

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

FELIPE DE OLIVEIRA FRANCO

**Abordagem de auxílio diagnóstico para o Transtorno do Espectro Autista:
integração de características semânticas em modelos de atenção visual**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

São Paulo

2024

FELIPE DE OLIVEIRA FRANCO

**Abordagem de auxílio diagnóstico para o Transtorno do Espectro Autista:
integração de características semânticas em modelos de atenção visual**

Versão Original

Tese de Doutorado apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de Concentração: Bioinformática
Orientadora: Dra. Helena Brentani
Coorientadora: Dra. Fátima de L. S. Nunes

São Paulo

2024

RESUMO

FRANCO, F. O. (2024). Abordagem de auxílio diagnóstico para o Transtorno do Espectro Autista: integração de características semânticas em modelos de atenção visual (Tese de Doutorado). Programa de Pós-graduação Interunidades em Bioinformática, Universidade de São Paulo, São Paulo.

Estima-se que 1% da população mundial se enquadre no Transtorno do Espectro Autista (TEA), o qual é caracterizado por alterações na comunicação social e presença de comportamentos estereotipados e repetitivos. Estudos sugerem a presença de alterações na conectividade cerebral no TEA, com um processamento atípico de informações sociais, relacionadas aos mecanismos *bottom-up* e *top-down*. Diversos classificadores para auxílio diagnóstico do TEA vêm sendo propostos a partir de dados de rastreamento do olhar. No entanto, a maioria dos trabalhos baseia-se em dados como duração e contagem das fixações do olhar. Outros estudos combinam dados de rastreamento do olhar e características extraídas da imagem para criar Modelos de Atenção Visual (MAVs) e mapas de saliência. Recentemente, nosso grupo propôs uma abordagem baseada em mapas de saliência para o auxílio diagnóstico do TEA. Considerando que alterações da conectividade *top-down* estão presentes no TEA, nossa hipótese é que um maior número de características semânticas nestes MAVs pode melhorar o desempenho destes classificadores. Portanto, o objetivo deste estudo é desenvolver classificadores para o auxílio diagnóstico do TEA, baseado em MAVs considerando diferentes níveis de características em mapas de saliência, com enfoque na integração de características semânticas. Para atingir os objetivos, o primeiro passo foi replicar o estudo anterior do grupo com uma nova amostra e novo estímulo visual pautado no mesmo paradigma de atenção visual. A replicação ocorreu de maneira satisfatória, sugerindo a robustez da abordagem prévia e a viabilidade das abordagens baseadas em MAVs para auxiliar no diagnóstico de TEA. No entanto, a troca de estímulo visual e amostras causaram diminuição no desempenho dos classificadores. Após a replicação, outros algoritmos indutores de classificação foram testados e o algoritmo *Random Forest* apresentou melhores desempenhos em todos os cenários, com a vantagem adicional de permitir a comparação de importância atribuída a cada uma das características consideradas. Foi realizada uma revisão da literatura para avaliar a viabilidade de inclusão de mais características semânticas e então, procedemos com a extração das novas características. A comparação de importância das características contribuiu na interpretabilidade dos classificadores, nesse sentido verificamos que: 1 - características semânticas apresentaram importância relativa menor que as demais, porém com maior importância para indivíduos com desenvolvimento típico; 2 - a característica "Pessoa" foi a que mais divergiu entre os grupos, sendo menos importante no TEA; 3 - as características centro da tela e centro das cenas foram as mais importantes e a exclusão delas impactou negativamente na classificação. De maneira geral, a inclusão de características semânticas não melhorou o desempenho de classificação. Porém, ao segmentar o *dataset* excluindo indivíduos mais novos, observou-se que o desempenho melhorou. Algumas limitações são discutidas e estudos adicionais são incentivados para testar outras condições e cenários.

Palavras-chave: Transtorno do Espectro Autista; Classificador; Atenção Visual; Mapa de Saliência; Características Semânticas.

ABSTRACT

FRANCO, F. O. (2024). Diagnostic aid approach for Autism Spectrum Disorder: integration of semantic features in visual attention models (Ph.D. Thesis). Graduate Program in Bioinformatics, Universidade de São Paulo, São Paulo.

It is estimated that 1% of the world's population has Autistic Spectrum Disorder (ASD), characterized by changes in social communication and the presence of stereotyped and repetitive behaviors. Studies suggest that the presence of changes in brain connectivity in ASD, with atypical processing of social information, related to bottom-up and top-down mechanisms. Several classifiers to diagnosis-aid of ASD have been proposed based on gaze-tracking data. However, most studies are based on data such as duration and count of gaze fixations. Other studies combine gaze-tracking data and extracted image features to create Visual Attention Models (VAM) and saliency maps. Recently, our group proposed an approach based on saliency maps to diagnosis-aid of ASD. Considering that changes in top-down connectivity are present in ASD, our hypothesis is that a greater number of semantic features in these VAM can improve the performance of these classifiers. Therefore, the aim of this study is to develop classifiers to diagnosis-aid of ASD, based on VAM considering different levels of features in saliency maps, focusing on the integration of semantic features. To achieve the objectives, we began by replicating the group's previous study with a new sample and new visual stimulus based on the same visual attention paradigm. The replication occurred satisfactorily, suggesting the robustness of the previous approach and the viability of approaches based on VAM to aid in the diagnosis of ASD. However, switching visual stimuli and samples had a negative impact on the performance of the classifiers. Although other factors need to be considered, the difference in performance between the previous study and the replication may be due to the difference in the ages of the individuals. After replication, other classification-inducing algorithms were tested and the Random Forest algorithm showed better performance in all scenarios, with the additional advantage of allowing comparison of the importance attributed to each of the features considered. A literature review was carried out to evaluate the feasibility of including more semantic features and then we proceeded with the extraction of new features. The comparison of the importance of the features contributed to the interpretability of the classifiers, in this sense, we verified that: 1 - semantic features presented lower relative importance than the others, but with greater importance for individuals with typical development; 2 - the "Person" feature differed more between groups, being less important in ASD; 3 - the features center of the screen and center of the scenes were the most important and their exclusion had a negative impact on the classification. In general, the inclusion of semantic features did not improve classification performance. However, when segmenting the dataset excluding younger individuals, it was observed that the performance improved. Some limitations are discussed and additional studies are encouraged to test other conditions and scenarios.

Keywords: Autism Spectrum Disorder; Classifier; Visual Attention; Saliency Map; Semantic Features.

Sumário

1	INTRODUÇÃO E REFERENCIAL TEÓRICO	10
1.1	Transtorno do Espectro Autista	10
1.1.1	<i>Caracterização</i>	10
1.1.2	<i>Dados epidemiológicos</i>	10
1.1.3	<i>Diagnóstico</i>	10
1.2	Rastreamento do olhar e paradigma visual	11
1.3	Processamento e atenção visual	13
1.3.1	<i>Vias de processamento visual</i>	13
1.3.2	<i>Atenção Visual</i>	14
1.4	Mapas de saliência	15
1.5	Características semânticas	17
1.6	Aprendizado de Máquina	18
1.6.1	<i>Definição</i>	18
1.6.2	<i>Métricas de avaliação</i>	19
1.6.3	<i>Aprendizado supervisionado e algoritmos indutores de classificação</i>	22
1.6.3.1	SVM	22
1.6.3.2	NN	22
1.6.3.3	Árvores de Decisão	22
1.6.3.4	RF	23
1.7	Aprendizado de máquina com rastreamento do olhar aplicado ao TEA	23
1.8	Estudos anteriores do grupo	26
2	JUSTIFICATIVA E PROPOSTA	29
3	HIPÓTESE	30
4	OBJETIVOS	30
4.1	Geral	30
4.2	Específicos	30
5	MATERIAIS E MÉTODOS	30
5.1	Ética e armazenamento de dados	31
5.2	Amostra	32
5.3	Dados de rastreamento do olhar	34

5.4	Paradigma e estímulo visual	35
5.5	Construção do classificador	37
5.5.1	<i>Construção de mapas de saliência</i>	38
5.5.2	<i>Classificação dos indivíduos</i>	38
5.6	Validação do estudo anterior	39
5.6.1	<i>Experimento 1</i>	39
5.6.2	<i>Experimento 2</i>	39
5.7	Alterações Metodológicas	40
5.8	Substituição do algoritmo indutor de classificação	41
5.8.1	<i>SVM</i>	41
5.8.2	<i>Random Forest</i>	41
5.9	Adição de novas características	42
5.9.1	<i>Levantamento de características semânticas</i>	42
5.9.1.1	<i>Elementos faciais</i>	44
5.9.1.2	<i>Expressões faciais</i>	45
5.9.1.3	<i>Direção do olhar</i>	45
5.9.1.4	<i>Gestos manuais</i>	45
5.9.2	<i>Extração das novas características</i>	45
5.10	Normalização de características	48
5.11	Análise Estatística	49
6	RESULTADOS E DISCUSSÃO	49
6.1	Análises com o <i>dataset I</i> - Todos os indivíduos	49
6.1.1	<i>Caracterização do dataset I</i>	49
6.1.2	<i>Validação do estudo anterior</i>	50
6.1.3	<i>Desempenho dos classificadores com diferentes datasets</i>	52
6.1.4	<i>Adição de características semânticas e avaliação de desempenho dos classificadores de indivíduos - Dataset I</i>	53
6.1.5	<i>Contribuição das características - Dataset I</i>	54
6.2	Análises com o <i>dataset II</i> - Exclusão de indivíduos menores de 4 anos	58
6.2.1	<i>Caracterização do dataset II</i>	58
6.2.2	<i>Desempenho dos classificadores de indivíduos - Dataset II</i>	59
6.2.3	<i>Contribuição das características - Dataset II</i>	61

6.3	Análises com o <i>dataset</i> III - Exclusão de TEA leve	64
6.3.1	<i>Caracterização do dataset III</i>	64
6.3.2	<i>Desempenho dos classificadores de indivíduos - Dataset III</i>	64
6.3.3	<i>Contribuição das características - Dataset III</i>	65
6.4	Análises com o <i>dataset</i> IV - Exclusão de DI	69
6.4.1	<i>Caracterização do dataset IV</i>	69
6.4.2	<i>Desempenho dos classificadores de indivíduos - Dataset IV</i>	69
6.4.3	<i>Contribuição das características - Dataset IV</i>	71
6.5	Análises com a exclusão das características de centro	74
6.6	Principais achados	76
6.7	Limitações do estudo	77
7	CONCLUSÕES	79
8	PUBLICAÇÕES	80
	REFERÊNCIAS	81
	APÊNDICE A - TABELA COMPLEMENTAR COM O LEVANTAMENTO DE CARACTERÍSTICAS SEMÂNTICAS	94
	APÊNDICE B - GEOLOCALIZAÇÃO DOS INDIVÍDUOS	95
	APÊNDICE C - DESEMPENHO DOS CLASSIFICADORES CONSIDERANDO APENAS CARACTERÍSTICAS SEMÂNTICAS	96
	APÊNDICE D - DOCUMENTAÇÃO DO CÓDIGO BASE	97
	ANEXO A - APROVAÇÃO PELO COMITÊ DE ÉTICA	106

Lista de Figuras

1	Representação do processamento visual neuronal em diferentes níveis.	15
2	Representação simplificada da elaboração de mapas de saliência.	16
3	Tipos de aprendizado de máquina.	19
4	O <i>trade-off</i> entre sensibilidade e especificidade representado na Curva ROC. . .	21
5	Representação esquemática das etapas de desenvolvimento do classificador. . .	26
6	Identificação dos classificadores de pixel e de indivíduos.	29
7	Fluxograma sobre os indivíduos incluídos e excluídos da pesquisa.	33
8	Descrição dos <i>datasets</i>	33
9	Intersecção entre os <i>datasets</i> para o grupo TEA	34
10	Exemplos de <i>frames</i> do estímulo utilizado.	36
11	Exemplos de mapas de fixação.	39
12	Representação das alterações metodológicas realizadas.	41
13	Malha de pontos utilizada - <i>landmarks</i>	46
14	Exemplos da detecção de gestos manuais.	48
15	Exemplos das novas características extraídas.	53
16	Importância relativa das características - <i>Dataset I</i> : 28 características.	56
17	Importância relativa das características - <i>Dataset I</i> : 34 características.	57
18	Desempenho dos classificadores com <i>Random Forest</i> e adição das novas caracte- rísticas para os <i>datasets I, II, III e IV</i>	60
19	Importância relativa das características - <i>Dataset II</i> : 28 características.	62
20	Importância relativa das características - <i>Dataset II</i> : 34 características.	63
21	Importância relativa das características - <i>Dataset III</i> : 28 características.	66
22	Importância relativa das características - <i>Dataset III</i> : 34 características.	68
23	Importância relativa das características - <i>Dataset IV</i> : 28 características.	72
24	Importância relativa das características - <i>Dataset IV</i> : 34 características.	73
25	Desempenho dos classificadores com a adição das novas características para os <i>datasets I, II, III e IV</i>	74
26	Mapas com a geolocalização dos indivíduos do estudo.	95

Lista de Tabelas

1	Matriz de confusão para classificador binário.	19
2	Levantamento da literatura em relação ao uso de rastreamento ocular no Trans- torno do Espectro Autista	25
3	Descrição das 28 características utilizadas no estudo prévio (Oliveira <i>et al.</i> [1]), e que são consideradas no presente estudo.	27
4	Comparação de valores médios para as características extraídas dos vídeos. . . .	37
5	Levantamento de características semânticas passíveis de serem extraídas no es- tímulo utilizado.	43
6	Caracterização da amostra - <i>dataset</i> Anterior (descrito em detalhes em Oliveira <i>et al.</i> [1]) versus I (<i>dataset</i> atual considerando todos os indivíduos).	50
7	Resultados de replicação do estudo prévio e implementação do Random Forest.	50
8	Desempenho dos classificadores com a adição das novas características semânti- cas ao <i>dataset</i> I.	52
9	Caracterização da amostra - <i>dataset</i> I versus II.	59
10	Desempenho dos classificadores com o <i>dataset</i> II.	60
11	Caracterização da amostra - <i>dataset</i> I versus III.	64
12	Desempenho dos classificadores com o <i>dataset</i> III.	65
13	Caracterização da amostra - <i>dataset</i> I versus IV.	69
14	Desempenho dos classificadores com o <i>dataset</i> IV.	70
15	Desempenho dos classificadores com a exclusão das características de centro para os <i>datasets</i> I, II, III e IV.	75
16	Características diferentemente importantes entre os grupos.	77
17	Desempenho dos classificadores considerando apenas características semânti- cas.	96

Siglas

ACM	Association for Computing Machinery Database	39
ADI-R	Autism Diagnostic Interview-Revised	11
ADOS	Autism Diagnostic Observation Schedule	10
AU	Action Unit	42
AUC	Área Sob a Curva ROC	13
CARS	Childhood Austim Rating Scale	11
CBCL	Child Behavior Checklist	11
CID	Classificação Internacional de Doenças da OMS	10
DI	Deficiência Intelectual	31
DSM	Manual Diagnóstico e Estatístico de Transtornos Mentais	10
DT	Desenvolvimento Típico	13
FN	Falso Negativo	19
FP	Falso Positivo	19
GeoPref	Paradigma de Preferência Visual com Cenas Geométricas	13
IC	Intervalo de Confiança	46
IDE	Integrated Development Environment	35
IEEE	Institute of Electrical and Electronics Engineers Database	39
LApIS	Laboratório de Aplicações de Informática em Saúde	26
LGPD	Lei Geral de Proteção de Dados	31
MAV	Modelo de Atenção Visual	13
N	Negativo Observado	19
N'	Negativo Predito	19

NN	Neural Network	22
OOB	Out-of-Bag.....	23
P	Positivo Observado	19
P'	Positivo Predito.....	19
PRONAS/PCD	Programa Nacional de Apoio à Atenção da Saúde da Pessoa com Deficiência	31
PROTEA	Programa de Transtornos do Espectro Autista.....	26
QI	Quociente de Inteligência.....	11
RF	Random Forest	22
ROC	Receiver Operating Characteristic.....	20
ROI	Region of Interest	13
SON-R	Snijders-Oomen Non-verbal Intelligence Test - Revised	31
SVM	Support Vector Machine	22
TEA	Transtorno do Espectro Autista	10
TPR	Taxa de Verdadeiro Positivo	20
USP	Universidade de São Paulo	31
VC	Vineland Communication.....	31
VDL	Vineland Daily Living	31
VL	Vineland Locomotion.....	31
VN	Verdadeiro Negativo.....	19
VP	Verdadeiro Positivo	19
VS	Vineland Socialization	31
WISC	Wechsler Intelligence Scale for Children	31

1 INTRODUÇÃO E REFERENCIAL TEÓRICO

1.1 Transtorno do Espectro Autista

1.1.1 Caracterização

O Transtorno do Espectro Autista (TEA) é um transtorno do neurodesenvolvimento, caracterizado pela deficiência na comunicação social e padrões comportamentais restritos e repetitivos [2–5]. A dificuldade em interpretar expressões faciais e linguagem corporal, pode prejudicar o aprendizado e interações sociais de indivíduos com TEA [6, 7].

O aprendizado baseado na observação social atende dois critérios básicos: 1 - concentrar a atenção nas partes mais informativas da cena; 2 - adaptar seu comportamento em tempo real de acordo com o contexto [8]. A crescente complexidade de interação social é um produto da seleção natural de um sistema de processamento visual cada vez mais elaborado [6].

1.1.2 Dados epidemiológicos

Dados epidemiológicos de populações de diferentes regiões do mundo podem apresentar uma grande variabilidade. No Brasil, apesar dos dados sobre TEA serem escassos, um estudo piloto de 2011 estimou uma prevalência de 0,3% em crianças de 7 a 12 anos de idade de um município no Estado de São Paulo [9]. Na população estadunidense, em 2021, estimou-se que a prevalência do TEA aos 8 anos de idade era de 1 para 44 indivíduos [10]. Globalmente a estimativa atual de prevalência é de 1/100 crianças. As estimativas aumentaram ao longo do tempo, possivelmente por alterações dos critérios diagnósticos [11], maior conhecimento do quadro pelos profissionais de saúde e da população em geral.

A prevalência do TEA entre os sexos é desproporcional, com números que variam de 2,3 a 3,5 indivíduos do sexo masculino para cada indivíduo do sexo feminino. Até o momento, os motivos para essa diferença ainda não foram plenamente elucidados [12–14].

1.1.3 Diagnóstico

O diagnóstico do TEA é clínico, realizado com base nos critérios estabelecidos no Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM) 5 [2] e na Classificação Internacional de Doenças da OMS (CID) 11 [15]. Existem dois instrumentos usados como “padrão-ouro” para o diagnóstico: o Sistema Diagnóstico de Observação do Autismo/*Autism Diagnostic Observation Schedule* (ADOS) 2 [16] e a Entrevista Diagnóstica de Autismo Revisada/*Autism Diagnostic*

Interview-Revised (ADI-R) [17]. A ADI-R consiste em uma entrevista com 93 itens aplicada ao cuidador, enquanto que o ADOS consiste em um protocolo de observação comportamental. Ambos instrumentos são aplicados por profissional treinado e a aplicação pode demorar até 2,5 horas para conclusão [18]. Além da demora e do custo envolvido, o acesso a esses profissionais treinados é escasso longe de grandes centros, os quais possuem área de saúde mais desenvolvida.

Por tratar-se de um espectro, o TEA contempla uma ampla heterogeneidade fenotípica de apresentação, com diferentes níveis de comprometimento de interação social, linguagem, cognição e habilidades motoras. Embora o DSM-5 classifique o grau de severidade em três níveis, de acordo com o nível de apoio necessário, esses níveis não são utilizados para a provisão de serviços ou tratamentos [2]. Na literatura, de maneira genérica a severidade é avaliada pela escala *Childhood Autism Rating Scale* (CARS) [19] e o comprometimento cognitivo é avaliado por testes de Quociente de Inteligência (QI). Para o prognóstico e direcionamento do tratamento, vários instrumentos e escalas são aplicados para avaliar os eixos, como por exemplo a avaliação cognitiva funcional por meio da Escala de Comportamento Adaptativo Vineland [20], bem como a presença de comorbidades, realizada por meio do *Child Behavior Checklist* (CBCL) [21]. Dentro deste contexto, fica clara a necessidade de uma equipe multidisciplinar altamente treinada para avaliar esses aspectos. Assim, a busca por biomarcadores e ferramentas computacionais de auxílio diagnóstico são essenciais no cenário atual, e a análise do rastreamento do olhar tem se mostrado um método promissor [22–24] para avançar no acesso ao diagnóstico do TEA.

1.2 Rastreamento do olhar e paradigma visual

O rastreador do olhar é um sistema que possibilita avaliar a atenção espontânea e a preferência do olhar registrando o foco visual do usuário [25]. O método mais comum de rastreamento do olhar baseia-se na estimativa da posição relativa da pupila em relação ao reflexo da córnea, na qual a posição do olhar é estimada com precisão [26]. Enquanto imagens ou vídeos são exibidos ao indivíduo, o posicionamento do olhar é estimado a partir de câmeras remotas, constituindo um método não invasivo. Não há, portanto, necessidade de equipamentos fixados na cabeça ou outros dispositivos que podem causar desconforto ao usuário, facilitando o uso em crianças [27] e, ainda, em indivíduos com TEA.

De maneira simplificada, os movimentos oculares podem ser interpretados como uma sequên-

cia de fixações e sacadas. Uma fixação é um período em que o olhar permanece em um determinado local. Uma sacada, por outro lado, é um movimento rápido dos olhos entre duas fixações consecutivas [28]. Essas duas métricas são as mais comuns em pesquisas de rastreamento do olhar [29].

A análise do rastreamento do olhar para detectar os sinais presentes no TEA envolve a escolha do paradigma a ser utilizado, o qual depende dos objetivos da análise e do público-alvo. Muitos paradigmas têm sido propostos na literatura, tais como:

- Rastreamento de objetos múltiplos, em que o observador deve acompanhar um subconjunto de objetos que se movem em diferentes trajetórias [30, 31];
- Campo de visão útil, que baseia-se na região do campo visual da qual um observador pode extrair informações em um relance [32];
- Contingente ao olhar, em que a imagem é desfocada, exceto uma pequena região, um *spot*. Nessa condição, o espectador tem dificuldade em processar informações periféricas e processa principalmente informações por meio de sua atenção focal [33, 34];
- Visualização livre, em que o observador explora o estímulo livremente. No TEA é comum a apresentação de cenas sociais, principalmente de faces [35].
- Preferência visual estática, que consiste na apresentação de imagens simultâneas na tela, sendo um lado com estímulo social e o outro não social, geralmente formas geométricas [36];
- Preferência visual dinâmica, semelhante ao anterior, porém com a apresentação de vídeos ao invés de imagens [36–38];
- Atenção compartilhada, em que o indivíduo recebe dicas gestuais de outro indivíduo, direcionando para onde ele deve olhar [39].

Os paradigmas que consideram a atenção social do indivíduo, como por exemplo a preferência visual, têm sido amplamente utilizados no estudo do TEA [36, 38, 40–42], assim como paradigmas de atenção compartilhada [43–45], uma vez que a habilidade social depende principalmente da interpretação acurada do ambiente social [46] e no TEA essa interpretação parece ser deficitária [47], além de haver uma grande variabilidade interindividual da percepção social [48].

Estudos de rastreamento do olhar mostram que indivíduos com Desenvolvimento Típico (DT) têm uma tendência inata de olhar para a região facial de outros humanos, destacadamente os olhos, característica já presente aos três meses de idade [49]. Também foi evidenciada uma preferência por movimento biológico, ou seja, cenas com pessoas em movimento [50, 51], que são fatores que favorecem o aprendizado baseado na observação social. Por outro lado, no TEA isso não ocorre da mesma maneira [6, 52]. No TEA há uma maior preferência por imagens geométricas [53] e menor fixação na região dos olhos, com esse último possivelmente correlacionado com o nível de deficiência social [52].

Pierce *et al.* [53] utilizaram o paradigma de preferência visual com cenas geométricas (Geo-Pref) e obtiveram uma Área Sob a Curva/*Area Under the ROC Curve* (AUC) de 0,71 na classificação de DT e TEA. Moore *et al.* [54] levantaram a hipótese de que cenas sociais mais complexas aumentariam a discriminação entre DT e TEA, e propuseram o GeoPref complexo. Porém, não observaram melhora na classificação e apontaram que um dos possíveis motivos é não terem considerado diferenças nas propriedades visuais de baixo nível, como cor e contraste. Os Modelos de Atenção Visual (MAVs) obtiveram relevância para melhor compreender o TEA uma vez que abordagens computacionais podem ser implementadas considerando características do modelo visual humano em vez das Regiões de Interesse/*Regions of Interest* (ROIs) [55].

1.3 Processamento e atenção visual

1.3.1 *Vias de processamento visual*

Os dados visuais processados pela retina são transmitidos ao cérebro pelo nervo óptico, passando pelo quiasma óptico, onde os nervos ópticos direito e esquerdo se unem. Esses dados seguem para o tálamo e, em seguida, para o córtex visual no lobo occipital, ativando a área visual V1 [56, 57]. A partir desta área sinais são transmitidos por duas vias diferentes: a dorsal e a ventral. O fluxo ventral é, principalmente, responsável pela identificação de objetos, levando informação, por exemplo, sobre a forma e a cor. A via dorsal é dedicada a guiar o movimento visual, com células seletivas para a direção do movimento. Estas vias não são estritamente separadas e existe uma interligação substancial entre elas [58–60].

A integração de entradas (*inputs*) sensoriais com sinais neuronais provenientes de áreas associativas cerebrais é um processo dinâmico que envolve a coordenação motora do sistema visual, assim como do movimento da cabeça, que visa a adquirir o menor erro de predição da entrada sensorial pelas áreas associativas, de acordo com o modelo de codificação preditiva

[61]. Alterações neste fino processo pode alterar a interpretação da informação. Sabe-se que no TEA o processamento e a integração sensorial são atípicos [62, 63]. Muitos estudos têm se dedicado a estudar este ponto [64–66]; no entanto, a natureza desses problemas sensoriais permanece incerta.

1.3.2 *Atenção Visual*

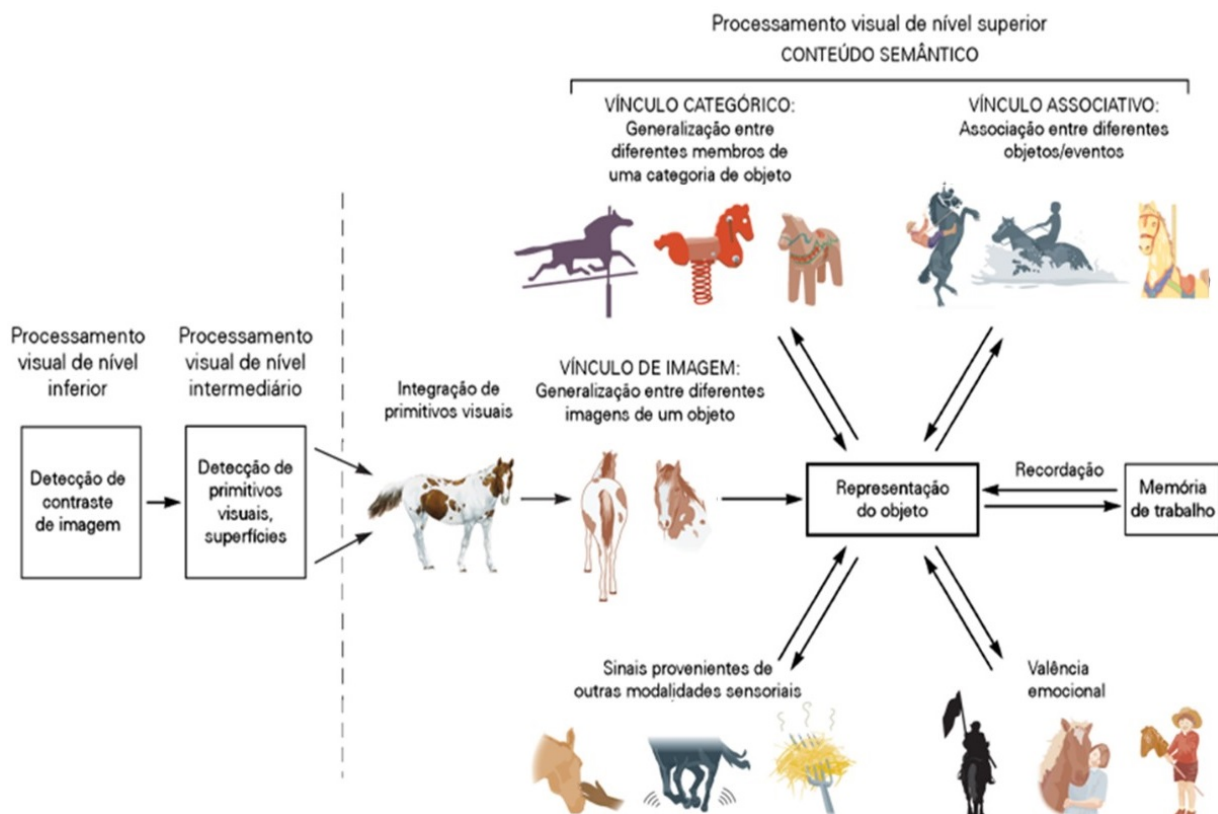
Constantemente recebemos uma grande quantidade de informações visuais e processar todos esses dados em tempo real seria uma tarefa difícil [67, 68]. A atenção visual, de maneira seletiva, direciona nosso olhar rapidamente para objetos de interesse em nosso ambiente visual. Essa capacidade de selecionar objetos salientes em uma cena visual desordenada confere uma vantagem evolutiva, possibilitando a detecção rápida de possíveis presas, companheiros ou predadores [69]. O termo atenção visual é bastante amplo, incluindo muitas formas, como a atenção espacial, temporal, baseada em objetos e em características [70], sendo que essas não são mutuamente exclusivas [67].

A rápida varredura periférica de um campo visual é essencialmente uma tentativa de ligar os pontos focais, ou microestruturas, de acordo com a percepção do observador [71]. Essa análise exploratória resulta em uma consciência cognitiva e afetiva, consciente e inconsciente dos elementos visuais (como cor, textura e forma) dentro da estrutura microestrutural do campo visual, que dependem da percepção e das experiências estéticas evocadas no espectador durante o ato de construção de significado [71].

Conforme ilustrado na figura 1, para reconhecer, entender e interagir com o mundo ao nosso redor, o cérebro cria representações de objetos. Essas representações são compostas de informações visuais, auditivas, táteis e outras, incluindo a forma, o tamanho, a cor, a textura, a valência emocional e a associação a outros objetos ou eventos. As representações de objetos são armazenadas na memória de trabalho e de longo prazo, e estão constantemente sendo refinadas à medida que interagimos com o mundo ao nosso redor [58].

Podemos dividir a atenção visual em dois principais mecanismos, o *bottom-up* e o *top-down*. No *bottom-up*, estímulos externos atraem a atenção devido a suas características inerentes como, por exemplo, o contraste visual em relação ao fundo [70]. Neste mecanismo, a atenção é orientada principalmente por dados que necessitam de um processamento visual de baixo nível, baseando-se em características da imagem, como cor, orientação e intensidade [60, 72]. Por outro lado, o mecanismo *top-down* é impulsionado pela orientação interna da

Figura 1: Representação do processamento visual neuronal em diferentes níveis.



Fonte: adaptado de Kandel *et al.* [58].

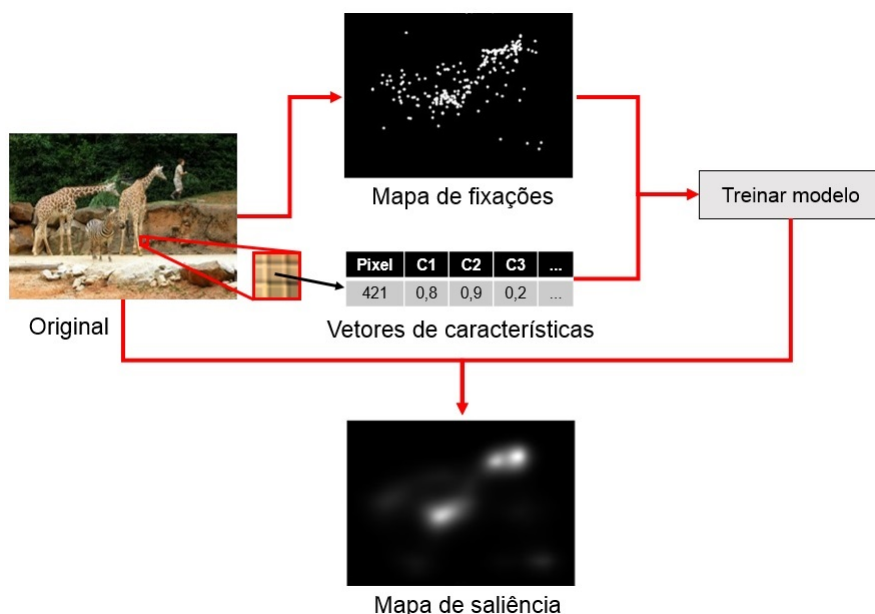
atenção, com base em conhecimentos prévios, planos e metas atuais [68, 73], ou seja, utiliza informações semânticas ou relativas a um contexto [74].

1.4 Mapas de saliência

Ao considerarmos padrões de atenção visual, podemos elaborar um mapeamento da importância relativa de cada parte de uma cena, conhecido como mapas de saliência [73, 75], exemplificado na figura 2. Na representação temos uma imagem original, que é composta por vários pixels e cada pixel possui um vetor de características. Esses vetores juntamente com o mapa de fixações (*i.e.*, representação bidimensional das localizações de todas as fixações a serem analisadas [76]) específico da imagem em questão são utilizados no treinamento do modelo. Após isso, ao aplicar o modelo já treinado em uma imagem (podendo ser utilizada ou não para o treinamento), obtemos o mapa de saliência [73]. Assim, quando mapeamos a importância relativa de partes de uma cena, estamos buscando os pixels mais salientes [77].

Os mapas de saliência são obtidos por modelos que consideram muitos fatores, por exem-

Figura 2: Representação simplificada da elaboração de mapas de saliência.



As setas vermelhas indicam o fluxo do processo. Abreviação: C, característica.

Fonte: adaptado de Cornia *et al.* [78].

plo, **como**, **quando** e o **porquê** selecionamos elementos da imagem que são relevantes. Uma abordagem geral é inspirar-se na anatomia e na funcionalidade do sistema visual humano, que é altamente adaptado para resolver esses problemas [67]. Estudos de neuroimagem têm demonstrado que no TEA existem alterações da conectividade cerebral fazendo com que a integração de sistemas *bottom-up* e *top-down* fique prejudicada [79, 80]. Isto pode acarretar em alterações na integração entre as entradas dos sistemas sensoriais [81] e circuitos cerebrais de alto nível responsáveis pelo foco atencional, valoração e tomada de decisão [82].

O grupo do professor Olivier Le Meur tem contribuído significativamente para o desenvolvimento de algoritmos de predição de mapas de saliência, buscando prever especificamente as fixações de crianças com TEA, ou seja, os pixels mais relevantes da imagem de acordo com as características selecionadas, sem classificar o indivíduo. O método propõe a utilização de redes convolucionais para a criação de três mapas de atenção para posterior concatenação em um único mapa de saliência [83, 84].

Em 2015, Wang *et al.* [85] propuseram um modelo de saliência em três camadas, considerando características da imagem em nível de pixel (exemplo: cor), objeto (exemplo: forma) e semântico (exemplo: faces), visto que, muitas vezes a interpretação de contexto é deficiente no TEA e que a maioria dos estudos de rastreamento do olhar baseiam-se no tempo de fixação em áreas de interesse definidas *a priori*. Confirmando a importância dessa abordagem, os

autores mostraram que existe diferença principalmente quanto às características semânticas em indivíduos com TEA quando comparados ao controle. No entanto, o estudo não objetivou a classificação de TEA e controles ou para diferenciar subtipos de TEA, apenas compreender melhor possíveis alterações na alocação espontânea de atenção entre os grupos.

1.5 Características semânticas

Conceitos abstratos podem gerar dubiedade [86, 87], por isso, seu emprego pode variar de acordo com o autor ou contexto. Para fins de clareza, a definição de “semântica” ora adotada foi a apresentada por Sheth *et al.* [88]:

Semântica é o significado atribuído aos conceitos e suas relações dentro da mente. A rede de conceitos e relações é utilizada para representar o conhecimento sobre o mundo, que por sua vez possibilita a cognição e a percepção necessárias para interpretar as experiências cotidianas. Os conceitos semânticos representam, unificam, agrupam ou mapeiam vários padrões de dados e, assim, fornecem uma abstração conceitual que oculta as diferenças sintáticas e representacionais dos dados. Geralmente, isso envolve o mapeamento de observações de vários estímulos físicos, como sinais visuais ou de fala, para conceitos e relacionamentos como os humanos os interpretam e comunicam.

Ainda neste sentido, podemos complementar com a visão de Bussmann [89], de que recursos semânticos são construções abstratas que representam as condições cognitivas e sociais segundo as quais o mundo circundante é categorizado pelos humanos. O uso de categorizações ou símbolos é fundamental no contexto de objetos e ações para estabelecer vínculos semânticos referenciais e, sem quaisquer relações referenciais, o conhecimento do significado é impossível. Esta questão é conhecida como problema da fundamentação simbólica [90].

Em suma, características semânticas são elementos ou estruturas presentes na imagem que tenham um significado para o observador, como por exemplo, a presença de um carro na imagem, que em condições normais, pode fornecer dicas sobre o contexto da cena, trazendo ideias de urbanização, movimento, poder aquisitivo de pessoas ao redor, dentre outras.

A investigação de características semânticas em abordagens computacionais para auxílio ao diagnóstico do TEA é fundamental, visto que esses indivíduos podem envolver menos efetivamente os mecanismos frontais no controle *top-down* do conhecimento semântico, priorizando processos visuais no córtex occipital que estão desconectados da semântica [91]. Adicionalmente, no TEA ocorre uma deficiência na comunicação social [2–5] e há evidências de que a

sociabilidade se reflete na organização dos sistemas neurais que suportam o processamento semântico [92].

Kiani *et al.* [93], mostraram que as pessoas com TEA prestam menos atenção a partes das imagens que incluem características semânticas e comunicativas. Por outro lado, passam mais tempo prestando atenção às partes de fundo da imagem que incluem padrões repetitivos.

1.6 Aprendizado de Máquina

1.6.1 Definição

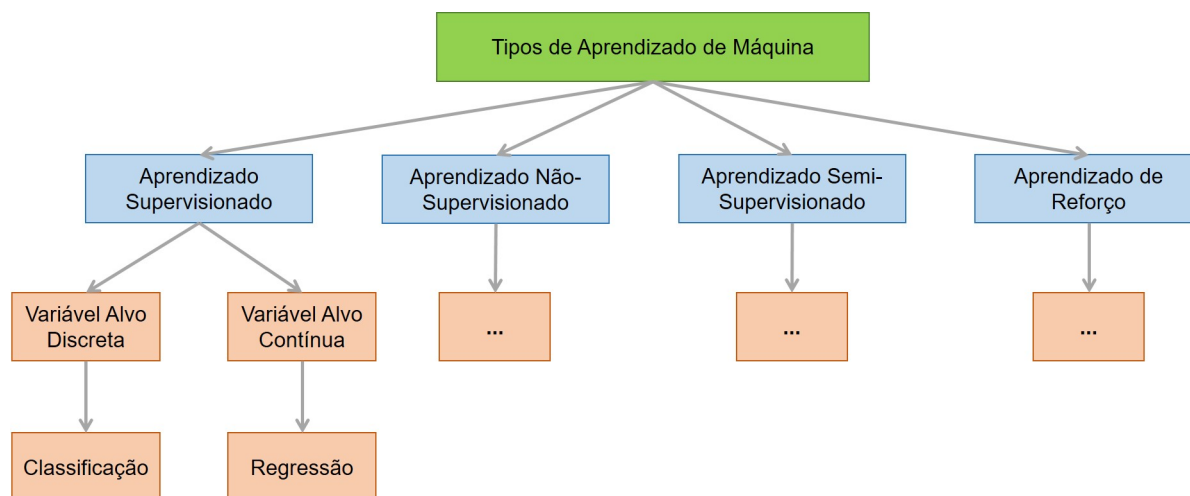
Segundo Arthur Samuel, um dos pioneiros da área, o aprendizado de máquina pode ser definido como o campo de estudo que fornece aos computadores a capacidade de aprender sem serem explicitamente programados [94]. O aprendizado de máquina refere-se ao processo de ajustar modelos de acordo com os dados ou identificar agrupamentos que sejam informativos, portanto tentar reconhecer padrões em dado conjunto de dados [95] e aplicar esses padrões “aprendidos” em novos dados com características similares. Esse processo é particularmente útil quando o conjunto de dados a analisar é muito grande ou complexo para a capacidade humana e/ou para automatizar a análise de dados [95]. A automatização possibilita a confecção de fluxos de análise (*pipelines*) reproduzíveis, além da possibilidade de redução de tempo de análise [95, 96].

