

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES  
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

JÉSSICA DOS SANTOS DE OLIVEIRA

**Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional  
de atenção visual**

São Paulo

2018

JÉSSICA DOS SANTOS DE OLIVEIRA

**Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional  
de atenção visual**

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 07 de dezembro de 2017. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques

São Paulo

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

CRB 8 -4936

Oliveira, Jéssica dos Santos de

Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional de atenção visual / Jéssica dos Santos de Oliveira ; orientadora, Fátima de Lourdes dos Santos Nunes Marques. – 2018.

134 f.

Dissertação (Mestrado em Ciências) - Programa de PósGraduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo.  
Versão corrigida

1. Reconhecimento de padrões. 2. Inteligência artificial. 3. Processamento de imagens. 4. Percepção da face. 4. Autismo - Diagnóstico. I. Marques, Fátima dos Santos Nunes, orient. II. Título.

CDD 22.ed.– 006.4

Dissertação de autoria de Jéssica dos Santos de Oliveira, sob o título “**Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional de atenção visual**”, apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovado em 07 de dezembro de 2017 pela comissão examinadora constituída pelos doutores:

---

**Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques**  
Universidade de São Paulo  
Presidente

---

**Prof. Dr. Prof. Dr. Luciano Antonio Digiampietri**  
Universidade de São Paulo

---

**Prof. Dr. Orestes Vicente Forlenza**  
Universidade de São Paulo

---

**Prof. Dr. Fabrício Martins Lopes**  
Universidade Tecnológica Federal do Paraná

## **Agradecimentos**

Agradeço a todas as pessoas que fazem parte da minha vida e que me apoiaram na realização deste trabalho.

Agradeço a minha mãe, a minha vó e a minha família pelo apoio e incentivo aos estudos desde sempre em minha vida.

Agradeço à professora Fátima, que dedicou seu esforço e trabalho para me orientar, desde a graduação. E a professora Ariane, que compartilhou seus conhecimentos na minha coorientação. Ambas exemplos de mulheres a serem seguidos.

Agradeço ao pessoal do Instituto de Psiquiatria - IPq, por todo o conhecimento sobre o autismo. Em especial à Dra Helena e à Mirian, que auxiliaram na concepção do projeto.

Agradeço ao grupo PROTEA, em especial à Mirian e à Andrea, e aos colegas Tuany e Thiago pelo auxílio na coleta de dados.

Agradeço à Marta, à Marina e ao professor Orestes do LIM-27 pelas orientações e utilização do equipamento Tobii, que permitiu a coleta de dados.

Agradeço aos colegas de mestrado Vivian, Bárbara, Priscila, Davi, Renan e Tuany e aos demais colegas do LApIS, que compartilharam as experiências durante o mestrado, oferecendo consolo e companheirismo.

Agradeço aos colegas de trabalho e aos demais amigos da vida que ofereceram apoio e incentivo durante essa jornada.

Agradeço à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro, que permitiu realizar parte deste trabalho com dedicação integral.

*“Os olhos são a janela da alma e o espelho do mundo.”*

*(Leonardo da Vinci)*

## Resumo

Oliveira, Jéssica dos Santos de. **Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional de atenção visual**. 2018. 134 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2017.

O Transtorno do Espectro do Autismo (TEA) é um transtorno do neurodesenvolvimento que afeta principalmente as habilidades de interação social dos indivíduos. Para auxiliar no diagnóstico, estudos têm utilizado técnicas de rastreamento de olhar, mas, em geral, os sinais são interpretados manualmente ou a automatização considera somente alguns dos indícios que podem definir a presença do TEA em indivíduos. O presente projeto visa a preencher esta lacuna, propondo utilizar os dados do rastreamento do olhar para o desenvolvimento de um modelo de atenção visual para TEA e para Desenvolvimento Típico (DT). A partir desses modelos, é construído um classificador que visará auxiliar no diagnóstico do TEA. Para alcançar o objetivo proposto, foram realizadas as seguintes etapas: revisão bibliográfica, aprovação do comitê de ética, definição do modelo de atenção visual, implementação do modelo, análise e publicação dos resultados. Os métodos foram testados com técnica de validação cruzada e construção de curva ROC. Os resultados mostraram que o modelo de atenção visual desenvolvido é capaz de prever a atenção visual do grupo TEA e do grupo DT, e que o método para classificação desenvolvido consegue classificar um indivíduo com TEA com média de 90% de precisão e 83% de especificidade, atingindo no melhor resultado 96% de precisão e 93% de especificidade. Espera-se que o método possa ser utilizado por profissionais da área de saúde e que sirva de base também para outras aplicações.

Palavras-chaves: Autismo. Rastreamento de olhar. Aprendizado de Máquina. Modelo de Atenção Visual.

## Abstract

Oliveira, Jéssica dos Santos de. **Classifier to aid ASD diagnosis based on a computational model of visual attention**. 2018. 134 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2017.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects mainly social interaction skills of individuals. In order to aid the diagnosis, studies have used eye-tracking techniques, but, in general, the signals are interpreted manually or the automation considers only some of the evidence that may define the presence of the ASD in individuals considering eye-tracking signals. This project aims at contributing to this challenge, processing data provided from eye-tracking based on a visual attention model for ASD and Typical Development (TD). From this model its build a classifier that will aid ASD diagnosis. To achieve the proposed goal, the following steps were performed: literature review, approval of the ethics committee, definition of a visual attention model, model implementation, analysis and publication of results. The methods were tested with cross-validation technique and ROC curve construction. The results showed that the developed model is able to predict visual attention of the TEA group and the DT group, and that the developed training method can classify an individual with TEA with an average of 90% of precision and 83% of specificity. In the best result was achive 96% of accuracy and 93% of specificity. It is expected that the method can be used by health professionals and also serve as a basis for other applications.

Keywords: Autism. Eye tracking. Machine Learning. Model of visual attention.

## Lista de figuras

Figura 1 – Representação genérica de uma rede neural. . . . .	25
Figura 2 – Representação gráfica do SVM, com o hiperplano $H_1$ que delimita a região $f(x) = 1$ e o hiperplano $H_2$ que delimita a região $f(x) = -1$ . . . . .	26
Figura 3 – Representação gráfica da maldição da dimensionalidade. A partir de certo ponto, aumentar a dimensão faz com que o erro do classificador aumente devido ao aumento do erro de estimação. . . . .	27
Figura 4 – Imagem 1024 x 1024 <i>pixels</i> sub-amostrada até 32 x 32 <i>pixels</i> . . . . .	30
Figura 5 – Pirâmides de imagens de um <i>frame</i> gerada com quatro escalas e três orientações	32
Figura 6 – Características haar-like utilizadas no algoritmo de Viola e Jones (2001) . . .	33
Figura 7 – Imagem em que há um objeto em “destaque”, com maior saliência, exemplificando a atenção <i>bottom-up</i> . . . . .	34
Figura 8 – Diagrama dos modelos baseados em mapa de saliência de a) Koch e Ullman (1987) e b) Itti, Koch e Niebur (1998) . . . . .	36
Figura 9 – Uma ilustração hierárquica dos principais modelos de atenção visual e suas categorias. . . . .	39
Figura 10 – Ilustração do modelo de Peters e Itti (2007) . . . . .	40
Figura 11 – Ilustração do modelo de Borji (2012) . . . . .	42
Figura 12 – Diagrama de fluxo resumindo o processo de seleção dos estudos . . . . .	44
Figura 13 – Distribuição dos artigos incluídos considerando o tamanho da amostra de participantes nos experimentos. . . . .	46
Figura 14 – Distribuição dos artigos incluídos considerando intervalos de idades dos participantes nos experimentos . . . . .	47
Figura 15 – Distribuição dos métodos que mais se destacam entre os artigos incluídos. . .	48
Figura 16 – <i>Frames</i> das combinações presentes no vídeo do estímulo visual. . . . .	61
Figura 17 – Imagem mostrando uma criança realizando o procedimento de rastreamento do olhar posicionada no colo de um cuidador. . . . .	62
Figura 18 – Diagrama com as etapas desenvolvidas até a classificação de um indivíduo como TEA ou DT . . . . .	63
Figura 19 – Etapas de pré-processamento do vídeo até a extração de características . . .	64

Figura 20 – Exemplificação do processo de geração do mapa de fixação. Inicialmente são somados os valores das fixações de todos os indivíduos de um grupo (neste exemplo representados apenas por dois indivíduos). Após o somatório, é aplicado um filtro Gaussiano, que realizará uma suavização na imagem. A parte de cima da figura mostra a representação numérica das imagens. . . .	66
Figura 21 – Exemplos de Mapas de Fixação para o <i>frame</i> 198. Em (a) um mapa de fixação gerado a partir dos dados de indivíduos com DT, em (b) um mapa de fixação gerado a partir dos dados de indivíduos com TEA. . . . .	67
Figura 22 – Exemplificação da formação do vetor de características para um <i>frame</i> . Na figura é mostrado o exemplo de um <i>frame</i> utilizado, um mapa de fixação do grupo TEA e o exemplo de duas linhas do vetor de características finais, uma para a classe positiva e outra para a classe negativa. . . . .	70
Figura 23 – Exemplo de mapa de saliência: à esquerda um mapa de saliência de um <i>frame</i> gerado para o grupo DT; à direita os pontos em vermelho são os <i>pixels</i> que de fato foram fixados por esse grupo no <i>frame</i> . . . . .	71
Figura 24 – Representação gráfica de uma curva ROC com o índice de Youden (J), ponto que máxima a distância entre a curva e a linha de equidade (em cinza). . . .	73
Figura 25 – Fluxograma com o funcionamento do modelo proposto. São divididos em dois processos: Modelo de Atenção Visual e Classificador de TEA. As saídas dos processos, representadas em verde, são: para o modelo de atenção visual um modelo treinado, esse processo é repetido para cada um dos grupos, tendo ao final um modelo treinado para TEA e um modelo treinado para DT; para o classificador TEA a classe predita (1 para TEA e 0 para não-TEA), o classificador tem como entrada, além dos dados de fixação do indivíduo e do estímulo visual assistido, os modelos treinados no processo anterior. . . . .	75
Figura 26 – Gráfico de <i>box-plot</i> exibindo as distribuições das AUC's em cada um dos métodos executados. . . . .	82
Figura 27 – Curvas ROC obtidas para todos os modelos nas 5 execuções da validação cruzada . . . . .	84

## Lista de tabelas

Tabela 1 – Exemplo de um <i>template</i> que implementa o filtro gaussiano no domínio espacial de imagens com desvio padrão igual a 1 e x e y variando de -2 a 2 .	31
Tabela 2 – Algumas das principais métricas retornadas pelo equipamento Tobii TX300	59
Tabela 3 – Características selecionadas pelos dois métodos de seleção utilizados. . . .	79

## Lista de abreviaturas e siglas

ADI-R	<i>Autism Diagnostic Interview-Revised</i>
ADOS	<i>Autism Diagnostic Observation Schedule</i>
AUC	<i>Area Under the Curve</i>
DMS	<i>Diagnostic and Statistical Manual of Mental Disorders</i>
DT	Desenvolvimento Típico
FIT	<i>Feature Integration Theory</i>
IPq-USP	Instituto de Psiquiatria - Universidade de São Paulo
LApIS	Laboratório de Aplicações de Informática em Saúde
LCC	<i>Linear Correlation Coefficient</i>
NSS	<i>Normalized Scanpath Saliency</i>
QI	Quociente de Inteligência
ROC	<i>Receiver Operating Characteristic</i>
ROI	<i>Region of Interest</i>
SVM	<i>Support Vector Machine</i>
TEA	Transtorno do Espectro Autista
WTA	<i>Winner-take-all</i>
SVM	<i>Support Vector Machine</i>
PROTEA	Programa Transtorno do Espectro Autista
RNA	Redes Neurais Artificiais
GBVS	<i>Graph-Based Visual Saliency</i>

## Sumário

<b>1</b>	<b>Introdução</b>	15
1.1	<i>Objetivo e hipótese</i>	18
1.2	<i>Organização do documento</i>	18
<b>2</b>	<b>Aspectos conceituais</b>	20
2.1	<i>Transtorno do Espectro do Autismo - TEA</i>	20
2.1.1	<i>Padrões visuais atípicos no TEA</i>	23
2.2	<i>Aprendizado de máquina</i>	23
2.2.1	<i>Seleção de características</i>	27
2.3	<i>Processamento de imagens digitais</i>	29
2.3.1	<i>Filtros de imagem</i>	30
2.3.2	<i>Métodos de extração de características</i>	31
<b>3</b>	<b>Modelos de atenção visual</b>	34
3.1	<i>Taxonomias dos modelos de atenção visual</i>	36
3.2	<i>Modelos de classificação de padrões</i>	38
<b>4</b>	<b>Revisão bibliográfica sistemática</b>	43
4.1	<i>Equipamentos de rastreamento de olhar</i>	43
4.2	<i>Participantes no experimento</i>	45
4.3	<i>Técnicas e métodos</i>	48
4.3.1	<i>Tratamento de estímulos</i>	48
4.3.2	<i>Seleção e Análise de Dados</i>	49
4.4	<i>Parâmetros para auxílio ao diagnóstico</i>	51
4.5	<i>Avanços, limitações, oportunidades e conclusões</i>	53
4.5.1	<i>Métodos</i>	53
4.5.2	<i>Participantes dos experimentos</i>	54
4.5.3	<i>Estímulos</i>	55
4.5.4	<i>Conclusões da revisão sistemática</i>	55
<b>5</b>	<b>Procedimentos metodológicos</b>	57
5.1	<i>Etapas de pesquisa</i>	57

5.2	<i>Coleta de dados</i>	58
5.2.1	Equipamento para aquisição de dados	58
5.2.2	Participantes	59
5.2.3	Estímulo visual	60
5.2.4	Procedimentos	60
5.3	<i>Desenvolvimento</i>	61
5.3.1	Pré-processamento e extração de características dos vídeos	62
5.3.2	Construção dos mapas de fixação	66
5.3.3	Seleção de características	67
5.3.4	Seleção de <i>pixels</i>	68
5.3.5	Modelo de atenção visual	69
5.3.6	Método de classificação de indivíduos	71
5.3.7	Avaliação do modelo desenvolvido	72
5.3.8	Considerações finais	73
<b>6</b>	<b>Classificador de TEA baseado em um modelo de atenção visual para auxílio ao diagnóstico</b>	74
6.1	<i>Modelo de atenção visual</i>	74
6.2	<i>Classificador TEA versus DT</i>	77
<b>7</b>	<b>Resultados e Discussões</b>	79
7.1	<i>Seleção de características</i>	79
7.1.1	Algoritmo genético	79
7.1.2	Algoritmo <i>ReliefF</i>	80
7.2	<i>Desempenho dos classificadores</i>	81
7.2.1	Abordagens de Preparação de Dados	81
7.2.2	Comparação de métodos de seleção e aprendizado	82
7.3	<i>Experimento e coleta de dados</i>	86
7.4	<i>Limitações do Método</i>	87
7.5	<i>Avanços encontrados</i>	88
<b>8</b>	<b>Conclusão</b>	89
8.1	<i>Propósitos, Vantagens e Limitações</i>	89
8.2	<i>Trabalhos Futuros</i>	90

8.3	<i>Impactos</i> . . . . .	90
	<b>Referências<sup>1</sup></b> . . . . .	92
	<b>APÊNDICES</b>	<b>102</b>
	<b>Apêndice A – TCLE - Termo de Consentimento Livre e Esclarecido</b> . .	103
	<b>Apêndice B – Artigo de Revisão Sistemática</b> . . . . .	107
	<b>ANEXOS</b>	<b>130</b>
	<b>Anexo A – Parecer do Comitê de Ética</b> . . . . .	131

---

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

## 1 Introdução

O Transtorno do Espectro Autista (TEA), segundo [Baio \(2012\)](#), afeta uma entre 68 crianças nos Estados Unidos. Um estudo piloto brasileiro estimou uma prevalência de 0,3% em crianças de 7 a 12 anos de idade de um município típico do Estado de São Paulo ([PAULA et al., 2011](#)). Acredita-se que a estimativa é de dois milhões de casos no país ([RIBAS, 2013](#)).

O TEA é um transtorno do neurodesenvolvimento, caracterizado por deficiências no âmbito dos domínios da comunicação, interação social e por comportamentos e interesses que tendem a ser estereotipados, repetitivos e ritualísticos ([AMERICAN PSYCHIATRIC ASSOCIATION, 2014](#)).

Quanto mais cedo for iniciado o tratamento, maiores serão as chances do indivíduo com TEA apresentar melhoras significativas nas áreas afetadas ([DAWSON et al., 2010](#)). Porém, para que o tratamento possa ser iniciado é necessário que o diagnóstico também seja realizado precocemente.

Atualmente, os dois instrumentos usados como “padrão-ouro” para o diagnóstico são: o Sistema Diagnóstico de Observação do Autismo, conhecido como ADOS (*Autism Diagnostic Observation Schedule*) ([LORD et al., 2000](#)) e a Entrevista Diagnóstica de Autismo Revisada, conhecida como ADI-R (*Autism Diagnostic Interview-Revised*) ([LORD; RUTTER; COUTEUR, 1994](#)). A ADI-R consiste em uma entrevista com 93 itens aplicada ao cuidador. O ADOS consiste em um protocolo de observação comportamental. Esses instrumentos são aplicados em uma clínica por um profissional treinado e sua aplicação pode demorar até 2,5 horas para ser concluída ([WALL et al., 2012](#)).

Um ponto importante relacionado ao diagnóstico refere-se às questões socioeconômicas e culturais, para as quais esses instrumentos não foram projetados para se adaptar ([BERNIER; MAO; YEN, 2010](#)), tornando-os subjetivos. Como exemplo, um estudo realizado na Arábia Saudita mostrou que as meninas com TEA tiveram o diagnóstico realizado muito mais tarde do que os meninos. Os pesquisadores sugerem que esse atraso pode ser devido à sociedade esperar que os meninos sejam mais extrovertidos, enquanto aquelas meninas com déficit social estariam sendo consideradas apenas tímidas ([AL-SALEHI; AL-HIFTHY; GHAZIUDDIN, 2009](#)).

No Brasil, há uma escassez de instrumentos que possam ser usados de uma maneira ampla e livre, ou seja, faltam instrumentos que tenham sido traduzidos, adaptados, padronizados, validados e com direitos autorais liberados ([MARQUES; BOSA, 2015](#)). Os treinamentos para as

aplicações do ADI-R e do ADOS são caros, além de ser necessário o pagamento dos direitos autorais cada vez que um deles é aplicado.

No TEA, o contato visual atípico não é apenas um critério de diagnóstico ([AMERICAN PSYCHIATRIC ASSOCIATION, 2014](#)), mas também um dos primeiros sinais patológicos, pois pode ser utilizado como indício da presença do transtorno em indivíduos de praticamente qualquer idade ou nível funcional, iniciando por volta dos seis meses de idade. Adicionalmente, dependendo dos estímulos visuais utilizados, este indício pode ser observado independentemente do ambiente cultural no qual o indivíduo se encontra.

Alguns paradigmas visuais estão sendo analisados, a partir do rastreamento do olhar, para serem utilizados no diagnóstico do TEA considerando os indícios que definem o contato visual atípico. Como exemplo, o indivíduo com TEA apresenta atenção compartilhada prejudicada, isto é, não consegue seguir com o olhar a direção do olhar ou o gesto de apontar de outra pessoa.

Na revisão bibliográfica, relatada no capítulo 4, foram encontrados alguns trabalhos que definiram critérios para distinguir entre os grupos de indivíduos com e sem TEA utilizando padrões visuais atípicos. [Pierce et al. \(2011\)](#), [Shi et al. \(2015\)](#) e [Pierce et al. \(2015\)](#) utilizam critérios e estímulos similares, baseados no paradigma de movimento biológico, em que definem a duração total da fixação do olhar em movimentos não biológicos como critério de diagnóstico, porém [Pierce et al. \(2011\)](#) avaliam crianças de um a três anos, enquanto [Shi et al. \(2015\)](#) avaliam crianças de quatro a seis anos.

Quando [Shi et al. \(2015\)](#) aplicaram o paradigma criado por [Pierce et al. \(2011\)](#) que continha apenas uma criança no vídeo, perceberam que para diferenciar a atenção de crianças mais velhas era necessário acrescentar cenas interativas, ou seja, com mais crianças brincando entre si. Com essa abordagem, [Pierce et al. \(2015\)](#) conseguiram distinguir os grupos com uma sensibilidade de 21% e especificidade de 98% para um grupo de 334 crianças, enquanto [Shi et al. \(2015\)](#), utilizando uma curva característica de operação do receptor (ROC - do inglês *Receiver Operating Characteristic*) para medir os resultados, atingiu uma área sobre a curva (AUC - do inglês *Area Under de Curve*) de 0,86 com um grupo de 33 crianças.

[Chevallier et al. \(2015\)](#) utilizaram um estímulo em que são apresentadas apenas crianças brincando. Neste trabalho, foi considerada a duração total da fixação do olhar na face como critério de diagnóstico, obtendo uma AUC de 0,71 para um grupo de 81 crianças com idades entre seis e 17 anos.

Apesar do trabalho apresentado por [Shi et al. \(2015\)](#) ter uma AUC considerada como boa para a área de psiquiatria ([RICE; HARRIS, 2005](#)), sua amostra é pequena e contém crianças com

idades muito específicas, apenas de um a três anos, não podendo ser utilizado para o diagnóstico em diferentes fases da vida. Já em [Chevallier et al. \(2015\)](#), apesar de ter uma amostra mais abrangente, o trabalho não conseguiu resultados tão significativos se comparado com [Shi et al. \(2015\)](#). No presente projeto a amostra utilizada foi de 108 crianças, com idades de três a 18 anos e obteve-se uma AUC média de 0,82.

Uma das técnicas empregadas para melhorar a análise dos dados do rastreamento ocular, evitando, por exemplo que o especialista selecione a área a ser analisada no estímulo visual e com isso perca informações, é a implementação de modelos computacionais de atenção visual. Tais modelos fazem uma descrição do comportamento observado e/ou previsto da atenção visual ([TSOTSOS; ROTHENSTEIN, 2011](#)). Os trabalhos de [Neumann et al. \(2006\)](#), [Shic et al. \(2007\)](#) e [Freeth, Foulsham e Chapman \(2011\)](#) utilizam o modelo de atenção visual *bottom-up* de [Itti, Koch e Niebur \(1998\)](#) para comparar os mapas de saliência<sup>1</sup> gerados com a atenção visual atípica dos indivíduos com TEA. Já [Wang et al. \(2015b\)](#) definem um modelo de atenção visual que leva em consideração todos os pontos de fixação do olhar realizados por cada grupo de indivíduos, criando um mapa de fixação para o grupo com TEA e outro para o grupo de controle, o que permitiu comparar a atenção dos grupos em relação a determinados aspectos da imagem, conforme explicado na seção 4.3.2. Apesar das análises apresentadas por estes trabalhos serem significativas para o estudo do transtorno, nenhum deles utilizou um modelo de atenção visual para definir um critério que conseguisse auxiliar na classificação do TEA. É esta lacuna que o presente projeto pretende preencher, conforme detalhado nos capítulos posteriores.

Este projeto faz parte de um trabalho maior em andamento no Laboratório de Aplicações de Informática em Saúde (LApIS), sediado na Escola de Artes, Ciências e Humanidades (EACH-USP) em parceria com o PROTEA (Programa Transtorno do Espectro Autista) do Instituto de Psiquiatria da Universidade de São Paulo (IPq-USP) que visa a contribuir com o diagnóstico precoce de TEA. O projeto em questão se divide em vários módulos, abrangendo desde aspectos de gestação até aspectos sociais dos portadores de TEA. Um dos módulos se propõe a estudar coortes de alto risco e possíveis mecanismos biológicos associados, relacionando fatores ambientais ao risco de transtornos do neurodesenvolvimento, bem como estudar sistemas computacionais e sua contribuição para classificadores fenotípicos e de diagnóstico precoce para TEA. O presente projeto integra este módulo, oferecendo uma contribuição para o diagnóstico de autismo.

---

<sup>1</sup> Imagens, geralmente em tons de cinza, que representam as regiões mais salientes de uma imagem, isto é, as regiões que se sobressaem, que mais chamam a atenção.

### *1.1 Objetivo e hipótese*

O presente projeto de pesquisa tem como objetivo definir, implementar e validar um classificador, baseado em modelo de atenção visual, visando classificar se um indivíduo é portador do TEA ou não. Para alcançar tal objetivo foram estabelecidos os seguintes objetivos específicos:

- mapear os indícios da atenção visual atípica que auxiliam no diagnóstico do TEA;
- definir os estímulos, utilizando vídeos, de acordo com os indícios visuais atípicos, que melhor auxiliem na detecção do TEA;
- implementar e avaliar um modelo computacional de atenção visual que realize a predição da atenção visual do TEA e de DT;
- implementar e avaliar um classificador que consiga identificar, utilizando o modelo desenvolvido, se um indivíduo é portador de TEA;
- avaliar a eficiência da técnica utilizada com casos reais.

A hipótese deste projeto é “Utilizando um modelo computacional de atenção visual juntamente com técnicas de aprendizado de máquina para processar dados de rastreamento do olhar é possível classificar, com maior sensibilidade do que os trabalhos computacionais que não utilizam tais técnicas, se um indivíduo é portador do TEA”.

### *1.2 Organização do documento*

Além deste capítulo introdutório, o presente documento está organizado da seguinte forma: no capítulo 2 são apresentados os aspectos conceituais fundamentais para compreensão do trabalho, contendo informações básicas sobre o Transtorno do Espectro Autista, Aprendizado de Máquina e Processamento de Imagens. No capítulo 3 são apresentadas as informações conceituais sobre Modelos de Atenção Visual. No capítulo 4 é apresentada a revisão bibliográfica sistemática conduzida sobre o estado da arte em relação às pesquisas que utilizam rastreamento de olhar para análise do TEA. No capítulo 5, são apresentados todos os métodos analisados e etapas percorridas para o desenvolvimento do modelo almejado. No capítulo 6 é apresentado o modelo de atenção visual desenvolvido e do classificador de TEA. Por fim, no capítulo 7 são apresentados os resultados encontrados e as discussões sobre cada um deles. No anexo A encontra-se o termo de aprovação do Comitê de Ética e no apêndice A está o Termo de Consentimento Livre e

Esclarecido (TCLE). Finalmente, no apêndice B é apresentado o artigo resultante da revisão sistemática conduzida, cujos principais resultados estão apresentados no capítulo 4.

## 2 Aspectos conceituais

Este capítulo pretende elucidar os principais conceitos envolvidos neste projeto. A seção 2.1 apresenta conceitos relacionados ao TEA e os principais paradigmas visuais associados ao transtorno. A seção 2.2 apresenta conceitos sobre aprendizado de máquina necessários para a compreensão do desenvolvimento do modelo proposto. A seção 2.3 apresenta os conceitos básicos sobre processamento de imagens, necessários para o entendimento do projeto. Por fim, a seção 3 apresenta conceitos relacionados aos modelos de atenção visual e os principais modelos relacionados ao projeto.

### 2.1 *Transtorno do Espectro do Autismo - TEA*

O TEA é um transtorno do neurodesenvolvimento, que compromete, em diferentes níveis, duas áreas: comunicação social e comportamento (RUTTER, 2011; AMERICAN PSYCHIATRIC ASSOCIATION, 2014). Os critérios clínicos a serem preenchidos para o diagnóstico englobam déficits na comunicação social, compreendendo falta de reciprocidade social e dificuldade em fazer e manter amizades, de acordo com o estágio de desenvolvimento. Envolvem também padrões restritos e repetitivos de comportamento, manifestados por excessiva aderência a rotina, comportamentos ritualizados, interesses restritos, fixos e intensos e comportamentos sensoriais incomuns. Os sintomas são precoces, aparecendo no início da infância, mas podem ser mais bem observados com o aumento da demanda social (por exemplo, no início da vida escolar), e comprometem o indivíduo ao longo de toda a vida (AMERICAN PSYCHIATRIC ASSOCIATION, 2014).

A avaliação é clínica, feita a partir da anamnese com os pais ou cuidadores da criança, aliada à observação da mesma. Os instrumentos de diagnóstico “padrão-ouro” aceitos para o diagnóstico de TEA são a ADOS (LORD et al., 2000) e a ADI-R (LORD; RUTTER; COUTEUR, 1994). A ADIR é aplicada aos pais enquanto a ADOS é uma entrevista semiestruturada que engloba a observação da criança.

É importante ressaltar que dificuldades cognitivas estão presentes em TEA. Assim, a avaliação neuropsicológica contribui na determinação da capacidade funcional global (FRITH; FRITH, 2003; FOMBONNE, 2009; SUN; BUYS, 2012). A avaliação de fonoaudiologia também é importante, uma vez que cerca de 20-30% dos indivíduos com TEA são não verbais e, de 20

a 25% deles apresentam uma história de desenvolvimento de regressão marcada com perda da linguagem previamente adquirida (LORD; SHULMAN; DILAVORE, 2004).

Estudos têm proposto e mostrado a eficácia da intervenção precoce no TEA (LOVAAS, 1987; COHEN; AMERINE-DICKENS; SMITH, 2006). O *National Research Council* (LORD; MCGEE, 2001) aponta como característica para intervenção eficaz em crianças com TEA, o tratamento intensivo e com o início do programa de intervenção o mais cedo possível, evidenciando a necessidade de se buscar sinais precoces preditivos de TEA.

Estudos prospectivos com filmes caseiros pontuam que, no primeiro ano, crianças que posteriormente receberam o diagnóstico de TEA têm menos resposta ao nome e olham menos para a face. A partir do segundo ano as diferenças aumentam, com diminuição do apontar e de mostrar objetos (PALOMO; BELINCHÓN; OZONOFF, 2006). O comportamento de compartilhar atenção é um marcador importante para diferenciar crianças com TEA a partir dos 12 meses (BEDFORD et al., 2012) podendo ser considerado como um sinal precoce para o transtorno (CHAWARSKA; KLIN; VOLKMAR, 2003).

Os sintomas observados entre os 6 e 12 meses são: exploração visual e fixação visual incomuns, exploração de objetos de forma repetitiva, diminuição do contato ocular, da coordenação do direcionamento visual à voz, do engajamento, do interesse e do sorriso social (DAWSON et al., 2010). A partir dos dois anos, observa-se redução da atenção compartilhada, contato ocular atípico, redução da resposta ao nome e atraso de linguagem (FALCK-YTTER; BÖLTE; GREDEBÄCK, 2013). Estudos apontam a possibilidade de um diagnóstico confiável de TEA entre 18 e 24 meses de idade (GUTHRIE et al., 2013). Porém, segundo alguns autores, na prática, a idade média do diagnóstico é muito maior, por volta dos três anos de idade (MANDELL; NOVAK; ZUBRITSKY, 2005) ou até mesmo no início da idade escolar ou mais tarde (CAMARATA, 2014). Apesar das evidências de alterações no desenvolvimento neurobiológico no TEA começarem durante a fase pré-natal, Landa, Holman e Garrett-Mayer (2007) apontam que muitas vezes o diagnóstico raramente acontece antes dos 3 anos de idade. Klin, Klaiman e Jones (2015) afirmam que a idade do diagnóstico na comunidade ainda é muito tardia em relação à “janela de oportunidade” proporcionada pela neuroplasticidade nos primeiros anos de vida.

Dificuldades para a realização do diagnóstico precoce no TEA de forma mais ampla se devem à própria natureza complexa e multifatorial do transtorno, assim como à complexidade do processo diagnóstico que necessita de uma abordagem compreensiva, sistemática e baseada no trabalho de uma equipe multiprofissional (LANDA; HOLMAN; GARRETT-MAYER, 2007; KLIN et al., 2002). Além disso, o uso das ferramentas “padrão-ouro de diagnóstico”, como a

ADOS ou ADI-R, é muito limitado devido especialmente aos custos e à necessidade de um grande treinamento e especialização dos avaliadores (KLIN; KLAIMAN; JONES, 2015). Esforços têm sido envidados para encontrar marcadores biológicos seguros e, assim, auxiliar na detecção de sinais do TEA, mesmo antes dos sintomas comportamentais aparecerem (ELSABBAGH; JOHNSON, 2010). Neste contexto, observa-se um número crescente de estudos que empregam exames médicos ou mecanismos tecnológicos para auxílio à detecção precoce do TEA.

O uso de eletroencefalograma (EEG) e rastreamento de olhar têm sido destacados neste cenário, pois mostram diferenças significativas em crianças com TEA comparadas às crianças com desenvolvimento típico (DT). O EEG disponibiliza informações sobre o padrão de conectividade cerebral (OREKHOVA et al., 2014). Análises de EEG mostraram que bebês de 6 a 24 meses, irmãos de afetados, portanto com maior risco para TEA, apresentam um padrão de conectividade diferente de DT (BOSL et al., 2011). Também já se demonstrou a associação positiva com a resposta evocada de mudança do contato ocular dinâmico dos 6 aos 10 meses com o diagnóstico de TEA aos 36 meses (ELSABBAGH et al., 2012).

A tecnologia do rastreamento de olhar permite, de forma objetiva, avaliar diversos aspectos do olhar, fornecendo informações confiáveis a respeito do padrão visual do indivíduo. Tem a vantagem de não ser invasiva e poder ser realizada com indivíduos de idades e níveis de funcionamento variados. Vários estudos têm sido realizados e as análises definiram padrões visuais típicos de indivíduos com TEA. Os principais padrões são definidos na seção 2.1.1.

É importante ressaltar que o diagnóstico precoce contribui não somente com a identificação do transtorno, mas permite uma melhor compreensão do perfil da criança, nas suas forças e fraquezas cognitivas, na identificação de comportamentos mal adaptativos e comorbidades. Permite, assim, o planejamento de uma intervenção direcionada às necessidades dessa criança, com objetivo a longo prazo de melhorar os sintomas do TEA (GUTHRIE et al., 2013; CAMARATA, 2014; BRENTANI et al., 2013). Neste sentido, o uso de rastreamento de olhar vem contribuindo para ampliar a possibilidade de diagnóstico precoce. No entanto, não existem muitos estudos sobre técnicas de análise dos sinais fornecidos pelo rastreamento de olhar, uma área ainda carente de investigação.

### 2.1.1 Padrões visuais atípicos no TEA

Alguns padrões visuais atípicos no TEA são utilizados como critério de diagnóstico. Os principais, citados pelo Manual Diagnóstico e Estatístico de Transtornos Mentais (*Diagnostic and Statistical Manual of Mental Disorders - DSM*) são: o uso reduzido, ausente ou atípico de contato visual e a atenção compartilhada prejudicada. Nesta última categoria, os principais aspectos são a ausência do gesto de apontar, mostrar ou trazer objetos para compartilhar o interesse com outros e a dificuldade para seguir o olhar ou gesto de apontar de outras pessoas ([AMERICAN PSYCHIATRIC ASSOCIATION, 2014](#)).

Por meio de sistemas de rastreamento de olhar é possível analisar de forma mais detalhada os padrões citados. Na pesquisa realizada por [Falck-Ytter e Hofsten \(2011\)](#), que analisa o paradigma da atenção compartilhada prejudicada, percebe-se de forma mais clara a dificuldade do autista de acompanhar o interesse do outro, isto é, de seguir os acontecimentos sociais.

Os estudos que analisam a fixação do olhar na face, como o realizado por [Kliemann et al. \(2010\)](#), mostram que autistas olham menos para a face dos participantes da cena do que os indivíduos com desenvolvimento típico e, quando olham, focam mais na região da boca do que dos olhos, possivelmente devido ao movimento realizado pela boca.

Além desses paradigmas, outros estão sendo analisados por meio de sistemas de rastreamento de olhar, como o estudo realizado por [Klin et al. \(2009\)](#) que mostra que os autistas, diferentemente das crianças com desenvolvimento típico, não apresentam uma preferência em observar movimentos biológicos; ao contrário, tendem a olhar para movimentos e cenas não sociais. O trabalho de [Pierce et al. \(2011\)](#), detalhado no capítulo 4, confirma este paradigma ao mostrar que autistas fixam o olhar por um tempo significativamente maior em movimentos geométricos do que em movimentos biológicos.

