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

THIAGO VICTOR CARDOSO

**Auxílio ao diagnóstico do Transtorno do Espectro Autista com base em
trajetórias provenientes de rastreamento do olhar**

São Paulo

2020

THIAGO VICTOR CARDOSO

Auxílio ao diagnóstico do Transtorno do Espectro Autista com base em trajetórias provenientes de rastreamento do olhar

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 03 de dezembro de 2020. 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.

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

Coorientadora: Profa. Dra. Arianne Machado Lima

São Paulo

2020

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

Cardoso, Thiago Victor

Auxílio ao diagnóstico do Transtorno do Espectro Autista com base em trajetórias provenientes de rastreamento do olhar / Thiago Victor Cardoso ; orientadora, Fátima de Lourdes dos Santos Nunes Marques ; coorientadora, Arianne Machado Lima. – 2020.

103 f : il.

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

Versão corrigida

1. Reconhecimento de padrões. 2. Processamento de imagens. 3. Transtorno do Espectro Autista - Diagnóstico. 4. Aprendizado computacional. 5. Inteligência artificial. I. Marques, Fátima de Lourdes dos Santos Nunes, orient. II. Lima, Arianne Machado, coorient. III. Título

CDD 22.ed.— 006.4

Dissertação de autoria de Thiago Victor Cardoso, sob o título “**Auxílio ao diagnóstico do Transtorno do Espectro Autista com base em trajetórias provenientes de rastreamento do olhar**”, 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, na área de concentração Metodologia e Técnicas da Computação, aprovada em 03 de dezembro de 2020 pela comissão julgadora constituída pelos doutores:

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

**Profa. Dra. Rosa Maria Esteves Moreira da
Costa**
Universidade Federal do Rio de Janeiro

Profa. Dra. Deller James Ferreira
Universidade Federal de Goiás

À minha esposa Priscila, meu filho Miguel, meus pais Walter e Eliane, e a toda minha família, pela paciência e incentivo que tiveram comigo para que este projeto pudesse ser realizado.

Agradecimentos

Agradeço primeiramente a Deus, pelo dom da vida que me proporcionou, e com Seu infinito amor e misericórdia, me abençoou, por meio dos meus pais que me conduziram acerca dos princípios, caráter, fé e temor no Senhor! Sou grato a Deus, por ter me preparado, a minha ajudadora, excelente esposa e mãe do meu menino Miguel, que sem a sua ajuda e dedicação, este trabalho dificilmente seria concluído.

Agradeço a meus pais, Walter e Eliane, aos meus irmãos, Walter e Liliane, ao meu sogro Guilherme e sogra Maria Inês, por me apoiarem em todo o momento da minha vida, pelas constantes orações e pela imensa amizade. Agradeço também à minha prima Raquel, que não mediu esforços para me ajudar na revisão gramatical da escrita deste projeto.

Agradeço à professora Fátima, pela oportunidade, orientação, esforço, dedicação e pela colaboração com quem sempre contei para a realização dos estudos pertinentes a esta pesquisa. Seu engajamento em prol da minha formação é uma qualidade que me inspira. De igual modo, agradeço à professora Ariane, pelo esforço e contribuições exercidas para o aprimoramento dos resultados deste projeto.

À professora Helena, agradeço pelas considerações pertinentes ao estudo do autismo, que por meio de seu incrível engajamento e habilidade, foi possível desenvolver este estudo com mais assertividade e obter melhores resultados. Também agradeço a sua equipe, que durante as reuniões e atividades se comprometeram no apoio desta pesquisa.

Aos colegas de laboratório Renan e Leonardo, que me ajudaram com seus conhecimentos técnicos. Também agradeço à Jéssica, pela explicação de seu trabalho correlato e por fornecer os dados de experimentos. Com certeza, esses dados me pouparam um tempo considerável na pesquisa.

Por fim, agradeço aos professores, colegas e a todas as pessoas que, mesmo não sendo citadas diretamente, me apoiaram e incentivaram quanto ao desenvolvimento e conclusão deste trabalho.

“SENHOR, quem habitará no teu tabernáculo? Quem morará no teu santo monte? Aquele que anda em sinceridade, e pratica a justiça, e fala verazmente, segundo o seu coração.”

(Salmos 15 1:2)

Resumo

Cardoso, Thiago Victor. **Auxílio ao diagnóstico do Transtorno do Espectro Autista com base em trajetórias provenientes de rastreamento do olhar**. 2020. 103 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2020.

O uso do *Eye Tracking* (ET) tem sido investigado como mecanismo auxiliar para diagnóstico do Transtorno do Espectro Autista (TEA). Por meio de imagens estáticas ou vídeos usados como estímulo, o olhar é rastreado e os dados obtidos são analisados por técnicas diversas. Um dos paradigmas investigados usando ET é a Atenção Compartilhada (AC), a qual refere-se a momentos em que indivíduos estão focados no mesmo alvo, de modo que ambos estão cientes de que o foco de atenção é compartilhado. As ferramentas computacionais que visam auxiliar o diagnóstico de TEA têm utilizado técnicas de Processamento de Imagens e Aprendizado de Máquina para processar imagens, vídeos e sinais de ET. No entanto, o paradigma de AC ainda é pouco explorado e apresenta desafios, pois é necessário analisar a trajetória do olhar, exigindo abordagens inovadoras. O objetivo deste projeto é propor um modelo capaz de extrair características de um vídeo usado como estímulo para verificar a AC e classificar indivíduos como pertencentes ao grupo TEA ou Desenvolvimento Típico (DT). O principal diferencial em relação às abordagens da literatura é a definição e implementação do conceito de Regiões de Interesse flutuantes, que possibilitam o acompanhamento do olhar em relação a um objeto, considerando a sua semântica, mesmo que o objeto apresente características distintas ao longo do vídeo. Um modelo baseado em Comitê de Máquinas usando o algoritmo *Random Forest* foi implementado para classificar indivíduos como TEA ou DT usando as características de trajetória extraídas dos sinais de ET. O método alcançou 0,75 de acurácia e 0,82 de *F1-score*, indicando que a abordagem proposta, baseada em trajetória e AC, apresenta potencial para aplicação no auxílio ao diagnóstico de TEA.

Palavras-chaves: Transtorno do Espectro Autista. Comitê de Máquinas. Processamento de imagens. Aprendizado de Máquina. Atenção compartilhada. Rastreamento de olhar. *Random Forest*. Trajetória. Desenvolvimento Típico.

Abstract

Cardoso, Thiago Victor. **Aid to diagnosis of Autism Spectrum Disorder based on trajectories of eye tracking**. 2020. 103 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2020.

The use of Eye Tracking (ET) has been investigated as an auxiliary mechanism to diagnose Autism Spectrum Disorder (ASD). Through static images or videos used as a stimulus, the gaze path is tracked and the data obtained are analyzed by different techniques. One of the paradigms investigated using ET is Joint Attention (JA), which refers to moments when two individuals are focused on the same object/event so that both are aware that the focus of attention is shared. The computational tools that assist in the diagnosis of ASD have used Image Processing and Machine Learning techniques to process images, videos and ET signals. However, the JA paradigm is still little explored and presents challenges, as it requires analyzing the gaze trajectory and needs innovative approaches. The purpose of this project is to propose a model capable of extracting characteristics from a video used as a stimulus to verify JA and classify individuals as belonging to the ASD or Typical Development (TD) group. The main differential in relation to the approaches in the literature is the definition and implementation of the concept of floating Regions of Interest, which allows monitoring the gaze in relation to an object, considering its semantics, even if the object presents different characteristics throughout the video. A model based on Ensemble Method using the Random Forest algorithm was implemented to classify individuals as ASD or TD using the trajectory characteristics extracted from the ET signals. The method reached 0.75 accuracy and 0.82 F1-score, indicating that the proposed approach, based on trajectory and JA, has the potential to be applied to assist in the diagnosis of ASD.