Em Psicologia Clínica e Psiquiatria, o aprendizado de máquina tem sido aplicado a estruturas de dados complexas para auxiliar na conceituação de transtornos mentais, detectar e prever o risco e a trajetória dos sintomas e estudar o resultado do tratamento, resposta diferencial a tratamentos, dentre outras aplicações [97].

Os algoritmos de aprendizado de máquina podem ser divididos em quatro categorias, baseando-se no tipo de treinamento: aprendizado supervisionado, aprendizado não-supervisionado, aprendizado semi-supervisionado e aprendizado por reforço [94, 98], conforme representado na figura 3. Esta é uma possível classificação dos algoritmos de acordo com Sarker [98], outras classificações são propostas, de acordo com outros critérios. No presente estudo, focaremos exclusivamente no aprendizado supervisionado.

A eficácia e a eficiência de uma solução de aprendizado de máquina dependem da natureza das características dos dados e do desempenho dos algoritmos de aprendizado especificamente para o tipo de dado utilizado [98]. A avaliação de desempenho de uma solução pode ser realizada por diferentes métricas que podem ser mais úteis de acordo com os objetivos da proposta

Figura 3: Tipos de aprendizado de máquina.



As reticências indicam que existem outras subdivisões que não foram representadas na figura.

Fonte: adaptado de Sarker [98].

apresentada.

1.6.2 Métricas de avaliação

As medidas de avaliação de desempenho têm papel essencial no desenvolvimento, seleção e avaliação do modelo de aprendizado de máquina [99]. Há diversas métricas para mensurar o desempenho de classificação, muitas delas baseadas no conceito de matriz de confusão, algumas das principais são: AUC, acurácia, sensibilidade, especificidade, precisão e F1-score [99].

A matriz de confusão é comumente utilizada na avaliação de tarefas de classificação. Para classificação binária, ela constitui em uma matriz dois por dois com a classe observada (real) e classe predita pelo classificador [99], conforme a tabela 1. Na diagonal principal são expressos os acertos (verdadeiros positivos + verdadeiros negativos) e na diagonal secundária são expressos os erros de classificação (falsos positivos + falsos negativos).

Tabela 1: Matriz de confusão para classificador binário.

		Predito		Total
		Positivo	Negativo	
Observado	Positivo	VP	FN	P
	Negativo	FP	VN	N
Total		P'	N'	

Abreviações: VP, Verdadeiro Positivo; FN, Falso Negativo; P, Positivo Observado; FP, Falso Positivo; VN, Verdadeiro Negativo; N, Negativo Observado; P', Positivo Predito; N', Negativo Predito.

As métricas baseadas na matriz de confusão são descritas a seguir:

- **Acurácia:** é a medida mais simples e comum derivada da matriz de confusão, porém essa métrica é tendenciosa para a classe majoritária [99], o que é um ponto fraco ao trabalhar com classes desbalanceadas. A acurácia é dada pela equação 1.

$$Acurácia = \frac{VP + VN}{P + N} = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

- **Sensibilidade:** também é chamada de Taxa de Verdadeiro Positivo (TPR) ou *recall* e é definida como a taxa de verdadeiros positivos dentre todos os reais positivos [99]. A sensibilidade é dada pela equação 2.

$$Sensibilidade = \frac{VP}{P} = \frac{VP}{VP + FN} \quad (2)$$

- **Especificidade:** é definida como a taxa de verdadeiros negativos dentre todos os reais negativos [99]. A especificidade é dada pela equação 3.

$$Especificidade = \frac{VN}{N} = \frac{VN}{FP + VN} \quad (3)$$

- **Precisão:** é a probabilidade de uma presença ser observada dado que foi prevista [100]. A diferença entre sensibilidade e precisão é que na sensibilidade calcula-se os valores verdadeiramente positivos dentre todos as observações positivas reais, enquanto que na precisão calcula-se os valores verdadeiramente positivos dentre todos as observações previstas como positivas pelo classificador. A precisão é dada pela equação 4.

$$Precisão = \frac{VP}{P'} = \frac{VP}{VP + FP} \quad (4)$$

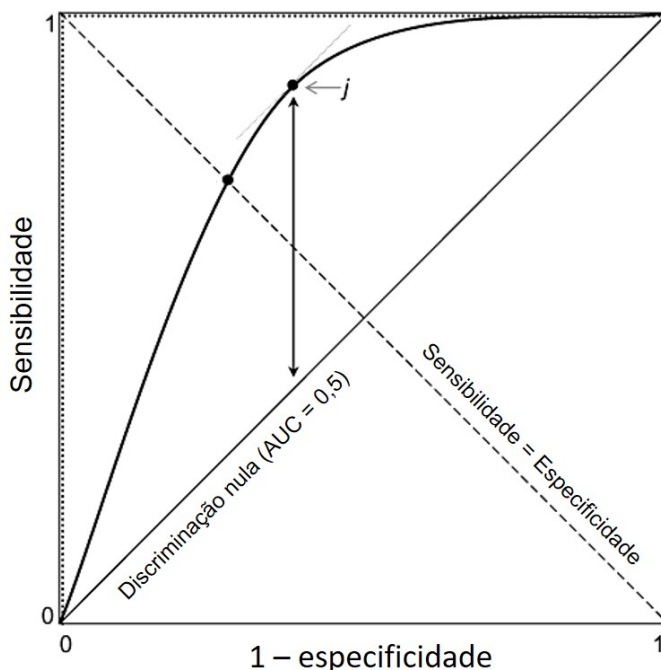
Abreviações: VP, Verdadeiro Positivo; P', Predito como Positivo; FP, Falso Positivo.

- **F1-score:** é uma métrica de desempenho equilibrada, que considera tanto a sensibilidade quanto a precisão. A medida F1-score é comumente conhecida como média harmônica, dada pela equação 5.

$$F1\text{-score} = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade} \quad (5)$$

- AUC: é uma medida robusta da capacidade discriminatória de modelos de classificação dada pelo cálculo da área sob a curva ROC (*Receiver Operating Characteristic*), ou seja, do gráfico de sensibilidade *versus* 1-especificidade entre os limiares [100], conforme representado na figura 4. Um modelo binário pode ser considerado para discriminar melhor do que o acaso se a curva estiver acima da diagonal de discriminação nula ($AUC = 0,5$), situação em que o modelo não fornece nenhum poder de separação [101]. O ROC é considerado como um *trade-off* entre a taxa de verdadeiros positivos e a taxa de falsos positivos. Uma vantagem do uso da curva ROC é que ela é pouco sensível a alterações de balanceamento entre as classes. Outra vantagem é que as curvas esgotam todos os limiares possíveis, contendo assim mais informações do que métricas dependentes de um limiar estático, como as métricas anteriores (acurácia, sensibilidade, especificidade, precisão e F1-score) [99].

Figura 4: O *trade-off* entre sensibilidade e especificidade representado na Curva ROC.



A curva ROC (*Receiver Operating Characteristic*) ideal teórica é representada pela linha pontilhada (curva de um modelo mostrando discriminação perfeita). Na linha preta mais grossa é mostrada uma curva para um modelo mostrando discriminação imperfeita (exemplo real). A diagonal de discriminação nula ($AUC = 0,5$) corresponde a uma classificação binária aleatória. “j” representa o índice de Youden, onde a distância vertical da curva ROC até a diagonal de discriminação nula atingem seus valores máximos, ou seja um ponto de otimização para a escolha do limiar de *trade-off* entre sensibilidade e especificidade.

Fonte: adaptado de Jiménez-Valverde [101].

1.6.3 *Aprendizado supervisionado e algoritmos indutores de classificação*

O aprendizado supervisionado engloba vários métodos que usa dados de treinamento com classes ou valores conhecidos para inferir uma função. Conforme representado na figura 3, as tarefas supervisionadas podem ser: 1 - “classificação”, para a separação de dados em classes predefinidas ou categorias; 2 - “regressão”, para predição de valores contínuos [98, 102].

A tarefa de classificação pode ser binária, quando há apenas duas classes; multiclasse, quando há mais de duas classes; ou multirótulo, quando uma observação pode ter mais de um rótulo ao mesmo tempo e, portanto, pertencer a mais de uma classe. Esses rótulos geralmente são fornecidos através de um julgamento humano [94].

Existem muitos algoritmos indutores de classificação, alguns dos principais são: *Support Vector Machine* (SVM); *Neural Network* (NN); Árvore de Decisão e *Random Forest* (RF) [94, 103].

1.6.3.1 SVM: é um algoritmo complexo capaz de realizar tarefas de classificação de dados linearmente ou não-linearmente separáveis. O SVM é baseado no conceito de maximizar a distância mínima do hiperplano (fronteira de decisão que divide o espaço amostral) ao ponto amostral mais próximo [94]. Algumas das vantagens é a robustez para dados de alta dimensão (*i.e.*, conjuntos de dados contendo mais características do que observações [104]) e a boa capacidade de generalização (*i.e.*, manter o desempenho para novas amostras [94]). No entanto, seu desempenho é altamente dependente da escolha dos parâmetros [94].

1.6.3.2 NN: são estratégias computacionais vagamente baseados na estrutura neuronal animal, mas em escalas muito menores [105]. Uma NN contém um conjunto de elementos de processamento chamados neurônios que são componentes inter-relacionados distribuídos em camadas [106]. Algumas vantagens são a aplicação em problemas bastante dinâmicos e capacidade de modelar funções complexas [94], a depender da densidade de neurônios e camadas. Por outro lado, o desempenho também é sensível aos parâmetros escolhidos e sobreajuste (*overfitting*) [94].

1.6.3.3 Árvore de Decisão: a base teórica para a construção de árvores de decisão é simples. Ao dividir o conjunto de dados em subconjuntos menores, a árvore é construída de forma incremental [107]. Para essas divisões uma característica é considerada por vez e categorizada para que o conjunto atual seja subdividido. Esse processo acontece de forma recorrente até que algum critério de parada seja atingido ou os subconjuntos gerados não possam mais serem

divididos. As árvores de decisão são fáceis de interpretar e explicar e por ser um método não paramétrico, os *outliers* não afetam tanto o modelo. No entanto, é difícil lidar com dados de alta dimensão e os erros se propagam pelas árvores, o que pode se tornar um problema sério. Além disso, podem facilmente se sobreajustar, por isso, modelos que consideram um conjunto de árvores de decisão foi desenvolvido, sendo o RF o mais popular [94].

1.6.3.4 RF: é uma versão estendida da árvore de decisão que emprega vários classificadores em vez de um único classificador, melhorando assim o desempenho do resultado final [107]. RF é um método de comitê (*ensemble*) que opera treinando um número de árvores de decisão e retornando a classe com a maioria sobre todas as árvores do comitê [94]. A agregação das respostas de milhares de pessoas aleatórias a uma pergunta complexa pode gerar uma resposta mais precisa do que a resposta de um especialista. Isso é conhecido como sabedoria da multidão. Da mesma forma, a agregação das previsões de um grupo de preditores pode gerar previsões mais precisas do que as previsões de um preditor individual [94]. O algoritmo RF é escalável, robusto a ruído, não se ajusta demais, fácil de interpretar e visualizar sem muitos parâmetros para configurar. No entanto, à medida que o número de árvores aumenta, o algoritmo torna-se lento [94]. Outra vantagem do algoritmo RF é o uso de subamostragem com substituição para criar amostras de treinamento para o “modelo aprender”. Dessa forma, possibilita a comparação de importância atribuída a cada uma das características consideradas, por meio do erro *Out-of-Bag* (OOB) ao excluir uma determinada característica. O erro OOB é o erro médio de predição em cada amostra de treinamento, considerando apenas as árvores que não possuem uma determinada característica em sua amostra [104].

1.7 Aprendizado de máquina com rastreamento do olhar aplicado ao TEA

Embora muitas abordagens tenham sido propostas para a classificação do TEA via rastreamento do olhar, a maioria delas baseia-se na demarcação manual de ROIs, como Wan *et al.* [108], que utilizaram um vídeo de dez segundos em que aparecia uma jovem falando, onde as medidas avaliadas foram os tempos de fixação em seis ROIs. Após a aplicação de análise discriminante os tempos de fixação na boca e no corpo puderam discriminar significativamente DT e TEA, com acurácia de classificação de 0,85. Outras abordagens apresentaram resultados comparáveis ou até mesmo superiores, como Jiang *et al.* [109], que propuseram um classificador diagnóstico utilizando reconhecimento de emoções (raiva, nojo, medo, felicidade, tristeza e

surpresa), onde os indivíduos observaram uma transição dos rostos, em que as expressões faciais mudavam lentamente no vídeo. O método baseou-se principalmente no tempo de resposta a cada emoção e na quantidade de fixações na tela. O estudo apresentou uma boa precisão de classificação entre DT e TEA, com acurácia de 0,86.

A fim de verificar o panorama da literatura em relação ao uso de rastreamento ocular e estímulos visuais no TEA, realizamos uma revisão, conforme a tabela 2. A maioria dos estudos utilizaram estímulos dinâmicos (65%), com cenas sociais (77%) e tiveram finalidade de auxílio diagnóstico (69%). É importante destacar que 85% dos estudos apresentaram n amostral menor que 80. Ao encontro do tema, Mastergeorge *et al.* [40] publicaram em 2021 uma revisão sistemática abordando essa temática. Corroborando nossos achados, a maioria dos trabalhos utilizaram estímulos com cenas sociais (89%) e estímulos dinâmicos (69%) [40].

Tabela 2: Levantamento da literatura em relação ao uso de rastreamento ocular no Transtorno do Espectro Autista

Descrição	Grupos	n	Paradigma	Estímulo	Finalidade	Referência
Busca por marcadores prognósticos	TEA	49	GeoPref	Vídeo	Prognóstico	Bacon et al. 2020
Verifica efeitos de intervenção	TEA e Controle	51	GeoPref	Vídeo	Intervenção	Bradshaw et al. 2019
Verifica efeitos de intervenção	TEA	34	Tarefa	Jogo	Intervenção	Scherf et al. 2018
Compara preferência visual entre grupos	TEA e Controle	38	GeoPref	Imagem	Descritivo apenas	Vacas et al. 2021
Compara preferência visual entre subtipos de TEA	TEA e Controle	227	GeoPref	Vídeo	Descritivo / Diagnóstico	Moore et al. 2018
Busca predizer traços autísticos	TEA	22	Social	Imagem	Diagnóstico	Lin et al. 2021
Acompanha irmãos de TEA e diagnóstica posteriormente	TEA e Risco	47	Cena Invertida	Vídeo	Diagnóstico precoce	Falck-Ytter et al. 2018
Explora associações entre imagens sociais e não sociais	TEA e Controle	44	Cenas sociais e não sociais	Imagem	Descritivo apenas	Greene et al. 2019
Explora a preferência geométrica de maneira robusta	TEA, DT e outros	334	GeoPref	Vídeo	Diagnóstico	Pierce et al. 2016
Classifica e correlaciona com severidade	TEA e Controle	201	Múltiplos	Imagem e Vídeo	Diagnóstico	Frazier et al. 2018
Classifica e correlaciona com severidade	TEA e Controle	69	GeoPref	Imagem e Vídeo	Diagnóstico	Kou et al. 2019
Classifica e correlaciona com severidade	TEA e Controle	79	Múltiplos	Imagem e Vídeo	Diagnóstico	Frazier et al. 2016
Verifica fixações repetidas em áreas geométricas	TEA, DT e outros	66	GeoPref	Vídeo	Diagnóstico precoce	Vargas-Cuentas et al. 2016
Verifica padrão ao explorar páginas web, utiliza alto func.	TEA e Controle	68	Tarefas / Visualização livre	Página Web	Diagnóstico	Yaneva et al. 2020
Utiliza uma regressão logística para classificação	TEA e Controle	59	Cena Social com Áudio	Vídeo	Diagnóstico	Carette et al. 2018
Clusteriza com K-means para classificação	TEA e Controle	59	Cena Social com Áudio	Vídeo	Diagnóstico	Elbattah et al. 2019
Compara atenção entre ROIs sociais e não sociais	TEA e Controle	37	ROI Social vs Não Social	Vídeo	Descritivo apenas	Banire et al. 2020
Clusteriza com K-means para classificação	TEA e Controle	59	Cena Social com Áudio	Vídeo	Diagnóstico	Ali et al. 2021
Classifica utilizando fixações com diferentes modelos	TEA e Controle	28	Cenas aleatórias	Imagem	Diagnóstico	Anden & Linstead 2020
Classifica utilizando padrões de sacadas em CNN	TEA e Controle	59	Cena Social com Áudio	Vídeo	Diagnóstico	Elbattah et al. 2020
Prediz características autísticas no reconhec. de emoções	TEA, DT e outros	60	Reconhec. de emoções	Vídeo	Diagnóstico	Jiang et al. 2020
Modela a atenção utilizando imagens aleatórias	TEA e Controle	28	Cenas aleatórias	Imagem	Descritivo apenas	LeMeur et al. 2020
Modela a atenção e classifica utilizando múltiplos datasets	TEA e Controle	189	Múltiplos datasets	Imagem	Diagnóstico	Rahman et al. 2021
Extraí métricas simples de vídeos e classifica indivíduos	TEA e Controle	60	Cenas sociais e não sociais	Vídeo	Diagnóstico	Ahuja et al. 2020
Explora a compreensão semântica e sintática de cenas	TEA e Controle	65	Cenas não naturais	Imagem	Descritivo apenas	Almourad & Bataineh 2020
Classifica utilizando padrões de sacadas em CNN e SVM	TEA e Controle	74	Cenas sociais e não sociais	Imagem	Diagnóstico	Xia et al. 2020

Panorama da literatura em relação ao uso de rastreamento ocular no Transtorno do Espectro Autista para diagnóstico, prognóstico e outras finalidades. As bases de consulta utilizadas foram PubMed, IEEE e ACM com as palavras-chave, "Autism"/"ASD"AND "Eye tracking"/"Gaze tracking"AND "Prognosis"/"Prognostic"/"Severity", com preendendo o período de 2016 à 2021.

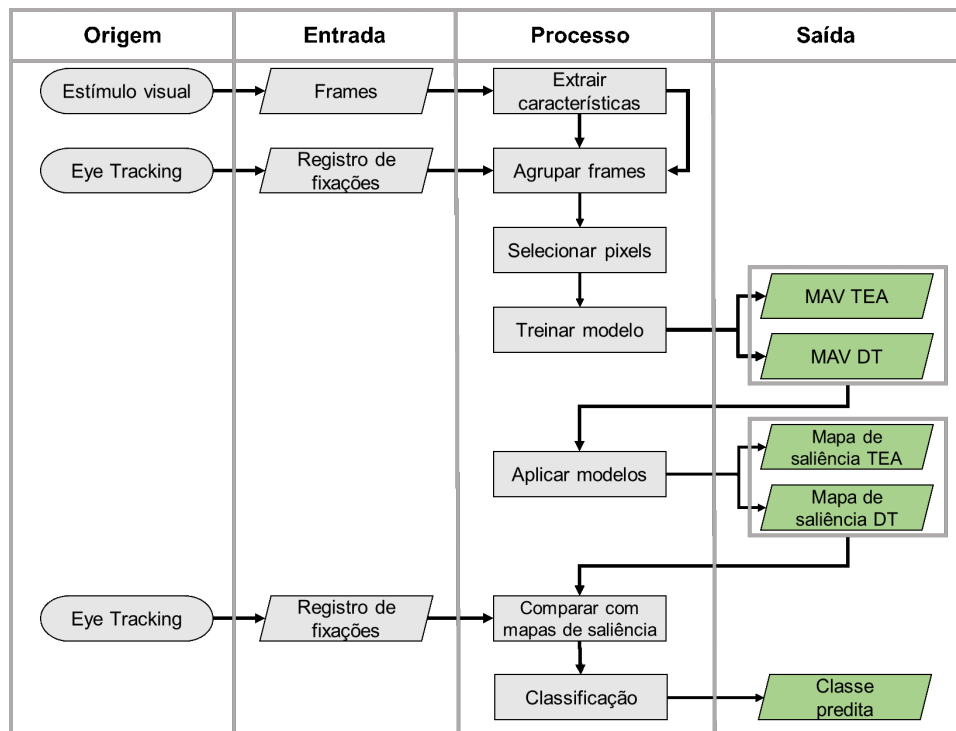
1.8 Estudos anteriores do grupo

Pesquisas preliminares do nosso grupo, Programa de Transtornos do Espectro Autista (PRO-TEA) e Laboratório de Aplicações de Informática em Saúde (LApIS), propuseram um classificador para casos DT e TEA utilizando o rastreamento do olhar, pautado em modelos de atenção visual gerados com base em mapas de fixação e vetores de características. O estímulo foi um vídeo com aproximadamente 54 segundos de duração, em que a tela foi dividida ao meio, sendo que em metade foram apresentados movimentos geométricos e metade cenas sociais. O estímulo foi baseado no paradigma GeoPref complexo, descrito por Moore *et al.* [54], neste caso, descrevendo cenas de várias crianças interagindo, fazendo ioga e pulando.

Com a finalidade de contextualizar o escopo do estudo, a abordagem metodológica é introduzida a seguir e complementada na seção 5.5.

Conforme descrito em Oliveira *et al.* [1] e representado na figura 5, inicialmente, os estímulos visuais em formato de vídeo foram submetidos a uma fase de pré-processamento, que incluiu a divisão em *frames*, redimensionamento e remoção de *frames* de transição entre cenas.

Figura 5: Representação esquemática das etapas de desenvolvimento do classificador.



Abreviações: MAV, Modelo de Atenção Visual; TEA, Transtorno do Espectro Autista; DT, Desenvolvimento Típico. Fonte: Adaptado de Oliveira [72].

A etapa subsequente envolveu a agregação de *frames* e dados brutos, utilizando a análise de fluxo óptico para agrupar *frames* consecutivos com base em valores médios de movimento.

Este processo visa superar a limitação de não haver pontos de fixação suficientes em cada *frame*. Mapas de fixação específicos para os grupos TEA e DT são então criados, destacando as áreas que foram mais fixadas pelos respectivos grupos (conforme ilustrado na figura 2, da seção 1.4). Esses mapas são suavizados com um filtro gaussiano para gerar imagens de saliência.

A fase de seleção de pixels no mapa de fixação visa criar amostras de treinamento equilibradas para os MAVs. Este processo envolve a escolha de 350 pixels com fixações (exemplos positivos) e 350 sem fixações (exemplos negativos), garantindo representatividade para ambas as classes. A extração de características dos pixels foi realizada com base em 28 características.

As 28 características utilizadas, listadas na tabela 3, são categorizadas pelo nível de processamento, descritas por baixo, médio e alto, que correspondem aos níveis de pixel, objeto e semântico, respectivamente. Na segunda coluna temos a quantidade de características consideradas inicialmente no modelo, onde vemos que as características a nível de pixel são sobrerrepresentadas em detrimento de características semânticas. Conforme a última coluna, as características consideradas derivaram principalmente dos estudos de Borji [110] e Judd *et al.* [111].

Tabela 3: Descrição das 28 características utilizadas no estudo prévio (Oliveira *et al.* [1]), e que são consideradas no presente estudo.

Nível	Quant.	Características	Descrição	Referência
Pixel	13	Pirâmides orientáveis	Complexidade / textura	[110, 111]
Pixel	1	Contraste de cor	Proporção entre canais	[110]
Pixel	1	Intensidade	Intensidade dos pixels	[110, 111]
Pixel	1	Orientação	Proporção entre canais	[110, 111]
Pixel	3	Coloração	Canais isolados (RGB)	[110, 111]
Objeto	1	Presença de movimento	Alteração entre <i>frames</i>	[111]
Objeto	1	Centro da tela	Distância até o centro da tela	[110, 111]
Objeto	1	Cena geométrica	Delimitação da cena geométrica	[72]
Objeto	1	Cena biológica	Delimitação da cena biológica	[72]
Objeto	1	Centro das cenas	Distância até o centro da cena	[72]
Objeto	1	Linha do horizonte	Faixa naturalmente direcionada	[110, 111]
Semântico	1	Face	Presença de face	[110, 111]
Semântico	1	Pessoa	Presença de pessoas	[110, 111]
Semântico	1	Pele	Presença de pele	[110]

As características estão organizadas conforme os níveis de pixel, objeto e semântico, que correspondem aos níveis de processamento baixo, médio e alto, respectivamente. Abreviações: Quant., Quantidade; RGB, vermelho/verde/azul.

Desta forma, foram consideradas apenas 3 características para o nível semântico e 19 representantes a nível de pixel, com 13 desses correspondendo a pirâmides orientáveis. Pirâmide orientável é uma técnica de processamento de imagens multiorientação e multiescala que foi desenvolvida na década de 1990 por Simoncelli *et al.* [112]. O fundamento das pirâmides

orientáveis é a capacidade de adaptar a orientação dos filtros usados na decomposição, permitindo analisar imagens em diferentes direções. Essa flexibilidade na escolha das orientações dos filtros torna as pirâmides orientáveis úteis em várias aplicações, como detecção de bordas, reconhecimento de padrões e processamento de texturas em imagens. Para os estudos foram utilizadas 13 pirâmides orientáveis, considerando quatro orientações e três escalas.

O aprendizado dos MAVs foi executado por meio de redes neurais com 10 neurônios, utilizando 15 características selecionadas pelo Algoritmo Relief ou Algoritmo Genético. A fase de diagnóstico ocorre em *frames* independentes, e mapas de saliência binários são gerados pelos MAVs para classificar indivíduos. A validação cruzada foi realizada em cinco *folds*.

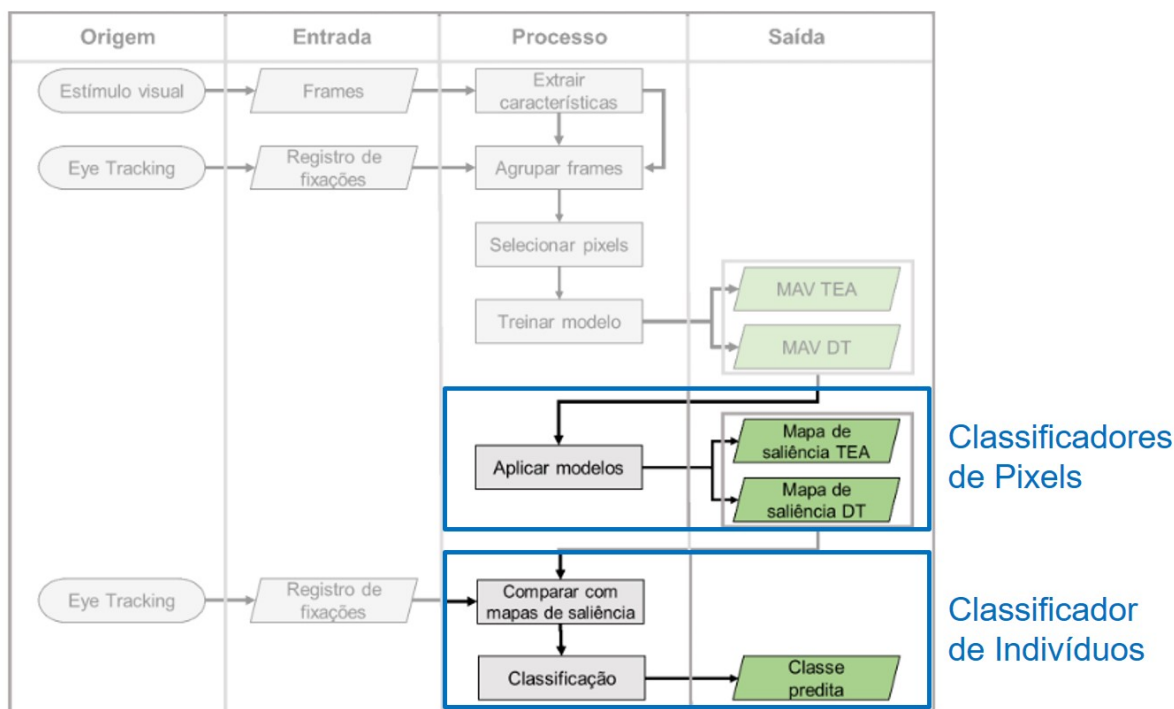
Assim, ao aplicar os MAVs aos *frames* do vídeo, mapas de saliência foram obtidos, os quais foram posteriormente comparados aos mapas de fixação de cada indivíduo para verificação de similaridade e, conseqüentemente, a classificação do indivíduo.

Portanto, na abordagem existe um classificador para cada grupo (DT e TEA) para pixels, que classifica cada pixel entre fixado e não fixado e o classificador geral para indivíduos, que classifica os indivíduos entre DT e TEA, de acordo com a similaridade dos mapas de fixações de cada indivíduos e os respectivos mapas de saliência. A fim de facilitar a compreensão, adotemos os termos classificadores de pixels e classificador de indivíduos, conforme representado na figura 6.

Dois estudos derivaram dessa abordagem: o primeiro contou com 30 indivíduos DT e 78 indivíduos com TEA e apresentou acurácia de 0,80 e AUC de 0,82 na classificação da presença do TEA [72]; o segundo contou com 49 indivíduos com TEA não grave (leve ou moderado) e 39 grave e obteve acurácia de 0,88 e AUC de 0,85 na classificação entre indivíduos com TEA não grave e grave [37].

Embora os trabalhos citados tenham apresentado resultados promissores, existem algumas limitações, destacadamente o desbalanceamento entre as classes no primeiro estudo (30 DT para 78 TEA), o que pode distorcer ou enviesar a classificação de novos indivíduos e o desbalanceamento das características, havendo uma sobrerrepresentação de características a nível de pixel em detrimento do nível semântico.

Figura 6: Identificação dos classificadores de pixel e de indivíduos.



Abreviações: MAV, Modelo de Atenção Visual; TEA, Transtorno do Espectro Autista; DT, Desenvolvimento Típico.

2 JUSTIFICATIVA E PROPOSTA

Estima-se que 1% da população mundial se enquadre no espectro autista [11, 14]. O TEA contempla uma ampla heterogeneidade fenotípica de apresentação e tal amplitude dificulta o diagnóstico e a avaliação de prognóstico do transtorno [38], necessitando de uma equipe multidisciplinar altamente treinada. A análise do rastreamento do olhar tem se mostrado um método promissor [22, 23]. MAVs considerando características semânticas têm contribuído para a compreensão da alteração do processamento visual no TEA e a classificação de casos e controles. No entanto, até o momento não existem estudos propondo a integração de diferentes tipos de características (pixel, objeto, semântica) e avaliando a contribuição relativa destas para classificação diagnóstica do TEA e compreensão da heterogeneidade fenotípica.

Com base no exposto, a proposta de ineditismo deste projeto é integrar características semânticas e verificar sua importância relativa em MAVs para a classificação do TEA, visando tanto a otimização dos classificadores, quanto sua interpretabilidade por meio da avaliação da importância relativa das características.

3 HIPÓTESE

A partir do cenário apresentado, a hipótese a ser verificada na presente tese é: “A inclusão de características semânticas em Modelos de Atenção Visual melhoram o desempenho de classificação entre Desenvolvimento Típico e Transtorno do Espectro Autista”.

4 OBJETIVOS

4.1 Geral

Desenvolver classificadores para o auxílio diagnóstico do TEA, baseado em MAVs considerando diferentes níveis de características em mapas de saliência, com enfoque na integração de características semânticas.

4.2 Específicos

A - Realizar uma revisão de literatura para mapear a lacuna acerca do uso de características semânticas e mapas de saliência em MAVs para a classificação do TEA com rastreamento do olhar;

B - Replicar e validar a abordagem do estudo prévio do grupo;

C - Compor um levantamento de características semânticas passíveis de serem extraídas de forma automatizada do estímulo visual;

D - Implementar novas características semânticas nos MAVs para DT e TEA;

E - Avaliar novos algoritmos indutores de classificação à nova abordagem;

F - Avaliar a importância das diferentes categorias de características para cada um dos grupos (DT e TEA).

5 MATERIAIS E MÉTODOS

O presente projeto de pesquisa foi conduzido de acordo com as seguintes etapas:

- Ética, Armazenamento de Dados e Amostra: nestas seções são descritos os princípios éticos, as práticas de armazenamento de dados, os critérios para seleção da amostra e a divisão dos *datasets* utilizados.

- Dados de Rastreamento do Olhar e Paradigma Visual: metodologia para coleta e análise dos dados de rastreamento do olhar. Além disso, apresenta o paradigma visual e as diferenças entre os estímulos visuais empregados no estudo anterior e no atual.
- Construção do Classificador e Validação do Estudo Anterior: abordagem geral para construção dos classificadores, com ênfase nas etapas de construção de mapas de saliência e classificação dos indivíduos. Adicionalmente, a validade dos resultados anteriores foi analisada por meio de dois experimentos distintos.
- Alterações Metodológicas: a fim de alcançar os objetivos propostos alterações metodológicas foram propostas. Nesta seção, as alterações realizadas são apresentadas de maneira ilustrativa.
- Substituição do Algoritmo Indutor de Classificação: detalha a substituição do algoritmo indutor de classificação, adotando-se outra implementação do algoritmo SVM e o algoritmo RF.
- Adição de Novas Características e Normalização: aborda a inclusão de características semânticas adicionais, desde o levantamento de características semânticas até a extração e incorporação destas. Além disso, foi incluída a etapa de normalização das características, permitindo a comparabilidade das características.
- Análise Estatística: abordagem estatística utilizada para analisar os resultados, incluindo os testes estatísticos e critérios de significância adotados.

5.1 Ética e armazenamento de dados

A presente proposta está inserida no projeto “Classificadores para diagnóstico precoce do Transtorno do Espectro Autista usando o rastreamento do olhar” financiado pelo Programa Nacional de Apoio à Atenção da Saúde da Pessoa com Deficiência (PRONAS/PCD), sob registro 00.530.493/0001-71 no Ministério da Saúde. O projeto tem parecer de ética em pesquisa aprovado (Anexo A), sob número 203 40819.7.0000.0068, pela Faculdade de Medicina da Universidade de São Paulo (USP) e cadastrado na Plataforma Brasil.

Os dados obtidos dos participantes seguiram as normas da Lei Geral de Proteção de Dados (LGPD), lei nº 13.709, de 14 de agosto de 2018, alterada pela lei nº 13.853, de 2019. Os dados foram anonimizados em publicações e utilizados apenas para fins de pesquisa. O geren-

ciamento desses dados foi realizado por meio da criação de um banco de dados relacional, na plataforma RedCap [113].

5.2 Amostra

O estudo contou com indivíduos de ambos os sexos, com idade entre 1 e 17 anos. Foram coletadas amostras de 50 indivíduos que fecharam critérios para TEA, avaliados no ambulatório do PROTEA, do Instituto de Psiquiatria da Faculdade de Medicina da USP. O grupo DT (controle) foi composto por 49 indivíduos recrutados em escolas públicas e privadas do estado de São Paulo, pareados por sexo e idade.

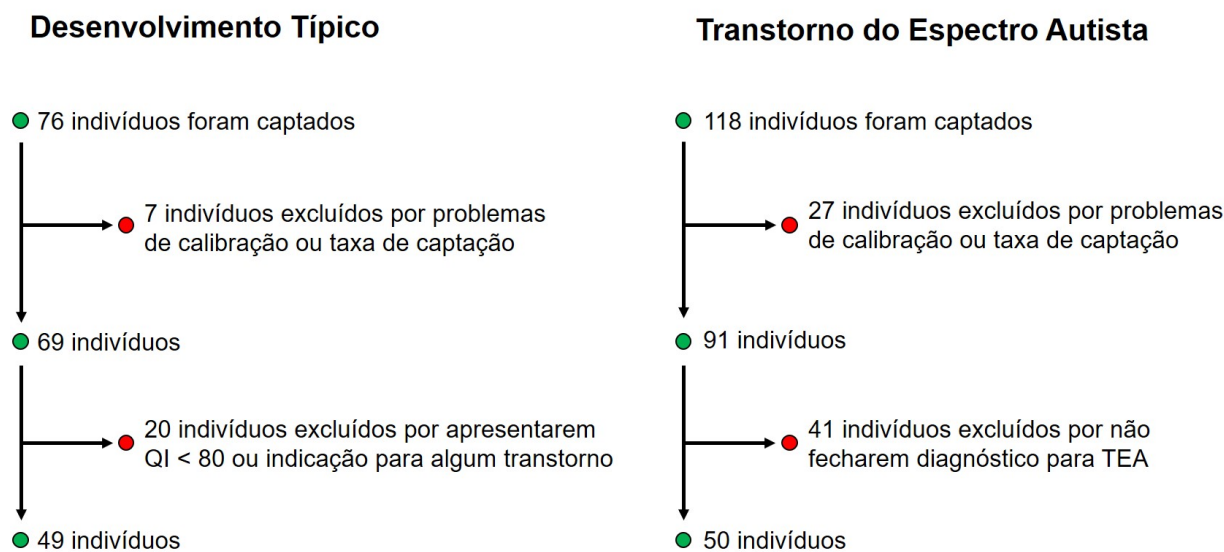
Para todos os participantes foram realizados os seguintes procedimentos: aplicação do instrumento ADOS-2 [16], histórico familiar, avaliação clínica por psiquiatra infantil e diagnóstico de comorbidades psiquiátricas com abordagem multidimensional através da pontuação da escala do CBCL [21]. Os critérios diagnósticos seguem o DSM-5 [114] e a CID-11 [15]. Os escores brutos foram transformados em t-escores.

Os casos diagnosticados como TEA foram avaliados pela escala CARS, traduzida e adaptada para a população brasileira [19], para verificar a gravidade do TEA, e também o teste de QI. O QI foi avaliado pela Escala de Inteligência Wechsler para Crianças/*Wechsler Intelligence Scale for Children (WISC) IV* [115] ou pelo Teste de Inteligência Não Verbal de Snijders-Oomen/*Snijders-Oomen Nonverbal Intelligence Test - Revised SON-R 2¹/₂-7[a]*, padronizado e validado para a população brasileira [116], de acordo com os critérios de idade recomendados. Indivíduos com QI < 80 foram considerados com baixo QI [117, 118] ou Deficiência Intelectual [2] (DI). A avaliação cognitiva funcional foi realizada por um neuropsicólogo treinado, por meio da Escala de Comportamento Adaptativo Vineland [20], considerando seus quatro domínios: *Vineland Communication (VC)*; *Vineland Daily Living (VDL)*; *Vineland Socialization (VS)*; *Vineland Locomotion (VL)*.

Na figura 7 é apresentado o fluxograma sobre os indivíduos incluídos e excluídos da pesquisa para cada grupo.

A fim de isolar possíveis influências e investigar possíveis efeitos de subgrupos, o *dataset* foi dividido para verificar: se a idade impacta no desempenho de classificação; se indivíduos com TEA leve (CARS < 36) dificultam a classificação; se as condições de DI ou baixo QI impactam no desempenho de classificação, uma vez que podem gerar confusão. Nesses cenários mais específicos, podemos verificar também a contribuição das características, pois os subgrupos

Figura 7: Fluxograma sobre os indivíduos incluídos e excluídos da pesquisa.



Abreviações: QI, Quociente de Inteligência; TEA, Transtorno do Espectro Autista.

podem interpretar/explorar de maneira diferente as características. Portanto, estratificamos nossa amostra conforme representado na figura 8.

No *dataset* I, temos todos os indivíduos: 49 DT e 50 TEA.

No *dataset* II, temos todos os indivíduos maiores de 4 anos: 33 DT e 28 TEA.

No *dataset* III, temos a exclusão dos indivíduos com TEA leve: 49 DT e 31 TEA.