## 2.2 Aprendizagem de máquina

Aprendizado de máquina pode ser definido como: “o ato de ‘ensinar’ computadores por meio de experiências, de forma que eles possam usar esse conhecimento para realizar outras tarefas” ([CONWAY; WHITE, 2012](#)). De forma genérica, aprendizado de máquina trata-se do processamento de um conjunto de dados a fim de encontrar um comportamento padrão entre esses dados.

Atualmente este tipo de técnica vem sendo aplicado a vários problemas, tais como pre-dizer se um paciente terá ou não uma doença com base em variáveis demográficas e medidas clínicas; identificação de números escritos à mão, por meio de imagens digitalizadas; agrupamento de genes que apresentam expressões semelhantes; entre outros (CONWAY; WHITE, 2012).

As técnicas utilizadas para a resolução desses problemas podem ser divididas em dois tipos: supervisionadas e não supervisionadas. As técnicas supervisionadas utilizam dados que possuem uma variável alvo, denominada rótulo, já definindo a que classe as observações pertencem. Já as técnicas não supervisionadas tentam agrupar as observações de acordo com as características, sem saber a que classes pertencem ou, em alguns casos, nem mesmo quantas classes existem (AGGARWAL, 2014). Neste projeto apenas algoritmos supervisionados serão utilizados, pois sabe-se, na população estudada, quem são as crianças com TEA e com DT.

Dois algoritmos diferentes foram utilizados para comparar a eficácia na classificação dos grupos: Redes Neurais Artificiais (RNA) e Máquina de Vetores de Suporte (SVM - do inglês *Support Vector Machine*). Esses métodos foram escolhidos por serem os mais utilizados na área de modelos computacionais de atenção visual e normalmente retornarem bons resultados de classificação (BORJI, 2012; JUDD et al., 2009; XU et al., 2014; KÜMMERER; WALLIS; BETHGE, 2016; KRUTHIVENTI; AYUSH; BABU, 2015).

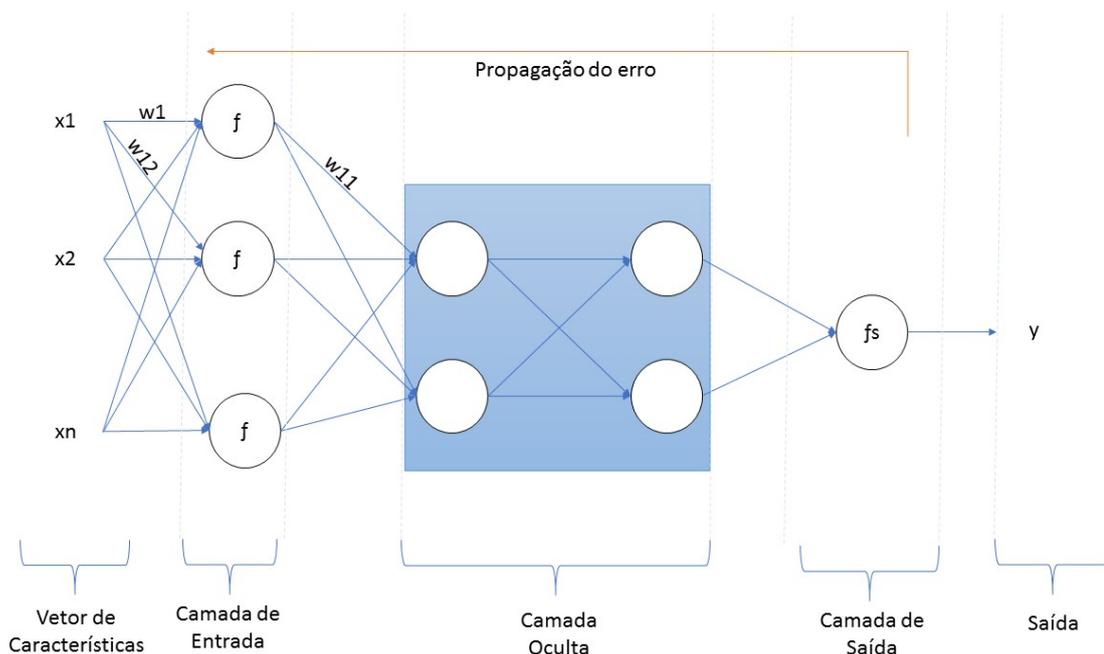
As RNA são baseadas na biologia, tendo como unidade principal o neurônio artificial, que simula o comportamento do neurônio biológico. No modelo computacional de um neurônio, os sinais interagem entre os neurônios, de acordo com o peso dado à relação entre eles. A ideia é que os pesos sejam aprendidos e controlem a força de influência de um neurônio em outro. Essa interação é modelada por uma função *net*, que geralmente assume a forma de uma soma ponderada. A taxa de disparo de um neurônio é modelada por uma função de ativação, que pode ter diferentes representações. Os três tipos básicos de função de ativação são: limiar, linear e sigmóide (HAYKIN; NETWORK, 2004; AGGARWAL, 2014).

A figura 1 apresenta a estrutura básica de uma RNA multicamadas, na qual os vetores das características são a entrada da rede, que interagem com os pesos do neurônio da camada de entrada de acordo com a função *net*. As saídas geradas pelas funções de ativação dos neurônios da camada de entrada passam pelos neurônios da camada oculta, também interagindo com seus pesos. Após chegarem na camada de saída, de acordo com o erro encontrado e com a condição de parada, a rede recomeça recalculando os pesos de acordo com a função de aprendizado, podendo propagar o erro para trás fazendo o caminho de volta (*backpropagation*) ou apenas da camada de

entrada para frente (*feedforward*) (FAUSETT, 1994). Após o aprendizado da rede, para calcular o valor de uma nova entrada, os cálculos são realizados em cima dos pesos já aprendidos.

A partir dessa estrutura básica, uma RNA pode variar quanto a: quantidade de neurônios, quantidade de camadas ocultas, função de ativação e função de aprendizado, forma de propagação e condição de parada, esta última sendo geralmente em relação ao erro mínimo ou à quantidade de épocas - execuções do início ao fim - que a rede atingirá. As especificações da rede utilizada neste projeto são apresentadas na seção 4.3.

Figura 1 – Representação genérica de uma rede neural.



Fonte: Jéssica dos Santos de Oliveira, 2017

Já o classificador SVM é definido da seguinte maneira: dadas duas classes e um conjunto de pontos que pertencem a estas, a SVM determina o hiperplano que os separa, de forma a colocar a maior quantidade possível de pontos da mesma classe do mesmo lado, ao mesmo tempo em que a distância de cada classe a esse hiperplano é maximizada (VAPNIK; VAPNIK, 1998). A SVM pode ser linear ou não linear. Quando não linear, pode assumir funções de kernel como polinomiais, gaussianas ou de tangente hiperbólica. Neste projeto apenas a SVM linear será utilizada, por ser a mais utilizada e normalmente atingir bons resultados (BORJI, 2012; JUDD et al., 2009; XU et al., 2014).

A SVM linear recebe como dados de entrada o vetor de características  $x$  e a classe correspondente  $y$ , que assume valores 1 ou -1. O objetivo é encontrar o hiperplano de margem

máxima que separe os pontos  $x$  com valores  $y = 1$  dos pontos do grupo que possuem  $y = -1$ , maximizando a distância entre os dois grupos. O hiperplano pode ser definido pela equação 1, em que  $w$  é o vetor normal ao hiperplano descrito e  $\frac{b}{\|w\|}$  corresponde à distância do hiperplano em relação à origem, com  $b \in \mathbb{R}$  (MÜLLER et al., 2001).

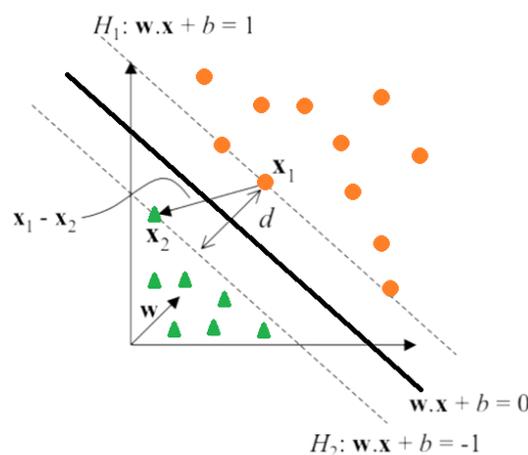
$$\text{Hiperplano} : wx + b = 0 \quad (1)$$

Considerando a equação 1 que define o hiperplano, uma função sigmoide empregada para a classificação, conforme descrito na equação 2

$$g(x) = \text{sgm}(f(x)) = \begin{cases} +1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ -1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b < 0 \end{cases} \quad (2)$$

A figura 2 apresenta uma representação gráfica da SVM, mostrando os hiperplanos  $H_1$  e  $H_2$ , que delimitam as regiões para  $f(x) = 1$  e  $f(x) = -1$ , respectivamente. A região limitada entre esses dois hiperplanos é chamada de margem e o hiperplano de margem máxima é o que fica entre os dois. As SVMs lineares podem ter margens rígidas ou suaves, que é quando se permite que alguns dados fiquem entre os hiperplanos  $H_1$  e  $H_2$  e também a ocorrência de alguns erros de classificação, o que é mais adequado a situações reais (MÜLLER et al., 2001).

Figura 2 – Representação gráfica do SVM, com o hiperplano  $H_1$  que delimita a região  $f(x) = 1$  e o hiperplano  $H_2$  que delimita a região  $f(x) = -1$ .

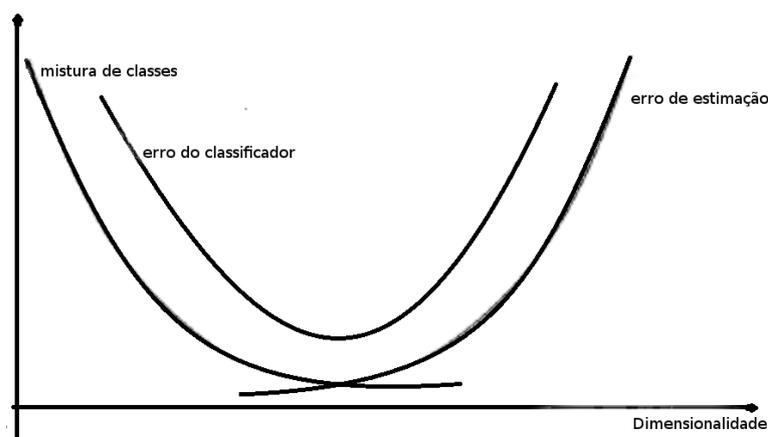


Fonte: Jéssica dos Santos de Oliveira, 2017

### 2.2.1 Seleção de características

Um dos grandes problemas encontrados em aprendizado de máquina refere-se à dimensão do vetor de características utilizado para o aprendizado. Nem sempre aumentar o número de características significa melhorar o desempenho do classificador; ao contrário, adicionar características que não são relevantes ou que estejam correlacionadas pode piorar o desempenho do classificador. Esse problema é chamado de “maldição da dimensionalidade”, termo que apareceu pela primeira vez em [Bellman \(1961\)](#), e que explica que funções de mais alta dimensão são mais complexas do que as de dimensões menores, exigindo que mais parâmetros sejam estimados. Por isso seriam necessários mais dados conhecidos, ou seja, mais amostras, para que seja possível solucionar o problema sem o decréscimo do erro de classificação ([DUDA; HART; STORK, 2012](#)). A figura 3 esquematiza essa relação.

Figura 3 – Representação gráfica da maldição da dimensionalidade. A partir de certo ponto, aumentar a dimensão faz com que o erro do classificador aumente devido ao aumento do erro de estimação.



Fonte: Jéssica dos Santos de Oliveira, 2017

Para tentar resolver este problema existem algoritmos para selecionar, dentre um vetor de características, aquelas que devem ser utilizadas para o aprendizado do classificador. Neste projeto foram utilizados dois algoritmos distintos de seleção, que também foram comparados com o aprendizado realizado com todas as características. Os algoritmos utilizados foram: *relief* e um algoritmo genético.

O *relief* é um algoritmo de seleção de características heurístico. De forma sucinta, ele funciona da seguinte forma: a partir de uma amostra de instâncias (dados) rotuladas (vetores de  $N$  características contendo a informação da classe), um peso  $W_j$  é calculado para cada característica  $j$  de forma que quanto maior o peso, maior a importância daquela característica na diferenciação

entre as classes. Tal peso representa o quão similares são os valores daquela característica para instâncias de uma classe e o quão diferentes são esses valores para instâncias de classes distintas. Para isso, o algoritmo seleciona aleatoriamente um subconjunto de tamanho  $nAmostra$  da amostra inicial e, para cada instância  $x$  deste subconjunto: 1) identifica os dois vizinhos mais próximo de  $x$ , um da mesma classe (*NearHit*) de  $x$  e um da outra classe (*NearMiss*) e 2) o peso da característica em questão ( $W_j$ ) é então decrementado no quadrado da diferença entre o valor da característica para a instância  $x$  e o valor dessa característica para seu *NearHit* e incrementado com o quadrado da diferença entre o valor da característica para a instância  $x$  e o valor dessa característica para seu *NearMiss*. Por fim, as características selecionadas são aquelas que ficam acima de um determinado limiar de peso (KIRA; RENDELL, 1992).

O pseudocódigo descrito no algoritmo 1 apresenta os passos para o *relief*:

---

**Algoritmo 1** *relief* pseudo-algoritmo

---

```

1: for  $i$  : 1 até  $nAmostra$  do
2:   Seleciona aleatoriamente uma instância  $x$ 
3:   Encontra NearHit e NearMiss de  $x$ 
4:   for  $j$  : 1 até  $N$  do
5:      $W_j \leftarrow W_j - diff(x_j, NearHit_j)^2 + diff(x_j, NearMiss_j)^2$ 
return  $W$ 

```

---

Algoritmos genéticos são baseados nas ideias evolucionistas de seleção natural e genética. São procedimentos de pesquisa probabilística projetados para funcionar em grandes espaços de dados, incluindo aqueles representados por *strings* (GOLDBERG; HOLLAND, 1988).

Inicialmente o algoritmo genético gera uma população de indivíduos de forma aleatória, cada indivíduo representando uma possível solução formada por um conjunto de bits, em que cada bit é um cromossomo. Na sequência, três processos ocorrem: seleção, que equivale à sobrevivência dos indivíduos mais fortes; reprodução (ou *crossover*), que equivale à mistura de dois pais; e mutação, que introduz modificações randômicas nos indivíduos. Para realizar a seleção é utilizada uma função de desempenho (ou *fitness*). O valor de desempenho dos cromossomos define quais deles serão selecionados para realizar a reprodução, de acordo com uma dada probabilidade. Após a reprodução, alguns indivíduos são mutados gerando uma nova população, o que constitui uma geração. Esse processo é repetido até uma determinada condição de parada (WHITLEY, 1994). Na utilização de algoritmos genéticos para seleção de características, cada possível solução é constituída por um vetor de características, a operação de *crossover* é realizada selecionando aleatoriamente os genes dos indivíduos pais, e a função *fitness* utiliza informação mútua, conforme descrito por Ludwig e Nunes (2010).

### 2.3 Processamento de imagens digitais

Uma imagem digital pode ser definida como uma função bidimensional de intensidade  $f(x, y)$ . Considerando o domínio espacial, uma imagem é uma matriz em que cada uma de suas células, chamadas *pixels* (*picture elements*), contém o valor de intensidade naquele ponto (GONZALEZ; WOODS, 2006; NIXON; AGUADO, 2012). Por sua vez, um vídeo é uma sequência temporal de imagens (TEKALP, 2015). As imagens que formam os vídeos são chamadas de *frames*, porém neste capítulo não será feita distinção entre uma imagem e um *frame*.

Os *pixels* que formam as imagens são os valores correspondentes à intensidade luminosa daquele ponto na cena. Esse valor pode ser representado por  $m$  bits. Assim, o valor de intensidade em geral é representado por um número inteiro no intervalo que varia de zero a uma potência de dois, considerando a quantidade  $m$  de bits como expoente. Por exemplo, uma imagem que usa oito bits para armazenar o valor de cada *pixel* considera os valores de 0 a 255 para cada *pixel*.

Para representação de imagens coloridas existem os modelos de cores, que permitem especificar cores em um formato padronizado, e geralmente variam de acordo com o *hardware* utilizado. Os modelos mais utilizados para representação de cores são: RGB (*red, green, blue*), CMY (*cyan, magenta, yellow*), CMYK (variante do modelo CMY, onde K denota o preto - *black*), YCbCr (padrão normalizado pela recomendação ITU-R BT.601 e utilizado em várias técnicas de compressão de vídeo), YIQ (padrão NTSC de TV em cores) e HSI (*hue, saturation, intensity*), às vezes também denominado HSV (*hue, saturation, value*). Neste projeto apenas o modelo RGB foi utilizado, por ser o modelo utilizado em monitores e câmeras de vídeo. No caso de imagens coloridas padrão RGB, é necessário uma função  $f(x, y)$  para cada uma de suas cores.

A resolução espacial de uma imagem determina o tamanho do menor objeto possível de ser identificado na imagem. Leva em consideração o tamanho do *pixel* da imagem, sendo especificada em pontos por polegada (*dots per inch - dpi*). Um valor típico de resolução é 72 dpi, suficiente para exibir uma imagem de 1024 x 1024 *pixels* em um monitor de 19 polegadas ou uma imagem de 640 x 400 pontos em uma tela cuja diagonal meça 12 polegadas. A figura 4 mostra a mesma imagem em diferentes resoluções.

A área de processamento de imagens digitais trata do processamento de imagens por meio do computador.

Figura 4 – Imagem 1024 x 1024 *pixels* sub-amostrada até 32 x 32 *pixels*



Fonte: (GONZALEZ; WOODS, 2006)

Uma forma de classificar as técnicas de processamento de imagens é considerar o nível de processamento como baixo, médio e alto. O processamento de baixo nível consiste em técnicas matemáticas ou de operação lógica para realizar tarefas simples como reduzir ruídos ou melhorar a qualidade da imagem. O processamento a nível médio combina essas operações de baixo nível para extrair alguns atributos da imagem, como segmentação e contorno. Já o processamento de alto nível é utilizado para interpretar e reconhecer objetos na cena, como pessoas, por exemplo (GONZALEZ; WOODS, 2006).

As subseções a seguir explicam conceitos básicos relacionados a imagens e alguns métodos de baixo, médio e alto nível de processamento, utilizados neste projeto para extração e reconhecimento de informações na imagem.

### 2.3.1 Filtros de imagem

Técnicas de filtragem em uma imagem podem ser tratadas no domínio espacial ou da frequência da imagem. As técnicas no domínio espacial atuam diretamente sobre a matriz de pixels, aplicando um *template*, ou máscara de convolução, na imagem. Existem vários tipos de filtros para diferentes aplicações na imagem, os mais comuns são os filtros para realce e os para suavização da imagem (Marques Filho; NETO, 1999).

Os filtros para realce destacam detalhes finos na imagem. Já os filtros para suavização eliminam pequenos detalhes da imagem, destacando grandes estruturas presentes na mesma, esse processo é geramente chamado de borramento da imagem. Os filtros de suavização também podem ser utilizados com o intuito de remover ruídos nas imagens (Marques Filho; NETO, 1999).

Há várias filtros utilizados para suavização: filtro de média, mediana, múltiplas imagens e filtro gaussiano. Este último foi o escolhido para ser utilizado neste projeto, mais especificamente na etapa de geração dos mapas de fixação, descrita na seção 5.3.2.

O filtro gaussiano é obtido através da geração de uma matriz gaussiana que será utilizada como máscara na convolução da imagem. Esta matriz é obtida através da equação 3, atribuindo os valores de  $x$ ,  $y$  e  $\alpha$ . Desta forma, para uma matriz de dimensão 5, com desvio padrão igual a 1 e ambos,  $x$  e  $y$  variando entre -2 e 2, tem-se a estrutura matricial constituída dos valores mostrados na tabela 1 (JESUS; Costa Jr., 2015).

$$G(x, y) = \frac{1}{2\pi\alpha^2} e^{-\frac{x^2+y^2}{2\alpha^2}} \quad (3)$$

Tabela 1 – Exemplo de um *template* que implementa o filtro gaussiano no domínio espacial de imagens com desvio padrão igual a 1 e  $x$  e  $y$  variando de -2 a 2

0,00078631	0,00655952	0,01330347	0,00655952	0,00078631
0,00655952	0,05472049	0,11097944	0,05472049	0,00655952
0,01330347	0,11097944	0,22507904	0,11097944	0,01330347
0,00655952	0,05472049	0,05472049	0,05472049	0,00655952
0,00078631	0,00655952	0,01330347	0,00655952	0,00655952

### 2.3.2 Métodos de extração de características

Nesta seção são apresentados alguns dos principais métodos de extração de características utilizados neste projeto.

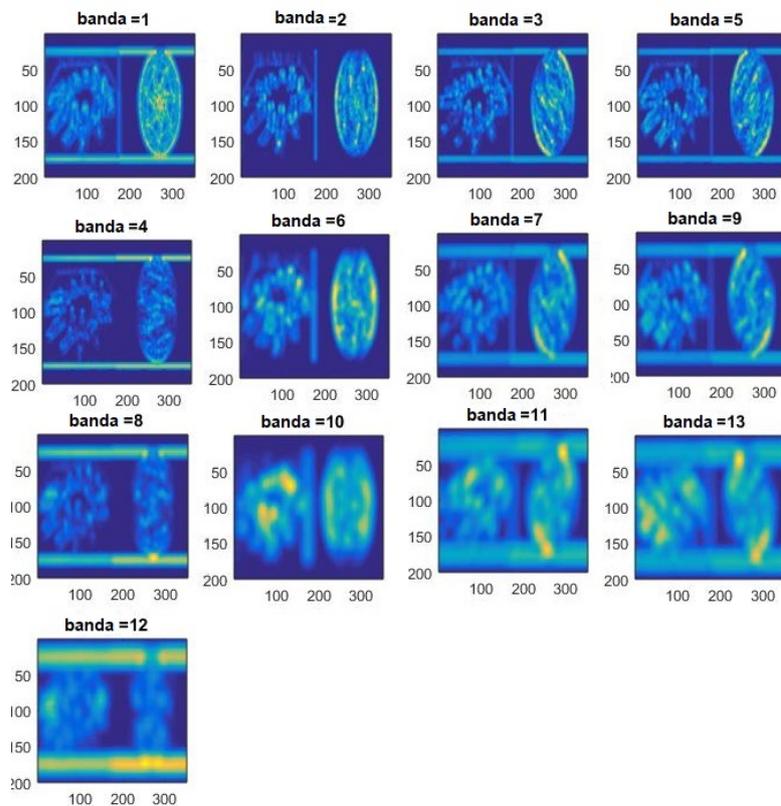
- **Processamento multiresolução: pirâmides de imagens**

As pirâmides de imagens consistem em um conjunto de imagens de resoluções espaciais cada vez menores, organizadas no formato de uma pirâmide. Quanto mais alto o nível na pirâmide, menor a resolução da imagem.

Há três técnicas principais para criação de pirâmides de imagens: gaussiana, em que os níveis são construídos por filtros gaussianos aplicados de forma sucessiva; laplaciana: similar à gaussiana, mas guarda a diferença entre as imagens de cada nível e, orientáveis (*steerable*), que são decomposições lineares da imagem em multiescala e multiorientação .

A figura 5 mostra as imagens extraídas de uma pirâmide de imagens orientável em quatro escalas e três orientações.

Figura 5 – Pirâmides de imagens de um *frame* gerada com quatro escalas e três orientações



Fonte: Jéssica dos Santos de Oliveira, 2017

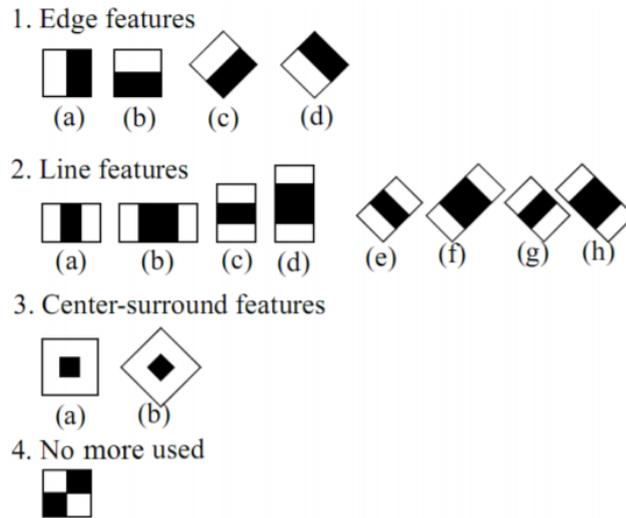
- **Detecção de faces e pessoas**

Várias informações alto nível podem ser extraídas de imagens, como pessoas, faces, carros, animais, etc. Essas informações têm aplicações variadas, sendo utilizadas desde a área de segurança, com detectores de movimento, até na área de robótica. Aqui, as informações utilizadas são faces e pessoas, selecionadas devido à diferença na atenção biológica entre pessoas com TEA e com DT. Apesar de existirem diversos métodos para essa tarefa, como o detector de pessoas de Felzenszwalb ([FELZENSZWALB; MCALLESTER; RAMANAN, 2008](#)), neste projeto foi utilizado o algoritmo de [Viola e Jones \(2001\)](#), mais conhecido na área.

O algoritmo de [Viola e Jones \(2001\)](#) utiliza-se de características baseadas em *Haar wavelets*. Essas características retornam um valor único obtido pela soma dos valores brancos menos a soma dos valores pretos dos *pixels* que ficam sobre os retângulos representados na figura 6.

Normalmente o algoritmo é utilizado para reconhecer faces em imagens, neste projeto também foi utilizado para reconhecer pessoas. Para o funcionamento do algoritmo são necessários muitos exemplos positivos e muitos exemplos negativos para aprender a

Figura 6 – Características haar-like utilizadas no algoritmo de Viola e Jones (2001)



Fonte: (VIOLA; JONES, 2001)

reconhecer o objeto em questão. Porém, aplicar todas as características de *Haar* em toda a imagem de cada um dos exemplos demandaria muito processamento computacional. Para melhorar o processamento foi introduzido o conceito de classificador em cascata. Inicialmente o algoritmo descarta janelas em que não há chances de ter o objeto em questão, para isso as características são separadas em grupos, que são aplicados em diferentes estágios. Se uma janela falhar no primeiro estágio, as próximas características não são aplicadas a ela, e ela é descartada; se falhar no segundo estágio, o mesmo ocorre de forma sucessiva. O algoritmo contém mais de 6000 características, para uma janela de 24x24, divididas em 38 estágios, sendo crescente o número de características em cada estágio. Os cinco primeiros estágios contêm 1, 10, 25, 25 e 50 características, respectivamente. As características são selecionadas utilizando o algoritmo *AdaBoosting* (VIOLA; JONES, 2001).

### 3 Modelos de atenção visual

O ser humano recebe muitas informações no seu campo visual a cada segundo, e processar todas essas informações seria muito custoso (BORJI; ITTI, 2013). Assim, se torna necessário um mecanismo que consiga selecionar os estímulos de maior interesse no ambiente visual. Este é o papel da atenção visual (TREISMAN; GELADE, 1980; BORJI; ITTI, 2013; ITTI; KOCH, 2001). Este mecanismo tem um importante significado evolutivo, pois permite detectar rapidamente uma possível presa, companheiros ou predadores no mundo visual (ITTI; KOCH, 2001).

Há duas categorias básicas de fatores que orientam a atenção visual: *bottom-up* e *top-down*. Fatores *bottom-up* são baseados principalmente em características da cena, como cor, orientação e intensidade (NOTHDURFT, 2005). São rápidos, involuntários e geralmente *feed-forward* (sem realimentação). Já fatores *top-down* são determinados por fenômenos cognitivos como o conhecimento, as expectativas, recompensas e metas atuais. São geralmente mais lentos, orientados a tarefa, voluntários e *closed-loop* (ciclo fechado) .

A figura 7 mostra um exemplo que ilustra as categorias citadas (*bottom-up* e *top-down*). Ao olhar para a figura, a atenção é automaticamente direcionada para a camisa vermelha, neste caso é definido que a saliência da roupa vermelha é maior do que o restante da imagem – este é um típico exemplo de atenção visual *bottom-up*. Entretanto, se a pergunta “Qual a cor do armário?” for realizada, a atenção será direcionada para o armário e não mais para a roupa vermelha, pois haverá uma tarefa a ser cumprida. Este é um exemplo de atenção *top-down*.

Figura 7 – Imagem em que há um objeto em “destaque”, com maior saliência, exemplificando a atenção *bottom-up*.



Fonte: Adaptado de Rigo (2010)

Para tentar explicar o funcionamento da atenção visual humana, Treisman e Gelade (1980) desenvolveram a Teoria das Características Integradas (*Feature Integration Theory* – FIT).

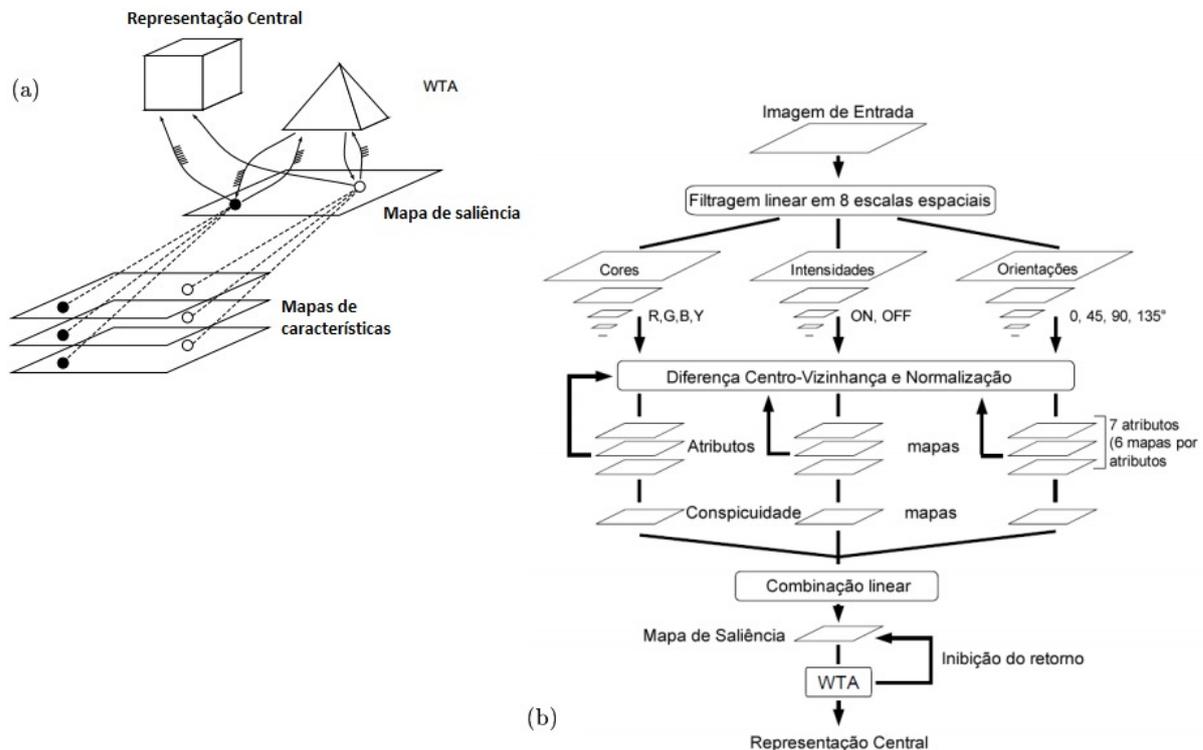
Essa teoria afirma que os estímulos captados no campo visual são processados em dois estágios. No primeiro estágio são organizados mapas mentais de características básicas da cena, como cor, forma e tamanho. Com a junção do resultado desses mapas de características é formado um mapa mestre que mostra onde estão os objetos, mas não o que eles são. Por meio deste mapa são selecionados os locais onde se deve focar a atenção, e só depois os objetos dessas regiões passam pela fase de reconhecimento.

Baseado na FIT, Koch e Ullman (1987) criaram o primeiro modelo computacional de atenção visual amplamente aceito. Um modelo de atenção visual pode ser definido como uma descrição do comportamento observado e/ou previsto da atenção visual primata humano ou não-humano (TSOTSOS; ROTHENSTEIN, 2011). No modelo de Koch (figura 8(a)), características visuais como cor, orientação e intensidade, são computadas em paralelo em um conjunto de mapas de características pré-atentivas. Todos os mapas de características são combinados originando um mapa topográfico de saliências. E, no final do processo, uma rede *winner-take-all* (WTA) detecta o local mais saliente e dirige a atenção para ele, de modo que apenas características deste local alcancem uma representação mais central para posterior análise.

O modelo de Itti, Koch e Niebur (1998) (figura 8(b)) foi o primeiro a fazer uma implementação e verificação completa do modelo de Koch e Ullman. Possui três características básicas: cor, intensidade e orientação. Estas características visuais são calculadas usando um filtro linear composto por oito escalas espaciais, de 1:1 até 1:256. Na sequência, é realizado o cálculo de diferenças centro-vizinhança, que computam o contraste espacial local para cada uma das características resultando em um total de 42 mapas: seis de intensidade, 12 de cores e 24 de orientação. Um esquema de inibição lateral interativa inicia uma competição por relevância dentro de cada mapa de característica. Depois da concorrência, os mapas de características são combinados em um único “mapa de conspicuidade” para cada tipo de característica, ou mapa de saliência em que as regiões mais claras representam as regiões em que há mais chances de receberem atenção. Os mapas de conspicuidades são somados gerando um único mapa de saliência. Para detectar o ponto mais saliente do mapa é utilizada uma rede WTA.

O modelo de Itti, Koch e Niebur (1998), apesar de amplamente utilizado, é um modelo estritamente *bottom-up*. Os modelos *top-down*, apesar de estarem disponíveis em menor quantidade, também vêm sendo estudados nas últimas décadas. Segundo Borji e Itti (2013), três grandes fontes de influências são exploradas pelos modelos *top-down*: características da imagem, contexto e tarefas. Alguns modelos abordam a busca visual baseando-se nas características do objeto que está sendo procurado. Por exemplo, quando o alvo é um objeto vermelho, um peso

Figura 8 – Diagrama dos modelos baseados em mapa de saliência de a) Koch e Ullman (1987) e b) Itti, Koch e Niebur (1998)



Fonte: Adaptado de Itti (2000)

maior seria atribuído à cor vermelha. Os trabalhos de Frintrop (2006) e Navalpakkam e Itti (2006) são exemplos de trabalhos que utilizam essa abordagem.

Outros modelos investigam o papel do contexto ou essência da cena para restringir os locais que o olhar se fixa. Entre os modelos que utilizam este tipo de informação estão os trabalhos de Oliva e Torralba (2001) e Walker e Malik (2002). Por fim, alguns modelos se concentram exclusivamente nas exigências de determinadas tarefas, por exemplo, analisando qual o foco visual quando se está dirigindo e qual o efeito do movimento dos olhos ao ajustar o rádio durante essa tarefa. Neste contexto estão os trabalhos de Hayhoe e Ballard (2005) e Peters e Itti (2007).

### 3.1 Taxonomias dos modelos de atenção visual

Borji e Itti (2013) classificam os modelos de atenção visual de acordo com o mecanismo utilizado para obter a saliência. De acordo com essa característica, são definidas oito categorias explicadas a seguir e ilustradas na figura 9.

1. **Modelos Cognitivos:** quase todos os modelos de atenção estão direta ou indiretamente inspirados em conceitos cognitivos. Estes modelos foram os primeiros a abordar o problema de algoritmos computacionais de saliência nas imagens digitais. O modelo de [Itti, Koch e Niebur \(1998\)](#) tem sido a base de modelos posteriores e o padrão de referência para comparação.

Modelos cognitivos têm a vantagem de expandir a visão sobre as bases biológicas da atenção visual. Isso ajuda ainda mais no entendimento dos princípios computacionais ou mecanismos neurais deste processo, bem como outros processos complexos dependentes, tais como o reconhecimento de objetos.