Keywords: Autism Spectrum Disorder. Ensemble Method. Eye tracking. Image processing. Joint attention. Machine Learning. Random Forest. Trajectory. Typical Development

Lista de figuras

Figura 1 – Exemplo de mapa de saliência: à esquerda, o mapa de saliência gerado e à direita, os <i>pixels</i> mais fixados (em cor vermelho).	26
Figura 2 – Resumo do processo realizado na RS	29
Figura 3 – Quantidade de artigos publicados e aceitos nessa revisão por ano. . . .	30
Figura 4 – Fases gerais das abordagens.	35
Figura 5 – Porcentual dos artigos que utilizam técnicas de pré-processamento. Legenda: I-VT = <i>Identification Velocity-Threshold</i>	36
Figura 6 – Métricas utilizadas nos artigos. Legenda: (ROC = <i>Receiver Operating Characteristic</i> , CC = Coeficiente de Correlação, NSS = <i>Normalized Scanpath Saliency</i> , AUC = <i>Area Under the ROC Curve</i>).	42
Figura 7 – Porcentual dos trabalhos realizados por área de aplicação.	44
Figura 8 – Exemplos de vídeos utilizados como estímulo: no vídeo A, apenas o olhar da terapeuta se movimenta; no vídeo B, o olhar se movimenta em relação ao objeto de interesse; no vídeo C, a mão aponta para o objeto de interesse.	54
Figura 9 – Exemplo da ferramenta de visualização de trajetórias exibindo o rastreamento do olhar de 15 indivíduos DT (lado esquerdo) e TEA (lado direito).	56
Figura 10 – Exemplo da visualização de trajetórias e as ROIs em movimento (retângulos demarcados em vermelho nas regiões do rosto, panda e mão). Rastreamento do olhar, por meio de variações das cores em azul (lado esquerdo) e vermelho (lado direito). Círculos em tons de azul do lado esquerdo correspondem a indivíduos do grupo DT e círculos em tons de vermelho do lado direito correspondem a indivíduos do grupo TEA.	57
Figura 11 – Visão geral da abordagem proposta.	58
Figura 12 – Exemplo de demarcação da ROI na imagem.	59
Figura 13 – Exemplo de ROIs flutuantes: o mesmo objeto apresenta diferentes características ao longo do vídeo.	60
Figura 14 – Processo para identificar objetos presentes nas ROIs flutuantes.	61

Figura 15 – Exemplo de cálculo para obter a antecipação ou atraso dos dados de ET a partir da ROI.	65
Figura 16 – Classificação por configuração e pesos homogêneos. Resultado da acurácia variando o limiar de 1 a 9.	75
Figura 17 – Classificação por configuração homogênea e pesos ponderados. Resultado da acurácia variando o limiar de 1,0 a 5,9.	76
Figura 18 – Classificação por configuração variável e pesos homogêneos. Resultado da melhor acurácia considerando cada uma das 512 configurações (eixo das abscissas).	76
Figura 19 – Gráfico de <i>boxplot</i> exibindo as distribuições das acurácias em cada uma das abordagens.	79

Lista de algoritmos

Algoritmo 1 – Classificação por configuração e pesos homogêneos.	70
Algoritmo 2 – Classificação por configuração homogênea e pesos ponderados.	71
Algoritmo 3 – Classificação por configuração variável e pesos homogêneos.	72

Lista de tabelas

Tabela 1 – Retorno dos resultados das máquinas de busca	28
Tabela 2 – Características extraídas dos artigos incluídos	31
Tabela 3 – Características utilizadas provenientes do rastreamento de olhar. . . .	55
Tabela 4 – Parâmetros usados no processo de calibração do modelo.	67
Tabela 5 – Parâmetros das melhores acurácias pelo método de calibração.	77
Tabela 6 – Parâmetros referentes aos melhores resultados do CM.	78
Tabela 7 – Comparação dos resultados entre o trabalhos relacionados. Legenda: AV - Atenção Visual; AC - Atenção Compartilhada	81

Lista de abreviaturas e siglas

ADI-R	<i>Autism Diagnostic Interview-Revised</i>
AC	Atenção Compartilhada
ACC	Acurácia
AB	<i>AdaBoost</i>
ACM	<i>Association for Computing Machinery</i>
ADOS	<i>Autism Diagnostic Observation Schedule</i>
AM	Aprendizado de Máquina
AUC	<i>Area Under the Curve</i>
ASD	<i>Autism Spectrum Disorder</i>
CARS	<i>Childhood Autism Rating Scale</i>
CC	Coefficiente de Correlação
CM	Comitê de Máquinas
DSM	<i>Diagnostic and Statistical Manual of Mental Disorders</i>
DT	Desenvolvimento Típico
EACH-USP	Escola de Artes, Ciências e Humanidades (EACH-USP)
EM	<i>Ensemble Method</i>
ET	<i>Eye Tracking</i>
FIT	<i>Feature Integration Theory</i>
fMRI	<i>Functional Magnetic Resonance Imaging</i>
F1	<i>F1-score</i>
HMM	<i>Hidden Markov model</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>

IPq-USP	Instituto de Psiquiatria - Universidade de São Paulo
I-VT	<i>Identification Velocity-Threshold</i>
JA	<i>Joint Attention</i>
KNN	<i>K Nearest Neighbors</i>
LApIS	Laboratório de Aplicações de Informática em Saúde
LDA	<i>Linear Discriminant Analysis</i>
ML	<i>Machine Learning</i>
PROTEA	Programa de Transtorno do Espectro Autista
PPV	<i>Positive Predictive Value</i> (ou Precisão)
NSS	Normalized Scanpath Saliency
ROC	<i>Receiver Operating Characteristic</i>
ROI	<i>Regions of interest</i>
RS	Revisão Sistemática
RF	<i>Random Forest</i>
Scielo	<i>Scientific Electronic Library Online</i>
SVM	<i>Support Vector Machine</i>
TEA	Transtorno do Espectro Autista
TD	<i>Typical Development</i>
TNR	<i>True Negative Rate</i> (ou Especificidade)
TPR	<i>True Positive Rate</i> (ou Sensibilidade)

Sumário

1	Introdução	17
1.1	<i>Motivação e justificativa</i>	18
1.2	<i>Objetivos e questão de pesquisa</i>	20
1.3	<i>Organização do documento</i>	21
2	Conceitos fundamentais	22
2.1	<i>Transtorno do Espectro Autista e atenção compartilhada</i>	22
2.2	<i>Rastreamento do olhar</i>	24
3	Revisão Bibliográfica	27
3.1	<i>Método de pesquisa</i>	27
3.2	<i>Visão geral dos estudos</i>	29
3.3	<i>Fases gerais das abordagens</i>	34
3.3.1	Coleta de dados	35
3.3.2	Pré-Processamento	35
3.3.3	Extração de características	37
3.3.4	Classificação	38
3.4	<i>Métricas de avaliação</i>	41
3.5	<i>Áreas de aplicação</i>	43
3.5.1	Classificação aplicada à área de saúde	44
3.5.2	Identificação biométrica para finalidade de segurança	45
3.5.3	Deteção de saliência	46
3.6	<i>Visualização de dados</i>	47
3.7	<i>Dispositivos</i>	48
3.8	<i>Discussão</i>	49
3.9	<i>Conclusão</i>	51
4	Metodologia	52
4.1	<i>Materiais</i>	52
4.1.1	Participantes	52
4.1.2	Estímulos	53

4.1.3	Protocolo	53
4.2	<i>Visualização de trajetórias</i>	55
4.3	<i>Classificação com base em trajetórias</i>	57
4.3.1	Identificação das ROIs flutuantes	58
4.3.2	Extração de características	61
4.3.3	Antecipação do olhar	64
4.3.4	Construção do modelo	65
5	Resultados e discussões	73
5.1	<i>Ferramentas e técnicas implementadas</i>	73
5.2	<i>Comparação das abordagens dos Comitês de Máquinas</i>	74
5.3	<i>Comparação dos classificadores</i>	80
5.4	<i>Limitações</i>	81
6	Considerações finais	82
6.1	<i>Contribuições</i>	83
6.2	<i>Trabalhos Futuros</i>	84
	REFERÊNCIAS	85
	ANEXOS	95
	Anexo A – Parecer do Comitê de Ética	96
	Anexo B – TCLE - Termo de Consentimento Livre e Esclarecido	100

1 Introdução

O Transtorno do Espectro Autista (TEA) é um transtorno do neurodesenvolvimento com apresentação clínica e impacto funcional variado (LORD *et al.*, 2020). O diagnóstico está principalmente associado a desafios centrais na comunicação social, bem como interesses restritos e comportamentos repetitivos (ASSOCIATION, 2014).

