito diferencial no campo magnético. Esse efeito gera uma redução de movimentação dos prótons e conseqüentemente aumentando o sinal recebido pelo equipamento de aquisição da imagem Bandettini et al. (1992a).

Um aumento do sinal recebido pelo equipamento, proveniente de determinada região, sugere um aumento do nível de oxigênio da mesma Bandettini et al. (1992a). Isso permite que seja estudado não apenas os processos sensoriais - tais como quais regiões são ativadas para determinada atividade - mas também fazer conjecturas sobre mecanismos relacionados a funções cognitivas mais complexas do cérebro (LOGOTHETIS, 2008).

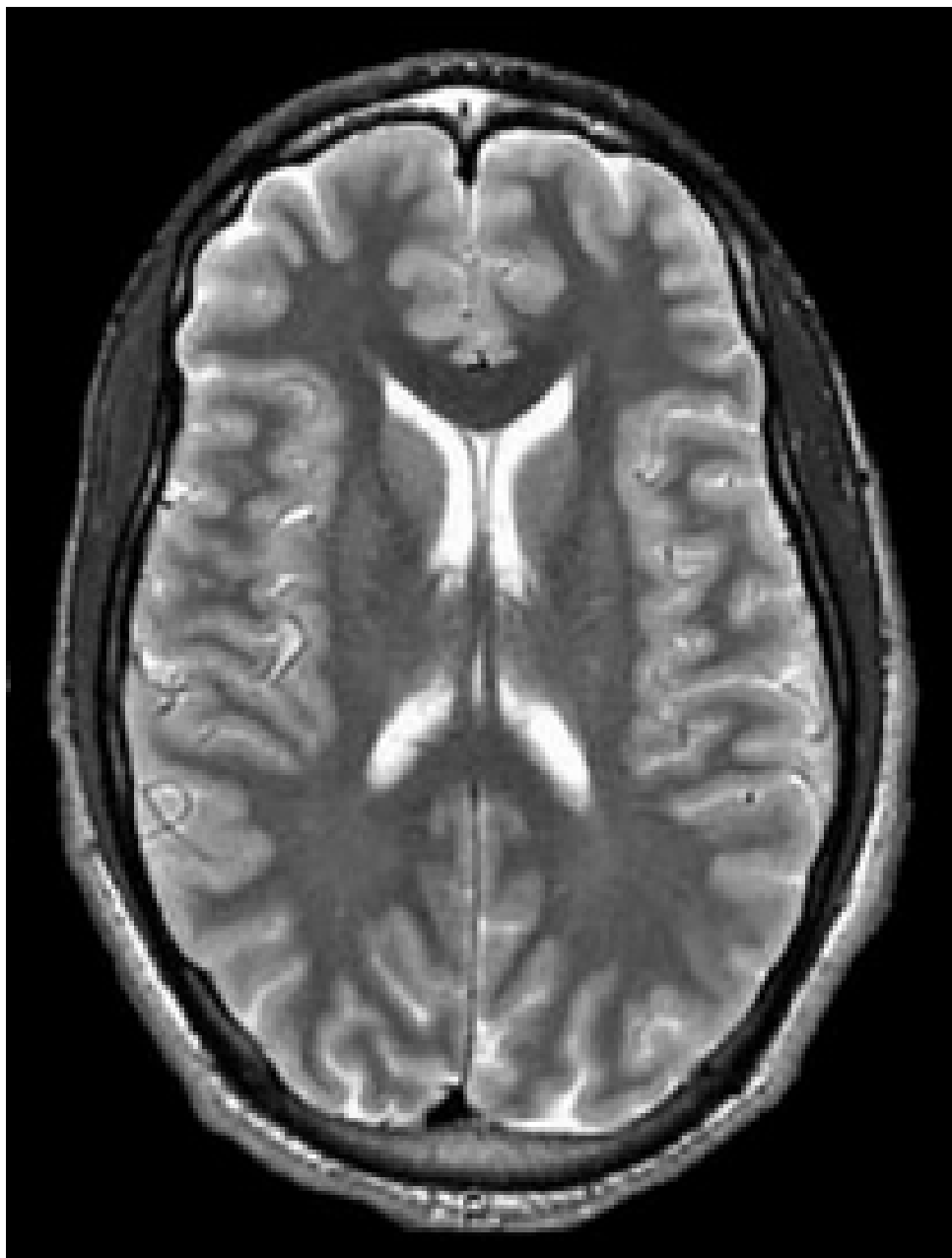


Figura 5 – Exemplo de MRI convencional em corte axial, FONTE: ([TANENBAUM et al., 2017](#))

A capacidade de medir a variação da oxigenação do cérebro humano provida pela fMRI, expandiu de forma significativa as possibilidades de estudo do funcionamento do cérebro ([DEYOE et al., 1994](#); [LOGOTHETIS, 2008](#)). O fato da fMRI permitir a aquisição de centenas de imagens de alta resolução de um mesmo indivíduo, permite análises detalhadas das respostas cerebrais ([DEYOE et al., 1994](#)).

Adicionalmente o fato de ser uma forma não-invasiva, sem riscos significativos para a saúde ([DEYOE et al., 1994](#); [YAN; ZANG, 2010](#); [BETTS et al., 2013](#)), permite que essa tecnologia seja aplicada a uma vasta gama de estudos, como Alzheimer ([WANG et al., 2007](#); [GOLBY et al., 2005](#); [ROMBOUTS et al., 2005](#)), concussão em atletas ([CHEN et al.,](#)

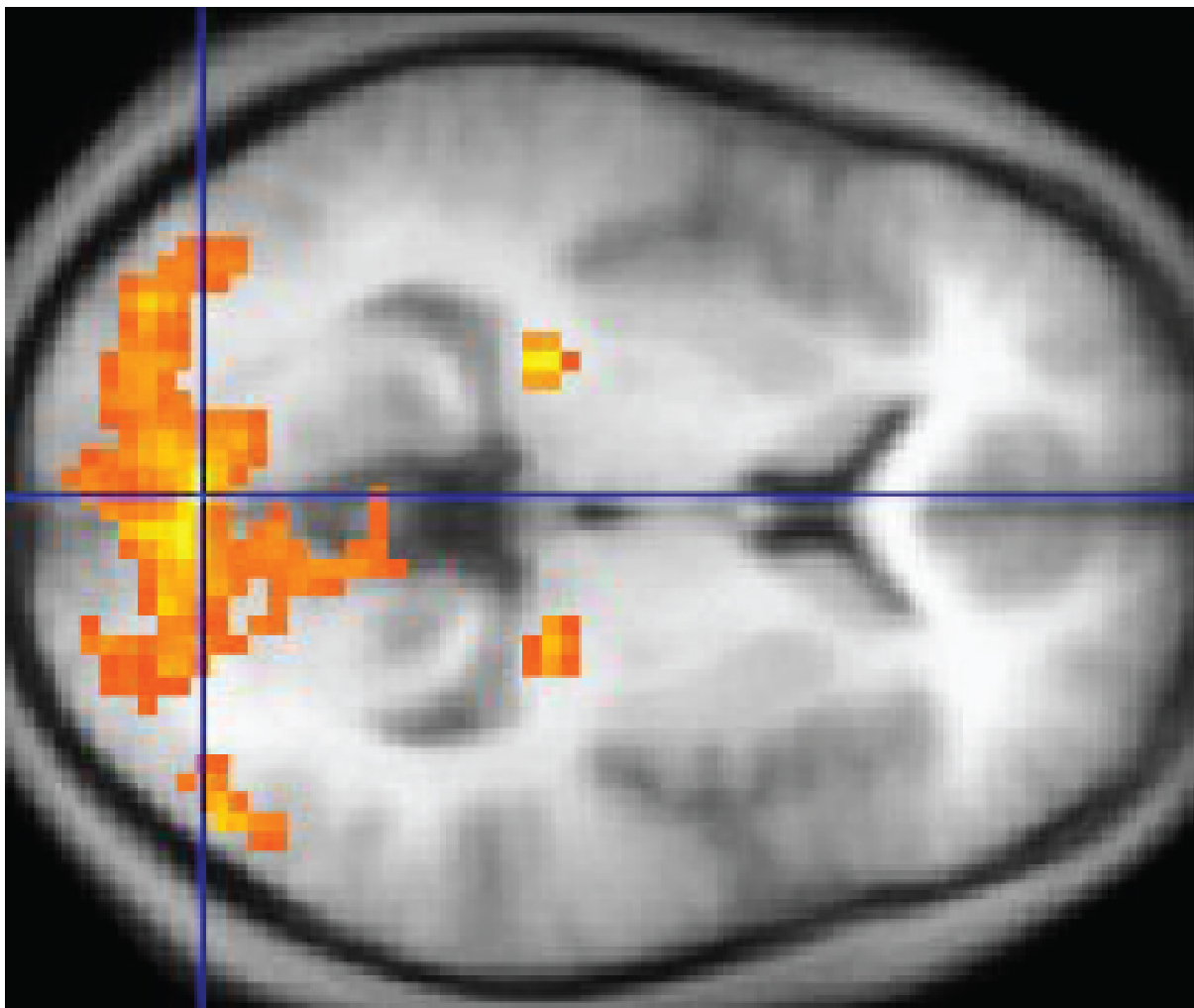


Figura 6 – Exemplo de imagem de ressonância magnética funcional para ativação do córtex durante estímulo visual Fonte: (BETTS et al., 2013), creditado a “Supersuk”/Wikimedia Commons

2004; NAUMAN et al., 2015), recuperação da coordenação motora após derrame (WARD et al., 2003), Transtorno do Espectro Autista (ITAHASHI et al., 2014; ABRAHAM et al., 2017; KIM et al., 2015; Chaitra; Vijaya, 2017), dentre diversos outros.

3.2.3 BOLD

Para um melhor desempenho na aquisição das imagens de fMRI, é comumente utilizado técnicas para realçar o contraste. O contraste mais comumente utilizado para estudos do cérebro é denominado Diferencial de nível de oxigênio no sangue (do inglês *Blood Oxygen Level Differential* - BOLD) (LOGOTHETIS; WANDELL, 2004).

BOLD é um contraste utilizado para avaliar mudanças neurológicas referente a concentração relativa do oxigênio no sangue (LOGOTHETIS; WANDELL, 2004). É importante ressaltar que a resposta *BOLD*, em alguns casos, não ocorre de forma linear em relação a atividade cerebral, apesar de possuir razoável proximidade, conforme (LO-

GOTHETIS; WANDELL, 2004).

A principal resposta do realce *BOLD* reflete a entrada de oxigênio e o processamento neurológico. De forma a ser medido pela correlação entre o Acoplamento Neurovascular. Esse acoplamento é a conexão entre os neurônios e a cadeia vascular, de onde os neurônios extraem sua fonte de energia (LOGOTHETIS; WANDELL, 2004).

O cérebro possui uma alta demanda média de energia quando comparado ao requerido por outros órgãos como o coração por exemplo. De forma a precisar de um suprimento contínuo de energia. Adicionalmente existe uma demanda regional que varia no tempo conforme entrega de oxigênio.

Apesar de bem estabelecida que a alteração da hemodinâmica reflete a atividade neural, a natureza dessa relação ainda não é por completo entendida (LOGOTHETIS; WANDELL, 2004). A existência desta relação é reforçada pelo fato de 90% do processo de consumo de energia no cérebro ser metabolizada de forma aeróbica criando um paralelo entre o consumo de energia e o consumo de oxigênio (LOGOTHETIS; WANDELL, 2004; FOX et al., 1988; BANDETTINI et al., 1992b). É importante ressaltar que diferentes regiões do cérebro podem possuir demandas diferentes, devido a sua estrutura, tamanho e nível de atividade (LOGOTHETIS; WANDELL, 2004).

A relação entre atividade sináptica e o consumo de energia é relatada por trabalhos como (PARRY; MATTHEWS, 2002; CMRR; UĞURBIL, 1997). Baseado nessa relação, diversos trabalhos utilizam o *BOLD* para fazer estimativas da atividade cerebral (LOGOTHETIS; WANDELL, 2004). Uma vez que as atividades vasculares são parte com imenso impacto sobre as imagens adquiridas utilizando o realce *BOLD* regiões com menor vascularização, ou onde essas são mais espaciais, podem apresentar pouca ou nenhuma resposta.

Sumarizando, o contraste *BOLD* reflete as mudanças no volume de sangue, fluxo de sangue e consumo de oxigênio do cérebro (LOGOTHETIS; WANDELL, 2004). Sendo razoável esperar uma correlação entre o contraste *BOLD* e a atividade sináptica de cada região do cérebro (LOGOTHETIS; WANDELL, 2004).

3.3 ABIDE

Um dos grandes desafios na pesquisa de diagnóstico de TEA utilizando aprendizado de máquina é a ausência de materiais que possibilitem uma análise em larga escala de dados. Esse é um problema comum da área de desordens psiquiátricas, conforme apontado por Wolfers et al. (2015). Um grande volume de dados a ser analisado possibilita que a ML alcance melhores resultados, assim como resultados mais confiáveis.