2. **Modelos Bayesianos:** nestes modelos, o conhecimento prévio (por exemplo, o contexto ou a essência da cena) e as informações sensoriais (por exemplo, características do alvo) são probabilisticamente combinados de acordo com a regra de Bayes (por exemplo, para detectar um objeto de interesse). Um dos principais benefícios de modelos Bayesianos é sua capacidade de aprender a partir de dados e sua capacidade de unificar muitos fatores em forma de princípios.
3. **Modelos baseados na Teoria da Decisão:** o ponto principal destes modelos é que a atenção visual deve ser impulsionada pela otimização em relação à tarefa final. [Gao e Vasconcelos \(2004\)](#) argumentam que para o reconhecimento, características marcantes são aquelas que melhor distinguem uma classe de objetos de interesse de todas as outras classes. Esses modelos têm sido muito bem sucedidos em aplicações de visão computacional, tais como classificação e predição da fixação do olhar. Algoritmos Bayesianos podem ser considerados um tipo especial de modelos teóricos de decisão.
4. **Modelos baseados na Teoria da Informação:** estes modelos baseiam-se na premissa de que a saliência computacional localizada serve para maximizar a informação de uma amostra do ambiente. Eles trabalham selecionando as partes mais informativas de uma cena e descartando o restante. Estes modelos atribuem valores superiores a regiões da cena com características raras. Embora, em teoria, seja viável a utilização de qualquer espaço de características, esses modelos geralmente utilizam um conjunto esparsos de funções de base aprendido a partir de um repositório de cenas naturais.
5. **Modelos baseados em Grafos:** um modelo baseado em grafo é uma estrutura probabilística em que um grafo denota a estrutura de independência condicional entre variáveis aleatórias. Modelos de atenção nesta categoria tratam os movimentos dos olhos como uma série temporal.

Esses modelos podem ser vistos como uma versão generalizada de modelos Bayesianos. Isto lhes permite modelar mecanismos de atenção mais complexos ao longo do espaço e do tempo, o que resulta em um bom poder de previsão. Os inconvenientes residem na complexidade do modelo, especialmente quando se trata de treinamento e legibilidade.

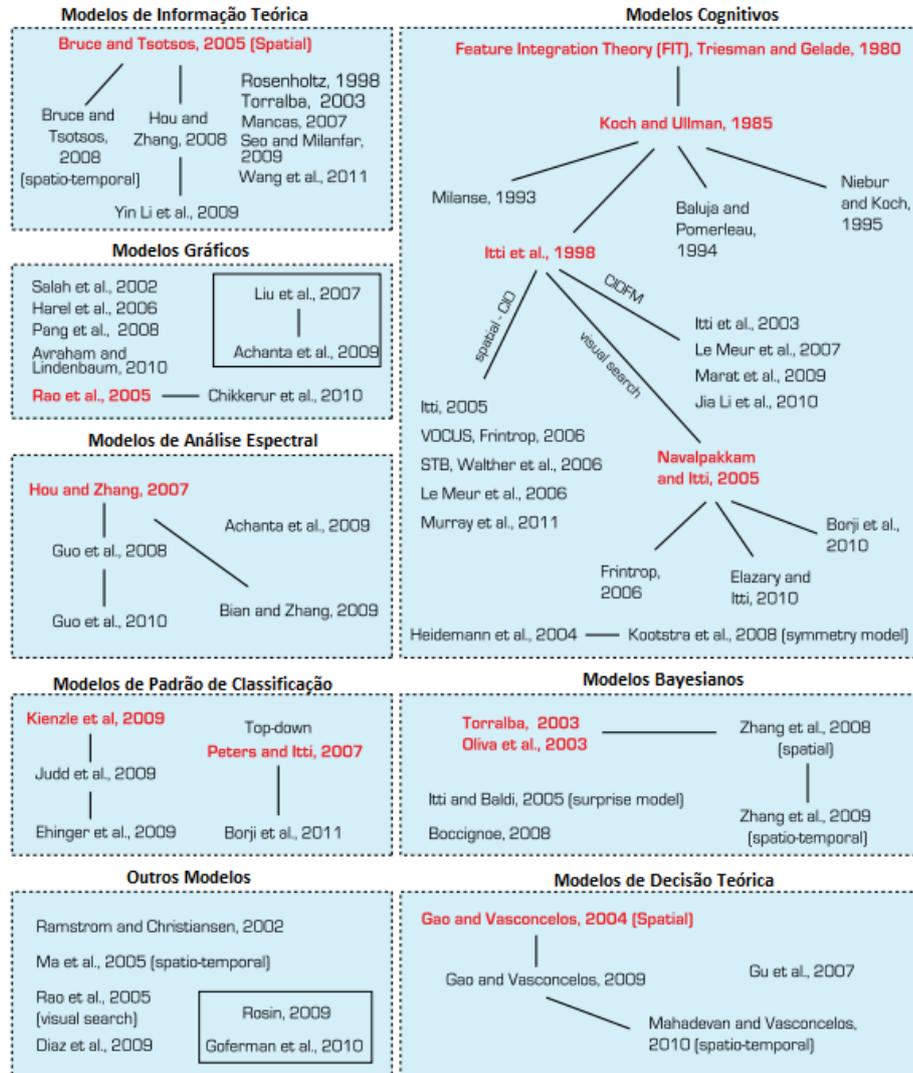
6. **Modelos de Análise Espectral:** ao invés de processar uma imagem no domínio espacial, modelos nesta categoria derivam a saliência no domínio da frequência. Modelos de análise espectral são simples de explicar e implementar. Embora bem sucedidos, a plausibilidade biológica desses modelos não é muito clara.
7. **Modelos de Classificação de Padrões:** os modelos desta categoria usam técnicas de aprendizado de máquina para aprender a saliência dos estímulos por meio de registros dos movimentos do olhar ou de rótulos nas regiões salientes. Várias técnicas diferentes podem ser utilizadas para tal objetivo, por exemplo: *patches* de imagem (KIENZLE et al., 2006), essência da cena (PETERS; ITTI, 2007), vetor de características em cada *pixel* (JUDD et al., 2009) ou mapas de saliências (BORJI, 2012) podem ser utilizados para extrair os dados da cena. SVM clássico e classificadores de Regressão podem ser usados para aprender a saliência. Alguns desses métodos são detalhados na seção 3.2.
8. **Outros Modelos:** existem alguns outros modelos que não se enquadram facilmente nessa categorização. Por exemplo, modelos que utilizam operadores de simetria para medir a saliência (KOOTSTRA; NEDERVEEN; BOER, 2008) ou que utilizam auto semelhança da estrutura de imagem local para a detecção de saliência (SEO; MILANFAR, 2009).

### 3.2 Modelos de classificação de padrões

Os modelos de classificação de padrões utilizam o movimento do olhar ou o rótulo das regiões salientes, combinados com algumas características, para aprender a prever a atenção visual humana. Por este motivo, o modelo a ser desenvolvido neste trabalho fará parte desta categoria e utilizará um vetor de características para guardar as informações extraídas da imagem. Nesta seção serão explicados alguns dos principais modelos que utilizam esse tipo de abordagem e que mais se assemelham ao modelo proposto.

O método de Peters e Itti (2007), ilustrado na figura 10, gera um vetor de características com baixa dimensão, que é definido como a “essência” da cena e combina esse vetor com os

Figura 9 – Uma ilustração hierárquica dos principais modelos de atenção visual e suas categorias.



Fonte: Adaptado de Borji e Itti (2013)

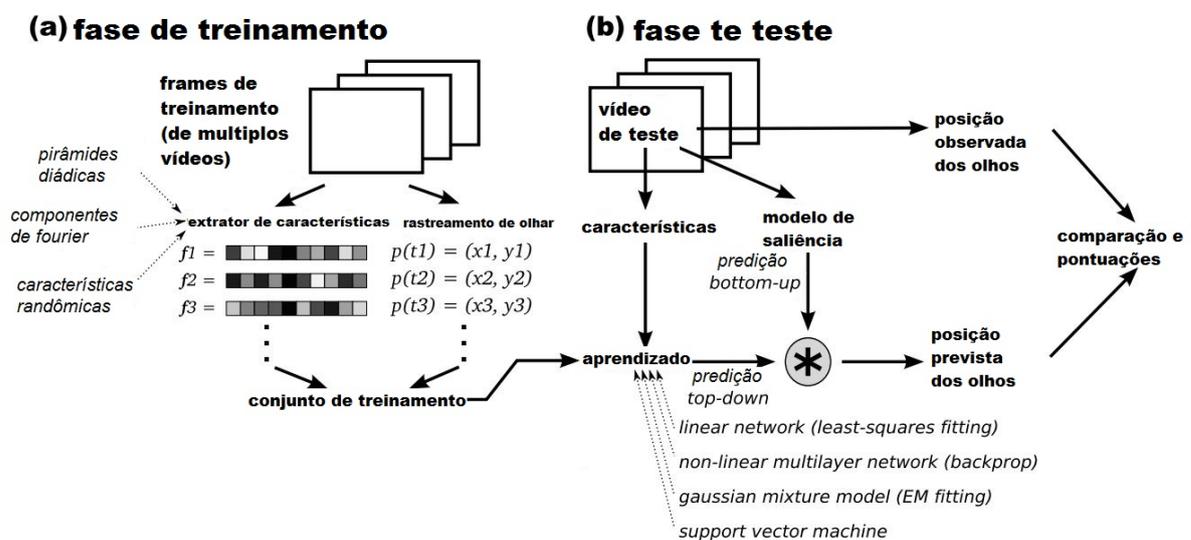
dados da posição do olhar. Duas abordagens são utilizadas para gerar vetores de características em Peters e Itti (2007): baseada em pirâmides e baseadas no método de Fourier.

A primeira abordagem baseia-se em sete pirâmides de características: luminância, oponência de cor vermelho/verde e azul/amarelo e quatro orientações. Para cada uma dessas pirâmides são extraídas duas escalas: uma mais fina e outra mais grossa. Em cada uma das pirâmides de escala é aplicada uma máscara 4x4 de média e outra de variância. No total o vetor de características contém 448 elementos.

O segundo vetor de características é baseado no método de Fourier. Para cada imagem é calculado o logaritmo da amplitude da Transformação rápida de Fourier (FFT - em inglês *Fast Fourier Transform*) da luminância da imagem. A matriz resultante é reamostrada para que os eixos passem a representar orientação e frequência espacial. O vetor de características consistiu

de 384 ( $24 * 16$ ) elementos, cada um representando a energia média entre 24 faixas de frequência espacial e 16 faixas de orientação. Também foi gerado um vetor de características aleatório, a fim de verificar se as características eram significativas. Para estimação, foi utilizado o método dos mínimos quadrados. Para avaliar o quão boa foi a predição do modelo, [Peters e Itti \(2007\)](#) utilizaram a Saliência *Scanpath* Normalizada (NSS - em inglês *Normalized Saliency Scanpath*). O vetor de características gerado pelo método de *Fourier* combinado com o modelo *bottom-up* teve o melhor desempenho em prever as posições dos olhos.

Figura 10 – Ilustração do modelo de [Peters e Itti \(2007\)](#)



Fonte: Adaptado de [Peters e Itti \(2007\)](#)

Em [Judd et al. \(2009\)](#) os dados de fixação de 15 indivíduos olhando para 1003 imagens foram utilizados para gerar um modelo baseado em vetores de características. As características foram escolhidas após analisar as fixações nas imagens. No nível mais baixo foram incluídas as características fisiologicamente plausíveis. Foram incluídas pirâmides de sub-bandas em quatro orientações e três escalas, também foram incluídas as pirâmides usadas nos modelos de [Oliva e Torralba \(2001\)](#) e [Rosenholtz \(1999\)](#), assim como as características cor, intensidade e orientação usadas em [Itti, Koch e Niebur \(1998\)](#) e um canal para cada cor RGB e para suas probabilidades. No nível médio foi utilizado um detector de linha horizontal. No alto nível foram utilizados o detector de faces de Viola Jones ([VIOLA; JONES, 2001](#)) e o detector de pessoas de Felzenszwalb ([FELZENSZWALB; MCALLESTER; RAMANAN, 2008](#)). Também foi adicionada uma característica indicando a distância de cada *pixel* ao centro da imagem. O algoritmo utilizado para aprendizado foi o SVM com função de *kernel* linear. O desempenho verificado pelo modelo foi alto: o modelo com todas as características atingiu 88% de proximidade ao desempenho

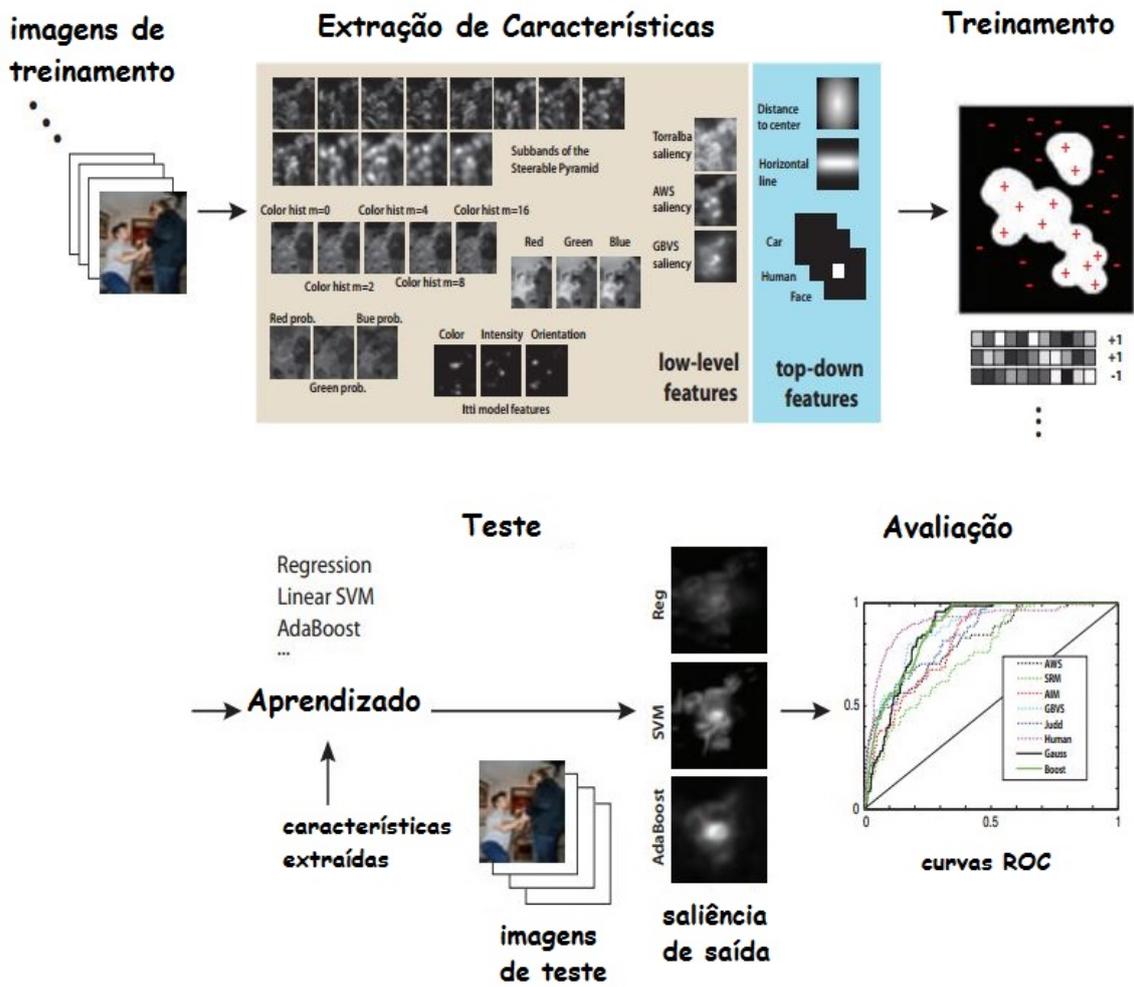
humano. Por exemplo, quando imagens são limiarizadas a 20% de saliência, o modelo teve um desempenho de 75%, enquanto que o dos seres humanos é de 85%.

Borji (2012) propôs um modelo parecido com o de Judd et al. (2009), seguindo a ideia de que a atenção dos seres humanos é atraída mais diretamente por alguns tipos específicos de características, como: faces, pessoas, carros, etc. A ilustração do modelo é mostrada na figura 11. Adicionalmente às características utilizadas em Judd et al. (2009), há os mapas de saliências de Torralba (TORRALBA, 2003), AWS (*Adaptive Whitening Saliency*)(GARCIA-DIAZ et al., 2009) e GBVS (*Graph-Based Visual Saliency*) (HAREL; KOCH; PERONA, 2006), utilizados como características de baixo nível e, carros, detectados pelo método de Felzenszwalb's *Deformable Part Model* (FELZENSZWALB et al., 2010). Na aprendizagem foram utilizados três classificadores diferentes: Regressão Linear, SVM e AdaBoosting. Foram utilizadas três abordagens para avaliar a performance de estimação do modelo: Área sobre a curva ROC (AUC), NSS e Coeficiente de Correlação Linear (LCC) . E três bases de dados distintas. O método utilizando AdaBoosting se sobrepôs aos outros métodos na maior parte dos casos, inclusive sobre outros modelos comparados. Em alguns testes o modelo de Tavakoli (TAVAKOLI; RAHTU; HEIKKILÄ, 2011), que utiliza uma técnica baseada na abordagem de centro-vizinhança, teve um melhor desempenho.

O método de Xu et al. (2014) utiliza três níveis de características: pixel, objeto e semântico. As características a nível de *pixel* são cor, intensidade e orientação. As características em nível de objeto são o tamanho, a convexidade, a solidez, a complexidade e a excentricidade. As características em nível semântico são face, emoção (faces com emoções óbvias), objetos tocáveis, objetos que estão sendo apontados, objetos que se movem, objetos que produzem som, objetos com cheiro, objetos que podem ser experimentados (comida, bebida, etc.), textos, objetos que podem ser assistidos (ex. telas) , objetos operáveis com as mãos. Para cada uma dessas características o objeto é marcado com um (que corresponde à existência desse atributo) ou zero (para indicar a ausência do atributo). A classificação foi realizada utilizando SVM. Com todas as características o modelo obteve uma AUC de 0,85.

Esses modelos são alguns dos principais dentre os métodos que utilizam padrões de classificação. O capítulo seguinte apresenta a revisão sistemática realizada com o objetivo de avaliar os métodos utilizados nas etapas referentes ao rastreamento de olhar do TEA.

Figura 11 – Ilustração do modelo de Borji (2012)



Fonte: Adaptado de Borji (2012)

## 4 Revisão bibliográfica sistemática

Este capítulo apresenta os principais resultados e discussões de uma Revisão Sistemática (RS) que teve como objetivo sistematizar o estado da arte em relação a trabalhos de rastreamento de olhar para auxílio à detecção de autismo, com foco na análise dos métodos e técnicas utilizados para seleção e análise de dados, nos dispositivos de rastreamento de olhar utilizados no experimento, na formação do grupo de participantes e nos paradigmas visuais estudados.

As questões definidas para essa pesquisa foram: “Quais métodos e técnicas são utilizados nos experimentos de rastreamento de olhar de pacientes com TEA, principalmente durante a fase de análise dos dados provenientes de rastreamento de olhar?”, “Quais as finalidades dos métodos e técnicas computacionais empregados?”, “Que dispositivos de rastreamento de olhar são empregados?”, “Como são compostos os grupos de pacientes participantes do experimento?” e “Quais os principais paradigmas visuais estudados e como são construídos seus estudos?”

As palavras chaves utilizadas foram “*eye tracking*” e “*autism*” e, as seguintes fontes de busca: IEEE (*Institute of Electrical and Electronics Engineers*) (IEEE, 2016), ACM (*Association for Computing Machinery*) (ACM, 2016), PubMed (PUBMED, 2016), *Science Direct* (DIRECT, 2016) e *Springer* (SPRINGER, 2016).

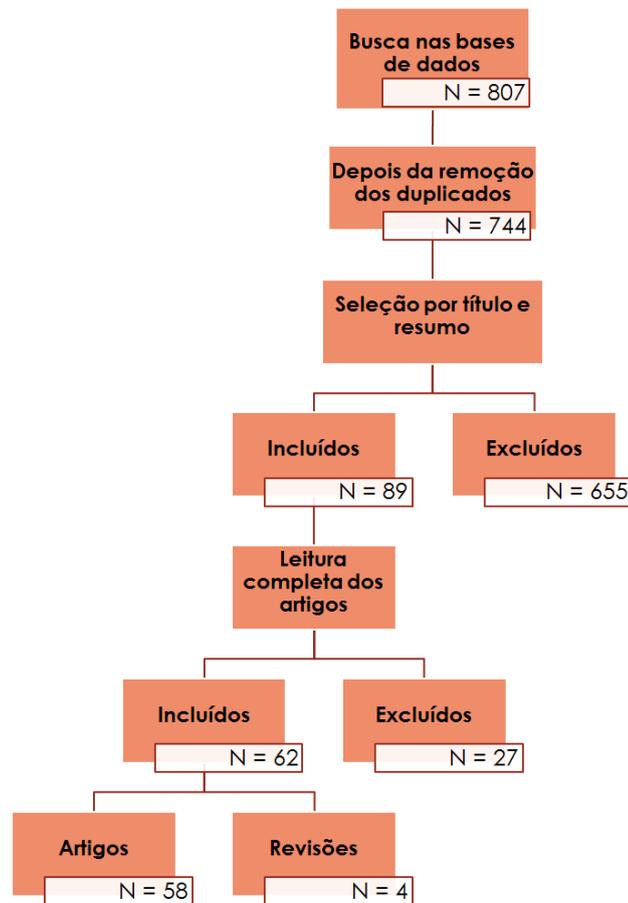
Foram encontrados 807 estudos e, após aplicação de critérios de inclusão e exclusão, 62 estudos foram aceitos para análise final. A figura 12 sumariza o processo realizado.

Para cada artigo foi preenchido um formulário com os dados definidos no protocolo, permitindo categorizar e analisar cada dimensão pesquisada. As principais análises tecidas durante a RS são apresentadas nas seções a seguir. O conteúdo completo foi apresentado em formato de artigo, e encontra-se no apêndice B.

### 4.1 Equipamentos de rastreamento de olhar

A escolha do dispositivo de rastreamento de olhar tem um papel importante na definição do experimento, pois pode facilitar ou dificultar a coleta de dados, além de definir limitações quanto ao tipo de paciente que será recrutado. Uma primeira classificação desses dispositivos, divide-os entre locais (aqueles que devem ser ajustados ao paciente) e remotos (que conseguem capturar os sinais a certa distância do paciente). Por exemplo, no caso de trabalhos que utilizam interação natural (MAGRELLI et al., 2013; NORIS et al., 2012), ou seja, o paciente não vê um

Figura 12 – Diagrama de fluxo resumindo o processo de seleção dos estudos



Fonte: Jéssica dos Santos de Oliveira, 2017

vídeo, mas sim interage diretamente com um pesquisador, são utilizados dispositivos do tipo *head-mounted* que se encaixam como um “capacete” na cabeça da criança. Porém, esse tipo de dispositivo é utilizado em apenas 16% dos trabalhos encontrados, provavelmente por ser necessário que o indivíduo fique sem se movimentar para que o equipamento seja ajustado e não o remova durante o experimento. Já os dispositivos remotos são utilizados em 81% dos artigos encontrados.

A maioria dos trabalhos (59%) utilizou os equipamentos da marca *Tobii* provavelmente porque todos os seus modelos são remotos e de alta precisão. O modelo mais utilizado foi o *Tobii T1750*, usado em 60% dos artigos encontrados.

Os trabalhos de [Guimard-Brunault et al. \(2013\)](#) e [Hernandez et al. \(2009\)](#) utilizam sistemas próprios de rastreamento de olhar, montados com duas câmeras e um computador. [Wang et al. \(2015a\)](#) utilizam uma *web câmera* para capturar o olhar dos participantes. Esses trabalhos não citam a precisão ou a quantidade de dados perdidos no experimento, não sendo possível comparar com os outros dispositivos.

Aproximadamente 35% dos artigos encontrados relataram a exclusão de alguns participantes devido a não conseguir calibrar o equipamento para estes participantes ou ao fato de a quantidade de dados capturados não atingir um critério determinado. Em geral, este critério é estabelecido em 50% do tempo total do experimento, como apontado por [Pierce et al. \(2015\)](#), [Shi et al. \(2015\)](#), [Pierce et al. \(2011\)](#). Já [Sasson et al. \(2011\)](#) excluem apenas indivíduos em que não foi possível captar pelo menos 15% dos dados do rastreamento de olhar em um dos testes.

#### 4.2 *Participantes no experimento*

O recrutamento de voluntários para experimentos de rastreamento de olhar no contexto do TEA é uma tarefa árdua, longa e, muitas vezes, cara, visto que pode ser exigida uma quantidade grande de indivíduos com TEA e com Desenvolvimento Típico (DT) para que o estudo possa tecer conclusões com significância estatística. Nos artigos incluídos foram observadas variações em relação a idade, quantidade de participantes, proporção de participantes com TEA e com DT e tipos de transtornos dos participantes dos experimentos.

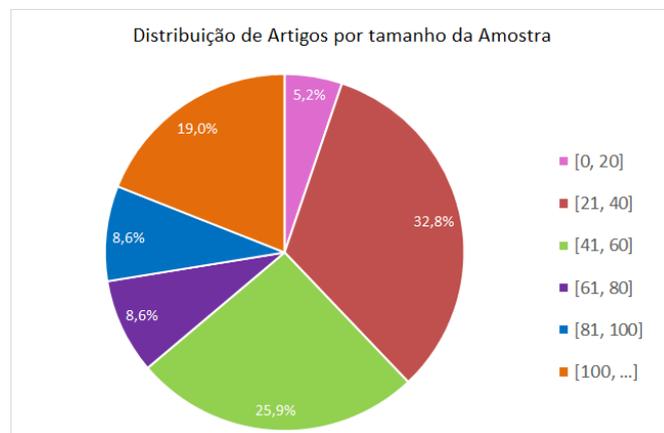
Esses aspectos variam, principalmente, de acordo com o objetivo do trabalho. No trabalho de [Pierce et al. \(2015\)](#), por exemplo, os indivíduos que participam do experimento têm idades entre 10 e 49 meses e, além dos participantes com TEA e do grupo de controle, também entraram no experimento crianças com atraso no desenvolvimento e outras condições como: nascimento prematuro e exposição a drogas durante a fase pré-natal. Essa seleção de indivíduos se adequa ao objetivo do trabalho de traçar um biomarcador precoce, utilizando crianças bem jovens e com outras condições, para que os resultados do TEA não se confundam com outras adversidades precoces. Também, por se tratar de um trabalho que propõe a predição, uma amostra muito pequena não traria resultados satisfatórios; neste caso a amostra de 334 participantes parece ser mais adequada ao proposto do que a de seu trabalho anterior ([PIERCE et al., 2011](#)), com 110 participantes.

Por outro lado, o trabalho de [Schwartzman et al. \(2015\)](#) pretendeu verificar a fixação do olhar em estímulos sociais, tecendo uma comparação entre pacientes com síndrome de Rett, TEA e controle. Neste caso, a pouca idade dos participantes já não era um requisito, permitindo participantes mais velhos e de idades mais variadas, entre 4 e 30 anos. Porém, é apontada como uma limitação do trabalho o fato de a amostra ser pequena, composta de apenas 42 participantes, sendo 14 com síndrome de Rett, 11 com TEA e 17 de controle.

A figura 14 mostra a porcentagem da quantidade de participantes por experimento, considerando intervalos a cada vintena de indivíduos. Cerca de 33% dos experimentos trabalham com um intervalo de 21 a 40 indivíduos no total (considerando todos os grupos exigidos pelo experimento), 25% utilizam de 41 a 60 pessoas e apenas 19% trabalham com mais de 100 indivíduos.

O tamanho da amostra é apontado por muitos artigos como uma limitação, muitas vezes pela dificuldade em recrutar participantes que preencham os critérios determinados pelo experimento, seja idade ou grau de autismo, além do fato de muitos trabalhos precisarem excluir dados. Aproximadamente 50% dos trabalhos encontrados precisam excluir os dados de alguns participantes, referentes ao grupo investigado ou ao grupo de controle. O principal motivo apresentado é a qualidade dos dados coletados, isto é, dados inadequados principalmente porque o equipamento não foi capaz de registrar a posição do olhar em alguns intervalos de tempo. Em geral, como citado na seção 4.1, são excluídos os casos em que não foram capturados corretamente metade ou mais dos dados do rastreamento de olhar.

Figura 13 – Distribuição dos artigos incluídos considerando o tamanho da amostra de participantes nos experimentos.



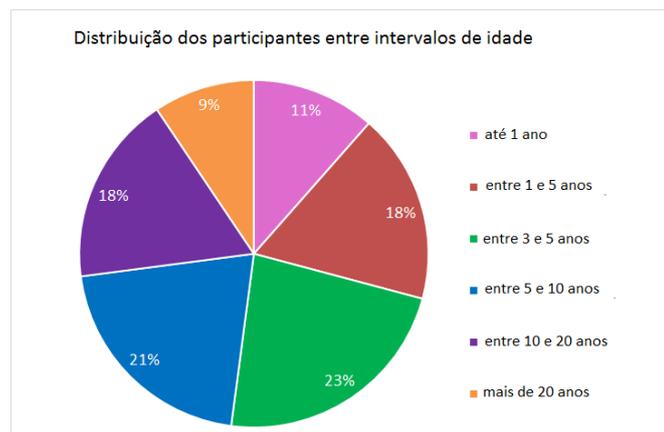
Fonte: Jéssica dos Santos de Oliveira, 2017

Em média, as amostras citadas nos trabalhos incluídos são balanceadas entre a quantidade de autistas e de controle, cerca de 79% dos trabalhos estão em um intervalo em que 30% a 65% dos participantes apresentam TEA. Dos que possuem menos que 30% com TEA, aproximadamente 54% possuem outros grupos de comparação, como o trabalho de (SCHWARTZMAN et al., 2015) que compara indivíduos com TEA, Síndrome de Rett e de controle. Cerca de 21% dos trabalhos fazem esse tipo de comparação com outros transtornos, sendo citadas as seguintes: transtorno do desenvolvimento pervasivo não específico, síndrome de Rett, síndrome de William, atraso no desenvolvimento, atraso na linguagem e síndrome de Down.

A figura 14 mostra a distribuição da idade dos participantes nos experimentos. Alguns trabalhos são incluídos em mais de uma faixa do gráfico, por utilizarem participantes de idades variadas.

Em relação à idade dos participantes, 23% dos trabalhos encontrados tem participantes com idade entre 3 e 5 anos, como mostrado na figura 14. Poucos trabalhos (apenas 11%) conseguem utilizar bebês menores de 1 ano para realizar o experimento; uma hipótese para essa baixa porcentagem é a dificuldade no uso do rastreamento de olhar, especialmente na fase de calibração, e a necessidade de um acompanhamento posterior para confirmar se os bebês com risco de TEA realmente apresentaram o transtorno. Apenas 9% utilizam pessoas com mais de 20 anos, provavelmente por não ser mais necessário o uso de rastreamento de olhar nessa idade, já que o diagnóstico se dá por volta dos 6 anos de idade.

Figura 14 – Distribuição dos artigos incluídos considerando intervalos de idades dos participantes nos experimentos



Fonte: Jéssica dos Santos de Oliveira, 2017

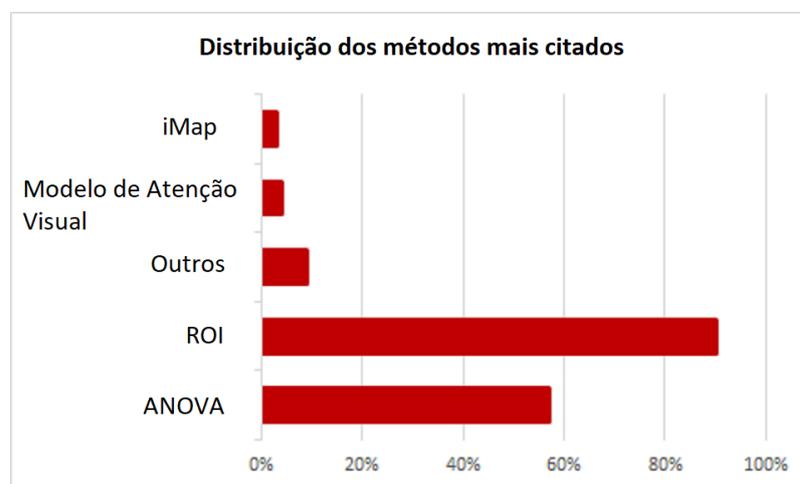
Há trabalhos que utilizam a idade cronológica como critério de pareamento<sup>1</sup> entre os grupos de TEA, controle e outros transtornos (SHIC et al., 2011; WANG et al., 2015a; ANDERSON; COLOMBO; SHADDY, 2006; GUILLON et al., 2015; SHI et al., 2015). Outros estudos utilizam níveis de inteligência, seja verbal ou de não-verbal (SPEER et al., 2007; FREETH; FOULSHAM; CHAPMAN, 2011). Noris et al. (2012), utiliza o comportamento adaptativo, medido por meio da *Vineland Adaptive Behavior Scale*. Segundo Noris et al. (2012) essa medida assegura que as crianças sejam agrupadas por semelhanças nas habilidades em realizar tarefas do dia-a-dia. A maioria dos trabalhos (31%) prefere utilizar mais de uma medida de pareamento, como idade cronológica, desenvolvimento verbal e não-verbal.

<sup>1</sup> critério para associação entre os participantes

### 4.3 Técnicas e métodos

Os trabalhos incluídos nesta revisão indicaram o uso de métodos computacionais ou estatísticos em várias etapas do experimento, com a finalidade de colaborar durante o processo de análise do movimento do olhar dos participantes TEA. As principais finalidades desses métodos são o tratamento dos estímulos visuais antes da utilização no experimento, a seleção das regiões para análise do estímulo após a captura de dados e a análise dos dados obtidos na captura. Nas subseções a seguir são descritos os principais métodos encontrados para cada uma dessas etapas. A figura 15 mostra a distribuição dos métodos que mais se destacaram entre os artigos incluídos.

Figura 15 – Distribuição dos métodos que mais se destacam entre os artigos incluídos.



Fonte: Jéssica dos Santos de Oliveira, 2017

#### 4.3.1 Tratamento de estímulos

Em 20% dos artigos incluídos foram usados métodos para tratamento do estímulo visual. Esses estímulos podem ser imagens, no caso de estímulos estáticos; vídeos, no caso de estímulos dinâmicos ou mesmo uma interação real entre o participante e um pesquisador, no caso de estímulos naturais ou seminaturais.

Um dos tratamentos aplicados em estímulos estáticos é o método *Bubbles* utilizado no trabalho de Neumann et al. (2006). Este trabalho tinha como intuito verificar se os indivíduos com TEA olhavam mais para a região da boca, mesmo quando ela não estava nítida na face. Para isso, foi aplicado o método *Bubble*, no qual as imagens de faces do estímulo visual eram transformadas por meio de métodos de suavização e realce, para que as faces fossem reveladas em pontos aleatórios e borradas em outros.

No trabalho de [Dundas, Gastgeb e Strauss \(2012\)](#), ao invés de mostrar vários estímulos diferentes, a imagem de uma única face foi manipulada seis vezes utilizando um programa de transformação facial, em que são alteradas as distâncias entre o nariz e a boca, o tamanho do nariz, o tamanho da testa e a largura dos lábios. Segundo [Dundas, Gastgeb e Strauss \(2012\)](#), foi manipulada a imagem da face de uma única pessoa, para maximizar a atenção nas características internas da face.

Em [Yi et al. \(2013\)](#) as imagens eram convertidas para escala de cinza, combinadas em relação a brilho e luminância e cortadas para ficarem no mesmo *template* normalizado. Já em [McPartland et al. \(2011\)](#) as imagens foram invertidas para verificar se a orientação alterava a atenção do indivíduos com TEA.