A interação de um indivíduo com seu meio e pares é um comportamento adaptativo fundamental. A sociabilidade é um processo temporal que reflete o desenvolvimento e o amadurecimento da conectividade cerebral. A capacidade rudimentar de seguir o direcionamento dos olhos está presente alguns dias após o nascimento em humanos (BILL *et al.*, 2020). Durante os primeiros dois anos de vida, os bebês humanos desenvolvem uma crescente compreensão referencial das pistas do olhar, as quais podem fornecer informações comunicativas sobre o conteúdo. A sensibilidade acerca dessas informações possibilita fazer inferências sobre intenções e motivações que direcionam o comportamento de outra pessoa, possibilitando responder a esses comportamentos e antecipar o comportamento subsequente (BILL *et al.*, 2020). O aprendizado baseado na atenção social atende dois critérios básicos: 1) concentrar a atenção nas partes mais informativas da cena; 2) adaptar seu comportamento em tempo real de acordo com o contexto (PARISH-MORRIS *et al.*, 2019). Desta forma, o sistema visual é muito importante para percepção do meio (AHISSAR; HOCHSTEIN, 2004) e a dificuldade em interpretar expressões faciais e linguagem corporal podem prejudicar o aprendizado e as interações sociais (FROST-KARLSSON *et al.*, 2019).

Em indivíduos portadores de TEA são observadas alterações de comportamentos precursores da sociabilidade como resposta menos frequente à invocação do próprio nome, olhar menos direcionado para a face, diminuição do apontar e de mostrar objetos (PALOMO; BELINCHÓN; OZONOFF, 2006). Um dos paradigmas investigados neste contexto é o comportamento de Atenção Compartilhada (AC), o qual refere-se a momentos em que dois indivíduos estão focados no mesmo alvo e ambos estão cientes de que o foco de atenção é compartilhado.

Segundo Gomes *et al.* (2015), o diagnóstico do TEA é, em geral, realizado por observações clínicas da criança, entrevistas com os pais e aplicações de instrumentos específicos (AMERICAN PSYCHIATRIC ASSOCIATION, 2014), como o Protocolo de Observação para Diagnóstico de Autismo, conhecido como ADOS (*Autism Diagnostic*

Observation Schedule), a Entrevista Diagnóstica para Autismo Revisada, conhecida como ADI-R (*Autism Diagnostic Interview-Revised*) e a Escala de avaliação para Autismo Infantil, conhecida como CARS (*Childhood Autism Rating Scale*). Estes instrumentos apresentam avaliações morosas, dispendiosas e dependente da experiência do profissional da saúde e da resposta do cuidador, o que geralmente pode resultar em uma diagnóstico subjetivo.

1.1 Motivação e justificativa

A tecnologia de *eye tracking* (ET) tem sido amplamente usada para investigar padrões visuais (BILLECI *et al.*, 2016) e tem se mostrado uma potencial ferramenta de auxílio para diagnóstico de TEA, viabilizando a análise e a identificação de padrões atípicos da atenção visual e oculomotores nestes indivíduos (SHIHAB; DAWOOD; KASHMAR, 2020). Crianças com autismo, diferentemente das crianças com desenvolvimento típico (DT), não apresentam uma preferência em observar faces e/ou movimentos biológicos; ao contrário, tendem a olhar para movimentos e cenas não sociais (KLIN *et al.*, 2009). Adicionalmente, há indícios de correlação entre a ausência de AC com prejuízos sociocomunicativos e atrasos de linguagem (BILLECI *et al.*, 2016). Os autores de (BILLECI *et al.*, 2016) sugerem também que pode haver dificuldades na alternância de estímulo e no desengajamento do olhar.

O processo para a realização da avaliação diagnóstica do TEA, com o uso do ET apresenta as seguintes vantagens:

- **Agilização do diagnóstico:** as avaliações existentes atualmente que resultam nos laudos de TEA costumam despende um tempo significativo, o que faz com que o processo apresente lentidão excessiva (SOCIEDADE BRASILEIRA DE PEDIATRIA, 2019).
- **Utilização da direção do olhar:** o dispositivo de ET define-se como eficiente, quando relacionado à precisão dos sinais de rastreamento do olhar, para avaliar indivíduos com TEA, porque captura sinais da direção do olhar de pacientes, constituindo-se uma das importantes características que definem o indivíduo com o espectro autista (FALCK-YTTER; HOFSTEN, 2011).

- **Acessibilidade financeira:** a ferramenta de ET é um dispositivo que pode estar disponível a um baixo custo, dependendo do modelo e fabricante, ou ainda, economicamente viável quando comparado com outros equipamentos, como o eletroencefalograma, facilitando o uso para instituições de saúde e profissionais dessa área (BARRETO, 2012).
- **Não invasividade:** ET não exige contato com o indivíduo, o que facilita seu uso principalmente em indivíduos que podem apresentar algum distúrbio.

Avaliações de ET da AC podem fornecer informações espaciais e temporais (HOFFSTEN *et al.*, 2009) de mudanças no monitoramento do ambiente, presentes precocemente no TEA. Considerando que a antecipação é a ação de fornecer uma visão do futuro imediato, com base no passado imediato (PALUMBO; BURNETT; JELLEMA, 2015), uma melhor compreensão de pistas e do contexto favorece a tarefa de antecipação, o que parece ser dificultado no TEA. No estudo de Hudson, Burnett e Jellema (2012), as respostas do TEA em relação a antecipação de ações sofreram mais influência das características dos objetos do que pelas intenções transmitidas. Ainda neste sentido, Hepach, Hedley e Nuske (2020) sugerem que o comportamento pró-social reduzido em crianças com TEA pode ser em parte devido a uma deficiência específica em antecipar o comportamento pró-social. Desta forma, investir em novos modelos para estudar a trajetória do olhar diante de dicas de intenção comunicativa, pode contribuir para o diagnóstico e prognóstico do TEA.

Embora possa contribuir sobremaneira com o diagnóstico precoce de TEA, a verificação da AC exige experiência, atenção e dispêndio de tempo do terapeuta, pois a observação do especialista sem auxílio de ferramentas é a prática mais comum (PARISH-MORRIS *et al.*, 2019; BEDFORD *et al.*, 2012). Mais recentemente alguns pesquisadores têm investigado o uso de ET como mecanismo de diagnóstico de TEA usando o paradigma da AC (BILLECI *et al.*, 2016; Little *et al.*, 2016). O estudo de Billeci *et al.* (2016) descreve os padrões visuais de crianças com TEA e DT durante as tarefas que induzem a AC respondente, bem como as tarefas que provocam a AC inicial. No estudo citado foi empregada a tecnologia de ET para monitorar o rastreamento do olhar por meio de vídeos envolvendo as duas tarefas citadas (BILLECI *et al.*, 2016).

Interpretar AC com auxílio de métodos computacionais envolve a dificuldade em mapear todas as situações presentes no estímulo do vídeo, delinear a trajetória percorrida durante a cena e processá-la adequadamente. Em geral, os trabalhos que utilizam sinais de

ET aliados a métodos computacionais não processam dados de trajetórias ou ignoram que o paradigma envolve o deslocamento das regiões de interesse, diferenciais contemplados no presente projeto. A partir do cenário apresentado, pretende-se processar dados de rastreamento de olhar previamente coletados, visando a definir um modelo capaz de classificar indivíduos como pertencentes ao grupo TEA ou ao grupo DT.

Esse estudo faz parte de um projeto mais abrangente que está sendo desenvolvido no Laboratório de Aplicações de Informática em Saúde (LApIS), da Escola de Artes, Ciências e Humanidades (EACH-USP), em colaboração com o Programa de Transtorno do Espectro Autista (PROTEA) pertencente ao Instituto de Psiquiatria da Universidade de São Paulo (IPq-USP), que pretende contribuir com o diagnóstico precoce do TEA.

1.2 Objetivos e questão de pesquisa

Este projeto tem o objetivo geral de propor um método computacional para classificar um indivíduo como pertencente ao grupo TEA ou ao grupo DT, a partir do processamento de características da trajetória percorrida pelo rastreamento do olhar, considerando vídeos utilizados como estímulos.