ABIDE é um consórcio que reúne 16 sites internacionais de imagens médicas, dis-

ponibilizando abertamente 539 imagens de indivíduos diagnosticados com TEA e 573 neurotípicos para controle. Todos os indivíduos possuem informações de fMRI em estado de repouso assim como informações de fenótipo. Os dados de fenótipo incluem idade, sexo, e diagnóstico para o TEA, dentre outros. Estes dados visam fomentar a pesquisa sobre o cérebro de indivíduos do espectro autista através do uso de imagens de ressonância magnética (CRADDOCK et al., 2013). ABIDE conta ainda com uma versão pré-processada das fMRI, tendo uma grande diversidade de combinações entre atlas e abordagem de pré-processamento.

3.3.1 Material disponibilizado

O ABIDE atualmente disponibiliza seus dados em um servidor Amazon S3 bucket, os quais são armazenados em pastas separadas conforme pipeline de processamento e estratégia de processamento. Cada uma dessas pastas possui um arquivo separado para cada indivíduo, visando fornecer maior flexibilidade de *download* para os pesquisadores (CRADDOCK et al., 2013).

Os arquivos individuais são resultantes do pré-processamento, conforme atlas (dentre os apresentados na Tabela 4), pipeline (dentre os apresentados na Seção 3.4) e a estratégia escolhida para o pipeline (dentre as apresentadas na Tabela 3). Para as estratégias de pipeline, a Regressão de Sinal global utiliza uma transformação dos valores baseado em uma média global do sinal de cada indivíduo, enquanto o filtro de Band-Pass (0.01 - 0.1 Hz) é uma medida aplicada para corrigir ruídos.

Tabela 3 – Opções de estratégia de pré-processamento disponíveis no ABIDE

Estratégia	Filtro Band-Pass	Regressão de Sinal Global
filt_global	Sim	Sim
filt_noglobal	SIM	Não
nofilt_global	Não	Sim
nofilt_noglobal	Não	Não

ABIDE disponibiliza seus arquivos em 19 versões diferentes de atlas de imagens, Tabela 3, com formas distintas de mapeamento do cérebro, conforme visto na Tabela 4. Cada atlas possui formatos de divisão do cérebro diferentes, variando em quantidades de regiões. Como exemplo o rois_aal possui 116 enquanto o rois_cc200 possui 200.

3.3.2 Trabalhos relacionados

O conjunto de dados ABIDE é amplamente utilizado em pesquisas envolvendo fMRI e TEA. Diversos trabalhos visam a classificação entre duas classes, sendo uma os

Tabela 4 – Atlas disponíveis no ABIDE

Sigla	Nome do atlas
alff	Amplitude of low frequency fluctuations
degree_binarize	Degree centrality with binarized weighting
degree_weighted	Degree centrality with correlation weighting
eigenvector_binarize	Eigenvector centrality with binarized weighting
eigenvector_weighted	Eigenvector centrality with correlation weighting
falff	Fractional ALFF
func_mask	Functional data mask
func_mean	Mean preprocessed functional image
func_preproc	Preprocessed functional image
lfcd	Local functional connectivity density
reho	Regional homogeneity
rois_aal	extraído de Automated Anatomical Labeling atlas
rois_cc200	extraído de Cameron Craddock's 200 ROI parcellation atlas
rois_cc400	400 ROI parcellation atlas
rois_dosenbach160	extraído de Dosenbach160 atlas
rois_ez	extraído de Eickhoff-Zilles atlas
rois_ho	extraído de Harvard-Oxford atlas
rois_tt	extraído de Talarach and Tournoux atlas
vmhc	Voxel-mirrored homotopic connectivity

indivíduos diagnosticados com TEA e outra os indivíduos DT, como em (ABRAHAM et al., 2017; BHAUMIK et al., 2018).

O trabalho Abraham et al. (2017), apresenta um estudo com o objetivo de identificar um biomarcador, que possibilite criar um diagnóstico TEA automatizado. Para isso 871 imagens de fMRI do ABIDE foram utilizadas, incluindo indivíduos TEA e DT. O trabalho foi dividido em cinco etapas, diferenciadas pelo tamanho da amostra, apresentadas do menor para o maior: i) apenas homens destros de 9 a 18 anos, destros, de 3 sites; ii) apenas homens destros de 9 a 18 anos; iii) apenas homens destros; iv) apenas o maior site; e v) todos os indivíduos. As etapas foram implementadas visando verificar o efeito do tamanho total da amostra, assim como a variação do uso de diferentes atlas. O melhor resultado alcançado foi 67% de precisão na classificação entre TEA e DT para a etapa iv, tendo o grupo v com o segundo melhor resultado. Os autores concluíram que o uso de uma amostra maior afeta positivamente os resultados.

Em relação aos atlas, Abraham et al. (2017) utilizou 8 tamanhos de divisão diferente, sendo os melhores resultados obtidos nos de maiores divisões. Conjectura-se que isso tenha ocorrido devido a uma melhor possibilidade de separação das regiões na hora do algoritmo analisar as variações, podendo inferir melhor sobre cada uma.

As regiões apontadas como possíveis biomarcadores por Abraham et al. (2017) foram *temporo-parietal*, *insula* e *inferior parietal*. O trabalho ressalta como limitação o

tamanho da amostra, apontando a necessidade de uma amostra maior para conclusões mais concretas.

O trabalho de [Bhaumik et al. \(2018\)](#), apresenta um estudo com 372 indivíduos, alcançando precisão na classificação entre TEA e DT de 62%. Para isso foi utilizado um algoritmo SVM em conjunto com processo de validação *k-fold cross validation*. Através da implementação do modelo o trabalho aponta regiões onde as diferenças entre indivíduos TEA e DT são mais relevantes na classificação, sendo elas: *Dorsolateral Prefrontal Cortex*, *Somatosensory Association Cortex*, *Primary Auditory Cortex*, *Inferior Temporal Gyrus*, e a área *Temporopolar*. Os resultados obtidos apontam para regiões relacionadas à fala como principais pontos de distinção entre indivíduos TEA e DT.

3.4 Pre-Processamento

O pré-processamento é uma etapa importante para o uso de técnicas de ML. Consiste em transformar a sequência de imagens contidas em uma fMRI em um conjunto de dados numéricos. Esses dados são então processados por um algoritmo de ML. Esse modelo de pré-processamento é aplicado após a escolha de um atlas, e geralmente o resultado deste pré-processamento é uma matriz numérica na qual cada linha representa uma imagem da fMRI e cada coluna uma ROI definida no atlas.

Alguns trabalhos utilizam pré-processamentos próprios, mas a maioria dos trabalhos encontrados na literatura sobre classificação de TEA usando fMRI e ML utilizam dados pré-processados pelos seguintes algoritmos:

- Connectome Computation System (CCS);
- Configurable Pipeline for the Analysis of Connectomes (CPAC);
- Data Processing Assistant for Resting-State fMRI (DPARSF);
- Neuroimaging Analysis Kit (NIAK).

Cada um desses pré-processamentos tem uma sequência de ações efetuadas sobre os dados originais para transformá-los em dados a serem processados por técnicas de ML. Algo que todos têm em comum é o tipo de dado obtido como resultado final, que é a matriz numérica já citada. Nos quatro modelos a matriz resultante do processo é constituída de números Reais, onde cada valor representa a taxa de oxigenação de uma região do cérebro em um determinado momento do escaneamento. Também possuem em comum o uso de linhas da matriz para representar o tempo de aquisição da fMRI e as colunas para representar cada ROI. Todos esses arquivos estão disponíveis para *download* em ([CRADDOCK et al., 2013](#)), juntamente com detalhamentos das operações efetuadas para alcançar a matriz final.

Cada um dos pré-processamentos existentes no ABIDE também possui uma variação de configurações diferentes, chamadas estratégias de pipeline. Essas configurações são referentes a:

- Filtro de ruído: remove ruídos potencialmente causados durante a leitura feita pelo equipamento, como por exemplo interferência nas ondas;
- Padronização de valor: transforma os valores globais encontrados em valores relativos ao invés de valores absolutos. Por exemplo, o maior valor de oxigênio encontrado na leitura de um indivíduo será convertida para 1.000 e o menor valor para -1.000, assim os outros valores encontrados serão padronizados para corresponder a valores relativos dentro dessa escala. Isso significa que o mesmo valor encontrado em ressonância de dois indivíduos não significa necessariamente que ambos possuíam o mesmo nível de oxigenação.

Essas opções de configuração são escolhas binárias que podem ser feitas em qualquer combinação, tal como Filtro de ruído: sim e Padronização de valor: sim; Filtro de ruído: não e Padronização de valor: sim; e assim por diante (CRADDOCK et al., 2013) e podem ser vistas na Tabela 3.

Diversos trabalhos na literatura apresentam resultados satisfatórios para a classificação de TEA usando ML e fMRI utilizando o DPARSF (BI et al., 2018a; SUBBARAJU et al., 2017; JUN; SUK, 2017; ZHU et al., 2016), um dos modelos disponíveis no ABIDE. Dessa forma, por ser vastamente utilizado, o DPARSF pode ser considerado um modelo validado para o tipo de pesquisa, e o fato de estar disponível no ABIDE, torna possível a reprodução de experimentos que o utilizam. Assim sendo, essas foram as razões pelas quais esse foi o modelo selecionado para utilização no presente trabalho.

3.4.1 DPARSF

Essa subseção apresenta o DPARSF, presente em (YAN; ZANG, 2010) e consiste em um *pipeline* de pré-processamento para fMRI, ou seja, uma sequência de passos efetuados para transformar os dados provenientes de uma fMRI em uma matriz a ser interpretada por um algoritmo de ML. DPARSF consiste de duas etapas básicas seguidas de uma configuração opcional. As duas etapas básicas são Pré-processamento Estrutural e Pré-processamento Funcional, respectivamente, nessa ordem. Cada passo é mais detalhado nos dois próximos tópicos 3.4.1.1 e 3.4.1.2. A etapa de configuração opcional ocorre após as duas etapas básicas e consiste na escolha de uso de filtro de ruído e padronização de valores.

3.4.1.1 Pré-processamento Estrutural

O pré-processamento estrutural consiste em quatro etapas que visam padronizar a parte estrutural do cérebro contido na fMRI, ou seja, a composição de tecidos, assim como o seu posicionamento. Essa etapa visa a comparação entre dois cérebros diferentes. As quatro etapas são:

- Remoção de crânio;
- Segmentação do cérebro por tecido;
- Restrição da segmentação de tecidos;
- Normalização das regiões.

A primeira etapa consiste na remoção da imagem do crânio da fMRI, possibilitando que a parte interna seja analisada, em seguida o cérebro é segmentado conforme os tipos de tecido. Em seguida, o tecido segmentado é convertido para uma padronização conforme sua localização espacial. Por fim, é feita uma normalização distribuindo as regiões em espaços (*voxels*) de 3mm por região.

O resultado desse processo é uma máscara representando o cérebro, permitindo associar o nível de oxigenação encontrado a uma região cerebral. Ou seja, um mapa que permita definir qual região do cérebro está representada em um determinado ponto da imagem.

3.4.1.2 Pré-processamento Funcional

O pré-processamento funcional consiste em uma série de etapas, utilizadas para possibilitar uma interpretação mais correta do fluxo de oxigênio. Essa etapa realiza as seguintes funções:

- Exclusão dos primeiros volumes;
- Correção temporal;
- Correção de movimento;
- Remoção de ruído;
- Normalização da intensidade.

O primeiro passo dessa etapa é a eliminação dos 4 primeiros volumes. O objetivo principal é evitar que ruídos causados pela inicialização do equipamento influenciem no resultado final da leitura (CRADDOCK et al., 2013).

O segundo passo, correção temporal, dá-se uma vez que durante o escaneamento da fMRI, a máquina faz uma leitura latitudinal do cérebro, mas ele não é lido totalmente de uma vez. Assim, a leitura feita nas extremidades pode possuir diferença de alguns segundos entre si, dependendo do equipamento utilizado. Dessa forma, uma das ações efetuadas no pré-processamento é uma correção temporal dessa leitura, que busca minimizar o efeito dessa diferença de tempo.

Além disso, ainda é possível que durante a leitura feita pelo equipamento, a pessoa submetida ao procedimento, tenha efetuado pequenos movimentos, mesmo que de forma involuntária, uma vez que até mesmo a respiração pode gerar uma movimentação que gere pequeno deslocamento da posição inicial (SPECK; HENNIG; ZAITSEV, 2006; JONES; BANDETTINI; BIRN, 2008). Para reduzir a interferência, proveniente dessa movimentação, uma das etapas do pré-processamento, consiste na correção dessa movimentação (POWER et al., 2012). O objetivo é evitar que o algoritmo de ML receba uma leitura que reflete uma movimentação do indivíduo e considere como uma alteração no fluxo de oxigênio (SATTERTHWAITE et al., 2012).