No caso de estímulos dinâmicos, foram encontrados três tipos de tratamento: inversão do estímulo, como em ([SHIC et al., 2007](#)), com o mesmo intuito do trabalho de [McPartland et al. \(2011\)](#) (verificar influência da orientação da imagem na fixação do olhar); remoção de som, também visto em ([SHIC et al., 2007](#)), pois o som não fazia parte da análise do trabalho e poderia influenciar nos resultados; e divisão da tela em duas ou mais partes, como em ([PIERCE et al., 2015](#); [PIERCE et al., 2011](#); [SHI et al., 2015](#); [CHEVALLIER et al., 2015](#)), com tipos de cenas distintas em cada parte, com o intuito, em geral, de analisar qual cena atrairia maior atenção.

#### 4.3.2 Seleção e Análise de Dados

Para selecionar as regiões do estímulo a serem processadas após o rastreamento de olhar, se destacam dois métodos: ROI (do inglês *regions of interest*, também chamado de AOI, do inglês *areas of interest*) e iMap. No caso do método ROI, as regiões de interesse são definidas manualmente, de acordo com o conhecimento *a priori* do pesquisador, e demarcadas manualmente ou com a ajuda de um software. Como mostrado na figura 15, a maioria dos trabalhos ainda utiliza esse tipo de abordagem. As regiões demarcadas, em geral, dependem do paradigma visual que está sendo analisado.

No método iMap não é necessário um conhecimento *a priori* do pesquisador, pois não é preciso delimitar as áreas no estímulo. O iMap gera mapas de fixação do olhar para cada participante e cada estímulo visual, resultando em um mapa médio de fixação. Pode então ser calculada a média dos mapas de fixação dos indivíduos em conjunto, resultando em um mapa de fixação do grupo. A principal vantagem do iMap é a utilização de todos os pontos de

fixação para geração dos mapas de fixação possibilitando comparações estatísticas destes dados (CALDARA; MIELLET, 2011). Os trabalhos de Yi et al. (2013) e de Shi et al. (2015) utilizam esses dois métodos, possibilitando uma análise tanto de pontos específicos do estímulo, como da visualização do estímulo como um todo. Segundo Shi et al. (2015) o iMap parece ser mais objetivo, assertivo e visual, além de prover uma melhor resolução espacial permitindo detectar mais facilmente a distinção entre os grupos na distribuição espacial da fixação. Porém, ambos os trabalhos utilizaram o iMap apenas como um método de confirmação dos resultados obtidos pelo ROI.

O trabalho de Amestoy et al. (2015) criou uma abordagem própria, baseada em métodos estatísticos espaciais, utilizando um Diagrama de Voronoi. Falck-Ytter et al. (2013) propõem um método diferenciado para visualizar os dados do rastreamento de olhar, uma medida nomeada D2R (*Distance to reference point*). Esta medida combina as duas dimensões espaciais em uma única dimensão, mas mantém a informação que é necessária para analisar os dados a olho nu em um gráfico espaço-temporal bidimensional. Este método também permite verificar padrões de movimento dos olhos entre os grupos. Em Nakano et al. (2010) foi utilizada uma escala multidimensional, por meio da qual era calculada a distância entre cada par de pontos de fixação. Se dois participantes olharam para o mesmo ponto, a distância era zero, se olharam para pontos diferentes a distância era a quantidade de *pixels* que separavam esses dois pontos. Esse procedimento foi repetido para cada par de participantes para definir uma matriz de distância. Com a matriz, foi aplicado o MDS para desenhar cada participante em um plano bidimensional. Nesse plano, foi possível verificar que os participantes com DT eram distribuídos no centro do plano, o que indicava um padrão de fixação. Já os participantes com TEA eram distribuídos nas periferias.

Para auxiliar na análise de dados, alguns trabalhos utilizaram modelos de atenção visual *bottom-up*, que se propõem a prever a atenção visual em um determinado estímulo. De um modo geral, esses modelos extraem as características visuais do estímulo, como cor, intensidade, orientação e, dependendo do modelo, movimento. Essas características são combinadas e processadas gerando um mapa de saliências, em que o maior valor representa o local em que a atenção deverá ser dirigida.

Um dos artigos que utiliza esse tipo de modelo é o trabalho de Neumann et al. (2006), que empregou o modelo *bottom-up* de Itti, Koch e Niebur (1998) para analisar a atenção atípica dos autistas em relação a imagens faciais. Comparando o mapa de saliência dado pelo modelo com a atenção dos indivíduos com TEA, percebeu-se que a atenção mais dirigida para a boca não se

devia às características *bottom-up* da imagem. O trabalho de [Freeth, Foulsham e Chapman \(2011\)](#) verifica se a saliência visual exerce influência na atenção de adolescentes com TEA da mesma forma que adolescentes com DT, também utilizando o modelo de [Itti, Koch e Niebur \(1998\)](#). Em ambos os grupos, a saliência na fixação foi maior que a média de saliência do mapa, indicando que os participantes, em geral, fixavam em regiões que eram abaixo da média visual de saliência. Também observou-se que os participantes com TEA fixavam o olhar na face mais tardiamente que os com DT. Em [Shic et al. \(2007\)](#) o modelo de [Itti, Koch e Niebur \(1998\)](#) foi utilizado para comparar a atenção de crianças com TEA e com DT, usando vídeos com orientações diversas e movimentos biológicos como estímulo. Foi observado que a orientação da imagem não afeta a atenção das crianças com TEA como acontece com as com DT e que as crianças com TEA atentavam-se menos ao movimento.

Em ([WANG et al., 2015b](#)) é utilizado um modelo de atenção visual que leva em consideração todos os pontos de fixação realizados por cada grupo de indivíduos, criando um mapa de fixação para o grupo com TEA e outro para o grupo de controle. De um modo geral, esses mapas são criados extraindo-se das imagens de entrada os atributos: em nível de *pixel* (cor, intensidade e orientação); em nível de objeto (tamanho, convexidade, solidez, complexidade e a excentricidade) e em nível semântico (face, emoção (faces com emoções óbvias)), entre outros. Foi utilizado um classificador SVM para combinar os pontos de fixação com os atributos extraídos e gerar os mapas de saliência. Analisando os mapas, foi possível observar que indivíduos com TEA olharam mais para o centro da imagem, mesmo quando não havia nenhum objeto naquela região e que eram mais lentos para se fixarem em faces, enquanto fixavam mais rapidamente em objetos mecânicos e manipuláveis.

Por fim, todos os trabalhos incluídos nesta revisão utilizam métodos estatísticos para a análise de dados e validação das hipóteses. O principal método utilizado para este fim é o ANOVA (57% dos trabalhos). Alguns trabalhos também utilizaram correlação de Pearson, como em [Johnels et al. \(2014\)](#), estimador de verossimilhança, como em [Elsabbagh et al. \(2014\)](#), testes não paramétricos, como o de Kruskal-Wallis [Falck-Ytter, Thorup e Bölte \(2014\)](#), entre outros.

#### 4.4 Parâmetros para auxílio ao diagnóstico

Seis trabalhos encontrados definem algum critério que consegue distinguir com certa sensibilidade os grupos de indivíduos com TEA e com DT.

O primeiro, dentre os trabalhos incluídos, a definir um critério foi o trabalho de [Nakano et al. \(2010\)](#). A distância entre a fixação do olhar dos participantes foi desenhado em um plano bidimensional, como descrito na seção [4.3.2](#), e utilizada para discriminar os grupos. Foram construídas duas curvas ROC com a variação dessas distâncias, uma para discriminar adultos com TEA e com DT e outra para discriminar crianças. Na primeira obteve-se uma AUC de 0,75 e na segunda de 0,87. Um problema observado neste método é que ele não considera as informações do estímulo visual para realizar a medida, o que pode levar à interpretação de que para qualquer estímulo seria encontrado o mesmo padrão. Além disso, por não considerar as informações visuais, outros transtornos poderiam se encaixar no mesmo perfil dos indivíduos com TEA.

O trabalho de [Pierce et al. \(2011\)](#), define um critério baseado no paradigma de movimento biológico. O estímulo consistia em uma tela dividida em duas partes iguais: uma delas apresentava cenas de crianças fazendo yoga e a outra metade disponibilizava figuras geométricas se movendo de acordo com um padrão. A partir das análises realizadas, foi concluído que, se a duração total da fixação nas figuras geométricas for de 69% ou mais, o indivíduo é classificado com 100% de acurácia como portador do TEA. No entanto, este trabalho não conseguiu chegar a nenhuma conclusão sobre os indivíduos que fixam o olhar menos que 69% do tempo, podendo estes apresentar ou não o transtorno. Em um trabalho posterior ([PIERCE et al., 2015](#)), os autores confirmam esses resultados em um experimento com maior número de crianças, obtendo como resultado uma sensibilidade de 21% e especificidade de 98%. Alterando o limiar da duração de fixação na área geométrica, os pesquisadores obtiveram uma curva ROC com AUC de 0,71.

[Shi et al. \(2015\)](#) comparam o estímulo criado por [Pierce et al. \(2015\)](#), com um estímulo com mais de uma criança na parte de movimentos biológicos. Neste, também alterando o limiar de duração de fixação na área geométrica, é obtida uma AUC de 0,865 para o segundo estímulo, e de 0,623 para o estímulo original de [Pierce et al. \(2015\)](#). Apesar de [Shi et al. \(2015\)](#) ter utilizado uma amostra pequena, de apenas 33 indivíduos, já é possível observar por este trabalho que quanto mais interações sociais presentes na cena, maior a capacidade de distinguir entre os grupos.

O mesmo ocorre com o trabalho de [Chevallier et al. \(2015\)](#). Comparando três tipos diferentes de estímulos, observa-se que o estímulo com interações entre crianças foi o que obteve maior capacidade de distinguir entre os grupos, com AUC igual a 0,71. Porém, neste trabalho o parâmetro utilizado para construção da curva foi o de duração total na fixação do olhar nas faces representadas no estímulo. Apesar dos avanços promissores nos resultados, estes valores

ainda não permitem que tais métodos sejam utilizados para auxílio ao diagnóstico. Amostras mais completas e métodos de análise mais robustos, podem ser passos iniciais para melhorar esses resultados.

#### 4.5 *Avanços, limitações, oportunidades e conclusões*

Com o avanço da tecnologia, novas possibilidades se abriram dentro do escopo da presente revisão, conforme apresentamos a seguir.

##### 4.5.1 Métodos

Um avanço encontrado foi a utilização de métodos como iMap e modelos de atenção visual ao invés de somente a habitual demarcação manual de regiões de interesse (ROI) . A utilização isolada de ROI pode causar perda de informação, pois depende da operação humana – geralmente do especialista – que deve demarcar as regiões do estímulo que serão mapeadas para captura e análise dos dados. Mesmo que tal demarcação possa contribuir para uma redução de dimensionalidade devido à seleção *a priori* dos dados que serão analisados, a natureza deste método pode fazer com que outras regiões não demarcadas sejam desconsideradas inadequadamente e, mesmo nas regiões demarcadas pode haver informações mais detalhadas que serão perdidas.

Há também os modelos de atenção visual, que permitem analisar não somente regiões específicas das cenas, mas também atributos presentes no estímulo, como: cor, movimento, orientação, pessoas, objetos, entre outros. Em (WANG et al., 2015b) [referencia](#), além de analisar esses atributos, é criado um mapa de saliência com todos os pontos de fixação, para cada um dos grupos analisados, possibilitando análise e comparação mais completas e aprofundadas.

Apesar do avanço citado, nos trabalhos incluídos foram encontrados poucos métodos computacionais que pudessem auxiliar na detecção dos padrões na atenção visual de indivíduos com TEA. Considerando a grande complexidade dos dados, verifica-se uma lacuna a ser preenchida: a utilização de métodos de aprendizado de máquinas como Redes Neurais, SVM, Árvores de Decisão, Naive Bayes, entre outros, que podem ser explorados para oferecer uma contribuição mais efetiva na análise dos sinais do rastreamento de olhar. Tais métodos poderiam ser utilizados, por exemplo, para identificar quais características dos estímulos (cor, tamanho, posição, etc.) poderiam melhor discriminar autistas e controles, ou métricas de análise mais adequadas (tempo

total de fixação, maior tempo de fixação contínua, ordem de fixação) e também para identificar uma combinação de diferentes estímulos que maximize essa discriminação.

#### 4.5.2 Participantes dos experimentos

Outro ponto de interesse que constitui oportunidades de pesquisa diz respeito à seleção de participantes para os experimentos. Alguns estudos citam o fato de só utilizarem participantes com autismo de alto funcionamento. Esse grupo de indivíduos muitas vezes é escolhido pela facilidade de conseguir com que prestem atenção aos estímulos apresentados e de usar certos tipos de equipamentos de rastreamento. Porém, ao testar as abordagens apenas com indivíduos de alto funcionamento, os resultados não podem ser generalizados para todo o espectro, pois estes representam apenas um pequeno número dos indivíduos com TEA. Portanto, o desenvolvimento de métodos que permitam considerar grupos mais abrangentes do espectro autista, assim como o desenvolvimento de tecnologias robustas, que permitam rastrear o olhar de forma menos invasiva e com menor perda de dados, constituem desafios que merecem aprofundamento em projetos de pesquisa.

Considerando ainda o aspecto de seleção de participantes nos experimentos, outros trabalhos apontam como limitação o fato contrário: por utilizarem um grupo heterogêneo de autistas, não conseguem separar os resultados quando o indivíduo tem comorbidades. A solução da questão citada passa pela necessidade de se obter uma ampla amostra de indivíduos em que pudessem ser analisados vários casos diferentes e seus respectivos resultados. O tamanho da amostra também é apontado como limitação na maioria dos estudos. Como mostrado na seção 4.2, apenas 19% dos estudos executaram experimentos com mais de 100 indivíduos participantes.

Também é citada como limitação o intervalo de idade dos indivíduos dos grupos participantes dos experimentos, visto que experimentos conduzidos com indivíduos com faixa etária específica não permitem generalizar os resultados encontrados.

As limitações apresentadas em relação aos participantes dos experimentos remete à ausência de uma base de dados pública e suficiente para testar as diversas abordagens e paradigmas propostos pelos pesquisadores. Além disso, os diferentes aspectos envolvidos nos experimentos, tais como paradigmas, métodos, dispositivos e características dos voluntários, torna muito difícil a comparação de resultados entre os diversos trabalhos e a elaboração de conclusões efetivas.

O recrutamento de voluntários para formar tal base de dados envolve questões de ordens ética, financeira e prática. Há uma grande dificuldade para se formar uma base de dados que considere uma grande variedade de casos em relação aos diversos aspectos envolvidos (níveis do TEA, presença de outros transtornos, idade, gênero, cognição, entre outros) e que, simultaneamente, possua casos de controle pareados considerando um ou mais dos aspectos citados. Neste sentido, em todas as fases do processo – desde o estabelecimento de protocolos adequados até a definição de formas eficientes de armazenamento e recuperação desses dados – há desafios de pesquisa que constituem importantes temas a serem investigados.

#### 4.5.3 Estímulos

Alguns estudos utilizam estímulos naturalísticos ou seminaturalísticos, pois citam que os estímulos dinâmicos e, mais ainda, os estáticos, estão longe da realidade, podendo apresentar resultados diferentes dos percebidos no dia-a-dia. Porém, para que seja possível a utilização de estímulos mais realistas, é necessária a utilização de dispositivos mais invasivos, como capacetes, que podem limitar, como já citado, o grupo de participantes do experimento.

A composição de estímulos adequados deve considerar aspectos como o público-alvo do experimento, o paradigma a ser avaliado e os métodos que serão empregados. É uma tarefa complexa que deve, obrigatoriamente, envolver especialistas de diversas áreas. Esta é uma boa prática de pesquisa que deveria ser considerada pelos diferentes grupos como uma forma de permitir avanço mais efetivo na área. A divulgação das fontes utilizadas para criação de estímulos, a formação e a disponibilização de bases públicas de estímulos, considerando os diversos paradigmas conhecidos, poderia ser uma contribuição para que o estado da arte avançasse mais rapidamente nesta área de pesquisa.

#### 4.5.4 Conclusões da revisão sistemática

Mesmo com as limitações apontadas, a utilização do rastreamento de olhar para diagnóstico do autismo é uma ferramenta promissora em relação às utilizadas hoje em dia. A utilização do rastreamento de olhar diminuiria a necessidade de tradução, adaptação e padronização dos instrumentos atuais para cada região, conforme citado na seção 2.1. Também seria mais rápido, pois a coleta e o processamento dos sinais poderia durar apenas alguns minutos.

Alguns trabalhos já têm conseguido definir critérios que conseguem distinguir indivíduos entre os grupos com TEA e com DT. [Pierce et al. \(2011\)](#), [Shi et al. \(2015\)](#) e [Pierce et al. \(2015\)](#) utilizam critérios e estímulos similares, baseados no paradigma de movimento biológico, em que definem a duração total da fixação do olhar em movimentos não-biológicos como critério de diagnóstico. Basicamente, quanto mais tempo o indivíduo fixar o olhar nesta região do estímulo, maiores as chances de apresentar TEA. Já [Chevallier et al. \(2015\)](#) utilizam um estímulo com crianças brincando, e consideram a duração total da fixação do olhar na face como critério de diagnóstico, tendo aqui a relação contrária: quanto menos tempo olhando para a face, maiores são as chances de apresentar TEA.

Modelos de atenção visual estão sendo utilizados para melhorar a análise dos sinais provenientes do rastreamento de olhar. Mas nenhum trabalho, por meio dos modelos de atenção visual, definiu e avaliou algum critério para distinguir os grupos, lacuna que este trabalho pretende preencher.

Neste projeto é desenvolvido um modelo de atenção visual para classificação de TEA. A amostra utilizada para o treinamento e teste do modelo é mista tanto em relação a idade dos participantes, quanto ao nível funcional e ao gênero dos indivíduos, a fim de desenvolver um método que não seja específico apenas a um grupo de TEA. O mesmo é suposto para o paradigma visual utilizado e para o equipamento de rastreamento. As informações referentes a amostra, equipamento e sobre o experimento de um modo geral, que foram extraídas desta revisão, serviram de base para a construção do experimento de coleta de dados, descrito na seção [5.2.4](#). Além disso, os trabalhos de [Freeth et al. \(2010\)](#), [Shic et al. \(2007\)](#) e posteriormente o trabalho de [Wang et al. \(2015b\)](#), auxiliaram na compreensão de que o modelo de atenção visual poderia trazer resultados significativos para a área, assim como os trabalhos descritos na seção [4.4](#) auxiliaram a definir os melhores tipos de estímulo visual para a classificação.

## 5 Procedimentos metodológicos

Como anteriormente enunciado, este projeto consiste no processamento dos sinais de rastreamento de olhar baseado em um modelo de atenção visual com a finalidade de auxiliar no diagnóstico do TEA.

Nas próximas seções são apresentados os métodos e etapas para alcançar o objetivo proposto.

### 5.1 Etapas de pesquisa

Este projeto é uma pesquisa experimental. A primeira fase do projeto constituiu-se de uma pesquisa bibliográfica. A segunda foi constituída pela coleta de dados e experimentação computacional, culminando em um modelo computacional e em um classificador de TEA. A terceira fase foi constituída pela avaliação e análise dos dados. Paralelamente às demais fases, foi estabelecida uma fase de divulgação, em que os resultados encontrados no projeto são apresentados à comunidade acadêmica. Essas fases são divididas nas seguintes atividades:

1. **Revisão Sistemática:** foi conduzida uma revisão sistemática da literatura a fim de conhecer quais métodos utilizados atualmente para o auxílio ao diagnóstico de TEA utilizam rastreamento de olhar, quais resultados foram obtidos por esses métodos e quais padrões visuais do TEA poderiam ser agregados ao modelo proposto no presente projeto. Os resultados da revisão foram apresentados no capítulo 4.
2. **Estudo e definição do modelo de atenção visual:** foi realizado um estudo aprofundado dos modelos de atenção visual para subsidiar a escolha da abordagem mais adequada para o projeto. Os conceitos referentes aos modelos estudados foram apresentados no capítulo 3.
3. **Submissão de projeto ao Comitê de Ética:** nesta atividade foi elaborado e submetido um projeto ao Comitê de Ética por se tratar de um projeto em que são necessários testes com seres humanos. A aprovação do Comitê de Ética encontra-se disponível no Anexo A.
4. **Coleta de Dados:** foram definidos os estímulos visuais plausíveis para análise da atenção visual no TEA com colaboração de integrantes do PROTEA (IPq-USP). Os dados foram obtidos com auxílio do equipamento *Tobii TX300*, fornecido em colaboração com o

Laboratório de Neurociência (LIM-27) do IPq-USP. Os detalhes da coleta de dados são apresentados na seção 5.2.

5. **Desenvolvimento do modelo de atenção visual e do classificador:** nesta etapa foram executadas atividades para definir o modelo de atenção visual e o classificador final. As atividades executadas estão detalhadas na seção 5.3 deste capítulo e o modelo final construído está apresentado no capítulo 6.
6. **Avaliação e análise dos resultados:** as técnicas empregadas para avaliar o modelo estão apresentadas na seção 5.3.7 deste capítulo. Os resultados são apresentados e discutidos no capítulo 7.
7. **Divulgação:** um artigo com os resultados completos da Revisão Sistemática sintetizada no capítulo 4 foi submetido a um periódico (OLIVEIRA et al., 2017) e também pode ser encontrado no Apêndice B. Um artigo com os resultados obtidos durante o projeto está em desenvolvimento. Um outro estudo utilizando dados de rastreamento do olhar de pacientes com Alzheimer foi conduzido em parceria com o LIM-27, este estudo serviu de base inicial para o entendimento dos dados de rastreamento de olhar e os resultados foram publicados em (PEREIRA et al., 2017).

## 5.2 Coleta de dados

O processo de coleta de dados incluiu a escolha do equipamento utilizado, definição da amostra e recrutamento de participantes, definição do estímulo visual e o procedimento da coleta em si. Todas essas etapas foram definidas de acordo com as informações obtidas na revisão sistemática descrita no capítulo 4 e estão detalhadas nas subseções a seguir.

### 5.2.1 Equipamento para aquisição de dados

O equipamento Tobii (Tobii Technology, 2016) Pro TX300 foi utilizado para capturar os dados da fixação do olhar dos indivíduos em resposta ao estímulo visual. Este rastreador consegue capturar o olhar com uma frequência de até 300Hz. O Pro TX300 inclui, além de uma unidade de rastreamento ocular, um monitor removível de 23 polegadas. O rastreador consegue capturar os movimentos visuais até quando os movimentos de cabeça estão a uma

velocidade máxima de 50cm/s. A distância em relação ao equipamento pode variar de 50 a 80cm do observador e um ângulo máximo do olhar a 35 graus em relação ao equipamento.

O equipamento retorna a porcentagem total de captura de dados em relação ao tempo de gravação. Um aproveitamento de captura de 80% ou superior é considerado de boa qualidade para análise dos dados.

O equipamento permite demarcar regiões de interesse nos estímulos. Essas demarcações, quando feitas pelo usuário, possibilitam que o software acoplado ao equipamento calcule métricas visuais para estas regiões, baseadas nas fixações e nas visitas realizadas. Uma fixação, como o nome mesmo remete, é quando os olhos ficam fixos em uma determinada região. A definição de quando uma fixação se inicia e termina varia de acordo com o algoritmo utilizado, mas geralmente está relacionada a velocidade do movimento dos olhos. Uma visita é definida como o intervalo de tempo entre a primeira fixação dentro da ROI e a próxima fixação fora da ROI. Algumas das principais métricas calculadas pelo Tobii são apresentadas na tabela 2. De cada uma dessas métricas podem ser extraídas, automaticamente, algumas medidas provenientes da Estatística Descritiva, tais como soma, média, mínimo, máximo, variância e desvio padrão.

Tabela 2 – Algumas das principais métricas retornadas pelo equipamento Tobii TX300

<b>Métrica</b>	<b>Descrição</b>
Duração da fixação	Mede o tempo de cada fixação individual na ROI demarcada.
Total da duração das fixações	Mede o tempo total das fixações na ROI demarcada.
Contagem de fixações	Mede quantas fixações foram feitas na ROI demarcada.
Duração da visita	Mede a duração total de cada visita individual na ROI demarcada.
Total da duração das visitas	Mede o tempo total de todas as visitas feitas na ROI demarcada.
Contagem de visitas	Mede quantas visitas foram feitas na ROI demarcada.

Também é possível extrair os dados brutos (*raw data*) que contêm as coordenadas dos *pixels* que foram fixados nos estímulos e o *timestamp* em que ocorreu a fixação (STUDIO, 2008).

### 5.2.2 Participantes

Dados de 108 participantes foram incluídos no projeto: 30 crianças e adolescentes com DT (10 do sexo feminino e 20 do sexo masculino, com idades entre 3 e 15 anos, com média de idade de 10 anos) e 78 crianças e adolescentes com TEA participaram da pesquisa (32 do sexo feminino e 46 do sexo masculino, com idades variando de 5 a 18 anos com média de 10 anos).

Os participantes com TEA foram recrutados pelo IPq-USP. O diagnóstico foi feito a partir do histórico de avaliação clínica do participante e confirmado pelos especialistas do Instituto. Os participantes também fizeram a escala CARS (*Childhood Autism Rating Scale*) para avaliação complementar e teste de nível de gravidade. O CARS varia de 15 a 60, sendo 30 o corte para autismo e 36 o corte para autismo grave.

Os indivíduos com TEA apresentaram CARS médio de 34,5, variando entre 33 a 45 pontos. Os participantes com DT também fizeram o CARS e apresentaram resultados dentro do normal. Todos os indivíduos selecionados tiveram captura total maior do que 80% do tempo de vídeo.

### 5.2.3 Estímulo visual

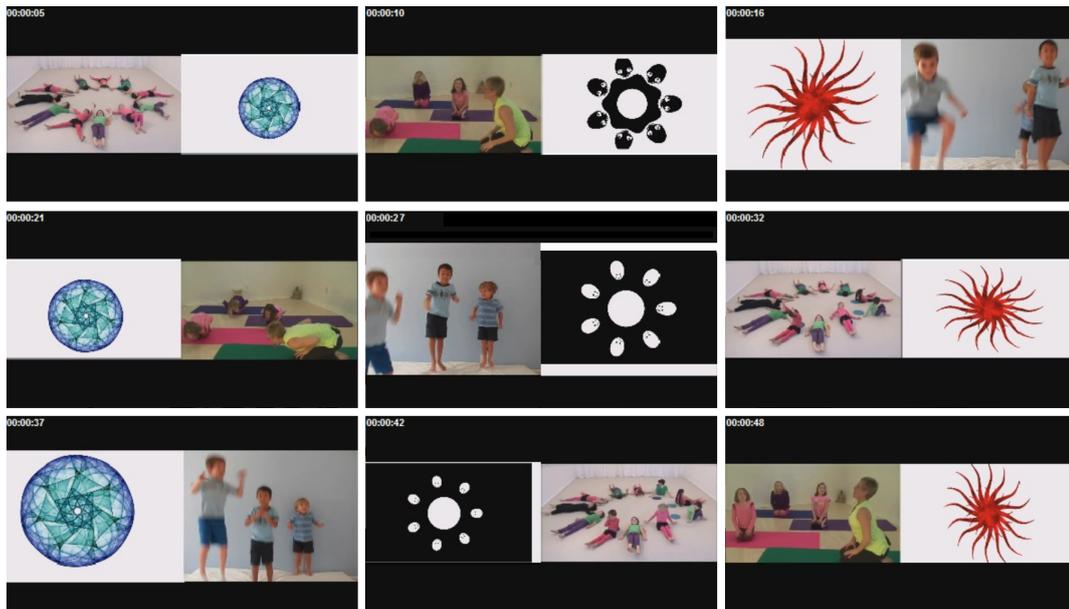
O estímulo visual foi definido com a colaboração de integrantes do grupo PROTEA. O estímulo definido consiste em vídeos curtos de aproximadamente 6 segundos cada, com dimensão de 1920 por 1080 *pixels*. Em cada vídeo a tela é dividida em duas partes: uma contendo movimentos biológicos em que há várias crianças interagindo, e outra contendo movimentos geométricos em que há um fractal em movimento. Três vídeos de movimentos biológicos e três de movimentos geométricos foram combinados, compondo nove vídeos exibidos sequencialmente, com tempo total de 54 segundos.

A figura 16 apresenta alguns *frames* do vídeo apresentado no estímulo. É possível perceber que a ordem e a posição entre as figuras com movimento biológico e movimento geométrico são alteradas para evitar que o participante fique condicionado.

### 5.2.4 Procedimentos

Para captura de dados, inicialmente eram explicados os procedimentos para o participante e para o responsável, que era convidado a ler e assinar o Termo de Consentimento Livre e Esclarecido aprovado pelo Comitê de Ética da Universidade de São Paulo (Apêndice A). O participante sentava-se em frente ao monitor do rastreador ocular, a uma distância entre 50cm e 70cm. Com o participante sentado em uma boa posição em relação ao equipamento era realizada a calibração de cinco pontos, na qual era mostrada na tela uma figura animada, em geral um ônibus ou ursinho, a qual o participante devia seguir com os olhos, de modo que o equipamento

Figura 16 – *Frames* das combinações presentes no vídeo do estímulo visual.



Fonte: Jéssica dos Santos de Oliveira, 2017

conseguisse reconhecer a localização do olhar. Caso houvesse falha na realização da calibração, era repetida e, em caso de nova falha, a criança era excluída do experimento.

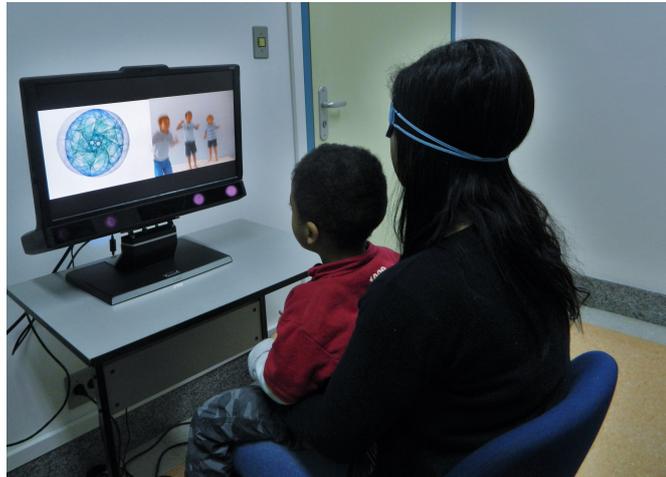
Após a calibração, era iniciada a coleta. Durante todo o tempo do experimento um médico ou cuidador ficava ao lado do participante para que ele permanecesse sentado e incentivava-o a olhar para tela se ele desviasse a atenção. Quando o participante tratava-se de uma criança pequena, era necessário, em alguns casos, que algum adulto (em geral, o cuidador) se sentasse com a criança em seu colo, seja para acalmá-la ou porque a altura da criança não permitia que ela ficasse na posição adequada para a captura. Nesses casos, o adulto utilizava uma venda nos olhos para não atrapalhar a captura, como mostrado na figura 17.

Após a coleta, os dados brutos (coordenadas dos *pixels* em que houve fixação e *timestamp* da fixação) e os valores dos parâmetros visuais gerados pelo Tobii são fornecidos pelo software disponibilizado junto ao equipamento. Esses dados foram utilizados para o desenvolvimento do modelo de atenção visual.

### 5.3 Desenvolvimento

Esta seção apresenta as etapas de desenvolvimento realizadas para a criação do modelo de atenção visual e para a classificação dos indivíduos. Todo o desenvolvimento do projeto foi realizado utilizando a plataforma Matlab (MATLAB, 2015) com auxílio do pacote *mexopenvcv*

Figura 17 – Imagem mostrando uma criança realizando o procedimento de rastreamento do olhar posicionada no colo de um cuidador.



Fonte: Jéssica dos Santos de Oliveira, 2017

(YAMAGUCHI, 2013) para extração das características. O desenvolvimento do método proposto é dividido em dois módulos: o primeiro compreende o desenvolvimento de modelos de atenção visual, para prever a atenção visual de indivíduos dos grupos TEA e de DT; o segundo módulo abrange a construção de um classificador que utiliza os modelos desenvolvidos no primeiro módulo para comparar a atenção visual de um indivíduo e definir se este pertence ao espectro do TEA. Com este propósito, foram definidas, implementadas e adaptadas técnicas computacionais, dentro das áreas de processamento de imagens e aprendizado de máquina.

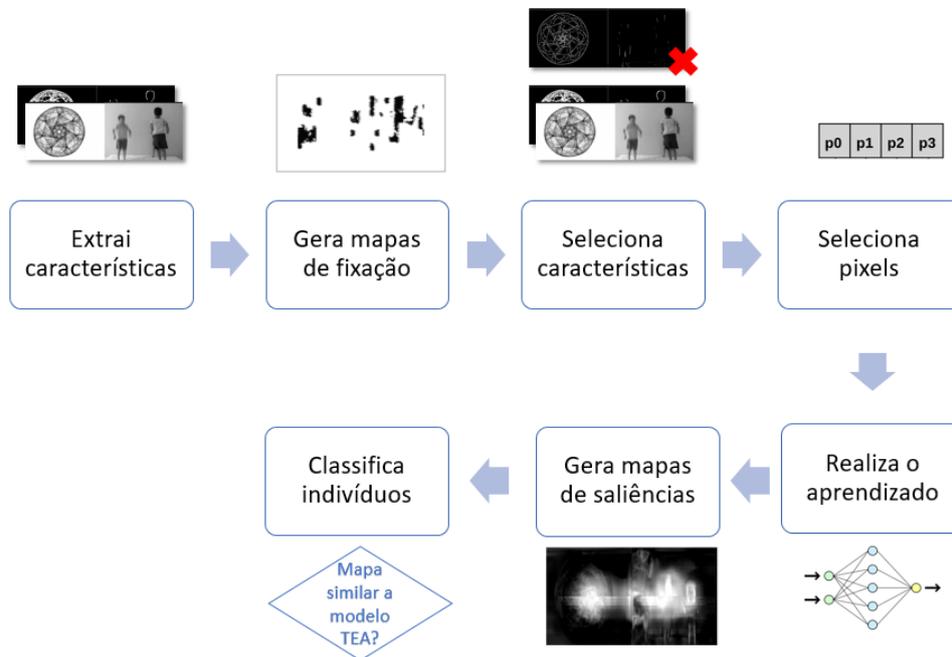
Como mostrado no diagrama apresentado na figura 18, as fases de desenvolvimento até o resultado final de classificação são: extração de características dos vídeos para formação do vetor de características; seleção das características extraídas; seleção dos *pixels* que serão utilizados para treinamento em cada um dos *frames*; construção dos mapas de fixação; definição do método de aprendizado para geração dos mapas de saliências dos grupos de TEA e DT; definição do método de classificação de TEA e avaliação dos resultados. Todas essas fases são detalhadas nas subseções a seguir.

### 5.3.1 Pré-processamento e extração de características dos vídeos

As etapas de pré-processamento e extração de características são apresentadas na figura 19.