Os seguintes objetivos específicos são definidos:

- mapear as características inerentes aos sinais de trajetória;
- desenvolver um protótipo que possibilite reproduzir visualmente as trajetórias dos participantes a fim de compreender os comportamentos dos dois grupos (TEA e DT);
- identificar e extrair características relevantes para caracterizar o comportamento do rastreamento de olhar considerando o paradigma de AC;
- definir uma forma automatizada de acompanhar objetos de interesse nos *frames* dos vídeos utilizados como estímulo;
- implementar e validar um método computacional para classificar indivíduos portadores e não portadores do TEA.

A partir dos objetivos enunciados é definida a seguinte questão de pesquisa: é possível distinguir indivíduos dos grupos TEA e DT, por meio da análise dos dados de *eye tracking* referentes a acompanhamento de trajetórias usando o paradigma de AC?

1.3 Organização do documento

Este documento está organizado da seguinte forma: neste presente capítulo são apresentados a contextualização, a motivação, o objetivo geral e os objetivos específicos; o capítulo 2 apresenta os fundamentos conceituais relacionados ao TEA, ao ET e aos estímulos; o capítulo 3 faz um levantamento bibliográfico de trabalhos contendo sinais de rastreamento de olhar, que usam imagens como estímulos, por meio de uma revisão sistemática; o capítulo 4 descreve a metodologia utilizada para a definição do modelo baseado em trajetória, explanando a proposta e o processo de desenvolvimento realizado; o capítulo 5 descreve os resultados obtidos pela proposta, além da avaliação efetuada para validar o modelo, e as discussões pertinentes a este processo; o capítulo 6 apresenta as considerações finais do trabalho. No anexo [A](#) encontra-se o termo de aprovação do Comitê de Ética referente aos dados do experimento fornecidos pelo trabalho de ([OLIVEIRA, 2017](#)). No anexo [B](#) é apresentado o Termo de Consentimento Livre e Esclarecido (TCLE).

2 Conceitos fundamentais

A presente pesquisa abrange diversas áreas do conhecimento, cujos conceitos fundamentais são apresentados neste capítulo. As seções 2.1 e 2.2 abordam, respectivamente, os princípios básicos relacionados ao TEA e ao dispositivo de rastreamento do olhar.

2.1 Transtorno do Espectro Autista e atenção compartilhada

Segundo [Klin \(2006\)](#), o TEA está entre os transtornos de desenvolvimento mais comuns, afetando aproximadamente um em cada 200 indivíduos. De acordo com o Manual Diagnóstico e Estatístico de Transtornos Mentais (*Diagnostic and Statistical Manual of Mental Disorders* - DSM), pessoas com TEA apresentam algumas dificuldades para desenvolver, manter e compreender relacionamentos.

Os sintomas estão presentes precocemente no período de desenvolvimento e causam prejuízo clinicamente significativo no funcionamento social, profissional, ou em outras áreas importantes da vida do indivíduo ([AMERICAN PSYCHIATRIC ASSOCIATION, 2014](#)). Conforme o Manual de Orientação do Departamento Científico de Pediatria do Desenvolvimento e Comportamento, da Sociedade Brasileira de Pediatria, trata-se de um transtorno pervasivo e permanente, não havendo cura, ainda que a intervenção precoce possa alterar o prognóstico e suavizar os sintomas ([SOCIEDADE BRASILEIRA DE PEDIATRIA, 2019](#)).

Consoante [Klin \(2006\)](#), o TEA é um comprometimento permanente e a maioria dos indivíduos afetados por esta condição permanece incapaz de viver de forma independente, necessitando de apoio familiar, comunidade ou institucionalização. Porém, [Klin \(2006\)](#) aponta que a maioria das crianças com TEA apresenta melhora nos relacionamentos sociais, na comunicação e nas habilidades de autocuidado quando crescem.

O TEA pode ser observado nos primeiros anos de vida. Apesar dos sintomas não serem homogêneos, podem ser aparentes logo no nascimento da criança ou identificados entre 12 a 24 meses de idade ([SOCIEDADE BRASILEIRA DE PEDIATRIA, 2019](#)). O diagnóstico pode ser realizado a partir dos 12 meses de idade; contudo, há evidências de que o diagnóstico do TEA ocorre em média aos 4 ou 5 anos de idade, o que pode

prejudicar nos ganhos do funcionamento cognitivo e adaptativo da criança (SOCIEDADE BRASILEIRA DE PEDIATRIA, 2019).

Atualmente, os métodos de avaliações utilizados para verificar a presença do TEA constituem-se dos seguintes instrumentos (SOCIEDADE BRASILEIRA DE PEDIATRIA, 2019):

1. **Protocolo de Observação para Diagnóstico de TEA (ADOS):** fornece uma escala de observação da comunicação e dos comportamentos típicos de indivíduos portadores de TEA.
2. **Entrevista Diagnóstica para Autismo Revisada (ADI-R):** é composta por uma escala obtida a partir de entrevistas como meio de avaliar uma sequência de comportamentos que são condizentes com pessoas que apresentam o TEA.
3. **Escala de avaliação para Autismo Infantil (CARS):** utiliza quinze grupos de características, comportamentos e capacidades que são comparados com os valores padrões de indivíduos com o TEA.

Considerando as limitações dos instrumentos citados e o avanço da tecnologia, alguns comportamentos têm sido objetos de estudo para o desenvolvimento de modelos computacionais visando a contribuir com o diagnóstico de TEA. Entre esses comportamentos está a Atenção Compartilhada.

O comportamento de Atenção Compartilhada refere-se a momentos em que dois indivíduos estão focados no mesmo alvo e ambos estão cientes de que o foco de atenção é compartilhado; ou seja, refere-se à atenção conjunta dos indivíduos coordenando o envolvimento mútuo com seu foco mútuo em uma terceira entidade (CARPENTER; NAGELL; TOMASELLO, 1998). AC tem sido 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). Há indícios de correlação entre a ausência de AC com prejuízos sociocomunicativos e atrasos de linguagem (BILLECI *et al.*, 2016). Os autores de (BILLECI *et al.*, 2016) sugerem também que pode haver dificuldades na alternância de estímulo e no desengajamento do olhar.

Avaliações de ET da AC podem fornecer informações sobre o espaço observado concomitantemente com o tempo de resposta de indivíduos (HOFSTEN *et al.*, 2009), podendo ser analisado o comportamento a respeito de mudanças no monitoramento

do ambiente, presentes precocemente no TEA. As crianças com DT, por exemplo, ao observarem uma pessoa pegar um copo, direcionam o olhar para a região da boca e da mão do sujeito que está alcançando o copo. Este acompanhamento da trajetória fornece pistas que a criança tem um entendimento das intenções motoras.

Krogh-Jespersen *et al.* (2018) obtiveram dados que mostravam uma tendência de predições baseadas em localização e não em intenção nas crianças com TEA. O desengajamento do olhar estava relacionado mais com pistas que o movimento humano pode fornecer do que com a compreensão da intenção do outro; quando isso não era claro o sujeito demorava mais para desengajar. Essas observações sobre o comportamento de AC, torna este comportamento passível de análise por meio de sinais de rastreamento de olhar.

2.2 Rastreamento do olhar

A técnica mais comumente utilizada para a obtenção dos movimentos oculares é denominada de *eye tracking* (ET). Esse recurso possibilita medir e registrar a posição e o comportamento do movimento visual, de modo que seja possível verificar a área onde foi fixada a atenção e, ainda, a duração da exploração visual (BARRETO, 2012).

O dispositivo de rastreamento do olhar é constituído de uma câmera e um emissor de luz infravermelha (TOBII, 2019). O conceito principal é o uso da fonte de luz para iluminar os olhos e causar uma alta reflexão visível, de modo que a câmera capture as imagens dos olhos mostrando essas reflexões. Assim, a imagem capturada pela câmera é usada para identificar as reflexões da luz entre as regiões da córnea e a pupila, o que resulta em coordenadas informadas pelo dispositivo.

O estudo dos movimentos oculares tem base no conceito *strong eye-mind*, segundo o qual, aquilo que uma pessoa visualiza é assumido como indicador de pensamento (BARRETO, 2012; JUST; CARPENTER, 1976a; JUST; CARPENTER, 1976b). A tecnologia do ET tem sido empregada em variados propósitos. Assim, conforme Barreto (2012), é possível observar um grande potencial de aplicação em uma variedade de disciplinas e áreas de estudo, com forte predominância no campo da Ciência da Comunicação.