Para exemplificar as implicações da movimentação durante a aquisição da imagem, suponha-se que o objetivo seja analisar a região α , β e ζ , no instante 1 e 2. Para o instante 1 as regiões registram os seguintes valores: $\alpha = 100$, $\beta = 200$ e $\zeta = 300$, e estão ocupando as coordenadas $\alpha = (0, 0)$, $\beta = (1, 1)$ e $\zeta = (2, 2)$. Para o instante 2, os valores das regiões são os mesmos, entretanto devido a uma movimentação as coordenadas ocupadas por cada região foram alteradas para $\alpha = (1, 1)$, $\beta = (2, 2)$ e $\zeta = (3, 3)$. Note que com essa movimentação a região α está ocupando a mesma posição que a região β estava no instante 1, enquanto β ocupa uma região que ζ ocupava no instante 1. Sem utilizar um processo de correção para movimentação, quando for analisado o instante 2, será considerado que a região β estava registrando uma queda de valor, saindo de 200 para 100, quando na verdade ela permaneceu igual, e o mesmo para a região ζ registrando uma falsa queda de 300 para 200. Esses eventos têm impacto direto no resultado do algoritmo.

A emissão de ondas elétricas, devido a ativação do equipamento, assim como outras interferências externas, podem influenciar na leitura, gerando pequenos ruídos. Esses ruídos podem ser interpretados equivocadamente como um valor correto, influenciando o algoritmo de ML a um resultado equivocado. Visando evitar esse tipo de erro, o pré-processamento elimina ruídos contidos na leitura.

Por fim, visando facilitar a comparação entre duas fMRI, é feita uma normalização da intensidade de oxigênio. Essa etapa consiste em transformar os valores brutos da leitura para valores relativos, no qual a média final é de 10 mil. Assim, a normalização feita, transforma os valores absolutos de forma que a soma de todas as regiões, para cada instante, tenha a média de 10 mil.

3.4.1.3 Resultado do Pre-Processamento

O produto resultante do algoritmo DPARSF é uma matriz (X, Y) , onde X é o número de colunas, e Y o número de linhas. Sendo gerada uma matriz por indivíduo onde cada linha representa um instante de tempo e cada coluna representa uma ROI, de acordo com o atlas escolhido. Para esse trabalho foi escolhido o ALL, com 116 ROI's, assim o X irá variar de 0 a 115. O número de linhas (Y) pode variar de uma fMRI para outra, mesmo utilizando o mesmo atlas, uma vez que esta depende do tempo total de aquisição da imagem. Por exemplo, em uma matriz DPARSF, o valor (X_i, Y_j) representa o nível de oxigênio da ROI i no instante j . Um exemplo de como essa matriz é composta é apresentado na Tabela 5, onde as 10 primeiras linhas e as cinco primeiras colunas de uma matriz são apresentadas, sendo a primeira linha um *label* da região e as demais os valores de oxigenação em um dado instante para a região daquela coluna.

Tabela 5 – Tabela Parcial Resultante do processamento DPARSF

#2001	#2002	#2101	#2102	#2111
-0.3234579861	-0.6951230168	0.0302159991	0.0172370002	0.4168649912
-0.3830829859	-1.1193660498	0.0989660025	0.0745910034	0.4997729957
-0.2349420041	-1.1271400452	0.0174979996	0.0037509999	0.3267379999
0.0175519995	-0.6673780084	-0.1136149988	-0.1676229984	0.0304879993
0.186214	0.0088099996	-0.1031569988	-0.2741340101	-0.2290749997
0.1825540066	0.5698459744	0.0972580016	-0.1905840039	-0.3765740097
0.1009519994	0.8479089737	0.314576	0.0216230005	-0.4025929868
0.1125459969	0.8928319812	0.3400399983	0.1850560009	-0.3098730147
0.2861660123	0.8442119956	0.1611009985	0.2136459947	-0.1292060018

4 Métodos

O presente Capítulo apresenta os métodos utilizados para elaboração do trabalho. A Seção 4.1 apresenta o material usado como entrada para o algoritmo de ML e a Seção 4.2 apresenta os valores mínimos utilizados para selecionar as possíveis regiões com relevância suficiente para classificar TEA e Autistas. Por fim a Seção 4.3 apresenta o método de validação selecionado para o trabalho.

4.1 Conjunto de Dados

O conjunto de dados adquirido através do site ABIDE consiste em 1100 arquivos fMRI disponibilizados de forma anônima, identificadas por um número ID, e um arquivo constituído de informações de fenótipos, que considera o mesmo número ID identificador. Dentre as informações fenotípicas destacam-se, sexo, idade, QI, diagnóstico TEA e nota ADOS. As informações de fenótipo não estão completas para todos os indivíduos, sendo essa uma das restrições da seleção de indivíduos. Ainda, o material gerado pelo DPARSF é incompatível com o tipo de entrada esperada pelo algoritmo de ML selecionado para esse trabalho, tornando necessário um processamento adicional como apresentado na Subseção 4.1.2. Por fim, na Subseção 4.1.1 é apresentado o método utilizado para agrupar as regiões do cérebro a fim de buscar diferença funcional entre os grupos.

4.1.1 Seleção de Indivíduos

Para verificar a hipótese proposta neste trabalho, os arquivos adquiridos do ABIDE foram filtrados, de forma a selecionar apenas os indivíduos diagnosticados como TEA-NA ou Autista os quais a informação de nota ADOS estava presente no arquivo de fenótipo. Dessa forma, todos os arquivos referentes a neurotípicos ou com inexistência de nota ADOS foram descartados. Resultando em 123 arquivos fMRI, sendo 24 TEA-NA conforme, Tabela 6 e 99 autistas, conforme Tabela 7.

Tabela 6 – Grupo Indivíduos TEA-NA

Índice de indivíduos do ABIDE					
51457	51461	51470	51471	50803	50958
50960	50967	50976	50981	50983	50991
50993	50995	51001	51007	51008	51011
51018	51019	51021	51026	51034	50142
50145	50146	50152	50025	50056	50057
50182	51210	51211	51224	51229	50499

Tabela 7 – Grupo Indivíduos Autistas

Índice de indivíduos do ABIDE					
50791	50792	50795	50798	50799	50800
50802	50804	50823	50824	50825	50954
50955	50956	50964	50965	50968	50969
50970	50972	50974	50977	50978	50979
50982	50984	50985	50986	50987	50988
50989	50990	50992	50998	51000	51002
51003	51006	51009	51010	51012	51014
51032	51033	51035	50143	50144	50148
50150	50153	50007	50014	50027	50029
50183	50184	50186	50188	50189	50190
50191	50212	51206	51208	51212	51214
51216	51217	51218	51221	51222	51223
51234	51235	51236	51239	51240	51241
51248	51249	51293	51294	51295	51298
51301	51302	50477	50486	50500	50504
50515	50516	50518	50519	50520	50521
50524	50528				

A distribuição fenotípica das amostras utilizadas, em relação a idade e QI, pode ser vista na Tabela 8, enquanto a distribuição em relação ao sexo pode ser vista na Tabela 9.

Tabela 8 – Distribuição por Fenótipo

Grupo	Idade Media	Idade Máximo	Idade Mínimo	Idade Desvio	QI Media	QI Máximo	QI Mínimo	QI Desvio
TEA-NA	16.47	38.76	8.0	7.90	108.35	132.0	76.0	13.66
Autista	17.63	55.4	7.13	8.91	104.81	148.0	65.0	16.76

Tabela 9 – Distribuição por Sexo

Grupo	Total	Homem	Mulher
TEA-NA	36	32	4
Autista	166	152	14

4.1.2 Adequação da Matriz DPARSF ao SVC

A matriz adquirida pelo pré-processamento (matriz DPARSF), precisou de processamentos adicionais para adequar a entrada esperada pelo algoritmo de ML. Esse trabalho utilizou-se do SVC, apresentado no Capítulo 5, esse algoritmo espera um vetor como entrada. Essa subseção descreve os passos feitos para adequar a matriz DPARSF á entrada esperada pelo SVC.

O primeiro passo foi a remoção da primeira linha do arquivo, contendo o identificador de cada coluna. Em seguida, duas opções de tratamento de dados para o restante

da informação foram avaliadas: i) construir um vetor onde cada posição da matriz (X_i, Y_j) é adicionada em uma posição (Z_{i+i*j}) ; e ii) adquirir os valores máximo, mínimo, média, e mediana para cada ROI, resultando num vetor:

$$(Z_{a^{max}}, Z_{a^{min}}, Z_{a^{med}}, Z_{a^{avg}}, \dots, Z_{b^{max}}, Z_{b^{min}}, Z_{b^{med}}, Z_{b^{avg}}),$$

onde a e b são respectivamente, o primeiro e o último ID das ROI's que compõe cada uma das Set of Regions (SoRs) presentes na Tabela 10.

Ambas as opções de tratamento de dados possuem vantagens e desvantagens. A primeira possui um processamento de transformação mais simples, mas em contrapartida possui uma maior necessidade de processamento computacional para a execução do algoritmo de ML, uma vez que uma maior quantidade de dados deverá ser analisada. Enquanto a segunda opção, possui a desvantagem de demandar um maior processamento prévio para transformar a matriz em um vetor, adquirindo dados de cada indivíduo para extrair os quatro valores para cada ROI. Entretanto, devido ao tamanho reduzido do vetor resultante da segunda opção, o algoritmo de ML fará uma análise em um conjunto de dados reduzido. Dessa forma, visando a escalabilidade, a segunda opção foi escolhida para esse trabalho.

Ademais, o fato de cada site do ABIDE possuir diferente quantidade de linhas (devido a diferentes tempos de aquisição), a escolha da opção i se torna ainda mais inviável, pois o SVC demanda vetores de mesmo tamanho para cada indivíduo.

Adicionalmente, ambas possuem a desvantagem de não permitir uma análise adequada de variação temporal, uma vez que o SVC faz a análise de cada posição sem considerar um possível deslocamento temporal. Para exemplificar esse deslocamento temporal, imagine que a mesma pessoa realizou duas aquisições de fMRI, na qual os resultados para a aquisição 1 foi 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 enquanto para o segundo foi 5, 6, 7, 8, 9, 10, 9, 8, 7, 6, uma vez que existe um deslocamento temporal, o algoritmo de SVC não será capaz de reconhecer esse deslocamento como sendo de indivíduos da mesma classe. Dado isso, o volume adicional de dados que seriam disponíveis pela opção i em relação a opção ii, não necessariamente teriam o efeito desejado.

4.1.3 Seleção de Regiões

Uma vez que o objetivo do estudo é descobrir regiões com potencial existência de diferenças funcionais entre TEA-NA e autista, visando facilitar a classificação do SVC, foram criados grupos de ROI's (do inglês Set of Regions - SoRs) a serem usadas como entradas para o SVC, conforme Tabela 10. Onde o grupo ID é um identificador do conjunto e ROI's ID é o identificador da ROI conforme a Tabela 2.

Dessa forma, o processo de treinamento e validação foi realizado para cada uma das SoRs dessa tabela, com o objetivo de identificar potenciais regiões com diferença

funcional entre TEA-NA e autistas.

O critério adotado para a criação dessas SoRs foi agrupar as regiões de mesmo nome (p.e. *Precentral Left* e *Precentral Right* como uma SoRs, *Angular left* e *Angular right* como outra SoRs, etc). Esse processo resultou em 35 SoRs contendo grupos de ROI's conforme suas regiões cerebrais. Adicionalmente, foi criada uma SoRs correspondendo a todas as regiões do cérebro.

Essa abordagem foi selecionada para permitir comparar os resultados obtidos com a função de cada região, assim como a relação dessa função com o TEA.

Tabela 10 – IDs das SoRs e os respectivos IDs das ROI's provenientes do atlas AAL

ID SoRs	ID das ROI's	ID SoRs	ID das ROI's
0	[0, 1]	18	[48, ...,53]
1	[2, ..., 5]	19	[54, 55]
2	[6, ..., 9]	20	[56, 57]
3	[10, ..., 15]	21	[58, ..., 61]
4	[16, 17]	22	[62, 63]
5	[18, 19]	23	[64, 65]
6	[20, 21]	24	[66, 67]
7	[22, 23]	25	[68, 69]
8	[24, 25]	26	[70, 71]
9	[26, 27]	27	[72, 73]
10	[28, 29]	28	[74, 75]
11	[30, ..., 35]	29	[76, 77]
12	[36, 37]	30	[78, 79]
13	[38, 39]	31	[80, ..., 89]
14	[40, 41]	32	[90, ..., 107]
15	[42, 43]	33	[108, ..., 115]
16	[44, 45]	34	[0, ..., 115]
17	[46, 47]		

$[X, \dots, Y]$ é uma sequência de incremento um, onde X é o limite inferior, enquanto Y o limite superior (p.e., $[1, \dots, 4]$ pode ser lido como $[1, 2, 3, 4]$).