Na etapa de pré-processamento os *frames* de transição entre dois vídeos são retirados da análise, devido ao movimento dos olhos que ocorre durante essa transição, que pode atrapalhar

Figura 18 – Diagrama com as etapas desenvolvidas até a classificação de um indivíduo como TEA ou DT

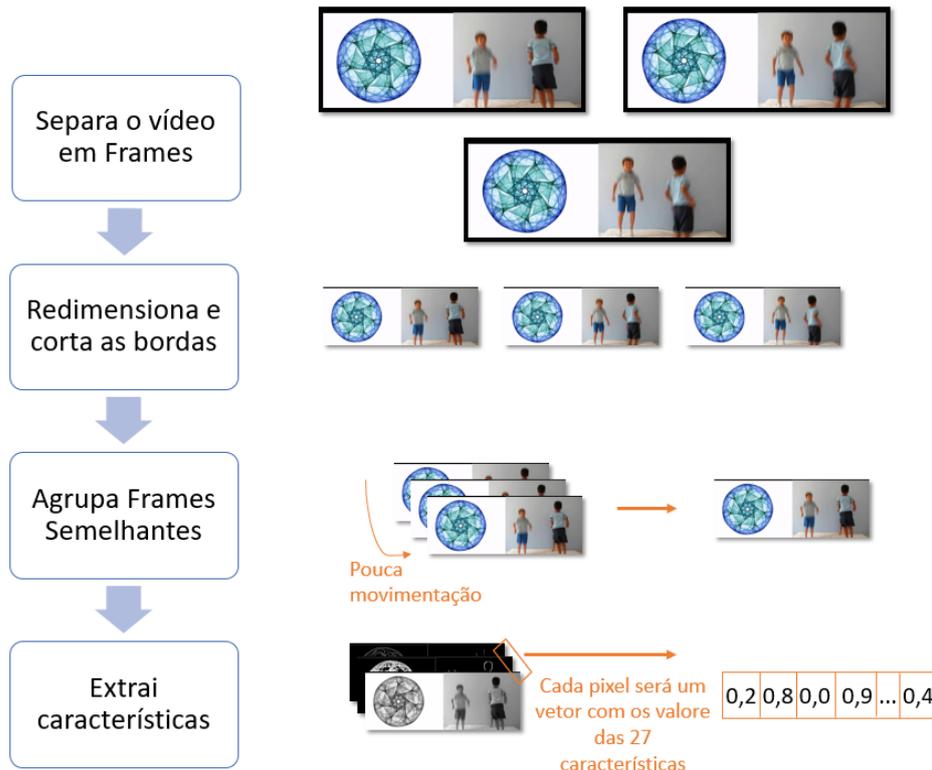


Fonte: Jéssica dos Santos de Oliveira, 2017

a análise. São removidos 10 *frames* do início e 10 *frames* do fim de cada vídeo, sobrando 1500 *frames* totais para análise. Em cada *frame* são removidas as bordas ao redor da imagem, principalmente nas partes superiores e inferiores, onde há somente *pixels* pretos, conforme mostrado na figura 19. Após os ajustes descritos os *frames* ficam com resolução de 1920x720 *pixels*. Os *frames* são redimensionados, resultando em 200x350 *pixels*, utilizando o método bicúbico, que reescala os *pixels* empregando média ponderada dos 4x4 vizinhos mais próximos. Após este tratamento inicial, as características são extraídas de todos os *frames* que restaram em cada vídeo, após os ajustes descritos.

As características inicialmente utilizadas neste projeto foram definidas com base nos principais modelos de aprendizado encontrados (JUDD et al., 2009; BORJI, 2012), porém com algumas alterações, conforme a necessidade do projeto. Por exemplo, a característica referente à detecção de carros na imagem, utilizada em (BORJI, 2012), não foi utilizada, pois além de não haver nenhum carro nos estímulos utilizados, não se adequava ao proposto do projeto. Por outro lado, características que definem distâncias ao centro das cenas e as características que definem separadamente, movimento nas regiões das cenas biológicas e movimento nas regiões das cenas geométricas foram adicionadas, pois são características que, de acordo com os estudos de atenção visual do TEA mostrados na seção 2.1.1, diferenciam a atenção visual entre os grupos.

Figura 19 – Etapas de pré-processamento do vídeo até a extração de características



Fonte: Jéssica dos Santos de Oliveira, 2017

Essas características são utilizadas como entrada para o modelo de atenção visual, conforme será detalhado ao longo das próximas seções.

As características são relativas a um dado *pixel*. São utilizadas como características:

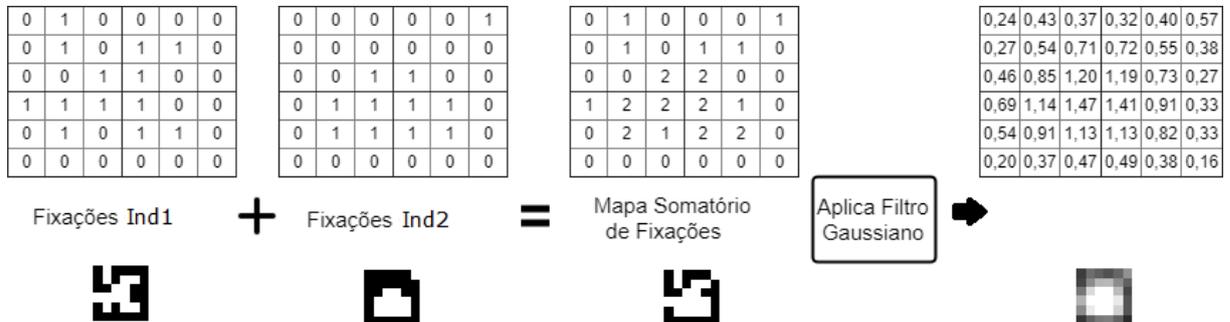
- **características 1 - 13:** pirâmides orientáveis, definidas na seção 2.3, implementadas por [Simoncelli e Freeman \(1995\)](#) com quatro escalas e três orientações;
- **características 14 - 17:** as quatro características *bottom-up*: cor, intensidade, orientação e pele, geradas utilizando a ferramenta *Saliency Toolbox* ([WALTHER; KOCH, 2006](#)). Essas características são baseadas no modelo tradicional desenvolvido por [Itti \(2000\)](#), detalhado no capítulo 3;
- **características 18 - 20:** três características que representam cores, uma para cada canal do modelo RGB descrito na seção 2.3;
- **característica 21:** a característica de linha do horizonte, em que é utilizado o detector de linha do horizonte utilizado em [Judd et al. \(2009\)](#), normalizado de zero a um, o qual assume valor máximo quando há linha horizontal;
- **característica 22:** característica referente à presença de face, extraída utilizando o algoritmo de [Viola e Jones \(2001\)](#);

- **característica 23:** característica referente à presença de pessoas, extraída também utilizando o algoritmo de [Viola e Jones \(2001\)](#);
- **característica 24:** a distância das bordas ao centro da imagem, normalizado de zero a um, o qual assume maior valor quanto mais próximo do centro;
- **característica 25:** característica de distância ao centro das cenas, adicionada porque o vídeo é composto de duas cenas lado a lado, como mostrado na seção [5.2.3](#);
- **característica 26:** uma característica referente à presença de movimento. No caso de vídeos digitais é extraído o fluxo óptico, um padrão de movimento aparente de objetos de imagem entre dois *frames* consecutivos causados pelo movimento do objeto ou da câmera. Para estimação desse fluxo foi utilizado o algoritmo de [Farneback \(2003\)](#), pois calcula o fluxo em todos os pontos da imagem.
- **características 27 - 28:** uma característica referente ao movimento biológico, ou seja, apresentam valores de movimento apenas os *pixels* que estão na região da tela onde há o movimento biológico. E uma característica referente ao movimento geométrico.

No total, 28 características foram utilizadas. Essas características foram escolhidas baseando-se nos modelos existentes na literatura, descritos no capítulo [3](#), que as definiram por meio de estudos da biologia e psicologia sobre a atenção visual humana. Além disso, algumas dessas características, como face, pessoas e movimento, também têm relevância com a atenção visual típica observada em indivíduos que pertencem ao espectro do TEA, conforme apresentado na seção [2.1.1](#).

Inicialmente o modelo foi desenvolvido com as características das imagens extraídas *frame a frame*, como se fossem imagens individuais. Porém, como mostrado nos resultados apresentados no capítulo [7](#), essa abordagem não foi eficiente para o objetivo proposto. Isso ocorre porque cada *frame* é exibido por 1/30 segundos, e neste intervalo de tempo não é possível obter fixações suficientes para uma boa análise. Portanto, a abordagem adotada foi a de agrupar os *frames* e suas características de acordo com o valor de movimento extraído entre os *frames*, pois quanto menor for o movimento, mais similares são os *frames*. Os valores de movimento são normalizados de 0 a 1. Se a média dos valores referentes ao movimento entre dois *frames* for menor do que 0,33 (valor escolhido de forma empírica analisando até o ponto que os *frames* não apresentavam uma mudança visual significativa), esses *frames* são agrupados; neste caso, os novos valores de *pixels* são compostos pela média dos *pixels* das imagens originais e suas respectivas características também são as médias das características originais.

Figura 20 – Exemplificação do processo de geração do mapa de fixação. Inicialmente são somados os valores das fixações de todos os indivíduos de um grupo (neste exemplo representados apenas por dois indivíduos). Após o somatório, é aplicado um filtro Gaussiano, que realizará uma suavização na imagem. A parte de cima da figura mostra a representação numérica das imagens.



Fonte: Jéssica dos Santos de Oliveira, 2017

### 5.3.2 Construção dos mapas de fixação

Utilizando os dados brutos extraídos do equipamento Tobii foram construídos os mapas de fixação, que são matrizes contendo os valores de fixação no *pixel* calculados como descrito a seguir e exemplificado na figura 20. Para isso, inicialmente os *pixels* em que houve fixação são separados de acordo com o *timestamp* dos respectivos *frames* (vale lembrar que cada *frame* refere-se a 1/30 segundos de vídeo). Em seguida, é construída uma matriz (mapa de fixação) do mesmo tamanho do *frame*, em que cada vez que um *pixel* foi fixado é adicionado o valor 1 na posição referente ao *pixel*. Neste mapa são somadas as fixações de todos os indivíduos pertencentes ao mesmo grupo. Portanto, há um mapa de fixação por *frame* para o grupo TEA e um mapa por *frame* para o grupo DT.

Como neste projeto houve um agrupamento de *frames*, como descrito na seção 5.3.1, os mapas com os somatórios das fixações dos *frames* que pertencem a um mesmo agrupamento também são somados.

Após a obtenção de um mapa com os somatórios de fixações para cada agrupamento de *frames*, é aplicado um filtro gaussiano, que realizará o espalhamento na imagem, deixando mais claras as regiões em que houve mais fixações e realizando um clareamento nas regiões próximas a ela. Este processamento é realizado para que as regiões muito próximas às áreas que foram fixadas não sejam desconsideradas.

A figura 21 mostra um exemplo de um mapa de fixação gerado para o grupo TEA e o mapa de fixação para o mesmo *frame* gerado para o grupo DT, bem como a imagem deste *frame*.

Figura 21 – Exemplos de Mapas de Fixação para o *frame* 198. Em (a) um mapa de fixação gerado a partir dos dados de indivíduos com DT, em (b) um mapa de fixação gerado a partir dos dados de indivíduos com TEA.



(a) Frame Original Exibido



(b) Mapa de Fixação Grupo DT



(c) Mapa de Fixação Grupo TEA

Fonte: Jéssica dos Santos de Oliveira, 2017

### 5.3.3 Seleção de características

Como mencionado na seção 2.2.1, a seleção de características se faz necessária em alguns problemas devido à maldição da dimensionalidade, quando a quantidade de características em relação ao tamanho da amostra faz com que o erro de estimação dos parâmetros prejudique o desempenho do classificador.

Os métodos de seleção de características utilizados foram o *relief* e o algoritmo genético descritos na seção 2.2.1.

O algoritmo utilizado para a seleção com o método *relief* foi o *ReliefF* (ROBNIK-SIKONJA; KONONENKO, 1997), uma variação que, ao invés de olhar um vizinho próximo, procura  $K$  vizinhos próximos da mesma classe e  $K$  vizinhos de classes distintas, o que aumenta a qualidade do estimador para os atributos.

Como limiar para a seleção de características, foi definido o valor da média de pesos retornados pelo algoritmo, ou seja, todas as características com peso maior que a média de pesos das características são selecionadas. Os pesos são calculados de acordo com o explicado na seção 2.2.1. O número de instâncias definido como parâmetro para o algoritmo foi 60. O valor de 60 instâncias foi escolhido com base nos resultados de experimentos empíricos, visto que tais resultados não melhoraram ao aumentar o número de instâncias.

Em uma primeira versão, o corte das características era feito ordenando-se os pesos de forma decrescente e selecionando-se as características até que apresentassem uma maior diferença de peso comparada com a próxima, porém esta abordagem não trouxe bons resultados.

O algoritmo genético utilizado foi o implementado por Ludwig e Nunes (2010) e utiliza como função *fitness* a informação mútua, com o objetivo de evitar a redundância entre as características e aumentar o poder de discriminação entre as classes. Para isso calcula a informação mútua entre as características e a classe menos a informação mútua entre as características. O algoritmo utilizado recebe como parâmetro a quantidade de características que devem ser retornadas. Foi delegado ao algoritmo selecionar 15 características, quantidade definida por meio de testes empíricos, visto que aumentar o número de características não aumentava o resultado da classificação.

Cada algoritmo gera, portanto, um subconjunto de características diferente. Os vetores de características correspondentes a esses subconjuntos e também o correspondente a todas as características foram utilizados como a entrada para o método de classificação especificado na seção 5.3.5, separadamente para cada subconjunto de características. As características selecionadas e os resultados da classificação utilizando tais características são apresentados no capítulo 7.

#### 5.3.4 Seleção de *pixels*

A partir dos *pixels* do mapa de fixação, representados por seus vetores de características, um modelo de atenção visual será aprendido. Para esse aprendizado, foi selecionada uma

quantidade igual de *pixels* positivos (fixados) e negativos (não-fixados) de cada imagem. Duas abordagens foram testadas para essa seleção.

A primeira abordagem, seguindo os algoritmos de [Borji \(2012\)](#) e [Judd et al. \(2009\)](#), seleciona tantos os *pixels* positivos quanto os negativos de forma aleatória. Inicialmente foram selecionados 15 *pixels* positivos e 15 *pixels* negativos de cada *frame*, seguindo as definições previstas nos trabalhos utilizados como base. No entanto, diferentemente da literatura, no presente trabalho são empregados vídeos. Por tratar-se de imagens dinâmicas, em que cada *frame* é exibido muito rapidamente, a quantidade de 30 *pixels* para cada agrupamento de *frames* mostrou-se pequena, sendo insuficiente para resultar em um aprendizado efetivo. Assim, a quantidade de *pixels* selecionados foi aumentada para 350 positivos e 350 negativos. O número 350 foi escolhido com testes empíricos, sendo estabelecido o menor valor que maximizava a acurácia do modelo, ou seja, a partir deste valor a acurácia se manteve constante. Além disso, esta abordagem de seleção completamente aleatória não trouxe bons resultados, sendo substituída por uma segunda abordagem.

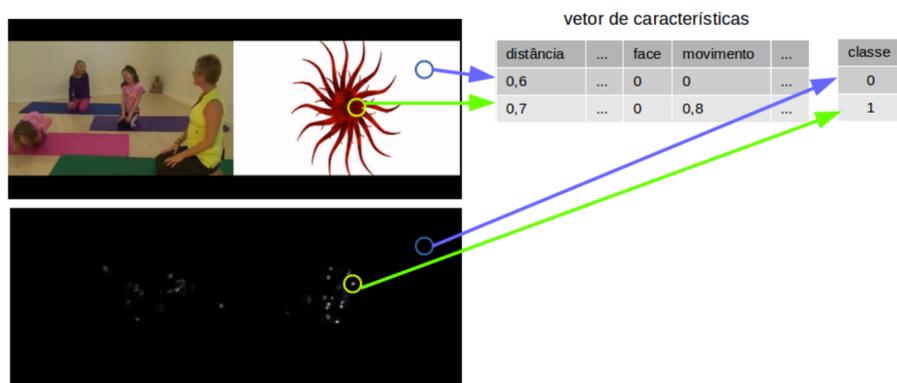
A segunda abordagem reorganiza os *pixels* do mapa de fixação em um vetor, considerando ordem decrescente, ou seja, os *pixels* com maior quantidade de fixação ficam no início do vetor e os *pixels* sem nenhuma fixação no final. A seleção dos *pixels* positivos considera, dentro de cada *frame*, os 350 *pixels* que foram mais fixados por todo o grupo, ou seja, são selecionadas as 350 primeiras posições do vetor referente a cada *frame*. Já os *pixels* negativos são selecionados aleatoriamente dentre todos os *pixels* negativos, ou seja, aqueles com valor zero, permitindo que qualquer uma das regiões não fixadas seja selecionada. Após a seleção, os *pixels* positivos recebem valor 1. Esses dois valores de fixação, 0 para *pixels* negativos e 1 para *pixels* positivos, representam os rótulos dessas duas classes de *pixels* a serem aprendidas para compor o mapa de saliência, conforme descrito na seção [5.3.5](#).

### 5.3.5 Modelo de atenção visual

De posse de uma amostra de *pixels* fixados (positivos) e não fixados (negativos) pelos dois grupos, TEA e DT, o próximo passo é aprender a prever a atenção desses dois grupos, ou seja, aprender um modelo de atenção visual para cada grupo. A [figura 22](#) mostra uma representação simples de como é formado o conjunto de treinamento. Em cada agrupamento de *frame* utilizado para treinamento, são selecionados os *pixels* positivos e negativos, e então as características

de cada um desses *pixels* são extraídas, formando o vetor de características. Utilizando essas informações, os métodos de aprendizado irão aprender uma função que, dado os valores das características, fornece a probabilidade daquele *pixel* ser fixado. Como já mencionado, será aprendida uma função para TEA e uma função para DT, cada uma prevendo a atenção visual de seu respectivo grupo. Os métodos de aprendizagem utilizados foram Redes Neurais e SVM, descritos na seção 2.2.

Figura 22 – Exemplificação da formação do vetor de características para um *frame*. Na figura é mostrado o exemplo de um *frame* utilizado, um mapa de fixação do grupo TEA e o exemplo de duas linhas do vetor de características finais, uma para a classe positiva e outra para a classe negativa.



Fonte: Jéssica dos Santos de Oliveira, 2017

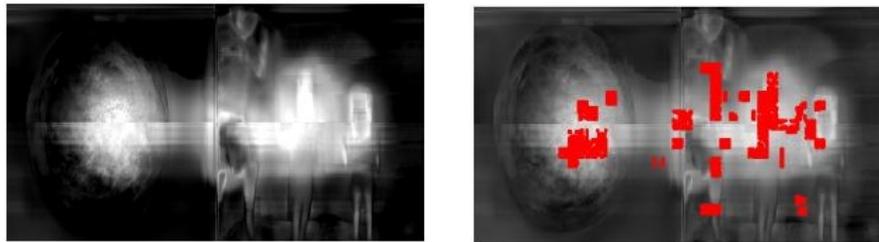
A Rede Neural utilizada tem como função de ativação a função sigmóide na camada oculta e a linear na camada de saída. O método de treinamento utilizado foi o de *backpropagation* com regularização *bayesiana*. Foram utilizadas 10 neurônios na camadas oculta e a condição de parada foi a de atingir 1000 épocas ou o erro ser menor que  $1e-7$ . Foram realizados testes aumentando-se a quantidade de neurônios na camada oculta para 20 e 50, porém isso não trouxe melhoras nos resultados e voltou-se para o valor inicial de 10. Já os parâmetros da condição de parada são os valores padrões do algoritmo utilizado e não foram alterados.

O algoritmo de SVM utilizado foi o linear, da biblioteca *liblinear* (FAN et al., 2008), que implementa uma SVM Linear. O parâmetro passado para o algoritmo foi o custo de 0,1. Também foram feitos testes com valores de custo de 0,01 e 0,001, porém, por não trazer melhoras nos resultados, foi mantido o valor inicial de 0,1.

Após o aprendizado, o modelo de atenção visual criado é capaz de reconhecer, dada uma imagem de teste, onde seria a saliência daquela imagem, ou seja, quais seriam os pontos de atenção. Para isso, para cada *frame* testado, a saída do modelo é uma matriz de mesma dimensão do *frame* com o valor predito para cada um dos *pixels*, normalizados no intervalo [0,1]. Essa

matriz é chamada de mapa de saliência, e quanto maior o valor de um *pixel*, maior a chance daquele *pixel* ser fixado. Para fins de visualização, esses valores podem ser transformados em uma imagem, na qual valores maiores indicam *pixels* mais claros. A figura 23 mostra um mapa de saliência gerado como saída para um dos *frames* por meio do modelo aprendido para o grupo DT, bem como os *pixels* que foram fixados pelo grupo DT, apresentados em vermelho.

Figura 23 – Exemplo de mapa de saliência: à esquerda um mapa de saliência de um *frame* gerado para o grupo DT; à direita os pontos em vermelho são os *pixels* que de fato foram fixados por esse grupo no *frame*.



Fonte: Jéssica dos Santos de Oliveira, 2017

### 5.3.6 Método de classificação de indivíduos

Para realizar a classificação de um indivíduo, um processo semelhante ao de geração de mapa de fixação detalhado na seção 5.3.2 é utilizado para gerar os mapas individuais de fixação.

Cada mapa de fixação individual é representado por uma matriz binária, que retrata o *frame* com o valor 1 nos *pixels* que foram fixados por aquele indivíduo e zero nos *pixels* restantes. Em média, um indivíduo fixa 200 *pixels* em um agrupamento de *frames*. Assim como nos mapas de fixação dos grupos (seção 5.3.2), é aplicado um filtro gaussiano ao mapa de fixação individual.

Na sequência, para cada agrupamento de *frames*, o mapa de fixação do indivíduo é comparado ao mapa de saliência de TEA e ao mapa de saliência de DT, ambos do respectivo agrupamento de *frames*, gerados conforme descrito na seção 5.3.5. Para realizar essa comparação, os mapas de saliência são antes limiarizados. O processo de limiarização empregado faz com que a matriz final obtida tenha somente dois valores: zeros ou uns. O valor de limiar usado foi 0,75, isto é, células com valores abaixo de 0,75 assumem o valor zero e as demais assumem o valor um. Assim, apenas os valores acima de 0,75 são considerados como os valores que foram fixados. Com este procedimento desprezam-se as células da matriz que apresentaram fixação pouco significativa dentro de cada grupo.

Após o procedimento de limiarização, passa-se à comparação entre o mapa de fixação individual e a matriz de saliência. Os *pixels* que foram preditos como pontos de fixação pelo mapa de saliência, ou seja, estão com valor um, e que foram fixados no mapa de fixação individual, assim como os que não foram preditos como pontos de fixação pelo mapa de saliência e que não foram fixados no mapa individual, são considerados como *matches*, ou seja, houve uma concordância entre o modelo de atenção visual aprendido e a fixação daquele indivíduo.

Assim, para um dado indivíduo é obtida a quantidade de *matches*, em cada agrupamento de *frame*, no modelo TEA e no modelo DT. Esse valor será utilizado para avaliar se um indivíduo teve a fixação mais próxima do mapa de saliência predito pelo modelo de TEA ou pelo predito pelo modelo de DT. Ou seja, se a quantidade total de *matches* for maior ao comparar o mapa de fixação individual com o mapa de saliência do grupo TEA do que com o mapa de saliência do grupo DT em um determinado agrupamento de *frames*, isso é contado como um voto para a classe de TEA. Se após todos os agrupamentos de *frames* darem um voto o indivíduo obtiver uma quantidade de votos de TEA maior do que um determinado limiar, então é classificado como pertencente ao TEA. Este limiar é alterado indo de 0 até a quantidade total de agrupamentos de *frames* (número máximo de votos), para a construção da curva ROC, conforme explicado na seção 5.3.7.

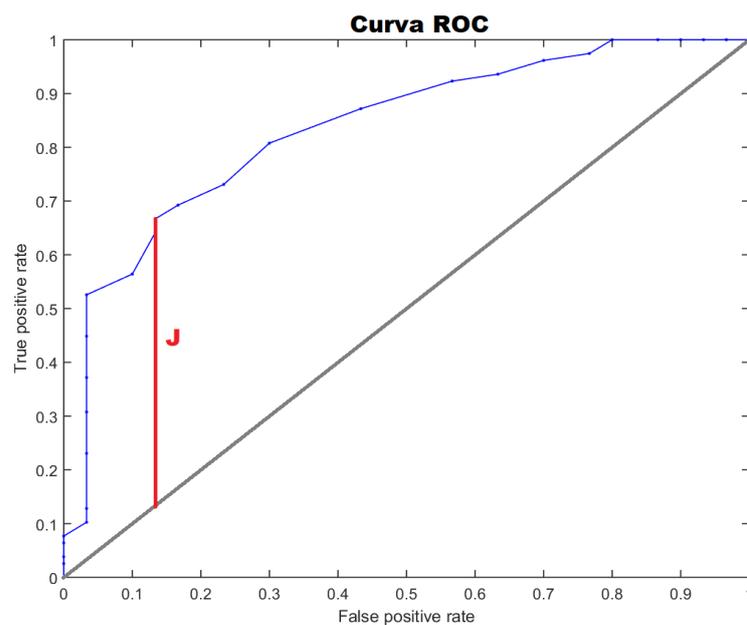
### 5.3.7 Avaliação do modelo desenvolvido

Para avaliação do classificador de indivíduos foi utilizado o método de validação cruzada (*k-fold cross-validation*). Esta técnica divide o conjunto de dados em  $k$  partes (*folds*) e executa  $k$  rodadas de treinamento e teste. Em cada execução, uma parte do conjunto de dados (ou seja,  $1/k$  dos dados) é utilizada para teste e as demais  $k - 1$  partes são utilizadas para treinamento. Neste trabalho foi utilizado  $k$  igual a 5. Desta forma, o conjunto de treinamento continha aproximadamente 135 agrupamentos de *frames* selecionados aleatoriamente em cada execução, e o restante dos *frames* eram utilizados para testes.

Para avaliar o desempenho de classificação, criou-se uma curva ROC variando o limiar da quantidade de votos TEA necessária para que o indivíduo fosse classificado como tal. Os resultados são mostrados no capítulo 7, no qual são comparados os métodos de aprendizado utilizados.

Também é calculado, utilizando a curva ROC, o limiar de votos TEA que retorna o melhor resultado de classificação. Este limiar é calculado utilizando o índice de Youden (YOUDEN, 1950), criado para avaliar desempenho de testes de diagnóstico e comumente utilizado para definir critério de ponto de corte na curva ROC. O objetivo do índice é maximizar a distância entre o TPR (*True Positive Rate*), equivalente à sensibilidade do classificador, e o FPR (*False Positive Rate*) equivalente a  $(1 - \text{especificidade})$  do classificador. Para isso calcula-se em cada ponto da curva o índice  $J = \text{sensibilidade} + \text{especificidade} - 1$  e encontra-se o máximo de J. A figura 24, exemplifica de maneira visual o índice de Youden, que como pode ser visto, também pode ser interpretado como o ponto que maximiza a distância na vertical entre a curva resultante do classificador e a linha diagonal que representa uma classificação puramente ao acaso.

Figura 24 – Representação gráfica de uma curva ROC com o índice de Youden (J), ponto que máxima a distância entre a curva e a linha de equidade (em cinza).



Fonte: Jéssica dos Santos de Oliveira, 2017

### 5.3.8 Considerações finais

Como descrito nesta seção, diversas abordagens foram testadas em cada uma das etapas do desenvolvimento, a fim de definir o método final que retornasse o melhor resultado para a classificação. O método final proposto é descrito no capítulo 6.

## 6 Classificador de TEA baseado em um modelo de atenção visual para auxílio ao diagnóstico

A figura 25 mostra o fluxograma do modelo final desenvolvido para classificação de indivíduos como pertencentes aos grupos de TEA ou DT. O modelo desenvolvido é composto de dois módulos: modelo de atenção visual e o classificador de TEA, detalhados na seções a seguir. As cores no diagrama representam quais dados de entrada são processados pela etapa. As etapas em amarelo processam os dados de fixação; as etapas em rosa processam os estímulos visuais; as em azul são responsáveis por processar todos os dados presentes no módulo (no caso do módulo do modelo de atenção, processa os dados de fixação e os estímulo, já no módulo de classificação, processa também o modelo treinado), a laranja processa os estímulos e os dados de fixação no módulo de classificação e a roxa processa os estímulos e o modelo treinado.

### 6.1 Modelo de atenção visual

O Modelo de Atenção Visual é responsável por predizer a atenção visual dos indivíduos. Para isso, processa dados de rastreamento do olhar com a finalidade de encontrar um padrão de atenção visual para cada grupo estudado (TEA e DT), considerando-se as características do estímulo disponibilizado.

Neste módulo foram processados os dados de rastreamento de olhar para a obtenção de dois modelos: um para predizer a atenção visual dos indivíduos com TEA e outro para predizer a atenção visual dos indivíduos com DT. As etapas para o desenvolvimento dos dois modelos são as mesmas:

1. **Separação e pré-processamento dos *frames* do estímulo:** esta atividade é responsável por dividir os estímulos visuais utilizados, que estão em formato de vídeo, em *frames*, e realizar o pré-processamento para que, posteriormente, as características sejam extraídas destes *frames*. O pré-processamento consiste em remover as bordas ao redor da imagem, redimensionar os *frames* para uma resolução de 200x350 *pixels* e remover os *frames* de transição entre dois vídeos. Essas etapas foram descritas na seção 5.3.1.
2. **Elaboração dos mapas de fixação:** nesta etapa os dados brutos extraídos do rastreador (as coordenadas dos *pixels* fixados e *timestamp*) são utilizados para gerar os mapas de fixação. Para isso, as coordenadas dos *pixels* que foram fixados no intervalo de tempo que corresponde a um *frame* recebem o valor 1 na matriz que representará o *frame*. As matrizes

Figura 25 – Fluxograma com o funcionamento do modelo proposto. São divididos em dois processos: Modelo de Atenção Visual e Classificador de TEA. As saídas dos processos, representadas em verde, são: para o modelo de atenção visual um modelo treinado, esse processo é repetido para cada um dos grupos, tendo ao final um modelo treinado para TEA e um modelo treinado para DT; para o classificador TEA a classe predita (1 para TEA e 0 para não-TEA), o classificador tem como entrada, além dos dados de fixação do indivíduo e do estímulo visual assistido, os modelos treinados no processo anterior.



Fonte: Jéssica dos Santos de Oliveira, 2017

de todas os indivíduos de um grupo (TEA ou DT) para cada *frame* são somadas gerando um único mapa com o somatório das fixações. Ao final é aplicado um filtro gaussiano. Mais detalhes sobre essa implementação foram descritos na seção 5.3.2.

3. **Extração de características:** neste passo são executadas técnicas de processamento de imagens para extrair características dos *frames* selecionados no passo anterior. As características finais extraídas foram apresentadas na seção 5.3.1.
4. **Agrupamento de *frames*:** as características extraídas dos *frames* e os mapas de fixação correspondentes a eles são agrupados quando o valor de movimento entre os *frames* for menor do que 0,33. O valor das características dos *frames* agrupados são as médias das características extraídas dos *frames* originais. Já com os mapas de fixação, são realizadas somas nos *pixels* fixados. Esta etapa foi apresentada com mais detalhes nas seções 5.3.1 e 5.3.2.
5. **Seleção de *pixels*:** para cada um dos agrupamentos de *frames* utilizados como treinamento são selecionados 350 *pixels* com valores positivos e 350 com valores negativos. Os *pixels* positivos são selecionados ordenando o mapa de fixação decrescentemente pelo número de fixações e escolhendo os maiores valores, ou seja, são selecionados os *pixels* que tiveram maiores fixações para o grupo. Os *pixels* negativos são selecionados aleatoriamente entre todos os *pixels* com valor zero. Após a seleção, os mesmos são binarizados, de forma que todos os *pixels* positivos recebam valor 1.
6. **Geração do vetor de características:** os *pixels* selecionados passam a representar a amostra rotulada de treinamento do modelo de aprendizado, sendo cada *pixel* rotulado como classe 1 (houve fixação) ou classe 0 (não houve fixação). As características extraídas para aquele *pixel*, e escolhidas previamente de acordo com a seleção realizada pelo Algoritmo Genético, formam o vetor de características do modelo. Portanto, para cada agrupamento de *frames* há uma matriz com 700 linhas (quantidade de *pixels* selecionados) e 28 colunas (número total de características). Os vetores de características, com suas respectivas classes, são as entradas para os métodos de aprendizado.
7. **Treinamento do modelo:** os vetores de características, com suas respectivas classes, são as entradas para os métodos de aprendizado, que têm como tarefa aprender a prever, a partir de novos valores de entrada, se aquele será ou não um ponto de saliência (atenção). O método de aprendizado utilizado na proposta final foi a Rede Neural, por apresentar melhor desempenho, conforme apresentado no capítulo 7. O modelo foi treinado e validado utilizando o método de *cross-validation*, sendo que em cada *fold* haviam aproximadamente 135 agrupamentos de *frames* para o treinamento.

Todas as etapas são realizadas duas vezes, uma com os dados dos indivíduos com TEA, gerando um modelo de atenção visual para TEA e outra com os dados dos indivíduos com DT, gerando um modelo de atenção visual para DT. Esses modelos são utilizados no classificador, predizendo a atenção visual dos dois grupos em relação aos estímulos visuais utilizados, possibilitando a comparação com a fixação do indivíduo a ser testado.

## 6.2 Classificador TEA versus DT

Este módulo é responsável por classificar a qual grupo (TEA ou DT) um indivíduo pertence. Para isso, utiliza as fixações realizadas por esse indivíduo em relação a um estímulo visual e compara com a atenção visual predita para o mesmo estímulo visual, considerando os modelos desenvolvidos para os dois grupos (TEA e DT). Os estímulos visuais utilizados na etapa de classificação foram os mesmos utilizados para a concepção do modelo de atenção visual, porém o método permite que outros estímulos sejam utilizados na classificação, já que o modelo deveria reconhecer a atenção em qualquer estímulo visual. As etapas realizadas para a classificação são descritas a seguir:

1. **Separação e pré-processamento dos frames:** as mesmas tarefas de separação e pré-processamento dos *frames* utilizadas para a geração do modelo de atenção visual são aplicadas, mas agora considerando os estímulos visuais utilizados para teste, caso sejam diferentes dos utilizados no modelo. As etapas são: separação do vídeo em *frames*; descarte dos *frames* de transição; corte das bordas e redimensionamento, descritos em mais detalhes na (seção 5.3.1).
2. **Geração dos mapas de fixação individual:** os mapas de fixação do indivíduo são gerados da mesma forma que o mapa de fixação utilizado no treinamento do modelo de atenção visual. Primeiro toda vez que um *pixel* é fixado por aquele indivíduo em um determinado *frame* agrupado, é somado 1 na posição do *pixel* na matriz de fixação que representa aquele *frame*. Após todos os *pixels* fixados serem preenchidos na matriz, é aplicado um filtro gaussiano, conforme detalhado na seção 5.3.6.
3. **Geração dos mapas de saliência:** aplicando os modelos de atenção visual treinados nos estímulos visuais utilizados no teste, o modelo de atenção visual treinado retorna para cada estímulo uma matriz contendo valores normalizados no intervalo [0,1], em que quanto mais próximo de 1 mais saliente aquele *pixel* é. Essa matriz é chamada de mapa de saliência.

4. **Comparação do mapa de saliência com o mapa de fixação:** os mapas de saliência são limiarizados a 0,75, ou seja, *pixels* com valores abaixo de 0,75 assumem o valor zero e os demais assumem o valor um. Os *pixels* que estão com valor um no mapa de saliência e que foram fixados segundo o mapa de fixação individual são somados aos *pixels* que estão zerados no mapa de saliência e que não foram fixados no mapa de fixação. Esses *pixels* são os *pixels* que representam um *match* entre o mapa de fixação do indivíduo e o mapa de saliência gerado pelo modelo. Quando maior a quantidade de *matches* mais similares são os mapas.
5. **Classificação por sistema de votação:** para classificar se um indivíduo pertence ao grupo de TEA ou de DT, é realizada uma votação, comparando os mapas de fixação aos mapas de saliência gerados pelos modelos treinados de TEA e de DT para cada agrupamento de *frames*. Se para um determinado agrupamento de *frames* houver uma maior quantidade de *pixels* que deram *match* entre o mapa de fixação do indivíduo e o mapa de saliência gerado pelo modelo de TEA do que entre o mapa de fixação do indivíduo e o mapa de saliência gerado pelo modelo de DT, naquele agrupamento de *frames* o voto será para TEA, caso contrário o voto será para DT. Se ao término da comparação de todos os agrupamentos de *frames* testados a quantidade de votos para TEA for maior do que 28 agrupamentos de *frames*, esse indivíduo é classificado como TEA. Esse valor limiar para a classificação de TEA foi definido por meio da análise da curva ROC gerada variando os limiares, a fim de avaliar o desempenho do método, conforme descrito na seção 5.3.7.

O modelo descrito foi avaliado considerando as métricas apresentadas na seção 5.3.7. Os resultados obtidos são apresentados no capítulo 7.

## 7 Resultados e Discussões

### 7.1 Seleção de características

Uma das primeiras tarefas do método proposto é a seleção de características. Nesta primeira etapa do trabalho trata-se de selecionar as características que mais separam os *pixels* fixados dos não fixados. A tabela 3 mostra quais das características extraídas (seção 5.3.1) foram selecionadas para cada um dos grupos utilizando os dois métodos testados, *reliefF* e algoritmo genético (seção 5.3.3), discutidas a seguir.