Na área da Saúde, o uso do dispositivo tem demonstrado grande apoio no auxílio de alguns transtornos como o TEA, pois com essa ferramenta é possível prover dados sobre as posições visuais, produzindo informações para serem posteriormente avaliadas.

Além disso, o ET é vantajoso em termos de custo e operacionalização, visto que não é uma ferramenta invasiva comparada a outros equipamentos de captura e análise, como o eletroencefalograma.

Em geral, os sinais coletados pelo rastreamento do olhar são obtidos usando-se estímulos aos quais os indivíduos são convidados a observar, tais como imagens ou vídeos. As principais informações obtidas do rastreamento do olhar são:

- **Fixação:** consiste das informações visuais dos eventos em que se mantém o olhar em uma posição por determinado tempo. Geralmente esse tempo é compreendido entre 225 a 250 milissegundos (SAEED, 2014a).
- **Sacadas:** são movimentos que ocorrem entre duas fixações, cerca de 3 a 4 vezes por segundo e apresentam duração de 30 a 80 milissegundos (BODALA *et al.*, 2014), (RICHARDSON; SPIVEY, 2008), (Larsson; Nyström; Stridh, 2013).
- **Número de piscadas:** são encontrados na literatura como *blink* e é expresso em piscadas por minuto (BPM). Sua taxa varia de 4.5 a 17 (BPM) (KUMAR *et al.*, 2016).
- **Perseguição suave:** trata-se de movimentos que são realizados para seguir um objeto que esteja também em movimento (LARSSON; NYSTROM; STRIDH, 2014). Segundo Saeed (2014b), *smooth pursuit* são movimentos mais lentos do que as sacadas e são usados para manter a fóvea (região central da retina do olho) alinhada com o estímulo móvel. De acordo com Kumar *et al.* (2016), estes movimentos são analisados em resposta aos estímulos visuais dinâmicos que se movimentam horizontalmente e verticalmente na tela do computador.

As informações citadas são fornecidas no formato de coordenadas cartesianas referentes à posição de localização do olhar em um determinado instante de tempo. Usualmente o *software* fornecido junto com os dispositivos de rastreamento fornecem tais coordenadas considerando uma periodicidade que varia de acordo com a frequência que o dispositivo é capaz de processar. Por meio desses dados é possível saber, por exemplo, se o indivíduo visitou (observou durante um intervalo curto de tempo) uma região de interesse ou se fixou o olhar (observou durante um intervalo maior de tempo) naquela região. No equipamento utilizado, é obtida uma medida a cada 3,3 milissegundos.

Geralmente, trabalhos que consideram o paradigma da atenção visual apresentam modelos computacionais empregando técnicas para geração de mapas de saliência. Segundo

Itti e Koch (2001), o mapa de saliência é um mapa escalar bidimensional que representa topograficamente a saliência visual, independentemente da dimensão do recurso que torna a localização saliente. Em geral é representado por uma imagem em que as regiões em que o olhar for mais direcionado ficam destacadas, conforme pode ser mostrado na figura 1.

Figura 1 – Exemplo de mapa de saliência: à esquerda, o mapa de saliência gerado e à direita, os *pixels* mais fixados (em cor vermelho).

Fonte – Jéssica dos Santos de Oliveira, 2017

3 Revisão Bibliográfica

Este capítulo apresenta a Revisão Sistemática (RS) conduzida no presente projeto, que tem como objetivo analisar os trabalhos contendo algoritmos para classificar, visualizar, extrair dados ou mensurar os resultados referentes a sinais do rastreamento de olhar, cujos estímulos são imagens dinâmicas.

3.1 Método de pesquisa

O processo da RS utilizado foi constituído de três etapas: Planejamento, com a elaboração do protocolo identificado na pesquisa; Condução, durante a qual foi conduzida a seleção e a avaliação dos estudos primários; e Análise dos estudos, fase responsável por extrair e compilar dados relevantes para a análise a fim de atingir o objetivo da RS.

O planejamento foi iniciado a partir de uma análise exploratória para selecionar estudos primários que contemplavam o uso de rastreamento de olhar usando vídeos (ou imagens dinâmicas) como estímulo. Foi inicialmente pesquisado o termo *eye tracking* e algumas variações do termo, como *eye-tracker*, *eye gaze*, *eye-controlled* e *eye movements*. As bases utilizadas nessa pesquisa foram IEEE (*Institute of Electrical and Electronics Engineers*) ([IEEE, 2019](#)), ACM (*Association for Computing Machinery*) ([ACM, 2019](#)), PubMed ([PUBMED, 2019](#)) e Scielo (*Scientific Electronic Library Online*) ([SCIELO, 2019](#)).

A partir da análise desses trabalhos, foi construída a formação da *string* de busca: (“eye gaze” OR “eye movements” OR “eye-controlled” OR “eye tracker” OR “eye tracking”) AND (analysis OR “features extraction”) AND (stimul*). Algumas alterações na *string* foram necessárias para adaptar a busca às bases selecionadas, conforme podem ser visualizadas na tabela 1.

Tabela 1 – Retorno dos resultados das máquinas de busca

Biblioteca	<i>String</i>	Filtros	Qtde. retornada
ACM	<i>(“eye gaze” OR “eye movements” OR “eye-controlled” OR (eye AND track*)) AND (analysis OR “features extraction”) AND (stimul*)</i>	Busca padrão/Full-text	46
IEEE	<i>(“eye gaze” OR “eye movements” OR “eye-controlled” OR (eye AND track*)) AND (analysis OR “features extraction”) AND (stimul*)</i>	Busca padrão/Metadados	196
Pubmed	<i>(“eye gaze” OR “eye movements” OR “eye-controlled” OR (eye AND track*)) AND (analysis OR “features extraction”) AND (stimul*)</i>	Busca padrão + filtros: 5 anos; espécies humanas; artigos em inglês e português.	481
SciELO	<i>(“eye gaze” OR “eye movements” OR “eye-controlled” OR (eye AND track*)) AND (analysis OR “features extraction”)</i>	Busca padrão + filtros: artigos em inglês e português.	32

Fonte – Thiago Victor Cardoso, 2020

Os critérios de inclusão definidos no protocolo foram:

- trabalhos que apresentam métodos, técnicas e algoritmos para processar sinais de rastreamento de olhar estimulados por imagens;
- trabalhos que citam características extraídas do rastreamento de olhar ou que façam a análise dessas características.

Os critérios de exclusão definidos no protocolo foram:

- trabalhos que não apresentam abstract/resumo;
- trabalhos que só apresentam questões de hardware;
- trabalhos que utilizam outras formas de monitoramento, além do rastreamento do olhar;
- trabalhos que não façam a análise do rastreamento do olhar;
- trabalhos que não utilizam imagem dinâmica como estímulo.

A figura 2 resume os passos que foram executados durante a revisão. Foram retornados 755 trabalhos, por meio das 4 bases pesquisadas. Dos trabalhos retornados, 131 foram aceitos na primeira fase, pois atenderam a pelo menos um dos critérios de inclusão. Foram excluídos 624 trabalhos devido a se encaixarem em pelo menos a um critério de exclusão. Após a fase de seleção, 131 artigos foram analisados e 52 estudos foram aceitos para análise final por corresponderem ao contexto desta revisão sistemática e apresentarem informações minimamente suficientes para esta pesquisa.

Figura 2 – Resumo do processo realizado na RS

Fonte – Thiago Victor Cardoso, 2020

3.2 Visão geral dos estudos

A atual revisão sistemática destaca uma concentração de trabalhos nas últimas décadas (figura 3), especialmente nos últimos nove anos, quando houve uma expansão de artigos publicados e analisados, envolvendo algoritmos para classificar e visualizar (seção 3.3) dados provenientes de dispositivos (seção 3.7) para rastreamento de olhar em determinado setor (seção 3.5).