4.2 Restrições de Precisão, Sensibilidade e Especificidade

Para o presente trabalho, foram impostas restrições à precisão (ACC), especificidade (SPC), e Sensibilidade (SNS), utilizando-se de valores mínimos exigidos para considerar uma diferença funcional entre as duas subclasses de TEA. O ponto de corte foi 60%, este limite foi selecionado com base nos valores obtidos por outros estudos de classificação TEA vs. DT (Chaitra; Vijaya, 2017; ABRAHAM et al., 2017; ZU et al., 2016; DODERO et al., 2015; Dodero et al., 2015; BHAUMIK et al., 2018).

Portanto, foram descartados resultados cuja ACC, SPC, ou SNS foram menores que 60%. Assim, os resultados apresentados no Capítulo 6 são apenas das SoRs que

cumpriram os requerimentos mencionados.

4.3 Processo de Validação

Para validar o modelo foi utilizado o processo de *k-fold cross-validation* (KCV) (BENGIO; GRANDVALET, 2004; RODRIGUEZ; PEREZ; LOZANO, 2009; FUSHIKI, 2011). Esse processo consiste em dividir a amostra a ser testada em k conjuntos, no qual um conjunto é separado para teste do algoritmo de ML e os outros $k - 1$ conjuntos utilizados para o treinamento do algoritmo de ML. Em seguida, o processo é repetido escolhendo-se um conjunto diferente para ser o de teste e utilizando o restante para treinamento. Esse processo é repetido até que todos os conjuntos tenham sido utilizados exatamente uma vez como conjunto de teste, e então é feita uma média dos resultados.

O processo de Cross-validation permite demonstrar o quanto a precisão da classificação pode ser generalizada (MARCOT; HANEA, 2020). Ou seja, permite verificar se o modelo realmente encontrou alguma característica da amostra que permite uma classificação que resolva o problema proposto, ou se apenas encontrou uma característica da amostra em questão.

Sem a utilização de um modelo de validação tal como o KCV, pode ocorrer o que é chamado *overfit* (MARCOT; HANEA, 2020). O processo de *overfit*, pode ser dividido em duas categorias i) utilizar uma abordagem mais complexa que o modelo demanda, ou seja um modelo mais flexível que o necessário; e ii) incluir no modelo componentes desnecessários para a classificação (HAWKINS, 2004). O *overfit*, pode levar o modelo a apresentar resultados piores quando usado em uma amostra diferente da qual o modelo foi treinado (HAWKINS, 2004).

Dessa forma, foi utilizado um KCV para validar o modelo proposto, com a escolha do valor de k como 10, que para o tamanho da amostra utilizada neste trabalho resulta em um valor aceitável (MARCOT; HANEA, 2020).

Visando evitar o enviesamento da amostra, foram adotados dois critérios adicionais para separação da amostra dentre os 10 *fold's* utilizados no experimento:

- Quantidade de indivíduos de cada classe por conjunto;
- Distribuição de QI por conjunto.

O primeiro critério visa garantir que todos os treinamentos e testes possuam uma quantidade relevante de indivíduos de cada grupo, de forma a evitar que o algoritmo de ML tenha um enviesamento de sempre responder à mesma classe durante a classificação, devido ao fato de existir apenas uma classe em um determinado conjunto. Esse enviesamento pode gerar uma falsa noção de precisão do algoritmo de ML, uma vez que ele não

	F1	F2	F3	...	F10
I1	Validação	Treinamento			
I2	Treinamento	Validação	Treinamento		
I3	Treinamento		Validação	Treinamento	
...					
I10	Treinamento				Validação

Figura 7 – Processo de Validação e treinamento utilizando *k-fold cross-validation*, com valor de $k = 10$, em cada nova interação (linha) é selecionado um *fold* diferente (coluna) para ser usado como amostra de validação, enquanto os outros 9 ($k-1$) são utilizados como amostra de treinamento. O processo é repetido 10 vezes (valor de k)

está sendo treinado para identificar uma classe, e sim a responder sempre classificando qualquer entrada como a classe existente no conjunto de treinamento.

O segundo critério visa garantir a capacidade de generalização do algoritmo de ML, dividindo os conjuntos de forma que a distribuição de QI dos indivíduos fosse de alguma forma equilibrada. Para isso, optou-se por ordenar cada classe por QI dos indivíduos, de forma decrescente, e em seguida realizar uma distribuição de 1 a 1 nos conjuntos. De forma que o maior QI de cada classe foi adicionado ao conjunto 1, o segundo maior ao conjunto 2, o décimo primeiro ao conjunto 1, o décimo segundo ao conjunto 2, e assim sucessivamente e de forma análoga para ambas as classes (TEA não autista e autista).

5 Aplicação

Esse capítulo contém o processo utilizado para a validação do experimento (Subseção 4.3) e a modelagem do algoritmo SVC utilizado para classificação (Subseção 5.1).

5.1 Support Vector Classification

A técnica de ML selecionada para este estudo foi o SVC, uma categoria de SVM. O SVM foi selecionado após uma revisão sistemática da literatura, entre os 50 artigos relevantes ao tema publicados entre 2005 e 2018, sendo a técnica mais utilizada na classificação entre neurotípicos e TEA.

O SVM é uma técnica de aprendizado supervisionado que consiste em um conjunto de métodos de ML, sendo utilizado para classificação (SVC) e regressão (SVR) (PEDREGOSA et al., 2011). Dessa forma, para classificação, o SVM pode ser configurado para previsão de classes de forma binária ou número real de proximidade. No primeiro caso, o retorno é um número representando a classe a qual o objeto está mais próximo. Enquanto no segundo, o retorno é um número real no intervalo fechado entre os números que representam cada classe, de forma que quanto mais próximo ao valor de uma classe, mais semelhante ele é ao conjunto em questão. Por exemplo, para duas classes 1 e 2 o SVC terá como saída os valores 1 ou 2, enquanto o SVR pode retornar valores como 1, 1.3, 1.8, 2 dependendo da similaridade da entrada com os valores de cada classe que foram usados para o treinamento, assim como do *kernel* escolhido para decisão do SVM.

De forma prática, O SVC cria uma superfície multi-planar, com X dimensões, onde X é o número de entradas (também chamadas *features*), em seguida, após uma etapa de treinamento (Subseção 5.1.2) e validação (Subseção 5.1.3), cria uma linha que divide a superfície e associa cada área dentro dessa divisão a uma classe. Dessa forma, para cada nova entrada, cria-se um ponto nas coordenadas referentes às suas *features*, classificando-as conforme a área a qual o ponto referente foi criado.

Tanto o SVC quanto o SVR podem ser adaptados para a classificação de mais de duas classes, entretanto essa aplicação do SVM está fora do escopo do presente trabalho.

Para a implementação do SVC utilizou-se a API *scikit-learn 0.23.1*, implementada em *Python*, apresentada em Pedregosa et al. (2011). Essa implementação possui cinco opções de *kernels* diferentes, assim como diversas opções de configuração.

5.1.1 Kernel

O kernel consiste basicamente na função matemática que será utilizada para determinar a separação entre os conjuntos a serem classificados, de forma que cada característica a ser utilizada na classificação possuirá um peso. Por exemplo, um kernel linear poderia ser $A.x_1 + B.x_2$ onde A é o peso para a entrada 1, x_1 é a posição 0 do vetor de entrada, B é o peso definido para a entrada 2 e x_2 é a segunda posição do vetor de entrada, de forma que tanto o valor de A quanto B serão definidos pelo SVC em sua fase de treinamento, visando a melhor margem de acerto possível.

As cinco opções de escolha para kernel são: linear, *poly*, RBF, *sigmoid*, *precomputed*. Entretanto, o presente trabalho utilizou-se apenas das 4 primeiras opções, as quais são explicadas a seguir.

- **Linear:** utiliza-se de uma função linear para a classificação, tal qual: (x, x') ;
- **Poly:** utiliza-se de uma função polinomial, tal qual: $(\gamma(x, x') + r)^d$ na qual d é o parâmetro de configuração *degree* e r o parâmetro de configuração *coef0*, ambos definidos antes da execução;
- **Radial Basis Function (RBF):** utiliza-se de uma função radial, tal qual: $\exp(-\gamma\|x - x'\|^2)$ na qual γ é o parâmetro *gamma* e deve ser maior que 0 enquanto \exp é uma função exponencial;
- **Sigmoid:** Utiliza-se de uma função tangencial, tal qual: $\tanh(\gamma(x, x') + r)$ na qual r é o parâmetro *coef0*.

Na Figura 8 são apresentados exemplos de como cada *kernel* cria a separação das classes, com a linha vermelha representando a linha de separação baseada na função escolhida e os pesos de treinamento, de forma que os objetos (representados pelas bolas brancas) que estiverem na área verde serão classificados em uma classe e os que estiverem na área azul serão classificados em outra classe. Para o *kernel* linear sempre será usada uma linha Linear, *Poly* uma linha polinomial, *RBF* uma linha radial e *Sigmoid* uma linha tangencial.

Para se obter a linha de separação utilizada na classificação, além da função que é predeterminada. O algoritmo atua em duas fases: Treinamento e Validação.

Antes de dar início a etapa de Treinamento, a amostra é dividida em duas: uma doravante denominada amostra de treinamento e outra doravante denominada amostra de validação. A API utilizada permite configurar a proporção de divisão da amostra, para o presente trabalho foi utilizado 70% da amostra para treinamento e 30% para validação.

Adicionalmente, a opção de balanceamento de amostra foi utilizada. Essa opção visa ponderar os pesos utilizados na decisão conforme o tamanho de cada classe, evitando

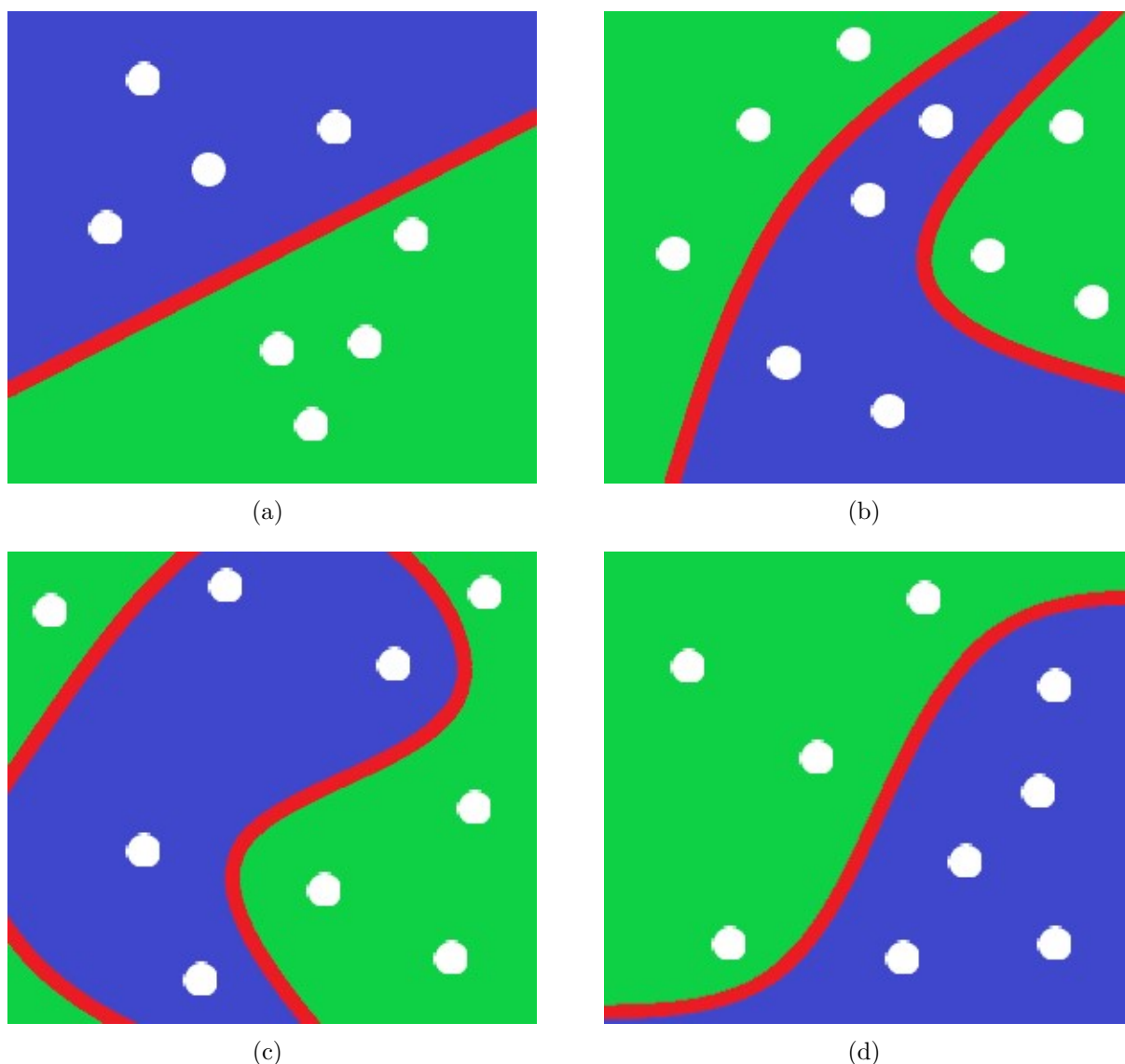


Figura 8 – Exemplo de separação usando *kernel* Linear (a), *kernel* Poly (b), *kernel* RBF (c) e *kernel* Sigmoid (d)

assim que o algoritmo fique enviesado devido a uma classe possuir um número de indivíduos muito maior que a outra. Como é o caso da amostra utilizada neste trabalho, uma vez que a classe TEA-NA representa apenas 19.5% da amostra total.