No *dataset* IV, temos a exclusão de indivíduos com DI: 49 DT e 38 TEA.

Figura 8: Descrição dos *datasets*.

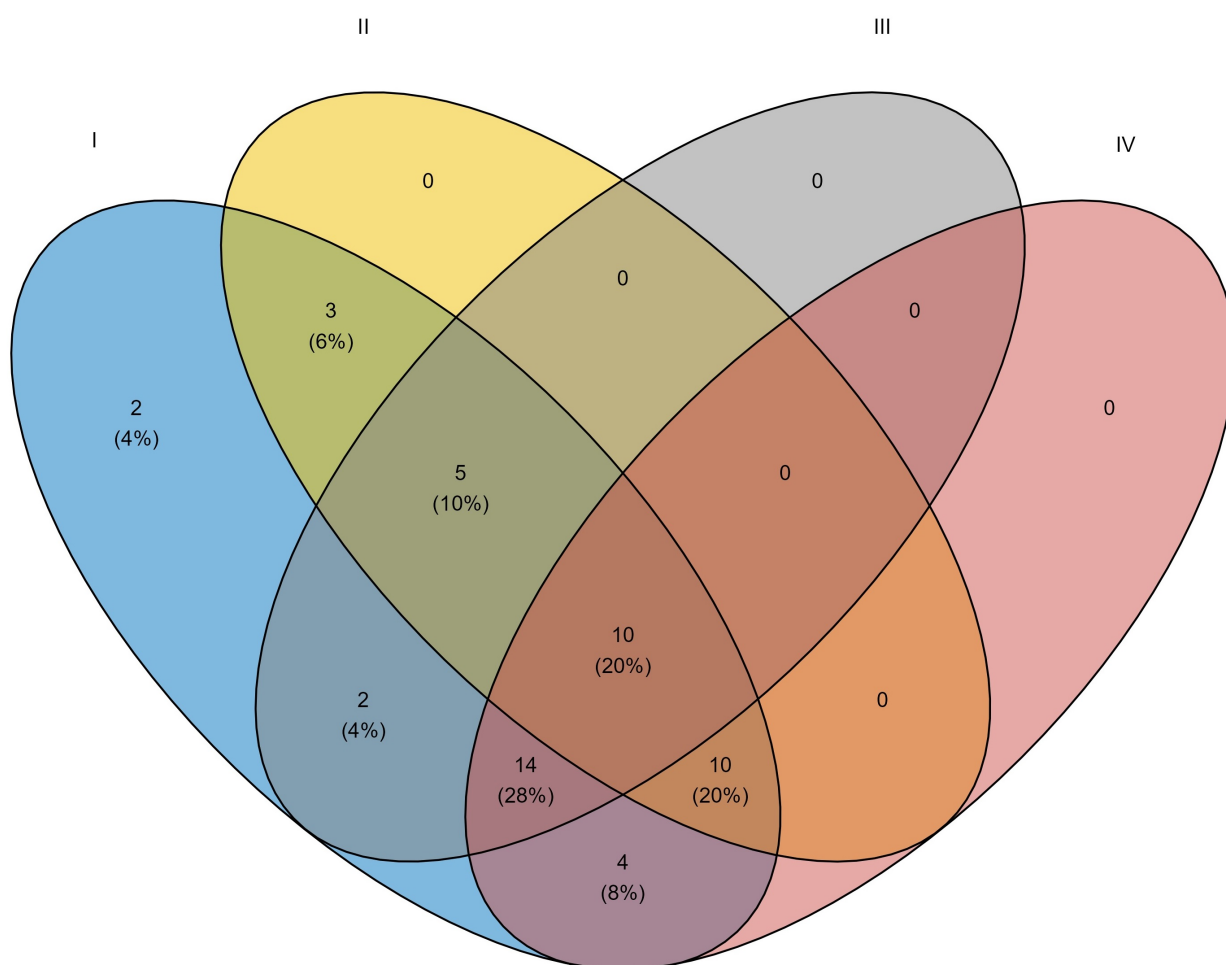
	DT	TEA	
I	49	50	Todos os indivíduos
II	33	28	Maiores de 4 anos
III	49	31	Exceto TEA leve
IV	49	38	Exceto DI

Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista; DI, Baixo QI ou Deficiência Intelectual.

Na figura 9 são representadas as intersecções que existem entre os *datasets* para o grupo TEA. O *dataset* I é composto por todos os indivíduos considerados na pesquisa (50 indivíduos). Apenas 2 indivíduos estavam presentes exclusivamente no *dataset* I. Portanto os demais 48 indivíduos foram compartilhados com outros *datasets*, desses: 3 indivíduos compuseram apenas

o *dataset* II (maiores de 4 anos); 2 indivíduos compuseram apenas o *dataset* III (sem TEA leve); 4 indivíduos compuseram apenas o *dataset* IV (sem QI baixo); 5 indivíduos estiveram presentes nos *datasets* II e III; 10 indivíduos estiveram presentes nos *datasets* II e IV; 14 indivíduos estiveram presentes nos *datasets* III e IV; 10 indivíduos estiveram presentes em todos os *datasets*. Para o grupo DT, os indivíduos considerados nos *datasets* I, III e IV foram os mesmos (49 indivíduos). Apenas o *dataset* II apresentou alteração, com a exclusão de 16 indivíduos com menos de 4 anos de idade.

Figura 9: Intersecção entre os *datasets* para o grupo TEA



“I” refere-se ao *dataset* com todos os indivíduos (50 TEA). “II” refere-se ao *dataset* com os indivíduos maiores de 4 anos (28 TEA). “III” refere-se ao *dataset* com a exclusão dos indivíduos com TEA leve (31 TEA). “IV” refere-se ao *dataset* com a exclusão dos indivíduos com baixo QI (38 TEA).

5.3 Dados de rastreamento do olhar

Nos estudos anteriores do grupo [1, 37], o sistema de captura utilizado foi o equipamento TX300 [119] e o *software* Tobii Studio [120]. No projeto atual, o equipamento Tobii Pro Fu-

sion [121], operando a 250 Hertz e o software de captura *Tobii ProLab* [122], com a utilização do filtro de fixação *I-VT* [123] foram utilizados. Os movimentos oculares foram classificados como fixação utilizando os parâmetros padrão do filtro (duração maior que 60 milissegundos e velocidade menor que 30°/segundo [124]).

Devido à alteração do sistema foi necessária a inclusão de um pré-processamento, para que o *output* do software fosse compatível com a abordagem desenvolvida anteriormente. Esta etapa foi conduzida na linguagem R [125], visando a sincronização da captura de dados com a execução do vídeo, bem como a alteração de nomes para corresponder aos utilizados anteriormente.

Para a captura, o observador permaneceu a uma distância de 50 a 80 cm do equipamento, possibilitando a captura de movimentos visuais com acurácia adequada para o processamento computacional, mesmo com movimentações da cabeça com velocidade de até 50 cm/segundo e angulação do olhar de até 35 graus.

Os critérios de inclusão para os dados de rastreamento do olhar foram: calibração com cinco pontos aceita pelo software *Tobii ProLab* e taxa de captação maior que 80%. Este último critério ajuda a garantir a qualidade dos dados, principalmente porque crianças tendem a desviar o olhar quando não estão mais interessadas [126].

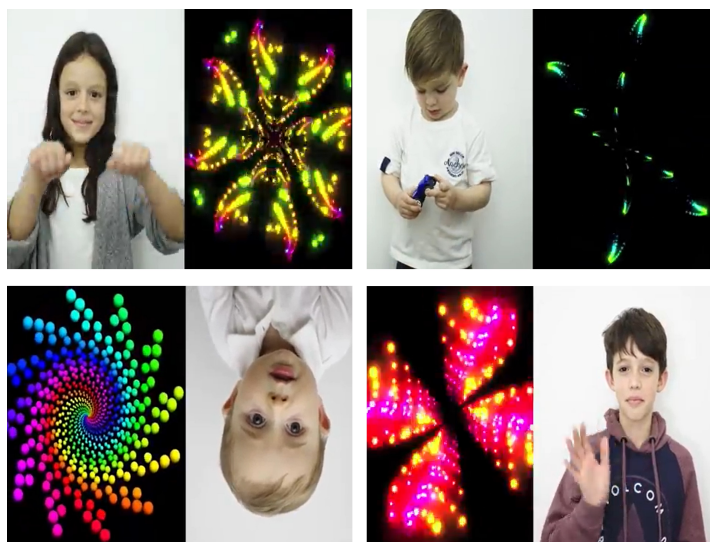
5.4 Paradigma e estímulo visual

O uso do paradigma de preferência visual permite explorar diferenças na atenção visual entre DT e TEA, uma vez que no TEA há: maior preferência por cenas geométricas [54], menor saliência para características semânticas [85] e maior dificuldade em desengajar o olhar (*i.e.*, desviar o olhar de algo que está no foco atual para focar em algo novo) [127].

O paradigma de preferência visual dinâmica foi estabelecido a partir de um estudo anterior do grupo, onde foi utilizado outro vídeo para a classificação de indivíduos com DT e TEA [72]. Tanto no vídeo anterior, quanto no atual, são apresentadas duas cenas simultâneas, de um lado movimentos biológicos, onde há a presença de crianças, e do outro lado, movimentos geométricos, com presença de fractais, conforme mostrado na figura 10. O vídeo atual tem 41 segundos de duração, sem acompanhamento de áudio, com apresentação de 30 *frames* por segundo.

Uma comparação entre as características extraídas dos vídeos é apresentada na tabela 4. Algo a ser considerado é a diferença de contraste entre os vídeos, representado pela caracte-

Figura 10: Exemplos de *frames* do estímulo utilizado.



rística Intensidade descritas por Itti [128], conforme a tabela 4. No vídeo atual temos valores menores de contraste e a percepção dele ocorre de maneira diferente entre DT e TEA, podendo ainda variar de acordo com a idade [129].

Tabela 4: Comparação de valores médios para as características extraídas dos vídeos.

Descrição	Vídeo Anterior	Vídeo Atual
Pirâmide 1	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 2	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 3	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 4	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 5	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 6	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 7	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 8	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 9	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 10	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 11	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 12	1,00 ± 0,00	1,00 ± 0,00
Pirâmide 13	1,00 ± 0,00	1,00 ± 0,00
Cor Itti	0,40 ± 0,08	0,16 ± 0,05
Intensidade	0,77 ± 0,17	0,20 ± 0,05
Orientação	0,44 ± 0,07	0,06 ± 0,01
Pele	0,22 ± 0,05	0,17 ± 0,05
Cor 1	0,54 ± 0,14	0,47 ± 0,09
Cor 2	0,52 ± 0,14	0,46 ± 0,11
Cor 3	0,51 ± 0,16	0,44 ± 0,10
Linha do Horizonte	0,50 ± 0,01	0,51 ± 0,00
Face	0,17 ± 0,09	0,07 ± 0,07
Pessoa	0,01 ± 0,01	0,07 ± 0,07
Centro da Tela	0,52 ± 0,00	0,54 ± 0,00
Movimento	0,05 ± 0,00	0,02 ± 0,00
Cena Biológica	0,48 ± 0,11	0,44 ± 0,16
Cena Geométrica	0,48 ± 0,11	0,44 ± 0,16
Centro das Cenas	0,54 ± 0,00	0,54 ± 0,00

Os valores são expressos como média ± desvio padrão. As primeiras 13 características correspondem às Pirâmides Orientáveis e possuem valor médio de 1 porque os valores foram normalizados, com variação apenas entre os pixels.

5.5 Construção do classificador

Nesta seção são apresentados os métodos base propostos por Oliveira *et al.* [1]. As alterações propostas no presente estudo são apresentadas na seção 5.7.

O classificador foi desenvolvido no Ambiente de Desenvolvimento Integrado/*Integrated Development Environment (IDE) MatLab R2015a* disponibilizado sob licença estudantil (532328) pela USP.

A figura 5 (apresentada na seção 1.8) sintetiza os passos para a construção do classificador, que podem ser divididos em duas etapas: construção dos mapas de saliência e classificação com base nos mapas de saliência. Essas fases são detalhadas nas seções a seguir.

5.5.1 Construção de mapas de saliência

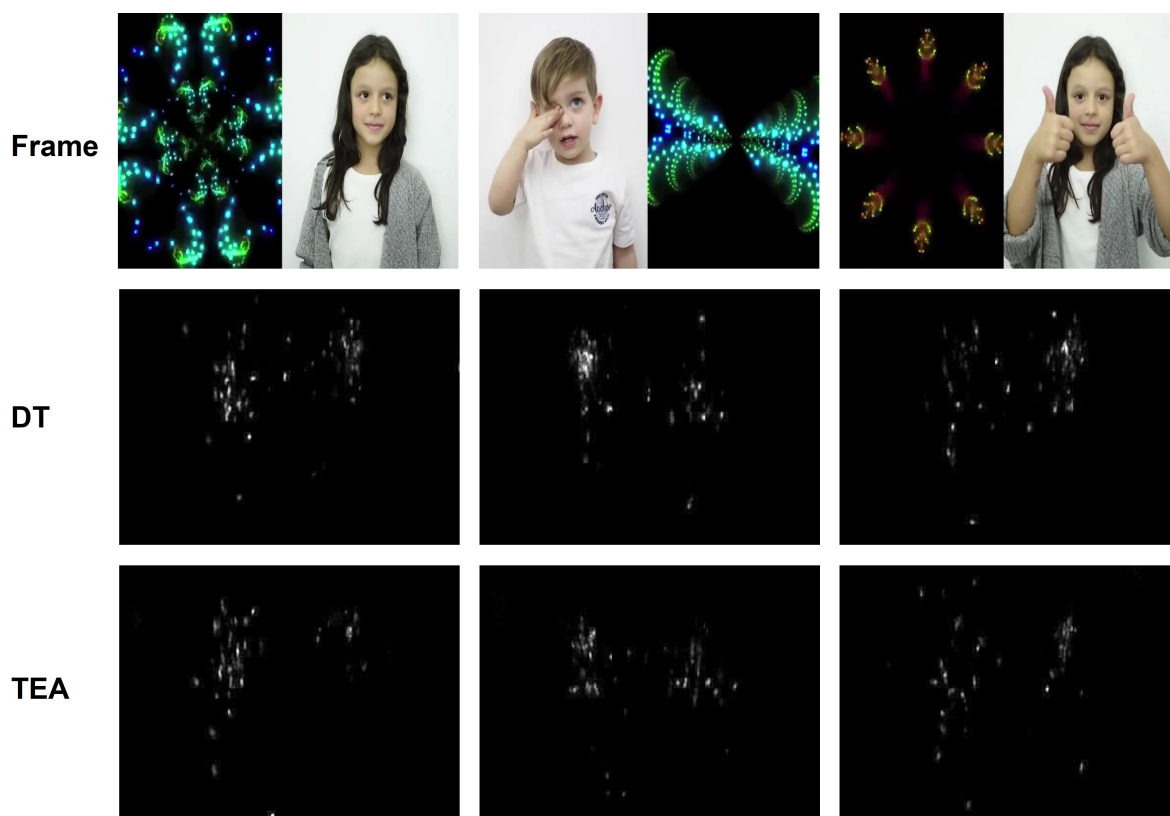
Inicialmente o estímulo é fragmentado em *frames* e de cada *frame* são extraídas as características, formando vetores de características para cada pixel que compõe a imagem (conforme representado na figura 2, apresentada na seção 1.8). Por se tratar de vídeos, cada *frame* não possui pontos de fixação suficientes para extrair informações. Para contornar isso, a informação de *frames* consecutivos é agregada com base no cálculo da densidade de fluxo óptico pelo algoritmo de Farneback [130], ou seja, a taxa de movimento entre eles. O fluxo óptico é calculado comparando um *frame* com o próximo e retornando um valor de movimento para cada pixel. Este valor leva em consideração principalmente a diferença de intensidade de um pixel no *frame* atual em relação ao pixel correspondente no próximo *frame*.

Ao exibir o estímulo para o indivíduo, os dados de rastreamento do olhar são coletados (conforme descrito na seção 5.3) e registrados os pontos da imagem em que o olhar do indivíduo foi fixado, para posterior sincronização com os *frames* agrupados. Na sequência, alguns pixels são selecionados para treinar o modelo, onde são selecionados os pixels com maior fixação (exemplos positivos) e os pixels que não foram fixados (exemplos negativos). Os MAVs são treinados paralelamente, sendo um para a classe DT e outro para a classe TEA. Os modelos agora treinados são aplicados no conjunto de *frames* do vídeo e são gerados os mapas de saliência para ambas as classes.

5.5.2 Classificação dos indivíduos

Com base nos dados de fixações dos indivíduos são sintetizados os mapas de fixações correspondentes aos *frames*. A figura 11 apresenta alguns *frames* de exemplo e seus respectivos mapas de fixação para ambos os grupos (DT e TEA). Cada mapa de fixações foi comparado aos mapas de saliência correspondentes (conforme representado na figura 5) e o *frame* foi classificado com o rótulo do mapa de saliência que apresentou maior similaridade, ou seja, acertos. Ao fim, teremos os *frames* já rotulados e os critérios para classificação do indivíduo foram testados, inicialmente a classe foi atribuída por um sistema de votação por maioria. O critério de decisão foi definido com base nos valores da curva ROC, utilizando o índice de Youden [131] (conforme representado e explicado na figura 4).

Figura 11: Exemplos de mapas de fixação.



Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

5.6 Validação do estudo anterior

A validação foi realizada em duas etapas, conforme descrito nos experimentos a seguir.

5.6.1 Experimento 1

Para verificar se o classificador é estímulo-específico, ou seja, seu poder de generalização, foi utilizado o modelo previamente treinado, conforme descrito por Oliveira [72] (descrito na seção 5.5) para classificar a nova amostra (*dataset I*) captada com um novo vídeo, conforme descrito na seção 5.4. Portanto, os mapas de saliência deste novo vídeo foram construídos com as mesmas características usadas no estudo anterior.

5.6.2 Experimento 2

Uma preocupação que tem ganhado destaque na comunidade científica é a replicabilidade dos achados [132] e a validade externa. Como existem diferentes definições de replicabilidade em diferentes áreas científicas, utilizaremos a definição de Patil *et al.* [133]: replicabilidade é “reexecutar o experimento e coletar novos dados”. Em suma, a replicabilidade envolve a co-

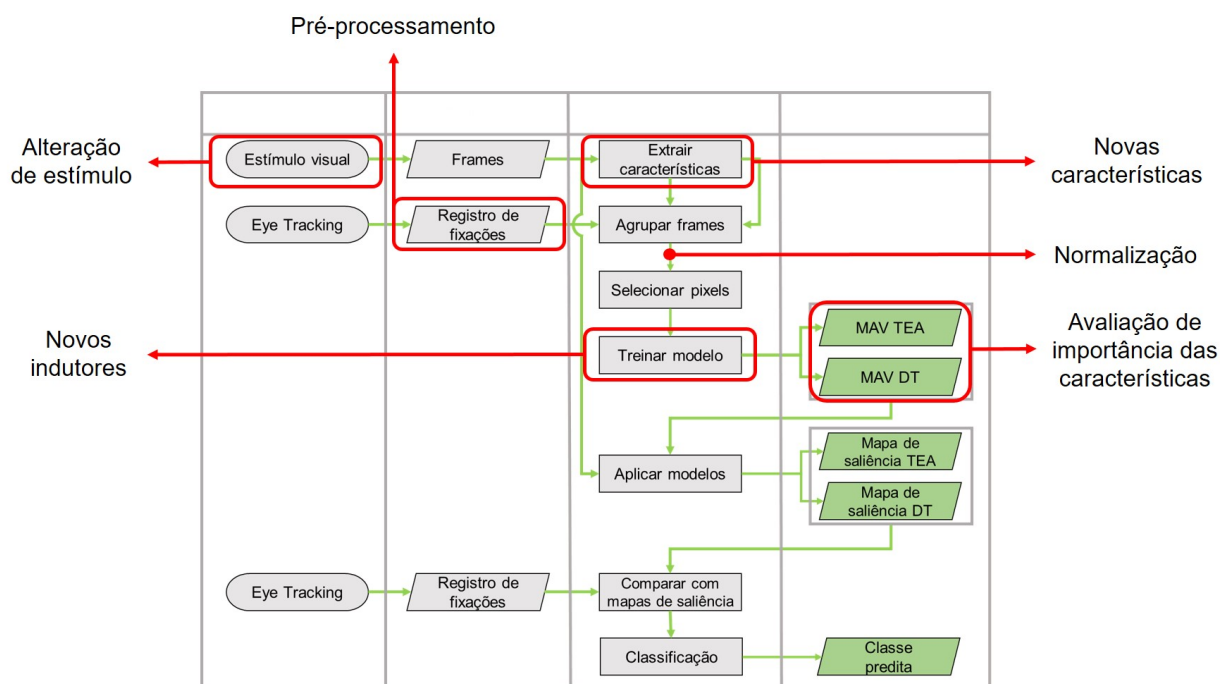
leta de novos dados e o uso de métodos iguais aos aplicados em estudos anteriores. Portanto, para verificar a robustez do modelo (*i.e.*, a capacidade do sistema de suprimir fontes de variação [134]) e a validade externa do método, a abordagem foi treinar o modelo com um vídeo semelhante ao usado anteriormente e testá-lo com as novas amostras. Portanto, as etapas de treinamento e teste ocorreram com o novo vídeo. Neste caso, os hiperparâmetros usados foram os mesmos do estudo anterior [1], tanto para os processos de extração e seleção de características, quanto para os algoritmos de indução de classificação.

Os algoritmos de indução classificatória utilizados foram NN e SVM com três conjuntos de características: 1 - Fixo, conjunto de 15 características em cada grupo, o mesmo descrito por Oliveira [72]; 2 - *Relief*, onde as características com relevância maior que a média geral no seletor *ReliefF* foram consideradas; 3 - Algoritmo Genético, onde o número de características consideradas foi 15. Em todas as situações, o desempenho de classificação foi estimado por meio de validação cruzada com $k - fold = 5$.

5.7 Alterações Metodológicas

Para alcançar os objetivos do estudo, a proposta apresentada contemplou seis alterações metodológicas na abordagem inicial do grupo, conforme representado na figura 12. Além disso, contribuímos com a criação de protocolos para a captação de dados de rastreamento do olhar e armazenamento de todos os dados coletados no ambulatório em uma base comum, o RedCap. A unificação dos registros no RedCap nos permitiu criar *pipelines* que facilitaram a visualização e análise de dados do grupo, contribuindo indiretamente com outros projetos.

Figura 12: Representação das alterações metodológicas realizadas.



As alterações realizadas estão destacadas na cor vermelha. Abreviações: MAV, Modelo de Atenção Visual; Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

5.8 Substituição do algoritmo indutor de classificação

Após a validação, seguimos com a substituição do algoritmo indutor de classificação.

5.8.1 SVM

Embora o algoritmo SVM já tenha sido utilizado anteriormente, a implementação utilizada permitia apenas o uso do *kernel* linear, através da biblioteca *Liblinear* [135]. Para a devida parametrização, a biblioteca foi substituída pela função *“fitcsvm”*. A nova implementação é oficial da empresa “MathWorks”, que mantém o MatLab, para o algoritmo indutor de classificação SVM. Os *kernels* suportados são *“linear”*, *“gaussian”*, *“rbf”* e *“polynomial”*. A função conta com uma parametrização automática ao definir o argumento *“OptimizeHyperparameters”* como *“auto”*.

5.8.2 Random Forest

O MatLab conta com uma implementação oficial para o algoritmo RF, dada pela função *“TreeBagger”* [136]. O algoritmo RF foi testado com todas as características. Não foram utilizados algoritmos seletores de características, pois já existe uma seleção interna de caracte-

rísticas no algoritmo RF. Para estabelecer o *mtry* (o número de preditores selecionados aleatoriamente em cada nó [137]), foram testados previamente valores de 3 a 9, com passo 1. Após os testes adotamos o *mtry* = 9. Para o número de árvores foi utilizado o valor 50, a partir de experimentos prévios que consideraram os valores 25, 50, 100 e 200. O parâmetro “*OOBPredictorImportance*” foi configurado na opção “*on*” para que o cálculo da importância relativa de cada característica fosse computado. Os demais parâmetros para a função “*TreeBagger*” [136] foram mantidos como padrão, ou seja, os valores pré-configurados da implementação.

A avaliação da importância relativa de cada característica refere-se a contribuição que cada uma delas tem para prever se um pixel foi ou não fixado para cada um dos grupos DT e TEA. A importância é estimada com base na ausência de determinado preditor (OOB) para o treinamento do modelo [136].

5.9 Adição de novas características

Um dos principais diferenciais deste projeto está nas características que foram consideradas, o que pode auxiliar na compreensão do TEA de acordo com mais ou menos características teoricamente associadas a conectividade cerebral do tipo *bottom-up* e *top-down*.

Conforme mencionado na introdução, apesar das limitações, estudos anteriores do grupo obtiveram bons resultados. Neste estudo, utilizamos as 28 características apresentadas na tabela 3 e expandimos a quantidade de características consideradas, incluindo outras de nível semântico.

5.9.1 Levantamento de características semânticas

Geralmente as regiões da imagem que contém características semânticas são demarcadas manualmente, por conta disso, foi conduzida uma revisão da literatura a fim de identificar métodos de extração automática de características semânticas que estiverem presentes no estímulo atual. Para o levantamento foram consideradas as bases de dados *MedLine*, *Association for Computing Machinery Database (ACM)*, *Institute of Electrical and Electronics Engineers Database (IEEE)* e *Google Scholar*, com a inclusão das seguintes palavras-chaves combinadas: “*semantic*”, “*image*”, “*feature*”, “*emotion*”, “*gesture*”, “*hand-gesture*”, “*computer vision*” e “*feature extraction*”. O período considerado engloba os anos de 2015 a 2022. A seleção dos artigos relevantes foi manual, com o propósito de encontrar possíveis soluções viáveis de serem implementadas

no presente projeto.

O levantamento de novas características semânticas passíveis de serem extraídas do estímulo são apresentadas na tabela 5 e detalhadas no Apêndice A.

Tabela 5: Levantamento de características semânticas passíveis de serem extraídas no estímulo utilizado.

Características presentes no vídeo	Referência
Reconhecimento de expressões faciais	Nwosu <i>et al.</i> , 2017 [138] Khairuddin & Chen, 2021 [139] Skiendziel <i>et al.</i> , 2019 [140] Yang <i>et al.</i> , 2018 [141] Baltrusaitis <i>et al.</i> , 2018 [142] Wang <i>et al.</i> , 2022 [143] Olivares-Mercado <i>et al.</i> , 2019 [144]
Reconhecimento de elementos faciais	Reimann <i>et al.</i> , 2021 [145] Nwosu <i>et al.</i> , 2017 [138] Baltrusaitis <i>et al.</i> , 2018 [142] Lugaresi <i>et al.</i> , 2018 [146]
Reconhecimento de gestos manuais	Deng <i>et al.</i> , 2020 [147] Wang <i>et al.</i> , 2020 [148] Köpüklü <i>et al.</i> , 2019a [149] Köpüklü <i>et al.</i> , 2019b [150] Dadashzadeh <i>et al.</i> , 2019 [151] Lugaresi <i>et al.</i> , 2018 [146]
Direcionamento do olhar do ator	Koutras & Maragos, 2015 [152] Mukherjee & Robertson, 2015 [153] Chong <i>et al.</i> , 2018 [154] Kellnhofer <i>et al.</i> , 2019 [155] Baltrusaitis <i>et al.</i> , 2018 [142]
Mapa afetivo (Emoção em regiões da imagem)	Rao <i>et al.</i> , 2018 [156] Chong <i>et al.</i> , 2018 [154] Zhao <i>et al.</i> , 2022 [157]

A partir do levantamento foram consideradas as características semânticas: elementos faciais, considerando como características separadas o T-Facial (região que engloba olhos + nariz + boca), região dos olhos, boca e nariz; expressões faciais; gestos manuais; direção do olhar; mapa afetivo. Dentre as características apresentadas, apenas mapa afetivo não foi extraída por limitações técnicas.

5.9.1.1 Elementos faciais: o rosto é uma das principais áreas de interesse na atenção visual, sendo naturalmente focalizado durante a interação social. O olhar humano é atraído para essas regiões, e a detecção da face auxilia na identificação de pessoas e compreensão de emoções. Portanto, essas características são importantes candidatas para a construção de mapas de saliência. Os elementos faciais compreendem a face como um todo e regiões específicas como o T-Facial, os olhos, a boca e o nariz. Aqui denominamos como T-Facial, a região central da face, compreendendo os pontos mais importantes da face estão localizados na região dos olhos, nariz e boca [158].

Entre outras pistas sociais no ambiente, as faces são provavelmente as mais importantes para humanos, uma vez que transmitem informações sociais relevantes, tais como identidade, idade, gênero, emoções [159], atuando como catalisadores do desenvolvimento cognitivo, social e emocional [160]. Recém-nascidos apresentam preferência por olhar para estímulos semelhantes a faces logo após o nascimento [161, 162].

A região central da face, compreendendo os pontos mais importantes da face estão localizados na região dos olhos, nariz e boca [158], foi denominado no presente estudo como T-Facial.

A preferência dos bebês pela região dos olhos ocorre provavelmente pelo fato de que o olhar tem importantes funções comunicativas nos seres humanos [163]. A morfologia única do olho humano, com maior exposição de esclera do que outros primatas e a esclera completamente branca, pode explicar parcialmente essa preferência [163].

Nos primeiros meses de vida, os bebês preferem olhar para os olhos e depois mudam gradualmente a sua atenção preferencial dos olhos para a boca durante o primeiro ano [164], quando começam a balbuciar [165]. Logo após esse período a atenção volta aos olhos [166].

Em um estudo clássico de Klin *et al.* [167], indivíduos adolescentes e jovens adultos com TEA focaram 2 vezes mais na região da boca e por outro lado, 2 vezes menos na região dos olhos. Sendo que o tempo de fixação na região da boca foi um forte preditor de medidas de competência social e associado a adaptação social do indivíduo.

Embora a região dos olhos tenda a ser mais saliente, o nariz está localizado em uma parte mais central da face. Ao visualizar faces em movimento, as fixações concentravam-se no nariz, sugerindo que ele sirva como uma âncora espacial [168].

5.9.1.2 Expressões faciais: a detecção de expressões faciais é um campo muito debatido na literatura [169–171]. Frequentemente a expressão facial é representada usando Unidades de Ação/*Action Units* (AUs), que tentam descrever as ativações dos músculos faciais [142]. AUs são as unidades elementares da codificação *Facial Action Coding System*, que é um sistema que tenta estabelecer medir as ações faciais observáveis com base em movimentos discerníveis da musculatura facial [170]. A ativação das AUs não podem ser interpretadas como expressões de emoção, apenas padrões de ativação da musculatura facial [170]. Ainda assim, os movimentos faciais transmitem uma série de informações sociais e são importantes para a comunicação social, emocional ou não [171].

5.9.1.3 Direção do olhar: por meio do direcionamento dos olhos do indivíduo, pode-se inferir a localização do foco visual do indivíduo e conseqüentemente, sua atenção visual. Inferir a atenção visual do outro, estimula a pessoa a olhar para a mesma região, processo conhecido como atenção compartilhada. A atenção compartilhada pode fornecer pistas cognitivas espaciais e sobre ações futuras do indivíduo, portanto configurando-se como dicas semânticas [172].

5.9.1.4 Gestos manuais: os gestos são principalmente movimentos das mãos (embora também ocorram nos movimentos da cabeça e faciais) que expressam algum tipo de pensamento ou processo de pensar [173]. A interpretação de gestos manuais não é exclusiva para humanos, outros animais como cães, focas e gorilas, também têm essa habilidade. Gestos manuais podem conter informações sociais e atrair a atenção [174]. Apesar de sua interpretação variar de acordo com a cultura do observador [173].

Crianças com TEA produzem menos gestos e apresentam déficits na interpretação de gestos de atenção compartilhada, havendo uma deficiência de integração do olhar e do gesto [175]. Além disso, crianças com TEA prestam pouca atenção visual a gestos de apontar e não conseguem distinguir o apontar intencional do incidental [176].

5.9.2 *Extração das novas características*

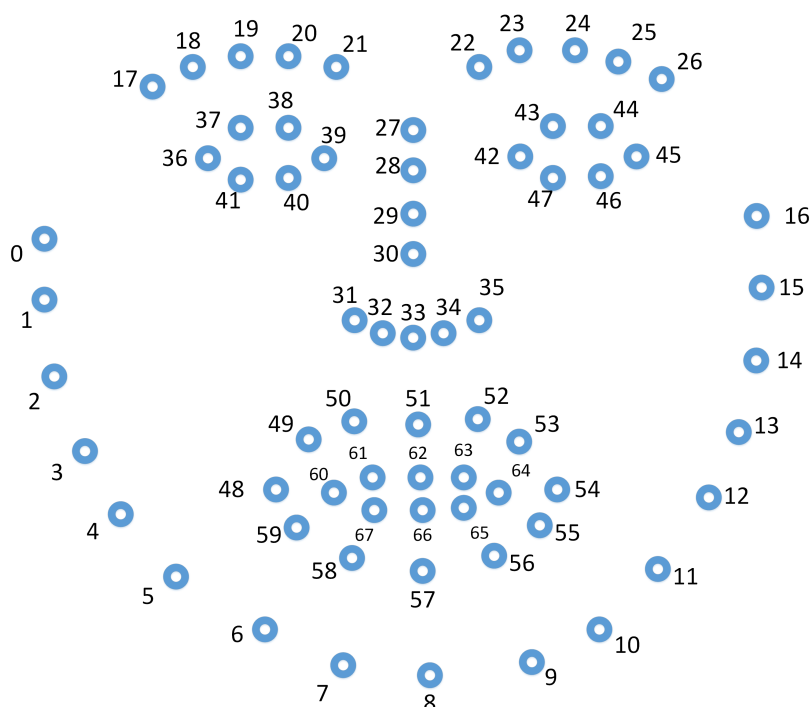
Conforme foi apresentado na seção anterior, o levantamento de literatura a respeito das características semânticas presentes no estímulo visual atual fomenta a implementação das seguintes características: 1 - Elementos faciais (Face, T-Facial, Olhos, Boca, Nariz); 2 - Expressões faciais - AUs; 3 - Direção do olhar; 4 - Gestos manuais.

A extração das características elementos faciais (Face, T-Facial, Olhos, Boca, Nariz), expressões faciais - AUs e direção do olhar, derivam do *Toolkit* OpenFace [177]. O OpenFace é um kit de ferramentas para a detecção de pontos de referência faciais (*landmarks*), estimativa de pose de cabeça, reconhecimento de AUs e estimativa de direcionamento do olhar com código-fonte disponível para execução e treinamento de modelos. A ferramenta pode ser executada a partir de fotos, vídeos ou captura via webcam [142].

A biblioteca OpenFace foi originalmente desenvolvida por Tadas Baltrušaitis em colaboração com o CMU MultiComp Lab liderado pelo Prof. Louis-Philippe Morency [142]. Atualmente o OpenFace está na versão 2.2 e continua sendo atualizado ativamente.

Dentre as aplicações da biblioteca, a detecção de *landmarks* faciais foi útil para a extração das características Face, T-Facial, Olhos, Boca e Nariz. A quantidade e localização dos *landmarks* podem variar de acordo com a referência utilizada. O OpenFace utiliza uma malha de *landmarks* com 68 pontos [142], conforme representado na figura 13, a qual é amplamente utilizada atualmente [178–180].

Figura 13: Malha de pontos utilizada - *landmarks*.



Fonte: Baltrušaitis [177].

A característica face já estava presente no estudo anterior com a implementação de Judd *et al.* [111], baseada no detector proposto por Viola & Jones [181]. No entanto, a antiga característica reconhece apenas 72,1% das faces no vídeo atual. Por conta disso, a detecção via OpenFace

foi adotada. Na nova detecção, 98,9% das faces foram reconhecidas, justificando a substituição da antiga implementação.

A detecção de face com o OpenFace foi realizada com o algoritmo *MTCNN*. O estímulo utilizado contém cenas de faces invertidas, o que dificulta a detecção das mesmas. Para superar essa dificuldade, a detecção foi realizada com o estímulo em seu posicionamento original e com o estímulo invertido (rotacionado em 180 graus). Posteriormente a taxa de confiabilidade de detecção foi comparada, assumindo-se o maior valor de confiabilidade. Caso a maior confiabilidade fosse maior considerando o estímulo invertido, a demarcação foi invertida para corresponder ao estímulo em sua versão original. Sem esta etapa de inversão, 75,2% das faces haviam sido reconhecidas adequadamente. Por outro lado, após esta etapa, 98,9% das faces foram reconhecidas.

O OpenFace apresenta três opções de algoritmo para detecção de *landmarks*, *CE – CLM*, *CLNF* e *CLM*. O algoritmo *CE – CLM* foi escolhido, uma vez que ele desempenha melhor para a detecção dos *landmarks* da face [142]. Em posse dos *landmarks* podemos combinar os pontos para corresponder aos elementos faciais. Para a característica olhos, foram considerados dois polígonos, abrangendo os pontos 36 ao 41 e 42 ao 47, conforme representado na figura 13. Para a característica nariz, foi plotado um polígono com os pontos 27 e 31 ao 35. Para a característica boca, foi plotado um polígono com os pontos 48 ao 59. Para a característica T-Facial, foi plotado um polígono com os pontos 26 ao 17, 36, 48, 59 ao 54 e 45.

No presente estudo adotamos a detecção do OpenFace [177] baseada em AUs. Ao todo 17 AUs são consideradas (AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45, conforme o sistema taxonômico para expressões faciais humanas, *Facial Action Coding System* [182]).

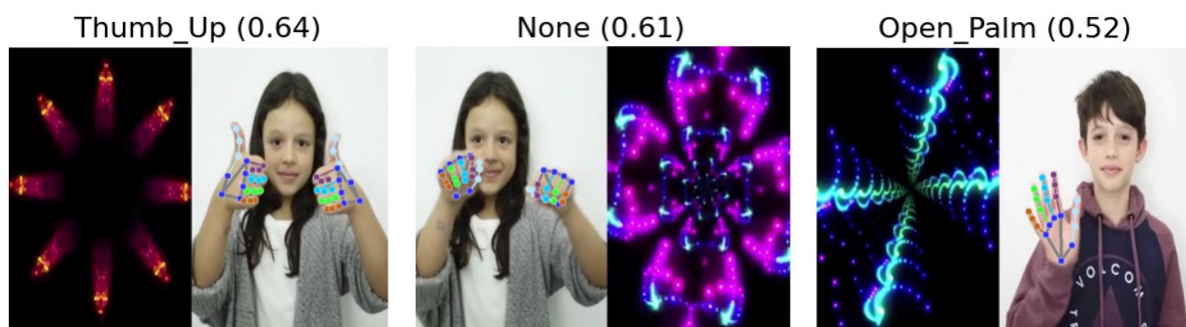
A característica direção do olhar foi estimada através do OpenFace [177], baseado em Wood *et al.* [183]. Existem três opções de dados a serem carregados, vídeo, imagens sequenciais e imagens (não sequenciais), a diferença no processamento é que nas duas primeiras opções, o posicionamento do olhar do *frame* atual é computado combinando o posicionamento do *frame* anterior, com peso 0,5 para cada. A opção imagens sequenciais foi escolhida para que as variações entre *frames* fossem suavizadas. O *output* do OpenFace fornece o posicionamento de cada olho separadamente, sendo as coordenadas resultantes plotadas em um polígono único.

A característica gestos manuais foi extraída a partir do *FrameWork* MediaPipe [146]. O *FrameWork* é *Open Source* e atualmente está na versão alfa 0.7. O MediaPipe oferece soluções

de aprendizado de máquina personalizáveis para reconhecimento de face, *landmarks* faciais, íris, mãos, gestos, poses corporais, objetos, dentre outras [184]. Os gestos considerados foram “Thumb_Up”, “Thumb_Down”, “Pointing_Up” e “Open_Palm”.

Por padrão, apenas uma mão é detectada, no entanto há o parâmetro *num_hands* para determinar o número máximo de mãos a ser considerado. Adotamos *num_hands* = 2. A confiabilidade de detecção é representada por *min_detection_confidence*, o qual foi adotado o valor 0,5. Na figura 14 são apresentados exemplos da detecção de gestos manuais por meio do MediaPipe. As coordenadas das regiões detectadas contendo mãos, bem como os gestos detectados foram salvos em um arquivo *JSON*. O qual foi pré-processado em linguagem R para redimensionamento das áreas demarcadas, adequando a resolução de 350x200 pixels e salvar a saída no formato *mat* por meio do pacote *R.matlab* (versão 3.7.0) [185], para posterior incorporação com as demais características em MatLab.