Tabela 3 – Características selecionadas pelos dois métodos de seleção utilizados.

Característica	Algoritmo Genético		ReliefF	
	TEA	DT	TEA	DT
<b>1 - 13 pirâmides</b>	3	4	4	5
<b>14 - 17 Itti</b>	4	4	2	2
<b>18 - 20 Cores</b>	0	1	3	2
<b>21 - Linha do Horizonte</b>	1	1	1	1
<b>22 - Face</b>	1	1	1	0
<b>23 - Pessoa</b>	1	1	0	0
<b>24 - Centro</b>	1	0	1	1
<b>25 - Movimento</b>	1	1	0	0
<b>26 - Movimento Lado Biológico</b>	1	0	0	0
<b>27 - Movimento Lado Geométrico</b>	1	1	0	0
<b>28 - Centro Lados</b>	1	1	1	1
<b>Total</b>	15	15	13	12

#### 7.1.1 Algoritmo genético

Utilizando o algoritmo genético nenhuma característica relacionada a cor foi selecionada para representar o grupo TEA. Já a característica relacionada ao centro da imagem foi selecionada apenas para o grupo TEA. Esses achados estão de acordo com os resultados encontrados por Wang et al. (2015b), que ao capturarem a fixação de participantes com TEA em várias imagens, perceberam que o grupo testado teve um maior foco ao centro da imagem, mesmo quando não havia nada no centro.

As características relacionadas ao modelo de Itti (2000) (seção 3) foram todas selecionadas para ambos os grupos, o que fornece indícios da relevância biológica de tais características, mesmo para indivíduos com um transtorno como o TEA. Para o grupo com DT, a característica

relacionada ao movimento na região biológica não foi selecionada. Este fato pode ser explicado pela característica ser genérica, cobrindo toda a região do vídeo que continha a parte biológica, e a atenção do grupo com DT ser mais focada, específica nas regiões onde haviam pessoas e faces, já cobertos pelas outras características.

As características selecionadas pelo algoritmo genético mostram-se plausíveis com os estudos anteriormente realizados, apresentados na seção 2.1.1 e no capítulo 4, referentes a importância no movimento biológico e geométrico, centro da imagem, pessoas e faces na atenção visual de indivíduos com TEA. Este grupo de características apresentou ótimos resultados utilizando Rede Neural, quando comparado aos resultados obtidos utilizando os demais grupos de características, como será detalhado na seção 7.2. Este grupo de características selecionado pelo algoritmo genético foi o escolhido para o modelo final.

### 7.1.2 Algoritmo *ReliefF*

O algoritmo *ReliefF*, além de selecionar as características, mostra a ordem das características que mais separam os grupos. Para ambos os grupos (TEA e DT), as quatro características mais relevantes foram: linha horizontal, centro da imagem, centro dos lados e característica relacionada às pirâmides de [Torralba \(2003\)](#).

As características selecionadas pelo *ReliefF* não são as esperadas de acordo com os estudos realizados. Por exemplo: nenhuma característica relacionada ao movimento foi selecionada e a característica relacionada à presença de face foi selecionada apenas para o grupo TEA. Apesar disso, os resultados apresentados por esse grupo de características foram razoáveis, com uma média de 0,78 de AUC utilizando Redes Neurais, conforme será discutido na seção 7.2. Isso indica que, apenas com as características iniciais básicas de [Itti \(2000\)](#) já é possível criar um modelo capaz de prever a atenção de ambos os grupos.

Porém, ao comparar o resultado obtido usando as características selecionadas pelo algoritmo *ReliefF* com o resultado dos métodos utilizando as características selecionadas pelo Algoritmo Genético, e mesmo com os resultados utilizando todas as características do modelo, o grupo de características selecionadas pelo *ReliefF* é o que apresenta os piores resultados, com uma média de AUC 0,04 menor que os outros métodos, conforme detalhado na seção 7.2.

A partir do cenário apresentado, ressalta-se o fato de que a utilização de características que se baseiem em estudos específicos para o objetivo proposto, no caso estudos da área da psiquiatria sobre paradigmas da atenção visual do TEA, trazem resultados mais satisfatórios.

## 7.2 Desempenho dos classificadores

Como descrito no capítulo 5, algumas diferentes abordagens foram testadas até a definição final do modelo proposto. Nesta seção são mostrados os resultados parciais dessas abordagens, bem como o resultado do modelo final com as diferentes técnicas de seleção e classificação. As abordagens na etapa de preparação de dados foram apenas testadas com o método de Redes Neurais Artificiais utilizando todas as características extraídas, já a versão final do modelo foi testada com Redes Neurais e SVM como classificadores e utilizando três grupos de características diferentes: todas as características, características selecionadas pelo método *ReliefF* e características selecionadas pelo algoritmo genético.

### 7.2.1 Abordagens de Preparação de Dados

A abordagem inicial de pré-processamento dos dados foi realizada baseando-se no método de Borji (2012). Foram utilizados todos os *frames* dos vídeos de forma individual, utilizando-os como se fossem imagens estáticas. Em cada *frame* foram selecionados 15 *pixels* positivos e 15 negativos e, foi utilizado como métrica o método de comparação de mapa de saliência com o mapa de fixação desenvolvido por Borji (2012) – método que divide o *frame* em várias pedaços e seleciona *pixels* aleatórios em cada pedaço para fazer a comparação entre o mapa de fixação e o mapa de saliência. Com este conjunto de técnicas a AUC foi de 0,55. Em outras palavras, com esta abordagem não foi possível distinguir os grupos, tendo como resultados respostas aleatórias.

Mudando apenas o método que calcula a métrica de comparação entre os mapas para o descrito na seção 5.3.6, a média da AUC aumentou para 0,62. Uma hipótese para essa melhora é que a simplificação do método, fazendo apenas a somatória dos *matches*, e o fato do mesmo comparar toda a imagem, e não somente alguns *pixels*, pode ter melhorado a acurácia na classificação.

Porém, ainda assim as combinações de métodos apresentadas estavam como uma AUC baixa para que pudessem ser utilizados como auxílio ao diagnóstico. Uma possível causa é que

pelo estímulo se tratar de um vídeo com 30 fps (*frames* por segundo), cada *frame* é observado somente em alguns milissegundos, o que pode fazer com que não tenham pontos de fixação suficientes em um *frame* para que a análise seja feita de forma correta. Para solucionar este problema foi feito o agrupamento dos *frames*, como descrito na seção 5.3.1. Assim, havia mais pontos de fixação nos *frames* agrupados, o que permitiu também selecionar mais pontos em cada *frame*. Foram extraídos 350 *pixels* positivos e 350 *pixels* negativos de cada *frame* agrupado. Com essa abordagem adotada o modelo aprendido obteve uma AUC média de 0,82.

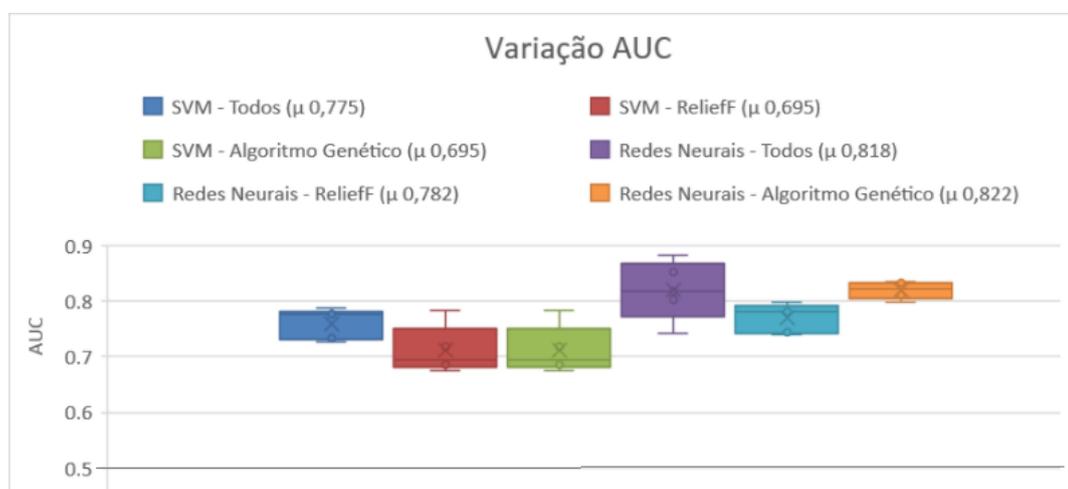
Esta última abordagem, com o agrupamento de *frames*, selecionando 350 *pixels* positivos e 350 *pixels* negativos em cada agrupamento e calculando ao final os *matches* entre o mapa de fixação e o mapa de saliência, foi a abordagem utilizada para comparação entre os métodos de seleção e aprendizado.

### 7.2.2 Comparação de métodos de seleção e aprendizado

Como citado, dois métodos de seleção, *reliefF* e algoritmo genético, e dois métodos de aprendizado, redes neurais e SVM, foram testados e comparados com o objetivo de desenvolver o melhor modelo de atenção visual e, conseqüentemente, o melhor classificador.

Os resultados das AUC obtidas em cada combinação de métodos são apresentados em forma de *boxplot* na figura 26, onde cada cor representa uma combinação. Os maiores valores são obtidos pela rede neural, tanto na média das AUC's, ficando acima de 0,78, quanto no valor máximo, atingindo 0,88, como é possível observar pela figura.

Figura 26 – Gráfico de *box-plot* exibindo as distribuições das AUC's em cada um dos métodos executados.



Fonte: Jéssica dos Santos de Oliveira, 2017

Cada uma das curvas ROC obtidas são apresentadas na figura 27. Cada linha representa uma execução da validação cruzada (seção 5.3.7).

Como é possível observar, em relação ao classificador, os métodos que utilizaram SVM (figuras 27d, 27e e 27f) apresentaram um desempenho mais baixo se comparados com os que utilizaram Redes Neurais (figuras 27a, 27b e 27c). De modo geral as curvas foram mais baixas, apresentando a maior AUC e 0,778 (figura 27f), obtida ao serem utilizadas todas as características e, a maior média foi de 0,775, também observada com todas as características.

As curvas utilizando Redes Neurais, além de terem áreas maiores, também apresentam maiores proximidades, com desvio padrão médio de 0,028, enquanto que a SVM teve desvio padrão médio de 0,34. Isso indica maior estabilidade entre o método com Redes Neurais. Apenas a execução utilizando todas as características (27c é que apresentam distâncias maiores entre as curvas.

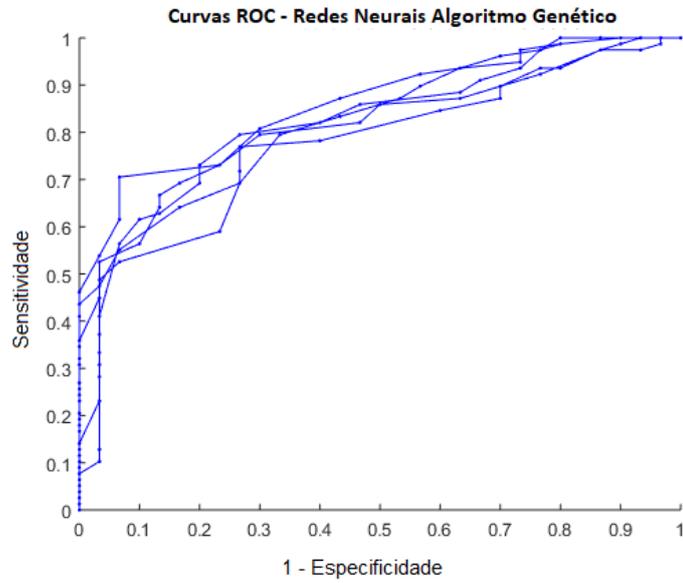
Em relação aos métodos de seleção de características, o ReliefF (figura 27b e 27e) apresentou os piores desempenhos, com médias de 0,69 com SVM e 0,78 com redes neurais, conforme mostrado na figura 26.

As médias entre Redes Neurais utilizando todas as características e utilizando Algoritmos Genéticos ficaram bem próximas, sendo a média da AUC com algoritmos Genéticos ficando apenas 0,004 acima da com todas as características. Porém, as variações das AUC's utilizando todas as características foi muito maior do que utilizando apenas as características selecionadas pelo Algoritmo Genético, indo de 0,741 a 0,882. Portanto, apesar de o método utilizando todas as características apresentar a maior AUC, foi escolhido o método utilizando apenas as características selecionadas pelo Algoritmo Genético como método final, por apresentar resultados mais estáveis em todas as execuções e ter uma média de AUC ligeiramente maior do que com todas as características.

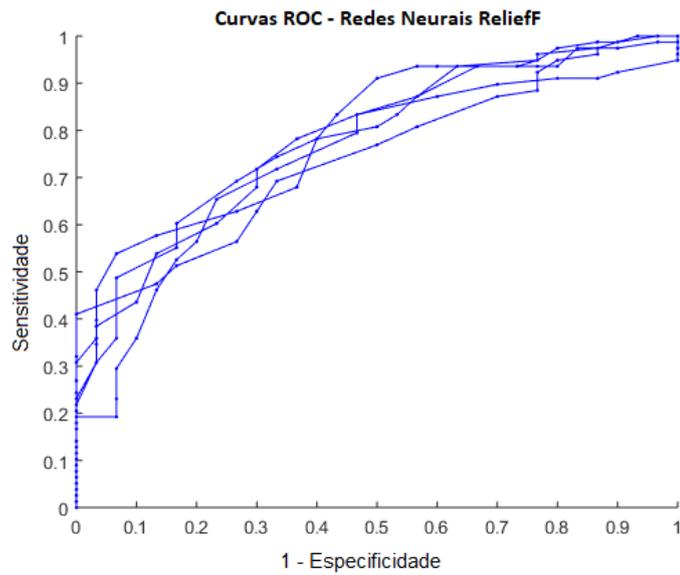
Utilizando a moda do índice de Youden nas curvas ROC do método com Redes Neurais e Algoritmo Genético, o melhor valor de limiar obtido foi igual a 28, sendo o limiar encontrado pelo índice de Youden em 3 das 5 curvas. O que significa que dos 50 *frames* utilizados de teste nesta rodada, se a fixação de um indivíduo for mais similar ao modelo de atenção visual de TEA do que o de CTRL em pelo menos 28 frames, esse indivíduo será classificado como TEA. Utilizando esse limiar, esse classificador correspondente à AUC máxima apresentou como valor de precisão de 0,96, especificidade de 0,93, sensibilidade de 0,70 e erro de 0,23.

Se este mesmo limiar for aplicado às outras rodadas do mesmo modelo com Rede Neural e todas as características, é obtida uma precisão média de 0,9, especificidade de 0,83, sensibilidade

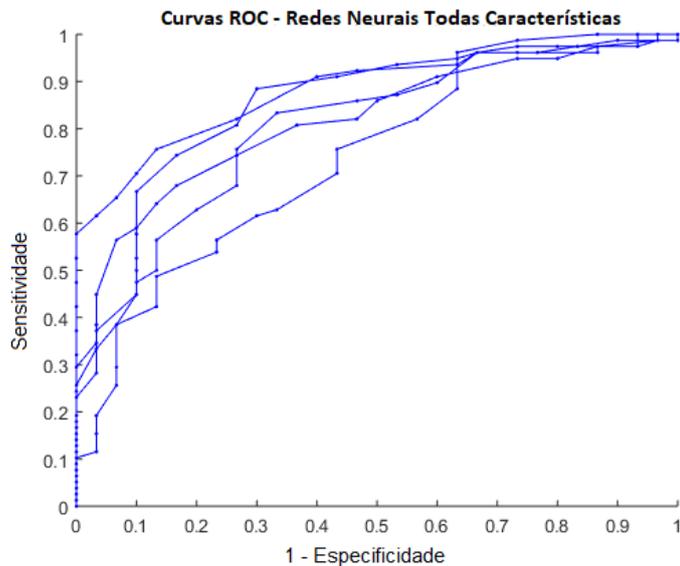
Figura 27 – Curvas ROC obtidas para todos os modelos nas 5 execuções da validação cruzada



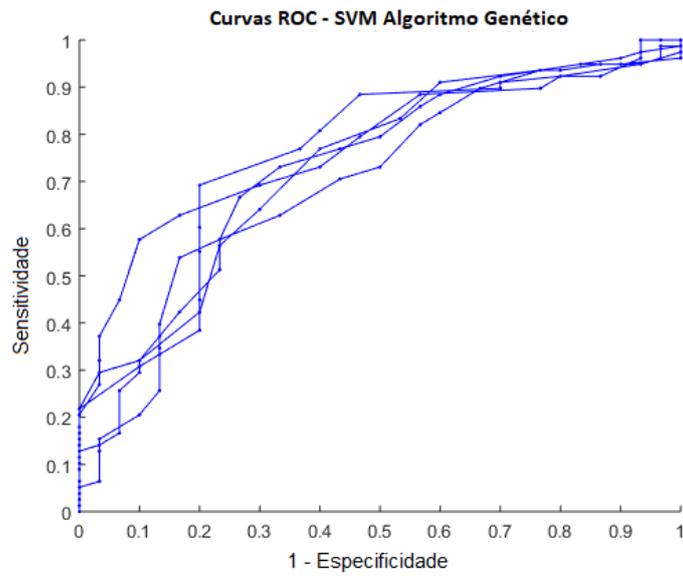
(a) Curvas ROC para Redes Neurais com as características selecionadas pelo Algoritmo Genético



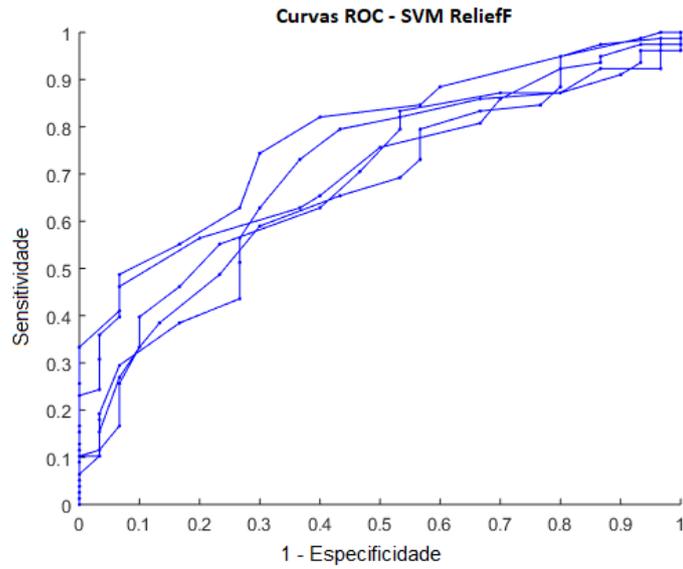
(b) Curvas ROC para Redes Neurais com as características selecionadas pelo ReliefF



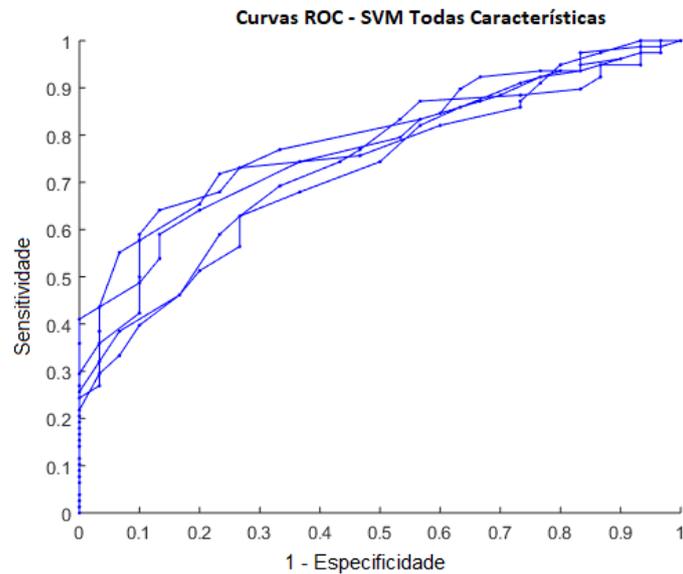
(c) Curvas ROC para Redes Neurais com todas as características



(d) Curvas ROC para SVM com as características selecionadas pelo Algoritmo Genético



(e) Curvas ROC para SVM com as características selecionadas pelo ReliefF



(f) Curvas ROC para SVM com todas as características

de 0,69 e erro de 0,27. Com o modelo já aprendido, o ideal seria que o teste fosse realizado em um novo vídeo, com paradigma similar, que pudesse ser aplicado por inteiro, para uma validação e avaliação mais precisa do modelo. Porém, não houve dados suficientes para testar essa abordagem.

As seções seguintes discutem alguns aspectos que podem ter influenciado no resultado final do modelo.

### 7.3 Experimento e coleta de dados

O experimento de coleta de dados foi parte crucial do projeto. Como visto na revisão sistemática (capítulo 4), não é uma tarefa simples fazer a coleta de dados com indivíduos com TEA; muitos deles apresentam *déficit* de atenção e até mesmo comportamentos agressivos. Porém, ter dados coletados adequadamente, com redução de *missing values*, é de extrema importância para que um bom modelo computacional seja construído.

Percebeu-se que o experimento sendo realizado com a presença de duas pessoas, uma para manusear o equipamento e um cuidador para manter a atenção exclusiva da criança ou adolescente com TEA fez com que as porcentagens do total de captura não fossem baixas, como pode ocorrer com TEA, casos muitas vezes mostrados na literatura. Porém, esse cuidador precisa ter conhecimentos de como agir com um indivíduo com TEA; para isso, o auxílio da equipe do PROTEA foi fundamental no desenvolvimento deste projeto. A calibragem do equipamento utilizando uma animação, e trocando a animação quando a primeira calibragem falhava, também foi um ponto crucial para a qualidade dos dados capturados. O protocolo de coleta de dados como um todo, definido na seção 5.2.4, foi definido para que ocorresse a menor perda de dados possível.

Outro aspecto em relação à captura de dados foi a falta de voluntários do grupo DT para participar do projeto, o que também dificultou o pareamento dos indivíduos e a formação de uma base balanceada. Este problema seria menor se os projetos citados na literatura, que fizeram o rastreamento do olhar com indivíduos com e sem o TEA, disponibilizassem suas bases de dados para outras pesquisas, bem como os estímulos visuais utilizados. A base utilizada neste projeto e os estímulos visuais estarão disponíveis no site do LAPIS<sup>1</sup> para acesso público.

Há um projeto sendo desenvolvido para organização e incremento desta base. O projeto atual adiciona na base as informações brutas de captura (*raw data*) dos 108 participantes (78

---

<sup>1</sup> <http://lapis.each.usp.br/>

com TEA e 30 com DT) do experimento, as informações de visita e fixação calculadas pelo Tobii e o estímulo visual utilizado. Também serão adicionadas informações do paciente, como: informações do CARS, sexo, idade e QI. Outros projetos incluirão futuramente fotos e medidas da face, além de incrementar com dados de mais participantes. Essa base de dados poderá ser utilizada para pesquisa em outros projetos.

Em relação ao estímulo visual utilizado, apesar de em teoria qualquer outro estímulo poder ser aplicado ao modelo desenvolvido, um estímulo inicial que esteja de acordo com os padrões visuais atípicos do TEA, facilitou na descoberta de diferenças nas características que separem os dois grupos. Um desenho animado comum, por exemplo, poderia não diferenciar tão bem a atenção visual e não apresentar bons resultados, pois não haveriam elementos que ressaltassem o padrão visual atípico do TEA. Por isso, mesmo que para a classificação o modelo aceite qualquer estímulo visual, é recomendado que se utilize vídeos com os elementos de movimento biológico versus movimento geométrico, discutidos na revisão.

#### 7.4 Limitações do Método

Todos os parâmetros definidos nas etapas de desenvolvimento foram definidos com testes empíricos para o conjunto de dados testados. Um trabalho futuro seria otimizar a forma de selecionar esses parâmetros, a fim de obter melhores resultados independentemente do conjunto de dados.

O agrupamento de *frames* e a mudança realizada na seleção de *pixels* foram feitos principalmente pelo estímulo visual ser dinâmico, ou seja, um vídeo. E, em cada *frame*, não há *pixels* fixados suficientes para uma boa análise. Existem modelos específicos para análise de atenção em vídeos, como [Zhong et al. \(2013\)](#) que utiliza análise espaço temporal. Porém, as alterações realizadas, com o agrupamento de *frames* e seleção de mais *pixels*, mostraram-se eficientes para o objetivo necessário, aumentando os resultados significativamente (de 0,62 para 0,82 de média de AUC), e técnicas mais avançadas podem ser testadas em trabalhos futuros.

Outro ponto limitante, seria a falta de uma validação pós treinamento e teste do modelo. Para isso, um novo conjunto de dados, ou seja, dados com novos participantes precisariam ser coletados e testados com o método desenvolvido para avaliar a classificação. Esta validação será feita em trabalhos futuros.

## 7.5 Avanços encontrados

A utilização do rastreamento de olhar para auxiliar o diagnóstico tem a vantagem de não ser invasivo e poder ser realizado em indivíduos de vários níveis funcionais e idades. Apesar de já existirem trabalhos que realizem a classificação do TEA baseando-se em dados de rastreamento de olhar (PIERCE et al., 2011; SHI et al., 2015; PIERCE et al., 2015; CHEVALLIER et al., 2015), o projeto atual consegue realizar essa classificação com AUC maiores do que a maioria dos projetos citados, e utilizando uma base de dados mista em relação a idade, sexo e CARS.

Além disso, a análise por meio de modelo de atenção visual evita a necessidade de demarcação de ROI's por parte de um especialista, o que pode acarretar em perda e enviesamento dos dados.

Apesar do modelo ter sido desenvolvido baseado em modelos anteriores, como os de Judd et al. (2009), Borji (2012), várias etapas do processo foram modificadas para a obtenção de melhores resultados. O agrupamento de *frames* utilizando as informações do movimento é uma contribuição que pode ser melhor explorada por outros pesquisadores e utilizada em outros métodos.

Em relação à etapa de comparação entre o mapa de fixação e o mapa de saliência, uma nova abordagem foi proposta. Essa abordagem foi proposta, pois além de ser mais simples, compara toda a imagem e não apenas *pixels* selecionados de maneira aleatória.

A proposta de classificação baseada em atenção visual é totalmente inovadora, não encontrada na literatura. Portanto, além do método apresentar bons resultados para a área da saúde com um método de classificação de TEA, ainda apresenta um novo método para a área computacional de visão computacional.

## 8 Conclusão

### 8.1 Propósitos, Vantagens e Limitações

O objetivo do projeto foi definir, implementar e validar um classificador, baseado em modelo de atenção visual, visando a prever se um indivíduo pertence ao grupo TEA ou DT. A hipótese do projeto era de que, com as técnicas utilizadas, o projeto conseguiria realizar a classificação com uma sensibilidade maior do que os trabalhos anteriores. O projeto cumpriu o propósito proposto e confirmou a hipótese, implementando um modelo de atenção visual e um classificador que indica se um indivíduo pertence ao grupo de TEA ou DT, de acordo com a atenção visual do mesmo, com uma sensibilidade média de 0,69, maior do que o trabalho de [Pierce et al. \(2011\)](#), por exemplo, que apresentou uma sensibilidade de 0,21.

Apesar do trabalho de [Shi et al. \(2015\)](#) apresentar uma AUC de 0,86, maior do que 0,82 obtida neste projeto, o projeto atual utilizou uma amostra maior e mais variada em relação a idade, QI, sexo e grau funcional dos participantes. O que aponta uma maior generalidade na aplicação do método.

O método desenvolvido possui a vantagem de não sofrer influência do profissional que o aplica, como ocorre nos métodos de análise utilizados atualmente na área, que demarcam as ROI's de acordo com o conhecimento *a priori* do profissional da saúde. Por ser um método que analisa as informações por meio de técnicas computacionais, apresentando maior objetividade na classificação.

Além disso, o método alcançou precisão de até 96%, o que confirma que a atenção visual é um biomarcador eficaz do TEA, que pode ser utilizado como forma de diagnóstico, desde que analisada com um método computacional validado.

Apesar das limitações citadas, principalmente na definição dos parâmetros, o método traz novas abordagens para a área que podem ser melhor exploradas por outros pesquisadores. Para isso disponibilizará a base de dados já citada, com os estímulos visuais e dados dos participantes e o protocolo de coleta de dados também poderá ser utilizado como base para novas pesquisas, já que a qualidade da captura foi satisfatória.

## 8.2 *Trabalhos Futuros*

Apesar de terem sido encontrados bons resultados com o desenvolvimento do modelo, foram percebidas algumas melhorias para serem realizadas em trabalhos futuros. Entre elas a realização de um teste com uma amostra maior de indivíduos, especialmente em relação ao grupo DT. Outra melhoria seria adicionar nesta amostra pessoas com outros tipos de transtorno, para que seja possível confirmar que as diferenças na atenção visual modelada são realmente típicas apenas do TEA.

Outro tópico que poderia ser explorado seria analisar atenção visual de forma temporal, analisando a trajetória do movimento ocular e não apenas os pontos de fixação. Uma hipótese é de que a atenção visual do TEA não seja diferente somente nas regiões que são mais fixadas, mas também na trajetória realizada entre essas regiões.

Como discutido nas seções anteriores, um método para definição mais assertiva dos parâmetros do modelo e técnicas mais avançadas de análise de vídeo também poderia ser estudado. Neste projeto os testes empíricos para definição dos parâmetros foram feitos de forma individual, uma combinação entre esses diferentes parâmetros e um estudo aprofundado sobre como definir cada um deles poderia melhorar ainda mais os resultados.

Além disso, estímulos visuais mais naturalísticos e com mais variações de cena possibilitariam o desenvolvimento de um modelo mais genérico que poderia ser aplicado a outros estímulos não treinados.

Para ser utilizado pela área de saúde, um trabalho futuro será a transformação do classificador em ferramenta, no qual o profissional de saúde poderá inserir os dados de rastreamento de um novo paciente e o classificador retornará se, de acordo com esses dados a atenção visual do paciente se enquadra com o atenção de TEA ou de DT, o que poderá auxiliar o médico na tomada de decisão, em caso de dúvida. Para isso, além da transformação em uma ferramenta adequada à rotina clínica, também se faz necessária a coleta de mais dados que possam ser utilizados para validação.

## 8.3 *Impactos*

O método desenvolvido processa os dados de rastreamento de olhar utilizando um modelo de atenção visual e posteriormente utiliza os dados do modelo para classificar se um indivíduo

pertence ao TEA ou possui DT. Esse classificador futuramente poderá ser transformado em uma ferramenta e utilizado para auxílio ao diagnóstico de TEA, podendo auxiliar o diagnóstico com menor custo e menor tempo, trazendo impactos na área econômica e da saúde.

Além disso, a metodologia desenvolvida poderá ser estudada e aplicada a outros transtornos que também apresentam diferenças na atenção visual, como Alzheimer e Esquizofrenia. Isso porque o projeto apresenta uma forma inovadora de analisar a atenção visual destes transtornos do que a utilizada hoje, tanto na parte da análise dos dados, utilizando um modelo computacional de atenção visual, quanto definindo um protocolo de estudo e captura de dados.

O método também traz alguns avanços e apresenta oportunidades de pesquisa para a área de computação visual, visto que apresenta diferentes abordagens em várias etapas do método desenvolvido, como: no agrupamento dos *frames*; seleção de *pixels*; método de comparação entre o mapa de fixação e o mapa de saliência e o método de classificação em si.

Além do artigo submetido com os resultados completos da Revisão Sistemática (OLIVEIRA et al., 2017) e do pôster publicado com o estudo da atenção visual de Alzheimer (PEREIRA et al., 2017), o artigo “*A method for aiding ASD diagnosis based in a computational model of visual attention*” contendo os resultados deste projeto está em desenvolvimento e será submetido em breve.

## Referências<sup>1</sup>

- ACM. ACM. 2016. Disponível em: <http://dl.acm.org/>. Acesso em: 05/06/2016. Citado na página 43.
- AGGARWAL, C. C. *Data Classification: Algorithms and Applications*. [S.l.]: Chapman and Hall/CRC, 2014. Citado na página 24.
- AL-SALEHI, S. M.; AL-HIFTHY, E. H.; GHAZIUDDIN, M. Autism in saudi arabia: presentation, clinical correlates and comorbidity. *Transcultural Psychiatry*, Sage Publications, v. 46, n. 2, p. 340–347, 2009. Citado na página 15.
- AMERICAN PSYCHIATRIC ASSOCIATION. *Manual diagnóstico e Estatístico de Transtornos Mentais-: DSM-5*. [S.l.]: Artmed Editora, 2014. Citado 4 vezes nas páginas 15, 16, 20 e 23.
- AMESTOY, A. et al. Developmental changes in face visual scanning in autism spectrum disorder as assessed by data-based analysis. *Frontiers in psychology*, Frontiers Media SA, v. 6, 2015. Citado na página 50.
- ANDERSON, C. J.; COLOMBO, J.; SHADDY, D. J. Visual scanning and pupillary responses in young children with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology*, Taylor & Francis, v. 28, n. 7, p. 1238–1256, 2006. Citado na página 47.
- BAIO, J. Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states, 2008. morbidity and mortality weekly report. surveillance summaries. volume 61, number 3. *Centers for Disease Control and Prevention*, ERIC, 2012. Citado na página 15.
- BEDFORD, R. et al. Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. *Journal of autism and developmental disorders*, Springer, v. 42, n. 10, p. 2208–2218, 2012. Citado na página 21.
- BELLMAN, R. *Adaptive control processes: a guided tour* princeton university press. *Princeton, New Jersey, USA*, 1961. Citado na página 27.
- BERNIER, R.; MAO, A.; YEN, J. Psychopathology, families, and culture: autism. *Child and adolescent psychiatric clinics of North America*, Elsevier, v. 19, n. 4, p. 855–867, 2010. Citado na página 15.
- BORJI, A. Boosting bottom-up and top-down visual features for saliency estimation. In: *IEEE. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 438–445. Citado 10 vezes nas páginas 8, 24, 25, 38, 41, 42, 63, 69, 81 e 88.
- BORJI, A.; ITTI, L. State-of-the-art in visual attention modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 35, n. 1, p. 185–207, 2013. Citado 4 vezes nas páginas 34, 35, 36 e 39.
- BOSL, W. et al. Eeg complexity as a biomarker for autism spectrum disorder risk. *BMC medicine*, BioMed Central, v. 9, n. 1, p. 1, 2011. Citado na página 22.
- BRENTANI, H. et al. Autism spectrum disorders: an overview on diagnosis and treatment. *Revista brasileira de psiquiatria*, SciELO Brasil, v. 35, p. S62–S72, 2013. Citado na página 22.

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

CALDARA, R.; MIELLET, S. imap: a novel method for statistical fixation mapping of eye movement data. *Behavior research methods*, Springer, v. 43, n. 3, p. 864–878, 2011. Citado na página 50.

CAMARATA, S. Early identification and early intervention in autism spectrum disorders: Accurate and effective? *International Journal of Speech-Language Pathology*, Taylor & Francis, v. 16, n. 1, p. 1–10, 2014. Citado 2 vezes nas páginas 21 e 22.

CHAWARSKA, K.; KLIN, A.; VOLKMAR, F. Automatic attention cueing through eye movement in 2-year-old children with autism. *Child development*, Wiley Online Library, v. 74, n. 4, p. 1108–1122, 2003. Citado na página 21.

CHEVALLIER, C. et al. Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research*, Wiley Online Library, v. 8, n. 5, p. 620–628, 2015. Citado 6 vezes nas páginas 16, 17, 49, 52, 56 e 88.

COHEN, H.; AMERINE-DICKENS, M.; SMITH, T. Early intensive behavioral treatment: Replication of the ucla model in a community setting. *Journal of Developmental & Behavioral Pediatrics*, LWW, v. 27, n. 2, p. S145–S155, 2006. Citado na página 21.

CONWAY, D.; WHITE, J. *Machine learning for hackers*. [S.l.]: "O'Reilly Media, Inc.", 2012. Citado 2 vezes nas páginas 23 e 24.

DAWSON, G. et al. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*, Am Acad Pediatrics, v. 125, n. 1, p. e17–e23, 2010. Citado 2 vezes nas páginas 15 e 21.