5.1.2 Treinamento

Primeiramente ocorre uma fase de treinamento, na qual o SVC recebe um conjunto de vetores contendo as características a serem analisadas, assim como um vetor possuindo a classe a qual cada vetor recebido pertence. A partir desses dados, SVC computa valores de pesos para cada posição do vetor, usando uma equação pré-definida (conforme o *kernel*), de tal forma que quando cada valor do vetor seja multiplicado pelo peso e inserido dentro da equação, gere a linha de separação, a partir da qual o maior número de objetos do

treinamento estejam corretamente na área correspondente a classe a qual ele pertence.

Essa computação de pesos é feita em etapas, cada etapa utiliza N objetos da amostra, no qual N é um número pré-definido entre 1 e X objetos, sendo X o número total da amostra. Em cada etapa, o algoritmo se baseia nos N objetos para ajustar a linha de separação, buscando melhorar a capacidade de classificação da linha. Ou seja, com base nos N elementos, o algoritmo ajusta a linha para otimizar os acertos na classificação dos objetos.

5.1.3 Validação

Após o treinamento criar essa linha de separação, inicia-se a fase de validação, a qual tem por objetivo verificar a capacidade de generalização da classificação. Ou seja, a capacidade que a linha de separação criada tem de classificar novos objetos das classes em questão. Essa validação visa garantir que a linha de separação é útil fora da amostra utilizada para o treinamento. Dessa forma, o algoritmo utiliza a amostra de validação para verificar a precisão de classificação.

Caso o acerto seja de 100% na validação, o algoritmo encerra a execução, salvando a linha de separação para futuras classificações, caso seja menor, ela executa a fase de treinamento novamente, onde a linha inicial de separação é a linha final alcançada na fase de treinamento anterior. Essa segunda etapa de treinamento é seguida por uma nova fase de validação, e compara o resultado com o obtido na validação anterior. Caso a melhoria do resultado seja menor que o estipulado previamente pelo índice de tolerância (tol), o algoritmo encerra a execução, caso seja superior a execução retorna a fase de treinamento e repete todo o procedimento, até que a melhoria do resultado seja menor que a tolerância determinado.

Esse processo pode ser consideravelmente demorado, dessa forma, o algoritmo disponível na *API* possui uma configuração de número máximo de execuções (Max_iter), que limita o número de vezes o qual o algoritmo irá repetir a execução do conjunto de fases treinamento e validação. Com isso, mesmo que o algoritmo nunca tenha uma melhoria inferior ao índice de tolerância, ou nunca alcance a precisão de 100% na classificação da amostra, a execução é encerrada. O valor padrão para essa configuração é -1, que significa que em teoria o algoritmo poderia ter o tempo de execução tendendo ao infinito.

5.1.4 Resultado Final

O resultado final do processo de Treinamento e Validação do SVC é um algoritmo capaz de receber um vetor contendo as características de um objeto e então retornar uma predição para sua classificação dentre as classes utilizadas no treinamento. A qualidade da precisão dessa classificação dependerá de diversos fatores, tais quais, a existência de

separação entre as classes comparadas segundo as características analisadas, o tamanho da amostra de treinamento assim como o quão generalizada a amostra utilizada para treinamento e validação era.

Para cada SoRs, o SVC passou pelo processo de Treinamento e Validação uma vez para cada um dos 10 conjuntos, conforme cross validation. O resultado é apresentado no Capítulo 6 como a média simples da precisão de acerto nas 10 execuções.

6 Resultados e Discussão

O presente trabalho avaliou a existência de regiões cerebrais potencialmente contendo diferença funcional entre TEA-NA e Autistas. Para isso utilizou-se imagens cerebrais (fMRI) e o algoritmo de classificação SVC. Todos os indivíduos utilizados no experimento possuíam um diagnóstico prévio, assim como atribuição de suas respectivas notas ADOS. Baseado nessas notas e seguindo as notas de corte determinadas pelo protocolo, conforme apresentadas na Tabela 1, os indivíduos foram separados nos dois grupos e utilizados para treinamento, teste e validação do SVC.

6.1 Classificação

O SVC avaliou cada um dos grupos de regiões (SoRs) apresentados na Tabela 10. Os resultados apontaram as regiões *Cingulate Gyrus* (SoRs 11 - Figura 9) e *Superior Frontal Gyrus* (SoRs 1 - Figura 9) como as com maior diferenças funcionais dentre as SoRs analisadas, chegando a até 73.8% de ACC na classificação.

Dessa forma, esta seção apresenta as regiões de destaque nos resultados do experimento de classificação TEA-NA vs. Autistas. No presente trabalho a sensibilidade (SNS) refere-se a precisão da predição de indivíduos diagnosticados como Autistas, ou seja, considerando os resultados sobre a amostra de teste, a Sensibilidade significa a porcentagem de indivíduos Autistas classificados corretamente. Enquanto a Especificidade (SPC) refere-se a precisão da predição de indivíduos diagnosticados como TEA-NA, que de forma análoga representa a porcentagem de indivíduos do grupo TEA-NA da amostra de teste classificados corretamente.

O experimento contou com um total de 202 amostras, sendo 166 Autistas e 36 TEA-NA (vide Tabelas 6 e 7, respectivamente). Essa proporção de classes representa 82.2% de autistas e 17.8% de TEA-NA, essa proporção foi mantida o mais próximo possível para cada um dos K conjuntos para a validação cruzada mencionados na Seção 4.3.

Considerando as SoRs que atenderam os requisitos mínimos apresentados pela Seção 4.2, os resultados para Precisão (ACC), Sensibilidade (SNS) e Especificidade (SPC), superiores a 60% encontram-se no Apêndice A, enquanto os 10 melhores resultados são apresentados na Figura 10.

A ACC variou de 60.3% (SoRs 32) a 73.8% (SoRs 11) (avg 64.5%, std 0.03%, mediana 64.3%). Enquanto a SNS variou de 60.1% (SoRs 10) a 76.5% (SoRs 11) (avg 64.9%, std 0.03%, mediana 64.3%). E finalmente SPC variou de 60.8% (SoRs 32) a 70% (SoRs 30) (avg 62.6%, std 0.02%, mediana 62.5%).

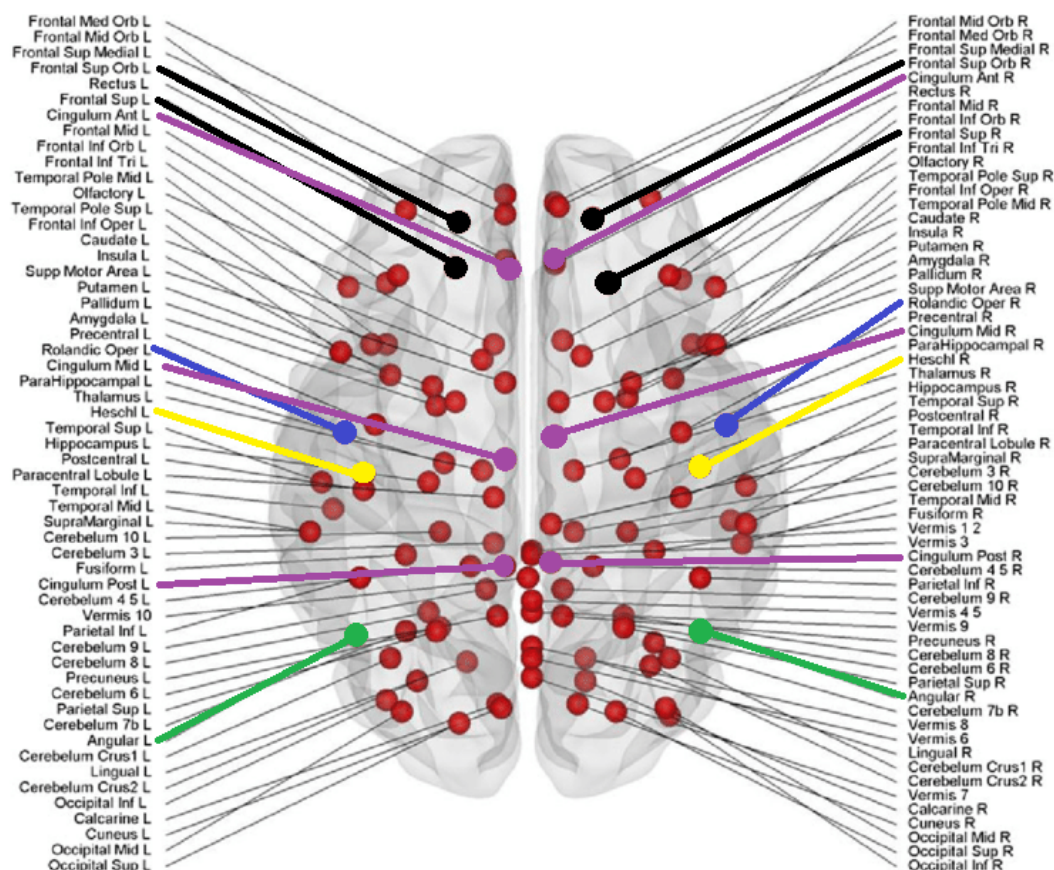


Figura 9 – SoRs com 5 melhores resultados de ACC. SoRs 11 - roxo, SoRs 23 - verde, SoRs 4 - azul, SoRs 1 - preto, SoRs 30 amarelo FONTE: (CAEYENBERGHS et al., 2012)

Dentre os quatro *Kernel* utilizados, o *linear* não alcançou os valores mínimos estipulados, enquanto os melhores resultados para cada região do *poly* se encontram no Apêndice B, os resultados para o *sigmoid* no Apêndice C e o *rbf* no Apêndice D.

Dentre os 4 *kernel* utilizados, o *poly* foi o que apresentou o melhor resultado para a maioria das SoRs, seguido pelo *sigmoid*, enquanto o *rbf* foi capaz de alcançar os resultados mínimos apenas para uma região (SoRs 30 - ACC 61.9%)

O *Kernel poly* foi o único a obter resultado superior ao mínimo estipulado para 14 SoRs, sendo o melhor em outras 3 para ACC e SNS e 4 para SPC, enquanto o *Kernel Sigmoid* foi o único a obter resultado superior em 3 SoRs, sendo melhor em outras 2 para ACC e SNS e em 0 para SPC. Ambos obtiveram resultados iguais para ACC e SNS em 8 SoRs e para SPC em 9. Enquanto em 5 SoRs nenhum *Kernel* obteve os valores mínimos.

Kernel poly foi o com maior quantidade de melhores resultados na classificação de SoRs. Comparando as SoRs onde ambos *poly* e *sigmoid* obtiveram alcançaram os requisitos mínimos (para ACC, SNS e SPC), nas SoRs onde *poly* foi superior a diferença média foi de 5% de ACC. Enquanto *Sigmoid* limitou-se a 1.2% a mais em relação às SoRs onde obteve melhor classificação. Vide Apêndice E para resultados mais detalhados.