Figura 14: Exemplos da detecção de gestos manuais.



Os valores entre parênteses estão em língua inglesa (troca da vírgula pelo ponto) e referem-se a confiabilidade de detecção do respectivo gesto ou ausência de gesto (*None*).

5.10 Normalização de características

A fim de garantir uma comparação adequada de cada característica para os classificadores de pixels, foi adicionada uma etapa de normalização das características. A normalização foi aplicada após a etapa de agrupamento de *frames*, imediatamente antes da separação dos dados em treino e teste. A normalização Min-Max foi adotada, onde a amplitude de valores resultante varia entre 0 e 1 [186]. A etapa de normalização que foi incluída não visa melhorar o desempenho dos classificadores, mas sim auxiliar na comparação das características.

5.11 Análise Estatística

Os dados de caracterização da amostra foram analisados na IDE *RStudio* versão 1.3.1093 [187]. A normalidade e a homogeneidade foram verificadas pelos testes de Shapiro-Wilk e Levene, respectivamente. A comparação entre os grupos foi determinada por teste t, exceto no caso de suposições sobre a normalidade e homogeneidade das distribuições da amostra não serem atendidas, neste caso, o teste U de Mann-Whitney foi adotado. Valores de $p < 0,05$ foram considerados significantes. Para a comparação de desempenho dos classificadores de indivíduos foi considerada a AUC com Intervalo de Confiança (IC) de 95%.

6 RESULTADOS E DISCUSSÃO

6.1 Análises com o *dataset I* - Todos os indivíduos

6.1.1 Caracterização do *dataset I*

O *dataset I* está representado na figura 8 (seção 5.2) e a caracterização da amostra está disponível na tabela 6. Os resultados dos testes de distribuição não satisfizeram as premissas para a aplicação do teste t, portanto as comparações para a caracterização da amostra foram conduzidas com o teste U. Em relação à comparação das idades, não houve diferença ($p = 0,082$) entre os grupos DT e TEA. A comparação entre os grupos TEA do estudo atual e anterior apresentou diferença ($p < 0,001$), os pacientes são 3,2 anos mais novos neste estudo. A principal variável de comparação entre os grupos de TEA foi o escore CARS, a qual não apresentou diferença entre os *datasets* ($p = 0,461$), sugerindo que a gravidade do TEA é comparável ao trabalho anterior. Adicionalmente, o QI e os domínios da Vineland não apresentaram diferença entre os *datasets* para o grupo TEA. Como esperado, a população de indivíduos do sexo feminino está sub-representada na amostra, pois a prevalência de transtornos do desenvolvimento é maior no sexo masculino [188].

Os indivíduos que participaram do estudo são residentes majoritariamente da zona metropolitana de São Paulo, com alguns indivíduos do Rio de Janeiro, Minas Gerais e interior de São Paulo, conforme representado na figura 26 do Apêndice B. A distribuição geográfica dos indivíduos de ambos grupos foi similar, não havendo grandes concentrações regionais em nenhum grupo.

Tabela 6: Caracterização da amostra - *dataset* Anterior (descrito em detalhes em Oliveira *et al.* [1]) versus I (*dataset* atual considerando todos os indivíduos).

Dataset	Anterior		I	
	DT	TEA	DT	TEA
Grupos				
n Amostral	30	76	49	50
Idade (anos)	9,8±2,8	8,8±3,7	6,2 ± 3,2	5,6 ± 3,8
Masc./Fem.	20/10	49/27	45 / 4	41 / 9
CARS	-	35,0 ± 04,0	-	36,1 ± 05,4
QI	-	89,0 ± 27,1	104,4± 15,8	88,4 ± 18,5
VDL	-	60,5 ± 10,7	93,0 ± 14,4	64,5 ± 11,3
VS	-	61,0 ± 12,6	94,8 ± 15,6	62,5 ± 10,7
VC	-	60,6 ± 13,4	91,9 ± 16,3	57,7 ± 15,4
VL	-	77,0 ± 15,5	88,2 ± 11,3	75,6 ± 10,7

Valores decimais expressos como média ± desvio padrão. Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista; Masc., Sexo Masculino; Fem., Sexo Feminino; CARS, Childhood Autism Rating Scale; QI, Quociente de Inteligência; VC, Vineland Communication; VDL, Vineland Daily Living; VS, Vineland Socialization; VL, Vineland Locomotion.

6.1.2 Validação do estudo anterior

No experimento 1, onde o modelo pré-treinado foi usado para classificar novas amostras, os valores de AUC, F1-Score, acurácia, sensibilidade, precisão e especificidade foram 0,56, 0,53, 0,58, 0,55, 0,51 e 0,61, respectivamente. A utilização do modelo pré-treinado não apresentou resultados satisfatórios, atingindo AUC de 0,56 contra 0,82 do valor de referência, conforme mostrado na tabela 7.

Tabela 7: Resultados de replicação do estudo prévio e implementação do Random Forest.

Dataset	Abordagem	AUC	F1-Score	Acurácia	Sens.	Precisão	Esp.
I (Atual)	Pré-treinado (NN)	0,56	0,53	0,58	0,55	0,51	0,61
	Replicação (NN)	0,65	0,61	0,59	0,61	0,61	0,57
	Random Forest	0,69	0,62	0,63	0,57	0,67	0,69
Anterior	Referência (NN)*	0,82	0,62	0,76	0,69	0,90	0,93
	Random Forest	0,94	0,87	0,83	0,79	0,97	0,94

* resultado descrito por Oliveira *et al.*, 2021 [1] e adicionado para efeito de comparação. Os valores são expressos como média, exceto os valores referentes ao classificador pré-treinado. Abreviações: AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Esp., Especificidade; NN, Rede Neural.

No experimento 2, o classificador foi retreinado, ou seja, tanto as amostras de treino quanto as amostras de teste foram coletadas com o mesmo estímulo. Nessa condição, as médias de AUC, F1-Score, acurácia, sensibilidade, precisão e especificidade obtidas foram de 0,65, 0,61, 0,59, 0,61, 0,61 e 0,57, respectivamente, conforme descrito na tabela 7 e identificado como “Replicação (NN)”. Portanto, o desempenho foi inferior ao valor de referência (AUC média de

0,82).

No teste com classificador pré-treinado, os resultados foram muito baixos quando aplicados a um novo estímulo, atingindo uma AUC de 0,56. Esse resultado sugere que o classificador pré-treinado não apresenta capacidade de generalização, sendo inadequado para ser utilizado para um vídeo diferente e sendo necessário um retreinamento dos MAVs ao mudar o estímulo. Por outro lado, os resultados do experimento de replicação foram superiores, ainda que inferiores aos apresentados no estudo de referência. Este resultado mostra que o método proposto pode ser utilizado mesmo sem reparametrização, ou seja, sem otimizar os parâmetros para cada conjunto de dados. No entanto, devemos considerar diferenças entre os estímulos utilizados e idade entre os *datasets*.

Nesse contexto, as principais diferenças entre os estímulos atual e anterior foram presença de menos pessoas nas cenas biológicas, menor taxa de movimento entre *frames* e cores mais vibrantes nas cenas geométricas no atual. A mudança de estímulo permite verificar a robustez do método, mesmo com a alteração de cores, o que pode atrair mais a atenção [189]. Assim, foi possível utilizar um vídeo pautado no mesmo paradigma apresentado por Oliveira *et al.* [1], desde que os MAVs sejam retreinados e com uma taxa de movimento semelhante. A taxa de movimento é importante para a abordagem, pois impacta diretamente na agregação de *frames*. Resultados apresentados em um artigo recente do grupo [190] mostraram que alterar o limiar de agregação de *frames* impacta no desempenho dos classificadores de indivíduos.

Em relação ao novo conjunto de dados, conforme mencionado na seção 5.3, o rastreador ocular usado tinha uma frequência diferente, operando a 250 Hertz em vez de 300 Hertz, o que leva a uma quantidade menor de registros de rastreamento ocular, com 17% menos dados. Também é importante notar que os indivíduos são mais jovens. Mais adiante, nos aprofundaremos nessa questão com o *dataset* II (exclusão dos indivíduos < 4 anos).

Apesar das limitações apresentadas, os resultados do experimento de replicação sugerem a robustez e a viabilidade de abordagens baseadas em MAVs para auxiliar no diagnóstico do TEA, desde que ocorra o retreinamento do classificador ao trocar o estímulo.

Mastergeorge *et al.*, [40] ao fazer uma revisão sistemática de estímulos visuais em paradigmas experimentais com utilização do rastreamento do olhar no TEA, concluíram que os principais problemas foram tamanhos de amostra baixos, heterogeneidade dos estímulos e métodos analíticos deficientes. Eles sugerem que esses problemas limitam o valor das descobertas e destacam a importância da replicabilidade. Nesse contexto, realizar a replicação do estudo

anterior como base para as alterações propostas se faz necessário.

6.1.3 Desempenho dos classificadores com diferentes datasets

Oliveira *et al.* [1] propuseram um método com o algoritmo indutor de classificação NN, que apresentou bons resultados, com AUC média de 0,82. Ao testar a abordagem com o algoritmo RF, obtivemos um resultado estatisticamente superior para o *dataset* anterior, com AUC média de 0,94 (tabela 7). Testes preliminares foram realizados com SVM com kernels não lineares, porém como os resultados parciais foram inferiores ao RF, não seguimos com as análises.

Ao considerar o *dataset* I, os resultados também foram melhores com algoritmo RF, conforme a tabela 8. Os ICs para RF e NN se interseccionaram, por outro lado, o RF foi superior ao SVM. A principal métrica de desempenho utilizada é a AUC, no entanto, ao considerar as demais métricas de desempenho, verifica-se que a condição com o algoritmo NN, 28 características e sem normalização apresentou os melhores valores de acurácia e F1-Score, ainda que sem diferença estatística comparado aos demais resultados.

Tabela 8: Desempenho dos classificadores com a adição das novas características semânticas ao *dataset* I.

Indutor	Caract.	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
RF	28	Não	0,69	0,57	0,69	0,67	0,63	0,62
RF	28	Sim	0,70	0,56	0,73	0,69	0,64	0,62
RF	34	Não	0,70	0,54	0,75	0,69	0,64	0,61
RF	34	Sim	0,70	0,80	0,50	0,63	0,65	0,70
NN	28	Não	0,63	0,71	0,73	0,79	0,66	0,75
NN	Relief 28	Não	0,65	0,61	0,57	0,61	0,59	0,61
NN	34	Não	0,64	0,60	0,58	0,62	0,59	0,61
NN	Relief 34	Não	0,63	0,60	0,61	0,63	0,60	0,61
SVM	28	Não	0,59	0,56	0,55	0,56	0,56	0,56
SVM	GA 28	Não	0,53	0,54	0,51	0,53	0,53	0,54
SVM	34	Não	0,57	0,68	0,41	0,55	0,55	0,61
SVM	Relief 34	Não	0,55	0,45	0,60	0,54	0,52	0,49

O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA). A condição “28 Características” representa as características basais e “34 Características” representa o conjunto com as novas características. Abreviações: RF, Random Forest; NN, Redes Neurais; SVM, Máquina de Vetores de Suporte; Caract., Características; Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

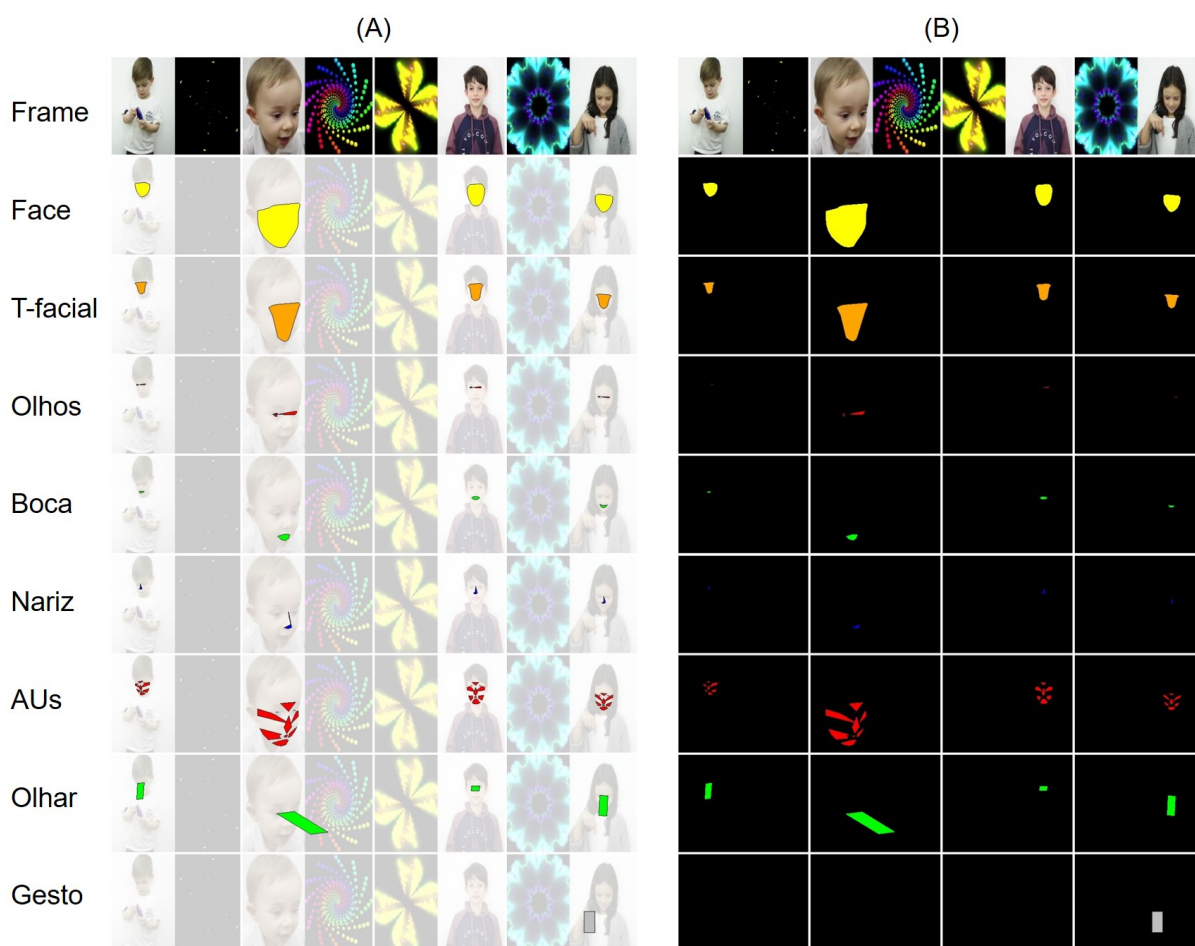
A alteração do método proposto melhorou o desempenho da classificação, evidenciando a viabilidade e superioridade do RF para a construção de MAVs. Provavelmente, a melhoria de desempenho se deva à seleção interna de características pelo algoritmo RF. Além disso, quatro das características são categóricas, o que pode ser uma vantagem para algoritmos baseados em árvores de decisão. Nesse processo de seleção de características, internamente a importância

relativa de cada característica é calculada, o que favorece o processo de interpretabilidade dos classificadores de pixels e contribui com os objetivos do projeto.

6.1.4 Adição de características semânticas e avaliação de desempenho dos classificadores de indivíduos - Dataset I

Para efeitos ilustrativos, as novas características extraídas são representadas na figura 15, são elas: Face (nova), T-Facial, Olhos, Boca, Nariz, AUs, Direção do olhar e Gesto Manual, conforme descrito na seção 5.9.2.

Figura 15: Exemplos das novas características extraídas.



A - Representação das características extraídas sobre os *frames* para facilitar a localização das mesmas; B - Representação das características extraídas sobre fundo preto para evidenciá-las. Observação: A característica Face mencionada na figura refere-se a nova implementação utilizada.

Ainda na tabela 8 (introduzida na seção anterior, 6.1.3), são apresentados os resultados de desempenho para o *dataset I*, considerando os algoritmos RF, NN e SVM, a inclusão da etapa de normalização e a adição das novas características. É importante destacar que as 28

características representam as basais (descritas nas tabelas 3 e 4), utilizadas no estudo anterior e ao adicionar as novas, temos 34 características. Ambos os conjuntos foram utilizados nesse estudo.

O melhor resultado foi utilizando o algoritmo RF com normalização e a adição das novas características, com AUC média de 0,70. No entanto, todos os resultados com o algoritmo RF variaram dentro do IC de 95%. Esse resultado mostra que não houve melhora de desempenho para o *dataset* I ao adicionar as características. O único fator que interferiu no desempenho foi o algoritmo indutor de classificação utilizado, conforme discutido anteriormente.

6.1.5 Contribuição das características - Dataset I

Ao comparar a importância relativa das características temos que levar em consideração que as características semânticas ocupam uma pequena área da imagem, com valores binários (0 e 1), ou seja, a característica está ou não presente em um determinado pixel. Por outro lado, características como cor, centro e movimento se estendem por toda a imagem, com valor contínuo para os pixels. Portanto, características semânticas podem contribuir menos para os classificadores de pixels por esse aspecto e não necessariamente por terem menor importância para a atenção visual.

Na figura 16-A verificamos a importância relativa da condição com 28 características consideradas para o *dataset* I com o algoritmo RF. Apenas a característica presença de pessoa apresentou diferença na importância relativa para prever se um pixel foi ou não fixado para cada um dos grupos DT e TEA, com menor importância relativa para o grupo TEA. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas.

Na figura 16-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos, com as características a nível de objeto (médio nível) contribuindo mais para o processo de classificação.

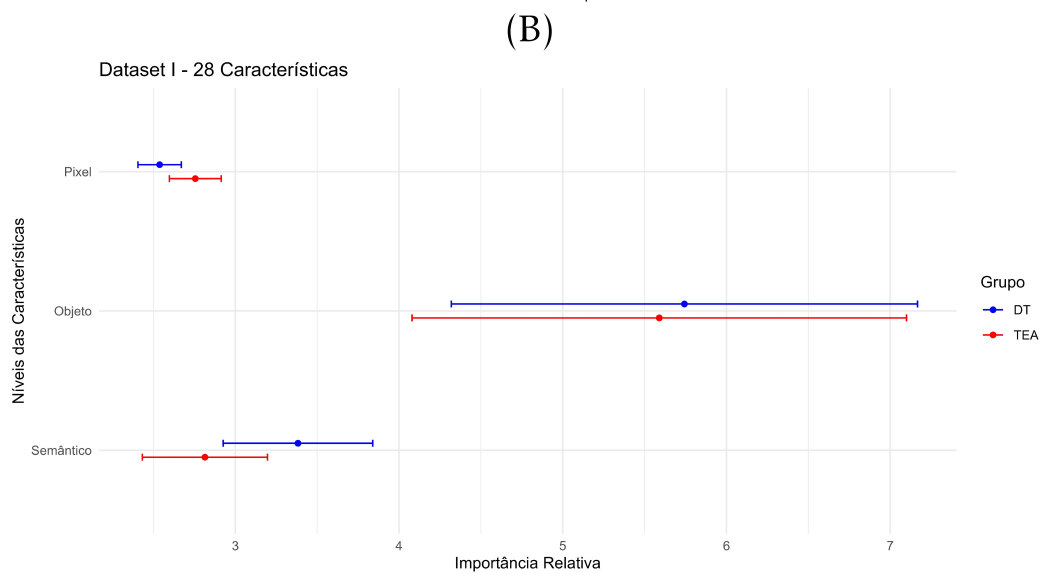
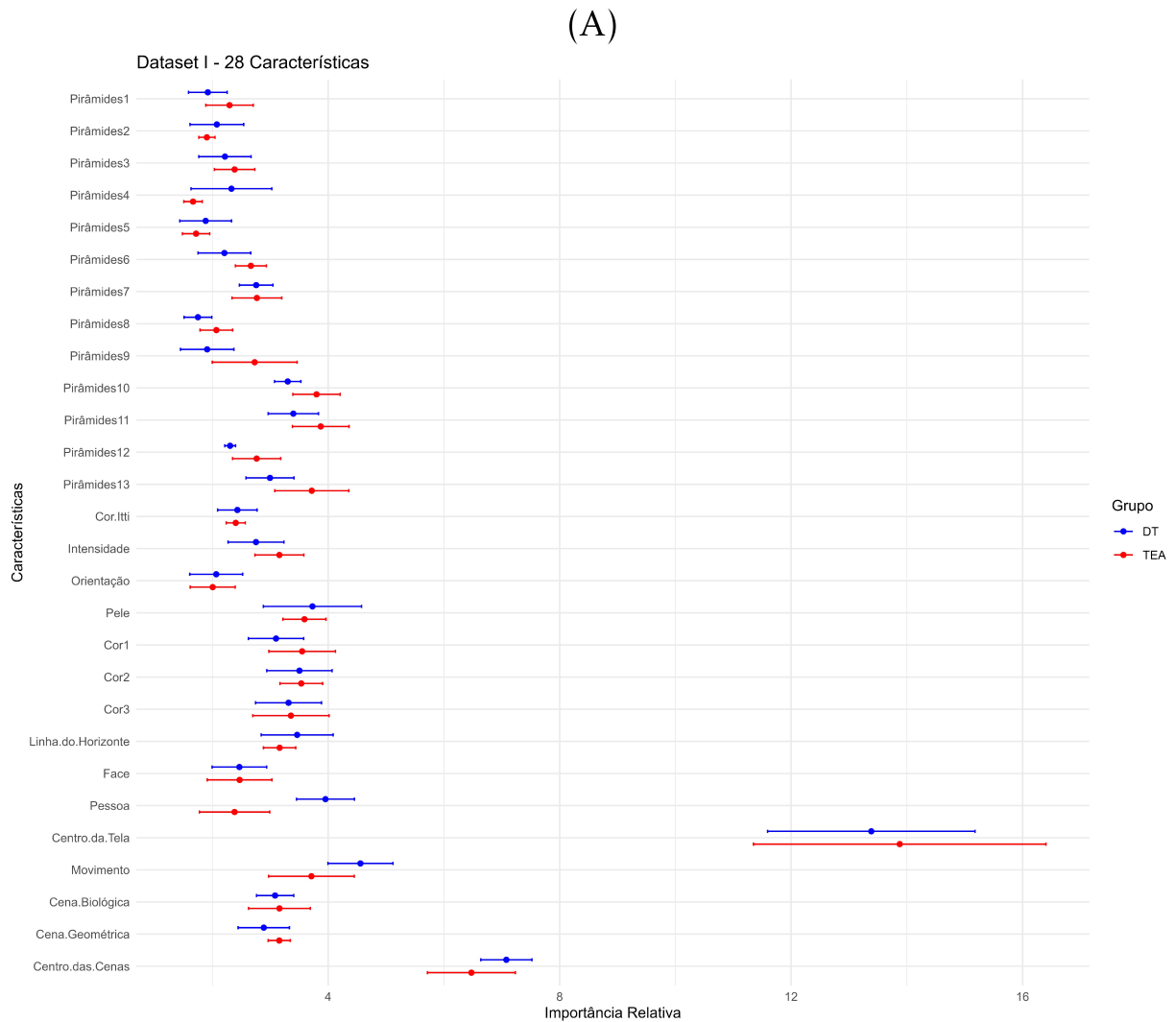
Na figura 17-A verificamos a importância relativa da condição com 34 características consideradas para o *dataset* I com o algoritmo RF. Além da diferença já observada para a característica presença de pessoa na condição com 28 características, verificamos também diferença para as características pirâmide orientável 12 e presença de movimento. A característica pirâmide orientável 12 apresentou maior importância relativa para o grupo TEA, enquanto que a presença de movimento apresentou menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas para ambos os grupos.

Conforme explicado na seção 1.8, as pirâmides orientáveis são multiescala e multiorientação. Desta forma, conserva-se a mesma resolução de imagem no primeiro nível de escala (o que é mais adequado para abordagens baseadas na percepção local), dividindo-se recursivamente para as próximas escalas [191]. Essas características são processadas pela área V1 do córtex visual e auxiliam tanto na percepção global quanto local da imagem [191].

A percepção local abrange situações em que o processamento é restrito a regiões menores do campo visual ou refere-se a características de um estímulo visual que podem ser processados isoladamente, de forma mais detalhada. A percepção global, entretanto, envolve um processamento dedicado a regiões maiores, permitindo abstrair detalhes e focar mais nos aspectos gerais para captar a imagem completa ou toda a cena visual. Ao visualizar uma cena é necessária a interação entre os diferentes aspectos perceptuais [192].

Portanto, a pirâmide orientável 12 por ser uma das últimas (de 1 a 13) possui uma escala pequena (*i.e.* menor granulosidade e resolução), representando uma visão mais global da imagem. Na literatura, possíveis diferenças na percepção global da imagem ainda é controversa para o TEA [192, 193]. O que parece haver é uma demanda maior de tempo para a percepção global em indivíduos TEA, sendo o processamento local mais espontâneo e automático para indivíduos com TEA [192].

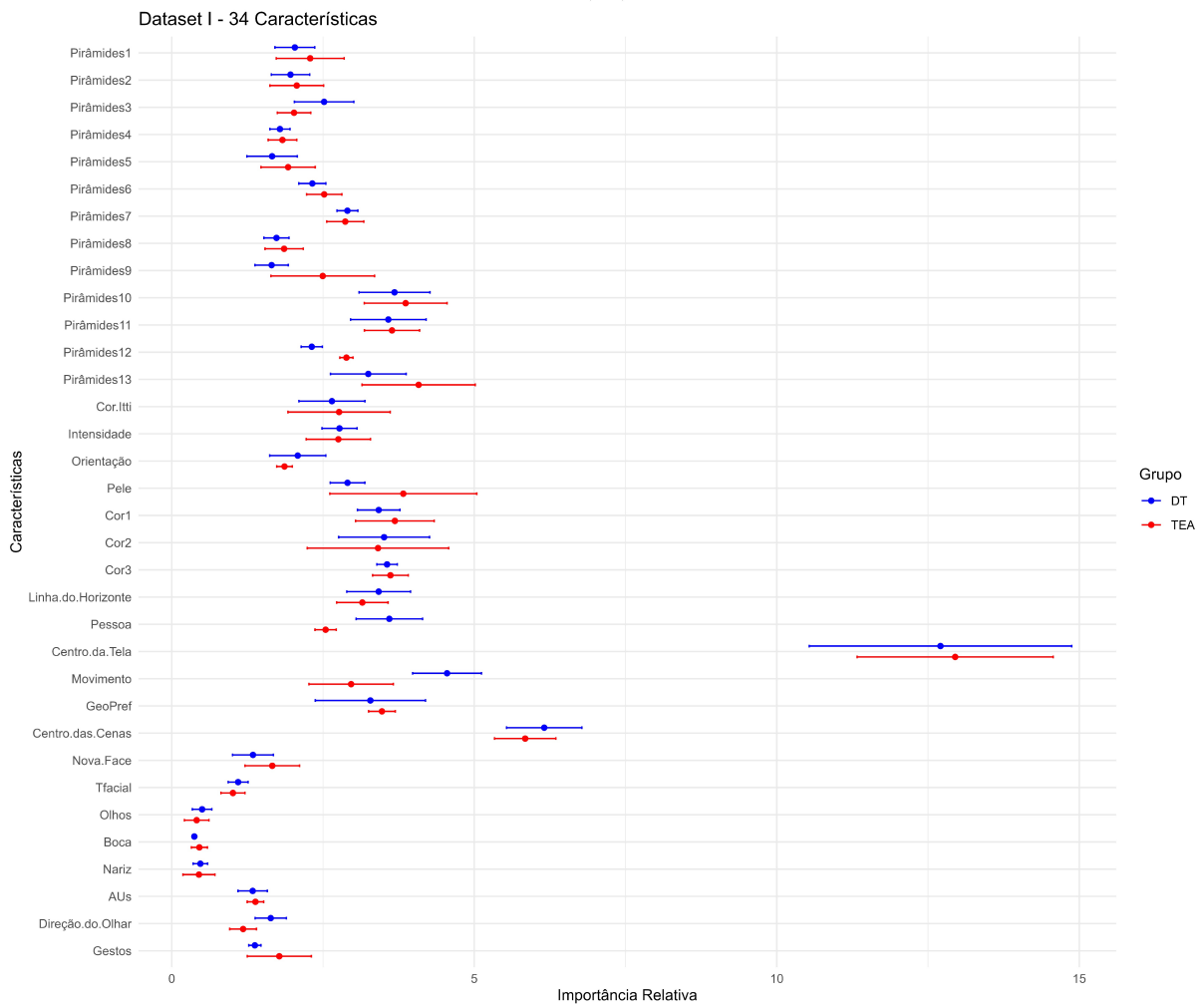
Figura 16: Importância relativa das características - Dataset I: 28 características.



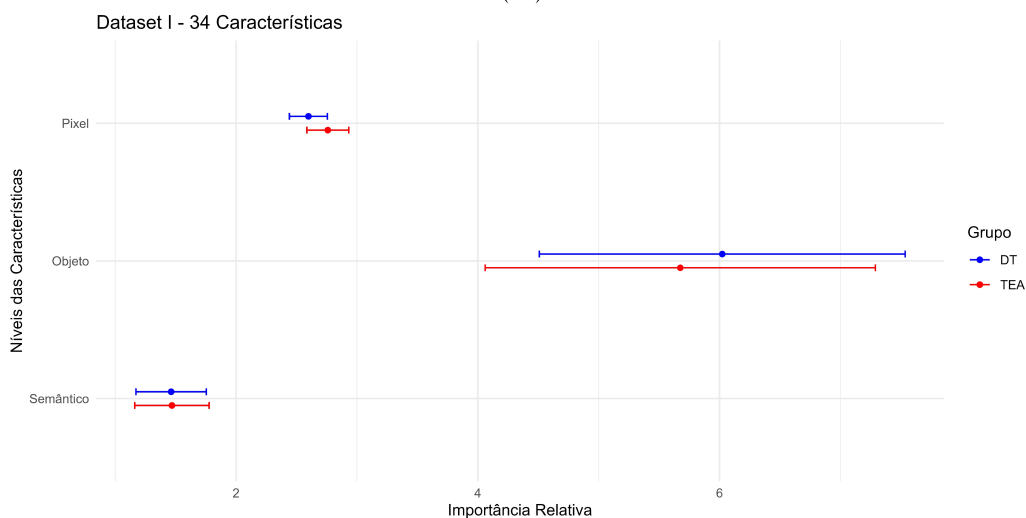
As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

Figura 17: Importância relativa das características - Dataset I: 34 características.

(A)



(B)



As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

Na figura 17-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos, com as características a nível de objeto contribuindo mais para o processo de classificação. Portanto, sem alteração do padrão exibido na condição anterior com 28 características.

Adicionalmente, foi verificada a possibilidade de classificar os indivíduos considerando apenas as características semânticas, porém os valores médios de AUC variaram entre 0,57 e 0,64, conforme o Apêndice C. Com base no baixo desempenho de classificação, não prosseguimos com os experimentos.

6.2 Análises com o *dataset II* - Exclusão de indivíduos menores de 4 anos

A fim de verificar se a idade dos indivíduos impactaria no desempenho de classificação, subdividimos o *dataset I*, excluindo indivíduos menores de 4 anos de ambos os grupos.

Pierce *et al.* [53] utilizaram o paradigma de GeoPref para classificar indivíduos DT e TEA, porém sem a construção de MAV, portanto sem considerar as características da imagem. Foi verificado que indivíduos mais velhos apresentaram tendência a olhar mais para cenas geométricas e no caso estudado, os autores afirmam que esse padrão observado pode diminuir a confiabilidade do teste GeoPref para indivíduos maiores de 4 anos. O limiar utilizado para estabelecer os indivíduos do *dataset II*, derivou dessa afirmação. Pelos indícios que tínhamos de testes prévios, isso não parecia ocorrer em nosso cenário, o que foi elucidado nesta seção.

6.2.1 Caracterização do *dataset II*

O *dataset II* está representado na figura 8 (seção 5.2) e a caracterização da amostra está disponível na tabela 9. Os pressupostos de homogeneidade e normalidade foram atendidos em todos os casos apresentados, com exceção da idade. Em relação à comparação das idades, não houve diferença ($p = 0,079$ no teste U) entre os grupos DT e TEA. A comparação entre os grupos DT dos *datasets I* e II apresentou diferença ($p = 0,011$ no teste U), os pacientes são em média 1,5 anos mais velhos no *dataset II*. De maneira semelhante, a comparação entre os grupos TEA dos *datasets I* e II apresentou diferença ($p = 0,002$ no teste U), os pacientes são 2,4 anos mais velhos no *dataset II*. As demais variáveis QI, CARS e os domínios da Vineland não apresentaram diferenças entre os *datasets*, considerando seus respectivos grupos. Portanto, a única diferença avaliada entre os *datasets* é a idade.

Tabela 9: Caracterização da amostra - *dataset* I versus II.

Dataset	I		II	
	DT	TEA	DT	TEA
Grupos				
n Amostral	49	50	33	28
Idade (anos)	6,2 ± 3,2	5,6 ± 3,8	7,8 ± 2,7	8,0 ± 3,6
Masc./Fem.	45 / 4	41 / 9	30 / 3	26 / 2
CARS	-	36,1 ± 5,4	-	33,8 ± 04,9
QI	104,4 ± 15,8	88,4 ± 18,5	105,5 ± 16,5	88,7 ± 23,1
VDL	93,0 ± 14,4	64,5 ± 11,3	97,0 ± 14,5	63,9 ± 09,8
VS	94,8 ± 15,6	62,5 ± 10,7	97,0 ± 15,2	61,5 ± 12,6
VC	91,9 ± 16,3	57,7 ± 15,4	93,5 ± 18,0	60,2 ± 19,2
VL	88,2 ± 11,3	75,6 ± 10,7	89,1 ± 13,2	70,0 ± 09,3

O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA) e o *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). Valores decimais expressos como média ± desvio padrão. Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista; Masc., Sexo Masculino; Fem., Sexo Feminino; CARS, Childhood Autism Rating Scale; QI, Quociente de Inteligência; VC, Vineland Communication; VDL, Vineland Daily Living; VS, Vineland Socialization; VL, Vineland Locomotion.

6.2.2 Desempenho dos classificadores de indivíduos - Dataset II

Na tabela 10 são apresentados os resultados de desempenho de classificação para o *dataset* II considerando os algoritmos RF, NN e SVM, a inclusão da etapa de normalização e a adição das novas características. O principal resultado foi com o algoritmo RF e a adição das novas características, com AUC média de 0,79, enquanto que sem as novas características a AUC média foi de 0,73. No entanto, essa flutuação ficou dentro do IC, portanto não foi estatisticamente diferente. O único fator que interferiu no desempenho foi o algoritmo indutor de classificação utilizado, em que o RF foi superior. Conforme discutido na seção 6.1.3, a melhoria de desempenho possivelmente se deva à seleção interna de características pelo algoritmo RF.

Na figura 18, os valores médios de AUC são apresentados juntamente com o IC para todos os *datasets* I, II, III e IV para efeito comparativo, considerando o algoritmo RF, o qual desempenhou melhor em todos os casos. Ao remover os indivíduos menores de 4 anos, o desempenho de classificação melhorou, com aumento fora do IC na comparação dos *datasets* I e II, considerando o conjunto de 34 características com normalização, conforme apresentado na figura 18. Esse resultado sugere que indivíduos mais velhos são melhores classificados por essa abordagem. Conforme mencionado ao discutir o resultado da replicação, no estudo anterior os indivíduos eram mais velhos e o desempenho de classificação foi superior aos obtidos na replicação. As duas principais mudanças foram no estímulo e na idade, com indivíduos mais novos na replicação. Levando em consideração esses dois achados, acreditamos que a idade seja uma variável chave no desempenho de classificação, uma vez que os indivíduos avaliados

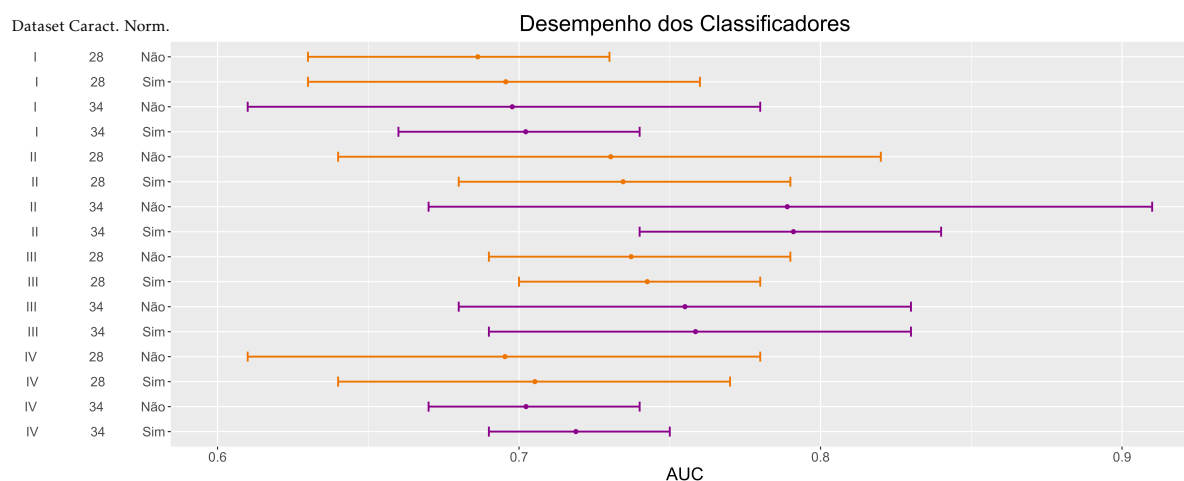
Tabela 10: Desempenho dos classificadores com o *dataset* II.

Indutor	Caract.	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
RF	28	Não	0,73	0,76	0,57	0,61	0,66	0,67
RF	28	Sim	0,73	0,56	0,79	0,69	0,68	0,62
RF	34	Não	0,79	0,80	0,66	0,67	0,72	0,73
RF	34	Sim	0,79	0,76	0,70	0,68	0,72	0,72
NN	28	Não	0,66	0,70	0,54	0,56	0,61	0,62
NN	GA 28	Não	0,67	0,61	0,61	0,58	0,61	0,59
NN	34	Não	0,72	0,51	0,76	0,68	0,64	0,58
NN	Relief 34	Não	0,71	0,70	0,62	0,65	0,66	0,67
SVM	28	Não	0,64	0,56	0,67	0,59	0,62	0,57
SVM	Relief 28	Não	0,61	0,72	0,51	0,56	0,61	0,63
SVM	34	Não	0,61	0,68	0,48	0,53	0,57	0,59
SVM	GA 34	Não	0,57	0,44	0,65	0,52	0,56	0,48

O *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). A condição “28 Características” representa as características basais e “34 Características” representa o conjunto com as novas características. Abreviações: RF, Random Forest; NN, Redes Neurais; SVM, Máquina de Vetores de Suporte; Caract., Características; Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

no presente estudo ainda têm um comportamento oculomotor em processo de maturação, o qual perdura até a adolescência [194].

Figura 18: Desempenho dos classificadores com *Random Forest* e adição das novas características para os *datasets* I, II, III e IV.



O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA). O *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). No *dataset* III houve a exclusão dos indivíduos com TEA leve (49 DT e 31 TEA). No *dataset* IV houve a exclusão de indivíduos com DI (49 DT e 38 TEA). A condição “28 Características” representa as características basais e “34 Características” representa o conjunto com as novas características. Abreviações: Caract., Características; Norm., Normalização; AUC, Área Sob a Curva ROC.

Conforme mencionado na seção 6.1.5, a literatura sobre possíveis diferenças na percepção local e global da imagem ainda é controversa para o TEA [192, 193]. No entanto, parece haver diferenças conforme o desenvolvimento do indivíduo. Guy *et al.* [193] verificaram diferenças

nas percepções local e global entre DT e TEA de acordo com a idade. Os autores concluíram que em indivíduos com TEA, a percepção global não se desenvolve na mesma extensão que no DT e que a percepção local pode permanecer dominante na idade adulta. Portanto, com o passar do tempo, as trajetórias se afastam nesse quesito, sendo essa uma possível justificativa para o melhor desempenho de classificação dos indivíduos mais velhos.