Na figura 3, é apresentada a quantidade de estudos publicados sobre o tema de 1947 a 2018, destacando a porção dos mesmos que foram aceitos para participarem dessa revisão. Com o aumento significativo de submissões a partir de 2006, o gráfico destaca um grande interesse no tema envolvendo dispositivos de *eye tracking*. De 2014 a 2017, houve um crescimento de quase oito vezes (480 trabalhos publicados), quando comparados com o quadriênio anterior (67 trabalhos publicados), indicando grande relevância na pesquisa

pela abordagem de algoritmos computacionais, estímulos e métricas atrelados aos sinais de rastreamento do olhar. Outros fatores podem estar diretamente ligados ao aumento expressivo dessas publicações, como a evolução do poder computacional que contribuiu principalmente para a adoção de novos métodos pertencentes ao campo de aprendizado de máquina, além de equipamentos de *eye tracking* com custos mais acessíveis e ferramentas *open source*, os quais refletiram em um número crescente de submissões.

Figura 3 – Quantidade de artigos publicados e aceitos nessa revisão por ano.

Fonte – Thiago Victor Cardoso, 2020

A tabela 2, apresenta os trabalhos incluídos na revisão. Os seguintes atributos são apresentados: dispositivos de ET, área de aplicação, algoritmo implementado, estímulo aplicado, sinais processados de rastreamento do olhar e finalidade da pesquisa. No geral, os trabalhos são categorizados em 3 finalidades: os que implementam uma forma de apenas visualizar os dados, propondo algoritmos para representar essas informações; os trabalhos que realizam uma análise dos sinais; e os trabalhos que estabelecem a classificação, por meio de algoritmos que fazem a predição para efetuar a distinção entre classes. Embora a maioria dos trabalhos que realizam apenas a visualização dos dados não mencionam o algoritmo implementado, alguns trabalhos adotam a presença de algoritmos proprietários ou tipicamente utilizados na área da aplicação.

Tabela 2 – Características extraídas dos artigos incluídos

#	Referência	Dispositivo	Área	Algoritmo	Estímulo	Sinais processados	Finalidade
1	Cercenelli <i>et al.</i> (2017)	Eyebrian	Saúde	<i>Algoritmo de reconhecimento de sacada</i>	Imagens	Fixação, sacada	V
2	Baranes, Oudeyer e Götzelieb (2015)	EyelinK 1000	Saúde	<i>RF</i>	Imagens	Fixação, sacada	C
3	Wang, Yuval-Greenberg e Heeger (2016)	EyelinK 1000	Saúde	Detecção de sacada	Vídeos	Fixação, sacada	C
4	Ohl e Rolfs (2017)	EyelinK 1000	Saúde	-	Imagens	Fixação, sacada	A
5	Otero-Millan <i>et al.</i> (2014)	EyeLink II, EyeLink 1000	Saúde	<i>K-means</i>	-	Fixação, sacada	C
6	Ohl e Kliegl (2016)	EyeLink II	Saúde	Modelo de riscos aditivos de Aalen	-	Fixação, sacada	A
7	Chakrabarti <i>et al.</i> (2017)	EyelinK II, SR Research	Saúde	-	Imagens	Sacada	A
8	Röhrbein <i>et al.</i> (2015)	Eyetracker	Saúde	-	Imagens	Fixação, sacada	C
9	Kumar <i>et al.</i> (2016)	EyeTribes	Saúde	SmartEye	Imagens	Fixação, smooth pursuit, <i>blink</i> , <i>vergence</i> movimento vestibulo ocular.	A
10	Dumitriu <i>et al.</i> (2018)	EyeTribes 101	Saúde	<i>AB, KNN, LDA e SVM</i>	-	-	C
11	Thibodeaux <i>et al.</i> (2014)	ISCAN infrared video tracking	Saúde	-	-	Sacada	A
12	Saavedra-Peña <i>et al.</i> (2018)	iTracker	Saúde	<i>CNN</i>	-	Sacada	C
13	Jansson e Medvedev (2013)	Smart Eye AB	Saúde	Função Gaussiana e estimativa de densidade de kernel (KDE)	Vídeos	Sacada, <i>smooth pursuit</i>	C
14	Wang <i>et al.</i> (2015)	-	Saúde	<i>SVM</i>	Imagens	Fixação, sacada	C
15	Jansson, Rosén e Medvedev (2013)	Smart Eye AB	Saúde	Função Gaussiana e estimativa de densidade de kernel (KDE)	Imagens	Sacada, <i>smooth pursuit</i>	C

Tabela 2 – *Continuação: Características extraídas dos artigos incluídos*

#	Referência	Dispositivo	Área	Técnica	Estímulo	Sinais processados	Finalidade
16	Hayes e Henderson (2018)	SR Re-search EyeLink 1000 plus	Saúde	-	Imagens	Fixação, sacada	A
17	Orsati e Fernanda Tebexreni (2009)	Tobii 1750	Saúde	-	Imagens	Fixação	A
18	Wieckowski e White (2017)	Tobii T60 XL	Saúde	-	Vídeos	Fixação	A
19	Lönnqvist et al. (2017)	Tobii TX300	Saúde	-	-	-	A
20	Hosseinkhani e Joslin (2018)	SMI iView 120Hz	Saúde	-	Imagens	-	A
21	Bodala et al. (2014)	EyelinK 1000	Segurança	Detecção de velocidade da sacada de pico	Imagens	Fixação, sacada, pupila, <i>blink</i>	C
22	Deng et al. (2016)	EyelinK 2000	Segurança	<i>Vanishing Point Feature and Detection Algorithm</i>	Imagens	Fixação	A
23	Josephson e Holmes (2010)	ISCAN RK-426PC	Segurança	Distância <i>Levenshtein</i>	Imagens	Fixação	A
24	Rigas, Economou e Fotopoulos (2012)	Ober2 250Hz	Segurança	KNN	Imagens	Fixação, sacada	C
25	Srivastava et al. (2015)	Ober2 eye tracker	Segurança	SVM	Imagens	Fixação, sacada	C
26	Saeed (2014a)	SMI iView X HiSpeed 1250 eye tracker	Segurança	Árvores de regressão	Vídeos	Sacada, <i>smooth pursuit</i> , <i>vergence</i> , movimento vestibulo ocular	C
27	Darwish e Pasquier (2013)	Tobii T120	Segurança	RF	Imagens	Fixação, sacada	C
28	Bodala et al. (2017)	Tobii TX300	Segurança	I-VT	Imagens	Fixação, sacada	C
29	Venugopal, Amudha e Jyotsna (2016)	EyeTribe	Comércio	-	Imagens	Fixação, sacada, pupila, <i>smooth pursuit</i> , <i>scanpath</i>	A

Tabela 2 – *Continuação: Características extraídas dos artigos incluídos*

#	Referência	Dispositivo	Área	Técnica	Estímulo	Sinais processados	Finalidade
30	Blascheck <i>et al.</i> (2016)	(SMI) head-mounted Eye Tracing Glasses (ETG) 2.0.	Comércio	-	Vídeos	Fixação, <i>scanpath</i>	V
31	Shahimin <i>et al.</i> (2014)	Tobii TX300	Comércio	-	Imagens	Fixação	V
32	Zhang <i>et al.</i> (2018)	Tobii EyeX e EyeTribe	Cultura	The Kernelized Correlation Filter (KFC)	Vídeos	Fixação.	V
33	Pflüger <i>et al.</i> (2015)	Tobii T60 XL	Cultura	AB	Imagens	Fixação	C
34	Goschy <i>et al.</i> (2014)	Eyelink 1000	Outras	Algoritmo para detecção de sacada (EyeLink)	Imagens	Fixação, sacada	A
35	Anderson <i>et al.</i> (2015)	Eyelink 1000	Outras	-	Imagens	Fixação, sacada, <i>scanpath</i>	A
36	Liang, Liang e Sun (2016)	Eyelink 1000	Outras	<i>Graph Based Visual Saliency (GBVS)</i>	Vídeos	Fixação	V
37	Marighetto <i>et al.</i> (2017)	Eyelink 1000	Outras	-	Vídeos e sons	Fixação	V
38	Coutrot, Hsiao e Chan (2018)	Eyelink 1000	Outras	HMM	Imagens e vídeos	Fixação, sacada, <i>blink</i> , <i>smooth pursuit</i> , pupila	C
39	Choi, Cormack e Bovik (2013)	Eyetracker	Outras	-	Vídeos	-	V
40	Alam e Jianu (2017)	SMI RED-120Hz	Outras		Vídeos	Fixação	V
41	Sogo (2017)	OptiTrack V100R2	Outras	-	-	-	V
42	Lee Barry B. (2011)	Infrared camera	Outras	SVR	Vídeos	Fixação, sacada, <i>blink</i>	A
43	Borji <i>et al.</i> (2013)	-	Outras	SVM	Imagens	Fixação, sacada	C
44	Li, Bai e Zhao (2017)	-	Outras	-	Vídeos	Fixação, sacada, <i>scanpath</i>	V
45	Larsson, Nyström e Stridh (2014)	SensoMotoric Instruments	Outras	I-VDT	Vídeos	Fixação, <i>smooth pursuit</i>	C