DIRECT, S. *Science Direct*. 2016. Disponível em: <http://www.sciencedirect.com/>. Acesso em: 05/06/2016. Citado na página 43.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012. Citado na página 27.

DUNDAS, E.; GASTGEB, H.; STRAUSS, M. S. Left visual field biases when infants process faces: A comparison of infants at high-and low-risk for autism spectrum disorder. *Journal of autism and developmental disorders*, Springer, v. 42, n. 12, p. 2659–2668, 2012. Citado na página 49.

ELSABBAGH, M. et al. What you see is what you get: contextual modulation of face scanning in typical and atypical development. *Social cognitive and affective neuroscience*, Oxford University Press, v. 9, n. 4, p. 538–543, 2014. Citado na página 51.

ELSABBAGH, M.; JOHNSON, M. H. Getting answers from babies about autism. *Trends in cognitive sciences*, Elsevier, v. 14, n. 2, p. 81–87, 2010. Citado na página 22.

ELSABBAGH, M. et al. Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current biology*, Elsevier, v. 22, n. 4, p. 338–342, 2012. Citado na página 22.

FALCK-YTTER, T.; BÖLTE, S.; GREDEBÄCK, G. Eye tracking in early autism research. *Journal of neurodevelopmental disorders*, BioMed Central Ltd, v. 5, n. 1, p. 28, 2013. Citado na página 21.

FALCK-YTTER, T.; HOFSTEN, C. von. How special is social looking in asd: a review. *Progress in brain research*, n. 189, p. 209–22, 2011. Citado na página 23.

FALCK-YTTER, T. et al. Visualization and analysis of eye movement data from children with typical and atypical development. *Journal of autism and developmental disorders*, Springer, v. 43, n. 10, p. 2249–2258, 2013. Citado na página 50.

FALCK-YTTER, T.; THORUP, E.; BÖLTE, S. Brief report: Lack of processing bias for the objects other people attend to in 3-year-olds with autism. *Journal of autism and developmental disorders*, Springer, v. 45, n. 6, p. 1897–1904, 2014. Citado na página 51.

FAN, R.-E. et al. Liblinear: A library for large linear classification. *Journal of machine learning research*, v. 9, n. Aug, p. 1871–1874, 2008. Citado na página 70.

FARNEBÄCK, G. Two-frame motion estimation based on polynomial expansion. *Image analysis*, Springer, p. 363–370, 2003. Citado na página 65.

FAUSETT, L. (Ed.). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1994. ISBN 0-13-334186-0. Citado na página 25.

FELZENSZWALB, P.; MCALLESTER, D.; RAMANAN, D. A discriminatively trained, multiscale, deformable part model. In: IEEE. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. [S.l.], 2008. p. 1–8. Citado 2 vezes nas páginas 32 e 40.

FELZENSZWALB, P. F. et al. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 32, n. 9, p. 1627–1645, 2010. Citado na página 41.

FOMBONNE, E. Epidemiology of pervasive developmental disorders. *Pediatric research*, Nature Publishing Group, v. 65, n. 6, p. 591–598, 2009. Citado na página 20.

FREETH, M. et al. Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with asd? evidence from eye-tracking. *Journal of autism and developmental disorders*, Springer, v. 40, n. 5, p. 534–547, 2010. Citado na página 56.

FREETH, M.; FOULSHAM, T.; CHAPMAN, P. The influence of visual saliency on fixation patterns in individuals with autism spectrum disorders. *Neuropsychologia*, Elsevier, v. 49, n. 1, p. 156–160, 2011. Citado 3 vezes nas páginas 17, 47 e 51.

FRINTROP, S. *VOCUS: A visual attention system for object detection and goal-directed search*. [S.l.]: Springer, 2006. v. 3899. Citado na página 36.

FRITH, U.; FRITH, C. D. Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, The Royal Society, v. 358, n. 1431, p. 459–473, 2003. Citado na página 20.

GAO, D.; VASCONCELOS, N. Discriminant saliency for visual recognition from cluttered scenes. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 481–488. Citado na página 37.

GARCIA-DIAZ, A. et al. Decorrelation and distinctiveness provide with human-like saliency. In: SPRINGER. *Advanced concepts for intelligent vision systems*. [S.l.], 2009. p. 343–354. Citado na página 41.

- GOLDBERG, D. E.; HOLLAND, J. H. Genetic algorithms and machine learning. *Machine Learning*, v. 3, n. 2, p. 95–99, Oct 1988. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1022602019183>. Citado na página 28.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X. Citado 2 vezes nas páginas 29 e 30.
- GUILLON, Q. et al. The importance of networking in autism gaze analysis. *PloS one*, Public Library of Science, v. 10, n. 10, p. e0141191, 2015. Citado na página 47.
- GUIMARD-BRUNAULT, M. et al. Back to basic: do children with autism spontaneously look at screen displaying a face or an object? *Autism research and treatment*, Hindawi Publishing Corporation, v. 2013, 2013. Citado na página 44.
- GUTHRIE, W. et al. Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 54, n. 5, p. 582–590, 2013. Citado 2 vezes nas páginas 21 e 22.
- HAREL, J.; KOCH, C.; PERONA, P. Graph-based visual saliency. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2006. p. 545–552. Citado na página 41.
- HAYHOE, M.; BALLARD, D. Eye movements in natural behavior. *Trends in cognitive sciences*, Elsevier, v. 9, n. 4, p. 188–194, 2005. Citado na página 36.
- HAYKIN, S.; NETWORK, N. A comprehensive foundation. *Neural Networks*, v. 2, n. 2004, p. 41, 2004. Citado na página 24.
- HERNANDEZ, N. et al. Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, Elsevier, v. 47, n. 4, p. 1004–1012, 2009. Citado na página 44.
- IEEE. *IEEE*. 2016. Disponível em: <http://ieeexplore.ieee.org/>. Acesso em: 05/06/2016. Citado na página 43.
- ITTI, L. *Models of bottom-up and top-down visual attention*. Tese (Doutorado) — California Institute of Technology, 2000. Citado 4 vezes nas páginas 36, 64, 79 e 80.
- ITTI, L.; KOCH, C. Computational modelling of visual attention. *Nature reviews neuroscience*, Nature Publishing Group, v. 2, n. 3, p. 194–203, 2001. Citado na página 34.
- ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 20, n. 11, p. 1254–1259, 1998. ISSN 0162-8828. Citado 8 vezes nas páginas 8, 17, 35, 36, 37, 40, 50 e 51.
- JESUS, E. O.; Costa Jr., R. A utilização de filtros gaussianos na análise de imagens digitais. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 3, n. 1, 2015. Citado na página 31.
- JOHNELS, J. Å. et al. Face-viewing patterns in young children with autism spectrum disorders: Speaking up for the role of language comprehension. *Journal of Speech, Language, and Hearing Research*, ASHA, v. 57, n. 6, p. 2246–2252, 2014. Citado na página 51.

JUDD, T. et al. Learning to predict where humans look. In: IEEE. *Computer Vision, 2009 IEEE 12th international conference on*. [S.l.], 2009. p. 2106–2113. Citado 9 vezes nas páginas 24, 25, 38, 40, 41, 63, 64, 69 e 88.

KIENZLE, W. et al. A nonparametric approach to bottom-up visual saliency. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2006. p. 689–696. Citado na página 38.

KIRA, K.; RENDELL, L. A. A practical approach to feature selection. In: *Proceedings of the ninth international workshop on Machine learning*. [S.l.: s.n.], 1992. p. 249–256. Citado na página 28.

KLIEMANN, D. et al. Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *The Journal of Neuroscience*, Soc Neuroscience, v. 30, n. 37, p. 12281–12287, 2010. Citado na página 23.

KLIN, A. et al. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, American Medical Association, v. 59, n. 9, p. 809–816, 2002. Citado na página 21.

KLIN, A.; KLAIMAN, C.; JONES, W. Reducing age of autism diagnosis: developmental social neuroscience meets public health challenge. *Revista de neurologia*, v. 60, n. S01, p. S3–S11, 2015. Citado 2 vezes nas páginas 21 e 22.

KLIN, A. et al. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, Nature Publishing Group, v. 459, n. 7244, p. 257–261, 2009. Citado na página 23.

KOCH, C.; ULLMAN, S. Shifts in selective visual attention: Towards the underlying neural circuitry. In: VAINA, L. (Ed.). *Matters of Intelligence*. Springer Netherlands, 1987, (Synthese Library, v. 188). p. 115–141. ISBN 978-94-010-8206-8. Disponível em: [http://dx.doi.org/10.1007/978-94-009-3833-5\\_5](http://dx.doi.org/10.1007/978-94-009-3833-5_5). Citado 3 vezes nas páginas 8, 35 e 36.

KOOTSTRA, G.; NEDERVEEN, A.; BOER, B. D. Paying attention to symmetry. In: THE BRITISH MACHINE VISION ASSOCIATION AND SOCIETY FOR PATTERN RECOGNITION. *British Machine Vision Conference (BMVC2008)*. [S.l.], 2008. p. 1115–1125. Citado na página 38.

KRUTHIVENTI, S. S.; AYUSH, K.; BABU, R. V. Deepfix: A fully convolutional neural network for predicting human eye fixations. *CoRR*, abs/1510.02927, 2015. Disponível em: <http://arxiv.org/abs/1510.02927>. Citado na página 24.

KÜMMERER, M.; WALLIS, T. S. A.; BETHGE, M. Deepgaze II: reading fixations from deep features trained on object recognition. *CoRR*, abs/1610.01563, 2016. Disponível em: <http://arxiv.org/abs/1610.01563>. Citado na página 24.

LANDA, R. J.; HOLMAN, K. C.; GARRETT-MAYER, E. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of general psychiatry*, American Medical Association, v. 64, n. 7, p. 853–864, 2007. Citado na página 21.

LORD, C.; MCGEE, J. *Committee on educational intervention for children with autism. Educating children with autism*. [S.l.]: Washington, DC: Nat Acad Press, 2001. Citado na página 21.

LORD, C. et al. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, Springer, v. 30, n. 3, p. 205–223, 2000. Citado 2 vezes nas páginas 15 e 20.

LORD, C.; RUTTER, M.; COUTEUR, A. L. 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*, Springer, v. 24, n. 5, p. 659–685, 1994. Citado 2 vezes nas páginas 15 e 20.

LORD, C.; SHULMAN, C.; DILAVORE, P. Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 45, n. 5, p. 936–955, 2004. Citado na página 21.

LOVAAS, O. I. Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of consulting and clinical psychology*, American Psychological Association, v. 55, n. 1, p. 3, 1987. Citado na página 21.

LUDWIG, O.; NUNES, U. Novel maximum-margin training algorithms for supervised neural networks. *IEEE Transactions on Neural Networks*, IEEE, v. 21, n. 6, p. 972–984, 2010. Citado 2 vezes nas páginas 28 e 68.

MAGRELLI, S. et al. Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings. *Frontiers in psychology*, Frontiers Media SA, v. 4, 2013. Citado na página 43.

MANDELL, D. S.; NOVAK, M. M.; ZUBRITSKY, C. D. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics*, Am Acad Pediatrics, v. 116, n. 6, p. 1480–1486, 2005. Citado na página 21.

MARQUES, D. F.; BOSA, C. A. Protocolo de avaliação de crianças com autismo: Evidências de validade de critério. *Psicologia: Teoria e Pesquisa*, SciELO Brasil, v. 31, n. 1, p. 43–51, 2015. Citado na página 15.

Marques Filho, O.; NETO, H. V. *Processamento digital de imagens*. [S.l.]: Brasport, 1999. Citado na página 30.

MATLAB. *R2015a*. Natick, Massachusetts: The MathWorks Inc., 2015. Citado na página 61.

MCPARTLAND, J. C. et al. Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of autism and developmental disorders*, Springer, v. 41, n. 2, p. 148–157, 2011. Citado na página 49.

MÜLLER, K.-R. et al. An introduction to kernel-based learning algorithms. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 12, n. 2, p. 181–201, 2001. Citado na página 26.

NAKANO, T. et al. Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, The Royal Society, v. 277, n. 1696, p. 2935–2943, 2010. Citado 2 vezes nas páginas 50 e 52.

NAVALPAKKAM, V.; ITTI, L. An integrated model of top-down and bottom-up attention for optimizing detection speed. In: IEEE. *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. [S.l.], 2006. v. 2, p. 2049–2056. Citado na página [36](#).

NEUMANN, D. et al. Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social cognitive and affective neuroscience*, Oxford University Press, v. 1, n. 3, p. 194–202, 2006. Citado 3 vezes nas páginas [17](#), [48](#) e [50](#).

NIXON, M. S.; AGUADO, A. S. *Feature extraction & image processing for computer vision*. [S.l.]: Academic Press, 2012. Citado na página [29](#).

NORIS, B. et al. Investigating gaze of children with asd in naturalistic settings. *PLoS One*, Public Library of Science, v. 7, n. 9, p. e44144, 2012. Citado 2 vezes nas páginas [43](#) e [47](#).

NOTHDURFT, H. "Salience of Feature Contrast", *Neurobiology of Attention*. Academic Press, 2005. Citado na página [34](#).

OLIVA, A.; TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, Springer, v. 42, n. 3, p. 145–175, 2001. Citado 2 vezes nas páginas [36](#) e [40](#).

OLIVEIRA, J. d. S. de et al. Eye tracking for autism detection: methods, devices, visual paradigms, and experiments (submetido). *Neuroscience & Biobehavioral Review*, 2017. Citado 2 vezes nas páginas [58](#) e [91](#).

OREKHOVA, E. V. et al. Eeg hyper-connectivity in high-risk infants is associated with later autism. *Journal of neurodevelopmental disorders*, BioMed Central, v. 6, n. 1, p. 1, 2014. Citado na página [22](#).

PALOMO, R.; BELINCHÓN, M.; OZONOFF, S. Autism and family home movies: a comprehensive review. *Journal of Developmental & Behavioral Pediatrics*, LWW, v. 27, n. 2, p. S59–S68, 2006. Citado na página [21](#).

PAULA, C. S. et al. Brief report: prevalence of pervasive developmental disorder in brazil: a pilot study. *Journal of Autism and Developmental Disorders*, Springer, v. 41, n. 12, p. 1738–1742, 2011. Citado na página [15](#).

PEREIRA, M. et al. Eye movement behavior in mci and ad: Using automatic classification algorithms to identify cognitive decline. In: *AAIC 2017*. London, UK: [s.n.], 2017. Citado 2 vezes nas páginas [58](#) e [91](#).

PETERS, R. J.; ITTI, L. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In: IEEE. *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. [S.l.], 2007. p. 1–8. Citado 5 vezes nas páginas [8](#), [36](#), [38](#), [39](#) e [40](#).

PIERCE, K. et al. Preference for geometric patterns early in life as a risk factor for autism. *Archives of general psychiatry*, American Medical Association, v. 68, n. 1, p. 101–109, 2011. Citado 8 vezes nas páginas [16](#), [23](#), [45](#), [49](#), [52](#), [56](#), [88](#) e [89](#).

PIERCE, K. et al. 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*, Elsevier, 2015. Citado 6 vezes nas páginas [16](#), [45](#), [49](#), [52](#), [56](#) e [88](#).

PUBMED. *PubMed*. 2016. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed>. Acesso em: 05/06/2016. Citado na página 43.

RIBAS, L. d. M. *Um estudo sobre o brincar de uma criança autista atípica: intervenções psicopedagógicas*. 24 p. — Universidade de Brasília, Brasília, 2013. Citado na página 15.

RICE, M. E.; HARRIS, G. T. Comparing effect sizes in follow-up studies: Roc area, cohen's d, and r. *Law and human behavior*, Springer, v. 29, n. 5, p. 615, 2005. Citado na página 16.

RIGO, G. V. *Modelando a atenção seletiva e a saliência visual através de redes complexas*. Tese (Doutorado) — Universidade de São Paulo, 2010. Citado na página 34.

ROBNIK-SIKONJA, M.; KONONENKO, I. *An adaptation of Relief for attribute estimation in regression*. 1997. Citado na página 68.

ROSENHOLTZ, R. A simple saliency model predicts a number of motion popout phenomena. *Vision research*, Elsevier, v. 39, n. 19, p. 3157–3163, 1999. Citado na página 40.

RUTTER, M. L. Progress in understanding autism: 2007–2010. *Journal of autism and developmental disorders*, Springer, v. 41, n. 4, p. 395–404, 2011. Citado na página 20.

SASSON, N. J. et al. Brief report: Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders*, Springer, v. 41, n. 2, p. 242–247, 2011. Citado na página 45.

SCHWARTZMAN, J. S. et al. The eye-tracking of social stimuli in patients with rett syndrome and autism spectrum disorders: a pilot study. *Arquivos de neuro-psiquiatria*, SciELO Brasil, v. 73, n. 5, p. 402–407, 2015. Citado 2 vezes nas páginas 45 e 46.

SEO, H. J.; MILANFAR, P. Static and space-time visual saliency detection by self-resemblance. *Journal of vision*, Association for Research in Vision and Ophthalmology, v. 9, n. 12, p. 15, 2009. Citado na página 38.

SHI, L. et al. Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PloS one*, Public Library of Science, v. 10, n. 3, p. e0122280, 2015. Citado 10 vezes nas páginas 16, 17, 45, 47, 49, 50, 52, 56, 88 e 89.

SHIC, F. et al. Limited activity monitoring in toddlers with autism spectrum disorder. *Brain research*, Elsevier, v. 1380, p. 246–254, 2011. Citado na página 47.

SHIC, F. et al. Measuring context: The gaze patterns of children with autism evaluated from the bottom-up. In: IEEE. *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*. [S.l.], 2007. p. 70–75. Citado 4 vezes nas páginas 17, 49, 51 e 56.

SIMONCELLI, E. P.; FREEMAN, W. T. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In: IEEE. *Image Processing, 1995. Proceedings., International Conference on*. [S.l.], 1995. v. 3, p. 444–447. Citado na página 64.

SPEER, L. L. et al. Face processing in children with autism effects of stimulus contents and type. *Autism*, Sage Publications, v. 11, n. 3, p. 265–277, 2007. Citado na página 47.

SPRINGER. *Springer*. 2016. Disponível em: <http://link.springer.com/>. Acesso em: 05/06/2016. Citado na página 43.

- STUDIO, T. 1.2 user manual. *Tobii Technology AB*, 2008. Citado na página 59.
- SUN, J.; BUYS, N. Early executive function deficit in preterm children and its association with neurodevelopmental disorders in childhood: a literature review. 2012. Citado na página 20.
- TAVAKOLI, H. R.; RAHTU, E.; HEIKKILÄ, J. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: *Image Analysis*. [S.l.]: Springer, 2011. p. 666–675. Citado na página 41.
- TEKALP, A. M. *Digital video processing*. [S.l.]: Prentice Hall Press, 2015. Citado na página 29.
- Tobii Technology. *Tobii*. 2016. Disponível em: [www.tobii.com](http://www.tobii.com). Citado na página 58.
- TORRALBA, A. Modeling global scene factors in attention. *JOSA A*, Optical Society of America, v. 20, n. 7, p. 1407–1418, 2003. Citado 2 vezes nas páginas 41 e 80.
- TREISMAN, A. M.; GELADE, G. A feature-integration theory of attention. *Cognitive psychology*, Elsevier, v. 12, n. 1, p. 97–136, 1980. Citado na página 34.
- TSOTSOS, J. K.; ROTHENSTEIN, A. Computational models of visual attention. *Scholarpedia*, v. 6, n. 1, p. 6201, 2011. Citado 2 vezes nas páginas 17 e 35.
- VAPNIK, V. N.; VAPNIK, V. *Statistical learning theory*. [S.l.]: Wiley New York, 1998. v. 1. Citado na página 25.
- VIOLA, P.; JONES, M. Robust real-time object detection. *International Journal of Computer Vision*, v. 4, 2001. Citado 6 vezes nas páginas 8, 32, 33, 40, 64 e 65.
- WALKER, L. L.; MALIK, J. When is scene recognition just texture recognition. *Journal of Vision*, Citeseer, v. 2, n. 7, p. 255, 2002. Citado na página 36.
- WALL, D. P. et al. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PloS one*, Public Library of Science, v. 7, n. 8, p. e43855, 2012. Citado na página 15.
- WALTHER, D.; KOCH, C. Modeling attention to salient proto-objects. *Neural networks*, Elsevier, v. 19, n. 9, p. 1395–1407, 2006. Citado na página 64.
- WANG, L.-H. et al. Children with autism spectrum disorders are less proficient in action identification and lacking a preference for upright point-light biological motion displays. *Research in Autism Spectrum Disorders*, Elsevier, v. 11, p. 63–76, 2015. Citado 2 vezes nas páginas 44 e 47.
- WANG, S. et al. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, Elsevier, v. 88, n. 3, p. 604–616, 2015. Citado 5 vezes nas páginas 17, 51, 53, 56 e 79.
- WHITLEY, D. A genetic algorithm tutorial. *Statistics and Computing*, v. 4, n. 2, p. 65–85, Jun 1994. ISSN 1573-1375. Disponível em: <https://doi.org/10.1007/BF00175354>. Citado na página 28.
- XU, J. et al. Predicting human gaze beyond pixels. *Journal of vision*, The Association for Research in Vision and Ophthalmology, v. 14, n. 1, p. 28–28, 2014. Citado 3 vezes nas páginas 24, 25 e 41.

YAMAGUCHI, K. Mexopencv. *Collection and a development kit of matlab mex functions for OpenCV library, available at <http://www.cs.stonybrook.edu/~kyamagu/mexopencv>*, 2013. Citado na página 62.

YI, L. et al. Abnormality in face scanning by children with autism spectrum disorder is limited to the eye region: Evidence from multi-method analyses of eye tracking data. *Journal of vision*, Association for Research in Vision and Ophthalmology, v. 13, n. 10, p. 5, 2013. Citado 2 vezes nas páginas 49 e 50.

YOU DEN, W. J. Index for rating diagnostic tests. *Cancer*, Wiley Online Library, v. 3, n. 1, p. 32–35, 1950. Citado na página 73.

ZHONG, S.-h. et al. Video saliency detection via dynamic consistent spatio-temporal attention modelling. In: *AAAI*. [S.l.: s.n.], 2013. p. 1063–1069. Citado na página 87.

# **Apêndices**

**Apêndice A – TCLE - Termo de Consentimento Livre e Esclarecido**

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

### DADOS DA PESQUISA

TÍTULO DA PESQUISA: PROCESSAMENTO DO RASTREAMENTO DE OLHAR COM BASE EM MODELO DE ATENÇÃO VISUAL PARA AUXÍLIO AO DIAGNÓSTICO DE AUTISMO

\_\_\_\_\_, neste ato apresentado por mim \_\_\_\_\_, está sendo convidado a participar de um estudo que pretende rastrear o olhar de crianças com Transtorno do Espectro de Autismo (TEA) e crianças com Desenvolvimento Típico (DT) e, utilizar esses dados para desenvolver um protótipo de sistema que auxilie no diagnóstico do TEA.

Neste experimento a criança sentará em frente a um equipamento de rastreamento de olhar e assistirá alguns vídeos curtos. A estimativa é de que toda o teste não leve mais de 30min.

Toda pesquisa pode oferecer riscos, porém esta pesquisa foi estruturada visando a minimização dos possíveis riscos, desconforto ou prejuízo para os participantes. A participação nesta investigação apresenta riscos mínimos, podendo o participante sentir alguns desconfortos, como cansaço visual, mal-estar, náusea e zozzeira.

Este estudo poderá trazer como benefício melhores resultados para auxílio ao diagnóstico do TEA, que beneficiará novos pacientes agilizando e facilitando o processo de diagnóstico.

O responsável pelo participante poderá acompanhá-lo durante toda a pesquisa. É assegurada a assistência do participante durante toda a pesquisa, bem como é garantido ao responsável o livre acesso a todas as informações e esclarecimentos adicionais sobre o estudo e suas consequências.

A privacidade do participante será respeitada, ou seja, seu nome ou qualquer outro dado ou elemento que possa, de qualquer forma, o (a) identificar, será mantido em sigilo. O mesmo tem todo o direito de recusar-se a participar ou retirar o seu consentimento em qualquer fase da pesquisa sem penalização alguma, de sigilo e privacidade.

Não tem despesas para os participantes da pesquisa. Também não há compensação financeira relacionada a sua participação.

Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis pela pesquisa para esclarecimento de dúvidas. O pesquisador responsável é Jéssica dos Santos de Oliveira, poderá ser encontrado pelo e-mail: [jessica.santos.oliveira@usp.br](mailto:jessica.santos.oliveira@usp.br). Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) da Escola de Artes, Ciências e Humanidades (EACH) – Rua Arlindo Béttio, 1000 - Ermelino Matarazzo, São Paulo - SP, 03828-000 - T14 - I1 – tel:(11) 3091-1046; e-mail: [cep-each@usp.br](mailto:cep-each@usp.br) – ou com o CEP do Hospital Universitário (HU) - Rua Ovídio Pires de Campos, 225 – 5º andar – tel: (11) 2661-6442 ramais 16, 17, 18, ou (11) 2661-7585; e-mail: [cappesq.adm@hc.fm.usp.br](mailto:cappesq.adm@hc.fm.usp.br) Fui suficientemente informado a respeito do estudo “Processamento do rastreamento de olhar com base em modelo de atenção visual para auxílio ao diagnóstico de autismo”. Eu discuti as informações acima com o Pesquisador Responsável (Jéssica dos Santos de Oliveira) ou pessoa (s) por ele delegada (s) (.....) sobre a minha decisão em participar nesse estudo. Ficaram claros para mim os objetivos, os procedimentos, os potenciais desconfortos e riscos e as garantias. Concordo voluntariamente em participar deste estudo, assino este termo de consentimento e recebo uma via rubricada pelo pesquisador.

Data \_\_/\_\_/201\_\_

---

Assinatura participante/representante legal

---

Assinatura do Pesquisador

---

**DADOS DE IDENTIFICAÇÃO (OU ETIQUETA INSTITUCIONAL DE IDENTIFICAÇÃO) DO PARTICIPANTE DA PESQUISA OU RESPONSÁVEL LEGAL**

1. NOME: .....

DOCUMENTO DE IDENTIDADE Nº : ..... SEXO : .M  F

DATA NASCIMENTO: ...../...../.....

ENDEREÇO ..... Nº ..... APTO: .....

BAIRRO: ..... CIDADE .....

CEP:..... TELEFONE: DDD (.....) .....

2. RESPONSÁVEL LEGAL .....

NATUREZA (grau de parentesco, tutor, curador etc.) .....

DOCUMENTO DE IDENTIDADE :.....SEXO: M  F

DATA NASCIMENTO.: ...../...../.....

ENDEREÇO: ..... Nº ..... APTO: .....

BAIRRO: ..... CIDADE: .....

CEP: ..... TELEFONE: DDD (.....).....

---

**Apêndice B – Artigo de Revisão Sistemática**

# Eye Tracking for Autism detection: methods, devices, visual paradigms, and experiments

---

## Abstract

The use of eye tracking to understand visual attention and to aid the diagnosis of autism has grown in the last two decades, and several research groups have investigated this subject. However, just a few works have been published to gather the results obtained. This article presents the results of a systematic review conducted to assess the main aspects of this type of study, including the eye tracking devices used, the formation of samples, and mainly the methods and techniques used to select and analyze the data. We selected 58 papers to be fully analysed. The results point to promising advances, such as the use of methods such as iMap and visual attention models. Further, we found some studies defining specific criteria to differentiate the groups. However, it is still necessary to fill some gaps, such as the use of more robust computer methods to analyze the signals, as well as the formation of suitable samples with individuals that represent the target under investigation in a significant way.

*Keywords:* Eye Tracking, Autism Spectrum Disorder, Visual Attention, Data Analysis Methods, Experiments

---

## 1. Introduction

The Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that affects, in different levels, two main areas: social communication and behaviour [1, 2]. The clinical criteria for diagnosis encompass the social communication deficit that, based on the development level, analyses the lack of social reciprocity. It also involves restricted and repetitive behaviour and interests, expressed by excessive adherence to routine, ritualized behaviour, intense and inflexible interests, and unusual sensory behaviours.

The clinical evaluation is made based on the anamnesis with parents/caregivers and the observation of the child. The “gold standard diagnosis”, instru-

ments accepted for the diagnosis of ASD, are the Autism Diagnostic Interview - reviewed (ADI-R) [3] and the Autism Diagnostic Observation Schedule (ADOS) [4]. The ADI-R is conducted with the parents while the ADOS is a semi-structured interview that comprises the observation of the child.

Some studies have proposed and showed the effectiveness of the earlier intervention in the ASD [5, 6]. The National Research Council [7] states that the effective intervention in ASD children depends on the intensive treatment and the early start of the intervention program, showing the need for earlier predictive signs of ASD. In practice, the average age to identify the disorder is around 3 years old, in the beginning of the scholar age, or even later [8]. Klin, Klaiman and Jones [9] state that the diagnosis age in the community as a whole is still too late in relation to the window of opportunity provided by the neuroplasticity in the first years of life.

In a wider sense, the difficulty in making an early ASD diagnosis is due to the complex and multifactorial nature of the disorder, which worsens with the complexity of the diagnosis [8, 10]. Moreover, the use of “gold-standard of diagnosis” tools is too limited due to the high financial costs and the need for specific training and specialization for evaluators [9]. Efforts have been made to find safe biological markers to help the detection of ASD [11].

In this context, a growing number of research studies have been using medical exams and technological mechanisms to aid the early detection of ASD. The use of eye tracking deserves attention because it can show significant differences in children with ASD when compared with children that do not have the disorder. In an objective way, eye tracking allows the evaluation of a wide variety of visual aspects, bringing reliable information about individuals visual standard. It was discovered, for instance, that children with ASD pay less attention to biological motion [12] and faces [13]. Not being invasive is also an advantage of this technology, which can be applied to almost any individual, at different ages or functionality levels.

Considering the given context, this paper aims to contribute with the systematization of works that help detecting autism using eye tracking. We focus on the analysis of the methods and techniques used in data selection and analysis. Our hypothesis is that with growth in the number of papers analysing the eye tracking data and with the increase of data, more computational methods will be used to help and improve the results. We also summarise and made a brief analysis of the eye-tracking devices and mechanisms used in experiments, and the formation of the group of participants.

During the research we found four different reviews that analyse the use of

eye tracking in autistic children [14, 15, 16, 17]. Boraston et al. [14] reviewed the use of eye tracking to investigate the strategies used by individuals with high-functioning autism, focusing on the tasks related to the processing of social information. Falck-Ytter et al. [15] revise the use of eye tracking in researches that focus on autism in children in their early ages. In Guillon et al. [16] the focus was papers that studied, with eye tracking technology, the differences in the social attention of individuals with ASD, from the infancy until the adulthood. The last review ([17]) examines papers that analyse autistic childrens eye movement when exploring faces. None of these reviews aim to verify the methods used to select and analyse data extracted from eye tracking, which is the purpose of the presented paper.

These reviews delineate the works according to the participants and the visual stimuli used in the experimental studies, while the current review is broader making no restrictions related to the stimuli type or participants.

## 2. Methodology

The research questions defined for this review were:

- Which methods and techniques were used in the eye tracking experiments with ASD patients, mainly during the data analysis phase provided by the eye tracking?
- Which eye tracking mechanisms were used?
- How were the groups of participants formed in the experiment?

The keywords “eye tracking” and “autism” were used to search the articles, followed by the search sources: ACM<sup>1</sup>, IEEE<sup>2</sup>, *Science Direct*<sup>3</sup>, PubMed<sup>4</sup>, and *Springer*<sup>5</sup>.

The inclusion criteria defined by the protocol were: (a) works that define any aspect of the autism-related visual attention using eye tracking, and (b) works that present a review about eye tracking and ASD. The exclusion

---

<sup>1</sup><http://dl.acm.org/>

<sup>2</sup><http://ieeexplore.ieee.org/>

<sup>3</sup><http://www.sciencedirect.com/>

<sup>4</sup><http://www.ncbi.nlm.nih.gov/pubmed>

<sup>5</sup><http://link.springer.com/>

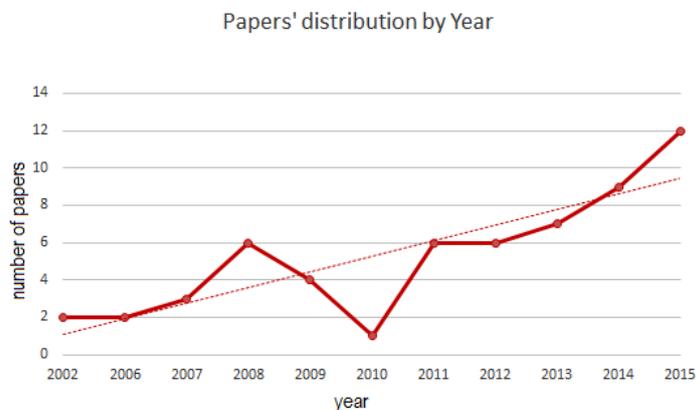


Figure 1: Distribution of included papers by year.

criteria defined by the protocol were: (a) articles published as abstracts or short papers; (b) works that do not focus on ASD characterization; (c) works that do not specifically define the technique or method used to analyse data; (d) works that do not use the obtained data from eye tracking; (e) works that do not approach the autistic children’s visual attention, and (f) works in which the individual performs another activity during the experiment despite of observing the scenes. The final analysis included 62 studies. A form was filled out with the data defined by the protocol for each article, allowing the categorization and analysis of each dimension researched, as made available in the section 3.

### 3. Results and Discussion

The number of studies about eye tracking in ASD has been growing, especially in the last five years, as shown in Figure 1, indicating the interest in this area. Probably one of the reasons for this trend is the availability of technology, which makes eye tracking more feasible when talking about the capture of the signs.

#### 3.1. Eye Tracking Devices

Choosing the eye tracking device is an important stage in the definition of the experiment, because it can enable or hamper data collection; further, it can define the limitations of recruiting patients. The first classification of

such mechanisms divide them into local (the ones that need to be adjusted to the patient) and remote (the ones that can capture signs while being distant from the patient, that do not need to be adjusted to him/her). If we look at the works that use natural interaction [18, 19], when the patient interacts directly with an experimenter instead of being asked to see a video, some head-mounted devices are used, worn as a helmet. This kind of device, however, is used in only 16% of the works analysed, probably because the individual should not move while the equipment is been adjusted and also because it cannot be removed during the experiment - two different conditions that may be a problem depending on the level of the autism. The remote devices were used in 81% of the works analysed.

Besides these two types of devices, the work developed by Hanley et al. [20] is the only one that uses chin rest, a kind of support with which the individual needs to stay with his/her chin held, avoiding head movement.

Some works use their own eye tracking systems, set with two cameras and one computer [21, 22]. Another work uses a web-cam to capture the participants eyes [23]. These works do not mention the accuracy or missing-data found in the experiment. Thus, it is not possible to compare them with other devices.

Authors of 35% of the included papers reported the exclusion of some of the participants due to the difficulty to adjust the equipment to these participants or due to a lack of sufficient data. At large, these criteria are established in 50% of the experiments overall time, as pointed out by Pierce et al. [24, 12] and Shi et al. [25]. Additionally, Sasson et al. [26] exclude information of individuals whose data was impossible to capture in at least 15% of the eye tracking data in one of the tests.

### *3.2. Participants in the experiments*

The recruitment of volunteers for eye tracking experiments in the ASD context is an arduous, long, and most of the times, an expensive task. It can be expensive because a large quantity of individuals with ASD and typical development (TD) is usually necessary to certify that the study will bring conclusions with statistical significance. Considering the papers analysed, we observed variations in relation to the age, the quantity of participants, the proportion of participants with and without ASD, and in other disorders that the participants might have.