6.2.3 *Contribuição das características - Dataset II*

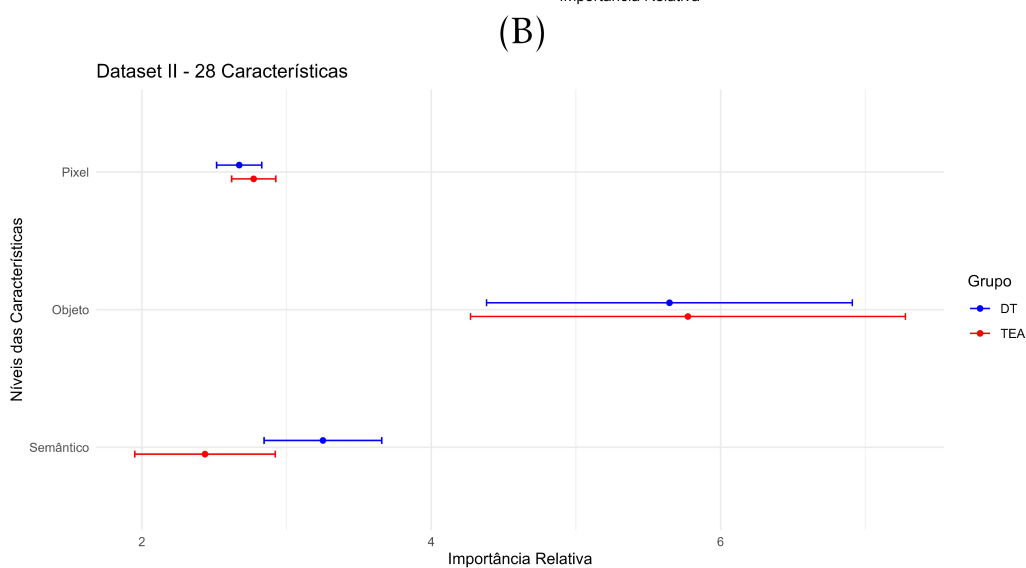
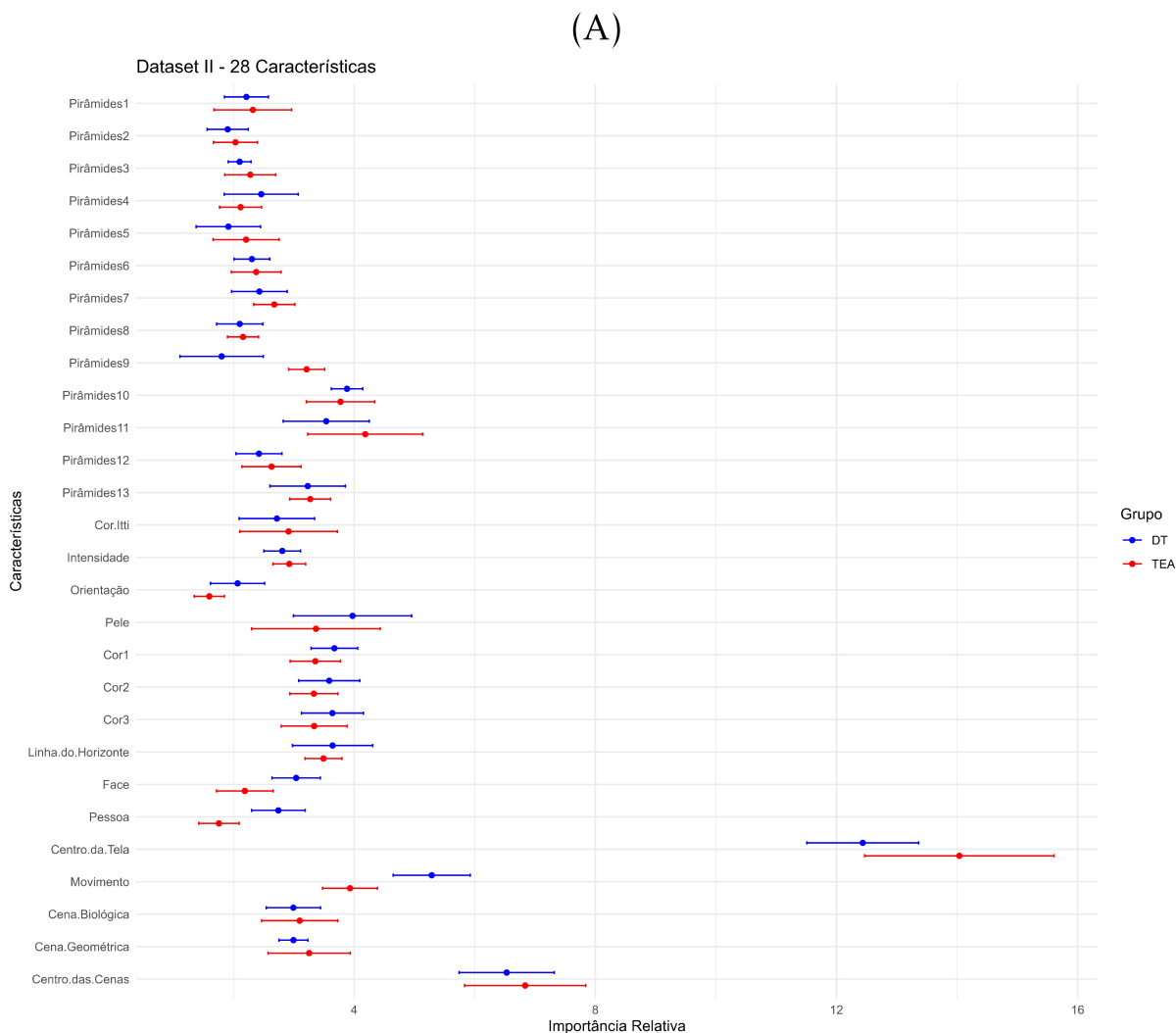
Na figura 19-A verificamos a importância relativa da condição com 28 características consideradas para o *dataset II* com o algoritmo RF. As características presença de pessoa, pirâmide orientável 9 e presença de movimento foram diferencialmente importantes entre os grupos. A característica pirâmide orientável 9 apresentou maior importância relativa para o grupo TEA, enquanto que a presença de pessoa e movimento apresentaram menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas.

Na figura 19-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos. Embora a importância relativa média para o nível semântico seja menor para o grupo TEA, a variação ficou dentro do IC. Assim como ocorreu com o *dataset I*, as características a nível de objeto foram as que mais contribuíram.

Na figura 20-A verificamos a importância relativa da condição com 34 características para o *dataset II* com o algoritmo RF. As características presença de pessoa, T-Facial e presença de movimento foram diferencialmente importantes entre os grupos. A característica T-Facial apresentou maior importância relativa para o grupo TEA, enquanto que presença de pessoa e movimento apresentaram menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas.

Na figura 20-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos, com as características a nível de objeto contribuindo mais para a construção dos mapas de saliência. Portanto, sem alteração do padrão exibido na condição anterior, com 28 características.

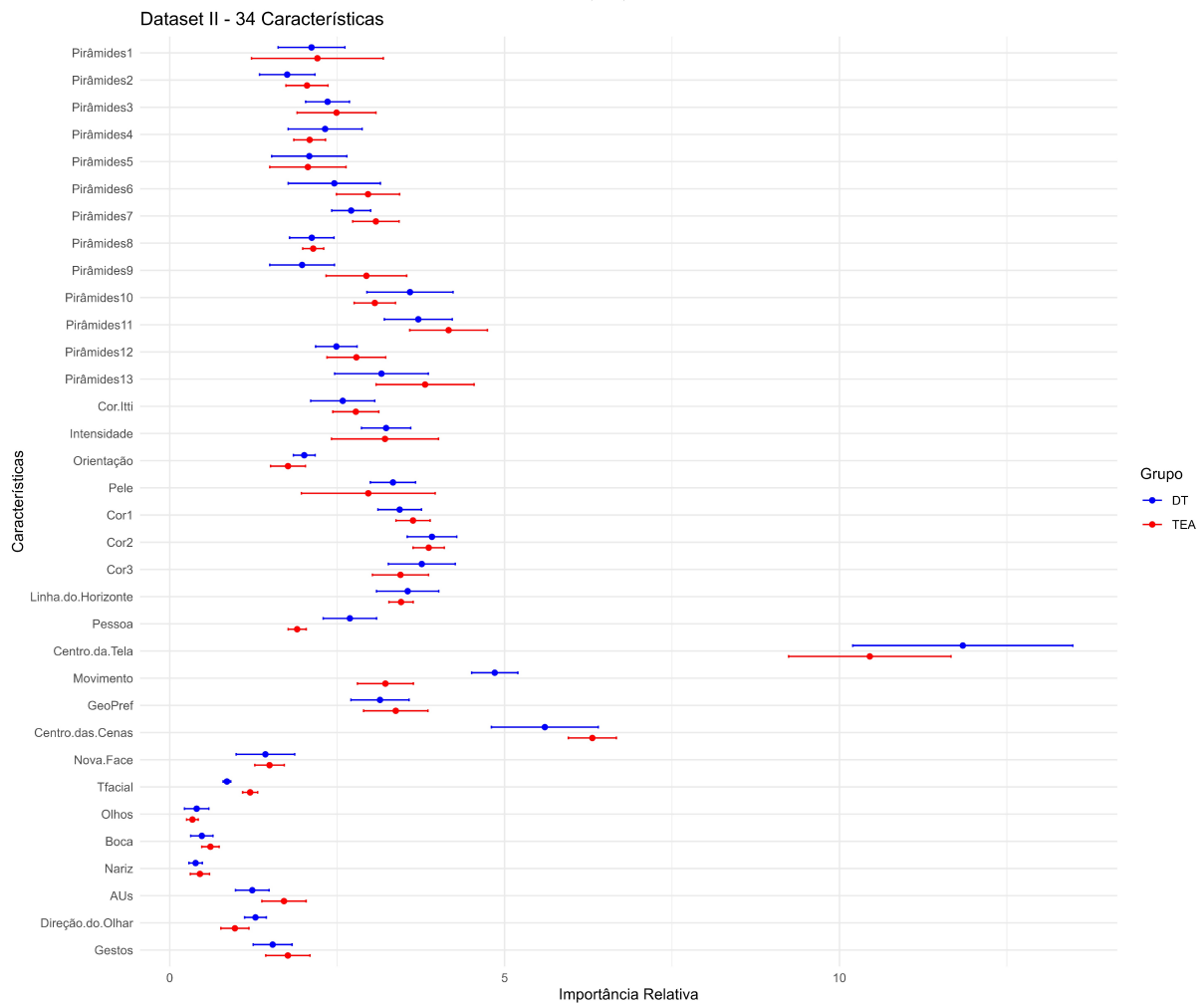
Figura 19: Importância relativa das características - Dataset II: 28 características.



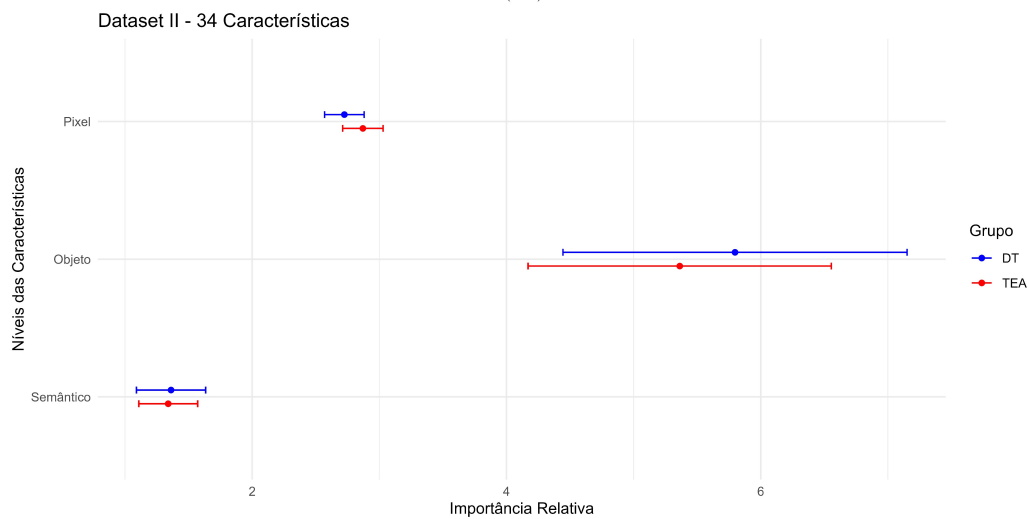
As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

Figura 20: Importância relativa das características - Dataset II: 34 características.

(A)



(B)



As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

6.3 Análises com o *dataset* III - Exclusão de TEA leve

Para verificar se indivíduos com TEA leve podem estar dificultando a classificação, excluímos os indivíduos sabidamente com TEA leve ($CARS > 36$) do *dataset* I. Portanto, no *dataset* III temos 19 indivíduos a menos no grupo TEA.

6.3.1 Caracterização do *dataset* III

O *dataset* III está representado na figura 8 (seção 5.2) e a caracterização da amostra está disponível na tabela 11. Os pressupostos de homogeneidade e normalidade foram atendidos em todos os casos apresentados, com exceção da idade. Em relação às comparações entre os grupos DT e TEA, a idade não apresentou diferença ($p = 0,060$ no teste U). O QI foi 12,4% (12,93 unidades) maior no grupo DT ($p = 0,009$ no teste t). A comparação de idade entre os grupos TEA dos *datasets* I e III não apresentou diferença ($p = 0,763$ no teste U). No *dataset* III a CARS foi 11,9 (4,3 unidades) maior que no *dataset* I ($p = 0,001$ no teste t). As demais variáveis não apresentaram diferença para o grupo TEA entre os *datasets*.

Tabela 11: Caracterização da amostra - *dataset* I versus III.

Dataset	I		III	
	DT	TEA	DT	TEA
Grupos				
n Amostral	49	50	49	31
Idade (anos)	$6,2 \pm 3,2$	$5,6 \pm 3,8$	$6,2 \pm 3,2$	$5,6 \pm 4,4$
Masc./Fem.	45 / 4	41 / 9	45 / 4	26 / 5
CARS	-	$36,1 \pm 5,4$	-	$40,4 \pm 02,9$
QI	$104,4 \pm 15,8$	$88,4 \pm 18,5$	$104,4 \pm 15,8$	$91,5 \pm 20,6$
VDL	$93,0 \pm 14,4$	$64,5 \pm 11,3$	$93,0 \pm 14,4$	$63,1 \pm 12,1$
VS	$94,8 \pm 15,6$	$62,5 \pm 10,7$	$94,8 \pm 15,6$	$60,8 \pm 11,2$
VC	$91,9 \pm 16,3$	$57,7 \pm 15,4$	$91,9 \pm 16,3$	$54,7 \pm 17,0$
VL	$88,2 \pm 11,3$	$75,6 \pm 10,7$	$88,2 \pm 11,3$	$75,5 \pm 12,4$

Valores decimais expressos como média \pm desvio padrão. Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista; Masc., Sexo Masculino; Fem., Sexo Feminino; CARS, Childhood Autism Rating Scale; QI, Quociente de Inteligência; VC, Vineland Communication; VDL, Vineland Daily Living; VS, Vineland Socialization; VL, Vineland Locomotion.

6.3.2 Desempenho dos classificadores de indivíduos - *Dataset* III

Na tabela 12 são apresentadas todas as medidas de desempenho avaliadas considerando o *dataset* III, com os algoritmos RF, NN e SVM, a inclusão da etapa de normalização e a adição das novas características. O principal resultado foi com o algoritmo RF e a adição das novas características, com AUC média de 0,76, enquanto que sem as novas características a AUC média foi de 0,74. No entanto, essa flutuação ficou dentro do IC, não sendo estatisticamente

diferente. O único fator que interferiu no desempenho foi o algoritmo indutor de classificação utilizado, em que o RF foi superior.

Tabela 12: Desempenho dos classificadores com o *dataset* III.

Indutor	Caract.	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
RF	28	Não	0,74	0,73	0,65	0,57	0,68	0,64
RF	28	Sim	0,74	0,64	0,67	0,55	0,66	0,59
RF	34	Não	0,76	0,68	0,67	0,58	0,67	0,62
RF	34	Sim	0,76	0,64	0,77	0,65	0,72	0,64
NN	28	Não	0,68	0,66	0,59	0,53	0,62	0,59
NN	Relief 28	Não	0,63	0,53	0,65	0,50	0,61	0,51
NN	34	Não	0,65	0,51	0,72	0,55	0,64	0,53
NN	GA 34	Não	0,61	0,50	0,67	0,49	0,61	0,50
SVM	28	Não	0,58	0,56	0,55	0,44	0,56	0,50
SVM	Relief 28	Não	0,55	0,51	0,59	0,45	0,56	0,48
SVM	34	Não	0,58	0,53	0,58	0,45	0,56	0,49
SVM	GA 34	Não	0,53	0,48	0,57	0,41	0,53	0,44

No *dataset* III houve a exclusão dos indivíduos com TEA leve (49 DT e 31 TEA). A condição “28 Características” representa as características basais e “34 Características” representa o conjunto com as novas características. Abreviações: RF, Random Forest; NN, Redes Neurais; SVM, Máquina de Vetores de Suporte; Caract., Características; GA, Algoritmo Genético; Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

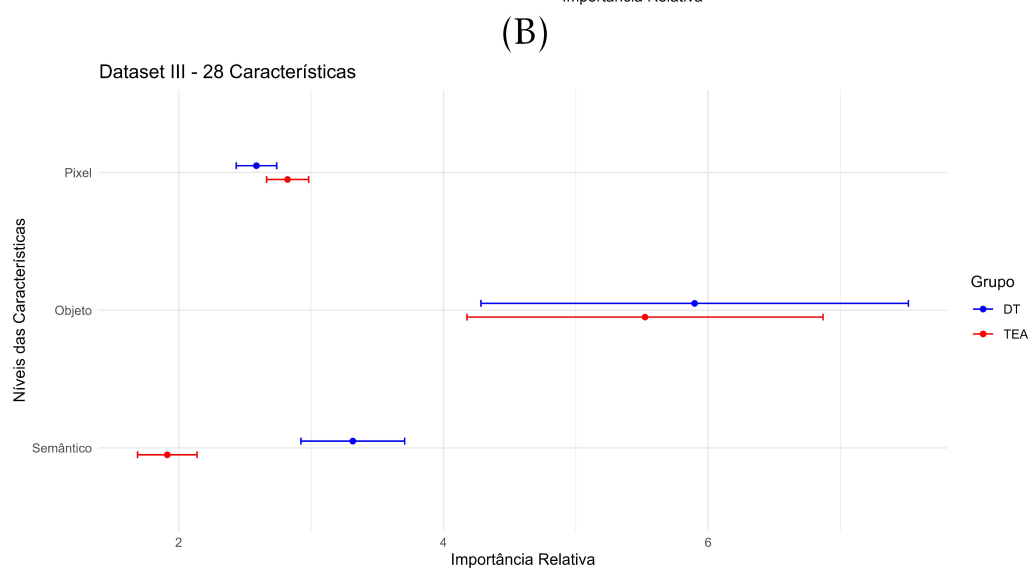
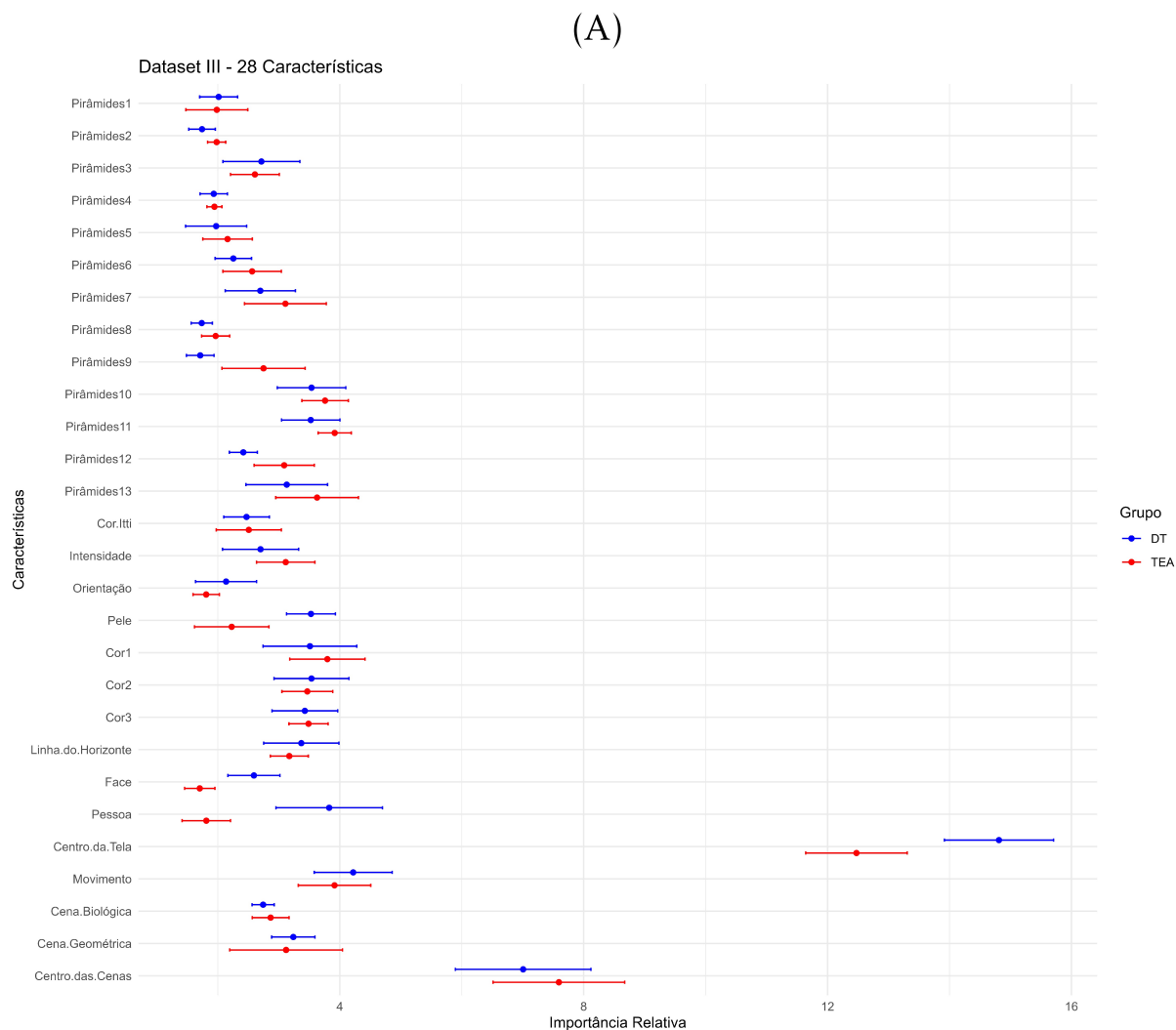
Ao remover os indivíduos com TEA leve, o desempenho médio de classificação foi superior aos do *dataset* I, no entanto, dentro do IC, conforme apresentado na figura 18.

Nossa hipótese para esse experimento foi que indivíduos com TEA leve pudessem ser classificados com maior erro, por apresentar um padrão intermediário entre DT e indivíduos com TEA mais severo. Ainda que a classificação não tenha melhorado, havendo apenas uma tendência, é importante avaliar a importância das características na ausência de TEA leve.

6.3.3 Contribuição das características - Dataset III

Na figura 21-A verificamos a importância relativa da condição com 28 características consideradas para o *dataset* III com o algoritmo RF. As características pirâmide orientável 9, presença de pele, pessoa, face e centro da tela foram diferencialmente importantes entre os grupos. A característica pirâmide orientável 9 apresentou maior importância relativa para o grupo TEA, enquanto que presença de pele, pessoa, face e centro da tela apresentaram menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas para ambos os grupos.

Figura 21: Importância relativa das características - Dataset III: 28 características.



As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

Resultados prévios do grupo, com o mesmo conjunto de 28 características, mostraram que ao considerar a severidade do TEA, as características presença de face e pele na imagem parecem ser menos importantes para o grupo com TEA grave [195]. De maneira semelhante, ao excluir os indivíduos com TEA leve da amostra, essas características passaram a apresentar importância relativa menor para o grupo TEA. Portanto, presença de face e pele parecem ser características discriminatórias relevantes nesse contexto.

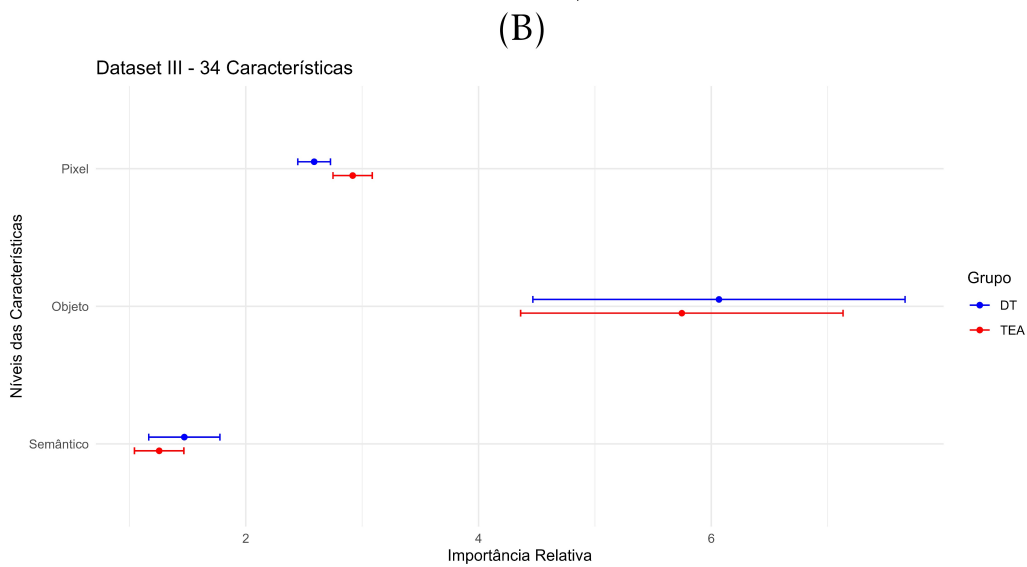
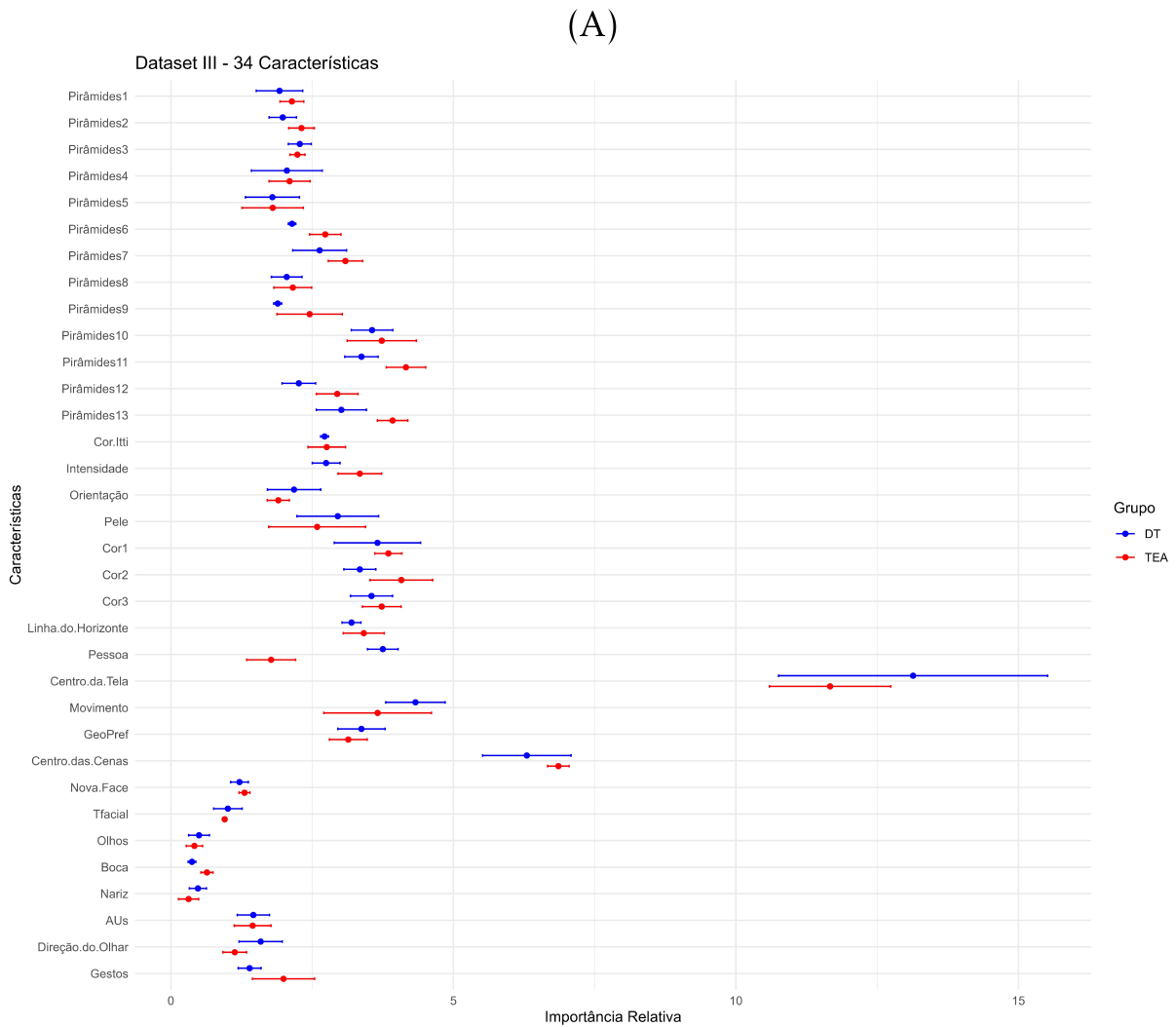
Na figura 21-B notamos que ao agrupar as características por nível de processamento visual, houve uma menor importância relativa das características de nível semântico para o grupo TEA. Para os demais níveis, não houve diferenças entre os grupos. Assim como ocorreu para os *datasets* I e II, as características a nível de objeto contribuíram mais para o processo de classificação.

Na figura 22-A verificamos a importância relativa da condição com 34 características consideradas para o *dataset* III com o algoritmo RF. As características pirâmides orientáveis 6, 11, 12, 13, presença de pessoa e boca foram diferencialmente importantes entre os grupos. As características pirâmides orientáveis 6, 11, 12, 13 e boca apresentaram maior importância relativa para o grupo TEA, enquanto que a presença de pessoa apresentou menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas para ambos os grupos.

Na figura 22-B notamos que ao agrupar as características por nível de processamento visual, houve uma maior importância relativa das características de nível de pixel (baixo nível) para o grupo TEA. Para os demais níveis, não houve diferenças entre os grupos. Assim como nas condições anteriores, as características a nível de objeto contribuíram mais para a construção dos mapas de saliência.

Ao excluir os indivíduos com TEA leve, verificamos que características de baixo nível foram diferencialmente importantes, com destaque para quatro delas (pirâmides orientáveis 6, 11, 12, 13), que foram mais importantes para predizer se o pixel foi ou não fixado por indivíduos TEA. Outra característica interessante de verificar foi Intensidade de Contraste, que embora não tenha apresentado diferença, houve uma tendência de aumento para o grupo TEA. Anteriormente havíamos discutido a percepção de contraste em relação a idade.

Figura 22: Importância relativa das características - Dataset III: 34 características.



As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

6.4 Análises com o *dataset IV* - Exclusão de DI

Para verificar se as condições de DI ou baixo QI impactariam no desempenho de classificação, subdividimos o *dataset I*, excluindo esses indivíduos. Portanto, no *dataset IV* temos 38 indivíduos no grupo TEA.

6.4.1 Caracterização do *dataset IV*

O *dataset IV* está representado na figura 8 (seção 5.2) e a caracterização da amostra está disponível na tabela 13. Os pressupostos de homogeneidade e normalidade foram atendidos em todos os casos apresentados, com exceção da idade e Vineland VL. Em relação às comparações entre os grupos DT e TEA do *dataset IV*, a idade não apresentou diferença ($p = 0,095$ no teste U). O QI foi 7,9% (8,27 unidades) maior no grupo DT ($p = 0,037$ no teste t). A comparação de idade entre os grupos TEA dos *datasets I* e *IV* não apresentou diferença ($p = 0,827$ no teste U). Apesar da exclusão dos indivíduos com baixo QI do *dataset I* para formar o *dataset IV*, não houve diferença estatística para QI, apenas uma tendência ($p = 0,055$ no teste U). Mesmo se o critério normalidade não fosse considerado para a aplicação do teste t, não há diferença para QI ($p = 0,062$). As demais variáveis não apresentaram diferença para o grupo TEA entre os *datasets*.

Tabela 13: Caracterização da amostra - *dataset I* versus *IV*.

Dataset	I		IV	
	DT	TEA	DT	TEA
n Amostral	49	50	49	38
Idade (anos)	6,2 ± 3,2	5,6 ± 3,8	6,2 ± 3,2	5,6 ± 4,1
Masc./Fem.	45 / 4	41 / 9	45 / 4	30 / 8
CARS	-	36,1 ± 5,4	-	35,9 ± 5,9
QI	104,4 ± 15,8	88,4 ± 18,5	104,4 ± 15,8	96,1 ± 4,1
VDL	93,0 ± 14,4	64,5 ± 11,3	93,0 ± 14,4	69,5 ± 8,9
VS	94,8 ± 15,6	62,5 ± 10,7	94,8 ± 15,6	65,1 ± 10,6
VC	91,9 ± 16,3	57,7 ± 15,4	91,9 ± 16,3	63,6 ± 12,3
VL	88,2 ± 11,3	75,6 ± 10,7	88,2 ± 11,3	76,9 ± 6,8

Valores decimais expressos como média ± desvio padrão. Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista; Masc., Sexo Masculino; Fem., Sexo Feminino; CARS, Childhood Autism Rating Scale; QI, Quociente de Inteligência; VC, Vineland Communication; VDL, Vineland Daily Living; VS, Vineland Socialization; VL, Vineland Locomotion.

6.4.2 Desempenho dos classificadores de indivíduos - *Dataset IV*

Na tabela 14 são apresentadas todas as medidas de desempenho avaliadas considerando o *dataset IV*, com os algoritmos RE, NN e SVM, a inclusão da etapa de normalização e a adição

das novas características. O principal resultado foi com o algoritmo RF e a adição das novas características, com AUC média de 0,72, enquanto que sem as novas características a AUC média foi de 0,71, com os ICs interseccionando-se. Assim como nos casos anteriores, o único fator que interferiu no desempenho foi o algoritmo indutor de classificação utilizado, em que o RF foi superior. Ao remover os indivíduos com DI ou baixo QI, o desempenho de classificação foi similar aos do *dataset I*, conforme apresentado na figura 18 (seção 6.2.2).

Tabela 14: Desempenho dos classificadores com o *dataset IV*.

Indutor	Caract.	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
RF	28	Não	0,70	0,71	0,56	0,56	0,62	0,63
RF	28	Sim	0,71	0,54	0,78	0,65	0,67	0,59
RF	34	Não	0,70	0,52	0,78	0,65	0,67	0,58
RF	34	Sim	0,72	0,67	0,65	0,60	0,66	0,63
NN	28	Não	0,59	0,42	0,69	0,53	0,57	0,47
NN	Relief 28	Não	0,61	0,55	0,64	0,55	0,60	0,55
NN	34	Não	0,62	0,48	0,68	0,58	0,59	0,53
NN	Relief 34	Não	0,63	0,57	0,58	0,53	0,58	0,55
SVM	28	Não	0,57	0,79	0,31	0,47	0,52	0,59
SVM	Relief 28	Não	0,55	0,53	0,60	0,51	0,57	0,52
SVM	34	Não	0,57	0,71	0,38	0,47	0,52	0,57
SVM	Relief 34	Não	0,55	0,33	0,77	0,54	0,58	0,41

No *dataset IV* houve a exclusão de indivíduos com DI (49 DT e 38 TEA). A condição “28 Características” representa as características basais e “34 Características” representa o conjunto com as novas características. Abreviações: RF, Random Forest; NN, Redes Neurais; SVM, Máquina de Vetores de Suporte; Caract., Características; Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

Nossa preocupação ao propor esse experimento foi que o fato de haver indivíduos com baixo QI ou DI pudesse interferir na classificação entre DT e TEA, de forma que os classificadores estivessem classificando os indivíduos baseando-se principalmente no QI como variável latente. No entanto, como o desempenho de classificação foi semelhante ao do *dataset I*, o QI parece não ter sido tão importante para a classificação dos indivíduos.

Apesar do TEA ter um amplo espectro de apresentação, a DI tem alta prevalência no TEA. Na Coreia do Sul, estima-se que cerca de 26% dos indivíduos com TEA apresentem DI [196]. Nos Estados Unidos da América, essa estimativa é de 35% [10]. Para a população brasileira ainda não temos esses dados disponíveis. Na nossa amostra, dos 50 indivíduos com TEA, 41 tinham informação de QI, sendo que desses, 12 apresentaram baixo QI. Portanto, 29% dos indivíduos com informação disponível continham baixo QI.