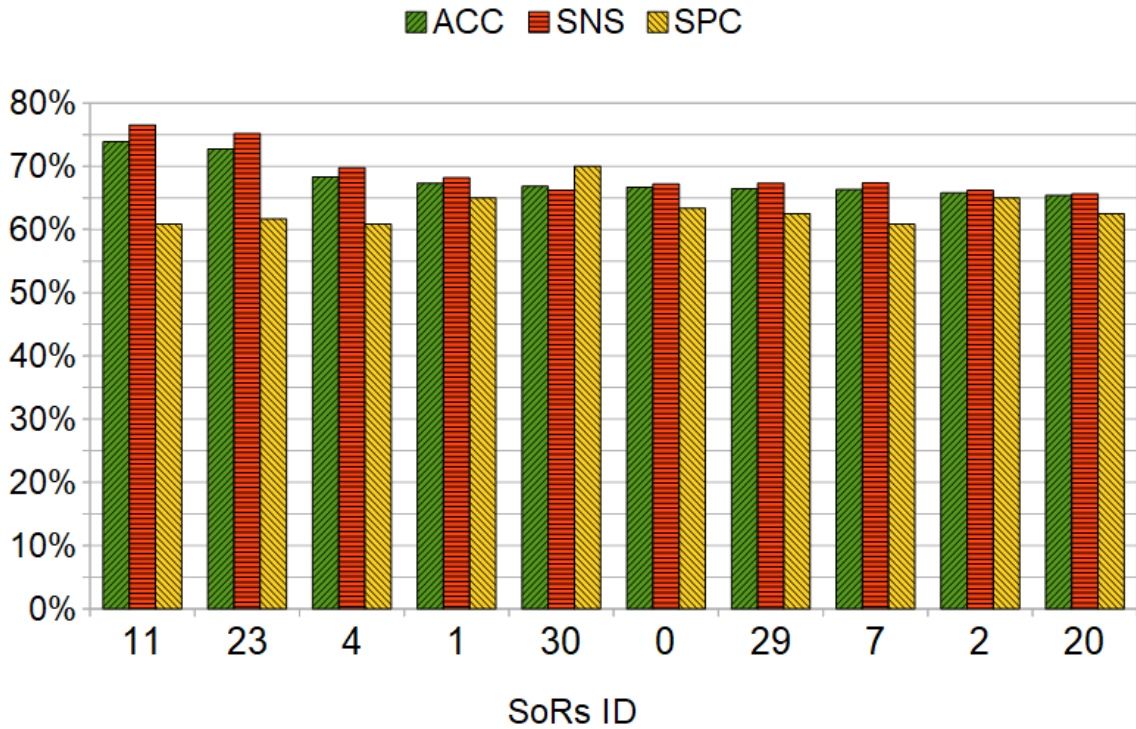


Figura 10 – SoRs com ACC, SNS, e SPC superior ou igual a 60%

Devido a maior capacidade de generalização do *Kernel poly* em relação aos outros três, é possível afirmar que este seja o mais indicado na resolução deste tipo de problema, sendo fortemente desencorajado o uso do *Kernel linear* e *RBF* para trabalhos futuros, utilizando métodos similares aos utilizados neste trabalho. Enquanto o *Kernel Sigmoid*, apesar de poder ser utilizado como auxiliar ao uso do *poly*, não apresentou grandes ganhos, podendo assim ser deixado de lado em detrimento de poupar recursos computacionais e tempo.

Conforme visto na Figura 10, a melhor precisão para o experimento foi de 73.8% referente a SoRs 11, a qual obteve 76.5% de SNS e 60.8% de SPC.

6.2 Efeito do Fenótipo

Adicionalmente utilizou-se de um teste final, no qual dados de fenótipo (idade e sexo) foram adicionadas como entradas em conjunto aos dados referentes a fMRI. Dessa forma o vetor de entrada para o SVC passa a ser:

$$(Z_a^{max}, Z_a^{min}, Z_a^{med}, Z_a^{avg}, \dots, Z_b^{max}, Z_b^{min}, Z_b^{med}, Z_b^{avg}, Z_{idade}, Z_{sexo}),$$

onde Z é o indivíduo, a e b são respectivamente, o primeiro e o último ID das ROI's que compõe cada uma das Set of Regions (SoRs) presentes na Tabela 10, enquanto Z_{idade} é a idade do indivíduo, e Z_{sexo} o sexo do indivíduo.

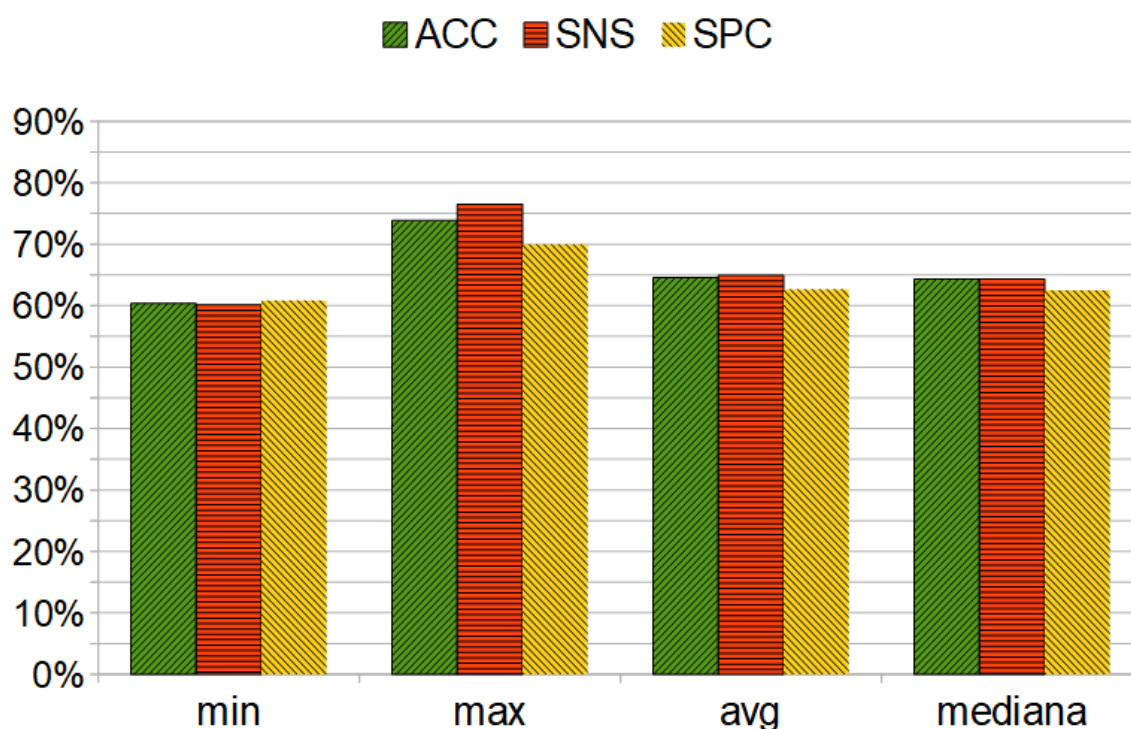


Figura 11 – Resultados de mínimo, máximo, média e desvio mediana respectivamente, para SoRs com ACC, SNS, e SPC superior ou igual a 60%

Esse experimento consistiu em utilizar as SoRs com melhor ACC encontradas por esse trabalho (vide Figura 10), e executar o experimento novamente, mas utilizando adicionalmente as informações de fenótipo relacionadas a cada indivíduo. o objetivo foi verificar os resultados para cada SoRs previamente selecionada, considerando o uso dessas informações adicionais no treinamento, teste e validação.

Entretanto, nenhuma SoRs testada apresentou resultados superiores aos requisitos mínimos apresentados em 4.2.

6.3 Discussão

Através da análise dos resultados obtidos no presente trabalho, nota-se a existência de diferenças funcionais a respeito de atividade máxima, mínima e média em determinados conjuntos de regiões do cérebro, entre indivíduos com diagnóstico TEA e autistas. Isto indica uma potencial existência de uma relação entre essas regiões e a severidade dos sintomas relacionados ao TEA.

Entretanto, uma das limitações do estudo foi o tamanho da amostra utilizada. Dessa forma, para verificar a capacidade de generalização da relação entre o funcionamento dessas regiões e a severidade dos sintomas TEA, é preciso uma amostra maior.

O tamanho da amostra, também possui implicações nos resultados de baixa precisão de predição. Dessa forma, para determinada região no experimento conduzido cuja precisão foi baixa, não necessariamente implica em uma inexistência de diferença funcional entre TEA-NA e Autista. É importante ressaltar que não se deve descartar totalmente as outras regiões. Mesmo que provavelmente não possuam diferença funcional para todos os indivíduos do espectro, tomando por base as características analisadas neste estudo.

Ressalta-se ainda que a proporção de distribuição da amostra, aumenta consideravelmente a importância das informações sobre sensibilidade e especificidade. Sendo assim, é ainda mais importante um equilíbrio entre SNS e SPC. Isto posto, a região *Cingulate Gyrus*, ganha ainda mais relevância nos resultados obtidos, uma vez que além de possuir melhor ACC foi a região com melhor resultado multiplicando-se sua SNS e SPC. Essa multiplicação aponta a capacidade de generalização na variação da proporção da amostra.

Ademais, os resultados apresentados no Capítulo 6, corroboram resultados encontrados em estudos prévios da literatura, que apontam alteração no sistema límbico de indivíduos com diagnóstico TEA em relação a neurotípicos (KIM et al., 2015; PASCUAL-BELDA; DÍAZ-PARRA; MORATAL, 2018; HAZNEDAR et al., 2000). Uma vez que duas das regiões encontradas fazem parte desse sistema (*Cingulate Gyrus* - SoRs 11 e *Thalamus* - SoRs 29), tal concordância reforça a possibilidade que exista uma variação funcional dessas regiões em relação a severidade dos sintomas do TEA.

Adicionalmente, a região *Thalamus*, obteve 63% na classificação, que assim como *Cingulate Gyrus* faz parte do sistema límbico. *Thalamus* está relacionado a fala e nomeação de objetos (OJEMANN; JR, 1971), assim como a questões visuais como busca e reconhecimento de padrão (CRICK, 1984). Sendo também apontada como uma região chave na classificação entre TEA e neurotípicos (IIDAKA, 2015; SUBBARAJU et al., 2017). Estando assim em conformidade com os resultados do presente estudo, de forma que além de possuir média geral de atividade menor no grupo autista em relação aos TEA-NA, a média de máximos de valores para a região esquerda foi menor enquanto para a região direita foi maior.

Por fim, diferente do esperado, as informações a respeito de sexo e idade não foram dados relevantes para uma classificação mais precisa, não apresentando melhoria nos resultados quando comparados com os resultados obtidos sem o uso dessas informações. Uma vez que a literatura apresenta trabalhos que apontam para a existência de diferenças funcionais baseadas em idade (CARVER et al., 2001; SURWILLO, 1963; DEKABAN; SADOWSKY, 1978), e sexo (SWAAB; FLIERS; PARTIMAN, 1985; ROBINSON et al., 1971; CORSI-CABRERA; HERRERA; MALVIDO, 1989).

Todavia, a ausência de melhoria de ACC, SPC e SNS de forma consistente para todas as regiões com a adição de informação de fenótipo, pode-se conjecturar três hipóte-

ses: i) as informações de fenótipo agiram como forma de *overfit* no SVC, de forma similar ao que é discutido por (HAWKINS, 2004); ii) as regiões que não apresentaram melhora são regiões do cérebro onde a atividade funcional possui menor efeito relacionado aos dados de fenótipos utilizados; iii) o volume de dados não foi suficiente para que o algoritmo encontre um padrão para classificação utilizando esses dados adicionais. Entretanto, a investigação dessa questão está fora do escopo proposto por esse trabalho, mas é uma investigação válida a se considerar para trabalhos futuros.

7 Conclusão

O algoritmo SVM utilizado obteve sucesso na classificação, identificando regiões com potencial de diferenciar o grau de severidade de indivíduos com diagnóstico TEA, com destaque para o *Cingulate Gyrus*.

O tamanho da amostra utilizada foi a maior limitação do estudo. Uma vez que a nota ADOS é uma informação crucial para o desenvolvimento do trabalho, o estudo limitou-se a 202 indivíduos cujas respectivas fMRI foram disponibilizadas pelo consórcio ABIDE juntamente com as notas da avaliação ADOS. o que não nos permite generalizar nossos resultados para todas as populações. Todavia, podemos especular sobre a existência dessas diferenças funcionais entre os subtipos de TEA.

A obtenção de 73.8% de precisão na classificação entre indivíduos TEA-NA e autistas, para a região do *Cingulate Gyrus* mostra uma região promissora como biomarcador da severidade do TEA. Principalmente por estar em conformidade com a literatura existente, como em (HAZNEDAR et al., 1997), que mostra indivíduos com diagnóstico TEA, possuem menor atividade metabólica na região do *Cingulate Gyrus* em comparação a indivíduos DT. Enquanto Apps, Rushworth e Chang (2016) aponta a região do *Cingulate Gyrus* como tendo um papel importante no processamento de informação social. Sendo a interação social uma das áreas avaliadas para se determinar a severidade do TEA.

Adicionalmente, o *Cingulate Gyrus* pode viabilizar uma avaliação computacional para contribuir com a melhora do diagnóstico médico, fornecendo mais ferramentas para a avaliação dos médicos, reduzindo a ambigüidade na análise comportamental. Adicionalmente, distinguir a gravidade em um indivíduo pode ajudar na seleção de intervenções, uma vez que as intervenções adequadas impactam significativamente a qualidade de vida individual.