Tabela 2 – *Continuação: Características extraídas dos artigos incluídos*

#	Referência	Dispositivo	Área	Técnica	Estímulo	Sinais processados	Finalidade
46	R-Tavakoli et al. (2015)	ASL eye tracker	Outras	SVM	-	Fixação, sacada	C
47	Hou et al. (2016)	Tobii T120	Outras	SVM	Vídeos	-	C
48	Kurzials et al. (2014)	Tobii T60 XL	Outras	-	Vídeos	Fixação	V
49	Kurzials, Heimerl e Weiskopf (2014)	Tobii T60 XL	Outras	Distância <i>Levenshtein</i>	Vídeos	Fixation, Scanpath	V
50	Kurzials et al. (2016b)	Tobii T60 XL	Outras	Distância <i>Levenshtein</i>	Imagens, vídeos	Fixação	V
51	Kurzials et al. (2016a)	Tobii T60 XL	Outras	-	Vídeos	Fixação, <i>Scanpath</i>	V
52	Haji-Abolhassani e Clark (2014)	Tobii X120	Outras	HMM	Imagens	Fixação, sacada	C

Legenda: AB - *Ada Boost*; KNN - *K-nearest neighbors*; RF - *Random Forest*; LDA - *Linear Discriminant Analysis*; SVM - *Support Vector Machine*; SVR - *Support Vector Regression*; CNN - *Convolutional Neural Networks*; HMM - *Hidden Markov Model*; I-VT - *identification velocity-threshold*; I-VDT - *velocity and dispersion threshold identification*; C - classificação; V - visualização; A - análise.

3.3 Fases gerais das abordagens

Conforme apresentado na figura 4, em geral, para construir um sistema para analisar sinais de rastreamento são necessárias quatro fases de desenvolvimento: coleta de dados, pré-processamento, extração de dados e classificação. Todos esses passos estabelecem condições relevantes para a acurácia dos resultados. O pré-processamento, por exemplo, pode aumentar significativamente a qualidade da classificação ([LARSSON; NYSTROM; STRIDH, 2014](#); [DARWISH; PASQUIER, 2013](#); [SAEED, 2014a](#)). Dos trabalhos abordados, os que têm como finalidade a classificação e representação gráfica dos dados, comumente, perpassam por essas etapas.

Figura 4 – Fases gerais das abordagens.

Fonte – Thiago Victor Cardoso, 2020

3.3.1 Coleta de dados

A fase de coleta de dados é destinada a capturar os dados de rastreamento do olhar por meio de *webcam* ou equipamento específico visando a obter uma representação das atividades de movimento ocular. Comumente, os dados recebidos desses movimentos são as coordenadas cartesianas obtidas de um determinado usuário que focalizou um estímulo. Durante o processo de captura, alguns dispositivos inserem dados ‘brutos’ (*raw data*), que são sinais recebidos diretamente do equipamento, sem passar por conversões de seus valores (ZHANG *et al.*, 2018).

Geralmente, os trabalhos, que utilizam ferramentas de rastreamento do olhar para a coleta de dados, adotam os fabricantes como o *Tobii* e *EyeLink*. Tais dispositivos são amplamente utilizados devido a sua boa precisão quanto aos dados capturados. Estes dispositivos serão explanados na seção 3.7.

3.3.2 Pré-Processamento

A figura 5 apresenta o percentual de trabalhos que realizam técnicas de pré-processamento. A fase de pré-processamento é encarregada de tratar os dados de en-

Figura 5 – Porcentual dos artigos que utilizam técnicas de pré-processamento. Legenda: I-VT = *Identification Velocity-Threshold*

Fonte – Thiago Victor Cardoso, 2020

trada relacionados aos estímulos utilizados durante a captura, usualmente imagens, como aplicações de filtros, realces, redimensionamentos, entre outros.

Os principais filtros empregados são: o filtro passa baixa, correspondendo a quatro artigos (13%); o filtro Gaussiano, tendo 10 trabalhos identificados (32%); *identification velocity-threshold* (I-VT), representado por 4 trabalhos (13%). Além disso, houve uma quantidade de artigos que implementaram algoritmos diversos, totalizando 13 artigos ou 42% desses trabalhos.

Como exemplo, em [Rigas, Economou e Fotopoulos \(2012\)](#), foi utilizado um processo de segmentação em cada exemplo de teste do modelo. [Blascheck et al. \(2016\)](#) usam opções de filtros definindo um limiar para reduzir ruídos visuais.

Com relação às técnicas de filtros, aplicadas aos sinais provenientes do rastreamento do olhar, em [Kurzahls et al. \(2016b\)](#), por exemplo, são aplicados filtros de fixação para melhorar a identificação do contexto local dos movimentos de micro sacadas. O trabalho de [Pflüger et al. \(2015\)](#) utiliza o filtro Gaussiano para distribuir a probabilidade das posições de saliências causadas pela fixação. De modo similar, o estudo de [Liang, Liang e Sun \(2016\)](#) aplica o filtro Gaussiano para converter mapas de saliências em imagens binárias segundo um limiar definido. O estudo de [Kurzahls et al. \(2016a\)](#) propõe usar uma técnica de visualização com um filtro interativo que permite a identificação das sequências de fixações de interesse. Em [Röhrbein et al. \(2015\)](#), são aplicados o filtro passa baixa Gaussiano, filtro de média circular e ruído branco Gaussiano aditivo para distorcer sistematicamente cenas

de paisagens naturais e feitas pelo homem. Em [Cercenelli et al. \(2017\)](#) é empregado um filtro de mediana de ordem selecionável, usado para remover picos indesejados nos sinais (função de filtro automático); uma função para reconhecer e remover automaticamente os artefatos que piscam usando o *Pupil Ratio Eye Data* (Remoção do *Blink*); uma janela temporal após cada apresentação de estímulo, aplicada para isolar as seções do arquivo de dados onde uma sacada é esperada (*Stimulus Trim*); e uma função para selecionar e remover manualmente seções do sinal (limpeza manual).

A técnica de *I-VT* trata-se de um recurso disponível nos pacotes de *software Tobii Studio*, como enunciado por [Bodala et al. \(2017\)](#).

Em [Darwish e Pasquier \(2013\)](#) é adotado o algoritmo para identificar sacadas e fixações, por meio de sinais filtrados baseados na velocidade da direção no deslocamento do olhar. Em [Haji-Abolhassani e Clark \(2014\)](#) são removidos os *blinks* e *outliers* dos dados de rastreamento do olhar, para serem classificados em sacadas ou fixações. O estudo de [Larsson, Nystrom e Stridh \(2014\)](#) emprega a mesma abordagem sobre um banco de dados, resultando em coordenadas cartesianas de fixação ou sacada ([SAEED, 2014a](#)). Já o trabalho de [Liang, Liang e Sun \(2016\)](#), aplica o pré-processamento nos mapas de saliências removendo informações redundantes e ambíguas.

3.3.3 Extração de características

Esta fase consiste em executar algoritmos capazes de interpretar os sinais do rastreamento de olhar e transformá-los em características (em geral, números) que possam ser processadas nas fases posteriores. Além da própria extração de características, também são contempladas atividades para transformar, simplificar ou reduzir o conjunto de características obtido.

Em vários trabalhos o conjunto de dados são transformados e reduzidos para melhor representatividade da informação. A extração de características é também considerada muito significativa ao determinar o poder de discriminação dos atributos, impactando diretamente no desempenho dos resultados quando a implementação tem natureza de classificação.