6.4.3 Contribuição das características - Dataset IV

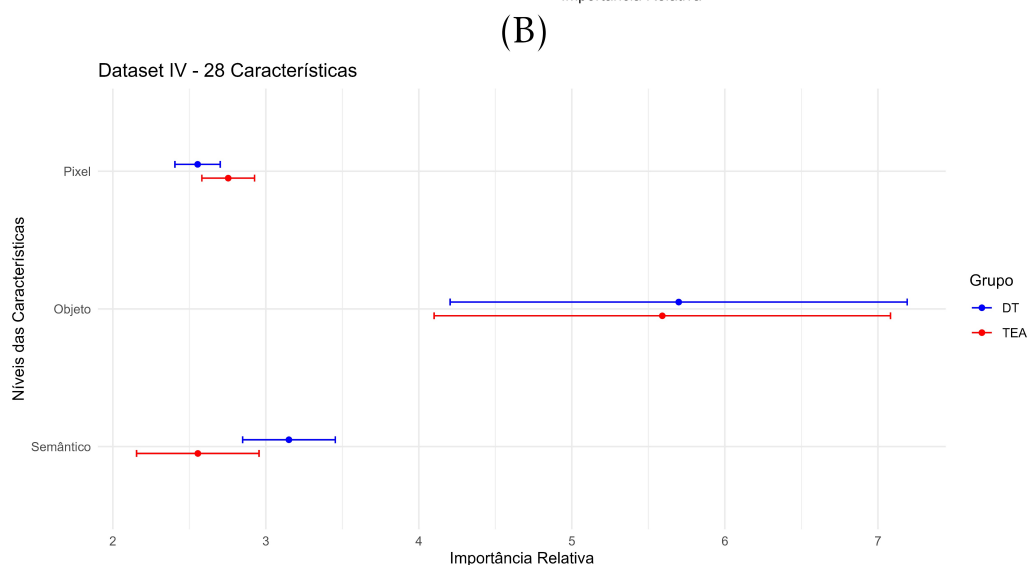
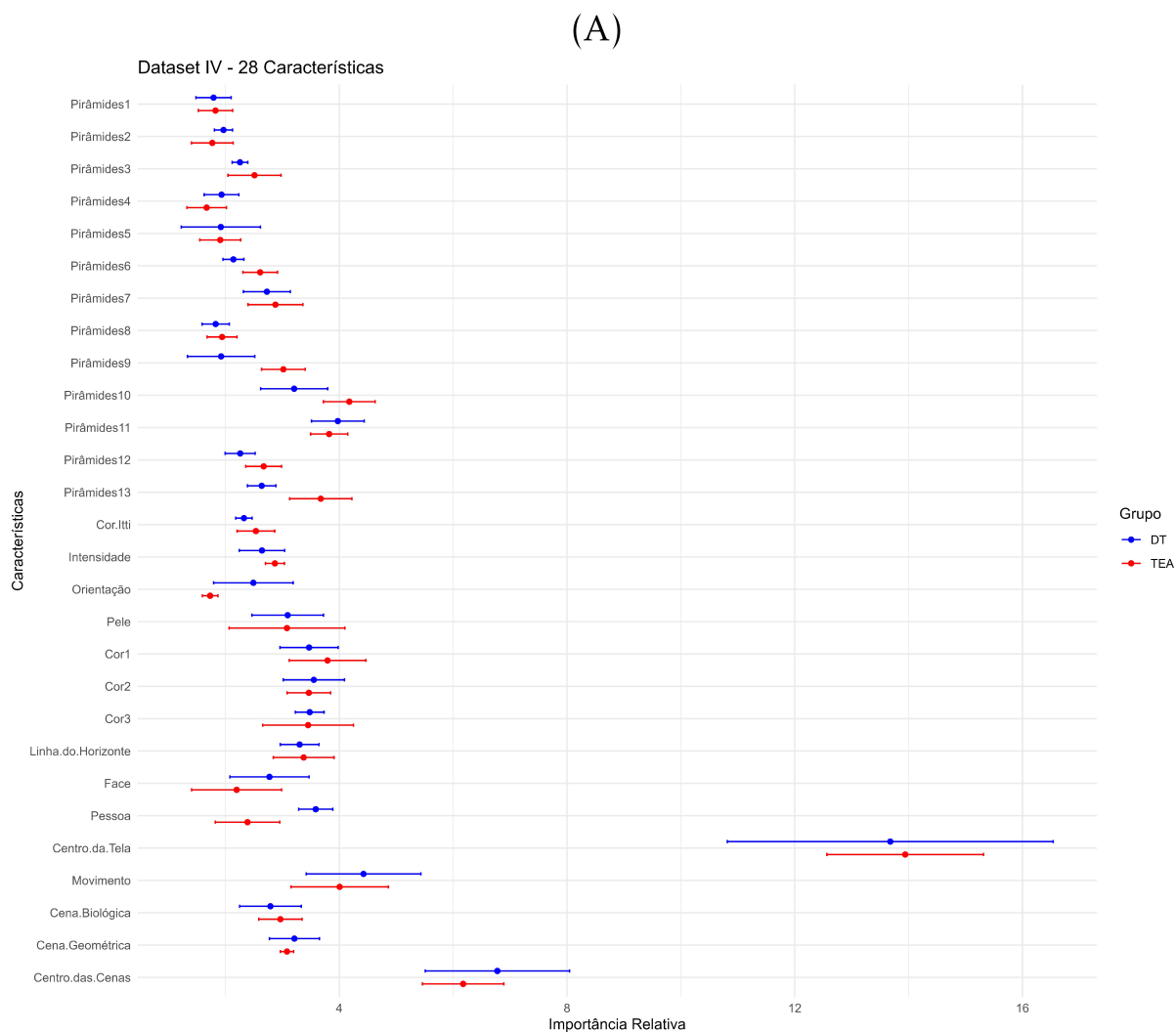
Na figura 23-A verificamos a importância relativa da condição com 28 características consideradas para o *dataset* IV com o algoritmo RF. As características pirâmides orientáveis 9 e 13 e presença de pessoa foram diferencialmente importantes entre os grupos. As pirâmides orientáveis 9 e 13 apresentaram maior importância relativa para o grupo TEA, enquanto que presença de pessoa apresentou menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas.

Na figura 23-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos. Assim como ocorreu para os *datasets* I, II e III, as características a nível de objeto contribuíram mais para o processo de classificação.

Na figura 24-A verificamos a importância relativa da condição com 34 características consideradas para o *dataset* IV com o algoritmo RF. Assim como no caso anterior, as características pirâmides orientáveis 9 e 13 e presença de pessoa foram diferencialmente importantes entre os grupos. As pirâmides orientáveis 9 e 13 apresentaram maior importância relativa para o grupo TEA, enquanto que presença de pessoa apresentou menor importância relativa. As características centro da tela e centro das cenas apresentaram as maiores importâncias relativas para ambos os grupos.

Na figura 24-B notamos que ao agrupar as características por nível de processamento visual, não houve diferença entre os grupos. Assim como nos casos anteriores, as características a nível de objeto contribuíram mais para o processo de classificação.

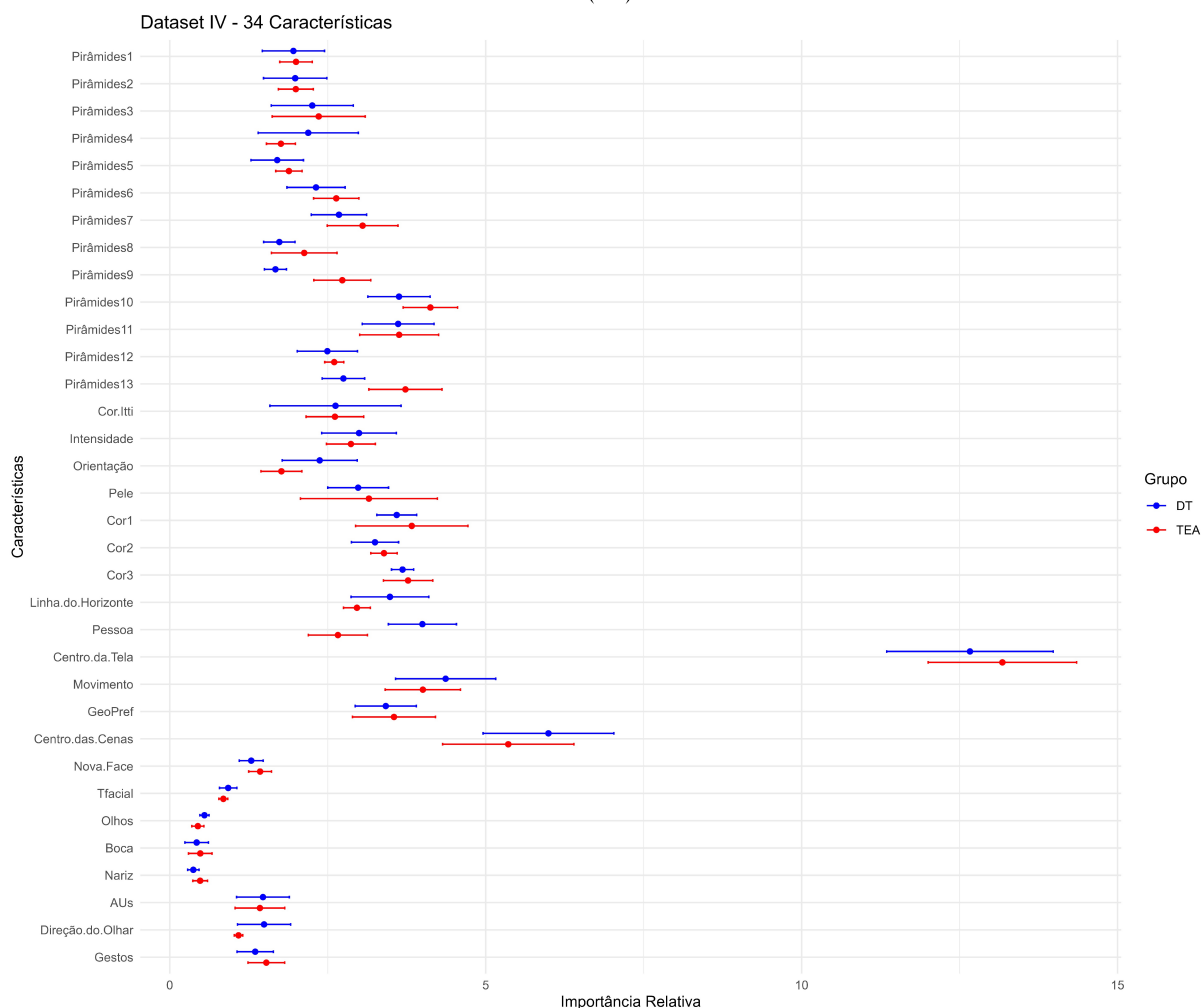
Figura 23: Importância relativa das características - Dataset IV: 28 características.



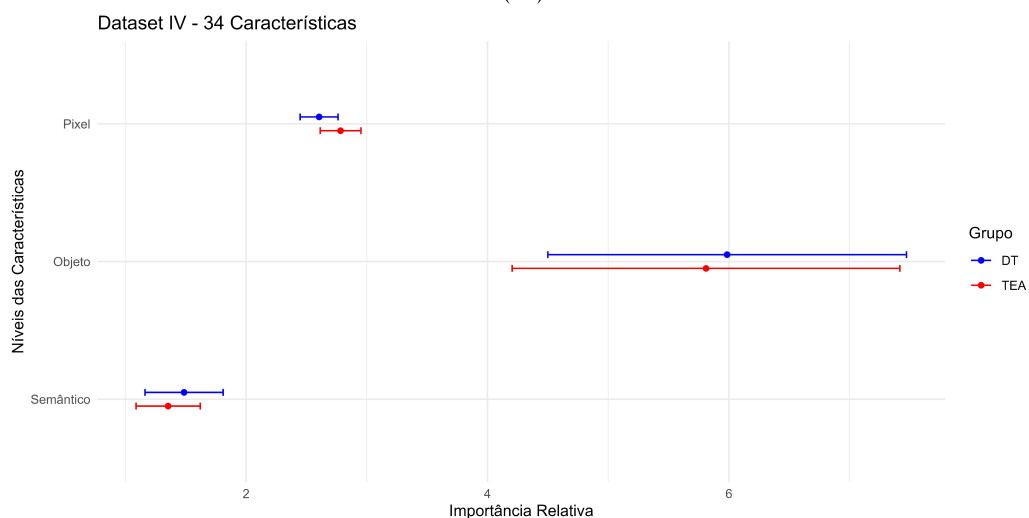
As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

Figura 24: Importância relativa das características - Dataset IV: 34 características.

(A)



(B)

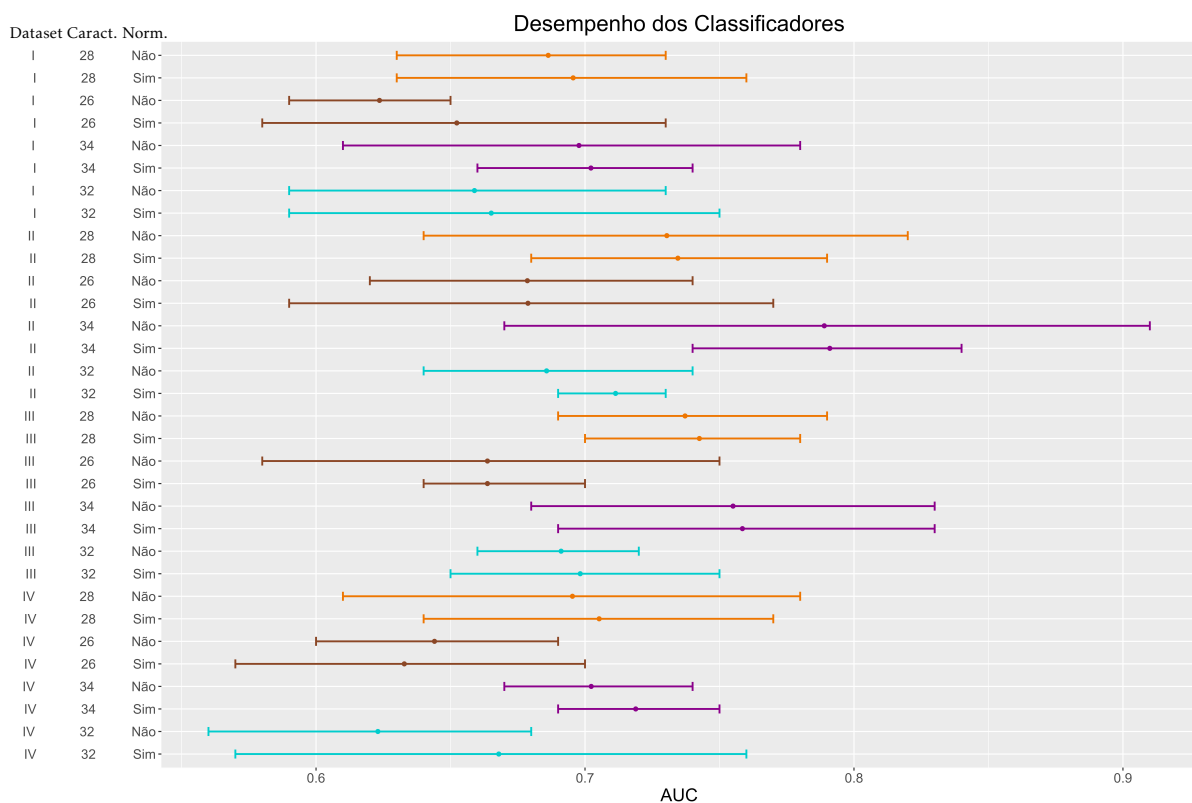


As barras representam o intervalo de confiança de 95% e o ponto central a média. A: Comparação da importância relativa das características para cada grupo. B: Comparação da importância relativa das características para cada grupo conforme o nível de processamento visual. Abreviações: AUs, Unidades de Ação para Expressões Faciais; DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

6.5 Análises com a exclusão das características de centro

Conforme apresentado na figura 25, ao eliminar as características de centro (tela e cena) houve uma diminuição do desempenho de classificação, sugerindo que elas foram fundamentais para a classificação. Embora esse padrão tenha se apresentado para todos os *datasets*, apenas para o *dataset* II na comparação entre os conjuntos 34 e 32 características, essa variação foi estatisticamente significativa. As demais medidas de desempenho são apresentadas na tabela 15.

Figura 25: Desempenho dos classificadores com a adição das novas características para os *datasets* I, II, III e IV.



O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA). O *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). No *dataset* III houve a exclusão dos indivíduos com TEA leve (49 DT e 31 TEA). No *dataset* IV houve a exclusão de indivíduos com DI (49 DT e 38 TEA). A condição “28 Características” representa as características basais. Abreviações: *Caract.*, Características; *Norm.*, Normalização; *AUC*, Área Sob a Curva ROC.

A partir dos nossos resultados, podemos responder a uma pergunta secundária: Em um paradigma com duas cenas apresentadas simultaneamente, o que é mais importante na classificação de DT e TEA, o centro da tela ou os centros das cenas? Os resultados foram consistentes em todos os cenários abordados (*datasets* I, II, III e IV), sugerindo que o centro da tela tem uma importância relativa maior que o centro das cenas, para ambos os grupos.

É amplamente evidenciado na literatura que o viés central é importante para prever fixa-

Tabela 15: Desempenho dos classificadores com a exclusão das características de centro para os *datasets* I, II, III e IV.

Indutor	Dataset	n (TEA+DT)	Caract.	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
RF	I	50+49	28	Não	0,69	0,57	0,69	0,67	0,63	0,62
RF	I	50+49	28	Sim	0,70	0,56	0,73	0,69	0,64	0,62
RF	I	50+49	26	Não	0,62	0,59	0,59	0,60	0,59	0,60
RF	I	50+49	26	Sim	0,65	0,57	0,60	0,60	0,58	0,58
RF	I	50+49	34	Não	0,70	0,54	0,75	0,69	0,64	0,61
RF	I	50+49	34	Sim	0,70	0,80	0,50	0,63	0,65	0,70
RF	I	50+49	32	Não	0,66	0,54	0,71	0,65	0,62	0,59
RF	I	50+49	32	Sim	0,67	0,57	0,69	0,66	0,63	0,61
RF	II	28+33	28	Não	0,73	0,76	0,57	0,61	0,66	0,67
RF	II	28+33	28	Sim	0,73	0,56	0,79	0,69	0,68	0,62
RF	II	28+33	26	Não	0,68	0,60	0,65	0,60	0,63	0,60
RF	II	28+33	26	Sim	0,68	0,59	0,65	0,60	0,62	0,59
RF	II	28+33	34	Não	0,79	0,80	0,66	0,67	0,72	0,73
RF	II	28+33	34	Sim	0,79	0,76	0,70	0,68	0,72	0,72
RF	II	28+33	32	Não	0,69	0,54	0,73	0,64	0,64	0,58
RF	II	28+33	32	Sim	0,71	0,59	0,75	0,66	0,67	0,62
RF	III	31+49	28	Não	0,74	0,73	0,65	0,57	0,68	0,64
RF	III	31+49	28	Sim	0,74	0,64	0,67	0,55	0,66	0,59
RF	III	31+49	26	Não	0,66	0,55	0,71	0,55	0,65	0,55
RF	III	31+49	26	Sim	0,66	0,55	0,71	0,55	0,65	0,55
RF	III	31+49	34	Não	0,76	0,68	0,67	0,58	0,67	0,62
RF	III	31+49	34	Sim	0,76	0,64	0,77	0,65	0,72	0,64
RF	III	31+49	32	Não	0,69	0,59	0,71	0,57	0,66	0,58
RF	III	31+49	32	Sim	0,70	0,69	0,58	0,51	0,62	0,59
RF	IV	38+49	28	Não	0,70	0,71	0,56	0,56	0,62	0,63
RF	IV	38+49	28	Sim	0,71	0,54	0,78	0,65	0,67	0,59
RF	IV	38+49	26	Não	0,64	0,53	0,67	0,56	0,61	0,54
RF	IV	38+49	26	Sim	0,63	0,48	0,71	0,57	0,61	0,52
RF	IV	38+49	34	Não	0,70	0,52	0,78	0,65	0,67	0,58
RF	IV	38+49	34	Sim	0,72	0,67	0,65	0,60	0,66	0,63
RF	IV	38+49	32	Não	0,62	0,53	0,61	0,53	0,57	0,53
RF	IV	38+49	32	Sim	0,67	0,60	0,69	0,61	0,65	0,61

O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA). O *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). No *dataset* III houve a exclusão dos indivíduos com TEA leve (49 DT e 31 TEA). No *dataset* IV houve a exclusão de indivíduos com DI (49 DT e 38 TEA). A condição “28 Características” representa as características basais. Abreviações: RF, Random Forest; NN, Redes Neurais; SVM, Máquina de Vetores de Suporte; n, número de observações; Caract., Características; Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

ções, tanto no DT [197–199] quanto no TEA [85, 93]. Porém, no paradigma aqui utilizado, a tela é dividida ao meio, sendo exibidas duas cenas simultaneamente. Os resultados apresentados mostram que, além do centro da tela, o centro das cenas também é importante para ambos os grupos. Eles não permitem conclusões sobre haver diferença da importância do viés de centro para os grupos, verificou-se diferença em apenas um caso analisado, na condição considerando o *dataset* III com as 28 características basais, conforme a figura 21. Neste caso, o viés de centro foi mais importante para predizer se um pixel foi ou não fixado para o grupo DT. Na literatura, existem estudos em ambos os sentidos, Wang *et al.* [85] e Kiani *et al.* [93], evidência de que o viés de centro é mais significativo para o grupo TEA. Por outro lado, Nebout *et al.* [83], veri-

ficou que o viés de centro é muito menos importante para pessoas com TEA do que para DT. Portanto, ainda não há consenso sobre o viés de centro entre os grupos.

É importante mencionar que existem algumas métricas que penalizam o viés central [200, 201]. Aqui não houve tal penalidade, sendo as características consideradas indiscriminadamente, independentemente de sua natureza.

Antes da remoção das características de centro, a contribuição das características em nível de objeto notadamente tinham maior importância. Ao remover as duas características com maior contribuição, centro da cena e centro da tela, as demais características a nível de objeto continuaram com maior importância para a predição. Isso enfatiza ainda mais a importância delas para a construção de mapas de saliência, seja considerando indivíduos DT ou TEA.

6.6 Principais achados

A replicação ocorreu de maneira satisfatória, sugerindo a robustez da abordagem prévia e a viabilidade das abordagens baseadas em MAVs para auxiliar no diagnóstico de TEA. Porém, a troca de estímulo visual e amostras teve um impacto negativo no desempenho dos classificadores. Embora outros fatores devam ser considerados, a diferença de desempenho entre o estudo anterior e a replicação pode ser devido a diferença de idade dos indivíduos. O resultado com o classificador pré-treinado sugere que a mudança de estímulo entre as etapas de treinamento e teste influencia significativamente o desempenho de classificação. Portanto, ao aplicar o classificador treinado em dados baseados em outro estímulo, deve-se haver um novo treinamento.

Após a replicação, alteramos o algoritmo indutor de classificação. O algoritmo RF apresentou melhores desempenhos em todos os cenários testados, com a vantagem adicional de permitir a comparação de importância atribuída a cada uma das características consideradas, por meio do erro OOB.

As características com importância relativa diferentes entre os grupos foram apresentadas e discutidas ao longo da seção 6. A fim de facilitar a visualização, esses achados foram resumidos na tabela 16. De maneira geral, as características a nível de objeto e semântico apresentaram proporcionalmente mais diferenças entre os grupos DT e TEA, ainda que características a nível de pixel sejam mais numerosas que as outras. Um terço das características semânticas apresentou diferença, contendo menor importância relativa para TEA. Nesse contexto, a característica presença de pessoa foi notadamente a mais frequente e, conseqüentemente, apre-

sentou maior consistência de resultados, sugerindo um maior poder preditivo para dizer se um indivíduo com TEA fixará ou não um determinado pixel, em comparação a indivíduos DT.

Tabela 16: Características diferentemente importantes entre os grupos.

Dataset	Caract.	Dif. Importantes Entre os Grupos	Prop. - Pixel	Prop. - Objeto	Prop. - Sem.
I	28	Pessoa	0	0	0,33
I	26	Pir. 6, 9 e 11, Pele, Hor., Face, Pessoa	0,16	0,25	1
I	34	Pir. 12, Pessoa e Movimento	0,05	0,20	0,10
I	32	Pessoa	0	0	0,10
II	28	Pir. 9, Pessoa e Movimento	0,05	0,17	0,33
II	26	Pir. 4, Orientação e Pessoa	0,11	0	0,33
II	34	Pessoa, T-Facial e Movimento	0	0,20	0,20
II	32	Pele, Pessoa, Movimento e T-Facial	0	0,33	0,30
III	28	Pir. 9, Pele, Face, Pessoa e Centro	0,05	0,17	1
III	26	Pir. 9, Pele, Face e Pessoa	0,05	0	1
III	34	Pir. 6, 11, 12 e 13, Pessoa e Boca	0,21	0	0,20
III	32	Pir. 6 e 9, Pele, Pessoa e Nova Face	0,11	0	0,30
IV	28	Pir. 9 e 13 e Pessoa	0,11	0	0,33
IV	26	Pir. 2 e 9, Pele e Pessoa	0,11	0	0,67
IV	34	Pir. 9 e 13 e Pessoa	0,11	0	0,10
IV	32	Pir. 8 e 9, Pele, Pessoa e Nova Face	0,11	0	0,30

O *dataset* I refere-se a todos os indivíduos (49 DT e 50 TEA). O *dataset* II refere-se aos indivíduos maiores de 4 anos (33 DT e 28 TEA). No *dataset* III houve a exclusão dos indivíduos com TEA leve (49 DT e 31 TEA). No *dataset* IV houve a exclusão de indivíduos com DI (49 DT e 38 TEA). A condição “28 Características” representa as características basais. Abreviações: Caract., Características; Dif., Diferentemente; Prop., Proporção de características diferentemente importantes; Sem., Semânticas; Pir., Pirâmides Orientáveis; Hor., Horizonte.

Características de centro foram as mais importantes e sua exclusão impactou negativamente na classificação e dentre elas, centro da tela contribuiu mais do que centro da cena para ambos os grupos. Características semânticas apresentaram importância relativa menor que as demais e com menor importância para TEA, em comparação a DT.

A exclusão de indivíduos com TEA leve ou baixo QI não impactou no desempenho de classificação. Por outro lado, a exclusão de indivíduos menores de 4 anos melhorou o desempenho de classificação, sugerindo uma influência da variável idade.

6.7 Limitações do estudo

Algumas coletas ocorreram em ambientes não controlados (escolas), podendo haver ruídos, variações de luminosidade e na sala de coleta.

Apesar desse estudo ter um tamanho amostral maior que a maioria dos trabalhos apresentados no levantamento descrito na tabela 2, recomenda-se que os achados apresentados sejam replicados com uma amostra maior. Dado o tamanho amostral, não foi possível controlar possíveis comorbidades, testar outras faixas etárias e os níveis de severidade do TEA.

Indivíduos do sexo feminino estão sub-representados, correspondendo a apenas 13% da

amostra e conforme evidenciado por Harrop *et al.* [202], existem diferenças na atenção visual no TEA entre os sexos. Portanto, limitando generalizações para os achados deste estudo.

Uma vez que o estímulo é dividido ao meio, onde duas cenas são apresentadas simultaneamente, seria relevante equalizar a taxa de movimento entre as cenas sociais e não sociais. Desta forma, proporcionar maior similaridade psicofísica entre as cenas para obter uma comparação mais adequada. Portanto recomenda-se um pareamento da taxa de movimento e equivalência de luminância entre as cenas. Outra extensão válida para trabalhos futuros seria estudar formas alternativas para a composição dos estímulos, com divisão horizontal entre as cenas e avaliar diferentes porcentagens da área ocupada por cada uma das cenas.

Todos os resultados foram obtidos utilizando o mesmo limiar de movimento para agregação de *frames* utilizado no estudo base (0,33). No entanto, um estudo recente do grupo [190] sugere variações nesse limiar podem influenciar no desempenho de classificação.

Por fim, as características semânticas avaliadas estão limitadas ao estímulo utilizado, desta forma, existem outras como texto e presença de objetos (com exceção de um brinquedo que aparece em apenas uma cena) que não estavam presentes no vídeo.

7 CONCLUSÕES

O presente projeto aborda uma temática multidisciplinar, contribuindo com um tema relacionado à área Biológica com ferramentas da Computação. A hipótese principal foi que características semânticas, não só são importantes, mas contribuem para melhorar o desempenho de classificação entre DT e TEA, a qual foi majoritariamente refutada no contexto avaliado.

A rejeição da hipótese foi fundamentada pela avaliação de desempenho nos quatro cenários (com os *datasets* I, II, III e IV), em nenhum dos casos o desempenho de classificação com a adição das novas características semânticas foi estatisticamente superior à condição basal, embora tenha apresentado tendência de melhora em todos os casos (conforme representado na figura 18, seção 6.2.2).

De maneira adjacente à hipótese principal, verificamos que:

- a algoritmo RF apresentou melhores desempenhos de classificação;
- a troca de estímulo visual e amostras teve um impacto negativo no desempenho dos classificadores;
- indivíduos mais velhos foram melhor classificados;
- características semânticas apresentaram importância relativa menor que as demais, porém com maior importância para indivíduos DT;
- a característica pessoa foi a característica que mais divergiu entre os grupos, sendo menos importante no TEA;
- características de centro foram as mais importantes e sua exclusão impactou negativamente na classificação.

Além das contribuições mencionadas anteriormente, existem ainda contribuições econômicas e sociais, visto que o diagnóstico assistido por computador pode ampliar a capacidade de atendimento do ambulatório PROTEA e em lugares onde não há profissionais especializados, proporcionando acessibilidade.

8 PUBLICAÇÕES

Durante o período do doutorado foram publicados cinco artigos relacionados à Bioinformática e o sexto está em desenvolvimento.

- Jessica Oliveira; **Felipe Franco**; Mirian Revers, Andréia Silva; Joana Portolese; Helena Brentani; Ariane Machado-Lima; Fátima Nunes. Computer-aided autism diagnosis based on visual attention models using eye tracking. **Scientific Reports**, 11(1):1–11, 2021. DOI: [10.1038/s41598-021-89023-8](https://doi.org/10.1038/s41598-021-89023-8).
- Mirian Revers; Jessica Oliveira; **Felipe Franco**; Joana Portolese; Thiago Cardoso; Andréia Silva; Ariane Machado-Lima; Fátima Nunes; Helena Brentani. Classification of autism spectrum disorder severity using eye tracking data based on visual attention model. In **34th International Symposium on CBMS**, 142–147, IEEE, 2021. DOI: [10.1109/CBMS52027.2021.00062](https://doi.org/10.1109/CBMS52027.2021.00062).
- Thiago Cardoso; Gabriel Michelassi; **Felipe Franco**; Fernando Sumiya; Joana Portolese; Helena Brentani; Ariane Machado-Lima; Fátima Nunes. Autism spectrum disorder diagnosis based on trajectories of eye tracking data. In **34th International Symposium on CBMS**, 50–55, IEEE, 2021. DOI: [10.1109/CBMS52027.2021.00016](https://doi.org/10.1109/CBMS52027.2021.00016).
- Anita Brito; **Felipe Franco**; Helena Brentani; Patrícia Beltrão-Braga. Assessment of vulnerability dimensions considering Family History and Environmental Interplay in Autism Spectrum Disorder. **BMC Psychiatry**, 23 (254): 1-13, 2023. DOI: [10.1186/s12888-023-04747-3](https://doi.org/10.1186/s12888-023-04747-3).
- **Felipe Franco**; Jessica Oliveira; Joana Portolese; Fernando Sumiya; Andréia Silva, Ariane Machado-Lima; Fátima Nunes; Helena Brentani. Computer-aided autism diagnosis approach based on visual attention models using eye tracking: a replication study and improvement proposal. **BMC Medical Informatics and Decision Making**, 23, (285), 1-9, 2023. DOI: [10.1186/s12911-023-02389-9](https://doi.org/10.1186/s12911-023-02389-9).
- **Felipe Franco et. al**; The role of semantic features in computer-aided autism diagnosis approach based on visual attention models. Situação: Em desenvolvimento.

REFERÊNCIAS

- [1] Jessica S Oliveira, Felipe O Franco, Mirian C Revers, Andréia F Silva, Joana Portolese, Helena Brentani, Ariane Machado-Lima, and Fátima LS Nunes. Computer-aided autism diagnosis based on visual attention models using eye tracking. *Scientific reports*, 11(1):1–11, 2021. doi: 10.1038/s41598-021-89023-8.
- [2] American Psychiatric Association et al. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Artmed Editora, 2014.
- [3] Ortal Slobodin and Michael Davidovitch. Gender differences in objective and subjective measures of ADHD among clinic-referred children. *Frontiers in Human Neuroscience*, 13:1–14, 2019. ISSN 1662-5161. doi: 10.3389/fnhum.2019.00441.
- [4] Helena Brentani, Cristiane Silvestre de Paula, Daniela Bordini, Deborah Rolim, Fabio Sato, Joana Portolese, Maria Clara Pacifico, and James McCracken. Autism spectrum disorders: an overview on diagnosis and treatment. *Brazilian Journal of Psychiatry*, 35:62–72, 2013. doi: 10.1590/1516-4446-2013-S104.
- [5] S Thomas, T Hinkley, LM Barnett, T May, and N Rinehart. Young children with ASD participate in the same level of physical activity as children without ASD: Implications for early intervention to maintain good health. *Journal of autism and developmental disorders*, 49(8):3278–3289, 2019.
- [6] Morgan Frost-Karlsson, Martyna Alexandra Galazka, Christopher Gillberg, Carina Gillberg, Carmela Miniscalco, Eva Billstedt, Nouchine Hadjikhani, and Jakob Åsberg Johnels. Social scene perception in autism spectrum disorder: an eye-tracking and pupillometric study. *Journal of Clinical and Experimental Neuropsychology*, 41(10):1024–1032, 2019. doi: 10.1080/13803395.2019.1646214.
- [7] Huiyu Duan, Xionguo Min, Yi Fang, Lei Fan, Xiaokang Yang, and Guangtao Zhai. Visual attention analysis and prediction on human faces for children with autism spectrum disorder. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(3):1–23, 2019. doi: 10.1145/3337066.
- [8] Julia Parish-Morris, Ashley A Pallathra, Emily Ferguson, Brenna B Maddox, Alison Pomykacz, Leat S Perez, Leila Bateman, Juhi Pandey, Robert T Schultz, and Edward S Brodtkin. Adaptation to different communicative contexts: an eye tracking study of autistic adults. *Journal of neurodevelopmental disorders*, 11(5):1–10, 2019. doi: 10.1186/s11689-019-9265-1.
- [9] Cristiane S Paula, Sabrina H Ribeiro, Eric Fombonne, and Marcos T Mercadante. Brief report: prevalence of pervasive developmental disorder in brazil: a pilot study. *Journal of autism and developmental disorders*, 41(12):1738–1742, 2011. doi: 10.1007/s10803-011-1200-6.
- [10] Matthew J Maenner, Kelly A Shaw, Amanda V Bakian, Deborah A Bilder, Maureen S Durkin, Amy Esler, Sarah M Furnier, Libby Hallas, Jennifer Hall-Lande, Allison Hudson, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, 70(11):1, 2021. doi: 10.15585/mmwr.ss7011a1.
- [11] Jinan Zeidan, Eric Fombonne, Julie Scora, Alaa Ibrahim, Maureen S Durkin, Shekhar Saxena, Afiqah Yusuf, Andy Shih, and Mayada Elsabbagh. Global prevalence of autism: a systematic review update. *Autism Research*, 15(5):778–790, 2022. doi: 10.1002/aur.2696.
- [12] Neal Halfon, Christopher B Forrest, Richard M Lerner, and Elaine M Faustman. *Handbook of life course health development*. Springer Nature, 2018. ISBN 978-3-319-47143-3. doi: 10.1007/978-3-319-47143-3.
- [13] Andrew Pickles, Deborah K Anderson, and Catherine Lord. Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry*, 55(12):1354–1362, 2014. doi: 10.1111/jcpp.12269.

- [14] Antonio Napolitano, Sara Schiavi, Piergiorgio La Rosa, Maria Camilla Rossi-Espagnet, Sara Pettillo, Francesca Bottino, Emanuela Tagliente, Daniela Longo, Elisabetta Lupi, Laura Casula, et al. Sex differences in autism spectrum disorder: Diagnostic, neurobiological, and behavioral features. *Frontiers in Psychiatry*, 13, 2022. doi: 10.3389/fpsy.2022.889636.
- [15] World Health Organization. ICD-11: International Classification of Diseases 11th Revision, 2023. URL <https://icd.who.int/en>.
- [16] Catherine Lord, M Rutter, P DiLavore, S Risi, K Gotham, S Bishop, et al. Autism diagnostic observation schedule–2nd edition (ADOS-2). Los Angeles, CA: Western Psychological Corporation, 284, 2012.
- [17] Catherine Lord, Michael Rutter, and Ann Le Couteur. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5):659–685, 1994.
- [18] Dennis P Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F DeLuca. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PloS one*, 7(8):e43855, 2012.
- [19] Alessandra Marques Pereira, Rudimar dos Santos Riesgo, and Mario Bernardes Wagner. Childhood autism: translation and validation of the childhood autism rating scale for use in Brazil. *Jornal de Pediatria*, 84(6):487–494, 2008. doi: 10.2223/JPED.1828.
- [20] Cameron R Pepperdine and Adam W McCrimmon. Test Review: Vineland Adaptive Behavior Scales, (Vineland-3) by Sparrow, SS, Cicchetti, DV, & Saulnier, CA. *Canadian Journal of School Psychology*, 33:157–163, 2018. doi: 10.1177/0829573517733845.
- [21] Thomas M Achenbach and Leslie Rescorla. Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment. Aseba Burlington, VT, 2001.
- [22] Emad Bataineh, Mohamed Basel Almourad, Farhi Marir, and Joana Stocker. Visual attention toward socially rich context information for autism spectrum disorder (ASD) and normal developing children: an eye tracking study. In *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia*, pages 61–65, 2018. doi: 10.1145/3282353.3282856.
- [23] Konstantinos-Filippos Kollias, Christine K Syriopoulou-Delli, Panagiotis Sarigiannidis, and George F Fragulis. The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review. *Electronics*, 10(23):2982, 2021. doi: 10.3390/electronics10232982.
- [24] George D Farmer, Paula Smith, Simon Baron-Cohen, and William J Skylark. The effect of autism on information sampling during decision-making: An eye-tracking study. *Judgment and Decision making*, 16(3):614–637, 2021. doi: 10.1017/S1930297500007762.
- [25] Mohamed Basel Almourad, Emad Bataineh, Joana Stocker, and Farhi Marir. Analyzing the behavior of autistic and normal developing children using eye tracking data. In *International Conference on Kansei Engineering & Emotion Research*, pages 340–349. Springer, 2018. doi: 10.1007/978-981-10-8612-0_36.
- [26] Zhaowei Li, Peiyuan Guo, and Chen Song. A review of main eye movement tracking methods. In *Journal of Physics: Conference Series*, volume 1802, page 042066. IOP Publishing, 2021. doi: 10.1088/1742-6596/1802/4/042066.
- [27] Terje Falck-Ytter, Sven Bölte, and Gustaf Gredebäck. Eye tracking in early autism research. *Journal of neurodevelopmental disorders*, 5(28):1–13, 2013. doi: 10.1186/1866-1955-5-28.
- [28] Bhanuka Mahanama, Yasith Jayawardana, Sundararaman Rengarajan, Gavindya Jayawardana, Leanne Chukoskie, Joseph Snider, and Sampath Jayarathna. Eye movement and pupil measures: A review. *frontiers in Computer Science*, 3:733531, 2022. doi: 10.3389/fcomp.2021.733531.

- [29] Florian Hauser, Jürgen Mottok, and Hans Gruber. Eye tracking metrics in software engineering. In *Proceedings of the 3rd European Conference of Software Engineering Education*, pages 39–44, 2018. doi: 10.1145/3209087.3209092.
- [30] Kami Koldewyn, Sarah Weigelt, Nancy Kanwisher, and Yuhong Jiang. Multiple object tracking in autism spectrum disorders. *Journal of autism and developmental disorders*, 43(6):1394–1405, 2013. doi: 10.1007/s10803-012-1694-6.
- [31] Kris Evers, Lee de Wit, Ruth Van der Hallen, Birgitt Haesen, Jean Steyaert, Ilse Noens, and Johan Wagemans. Brief report: reduced grouping interference in children with ASD: evidence from a multiple object tracking task. *Journal of autism and developmental disorders*, 44(7):1779–1787, 2014. doi: 10.1007/s10803-013-2031-4.
- [32] MD Rutherford, Eric Richards, Vanessa Moldes, and Allison Sekuler. Evidence of a divided-attention advantage in autism. *Cognitive Neuropsychology*, 24(5):505–515, 2007. doi: 10.1080/02643290701508224.
- [33] Qiandong Wang, Sio Pan Hoi, Yuyin Wang, Ci Song, Tianbi Li, Cheuk Man Lam, Fang Fang, and Li Yi. Out of mind, out of sight? investigating abnormal face scanning in autism spectrum disorder using gaze-contingent paradigm. *Developmental science*, 23(1):e12856, 2020. doi: 10.1111/desc.12856.
- [34] Soichiro Matsuda, Takahide Omori, Joseph P McCleery, and Junichi Yamamoto. Comparing reinforcement values of facial expressions: An eye-tracking study. *The Psychological Record*, 69(3):393–400, 2019. doi: 10.1007/s40732-019-00330-z.
- [35] Jacqueline A Riddiford, Peter G Enticott, Alex Lavale, and Caroline Gurvich. Gaze and social functioning associations in autism spectrum disorder: A systematic review and meta-analysis. *Autism Research*, 15(8):1380–1446, 2022. doi: 10.1002/aur.2729.
- [36] Juan Kou, Jiao Le, Meina Fu, Chunmei Lan, Zhuo Chen, Qin Li, Weihua Zhao, Lei Xu, Benjamin Becker, and Keith M Kendrick. Comparison of three different eye-tracking tasks for distinguishing autistic from typically developing children and autistic symptom severity. *Autism Research*, 12(10):1529–1540, 2019. doi: 10.1002/aur.2174.
- [37] Mirian de Cesaro Revers Biasão. Classificação da gravidade do transtorno do espectro autista baseada no padrão de rastreamento do olhar. *Dissertação: Universidade de São Paulo*, 2019.
- [38] Elizabeth C Bacon, Adrienne Moore, Quimby Lee, Cynthia Carter Barnes, Eric Courchesne, and Karen Pierce. Identifying prognostic markers in autism spectrum disorder using eye tracking. *Autism*, 24(3):658–669, 2020. doi: 10.1177/1362361319878578.
- [39] Barbora Siposova and Malinda Carpenter. A new look at joint attention and common knowledge. *Cognition*, 189:260–274, 2019. doi: 10.1016/j.cognition.2019.03.019.
- [40] Ann M Mastergeorge, Chanaka Kahathuduwa, and Jessica Blume. Eye-tracking in infants and young children at risk for autism spectrum disorder: A systematic review of visual stimuli in experimental paradigms. *Journal of Autism and Developmental Disorders*, 51(8):2578–2599, 2021. doi: 10.1007/s10803-020-04731-w.
- [41] Tianbi Li, Yewei Li, Yixiao Hu, Yuyin Wang, Cheuk Man Lam, Wei Ni, Xueqin Wang, and Li Yi. Heterogeneity of Visual Preferences for Biological and Repetitive Movements in Children With Autism Spectrum Disorder. *Autism Research*, pages 1–10, 2021. doi: 10.1002/aur.2366.
- [42] Selda Ozdemir, Isik Akin-Bulbul, Ibrahim Kok, and Suat Ozdemir. Development of a visual attention based decision support system for autism spectrum disorder screening. *International Journal of Psychophysiology*, 173:69–81, 2022. doi: 10.1016/j.ijpsycho.2022.01.004.
- [43] Pär Nyström, Emilia Thorup, Sven Bölte, and Terje Falck-Ytter. Joint attention in infancy and the emergence of autism. *Biological psychiatry*, 86(8):631–638, 2019. doi: 10.1016/j.biopsycho.2019.05.006.