Figure 2 shows the percentage of participants by experiment, considering intervals of twenty individuals. Around 33% of the experiments work with

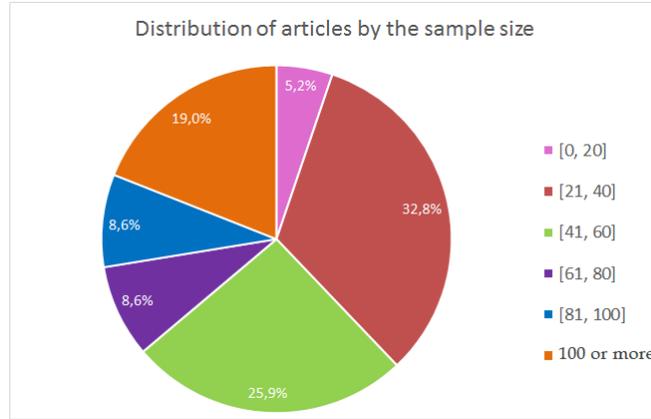


Figure 2: Distribution of included papers by sample size in the experiments

an interval of 21 to 40 individuals in total (considering all groups required by the experiment); 25% of the studies use from 41 to 60 people; and only 19% work with more than a hundred individuals.

The sample size is pointed out by many articles as a limitation, mostly because it is difficult to recruit participants that fulfil the criteria determined by the experiment, such as age or autism level, besides the fact that most of the works had to reject low quality data. Around 50% of the works analysed had to reject some of their participants data, related to the group that is investigated or the control group.

In general, as mentioned in section 3.1, cases in which the eye tracking data was not correctly captured, or not captured in its totality, were rejected. Other reasons that explain the exclusion of some participants are: visual impairments [27, 28], failure to meet all the criteria used for the study [29, 30, 31], the participants lack of attention caused by agitation, sleep, crying or complaining during the experiment [32, 33, 34] or the fact that the participant could not remain with the equipment – in case of the head-mounted device [35].

On average, the samples mentioned in the works analysed are balanced between the quantity of ASD and the quantity of TD groups. Around 79% of the works are in an interval in which 30-65% of the participants have ASD. From the works that have less than 30% participants with ASD, approximately 54% have other comparison groups. For example, the work of Schwartzman et al. [36] compares individuals with ASD, Rett syndrome, and

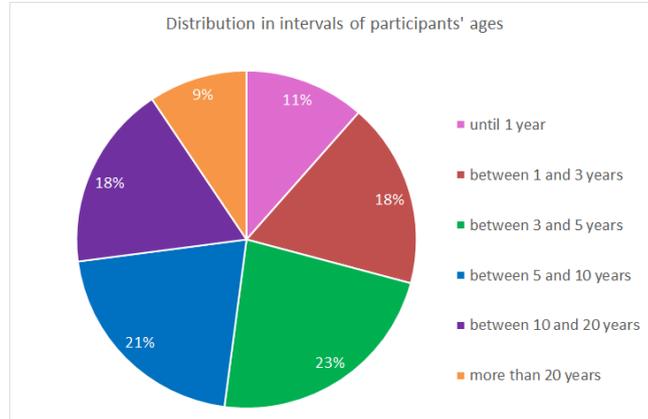


Figure 3: Distribution of papers included by participants' age range

the control group. Around 21% of the works examines these comparisons with other disorders, mentioning above: invasive development disorder with any other specification, Rett syndrome, Willians syndrome, development delay, speech delay, and Down syndrome.

Figure 3 shows the distribution of participants' age in the experiments. Some works are included in more than one range of the graphic, as they use participants with different ages. The work of Nakano et al. [33], for instance, compares individuals from 2-9 years and 22-36 years with the purpose of verifying if the atypical gaze of individuals with ASD changes with the development.

In relation to the age of the participants, 23% of the analysed works have some participants with 3-5 years, as shown in Figure 3. Very few works, only 11%, mention the inclusion of children younger than 1 year old to make the experiment. A hypothesis that could justify this percentage is the difficulty in using eye tracking, especially in the adjustment phase, and the need of a later follow-up to confirm if the toddlers with ASD risk really developed the disorder. Only 9% of the included studies used people older than 20 years.

There are some works that use the chronological age as a matching criteria among the groups with ASD, control group and other disorders group [37, 23, 27, 38, 25]. Other studies use verbal or nonverbal intelligence measurements, [39, 40]. [19] use the adaptive behaviour, measured by the Vineland Adaptive Behavior Scale. According to the authors, this measure ensures the grouping of children considering similarities in the abilities

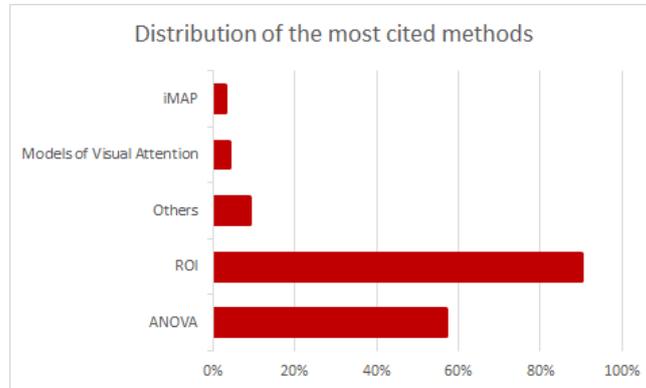


Figure 4: Distribution of the most cited methods to perform signal analysis in the included studies.

to perform day-to-day tasks. The majority of the works (31%) decided to use more than one matching measurement, as chronological age, and verbal and nonverbal development.

### 3.3. Methods and Techniques

The analysed works indicate the use of computational or statistical methods in different stages of the experiment, with the purpose of aiding the process of analysing the gaze movements of ASD participants. The main tasks of these methods are the treatment of visual stimuli before using the experiment, the selection of regions for the analysis of the stimuli after the data were captured, and then the data analysis obtained in the capture. In the following subsections the main methods found to each of these stages are described. Figure 4 shows the distribution of the methods that deserves attention among the analysed works.

#### 3.3.1. Stimuli Treatment

Only in 20% of the analysed studies some method to treat the visual stimuli was used. One of the treatments applied in static stimuli is the 'Bubbles' method used in Neumann et al. [41]. This work aims to verify if individuals with ASD looked more to the mouth region, even if this region was not clear on the face. In this regard, it was applied the Bubbles method, in which the stimuli images were transformed by the smoothing and enhancing methods, in a way that the faces could be revealed in random points and blurred in others.

Dundas et al. [42], instead of showing various different stimuli, manipulated a face image six times using a software for facial transformation, varying distances and sizes of nose, mouth, forehead and lip. According to the authors, the image of only one person was manipulated in order to maximize the attention to the internal characteristics of the face.

In Yi et al. [31] the images were converted to the grey scale, combined in relation to the shine and luminance, and were cut to match a normalized template. In McPartland et al. [35], in turn, images were inverted to check if the orientation shifted the attention of individuals with ASD.

In the case of dynamic stimulus, three types of treatment were found: (a) stimulus inversion, as in Shic et al. [43]; (b) sound removal, also described in Shic et al. [43]; and (c) the screen division in two or more parts, presenting different scene types [24, 12, 25, 44].

Either to maximise the attention as in Dundas et al. [42] or to verify some more specific hypothesis as in McPartland et al. [35] and in Neumann et al. [41], these treatments can transform the stimulus according to the goal of the study, facilitating the analysis.

### *3.3.2. Data Selection and Analysis*

With the purpose of selecting the stimulus regions to be processed after the eye tracking, two methods are highlighted: ROI (Regions of Interest), also called AOI (Areas of Interest), and iMap. The main issue of the use of the ROI method is that the regions of interest need to be defined according to the priori knowledge of the researcher, which can lead to loss of information. The regions are delimited manually or with the help of a software. As shown in Figure 4, the majority of the studies still use this type of approach. The delimited regions, in general, depend on the visual paradigm that is been analysed. Johnels et al. [45], for example, aims to analyse whether the time that the child spend looking at the mouth, eyes and other areas can determine which are the individuals with or without ASD. To help identifying it, a rectangle was marked in the eye region and an ellipse was marked in the mouth region.

In the iMap method, a priori knowledge of the researcher is not necessary, because there is no need to delimit the areas in the stimulus. The iMap creates fixations maps of each participant and each visual stimulus, resulting in an average fixation map. The average fixation maps of the individuals taken as a set can then be calculated, and ends up as a group fixation map. The main advantage of the iMap is the application of all the fixation points

to generate fixation maps that would allow the statistical comparison of these data [46]. The works of [31] and [25] use these two methods, enabling the analysis of specific points of the stimulus and of the stimulus visualization as a whole. According to [25], the iMap seems to be more objective, accurate and visual, besides providing a better spacial resolution. Further, it allows detecting distinctions among the groups in the spacial distribution of the fixation in an easier way. However, both works used iMap only as a method to confirm the results obtained by the ROI.

The work of [47] created its own approach, based on statistical spatial methods, using a Voronoi Diagram. According to the authors, this approach was able to define, a posteriori, the spatial location of fixation agglomeration, avoiding the need of the researcher to define the ROIs.

The work of [28] proposes a differentiated method to visualize the eye tracking data, a measure named D2R. The D2R is a measure that combines the two spatial dimensions in one, and maintain the information to analyse the data in a two-dimensional graphic that relates space and time. This method also favors the verification of movement patterns of the eyes among the groups.

In [33] a multidimensional scale (MDS) was used, in which the distance between each pair of fixation points was calculated. If the fixation point of two participants was the same, the distance was zero, if they looked at different points the distance was the quantity of pixels that separates these two points. A distance matrix was defined repeating this procedure for each pair of participants. With the matrix, it was applied the MDS to plot each participant in a two-dimensional plane. In this plane, it was possible to verify that the participants without ASD were distributed in the middle of the plane, which indicates a fixation pattern. Yet, the participants with ASD were distributed towards the periphery of the plane.

To support the data analysis, some works have used the visual attention bottom-up model, that proposes to predict the visual attention with one specific stimulus. In general, the bottom-up visual attention models draw the visual characteristics of stimulus, as color, intensity, orientation and even movement . These characteristics are combined and processed producing a saliency map, on what the higher value represents the local in which the attention will be draw.

One of the works that uses this model type is Neumann et al. [41], that employed the bottom-up model of Itti et al. [48] to analyse the atypical attention of the autistic in relation to facial images. Comparing the saliency

map provided by the model with attention from the ASD individuals, it was noted that the attention paid to the mouth was not related to the bottom-up characteristics of the image. The work of [40] verifies whether the visual saliency exerts influence in the attention to adolescents with and without ASD in the same way, also using the Itti's model. In both groups, participants fixated in regions that were above the average visual saliency, what was indicated by the fact that the saliency in the fixation was higher than the mean saliency value. The researchers also observed that the participants with ASD stared at faces later in the research. In [43] the Itti's model was used to compare the attention of children with and without ASD, using videos with various orientations and biological motions as stimulus. It was observed that the image orientation does not affect the attention of the children with ASD in the same way it happens with children without ASD that cared less about the movement.

In Wang et al. [49] it is used a visual attention model that takes into consideration all the fixation points marked by each group of individuals. Then, fixation maps are done related to the group with ASD and the control group. In general, these maps are created extracting the attributes from the entry images: pixel-level attributes, object-level attributes, and semantic attributes. It was used a Support Vector Machine (SVM) classifier to combine the fixation points with the extracted attributes aiming to create the saliency maps. By analysing the maps, it was possible to observe that individuals with ASD fixated more to the center of the image, even when there was nothing in that region. They were also slower to stare at faces, as they did it faster when talking about mechanic and manipulable objects.

Finally, all the analysed works in this review use statistical methods to analyse data and validate the hypothesis. The main method used for this purpose was the ANOVA, used in 57% of the works. Some works also used Pearson correlation [50], likelihood estimator [51], and nonparametric tests [52], among others.

Therefore, against the hypothesis of this paper, just a few computational methods were found. The iMap seems to be a promising method, appearing just a few years ago, in 2013 [31]. However, it is still not fully used to the analysis. One of the first methods to appear was the Model of Visual Attention, in the paper of Neumann et al. [41], analysing the bottom-up attention relation to the mouth. Another model more complex appears in 2015, in [49], revealing different discussions and conclusions. Showing the power of such methods to this kind of analysis.

#### *3.4. Parameters for aiding diagnosis*

Six of the works analysed define some criteria that can distinguish, among the groups, individuals with and without ASD with a certain sensitivity. The first work to define a criterion was Nakano et al. [33]. The distance between the fixation of the participants was plotted in a two-dimensional plan and, then, it was used to specify the groups, as described in the section 3.3.2. Two Receiver Operating Characteristic (ROC) curves were built considering these distance variations. One of the curves compared adults and the other, children with and without ASD. While the first Area Under the Curve (AUC) was 0.75, the second was 0.87. The only problem observed in this method is that it does not consider the visual stimulus data to determine the measure, which gives us the impression that the same pattern would be found for any kind of stimulus. Furthermore, as it does not use visual information, other disorders could fit in the same profile of the individuals with ASD.

The work of Pierce et al. [12] defines a criterion based on the paradigm of biological motion. The stimulus consists of a screen divided into two halves: one half with biological motion and the other half with geometrical motion. From the analysis, it was concluded that an individual is accurately classified as 100% autistic when 69% or more of the total fixation duration occurs in the geometric images. However, the authors could not draw a conclusion when the duration was less than 69% of the time. In a later work [24] the authors confirmed these results in an experiment with a higher number of children. The results demonstrated 21% of sensitivity and 98% of specificity. Changing the threshold of the fixation duration in the geometric area, Pierce obtained a ROC curve with AUC of 0.71.

Shi et al. [25] compare the stimulus created by Pierce et al. [24] with a stimulus applied to more than one child in the biological motions. In this case it was obtained an AUC of 0.865 to the second stimulus, and 0.623 to the Pierces original stimulus. Despite the fact that Shi et al. [25] use a small sample of only 33 individuals, this work enables us to examine that the more the social interactions were displayed in the scene, the greater was the capacity of the method to split the groups.

Chevallier et al. [44] use three different stimuli: the first with static images of people or objects, the second with silent videos with people, and objects and the last with children playing. According to the authors, the one that brings the interactions between children is the most reliable when distinguishing the groups, with an AUC of 0.71. In this work, the chosen parameter to build the curve was the total duration in the faces fixation.

In spite of the promising improvements of the results, these values cannot endorse such methods as a valid tool to help with the diagnosis. More complex samples and more effective methods of analysis can be a starting point to improve these results.

#### **4. Advances, limitations, opportunities, and conclusions**

Considering the advance of technology, new possibilities were revealed inside the scope of this current review, as presented below.

##### *4.1. Methods*

One of the advancements was the use of methods such as iMap and visual attention instead of the habitual manual definition of ROI. The isolated use of ROI can cause information loss as it depends on humans and, to be more specific, on a specialist who most of the times needs to delimitate the stimulus regions that will be mapped to capture and analyse data. Even though this demarcation could contribute reducing the dimensionality, the method in its essence may improperly ignore other regions that were not delimited and can also exclude more detailed information about the areas that were delimited.

There are also the visual attention models that enable the analysis of the specific regions of the scenes as well as the attributes of the stimulus. In [49], for example, besides analysing these attributes, it is created a saliency map with all the fixation points for each of the analysed groups, enhancing the analysis and a more complex and profound comparison.

Despite the improvement, in the analysed works we found only a few effective computational methods that could help with the detection of the patterns in the visual attention of individuals with ASD. Considering the complexity of the data, there is a gap to be fulfilled. The use of methods of pattern recognition or machine learning such as Neural Networks, SVM, Decision Tree, NaiveBayes, and others, can be explored to contribute in a more effective way to the analysis of the eye tracking signs. Such methods could be used, for instance, to identify which stimuli features (color, orientation, semantic, etc), or metrics (time of fixation, order of fixation, etc), or the combination of those, will be better suited to distinguish the groups.

##### *4.2. Participants of the experiments*

Another interesting point that constitutes a good research opportunity is related to the participants selection for the experiments. Some limitations

were presented by the studies, and one of them refers to the type of disorder considered. High-functioning autistics participants are most of the times chosen because they can easily pay attention to the stimulus that they are submitted to. Besides, with these groups of individuals, the use of eye tracking devices attached to their bodies (like head-mounted devices) is usually possible, and it does allow that the visual stimulus ends up turning into a real interaction with another person.

However, by testing the approaches only with this group of individuals, the results cannot be considered for the whole spectrum, because it only represents this specific group of autistics. For all that, the development of methods that consider broader groups of the autistic spectrum is a need. And the same happens with the development of more effective technologies capable of tracking the look in a less invasive way and with less data loss. These two are challenges that need to be more profoundly examined in research projects.

One of the main limitations of the studies in relation to the participants is the impossibility of differentiating, up to now, what is exclusively derived from the autism spectrum and what can be caused by comorbid disorders. The study with heterogeneous groups of autistics that uses eye tracking together with another biomarkers such as genetic evaluation, neuropsychological test, among others, is essential to identify subgroups inside ASD. Thus, besides the accurate and early diagnosis, it will be possible to indicate more targeted therapies to each case and predict the response of the individuals to these therapies.

The solution for the aforementioned question includes the need of obtaining a broad sample of individuals in which different cases and its results could be analysed. Nevertheless, the sample size is also pointed as a limitation in the majority of the studies [27, 32, 53, 52]. As demonstrated in section 3.2, only 19% of the studies made experiments with more than a hundred individuals. Another limitation mentioned in the works is related to the pairing of the autistic group and the control group. Some studies [54] say that because they do not have a sufficient sample, it was not possible to pair the group in relation to the IQ or age. Because of this, a more specific analysis is not possible. The age of the group that participates in the experiment is also seen as a limitation. Experiments conducted with individuals from a specific age range do not allow the generalization of the outcomes.

The aforementioned limitations related to the participants of the experiments refer to the absence of a public database good enough to test the

different approaches and paradigms proposed by the researchers. Furthermore, the different aspects involved in the experiments, methods, devices, and volunteer attributes, make it difficult to compare the outcomes among the various works and to complete an effective conclusion.

In order to form this database, the volunteer recruitment involves ethical, financial, and practical concerns. It is really hard to form a database that considers a great variety of cases in relation to the different aspects involved (disorder level, the existence of other syndromes, age, gender, among others) as it is hard to form a basis that have paired control cases considering one or more of these aspects. In these terms, in all the phases of the process – from the determination of appropriate protocols until the definition of efficient forms of storing and retrieving these data – there are research challenges that represent important themes in need of investigation.

#### *4.3. Stimuli*

In relation to the stimuli, some limitations mentioned here refer to the type of stimulus used. Some studies state that the dynamic stimuli and, more than that, the static stimuli, are far from reality, and can present different outcomes from the daily ones [55]. To make the use of more realistic stimuli possible, we need to use more invasive devices, such as helmets, that can limit, as said before, the participants group of the experiment.

Some studies using video and/or photos [25], cite that one limitation is the fact that the stimuli used show only Caucasian children, what could limit the results. The composition of the right stimuli must consider aspects such as the target audience of the experiment, the paradigm that will be evaluated, and the methods that will be applied. It is a complex task that necessarily needs to involve specialists from various areas. Some works disclose the sources they use to create stimuli. In [24] the social stimulus part was taken from the project YogaKids<sup>6</sup>, while [41] worked with the facial images of [56]. This is a good research practice that should be considered by different groups as a way to allow more effective improvements in this area. Besides that, the creation and the provision of bases of public stimuli, considering the various paradigms that are known, could be a contribution to the advancement of the state of the art in this research area.

---

<sup>6</sup>Yoga Kids 3; Gaiam, Boulder, Colorado, <http://www.gaiam.com>, created by Marsha Wenig, <http://yogakids.com/>

#### 4.4. Conclusions

Even with the limitations discussed, the use of eye tracking for autism diagnosis purposes is a promising tool when compared to the ones used nowadays. Using eye tracking would eliminate the necessity of translation, adaptation and standardization of tools in each region, which are essential to the gold standard devices. Moreover, the collection and processing of the signals could take only a few minutes.

To this end, some works have been able to define criteria which distinguish individuals with or without ASD. [12, 25] and [24] utilize criteria and similar stimuli based on the biological motion paradigm that defines the total duration of fixation in non-biological movements as the criterion for the diagnosis. In short, the more the individual stares at a spot in this region, the greater the chances of having ASD. [44] uses one stimulus with children playing, and considers the total duration of fixation in the face as the criterion to diagnose, having an opposite relation: the less time staring at a face, the higher the chances of showing ASD.

These results point to promising advances but it is still necessary to fill in some gaps. From our understanding, more research and collaboration are necessary to enable experiments considering bigger and better samples, as well as more robust analysis methods to effectively use the data provided from eye tracking to assist the diagnosis in clinical routine.

- [1] M. L. Rutter, Progress in understanding autism: 2007–2010, *Journal of autism and developmental disorders* 41 (4) (2011) 395–404.
- [2] AMERICAN PSYCHIATRIC ASSOCIATION, Manual diagnóstico e Estatístico de Transtornos Mentais-: DSM-5, Artmed Editora, 2014.
- [3] C. Lord, M. Rutter, A. 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) (1994) 659–685.
- [4] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles, M. Rutter, The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism, *Journal of autism and developmental disorders* 30 (3) (2000) 205–223.

- [5] O. I. Lovaas, Behavioral treatment and normal educational and intellectual functioning in young autistic children., *Journal of consulting and clinical psychology* 55 (1) (1987) 3.
- [6] H. Cohen, M. Amerine-Dickens, T. Smith, Early intensive behavioral treatment: Replication of the ucla model in a community setting, *Journal of Developmental & Behavioral Pediatrics* 27 (2) (2006) S145–S155.
- [7] C. Lord, J. McGee, Committee on educational intervention for children with autism. *educating children with autism* (2001).
- [8] R. J. Landa, K. C. Holman, E. Garrett-Mayer, Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders, *Archives of general psychiatry* 64 (7) (2007) 853–864.
- [9] A. Klin, C. Klaiman, W. Jones, Reducing age of autism diagnosis: developmental social neuroscience meets public health challenge., *Revista de neurologia* 60 (S01) (2015) S3–S11.
- [10] A. Klin, W. Jones, R. Schultz, F. Volkmar, D. 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) (2002) 809–816.
- [11] M. Elsabbagh, M. H. Johnson, Getting answers from babies about autism, *Trends in cognitive sciences* 14 (2) (2010) 81–87.
- [12] K. Pierce, D. Conant, R. Hazin, R. Stoner, J. Desmond, Preference for geometric patterns early in life as a risk factor for autism, *Archives of general psychiatry* 68 (1) (2011) 101–109.
- [13] K. Chawarska, S. Macari, F. Shic, Context modulates attention to social scenes in toddlers with autism, *Journal of Child Psychology and Psychiatry* 53 (8) (2012) 903–913.
- [14] Z. Boraston, S.-J. Blakemore, The application of eye-tracking technology in the study of autism, *The Journal of physiology* 581 (3) (2007) 893–898.

- [15] T. Falck-Ytter, S. Bölte, G. Gredebäck, Eye tracking in early autism research, *Journal of neurodevelopmental disorders* 5 (1) (2013) 28.
- [16] Q. Guillon, N. Hadjikhani, S. Baduel, B. Rogé, Visual social attention in autism spectrum disorder: insights from eye tracking studies, *Neuroscience & Biobehavioral Reviews* 42 (2014) 279–297.
- [17] E. A. Papagiannopoulou, K. M. Chitty, D. F. Hermens, I. B. Hickie, J. Lagopoulos, A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders, *Social neuroscience* 9 (6) (2014) 610–632.
- [18] S. Magrelli, P. Jermann, B. Noris, F. Ansermet, F. Hentsch, J. Nadel, A. Billard, Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings, *Frontiers in psychology* 4.
- [19] B. Noris, J. Nadel, M. Barker, N. Hadjikhani, A. Billard, Investigating gaze of children with asd in naturalistic settings, *PLoS One* 7 (9) (2012) e44144.
- [20] M. Hanley, M. McPhillips, G. Mulhern, D. M. Riby, Spontaneous attention to faces in asperger syndrome using ecologically valid static stimuli, *Autism* 17 (6) (2013) 754–761.
- [21] M. Guimard-Brunault, N. Hernandez, L. Roché, S. Roux, C. Barthélémy, J. Martineau, F. Bonnet-Brilhault, Back to basic: do children with autism spontaneously look at screen displaying a face or an object?, *Autism research and treatment* 2013.
- [22] N. Hernandez, A. Metzger, R. Magné, F. Bonnet-Brilhault, S. Roux, C. Barthelemy, J. Martineau, Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning, *Neuropsychologia* 47 (4) (2009) 1004–1012.
- [23] L.-H. Wang, S. H.-L. Chien, S.-F. Hu, T.-Y. Chen, H.-S. Chen, Children with autism spectrum disorders are less proficient in action identification and lacking a preference for upright point-light biological motion displays, *Research in Autism Spectrum Disorders* 11 (2015) 63–76.

- [24] K. Pierce, S. Marinero, R. Hazin, B. McKenna, C. C. Barnes, A. 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*.
- [25] L. Shi, Y. Zhou, J. Ou, J. Gong, S. Wang, X. Cui, H. Lyu, J. Zhao, X. Luo, Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders, *PloS one* 10 (3) (2015) e0122280.
- [26] N. J. Sasson, J. T. Ellison, L. M. Turner-Brown, G. S. Dichter, J. W. Bodfish, Brief report: Circumscribed attention in young children with autism, *Journal of Autism and Developmental Disorders* 41 (2) (2011) 242–247.
- [27] C. J. Anderson, J. Colombo, D. Jill Shaddy, Visual scanning and pupillary responses in young children with autism spectrum disorder, *Journal of Clinical and Experimental Neuropsychology* 28 (7) (2006) 1238–1256.
- [28] T. Falck-Ytter, C. von Hofsten, C. Gillberg, E. Fernell, Visualization and analysis of eye movement data from children with typical and atypical development, *Journal of autism and developmental disorders* 43 (10) (2013) 2249–2258.
- [29] N. J. Sasson, L. M. Turner-Brown, T. N. Holtzclaw, K. S. Lam, J. W. Bodfish, Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays, *Autism Research* 1 (1) (2008) 31–42.
- [30] L. Marsh, A. Pearson, D. Ropar, A. d. C. Hamilton, Predictive gaze during observation of irrational actions in adults with autism spectrum conditions, *Journal of autism and developmental disorders* 45 (1) (2015) 245–261.
- [31] L. Yi, Y. Fan, P. C. Quinn, C. Feng, D. Huang, J. Li, G. Mao, K. Lee, Abnormality in face scanning by children with autism spectrum disorder is limited to the eye region: Evidence from multi-method analyses of eye tracking data, *Journal of vision* 13 (10) (2013) 5.

- [32] N. Merin, G. S. Young, S. Ozonoff, S. J. Rogers, Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants, *Journal of autism and developmental disorders* 37 (1) (2007) 108–121.
- [33] T. Nakano, K. Tanaka, Y. Endo, Y. Yamane, T. Yamamoto, Y. Nakano, H. Ohta, N. Kato, S. Kitazawa, Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour, *Proceedings of the Royal Society of London B: Biological Sciences* 277 (1696) (2010) 2935–2943.
- [34] W. Jones, K. Carr, A. 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) (2008) 946–954.
- [35] J. C. McPartland, S. J. Webb, B. Keehn, G. Dawson, Patterns of visual attention to faces and objects in autism spectrum disorder, *Journal of autism and developmental disorders* 41 (2) (2011) 148–157.
- [36] J. S. Schwartzman, R. d. L. Velloso, M. E. F. DAntino, S. Santos, The eye-tracking of social stimuli in patients with rett syndrome and autism spectrum disorders: a pilot study, *Arquivos de neuro-psiquiatria* 73 (5) (2015) 402–407.
- [37] F. Shic, J. Bradshaw, A. Klin, B. Scassellati, K. Chawarska, Limited activity monitoring in toddlers with autism spectrum disorder, *Brain research* 1380 (2011) 246–254.
- [38] Q. Guillon, M. H. Afzali, B. Rogé, S. Baduel, J. Kruck, N. Hadjikhani, The importance of networking in autism gaze analysis, *PloS one* 10 (10) (2015) e0141191.
- [39] L. L. Speer, A. E. Cook, W. M. McMahon, E. Clark, Face processing in children with autism effects of stimulus contents and type, *Autism* 11 (3) (2007) 265–277.
- [40] M. Freeth, T. Foulsham, P. Chapman, The influence of visual saliency on fixation patterns in individuals with autism spectrum disorders, *Neuropsychologia* 49 (1) (2011) 156–160.

- [41] D. Neumann, M. L. Spezio, J. Piven, R. Adolphs, Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention, *Social cognitive and affective neuroscience* 1 (3) (2006) 194–202.
- [42] E. Dundas, H. Gastgeb, M. S. Strauss, Left visual field biases when infants process faces: A comparison of infants at high-and low-risk for autism spectrum disorder, *Journal of autism and developmental disorders* 42 (12) (2012) 2659–2668.
- [43] F. Shic, B. Scassellati, D. Lin, K. Chawarska, Measuring context: The gaze patterns of children with autism evaluated from the bottom-up, in: *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on, IEEE, 2007*, pp. 70–75.
- [44] C. Chevallier, J. Parish-Morris, A. McVey, K. M. Rump, N. J. Sasson, J. D. Herrington, R. T. Schultz, Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters, *Autism Research* 8 (5) (2015) 620–628.
- [45] J. Å. Johnels, C. Gillberg, T. Falck-Ytter, C. Miniscalco, Face-viewing patterns in young children with autism spectrum disorders: Speaking up for the role of language comprehension, *Journal of Speech, Language, and Hearing Research* 57 (6) (2014) 2246–2252.
- [46] R. Caldara, S. Mielle, imap: a novel method for statistical fixation mapping of eye movement data, *Behavior research methods* 43 (3) (2011) 864–878.
- [47] A. Amestoy, E. Guillaud, M. P. Bouvard, J.-R. Cazalets, Developmental changes in face visual scanning in autism spectrum disorder as assessed by data-based analysis, *Frontiers in psychology* 6.
- [48] L. Itti, Models of bottom-up and top-down visual attention, Ph.D. thesis, California Institute of Technology (2000).
- [49] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs, Q. Zhao, Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking, *Neuron* 88 (3) (2015) 604–616.

- [50] J. Å. Johnels, C. Gillberg, T. Falck-Ytter, C. Miniscalco, Face-viewing patterns in young children with autism spectrum disorders: Speaking up for the role of language comprehension, *Journal of Speech, Language, and Hearing Research* 57 (6) (2014) 2246–2252.
- [51] M. Elsabbagh, R. Bedford, A. Senju, T. Charman, A. Pickles, M. H. Johnson, B. Team, et al., What you see is what you get: contextual modulation of face scanning in typical and atypical development, *Social cognitive and affective neuroscience* 9 (4) (2014) 538–543.
- [52] T. Falck-Ytter, E. Thorup, S. Bölte, Brief report: Lack of processing bias for the objects other people attend to in 3-year-olds with autism, *Journal of autism and developmental disorders* 45 (6) (2014) 1897–1904.
- [53] G. Bird, C. Press, D. C. Richardson, The role of alexithymia in reduced eye-fixation in autism spectrum conditions, *Journal of autism and developmental disorders* 41 (11) (2011) 1556–1564.
- [54] T. Falck-Ytter, Face inversion effects in autism: a combined looking time and pupillometric study, *Autism Research* 1 (5) (2008) 297–306.
- [55] J. Van der Geest, C. Kemner, G. Camfferman, M. Verbaten, H. van Engeland, Looking at images with human figures: comparison between autistic and normal children, *Journal of autism and developmental disorders* 32 (2) (2002) 69–75.
- [56] P. Ekman, W. V. Friesen, Measuring facial movement, *Environmental psychology and nonverbal behavior* 1 (1) (1976) 56–75.

# **Anexos**

**Anexo A – Parecer do Comitê de Ética**

**PARECER CONSUBSTANCIADO DO CEP**

**DADOS DO PROJETO DE PESQUISA**

**Título da Pesquisa:** PROCESSAMENTO DO RASTREAMENTO DE OLHAR COM BASE EM MODELO DE ATENÇÃO VISUAL PARA AUXÍLIO AO DIAGNÓSTICO DE AUTISMO

**Pesquisador:** Jéssica dos Santos de Oliveira

**Área Temática:**

**Versão:** 2

**CAAE:** 57185516.9.0000.5390

**Instituição Proponente:** UNIVERSIDADE DE SAO PAULO

**Patrocinador Principal:** Financiamento Próprio

**DADOS DO PARECER**

**Número do Parecer:** 1.669.810

**Apresentação do Projeto:**

O Transtorno do Espectro do Autismo (TEA) é uma anomalia do neurodesenvolvimento que afeta principalmente as habilidades de interação social dos indivíduos. Para auxiliar no diagnóstico, estudos têm utilizado técnicas de rastreamento de olhar, mas, em geral, os sinais são interpretados manualmente ou a automatização considera somente alguns dos indícios que podem definir a presença do TEA em indivíduos. O presente projeto visa a preencher esta lacuna, propondo processar todos os indícios fornecidos pelos sinais do rastreamento do olhar com base em um modelo de atenção visual de forma a auxiliar no diagnóstico do TEA. Para alcançar o objetivo proposto, são previstas as seguintes etapas: revisão bibliográfica, aprovação do comitê de ética, definição do modelo de atenção visual, implementação do modelo, construção de um protótipo de sistema, análise e publicação dos resultados. Os resultados serão analisados com o auxílio de profissionais da saúde e avaliados por meio de técnicas estatísticas. Espera-se que os resultados obtidos sejam relevantes para profissionais da saúde, auxiliando-os no diagnóstico do TEA e contribuam com a área de Computação com o desenvolvimento de um modelo de atenção visual que posteriormente poderá ser utilizado em outras aplicações.

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**CEP:** 03.828-000

**UF:** SP

**Município:** SAO PAULO

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br

Continuação do Parecer: 1.669.810

**Objetivo da Pesquisa:**

O presente projeto de pesquisa tem como objetivo definir, implementar e validar técnicas para processamento do rastreamento de olhar com base em um modelo de atenção visual, visando ao auxílio no diagnóstico de autismo.

**Objetivo Secundário:**

- mapear os indícios da atenção visual atípica que auxiliam no diagnóstico do TEA;
- definir os estímulos, utilizando vídeos e/ou imagens, de acordo com os indícios visuais atípicos, que melhor auxiliem na detecção do TEA;
- implementar e avaliar um modelo computacional de atenção visual que realize a predição da atenção visual do TEA;
- desenvolver um protótipo de sistema para facilitar a validação e utilização do modelo desenvolvido;
- avaliar a eficiência da técnica utilizada com casos reais.

**Avaliação dos Riscos e Benefícios:**

**Riscos:**

O experimento pode apresentar riscos mínimos ao participante, como alguns desconfortos, cansaço visual, náusea e zonzeira.

**Benefícios:**

O benefício esperado é o desenvolvimento de um protótipo que auxilie no diagnóstico do TEA.

**Comentários e Considerações sobre a Pesquisa:**

Trata-se de projeto experimental de PG em Sistemas de Informação, Área de Concentração em Metodologias e Técnicas da Computação cuja amostra será constituída com cerca de 150 crianças de 3 a 5 anos. Sendo 75 com Transtorno do Espectro do Autismo (TEA) e 75 com Desenvolvimento Típico (DT). Os dados serão captados com auxílio do equipamento Tobii TX300, disponível no IPq-USP.

**Considerações sobre os Termos de apresentação obrigatória:**

O pesquisador apresentou o TCLE e o Termo de Assentimento conforme Resolução CNS CONEP 466/12. Todos outros termos foram apresentados.

**Recomendações:**

Um cuidado especial por envolver crianças.

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**UF:** SP

**Município:** SAO PAULO

**CEP:** 03.828-000

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br

Continuação do Parecer: 1.669.810

**Conclusões ou Pendências e Lista de Inadequações:**

Não há.

**Considerações Finais a critério do CEP:**

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_722214.pdf	04/07/2016 10:43:33		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TermoDeAssentimento.pdf	04/07/2016 10:42:24	Jéssica dos Santos de Oliveira	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE.pdf	04/07/2016 10:40:53	Jéssica dos Santos de Oliveira	Aceito
Folha de Rosto	FolhaDeRosto.pdf	08/06/2016 17:57:02	Jéssica dos Santos de Oliveira	Aceito
Projeto Detalhado / Brochura Investigador	Projeto.pdf	19/05/2016 17:52:06	Jéssica dos Santos de Oliveira	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SAO PAULO, 08 de Agosto de 2016

---

**Assinado por:**  
**Rosa Yuka Sato Chubaci**  
**(Coordenador)**

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**CEP:** 03.828-000

**UF:** SP

**Município:** SAO PAULO

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br