Outra limitação do estudo, foi o pipeline de processamento final, o qual foi um modelo bastante simples, consistindo em analisar apenas os máximos, mínimos, médias e desvio padrão para cada um dos conjuntos de regiões analisadas. Talvez um processamento mais elaborado possa levar a melhores resultados.

Para trabalhos futuros, um aumento nos dados disponíveis potencialmente melhoraria a precisão, contribuindo para o esclarecimento da capacidade de generalização dos resultados para todas as populações. Seria interessante para a comunidade de pesquisadores que os sites que disponibilizam esses dados também disponibilizassem mais dados fenotípicos, como a nota ADOS tanto para os indivíduos com diagnóstico TEA quanto para os neurotípicos, uma vez que isso permitiria investigações mais precisas.

7.1 Trabalhos Futuros

Dentre os possíveis trabalhos futuros sugere-se:

A criação de um modelo para agrupamento de ROI's baseando-se nas SoRs encontradas neste trabalho, gerando novas SoRs com maior potencial de classificação que as encontradas.

Utilizar diferentes técnicas de ML baseando-se nas regiões encontradas, para comparação das técnicas entre si na solução deste tipo de problema.

Utilizar técnicas de regressão linear, buscando quantificar as diferenças entre indivíduos, baseando-se nas regiões encontradas visando uma relação entre a nota ADOS e as variações funcionais nas SoRs apresentadas neste trabalho.

Utilizar modelos de pré-processamento mais complexos que o utilizado neste trabalho (média, mediana, máximo e mínimo), para investigar as potenciais diferenças.

Apêndices

APÊNDICE A – Tabela Melhores

Resultados Por Região

Tabela 11 – Melhores Resultados Por Região

ID	ACC	SNS	SPC	ID	ACC	SNS	SPC
0	66.666	67.169	63.333	18	60.904	60.367	63.333
1	67.309	68.161	65.000	19	63.809	64.375	60.833
2	65.785	66.213	65.000	20	65.380	65.588	62.500
3	*	*	*	21	61.380	61.580	60.833
4	68.309	69.742	60.833	22	63.833	63.529	63.333
5	63.404	64.191	61.666	23	72.738	75.147	61.666
6	63.809	64.264	60.833	24	61.952	61.507	62.500
7	66.333	67.352	60.833	25	64.880	65.147	63.333
8	64.380	64.080	67.500	26	61.380	61.544	60.833
9	*	*	*	27	63.928	64.154	63.333
10	60.380	60.110	61.666	28	60.952	60.367	63.333
11	73.857	76.507	60.833	29	66.404	67.316	62.500
12	*	*	*	30	66.809	66.213	70.000
13	*	*	*	31	65.333	65.551	63.333
14	60.952	60.919	60.833	32	60.357	60.330	60.833
15	*	*	*	33	64.809	65.441	62.500
16	61.904	62.132	60.833	34	64.380	64.558	63.333
17	64.285	64.264	62.500				

Valores apresentados em porcentagem, referente a classificação

Foram considerados apenas os valores das 3 primeiras casas decimais para cada região.

Regiões marcadas por * foram as que o algoritmo não alcançou os valores mínimos estipulados.

APÊNDICE B – Tabela Melhores Resultados Para *Kernel Poly*

Tabela 12 – Melhores Resultados Para Kernel Poly

ID	ACC	SNS	SPC	ID	ACC	SNS	SPC
0	66.67	67.17	63.33	18	*	*	*
1	67.31	68.16	65.00	19	63.81	64.38	60.83
2	65.79	66.21	65.00	20	65.38	65.59	62.50
3	*	*	*	21	60.38	60.22	62.50
4	68.31	69.74	60.83	22	63.83	63.53	63.33
5	63.40	64.19	61.67	23	72.74	75.15	61.67
6	63.81	64.26	60.83	24	61.95	61.51	62.50
7	66.33	67.35	60.83	25	64.88	65.15	63.33
8	64.38	64.08	67.50	26	*	*	*
9	*	*	*	27	63.93	64.15	63.33
10	60.38	60.11	61.67	28	60.95	60.37	63.33
11	73.86	76.51	60.83	29	66.40	67.32	62.50
12	*	*	*	30	66.81	66.21	70.00
13	*	*	*	31	65.33	65.55	63.33
14	60.95	60.92	60.83	32	*	*	*
15	*	*	*	33	64.81	65.44	62.50
16	60.48	60.44	62.50	34	64.38	64.56	63.33
17	64.29	64.26	62.50		*	*	*

Valores apresentados em porcentagem, referente a classificação

Foram considerados apenas os valores das 4 primeiras casas decimais para cada região.

Regiões marcadas por * foram as que o algoritmo não alcançou os valores mínimos estipulados.

APÊNDICE C – Tabela Melhores Resultados Para *Kernel Sigmoid*

Tabela 13 – Melhores Resultados Para *Kernel Sigmoid*

ID	ACC	SNS	SPC	ID	ACC	SNS	SPC
0	63.83	64.55	60.83	18	60.90	60.36	63.33
1	*	*	*	19	*	*	*
2	65.78	66.21	65.00	20	*	*	*
3	*	*	*	21	61.38	61.58	60.83
4	68.31	69.74	60.83	22	62.38	62.09	63.33
5	63.40	64.19	61.66	23	*	*	*
6	63.81	64.26	60.83	24	*	*	*
7	66.33	67.35	60.83	25	*	*	*
8	64.38	64.08	67.50	26	61.38	61.54	60.83
9	*	*	*	27	*	*	*
10	*	*	*	28	*	*	*
11	*	*	*	29	*	*	*
12	*	*	*	30	60.90	60.33	62.50
13	*	*	*	31	65.33	65.55	63.33
14	60.95	60.91	60.83	32	60.35	60.33	60.83
15	*	*	*	33	*	*	*
16	61.90	62.13	60.83	34	*	*	*
17	*	*	*				

Valores apresentados em porcentagem, referente a classificação

Foram considerados apenas os valores das 2 primeiras casas decimais para cada região.

Regiões marcadas por * foram as que o algoritmo não alcançou os valores mínimos estipulados.

APÊNDICE D – Tabela Melhores Resultados Para *Kernel* RBF

Tabela 14 – Melhores Resultados Para *Kernel* RBF

ID	ACC	SNS	SPC	ID	ACC	SNS	SPC
0	*	*	*	18	*	*	*
1	*	*	*	19	*	*	*
2	*	*	*	20	*	*	*
3	*	*	*	21	*	*	*
4	*	*	*	22	*	*	*
5	*	*	*	23	*	*	*
6	*	*	*	24	*	*	*
7	*	*	*	25	*	*	*
8	*	*	*	26	*	*	*
9	*	*	*	27	*	*	*
10	*	*	*	28	*	*	*
11	*	*	*	29	*	*	*
12	*	*	*	30	61.90	61.47	63.33
13	*	*	*	31	*	*	*
14	*	*	*	32	*	*	*
15	*	*	*	33	*	*	*
16	*	*	*	34	*	*	*
17	*	*	*				

Valores apresentados em porcentagem, referente a classificação

Foram considerados apenas os valores das 5 primeiras casas decimais para cada região.

Regiões marcadas por * foram as que o algoritmo não alcançou os valores mínimos estipulados.

APÊNDICE E – Tabela Comparação dos resultados entre *Kernel poly* e *Sigmoid*

Tabela 15 – Comparação dos resultados entre *Kernel poly* e *Sigmoid*

ID	ACC	SNS	SPC	ID	ACC	SNS	SPC
0	02.833	02.610	02.500	18	-60.905	-60.368	-63.333
1	67.310	68.162	65.000	19	63.810	64.375	60.833
2	00.000	00.000	00.000	20	65.381	65.588	62.500
3	*	*	*	21	-01.000	-01.360	01.667
4	00.000	00.000	00.000	22	01.452	01.434	00.000
5	00.000	00.000	00.000	23	72.738	75.147	61.667
6	00.000	00.000	00.000	24	61.952	61.507	62.500
7	00.000	00.000	00.000	25	64.881	65.147	63.333
8	00.000	00.000	00.000	26	-61.381	-61.544	-60.833
9	*	*	*	27	63.929	64.154	63.333
10	60.381	60.110	61.667	28	60.952	60.368	63.333
11	73.857	76.507	60.833	29	66.405	67.316	62.500
12	*	*	*	30	05.905	05.882	07.500
13	*	*	*	31	00.000	00.000	00.000
14	00.000	00.000	00.000	32	-60.357	-60.331	-60.833
15	*	*	*	33	64.810	65.441	62.500
16	-01.429	-01.691	01.667	34	64.381	64.559	63.333
17	64.286	64.265	62.500				

Valores apresentados em porcentagem, referente a classificação

Foram considerados apenas os valores das 3 primeiras casas decimais para cada região.

Regiões marcadas por * foram as que o algoritmo não alcançou os valores mínimos estipulados em nenhum *Kernel*.

Os valores foram gerados fazendo subtraindo dos resultados obtidos pelo *poly* pelo obtido pelo *Sigmoid* ($poly - Sigmoid$)

Para SoRs onde um kernel não tenha alcançado os requisitos é apresentado o resultado obtido pelo outro, com números positivos representando apenas o *poly* tendo obtido resultados e valores negativos representado apenas o *Sigmoid*

valores de 0.0 representam que ambos *Kernel* obtiveram o mesmo resultado.

Referências

- ABRAHAM, A. et al. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, v. 147, p. 736 – 745, 2017. ISSN 1053-8119. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1053811916305924>>.
- APPS, M. A.; RUSHWORTH, M. F.; CHANG, S. W. The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron*, Elsevier, v. 90, n. 4, p. 692–707, 2016.
- ARANI, A. et al. Measuring the effects of aging and sex on regional brain stiffness with mr elastography in healthy older adults. *NeuroImage*, v. 111, p. 59 – 64, 2015. ISSN 1053-8119. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1053811915001172>>.
- ASSOCIATION, A. P. et al. *DSM-5: Diagnostic and Statistical Manual of Mental Disorders*. [S.l.]: Artmed Editora, 2014.
- BAI, D. et al. Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA psychiatry*, American Medical Association, v. 76, n. 10, p. 1035–1043, 2019.
- BAIO, J. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2010. 2014.
- BANDETTINI, P. A. et al. Time course epi of human brain function during task activation. *Magnetic resonance in medicine*, Wiley Online Library, v. 25, n. 2, p. 390–397, 1992.
- BANDETTINI, P. A. et al. Time course epi of human brain function during task activation. *Magnetic resonance in medicine*, Wiley Online Library, v. 25, n. 2, p. 390–397, 1992.
- BANDINI, L. G. et al. Food selectivity in children with autism spectrum disorders and typically developing children. *The Journal of pediatrics*, Elsevier, v. 157, n. 2, p. 259–264, 2010.
- BEJARANO-MARTÍN Álvaro et al. Efficacy of focused social and communication intervention practices for young children with autism spectrum disorder: A meta-analysis. *Early Childhood Research Quarterly*, v. 51, p. 430 – 445, 2020. ISSN 0885-2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0885200620300041>>.
- BENGIO, Y.; GRANDVALET, Y. No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, v. 5, n. Sep, p. 1089–1105, 2004.
- BETTS, J. et al. Anatomy & physiology: Openstax college. *Rice University*, 2013.
- BHAUMIK, R. et al. Predicting autism spectrum disorder using domain-adaptive cross-site evaluation. *Neuroinformatics*, v. 16, 02 2018.

- BI, X.-a. et al. Analysis of asperger syndrome using genetic-evolutionary random support vector machine cluster. *Frontiers in Physiology*, Frontiers, v. 9, p. 1646, 2018.
- BI, X.-a. et al. Classification of autism spectrum disorder using random support vector machine cluster. *Frontiers in genetics*, Frontiers, v. 9, p. 18, 2018.
- Brazil's Ministry of Health. *Diretrizes de Atenção à Reabilitação da Pessoa com Transtorno do Espectro Autista (TEA)*. [S.l.]: Ministério da Saúde Brasília, 2014.
- CAEYENBERGHS, K. et al. Altered structural networks and executive deficits in traumatic brain injury patients. *Brain structure function*, v. 219, 12 2012.
- Carvalho, E. A. et al. Hidden markov models to estimate the probability of having autistic children. *IEEE Access*, v. 8, p. 99540–99551, 2020.
- CARVER, J. D. et al. The relationship between age and the fatty acid composition of cerebral cortex and erythrocytes in human subjects. *Brain Research Bulletin*, v. 56, n. 2, p. 79 – 85, 2001. ISSN 0361-9230. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S036192300100512>>.
- Chaitra, N.; Vijaya, P. A. Comparing univalent and bivalent brain functional connectivity measures using machine learning. In: *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*. [S.l.: s.n.], 2017. p. 1–5.
- CHEN, J.-K. et al. Functional abnormalities in symptomatic concussed athletes: an fmri study. *Neuroimage*, Elsevier, v. 22, n. 1, p. 68–82, 2004.
- CHISTOL, L. T. et al. Sensory sensitivity and food selectivity in children with autism spectrum disorder. *Journal of autism and developmental disorders*, Springer, v. 48, n. 2, p. 583–591, 2018.
- CMRR, S.-g. K.; UĞURBIL, K. Comparison of blood oxygenation and cerebral blood flow effect in fmri: Estimation of relative oxygen consumption change. *Magnetic Resonance in Medicine*, Wiley Online Library, v. 38, n. 1, p. 59–65, 1997.
- CORSI-CABRERA, M.; HERRERA, P.; MALVIDO, M. Correlation between eeg and cognitive abilities: Sex differences. *International Journal of Neuroscience*, Taylor Francis, v. 45, n. 1-2, p. 133–141, 1989. Disponível em: <<https://doi.org/10.3109/00207458908986226>>.
- CRADDOCK, C. et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives
. *Frontiers in Neuroinformatics*, v. 7, 2013.
- CRICK, F. Function of the thalamic reticular complex: the searchlight hypothesis. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 81, n. 14, p. 4586–4590, 1984.
- DAWSON, G. et al. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*, Am Acad Pediatrics, v. 125, n. 1, p. e17–e23, 2010.