- [44] David Ambrose, Diane E MacKenzie, Parisa Ghanouni, and Heather F Neyedli. Investigating joint attention in a guided interaction between a child with asd and therapists: A pilot eye-tracking study. *British Journal of Occupational Therapy*, 84(10):637–646, 2021. doi: 10.1177/0308022620963727.
- [45] Nathan Caruana, Heidi Stieglitz Ham, Jon Brock, Alexandra Woolgar, Nadine Kloth, Romina Palermo, and Genevieve McArthur. Joint attention difficulties in autistic adults: an interactive eye-tracking study. *Autism*, 22(4):502–512, 2018. doi: 10.1177/1362361316676204.
- [46] Randall L Morrison and Alan S Bellack. The role of social perception in social skill. *Behavior therapy*, 12(1):69–79, 1981. doi: 10.1016/S0005-7894(81)80107-4.
- [47] Maria Chiara Pino, Roberto Vagnetti, Francesco Masedu, Margherita Attanasio, Sergio Tiberti, Marco Valenti, and Monica Mazza. Mapping the Network of Social Cognition Domains in Children With Autism Spectrum Disorder Through Graph Analysis. *Frontiers in Psychiatry*, 11, 2020. doi: 10.3389/fpsyt.2020.579339.
- [48] Alice Vinçon-Leite, Ana Saitovitch, Herve Lemaître, Elza Rechtman, Jennifer Boisgontier, Ludovic Fillon, Anne Philippe, Marlène Rio, Isabelle Desguerre, Aurélie Fabre, et al. Identifying interindividual variability of social perception and associated brain anatomical correlations in children with autism spectrum disorder using eye-tracking and diffusion tensor imaging mri (dti-mri). *Cerebral Cortex*, page bhad434, 2023. doi: 10.1093/cercor/bhad434.
- [49] Elisa Di Giorgio, David Meary, Olivier Pascalis, and Francesca Simion. The face perception system becomes species-specific at 3 months: An eye-tracking study. *International Journal of Behavioral Development*, 37(2):95–99, 2013. doi: 10.1177/0165025412465362.
- [50] Robin Sifre, Lindsay Olson, Scott Gillespie, Ami Klin, Warren Jones, and Sarah Shultz. A longitudinal investigation of preferential attention to biological motion in 2-to 24-month-old infants. *Scientific reports*, 8(1):1–10, 2018. doi: 10.1038/s41598-018-20808-0.
- [51] Patricia Soto-Icaza, Francisco Aboitiz, and Pablo Billeke. Development of social skills in children: neural and behavioral evidence for the elaboration of cognitive models. *Frontiers in neuroscience*, 9:1–16, 2015. doi: 10.3389/fnins.2015.00333.
- [52] Warren Jones, Katelin Carr, and Ami Klin. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry*, 65(8):946–954, 2008. ISSN 0003-990X. doi: 10.1001/archpsyc.65.8.946.
- [53] Karen Pierce, Steven Marinero, Roxana Hazin, Benjamin McKenna, Cynthia Carter Barnes, and Ajith Malige. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biological psychiatry*, 79(8):657–666, 2016. doi: 10.1016/j.biopsych.2015.03.032.
- [54] Adrienne Moore, Madeline Wozniak, Andrew Yousef, Cindy Carter Barnes, Debra Cha, Eric Courchesne, and Karen Pierce. The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Molecular autism*, 9(1):1–13, 2018. doi: 10.1186/s13229-018-0202-z.
- [55] Guang Liu, Jing Yang, and Zuo Li. Content-based image retrieval using computational visual attention model. *Pattern Recognition*, 48(8):2554–2566, 2015. doi: 10.1016/j.patcog.2015.02.005.
- [56] Kacie Dougherty, Michael C Schmid, and Alexander Maier. Binocular response modulation in the lateral geniculate nucleus. *Journal of Comparative Neurology*, 527(3):522–534, 2019. doi: 10.1002/cne.24417.
- [57] Daniel Casper and Janet Sparrow. *Orbital and Ocular Anatomy*. Springer, 2019.
- [58] Eric Kandel, James Schwartz, Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. *Princípios de neurociências-5*. AMGH Editora, 2014.

- [59] Alireza Sokhandan and Amirhassan Monadjemi. Visual tracking in video sequences based on biologically inspired mechanisms. *Computer Vision and Image Understanding*, 2018. doi: 10.1016/j.cviu.2018.10.002.
- [60] J McAfoose and BT Baune. Exploring visual–spatial working memory: A critical review of concepts and models. *Neuropsychology review*, 19:130–142, 2009. doi: 10.1007/s11065-008-9063-0.
- [61] Hanneke Den Ouden, Karl Friston, Nathaniel Daw, Anthony McIntosh, and Klaas Stephan. A dual role for prediction error in associative learning. *Cerebral cortex*, 19(5):1175–1185, 2009. doi: 10.1093/cercor/bhn161.
- [62] Ryan A. Stevenson, Magali Segers, Susanne Ferber, Morgan D. Barense, and Mark T. Wallace. The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology*, 5:1–4, 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2014.00379.
- [63] Elysa J Marco, Leighton BN Hinkley, Susanna S Hill, and Srikantan S Nagarajan. Sensory processing in autism: a review of neurophysiologic findings. *Pediatric research*, 69(8):48–54, 2011. doi: 10.1203/PDR.0b013e3182130c54.
- [64] Myriam Vandenbroucke, Steven Scholte, Herman van Engeland, Victor Lamme, and Chantal Kemner. A neural substrate for atypical low-level visual processing in autism spectrum disorder. *Brain*, 131(4):1013–1024, 2008. ISSN 0006-8950. doi: 10.1093/brain/awm321.
- [65] Francisco Sanchez-Marin and Jose Padilla-Medina. A psychophysical test of the visual pathway of children with autism. *Journal of Autism and Developmental Disorders*, 38(7):1270–1277, 2008. doi: 10.1007/s10803-007-0507-9.
- [66] Boutheina Jemel, Daniel Mimeault, Dave Saint-Amour, Anthony Hosein, and Laurent Mottron. VEP contrast sensitivity responses reveal reduced functional segregation of mid and high filters of visual channels in autism. *Journal of Vision*, 10(6):1–13, 2010. ISSN 1534-7362. doi: 10.1167/10.6.13.
- [67] Ali Borji and Laurent Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2012. doi: 10.1109/TPAMI.2012.89.
- [68] Fumi Katsuki and Christos Constantinidis. Bottom-up and top-down attention: different processes and overlapping neural systems. *The Neuroscientist*, 20(5):509–521, 2014. doi: 10.1177/1073858413514136.
- [69] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001. doi: 10.1038/35058500.
- [70] Zhang Mengmi. COMPUTATIONAL MODELS OF BOTTOM-UP AND TOP-DOWN VISUAL ATTENTION. PhD thesis, National University of Singapore, 2019.
- [71] Peter Pericles Trifonas. The semiotics of visual perception and the autonomy of pictorial text: Toward a semiotic pedagogy of the image. *Educational Philosophy and Theory*, 53(7):696–705, 2021. doi: 10.1080/00131857.2020.1761329.
- [72] Jéssica dos Santos de Oliveira. Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional de atenção visual. *Dissertação: Universidade de São Paulo*, 2018.
- [73] Derrick Parkhurst, Klinton Law, and Ernst Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision research*, 42(1):107–123, 2002. doi: 10.1016/S0042-6989(01)00250-4.
- [74] Yuming Fang, Hanqin Huang, Boyang Wan, and Yifan Zuo. Visual attention modeling for autism spectrum disorder by semantic features. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 625–628. IEEE, 2019. doi: 10.1109/ICMEW.2019.00120.

- [75] Nelson Garcia de Paula. Detecção de região saliente em imagens usando dissimilaridade de cor e amostragem por pixels aleatórios. Tese: Universidade Tecnológica Federal do Paraná, 2015.
- [76] David S Wooding. Fixation maps: quantifying eye-movement traces. In Proceedings of the 2002 symposium on Eye tracking research & applications, pages 31–36, 2002. doi: 10.1145/507072.507078.
- [77] Marco Reggiannini and Davide Moroni. The use of saliency in underwater computer vision: A review. Remote Sensing, 13(1):22, 2021. doi: 10.3390/rs13010022.
- [78] Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. Predicting human eye fixations via an LSTM-based saliency attentive model. IEEE Transactions on Image Processing, 27(10):5142–5154, 2018. doi: 10.1109/TIP.2018.2851672.
- [79] Mark Plitt, Kelly Anne Barnes, and Alex Martin. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. NeuroImage: Clinical, 7:359–366, 2015. doi: 10.1016/j.nicl.2014.12.013.
- [80] Yu Li, Yuying Zhu, Benedictor Alexander Nguchu, Yanming Wang, Huijuan Wang, Bensheng Qiu, and Xiaoxiao Wang. Dynamic functional connectivity reveals abnormal variability and hyper-connected pattern in autism spectrum disorder. Autism Research, 13(2):230–243, 2020. doi: 10.1002/aur.2212.
- [81] Caroline E Robertson and Simon Baron-Cohen. Sensory perception in autism. Nature Reviews Neuroscience, 18(11):671–684, 2017. doi: 10.1038/nrn.2017.112.
- [82] Dragan Rangelov and Jason B Mattingley. Evidence accumulation during perceptual decision-making is sensitive to the dynamics of attentional selection. NeuroImage, 220:117093, 2020. doi: 10.1016/j.neuroimage.2020.117093.
- [83] Alexis Nebout, Weijie Wei, Zhi Liu, Lijin Huang, and Olivier Le Meur. Predicting saliency maps for ASD people. In 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pages 629–632. IEEE, 2019. doi: 10.1109/ICMEW.2019.00121.
- [84] Weijie Wei, Zhi Liu, Lijin Huang, Alexis Nebout, and Olivier Le Meur. Saliency prediction via multi-level features and deep supervision for children with autism spectrum disorder. In 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pages 621–624. IEEE, 2019. doi: 10.1109/ICMEW.2019.00119.
- [85] Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A Laugeson, Daniel P Kennedy, Ralph Adolphs, and Qi Zhao. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron, 88(3):604–616, 2015. doi: 10.1016/j.neuron.2015.09.042.
- [86] Guido Löhr. Embodied cognition and abstract concepts: Do concept empiricists leave anything out? Philosophical Psychology, 32(2):161–185, 2019. doi: 10.1080/09515089.2018.1517207.
- [87] Marcel Harpaintner, Natalie M Trumpp, and Markus Kiefer. The semantic content of abstract concepts: A property listing study of 296 abstract words. Frontiers in psychology, 9:1748, 2018. doi: 10.3389/fpsyg.2018.01748.
- [88] Amit Sheth, Pramod Anantharam, and Cory Henson. Semantic, cognitive, and perceptual computing: Paradigms that shape human experience. Computer, 49(3):64–72, 2016. doi: 10.1109/MC.2016.75.
- [89] Hadumod Bussmann. Routledge dictionary of language and linguistics. Routledge, 2006.
- [90] Friedemann Pulvermüller. Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. Topics in Cognitive Science, 10(3):590–620, 2018. doi: 10.1111/tops.12367.
- [91] Li-Ying Fan, James R Booth, Min Liu, Tai-Li Chou, and Susan Shur-Fen Gau. Developmental differences in neural connectivity for semantic processing in youths with autism. Journal of Child Psychology and Psychiatry, 62(9):1090–1099, 2021. doi: 10.1111/jcpp.13373.

- [92] Penny M Pexman, Veronica Diveica, and Richard J Binney. Social semantics: the organization and grounding of abstract concepts. *Philosophical Transactions of the Royal Society B*, 378(1870): 20210363, 2023. doi: 10.1098/rstb.2021.0363.
- [93] Mohammad Kiani, Hamidreza Pouretamad, and Jamal Amani Rad. Visual attention patterns in autism spectrum disorder. *Advances in Cognitive Science*, 2021. doi: 10.30514/icss.23.2.1.
- [94] A Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems 2nd ed.* O'Reilly Media. Sebastopol, CA, USA, 2019. ISBN 978-1-492-03264-9.
- [95] Joe G Greener, Shaun M Kandathil, Lewis Moffat, and David T Jones. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23(1):40–55, 2022. doi: 10.1038/s41580-021-00407-0.
- [96] Laura Wratten, Andreas Wilm, and Jonathan Göke. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nature methods*, 18(10):1161–1168, 2021. doi: 10.1038/s41592-021-01254-9.
- [97] Tammy Jiang, Jaimie L Gradus, and Anthony J Rosellini. Supervised machine learning: a brief primer. *Behavior Therapy*, 51(5):675–687, 2020. doi: 10.1016/j.beth.2020.05.002.
- [98] Iqbal H Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3):1–21, 2021. doi: 10.1007/s42979-021-00592-x.
- [99] Mingxing Gong. A novel performance measure for machine learning classification. *International Journal of Managing Information Technology (IJMIT) Vol*, 13, 2021. doi: 10.5121/ijmit.2021.13101.
- [100] Helen R Sofaer, Jennifer A Hoeting, and Catherine S Jarnevich. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577, 2019. doi: 10.1111/2041-210X.13140.
- [101] Alberto Jiménez-Valverde. Insights into the area under the receiver operating characteristic curve (auc) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4):498–507, 2012.
- [102] Vladimir Nasteski. An overview of the supervised machine learning methods. *Horizons. b*, 4: 51–62, 2017. doi: 10.20544/HORIZONS.B.04.1.17.P05.
- [103] Brett Lantz. *Machine learning with R: expert techniques for predictive modeling.* Packt publishing ltd, 2019.
- [104] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [105] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1310–1315. IEEE, 2016.
- [106] Maher G. M. Abdolrasol, S. M. Suhail Hussain, Taha Selim Ustun, Mahidur R. Sarker, Mahammad A. Hannan, Ramizi Mohamed, Jamal Abd Ali, Saad Mekhilef, and Abdalrhman Milad. Artificial neural networks based optimization techniques: A review. *Electronics*, 10(21), 2021. ISSN 2079-9292. doi: 10.3390/electronics10212689. URL <https://www.mdpi.com/2079-9292/10/21/2689>.
- [107] Khaled Fawagreh, Mohamed Medhat Gaber, and Eyad Elyan. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2(1):602–609, 2014. doi: 10.1080/21642583.2014.956265.
- [108] Guobin Wan, Xuejun Kong, Binbin Sun, Siyi Yu, Yiheng Tu, Joel Park, Courtney Lang, Madelyn Koh, Zhen Wei, Zhe Feng, et al. Applying eye tracking to identify autism spectrum disorder in children. *Journal of autism and developmental disorders*, 49(1):209–215, 2019. doi: 10.1007/s10803-018-3690-y.

- [109] Ming Jiang, Sunday M Francis, Diksha Srishyla, Christine Conelea, Qi Zhao, and Suma Jacob. Classifying individuals with ASD through Facial emotion recognition and eye-tracking. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 6063–6068. IEEE, 2019. doi: 10.1109/embc.2019.8857005.
- [110] Ali Borji. Boosting bottom-up and top-down visual features for saliency estimation. In 2012 IEEE conference on computer vision and pattern recognition, pages 438–445. IEEE, 2012. doi: 10.1109/CVPR.2012.6247706.
- [111] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In 2009 IEEE 12th international conference on computer vision, pages 2106–2113. IEEE, 2009. doi: 10.1109/ICCV.2009.5459462.
- [112] Eero P Simoncelli and William T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In Proceedings., International Conference on Image Processing, volume 3, pages 444–447. IEEE, 1995. doi: 10.1109/ICIP.1995.537667.
- [113] Paul A Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, and Jose G Conde. Research electronic data capture (redcap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics, 42(2):377–381, 2009. doi: 10.1016/j.jbi.2008.08.010.
- [114] Cristiane S Duarte, Isabel AS Bordin, Albeni De Oliveira, and Hector Bird. The CBCL and the identification of children with autism and related conditions in Brazil: pilot findings. Journal of autism and developmental disorders, 33(6):703–707, 2003. doi: 10.1023/b:jadd.0000006005.31818.1c.
- [115] David Wechsler. Wechsler intelligence scale for children—Fourth Edition (WISC-IV). The Psychological Corporation, 3, 2003.
- [116] Girlene Ribeiro de Jesus. Normatização e validação do teste não-verbal de inteligência SON-R 2 1/2-7 [a] para o Brasil. Tese: Universidade de Brasília, 2009.
- [117] Simon Whitaker. The stability of iq in people with low intellectual ability: An analysis of the literature. Intellectual and Developmental Disabilities, 46(2):120–128, 2008. doi: 10.1352/0047-6765(2008)46[120:TSOIP]2.0.CO;2.
- [118] Anne T Berg, Sigita Plioplys, and Roberto Tuchman. Risk and correlates of autism spectrum disorder in children with epilepsy: a community-based study. Journal of child neurology, 26(5): 540–547, 2011. doi: 10.1177/0883073810384869.
- [119] Tobii Technology. Tobii TX300 Eye Tracker, 2021. URL https://medialibrary.uantwerpen.be/oldcontent/container29268/files/Tobii_TX300_EyeTracker_UserManual.pdf.
- [120] Tobii Technology. Tobii Studio, 2021. URL <https://go.tobii.com/TSUM>.
- [121] Tobii Technology. Tobii, 2021. URL <https://www.tobiipro.com/product-listing/fusion/>.
- [122] Tobii Technology. Tobii prolab overview, 2021. URL <https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/design/tobii-pro-lab-overview/>.
- [123] Tobii Technology. Tobii, 2012. URL http://www.vinis.co.kr/ivt_filter.pdf.
- [124] Tobii Technology. Tobii Pro Lab Gaze Filter, 2023. URL https://connect.tobii.com/s/article/Gaze-Filter-functions-and-effects?language=en_US.
- [125] R Core Team. R: A language and environment for statistical computing, 2022. URL www.R-project.org/.
- [126] Kenneth Holmqvist, Saga Lee Örbom, Ignace TC Hooge, Diederick C Niehorster, Robert G Alexander, Richard Andersson, Jeroen S Benjamins, Pieter Bignaut, Anne-Marie Brouwer, Lewis L Chuang, et al. Eye tracking: empirical foundations for a minimal reporting guideline. Behavior research methods, pages 1–53, 2022. doi: 10.3758/s13428-021-01762-8.

- [127] Courtney E Venker, Dominik Neumann, and Fashina Aladé. Visual perceptual salience and novel referent selection in children with and without autism spectrum disorder. *Autism & Developmental Language Impairments*, 7:23969415221085476, 2022. doi: 10.1177/23969415221085476.
- [128] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998. doi: 10.1109/34.730558.
- [129] Jacalyn Guy, Laurent Mottron, Claude Berthiaume, and Armando Bertone. The developmental trajectory of contrast sensitivity in autism spectrum disorder. *Autism Research*, 9(8):866–878, 2016. doi: 10.1002/aur.1579.
- [130] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer, 2003. doi: 10.1007/3-540-45103-X_50.
- [131] William Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950. doi: 10.1002/1097-0142.
- [132] Brian Nosek, Tom Hardwicke, Hannah Moshontz, et al. Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 2022. doi: 10.1146/annurev-psych-020821-114157.
- [133] Prasad Patil, Roger D Peng, and Jeffrey T Leek. A visual tool for defining reproducibility and replicability. *Nature human behaviour*, 3(7):650–652, 2019. doi: 10.1038/s41562-019-0629-z.
- [134] SJ Hu, R Webbink, J Lee, and Y Long. Robustness evaluation for compliant assembly systems. *J. Mech. Des.*, 125(2):262–267, 2003. doi: 10.1115/1.1563633.
- [135] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *the Journal of machine Learning research*, 9:1871–1874, 2008. doi: 10.5555/1390681.1442794.
- [136] The MathWorks Inc. Treebagger, 2022. URL www.mathworks.com/help/stats/treebagger.html.
- [137] Robin Genuer, Jean-Michel Poggi, Robin Genuer, and Jean-Michel Poggi. *Random forests*. Springer, 2020. doi: 10.1007/978-3-030-56485-8_3.
- [138] Lucy Nwosu, Hui Wang, Jiang Lu, Ishaq Unwala, Xiaokun Yang, and Ting Zhang. Deep convolutional neural network for facial expression recognition using facial parts. In *15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pages 1318–1321, 2017. doi: 10.1109/DASC-PiCom-DataCom-CyberSciTec.2017.213.
- [139] Yousif Khairuddin and Zhuofa Chen. Facial emotion recognition: State of the art performance on fer2013. preprint arXiv:2105.03588, 2021.
- [140] Tanja Skiendziel, Andreas G Rösch, and Oliver C Schultheiss. Assessing the convergent validity between the automated emotion recognition software noldus facereader 7 and facial action coding system scoring. *PloS one*, 14(10):e0223905, 2019. doi: 10.1371/journal.pone.0223905.
- [141] Jufeng Yang, Dongyu She, Yu-Kun Lai, Paul L Rosin, and Ming-Hsuan Yang. Weakly supervised coupled networks for visual sentiment analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7584–7592, 2018. doi: 10.1109/CVPR.2018.00791.
- [142] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018.

- [143] Lijuan Wang, Xin Zhang, Ning Jiang, Haiying Wu, and Jufeng Yang. D 2 s: Dynamic distribution supervision for multi-label facial expression recognition. In 2022 IEEE International Conference on Multimedia and Expo (ICME), pages 1–6. IEEE, 2022. doi: 10.1109/ICME52920.2022.9859687.
- [144] Jesus Olivares-Mercado, Karina Toscano-Medina, Gabriel Sanchez-Perez, Jose Portillo-Portillo, Hector Perez-Meana, and Gibran Benitez-Garcia. Analysis of hand-crafted and learned feature extraction methods for real-time facial expression recognition. In 2019 7th International Workshop on Biometrics and Forensics (IWBF), pages 1–6. IEEE, 2019. doi: 10.1109/IWBF.2019.8739178.
- [145] Gabrielle E Reimann, Catherine Walsh, Kelsey D Csumitta, Patrick McClure, Francisco Pereira, Alex Martin, and Michal Ramot. Gauging facial feature viewing preference as a stable individual trait in autism spectrum disorder. Autism research, 14(8):1670–1683, 2021. doi: 10.1002/aur.2540.
- [146] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. Mediapipe: A framework for building perception pipelines. arXiv preprint, 2019. doi: 10.48550/arXiv.1906.08172.
- [147] Minwei Deng. Robust human gesture recognition by leveraging multi-scale feature fusion. Signal Processing: Image Communication, 83:115768, 2020. doi: 10.1016/j.image.2019.115768.
- [148] Ming Wang, Zheng Yan, Ting Wang, Pingqiang Cai, Siyu Gao, Yi Zeng, Changjin Wan, Hong Wang, Liang Pan, Jiancan Yu, et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. Nature Electronics, 3(9):563–570, 2020. doi: 10.1038/s41928-020-0422-z.
- [149] Okan Kopuklu, Yao Rong, and Gerhard Rigoll. Talking with your hands: Scaling hand gestures and recognition with cnns. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pages 1–10, 2019. doi: 10.1109/ICCVW.2019.00345.
- [150] Okan Köpüklü, Ahmet Gunduz, Neslihan Kose, and Gerhard Rigoll. Real-time hand gesture detection and classification using convolutional neural networks. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), pages 1–8, 2019. doi: 10.1109/FG.2019.8756576.
- [151] Amirhossein Dadashzadeh, Alireza Targhi, Maryam Tahmasbi, and Majid Mirmehdi. Hgr-net: a fusion network for hand gesture segmentation and recognition. IET Computer Vision, 2019. doi: 10.1049/iet-cvi.2018.5796.
- [152] Petros Koutras and Petros Maragos. Estimation of eye gaze direction angles based on active appearance models. In 2015 IEEE International Conference on Image Processing (ICIP), pages 2424–2428. IEEE, 2015. doi: 10.1109/ICIP.2015.7351237.
- [153] Sankha S Mukherjee and Neil Martin Robertson. Deep head pose: Gaze-direction estimation in multimodal video. IEEE Transactions on Multimedia, 17(11):2094–2107, 2015. doi: 10.1109/TMM.2015.2482819.
- [154] Eunji Chong, Nataniel Ruiz, Yongxin Wang, Yun Zhang, Agata Rozga, and James M Rehg. Connecting gaze, scene, and attention: Generalized attention estimation via joint modeling of gaze and scene saliency. In Proceedings of the European conference on computer vision (ECCV), pages 383–398, 2018. doi: 10.1007/978-3-030-01228-1_24.
- [155] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 6912–6921, 2019. doi: 10.1109/ICCV.2019.00701.
- [156] Tianrong Rao, Min Xu, and Huiying Liu. Generating affective maps for images. Multimedia Tools and Applications, 77(13):17247–17267, 2018. doi: 10.1007/s11042-017-5289-2.

- [157] Sicheng Zhao, Xuanbai Chen, Xiangyu Yue, Chuang Lin, Pengfei Xu, Ravi Krishna, Jufeng Yang, Guiguang Ding, Alberto L. Sangiovanni-Vincentelli, and Kurt Keutzer. Emotional semantics-preserved and feature-aligned cycleGAN for visual emotion adaptation. *IEEE Transactions on Cybernetics*, 52(10):10000–10013, 2022. doi: 10.1109/TCYB.2021.3062750.
- [158] Abu Sayeed Md Sohail and Prabir Bhattacharya. Detection of facial feature points using anthropometric face model. In *Signal Processing for Image Enhancement and Multimedia Processing*, pages 189–200. Springer, 2008. doi: 10.1007/978-0-387-72500-0_17.
- [159] Francesca Simion and Elisa Di Giorgio. Face perception and processing in early infancy: inborn predispositions and developmental changes. *Frontiers in psychology*, 6:969, 2015. doi: 10.3389/fpsyg.2015.00969.
- [160] Ana Maria Portugal, Charlotte Viktorsson, Mark J Taylor, Luke Mason, Kristiina Tammimies, Angelica Ronald, and Terje Falck-Ytter. Infants’ looking preferences for social versus non-social objects reflect genetic variation. *Nature Human Behaviour*, pages 1–10, 2023. doi: 10.1038/s41562-023-01764-w.
- [161] Carolyn C Goren, Merrill Sarty, and PY Wu. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56(4):544–549, 1975. doi: 10.1542/peds.56.4.544.
- [162] Mark H Johnson, Suzanne Dziurawiec, Hadyn Ellis, and John Morton. Newborns’ preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2):1–19, 1991. doi: 10.1016/0010-0277(91)90045-6.
- [163] Teodora Gliga and Gergely Csibra. Seeing the face through the eyes: a developmental perspective on face expertise. *Progress in brain research*, 164:323–339, 2007. doi: 10.1016/S0079-6123(07)64018-7.
- [164] Charlotte Viktorsson, Ana Maria Portugal, Danyang Li, Maja Rudling, Monica Siqueiros Sanchez, Kristiina Tammimies, Mark J Taylor, Angelica Ronald, and Terje Falck-Ytter. Preferential looking to eyes versus mouth in early infancy: heritability and link to concurrent and later development. *Journal of child psychology and psychiatry*, 64(2):311–319, 2023. doi: 10.1111/jcpp.13724.
- [165] David J Lewkowicz. Early experience and multisensory perceptual narrowing. *Developmental psychobiology*, 56(2):292–315, 2014. doi: 10.1002/dev.21197.
- [166] David J Lewkowicz and Amy M Hansen-Tift. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, 109(5):1431–1436, 2012. doi: 10.1073/pnas.1114783109.
- [167] Ami Klin, Warren Jones, Robert Schultz, Fred Volkmar, and Donald Cohen. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, 59(9):809–816, 2002. doi: 10.1001/archpsyc.59.9.809.
- [168] Melissa L-H Vö, Tim J Smith, Parag K Mital, and John M Henderson. Do the eyes really have it? dynamic allocation of attention when viewing moving faces. *Journal of vision*, 12(13):3–3, 2012. doi: 10.1167/12.13.3.
- [169] Paul Ekman. Facial expressions. In José Miguel Fernández Dols and James Albert Russell, editors, *The science of facial expression*, chapter 2, pages 39–56. Oxford University Press, 2017.
- [170] Erika L Rosenberg and Paul Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 2020.
- [171] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1):1–68, 2019. doi: 10.1177/1529100619832930.

- [172] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. Psychological bulletin, 133(4):694, 2007. doi: 10.1037/0033-2909.133.4.694.
- [173] David Matsumoto, Mark G Frank, and Hyi Sung Hwang. Nonverbal communication: Science and applications. Sage Publications, 2012.
- [174] Shuo Wang. Social Saliency: Visual Psychophysics and Single-Neuron Recordings in Humans. PhD thesis, California Institute of Technology, 2014.
- [175] QianYing Ye, LinRu Liu, ShaoLi Lv, SanMei Cheng, HuiLin Zhu, YanTing Xu, XiaoBing Zou, and HongZhu Deng. The gestures in 2–4-year-old children with autism spectrum disorder. Frontiers in psychology, 12:604542, 2021. doi: 10.3389/fpsyg.2021.604542.
- [176] Pauline Maes, Fanny Stercq, and Mikhail Kissine. Attention to intentional versus incidental pointing gestures in young autistic children: An eye-tracking study. Journal of experimental child psychology, 210:105205, 2021. doi: 10.1016/j.jecp.2021.105205.
- [177] Tadas Baltrušaitis, others. OpenFace, 2022. URL github.com/TadasBaltrušaitis/OpenFace.
- [178] Stefanos Zafeiriou, George Trigeorgis, Grigorios Chrysos, Jiankang Deng, and Jie Shen. The menpo facial landmark localisation challenge: A step towards the solution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 170–179, 2017. doi: 10.1109/CVPRW.2017.263.
- [179] Benjamin Johnston and Philip de Chazal. A review of image-based automatic facial landmark identification techniques. EURASIP Journal on Image and Video Processing, 2018(1):1–23, 2018. doi: 10.1186/s13640-018-0324-4.
- [180] Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, and Tao Mei. The elements of end-to-end deep face recognition: A survey of recent advances. ACM Computing Surveys (CSUR), 54(10s):1–42, 2022. doi: 10.1145/3507902.
- [181] Paul Viola and Michael Jones. Robust real-time object detection [j]. International journal of computer vision, 4:34–47, 2001.
- [182] Fernando De la Torre and Jeffrey F Cohn. Facial expression analysis. Visual analysis of humans: Looking at people, pages 377–409, 2011. doi: 10.1007/978-0-85729-997-0_19.
- [183] Erroll Wood, Tadas Baltrušaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. Rendering of eyes for eye-shape registration and gaze estimation. In Proceedings of the IEEE international conference on computer vision, pages 3756–3764, 2015. doi: 10.1109/ICCV.2015.428.
- [184] Google Team, others. MediaPipe, 2022. URL github.com/google/mediapipe.
- [185] Henrik Bengtsson, Andy Jacobson, Jason Riedy. R.matlab: Read and Write MAT Files and Call MATLAB from Within R, 2023. URL <https://CRAN.R-project.org/package=R.matlab>.
- [186] Peshawa Jamal Muhammad Ali, Rezhna Hassan Faraj, Erbil Koya, Peshawa J Muhammad Ali, and Rezhna H Faraj. Data normalization and standardization: a technical report. Mach Learn Tech Rep, 1(1):1–6, 2014.
- [187] RStudio Team. RStudio: Integrated Development Environment for R, 2020. URL www.rstudio.com.
- [188] Benjamin Zablotsky, Lindsey I Black, Matthew J Maenner, Laura A Schieve, Melissa L Danielson, Rebecca H Bitsko, Stephen J Blumberg, Michael D Kogan, and Coleen A Boyle. Prevalence and trends of developmental disabilities among children in the united states: 2009–2017. Pediatrics, 144(4), 2019. doi: 10.1542/peds.2019-0811.
- [189] Jasna Martinovic and Søren Andersen. Cortical summation and attentional modulation of combined chromatic and luminance signals. NeuroImage, 176:390–403, 2018. doi: 10.1016/j.neuroimage.2018.04.066.

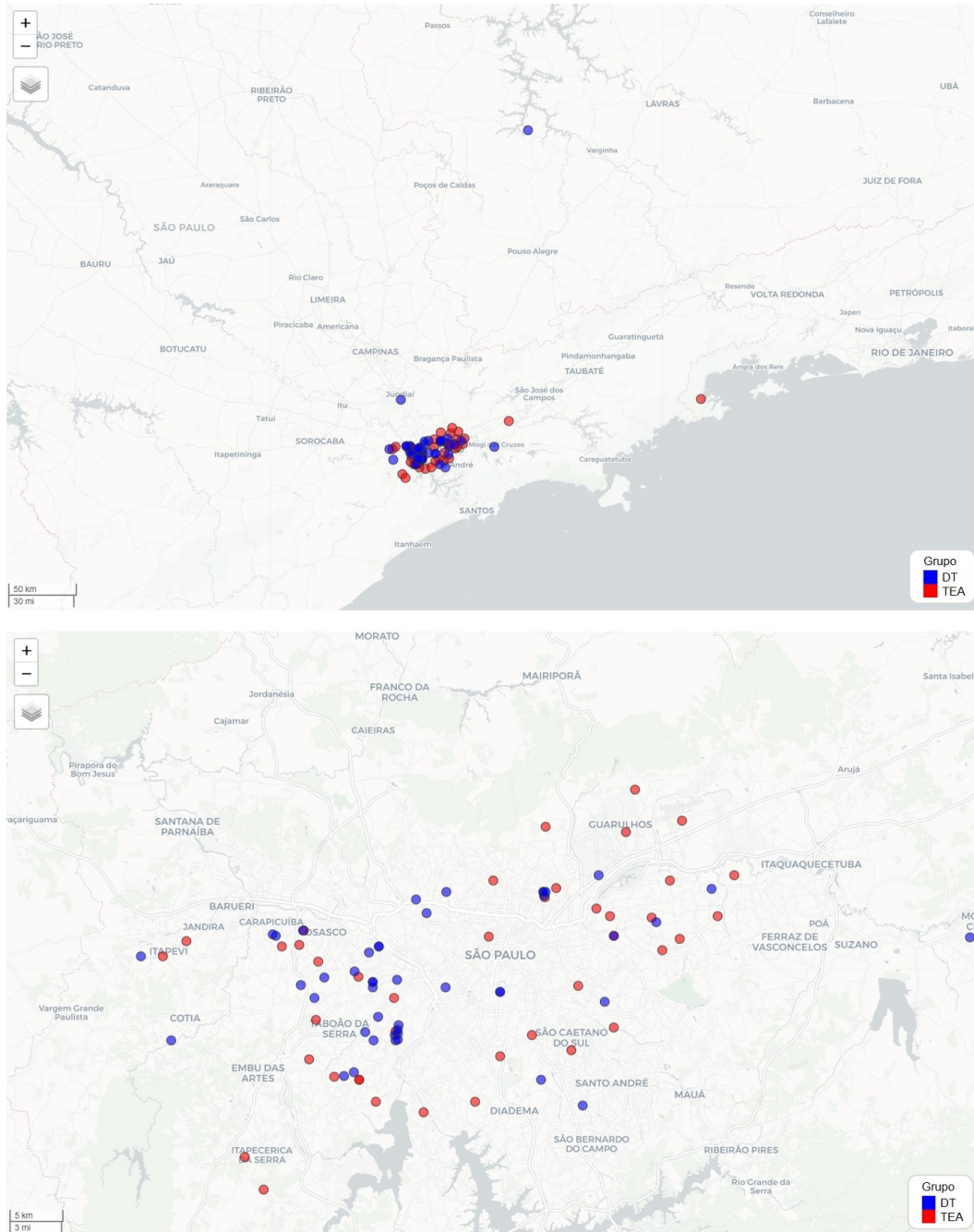
- [190] Felipe O Franco, Jessica S Oliveira, Joana Portolese, Fernando M Sumiya, Andréia F Silva, Ariane Machado-Lima, Fatima LS Nunes, and Helena Brentani. Computer-aided autism diagnosis using visual attention models and eye-tracking: replication and improvement proposal. BMC Medical Informatics and Decision Making, 23(1):1–9, 2023. doi: 10.1186/s12911-023-02389-9.
- [191] Mohamed El Aroussi, Mohammed El Hassouni, Sanaa Ghouzali, Mohammed Rziza, and Driss Aboutajdine. Local appearance based face recognition method using block based steerable pyramid transform. Signal Processing, 91(1):38–50, 2011. doi: 10.1016/j.sigpro.2010.06.005.
- [192] Ruth Van der Hallen, Kris Evers, Katrien Brewaeys, Wim Van den Noortgate, and Johan Wagemans. Global processing takes time: A meta-analysis on local–global visual processing in asd. Psychological bulletin, 141(3):549, 2015. doi: 10.1037/bul0000004.
- [193] Jacalyn Guy, Laurent Mottron, Claude Berthiaume, and Armando Bertone. A developmental perspective of global and local visual perception in autism spectrum disorder. Journal of Autism and Developmental Disorders, 49:2706–2720, 2019. doi: 10.1007/s10803-016-2834-1.
- [194] Andrea Helo, Sebastian Pannasch, Louah Sirri, and Pia Rämä. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. Vision research, 103:83–91, 2014. doi: 10.1016/j.visres.2014.08.006.
- [195] Mirian C Revers, Jessica S Oliveira, Felipe O Franco, Joana Portolese, Thiago V Cardoso, Andréia F Silva, Ariane Machado-Lima, Fátima LS Nunes, and Helena Brentani. Classification of autism spectrum disorder severity using eye tracking data based on visual attention model. In 34th International Symposium on CBMS, pages 142–147. IEEE, 2021. doi: 10.1109/CBMS52027.2021.00062.
- [196] Young Shin Kim, Bennett L Leventhal, Yun-Joo Koh, Eric Fombonne, Eugene Laska, Eun-Chung Lim, Keun-Ah Cheon, Soo-Jeong Kim, Young-Key Kim, HyunKyung Lee, et al. Prevalence of autism spectrum disorders in a total population sample. American Journal of Psychiatry, 168(9): 904–912, 2011. doi: 10.1176/appi.ajp.2011.10101532.
- [197] Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of vision, 7(14): 4, 2007. doi: 10.1167/7.14.4.
- [198] Po-He Tseng, Ran Carmi, Ian GM Cameron, Douglas P Munoz, and Laurent Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. Journal of vision, 9(7):4–4, 2009. doi: 10.1167/9.7.4.
- [199] Juan Xu, Ming Jiang, Shuo Wang, Mohan S Kankanhalli, and Qi Zhao. Predicting human gaze beyond pixels. Journal of vision, 14(1):28–28, 2014. doi: 10.1167/14.1.28.
- [200] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? IEEE transactions on pattern analysis and machine intelligence, 41(3):740–757, 2018. doi: 10.1109/TPAMI.2018.2815601.
- [201] Sen Jia and Neil DB Bruce. Revisiting saliency metrics: Farthest-neighbor area under curve. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2667–2676, 2020.
- [202] Clare Harrop, Desiree Jones, Shuting Zheng, Sallie Nowell, Robert Schultz, and Julia Parish-Morris. Visual attention to faces in children with autism spectrum disorder: are there sex differences? Molecular Autism, 10(1):1–10, 2019. doi: 10.1186/s13229-019-0276-2.