- DEKABAN, A. S.; SADOWSKY, D. Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, Wiley Online Library, v. 4, n. 4, p. 345–356, 1978.
- DEYOE, E. A. et al. Functional magnetic resonance imaging (fmri) of the human brain. *Journal of neuroscience methods*, Elsevier, v. 54, n. 2, p. 171–187, 1994.
- Dodero, L. et al. Kernel-based classification for brain connectivity graphs on the riemannian manifold of positive definite matrices. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2015. p. 42–45.
- DODERO, L. et al. Kernel-based analysis of functional brain connectivity on grassmann manifold. In: NAVAB, N. et al. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 604–611. ISBN 978-3-319-24574-4.
- DUDA, M.; KOSMICKI, J.; WALL, D. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Translational psychiatry*, Nature Publishing Group, v. 4, n. 8, p. e424–e424, 2014.
- DURKIN, M. S. et al. Autism screening and diagnosis in low resource settings: challenges and opportunities to enhance research and services worldwide. *Autism Research*, Wiley Online Library, v. 8, n. 5, p. 473–476, 2015.
- FALKMER, T. et al. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European child & adolescent psychiatry*, Springer, v. 22, n. 6, p. 329–340, 2013.
- FAN, L. et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, Oxford University Press, v. 26, n. 8, p. 3508–3526, 2016.
- FOX, P. T. et al. Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, American Association for the Advancement of Science, v. 241, n. 4864, p. 462–464, 1988.
- FUSHIKI, T. Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing*, Springer, v. 21, n. 2, p. 137–146, 2011.
- GOLBY, A. et al. Memory encoding in alzheimer’s disease: an fmri study of explicit and implicit memory. *Brain*, Oxford University Press, v. 128, n. 4, p. 773–787, 2005.
- GOTHAM, K. et al. The autism diagnostic observation schedule: revised algorithms for improved diagnostic validity. *Journal of autism and developmental disorders*, Springer, v. 37, n. 4, p. 613–627, 2007.
- GREITZ, T. et al. A computerized brain atlas: construction, anatomical content, and some applications. *Journal of computer assisted tomography*, v. 15, n. 1, p. 26–38, 1991.
- HAWKINS, D. M. The problem of overfitting. *Journal of chemical information and computer sciences*, ACS Publications, v. 44, n. 1, p. 1–12, 2004.
- HAZLETT, H. C. et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*, Nature Publishing Group, v. 542, n. 7641, p. 348, 2017.

- HAZNEDAR, M. M. et al. Anterior cingulate gyrus volume and glucose metabolism in autistic disorder. *American Journal of Psychiatry*, [Washington, etc., American Psychiatric Association], v. 154, n. 8, p. 1047–1050, 1997.
- HAZNEDAR, M. M. et al. Limbic circuitry in patients with autism spectrum disorders studied with positron emission tomography and magnetic resonance imaging. *American Journal of Psychiatry*, Am Psychiatric Assoc, v. 157, n. 12, p. 1994–2001, 2000.
- HEEGER, D. J.; RESS, D. What does fmri tell us about neuronal activity? *Nature Reviews Neuroscience*, Nature Publishing Group, v. 3, n. 2, p. 142–151, 2002.
- IIDAKA, T. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, Elsevier, v. 63, p. 55–67, 2015.
- ITAHASHI, T. et al. Altered network topologies and hub organization in adults with autism: a resting-state fmri study. *PloS one*, Public Library of Science, v. 9, n. 4, p. e94115, 2014.
- JIAO, Y. et al. Predictive models for subtypes of autism spectrum disorder based on single-nucleotide polymorphisms and magnetic resonance imaging. *Advances in medical sciences*, Elsevier, v. 56, n. 2, p. 334–342, 2011.
- JONES, T. B.; BANDETTINI, P. A.; BIRN, R. M. Integration of motion correction and physiological noise regression in fmri. *Neuroimage*, Elsevier, v. 42, n. 2, p. 582–590, 2008.
- JUN, E.; SUK, H.-I. Region-wise stochastic pattern modeling for autism spectrum disorder identification and temporal dynamics analysis. In: SPRINGER. *International Workshop on Connectomics in Neuroimaging*. [S.l.], 2017. p. 143–151.
- KAMP-BECKER, I. et al. Diagnostic accuracy of the ados and ados-2 in clinical practice. *European child & adolescent psychiatry*, Springer, v. 27, n. 9, p. 1193–1207, 2018.
- KIM, S.-Y. et al. Abnormal activation of the social brain network in children with autism spectrum disorder: an fmri study. *Psychiatry investigation*, Korean Neuropsychiatric Association, v. 12, n. 1, p. 37, 2015.
- LEVY, S. et al. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular autism*, Springer, v. 8, n. 1, p. 65, 2017.
- LEWIS, J. et al. Network efficiency in autism spectrum disorder and its relation to brain overgrowth. *Frontiers in human neuroscience*, v. 7, p. 845, 12 2013.
- LOGOTHETIS, N. K. What we can do and what we cannot do with fmri. *Nature*, Nature Publishing Group, v. 453, n. 7197, p. 869–878, 2008.
- LOGOTHETIS, N. K.; WANDELL, B. A. Interpreting the bold signal. *Annu. Rev. Physiol.*, Annual Reviews, v. 66, p. 735–769, 2004.
- LONG, Z. et al. A comparative atlas-based recognition of mild cognitive impairment with voxel-based morphometry. *Frontiers in Neuroscience*, v. 12, p. 916, 2018. ISSN 1662-453X. Disponível em: <<https://www.frontiersin.org/article/10.3389/fnins.2018.00916>>.

- LOOMES, R.; HULL, L.; MANDY, W. P. L. What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 56, n. 6, p. 466–474, 2017.
- LORD, C. et al. Autism diagnostic observation schedule–generic. *Dissertation Abstracts International Section A: Humanities and Social Sciences*.
- MAHANAND, B. et al. An enhanced effect-size thresholding method for the diagnosis of autism spectrum disorder using resting state functional mri. In: IEEE. *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*. [S.l.], 2016. p. 1–6.
- MARCOT, B. G.; HANEA, A. M. What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis? *Computational Statistics*, Springer, p. 1–23, 2020.
- NAUMAN, E. A. et al. Post-season neurophysiological deficits assessed by impact and fmri in athletes competing in american football. *Developmental neuropsychology*, Taylor & Francis, v. 40, n. 2, p. 85–91, 2015.
- OJEMANN, G. A.; JR, A. A. W. Speech representation in ventrolateral thalamus. *Brain*, Citeseer, v. 94, n. 4, p. 669–680, 1971.
- PARRY, A.; MATTHEWS, P. Functional magnetic resonance imaging (fmri): A “window” into the brain. *Retrieved July*, v. 25, p. 2005, 2002.
- PASCUAL-BELDA, A.; DÍAZ-PARRA, A.; MORATAL, D. Evaluating functional connectivity alterations in autism spectrum disorder using network-based statistics. *Diagnostics*, Multidisciplinary Digital Publishing Institute, v. 8, n. 3, p. 51, 2018.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- POWER, J. D. et al. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, Elsevier, v. 59, n. 3, p. 2142–2154, 2012.
- PROSPERI, M. et al. Behavioral phenotype of asd preschoolers with gastrointestinal symptoms or food selectivity. *Journal of Autism and Developmental Disorders*, Springer, v. 47, n. 11, p. 3574–3588, 2017.
- PROVOST, B.; LOPEZ, B. R.; HEIMERL, S. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. *Journal of autism and developmental disorders*, Springer, v. 37, n. 2, p. 321–328, 2007.
- ROBINSON, D. S. et al. Relation of Sex and Aging to Monoamine Oxidase Activity of Human Brain, Plasma, and Platelets. *Archives of General Psychiatry*, v. 24, n. 6, p. 536–539, 06 1971. ISSN 0003-990X. Disponível em: <<https://doi.org/10.1001/archpsyc.1971.01750120052009>>.
- RODRIGUEZ, J. D.; PEREZ, A.; LOZANO, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 32, n. 3, p. 569–575, 2009.

- ROLLS, E. T. et al. Automated anatomical labelling atlas 3. *NeuroImage*, p. 116–189, 2019.
- ROMBOUTS, S. A. et al. Altered resting state networks in mild cognitive impairment and mild alzheimer’s disease: an fmri study. *Human brain mapping*, Wiley Online Library, v. 26, n. 4, p. 231–239, 2005.
- SANDIN, S. et al. The heritability of autism spectrum disorder. *Jama*, v. 318, n. 12, p. 1182–1184, 2017.
- SATTERTHWAITE, T. D. et al. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage*, Elsevier, v. 60, n. 1, p. 623–632, 2012.
- SHIMABUKURO, T. T.; GROSSE, S. D.; RICE, C. Medical expenditures for children with an autism spectrum disorder in a privately insured population. *Journal of Autism and Developmental Disorders*, Springer, v. 38, n. 3, p. 546–552, 2008.
- SPECK, O.; HENNIG, J.; ZAITSEV, M. Prospective real-time slice-by-slice motion correction for fmri in freely moving subjects. *Magnetic Resonance Materials in Physics, Biology and Medicine*, Springer, v. 19, n. 2, p. 55, 2006.
- SUBBARAJU, V. et al. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. *Medical image analysis*, Elsevier, v. 35, p. 375–389, 2017.
- SURWILLO, W. W. The relation of simple response time to brain-wave frequency and the effects of age. *Electroencephalography and Clinical Neurophysiology*, v. 15, n. 1, p. 105 – 114, 1963. ISSN 0013-4694. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0013469463900439>>.
- SWAAB, D.; FLIERS, E.; PARTIMAN, T. The suprachiasmatic nucleus of the human brain in relation to sex, age and senile dementia. *Brain Research*, v. 342, n. 1, p. 37 – 44, 1985. ISSN 0006-8993. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0006899385913502>>.
- TANENBAUM, L. N. et al. Synthetic mri for clinical neuroimaging: results of the magnetic resonance image compilation (magic) prospective, multicenter, multireader trial. *American Journal of Neuroradiology*, Am Soc Neuroradiology, v. 38, n. 6, p. 1103–1110, 2017.
- WANG, K. et al. Altered functional connectivity in early alzheimer’s disease: A resting-state fmri study. *Human brain mapping*, Wiley Online Library, v. 28, n. 10, p. 967–978, 2007.
- WARD, N. et al. Neural correlates of motor recovery after stroke: a longitudinal fmri study. *Brain*, Oxford University Press, v. 126, n. 11, p. 2476–2496, 2003.
- WILSON, R. B.; ENTICOTT, P. G.; RINEHART, N. J. Motor development and delay: advances in assessment of motor skills in autism spectrum disorders. *Current opinion in neurology*, Wolters Kluwer, v. 31, n. 2, p. 134–139, 2018.

WOLFERS, T. et al. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, Elsevier, v. 57, p. 328–349, 2015.

YAN, C.; ZANG, Y. Dparsf: a matlab toolbox for "pipeline" data analysis of resting-state fmri. *Frontiers in Systems Neuroscience*, v. 4, p. 13, 2010. ISSN 1662-5137. Disponível em: <<https://www.frontiersin.org/article/10.3389/fnsys.2010.00013>>.

ZHU, Y. et al. Reveal consistent spatial-temporal patterns from dynamic functional connectivity for autism spectrum disorder identification. In: SPRINGER. *International conference on medical image computing and computer-assisted intervention*. [S.l.], 2016. p. 106–114.

ZU, C. et al. Identifying high order brain connectome biomarkers via learning on hypergraph. In: WANG, L. et al. (Ed.). *Machine Learning in Medical Imaging*. Cham: Springer International Publishing, 2016. p. 1–9. ISBN 978-3-319-47157-0.