APÊNDICE A - TABELA COMPLEMENTAR COM O LEVANTAMENTO DE CARACTERÍSTICAS SEMÂNTICAS

ID	Descrição do Estudo	Autores	Ano	Características de Interesse
1	Proposição de método para a extração de elementos gestuais manuais	Deng et al.	2020	15 gestos manuais
2	Extração de elementos gestuais manuais	Wang et al.	2020	10 gestos manuais
3	Compara a preferência de TEA e DT na visualização de elementos faciais	Reimann et al.	2020	11 partes do corpo
4	Reconhecimento de expressões faciais	Nwosu et al.	2017	7 expressões faciais, olhos e boca
5	Reconhecimento de expressões faciais	Khairuddin & Chen	2021	7 expressões faciais
6	Reconhecimento de expressões faciais e pontuação na FACS	Skiendziel et al.	2019	6 expressões faciais
7	Atribuição de emoção aos pixels que compõem a imagem	Rao et al.	2018	8 expressões
8	Estima a posição do olhar baseado em vídeo	Koutras & Maragos	2015	6 posições do olhar e angulação
9	Estima a posição do olhar baseado em vídeo	Mukherjee & Rob.	2015	8 posições do olhar e angulação
10	Estima a posição do olhar baseado em imagem	Chong et al.	2018	Angulação
11	Estima a posição do olhar baseado em vídeo	Kellnhofer et al.	2019	Angulação (360°)
12	Reconhecimento automático de ROI-emocional	Yang et al.	2018	4 expressões
13	Toolkit que fornece a extração de características relacionadas a face	Baltrusaitis et al.	2018	Landmarks + direção do olhar + AUs
14	Reconhecimento de Expressão Facial Multi-Rótulo	Wang et al.	2021	6 expressões faciais
15	Atribuição de emoção em imagens: distribuição e classificação de emoções	Zhao et al.	2022	8 emoções
16	Abordagem que classifica imagem em duas classes: com e sem gestos	Köpüklü et al.	2019a	Rótula presença e ausência de gesto
17	Reconhece 5 gestos manuais combinando sinais de imagem e som	Köpüklü et al.	2019b	5 gestos
18	Propõe uma rede para reconhecimento e segmentação de gestos	Dadashzadeh et al.	2019	9 gestos (não deixa claro quais são)
19	Avalia métodos de reconhecimento de expressões faciais	Benitez-Garcia et al.	2019	7 expressões faciais
20	FrameWork multipropósito para reconhecimento de gestos, landmarks, etc.	Lugaresi et al.	2019	Gestos, face e landmarks

APÊNDICE B - GEOLOCALIZAÇÃO DOS INDIVÍDUOS

Figura 26: Mapas com a geolocalização dos indivíduos do estudo.



Abreviações: DT, Desenvolvimento Típico; TEA, Transtorno do Espectro Autista.

APÊNDICE C - DESEMPENHO DOS CLASSIFICADORES CONSIDERANDO APENAS CARACTERÍSTICAS SEMÂNTICAS

Tabela 17: Desempenho dos classificadores considerando apenas características semânticas.

Dataset	Norm.	AUC	Sens.	Espec.	Prec.	Acurácia	F1-Score
I	Não	0,58	0,57	0,56	0,57	0,57	0,57
I	Sim	0,60	0,57	0,58	0,59	0,58	0,58
II	Não	0,64	0,61	0,60	0,49	0,60	0,54
II	Sim	0,63	0,59	0,64	0,51	0,62	0,55
III	Não	0,61	0,60	0,56	0,46	0,58	0,52
III	Sim	0,62	0,57	0,60	0,47	0,59	0,52
IV	Não	0,63	0,63	0,53	0,51	0,57	0,56
IV	Sim	0,57	0,65	0,46	0,48	0,54	0,55

Todos os resultados foram com o algoritmo Random Forest, considerando apenas as características semânticas. Abreviações: Norm., Normalização; AUC, Área Sob a Curva ROC; Sens., Sensibilidade; Espec., Especificidade; Prec., Precisão.

APÊNDICE D - DOCUMENTAÇÃO DO CÓDIGO BASE

Documentação do classificador de auxílio diagnóstico para o TEA, baseado em dados de rastreamento do olhar

Este arquivo descreve os passos necessários para execução/construção do classificador de auxílio diagnóstico ao Transtorno do espectro autista descrito em “Computer-aided autism diagnosis based on visual attention models using eye tracking”, disponível em:

<https://www.nature.com/articles/s41598-021-89023-8>

O classificador foi construído em MatLab, com a versão 2015a. Siga os passos a seguir:

1 – Downloads

Baixe os arquivos a seguir:

A – Códigos para a classificação diagnóstica:

[link a ser disponibilizado]

B – Dataset:

[link a ser disponibilizado]

C – OpenCv 3.0:

<https://sourceforge.net/projects/opencvlibrary/>

D – Pacotes MatLab:

<https://drive.google.com/drive/folders/1OtEiYcSU0qmEHaEYICankXV3o9rDonOW?usp=sharing>

E – Visual Studio 2012:

<https://www.microsoft.com/en-us/download/details.aspx?id=30679>

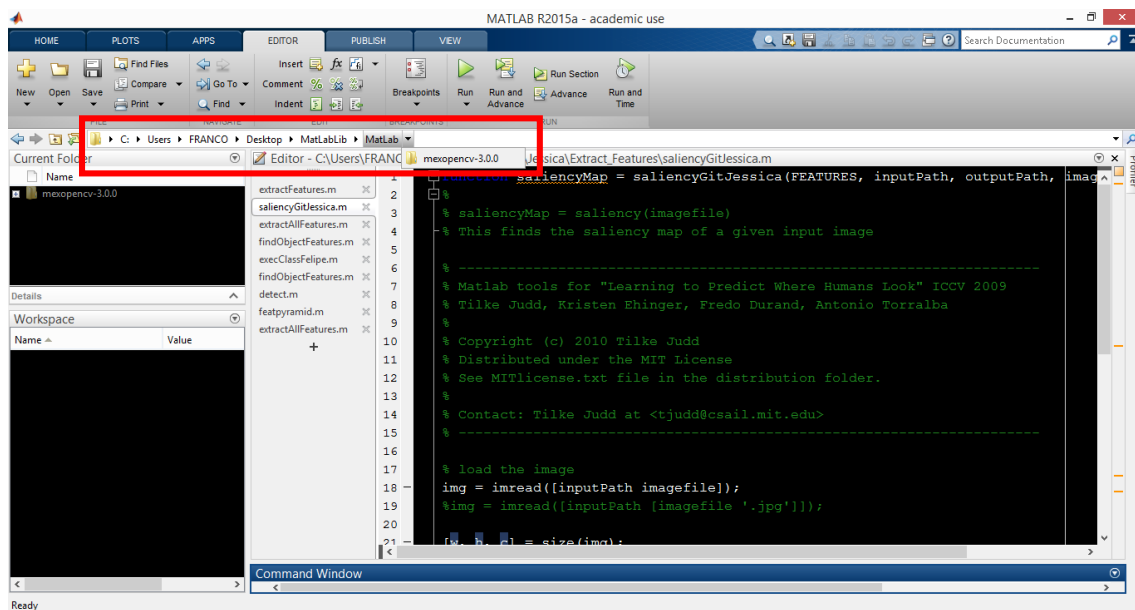
2 – Instalações

Instale o OpenCv e Visual Studio 2012 conforme as informações disponíveis no próprio site.

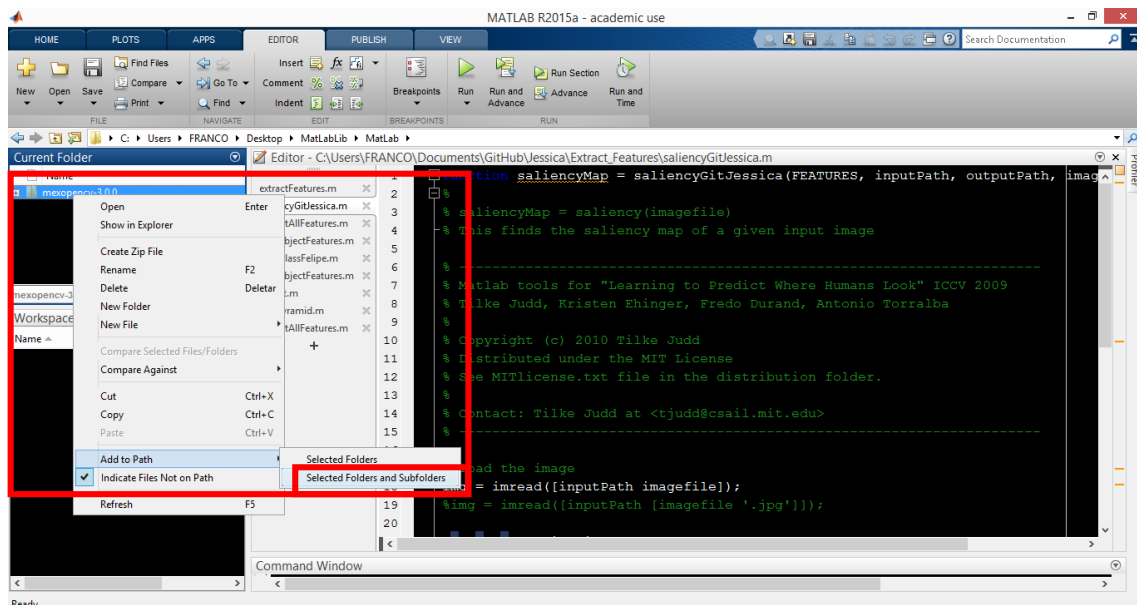
Após a instalação do OpenCv, você deve compilar os arquivos no MatLab usando o MexOpenCv:

https://drive.google.com/drive/folders/1NTEjG1_zO65SAU4s6DKwChCWrkZ624M3?usp=sharing

Para Compilar o MexOpenCv você deve adicionar o caminho da pasta, como no exemplo abaixo:



A seguir você deve adicionar as pastas e subpastas correspondentes, para isso clique com o botão direito na pasta e no menu que aparecerá, selecione a opção “Add to Path” e depois “Select Folders and Subfolders”, como na imagem abaixo:



Após a seleção das pastas e subpastas você deve compilar os arquivos, para isso utilize o comando abaixo:

```
mexopencv.make;
```

Observação: versões mais recentes do Visual Studio podem não ser compatíveis com o classificador. Não confundir com Visual Studio **Code**.

Pacotes: Para instalar os pacotes, siga as instruções do vídeo a seguir:

<https://www.mathworks.com/videos/package-a-custom-matlab-toolbox-106803.html>

3 – Extrair frames do vídeo:

```
videoObject =  
VideoReader('C:\Users\FRANCO\Documents\GitHub\Jessica\Stimuli\vídeos\Video 2 - Teste2.avi')  
numberOfFrames = videoObject.NumberOfFrames;  
output = 'C:\Users\FRANCO\Desktop\Frames\Numero\';
```

```
frame = cell(numberOfFrames);
for i = 1:numberOfFrames
    frame{i} = read(videoObject, i);
    imwrite(frame{i}, [output [num2str(i) '.jpg']]);
end
```

Não esqueça de alterar o caminho de origem dos arquivos e de criar uma pasta de saída para os arquivos.

Observação geral para os itens a seguir: Verifique em todos os scripts a necessidade de alterar o path para corresponder aos locais dos arquivos em sua máquina. Essa informação constará no topo de cada script, com os nomes “Input” e/ou “Output”. A organização das pastas deve ser mantida, pois há um encadeamento entre os scripts.

4 – Extrair características

Esta etapa de compreende a extração das características mencionadas na página 8 do artigo indicado no início do tutorial, com a adição de outras quatro características relacionadas referentes a relação entre os canais de cores, descritos inicialmente por Itti *et al.* e implementados por Judd *et. al.*, disponível em <https://people.csail.mit.edu/tjudd/WherePeopleLook/>.

O script utilizado aqui será o “extractAllFeatures.m”, que está na pasta “Extract_Features”.

A execução pode demorar. Comentário: em minha máquina (Dell I15-3567) leva aproximadamente três horas.

As características 30 (cena biológica), 31 (cena geométrica) e 32 (distância até o centro da cena), requerem a declaração de valores específicos referentes ao vídeo analisado no momento, segue o link para o documento com a esquematização destes valores:

<https://drive.google.com/file/d/1mrKcVpwX6QRjoUidWbQ6zNllcW7GDetD/view?usp=sharing>

5 – Índice “list”

No total existem 1604 frames, porém nem todos são informativos para a construção do classificador, portanto alguns deles serão descartados. A escolha de frames é feita com base na taxa de movimento entre eles, ou seja, frames com pouco movimento entre si são similares e portanto, pouco informativos. A taxa de movimento aqui é calculada com base no fluxo óptico entre dois frames e esse limiar está pré-definido para 0,33, para verificar o artigo original acesse: https://link.springer.com/chapter/10.1007%2F3-540-45103-X_50.

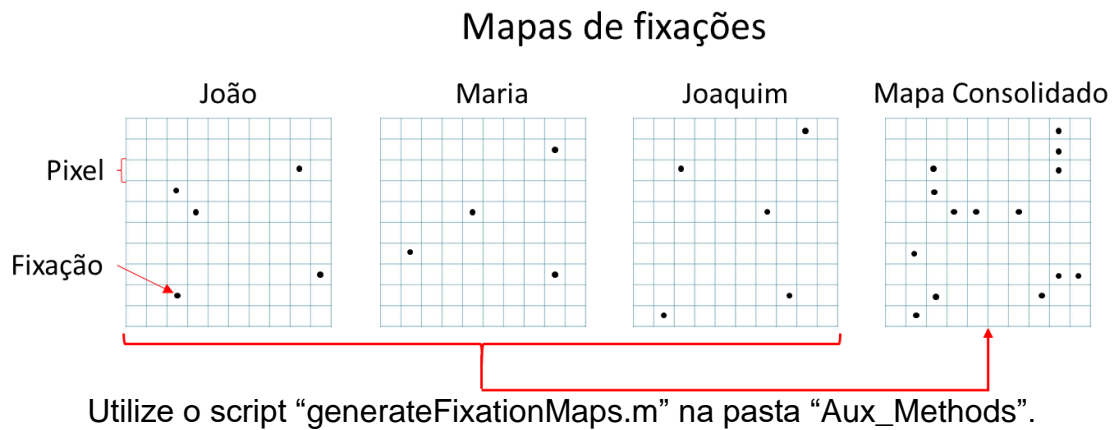
Na etapa anterior você criou um arquivo chamado “Frame” e com o nome interno “list”. Este arquivo contém os vetores (198 ao todo) de “agrupamento” de frames e também servirá na etapa 8 como índice de agrupamento entre mapas de fixações.

6 – Criar mapas de fixações:

Utilize o script “montaFixacao.m” na pasta “Aux_Methods” que você havia baixado anteriormente, no item “1 – A” deste tutorial.

7 – Consolidar mapas de fixação:

O objetivo desta etapa é sobrepor todos os mapas de fixações individuais em apenas um mapa para cada grupo, como ilustrado no esquema a seguir:



8 – Agrupamento de mapas de fixações:

Agora que criamos o arquivo “list”, temos os vetores para realizar o agrupamento dos mapas de fixações.

O script utilizado aqui é o “consolidaFelipe.m” na pasta “Aux_Methods”. Logo na primeira linha do código a variável “i” é definida com o valor 198, o qual corresponde ao número de frames considerados após a verificação da taxa de movimento, este valor pode ser alterado conforme o número de linhas geradas no arquivo “list”. O output desta etapa será armazenado dentro da pasta “Fixation_Maps” e você deve alterar a pasta interno conforme o grupo que você estiver trabalhando (TEA ou Controle, por exemplo).

9 – Rodando o classificador:

Agora que você já tem os arquivos de entrada para o modelo, utilize o script “execClassFelipe”. Você deverá alterar os caminhos (paths) declarados no início do código para coincidir com os caminhos da sua máquina. Verifique também os valores de “P” e “N”, que referem-se à quantidade de amostras positivas (no caso, TEA) e negativas (no caso, Controle), respectivamente.

Para reproduzir os resultados do projeto passe os parâmetros “execClass(1,1,15,350)” para executar a função.

Abaixo segue uma descrição com o significado de cada um desses parâmetros:

Primeiro valor (1): refere-se ao indutor de classificação utilizado, 1 – Rede Neural ou 2 – SVM;

Segundo valor (1): refere-se ao método de seleção das características, 1 – para utilizar as características selecionadas previamente (as mesmas do artigo), 2 – para Relief ou 3 – para Algoritmo Genético;

Terceiro valor (15): esse parâmetro é utilizado apenas quando se usa a opção de seleção pelo Algoritmo Genético, correspondendo ao número de características que deverão ser consideradas. Mesmo se não estiver utilizando o Algoritmo Genético, esse parâmetro deverá ser passado na função;

Quarto valor (350): refere-se ao número de pontos que deverão ser considerados na imagem para treinar o modelo.

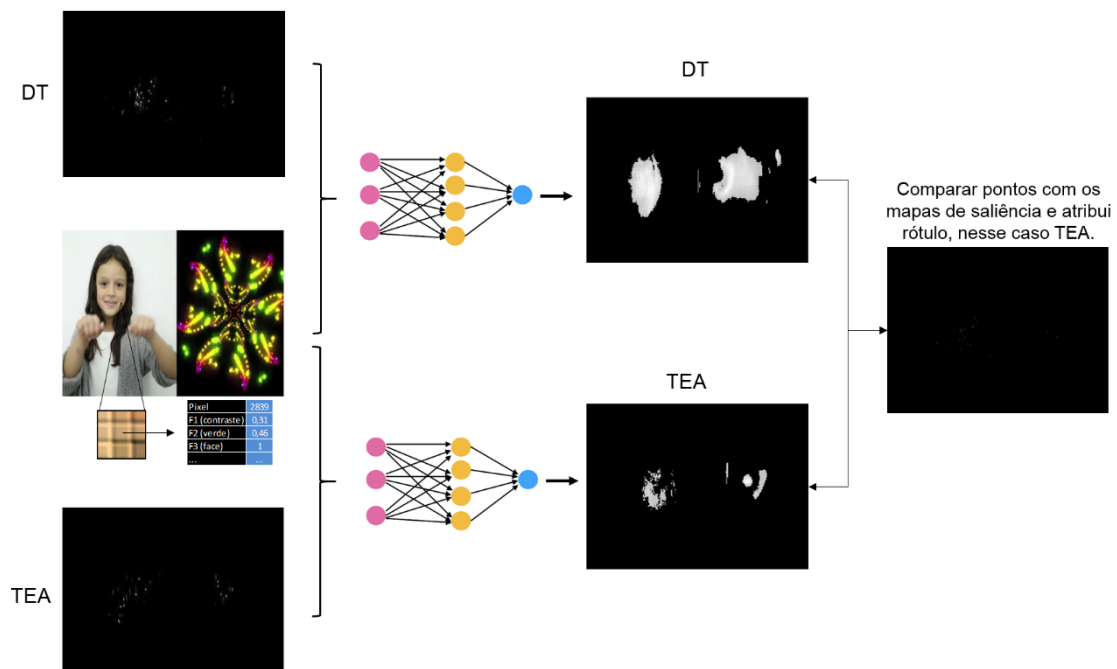
A execução pode demorar. Comentário: na minha máquina leva aproximadamente 9,5 horas.

10 – Saída do classificador:

Os parâmetros serão armazenados no “Workspace” do MatLab, você pode extrair apenas as informações que quiser ou salvar o Workspace inteiro para consulta posterior.

Observação: As métricas de saída referem-se à classificação com a validação-cruzada variando em função dos frames, portanto, considerando alguns frames para treino e outros para teste. Todos os indivíduos são considerados tanto grupo treino, quanto no grupo teste.

Segue abaixo um exemplo *Toy* para ilustrar o processo de Classificação:



11 – Avaliação do classificador:

As métricas de avaliação do classificador consideradas são AUC, acurácia, sensibilidade, especificidade e precisão.

Para obter essas métricas, utilize o código "etapaFinalMetricas".

Todas as métricas serão compiladas na variável "metricas" que estará no Workspace.

Informações complementares

Outras informações mais específicas foram adicionadas ao longo do script.

No caso de dúvidas entre em contato com Felipe de Oliveira Franco no email:

franco@ime.usp.br

ANEXO A - APROVAÇÃO PELO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Ferramenta de auxílio diagnóstico para transtorno do espectro autista: integração de características semânticas em modelos de atenção visual.

Pesquisador: Helena Paula Brentani

Área Temática:

Versão: 2

CAAE: 41895220.1.0000.0068

Instituição Proponente: Hospital das Clínicas da Faculdade de Medicina da USP

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.556.127

Apresentação do Projeto:

As informações constantes dos itens apresentação do projeto, objetivos e riscos e benefícios foram retiradas de PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1601652 e Projeto_Felipe_SGP.

O Transtorno do Espectro Autista (TEA) é um transtorno do neurodesenvolvimento, caracterizado pela deficiência na comunicação e interação social e padrões comportamentais restritos e repetitivos. A dificuldade em interpretar expressões faciais e linguagem corporal, características prevalentes no TEA, pode prejudicar o aprendizado e interações sociais. O aprendizado baseado na observação social atende dois critérios básicos: 1 - concentrar a atenção nas partes mais informativas da cena; 2 - adaptar seu comportamento em tempo real de acordo com o contexto. A crescente complexidade de interação social é um produto da seleção natural de um sistema de processamento visual cada vez mais elaborado. Estudos de rastreamento do olhar mostram que indivíduos com desenvolvimento neurotípico têm uma tendência inata de olhar para a região facial de outros humanos, destacadamente os olhos, característica já presente aos três meses de idade. Também foi evidenciada uma preferência por movimento biológico, ou seja, cenas com pessoas em movimento, ou seja, fatores que favorecem o aprendizado baseado na observação social. Por outro lado, no TEA isso não ocorre da mesma maneira. 1.2 Rastreamento do Olhar- O rastreador de olhar é um sistema que possibilita avaliar a atenção espontânea e a preferência do olhar registrando o foco visual do usuário. O método mais comum de rastreamento do olhar consiste na

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

análise da reflexão ocular, na qual a posição do olhar é estimada com alta precisão. O cálculo é realizado a partir de algoritmos que utilizam gravações de vídeo coletadas por câmeras remotas, as quais são colocadas na frente do observador. Não há, portanto, necessidade de equipamentos fixados na cabeça ou outros dispositivos que podem causar desconforto no usuário, facilitando o uso em crianças e, ainda, portadores de TEA. Falk-Ytter et al. mostram que sinais como tempo reduzido de observação de pessoas e rostos, falta de atenção e preferência por padrões geométricos parecem estar entre os primeiros sinais de TEA. Tal fato tem levado muitos profissionais da saúde e pesquisadores a utilizarem o rastreamento do olhar, dado seu potencial para diagnóstico e uso não invasivo. Grupos de pesquisa e profissionais da área ressaltam a importância do diagnóstico precoce e acompanhamento familiar para melhorar a qualidade de vida do indivíduo com TEA, impactando principalmente na cognição e linguagem. A análise do rastreamento do olhar para detectar os sinais presentes no TEA envolve a escolha do paradigma a ser utilizado, o qual depende dos objetivos da análise e do público-alvo. Muitos paradigmas têm sido propostos na literatura, tais como: Rastreamento de objetos múltiplos, no qual o observador deve acompanhar um subconjunto de objetos que se movem em diferentes trajetórias; Campo de visão útil, que baseia-se na região do campo visual da qual um observador pode extrair informações em um relance; Preferência visual estática, que consiste na apresentação de imagens simultâneas na tela, sendo um lado com estímulo social e o outro não social, geralmente formas geométricas; Preferência visual dinâmica, semelhante ao último, porém com a apresentação de vídeos ao invés de imagens simultâneas; dentre outros. Diferentes abordagens têm sido propostas para a classificação do TEA via rastreamento do olhar com paradigmas sociais e demarcação prévia de regiões de interesse (ROI). Jiang et al. propuseram um classificador diagnóstico utilizando reconhecimento de emoções (raiva, nojo, medo, felicidade, tristeza e surpresa), em que os indivíduos observam uma transição dos rostos, sendo que as expressões faciais mudam lentamente no vídeo. O método baseou-se principalmente no tempo de resposta à cada emoção e na quantidade de fixações na tela. O estudo apresentou uma boa precisão de classificação entre TEA e controle (desenvolvimento típico), com acurácia de 86%. Em outro estudo, Wan et al. utilizaram um vídeo de dez segundos em que aparecia uma jovem falando. As medidas avaliadas foram os tempos de fixação em seis ROIs. Após a aplicação da análise discriminante, os tempos de fixação na boca e no corpo puderam discriminar significativamente TEA e controle, com uma precisão de classificação de 85,1%. Podemos destacar também abordagens mistas, tal como o estudo de Carette et al., que combinaram diferentes estímulos, apresentando ao observador vários vídeos, com e sem áudio, cenas estáticas e dinâmicas, realistas ou com desenhos animados. Nesta

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

abordagem, o classificador teve a acurácia de predição de 92%. 1.3 Vias de processamento visual. Os dados visuais processados pela retina neural são transmitidos ao cérebro através do nervo óptico, constituído por um feixe de fibras axonais de células ganglionares. Os nervos ópticos direito e esquerdo se unem para formar o quiasma óptico. O trato óptico pós-quiasmático é composto por fibras combinadas de ambos os olhos, carregando informações derivadas do campo visual contralateral. Posteriormente, esses feixes de fibras seguem para o núcleo geniculado lateral dorsal do tálamo, o qual recebe informações de ambos os olhos e transmite esses sinais para o córtex visual no lobo occipital, onde ocorrerá a ativação da área visual V1. A partir desta área, sinais são transmitidos por duas vias diferentes, a via dorsal e a ventral. O fluxo ventral é, principalmente, responsável pela identificação de objetos, levando informação, por exemplo, sobre a forma e a cor. A via dorsal é dedicada a guiar o movimento visual, com células seletivas para a direção do movimento. Estas vias não são estritamente separadas e existe uma interligação substancial entre elas. A integração de entradas (inputs) sensoriais com sinais neuronais provenientes de áreas associativas cerebrais é um processo dinâmico que envolve a coordenação motora do sistema visual, assim como do movimento da cabeça, que visa adquirir o menor erro de predição da entrada sensorial pelas áreas associativas de acordo com o modelo de codificação preditiva [27]. Alterações neste fino processo pode alterar a interpretação da informação. Sabe-se que no TEA o processamento e a integração sensorial é atípica. Muitos estudos têm se dedicado a estudar este ponto, no entanto, a natureza exata desses problemas sensoriais permanece incerta. 1.4 Atenção Visual e Mapas de Saliência Constantemente recebemos uma grande quantidade de informações visuais e processar todos esses dados em tempo real seria uma tarefa difícil. A atenção visual, de maneira seletiva, direciona nosso olhar rapidamente para objetos de interesse em nosso ambiente visual. Essa capacidade de selecionar objetos salientes em uma cena visual desordenada confere uma vantagem evolutiva, possibilitando a detecção rápida de possíveis presas, companheiros ou predadores. O termo atenção visual é bastante amplo, incluindo muitas formas, como a atenção espacial, temporal, baseada em objetos e em características, sendo que essas não são mutuamente exclusivas. Podemos dividir a atenção visual em dois principais mecanismos, o bottom-up e o top-down. No bottom-up, estímulos externos atraem a atenção devido a suas características inerentes como, por exemplo, o contraste visual em relação ao fundo [36]. Neste mecanismo, a atenção é orientada por dados em nível de pixel, baseando-se em características da cena, como cor, orientação e intensidade. Por outro lado, o mecanismo top-down é impulsionado pela orientação interna da atenção, com base em conhecimentos prévios, planos e metas atuais, ou seja, utiliza informações semânticas ou relativas a um contexto. Ao

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

considerarmos padrões de atenção visual, podemos elaborar um mapeamento da importância relativa de cada pixel em uma imagem, conhecido como mapas de saliência [39], como exemplificado na figura 3. Os mapas de saliência são obtidos por Introdução:

Hipótese: O presente estudo visa superar essas limitações e comparar o modelo já usado pelo nosso grupo com algoritmos de classificação de mapas de saliência recentemente utilizados em TEA.

Os dados obtidos de pacientes seguirão as normas da Lei Geral de Proteção de Dados (LGPD), lei nº 13.709, de 14 de agosto de 2018, alterada pela lei nº 13.853, de 2019. A LGPD passa a vigorar partir de agosto de 2020, portanto, os dados serão anonimizados em publicações e utilizados apenas para fins de pesquisa. A classificação é realizada com base nas fixações do indivíduo. Para isso, o mapa de fixação é comparado aos mapas de saliência e será classificado com o rótulo do mapa de saliência que apresentar maior similaridade, ou seja, acertos. O critério de decisão será definido com base nos valores da curva Receiver Operating Characteristic (ROC), utilizando o índice de Youden[61]

Desfecho Primário: O classificador piloto será validado por meio da classificação de 50 novos pacientes sem repasse da classe. Na discriminação de indivíduos com e sem TEA, espera-se obter sensibilidade e especificidade comparáveis ou superiores ao estado da arte. Será avaliada a eficiência preditiva a partir do método de validação cruzada (k-fold cross-validation) para comparação estatística. As análises estatísticas serão realizadas em linguagem R, no IDE de domínio público RStudio Versão 1.2.5001 [62]. Desfecho Secundário: Após a adequação do classificador à ferramenta serão implementadas outras características valorizando a presença de características semânticas com possibilidade de cobertura específica do nível de pixel, objeto e semântico, de forma que a variação nesta composição possa auxiliar na classificação de subtipos de pacientes de acordo com mais ou menos características teoricamente associadas a conectividade cerebral do tipo bottom-up e top-down. As primeiras características testadas serão baseadas em Wang et al. [43], com extração semiautomática. Ao todo 13 características semânticas são sugeridas e apenas uma foi utilizada (rostro) no atual sistema de classificação, uma vez que as características iniciais derivaram dos estudos de Borji et al. [46] e Judd et al. [47]. Dentre as 12 características restantes, 5 delas estão presentes no vídeo estímulo e implementaremos em nosso classificador: 1 - Emoção (rostos com expressões emocionais); 2 - Objetos observados (objetos de cena visualizados, incluindo humanos); 3 Movimento (gestos significativos de posturas que implicam movimento); 4 - Objetos tocados (objetos com sensação tátil); 5 - Objetos rotulados (não incluídos em outras categorias). Após isso, a possibilidade de

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



Continuação do Parecer: 4.556.127

implementação de outras características será avaliada.

Está prevista a inclusão de 150 participantes HC-FMUSP.

Objetivo da Pesquisa:

Objetivo Primário: Desenvolver uma ferramenta classificatória para o TEA, baseado em modelos de atenção visual com integração de características em nível de pixel, objeto e semântico em mapas de saliência.

Objetivo Secundário: 1) Compor uma revisão de literatura para mapear a lacuna acerca do uso de características e mapas de saliência em modelos de atenção visual para a classificação do TEA com rastreamento do olhar; 2) Definir e implementar níveis de características, permitindo melhor compreensão do TEA; 3) Definir e implementar um modelo computacional com a adição de níveis de características, visando classificação do TEA; 4) Implementar uma interface para classificação automática de TEA para utilização de pesquisadores.

Avaliação dos Riscos e Benefícios:

Riscos: Os dados obtidos de pacientes seguirão as normas da Lei Geral de Proteção de Dados (LGPD), lei nº 13.709, de 14 de agosto de 2018, alterada pela lei nº 13.853, de 2019. A LGPD passa a vigorar partir de agosto de 2020, portanto, os dados serão anonimizados em publicações e utilizados apenas para fins de pesquisa. Assim, está pesquisa não oferece riscos ao paciente.

Benefícios: Desenvolver uma ferramenta classificatória para o TEA, baseado em modelos de atenção visual com integração de características em nível de pixel, objeto e semântico em mapas de saliência. Tal pesquisa implicará no auxílio ao diagnóstico precoce de TEA. Não há benefícios diretos para o participante.

Comentários e Considerações sobre a Pesquisa:

Os pesquisadores irão estudar o olhar de participantes com transtorno de espectro autista para validar um instrumento para o diagnóstico precoce desta classe de doenças. O estudo consiste em submeter esses participantes a assistir um vídeo e, depois, classificar as características do olhar. Com isso pretende-se avaliar padrões característicos encontrados nas pessoas com transtorno do espectro autista.

Esta é a resposta do pesquisador aos questionamentos da CAPPesq.

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



Continuação do Parecer: 4.556.127

Considerações sobre os Termos de apresentação obrigatória:

Os pesquisadores incluíram entre os termos de apresentação obrigatória:

1. Folha de rosto. Preenchida e assinada adequadamente;
2. Projeto completo;
3. TCLE que ainda precisa de alguns ajustes. Vide conclusões e pendências;
4. TALE. Adequado para o que se propõe;
5. Orçamento;
6. Cronograma.

Conclusões ou Pendências e Lista de Inadequações:

O estudo está bem elaborado e com métodos corretos, entretanto, ainda são necessários alguns esclarecimentos e adequações para que se possa dar o parecer final:

1. Os pesquisadores afirmam que “Riscos. Os dados obtidos de pacientes seguirão as normas da Lei Geral de Proteção de Dados (LGPD), lei nº 13.709, de 14 de agosto de 2018, alterada pela lei nº 13.853, de 2019. A LGPD passa a vigorar partir de agosto de 2020, portanto, os dados serão anonimizados em publicações e utilizados apenas para fins de pesquisa. Assim, **ESTÁ PESQUISA NÃO OFERECE RISCOS AO PACIENTE** (grifo do responsável pelo parecer. Isso também difere do que está escrito no TCLE. Segundo a resolução 466 de 2012 “item. V. Toda pesquisa com seres humanos envolve riscos em tipos e gradações” – Favor adequar.

RESPOSTA: A alteração encontra-se destacada na página 13 do projeto anexado.

“Riscos e desconfortos: O procedimento tem riscos mínimos como, por exemplo, cair da cadeira, sentir desconforto e ficar irritado pela novidade da tarefa. No caso de uma eventualidade durante o procedimento, contamos com a estrutura do Ambulatório PROTEA, bem como equipe preparada para o atendimento e, se necessário, encaminhamento a outro serviço especializado. O participante terá assistência imediata e integral no que se refere às complicações e danos decorrentes da pesquisa. Caso o participante se sinta desconfortável poderá desistir em qualquer momento do estudo, sem qualquer dano ou prejuízo. Você tem liberdade para retirar seu consentimento sem penalização alguma e sem prejuízo do cuidado que o participante recebe neste ou em outro serviço”.

CONCLUSÃO: PEDÊNCIA ATENDIDA.

2. TCLE: 2.1.

- a. “Despesas e compensações: Não há despesas pessoais para o participante em qualquer fase do

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

estudo, incluindo avaliações e consultas. Também não há compensação financeira relacionada à sua participação”. Indaga-se se haverá despesas com transporte e outras eventuais para o deslocamento dos participantes e acompanhantes e como serão ressarcidas? (resolução CNS 466 de 2012 – item IV .3 g).

RESPOSTA: As amostras dos grupos Caso (TEA) e Risco são coletadas a partir dos indivíduos que entram na triagem do ambulatório e portanto procuram o serviço para diagnóstico, o processo de diagnóstico do PROTEA é realizado por diferentes avaliações com equipe multidisciplinar e o retorno é dado aos pacientes por equipe de devolutiva constituída por médico psiquiatra infantil e psicóloga em período de até três meses, dependendo do número de avaliações feitas. As avaliações deste projeto são inseridas na rotina de diagnóstico do ambulatório. Para o grupo controle existem três possibilidades de avaliação, a equipe do projeto vai até o indivíduo para fazer a avaliação, quando o paciente vai na consulta da pediatria ele pode ser avaliado e finalmente pode ser contactado por telefone e se aceitar é convidado a vir para a avaliação. Para aqueles que se dispõem a vir ao PROTEA para realizar as avaliações recebem ajuda de custo no valor de R\$ 50,00 (cinquenta reais) conforme plano de trabalho aprovado.

CONCLUSÃO: PEDÊNCIA ATENDIDA.

b. Não foi localizada a afirmação de que o participante tem direito a indenização em decorrência de danos decorrentes da pesquisa (resolução CNS 466 de 2012 – item IV .3 h).

RESPOSTA: A alteração encontra-se destacada no TCLE anexado.

“Despesas e compensações: Não há despesas pessoais para o participante em qualquer fase do estudo, incluindo avaliações e consultas. Também não há compensação financeira relacionada à sua participação. Se existir qualquer despesa adicional, ela será absorvida pelo orçamento da pesquisa. O participante terá direito a indenização em caso de danos decorrentes da pesquisa.”

CONCLUSÃO: PEDÊNCIA ATENDIDA.

c. De acordo com a resolução 466 de 2012 Cap. V item 6 é necessário explicitar que o participante tem direito a assistência imediata e integral no que se refere às complicações e danos decorrentes da pesquisa.

RESPOSTA: A alteração encontra-se destacada no TCLE anexado.

“Riscos e desconfortos: O procedimento tem riscos mínimos como, por exemplo, cair da cadeira, sentir desconforto e ficar irritado pela novidade da tarefa. No caso de uma eventualidade durante o procedimento, contamos com a estrutura do Ambulatório PROTEA, bem como equipe preparada para o atendimento e, se necessário, encaminhamento a outro serviço especializado. O participante terá assistência imediata e integral no que se refere às complicações e danos

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

decorrentes da pesquisa. Caso o participante se sinta desconfortável poderá desistir em qualquer momento do estudo, sem qualquer dano ou prejuízo. Você tem liberdade para retirar seu consentimento sem penalização alguma e sem prejuízo do cuidado que o participante recebe neste ou em outro serviço".
CONCLUSÃO: PEDÊNCIA ATENDIDA.

Desse modo não há impedimentos éticos para a aprovação do projeto de pesquisa.

Considerações Finais a critério do CEP:

Em conformidade com a Resolução CNS nº 466/12 – cabe ao pesquisador: a) desenvolver o projeto conforme delineado; b) elaborar e apresentar relatórios parciais e final; c) apresentar dados solicitados pelo CEP, a qualquer momento; d) manter em arquivo sob sua guarda, por 5 anos da pesquisa, contendo fichas individuais e todos os demais documentos recomendados pelo CEP; e) encaminhar os resultados para publicação, com os devidos créditos aos pesquisadores associados e ao pessoal técnico participante do projeto; f) justificar perante ao CEP interrupção do projeto ou a não publicação dos resultados.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1601652.pdf	16/02/2021 10:24:56		Aceito
Outros	Carta_adequacao_comite.docx	16/02/2021 10:23:18	Helena Paula Brentani	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Felipe_adequacao_revisado.docx	16/02/2021 10:21:53	Helena Paula Brentani	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Comite_com_adequacoes.pdf	16/02/2021 10:18:04	Helena Paula Brentani	Aceito
Folha de Rosto	FR_Helena_assinado.pdf	06/01/2021 11:49:14	Helena Paula Brentani	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_22_12.docx	22/12/2020 15:41:23	Helena Paula Brentani	Aceito
Outros	Termo_para_Uso_de_Dados_do_Registro_Eletronico_de_Saude.docx	15/12/2020 22:01:51	Helena Paula Brentani	Aceito
Declaração de	Declaracao_mestrado_Thiago.pdf	15/12/2020	Helena Paula	Aceito

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS
CLÍNICAS DA FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.556.127

Pesquisadores	Declaracao_mestrado_Thiago.pdf	21:58:01	Brentani	Aceito
Declaração de Pesquisadores	Declaracao_compromisso_doutorado.pdf	15/12/2020 21:56:55	Helena Paula Brentani	Aceito
Outros	Declaracao_ausencia_de_custo.docx	15/12/2020 21:55:41	Helena Paula Brentani	Aceito
Outros	Check_List_Felipe.doc	15/12/2020 21:52:35	Helena Paula Brentani	Aceito
Outros	Carta_Projeto_Felipe.docx	15/12/2020 21:26:43	Helena Paula Brentani	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

SAO PAULO, 24 de Fevereiro de 2021

Assinado por:
ALFREDO JOSE MANSUR
(Coordenador(a))

Endereço: Rua Ovídio Pires de Campos, 225 5º andar

Bairro: Cerqueira Cesar

CEP: 05.403-010

UF: SP

Município: SAO PAULO

Telefone: (11)2661-7585

Fax: (11)2661-7585

E-mail: cappesq.adm@hc.fm.usp.br