Em [Lee et al. \(2011\)](#) é adicionada uma fase posterior à extração para executar fusão de características, durante a qual é analisado um mapa resultante que representa

a distribuição de saliência em determinada imagem, formando uma fusão não linear que captura as regiões salientes. Já no estudo de [Saeed \(2014a\)](#), os dados de coordenadas de fixação e sacada são submetidos ao método *k-means* para redução de grande quantidade de informações redundantes. [Darwish e Pasquier \(2013\)](#) nomeiam essa etapa de “Método multimodal de características”, associando os movimentos de olhar com os traços estáticos, como por exemplo, o reconhecimento da íris e a distância entre os olhos ([DARWISH; PASQUIER, 2013](#)).

Alguns trabalhos integram o procedimento de redução de dimensionalidade para discriminar os atributos mais relevantes, de maneira que correlacionam e desprezam as características que apresentam menor variabilidade dos dados. Assim, a utilização dessa técnica possibilita melhorar o desempenho do classificador. Para a realização dessa tarefa, o algoritmo mais empregado é o *Principal Component Analysis (PCA)*.

Em [Larsson, Nystrom e Stridh \(2014\)](#), o PCA é utilizado para medir a dispersão do segmento, pela determinação da sua direção e maior variância. No trabalho de [R-Tavakoli et al. \(2015\)](#), além da utilização do PCA foi utilizado um conjunto de técnicas de seleção/fusão de características, sub-selecionando componentes com mais de 90% de contribuição para a variância. O PCA também foi implementado em [Coutrot, Hsiao e Chan \(2018\)](#) para minimizar problemas quando os vetores de características presentes na abordagem do trabalho são preenchidos com muitos zeros, ou quando o número de características excedem o número de observações. Ainda o estudo de [Hayes e Henderson \(2018\)](#) utiliza PCA para reduzir o *successor representation (SR)*. O SR, é uma matriz de tamanho fixo utilizada para capturar regularidades estatísticas em padrões de varredura, no tratamento clínico das cenas de cada participante.

3.3.4 Classificação

Classificação é a tarefa que consiste em prever se um conjunto de dados pertence a determinada classe. Para que o processo de aprendizado ocorra, é necessário que se aplique algum algoritmo capaz de separar esses dados e fornecer condições suficientes para analisá-los e classificá-los. Assim, esses algoritmos usualmente estão inseridos na área de Aprendizado de Máquina.

A tabela 2 apresenta os algoritmos de classificação em estudos analisados na literatura. Em geral, os trabalhos empregaram técnicas de Aprendizado de Máquina, como *AdaBoost (AB)*, *K-Nearest Neighbors (KNN)*, *Linear Discriminant Analysis (LDA)*, *Support Vector Machine (SVM)* e *Hidden Markov model (HMM)*, conforme detalhado a seguir.

Aprendizado de Máquina

Nessa revisão foram identificados trabalhos que utilizam aprendizado supervisionado e não-supervisionado. No aprendizado supervisionado necessita-se de dados rotulados com a classe conhecida para treinamento e testes, ao contrário do aprendizado não supervisionado.

Os estudos que apresentam métodos de aprendizado supervisionado são compostos pelos seguintes algoritmos: *Random Forest (RF)*, como em Baranes, Oudeyer e Gottlieb (2015), que implementa o RF, usando 250 árvores randômicas, 8 características e validação cruzada 10X. Este trabalho foi conduzido pela análise de mineração separadamente para prever escalas de curiosidade, confiança e surpresa. Também, no trabalho de Darwish e Pasquier (2013), o RF é utilizado para investigar as características do rastreamento do olhar por biometria a fim de testar vários tipos de estímulos e compará-los com a sua viabilidade na identificação biométrica.

Em outros estudos Haji-Abolhassani e Clark (2014), Coutrot, Hsiao e Chan (2018), foi utilizado o modelo oculto de *Markov*, do inglês *Hidden Markov model (HMM)*, como ferramenta para analisar as trajetórias do movimento do olhar.

Em Haji-Abolhassani e Clark (2014), o HMM foi estabelecido no desenvolvimento do modelo de atenção visual em trajetórias oculares reais dependentes da tarefa, durante a visualização de imagens naturais. Já Coutrot, Hsiao e Chan (2018) fornecem uma solução para a modelagem e classificação do olhar, além de um conjunto de ferramentas associadas ao Matlab. Essa abordagem é baseada nos modelos de *Markov*, que integra influências provenientes de mecanismos *top-down* e *bottom-up* e vieses de visualização em um único modelo.

Alguns estudos consideraram utilizar o algoritmo de Máquina de Vetores de Suporte (*Support Vector Machine - SVM*). Em R-Tavakoli *et al.* (2015) foi utilizado SVM na classificação das imagens em três classes: desagradável, neutro e agradável, a partir das

características extraídas pelos dados de ET. As características que o autor emprega são essencialmente a fixação e sacada, para reconhecer a essência emocional de uma cena, como a sua valência. Também foram aplicadas três variações do kernel referente ao SVM: linear, polinomial e função base radial (RBF). Em [Hou et al. \(2016\)](#), o SVM é estabelecido para classificar as emoções em filmes. O método classifica os *trailers* em quatro escolhas: “bom”, “ruim”, “gostou” e “não gostou”. Foram utilizados 725 *trailers*, separados em partes de treinamento e testes. O trabalho de [Srivastava et al. \(2015\)](#) propõe determinar como os usuários podem ser identificados com base em suas características de movimento do olhar, de modo a prever rótulos de usuários desconhecidos em *datasets*, pelo uso de um modelo de classificação adequado. O trabalho de [Lee Barry B. \(2011\)](#) adotou o algoritmo *Support Vector Regression (SVR)* para o processo de treinamento, referente à etapa de predição da atenção visual de sinais de vídeos, pelo uso de características de alto nível e baixo nível.

Também em [Wang et al. \(2015\)](#), o SVM foi utilizado para os dados de *eye tracking*. Nesse trabalho foram construídas três camadas de arquitetura, incluindo as características de *pixel*, objeto e estágios semânticos.

SVM também foi utilizado como classificador no trabalho de [Borji et al. \(2013\)](#), que adotou multi-classes SVM usando a classificação binária *C-Support Vector Classification (C-SVC)*, com o kernel RBF, na predição do modelo de saliência visual. O C-SVC, trata-se de uma variação do SVM, ferramenta tipicamente utilizada para resolver problemas de classificação.

Com relação aos métodos de aprendizado não supervisionados, o estudo de [Otero-Millan et al. \(2014\)](#) abordou um método para detectar pequenas sacadas no movimento ocular. Este modelo não requer configuração específica de limiar e conta com técnicas de clusterização no método não supervisionado implementado.

Alguns trabalhos abordaram algoritmos específicos proprietários. Em [Larsson, Nystrom e Stridh \(2014\)](#), foi adotado o *Velocity and Dispersion Threshold Identification (I-VDT)*, que aplica a combinação de velocidade e dispersão de sacadas, e que foi capaz de detectar 94.3% dos estímulos de fixações das imagens. Em [Cercenelli et al. \(2017\)](#), foi implementado um algoritmo para reconhecimento das sacadas. Já em [Borji et al. \(2013\)](#), foi realizada uma construção para avaliar os *scanpaths*. Segundo [Coutrot, Hsiao e Chan \(2018\)](#), *scanpaths* são sequências de ET, como sacadas, *smooth pursuits* e fixações.

[Bodala et al. \(2014\)](#) utilizaram o *Source invariant feature transform (SIFT)*, que identifica a condução do veículo em aproximação do tráfego em cada *frame*. O *SIFT*

compreende um detector e descritor de características. [Liang, Liang e Sun \(2016\)](#) empregaram o *Graph Based Visual Saliency (GBVS)* encarregado em prever a área de saliência de forma precisa e eficiente na maioria dos quadros de vídeos. O algoritmo representa graficamente o mapa de saliência predito. Nesse trabalho, o GBVS é comparado com o *Adaptive Whitening Saliency (AWS)*, que tem a capacidade de identificar a área saliente.

[Kurzahls et al. \(2016b\)](#) fizeram uso do *Levenshtein*, que mensura a sequência da dissimilaridade, também executado como função de similaridade em [Kurzahls, Heimerl e Weiskopf \(2014\)](#). No estudo de [Ohl e Kliegl \(2016\)](#) é proposto o modelo de riscos aditivos de Aalen para delinear o curso do tempo das principais influências na taxa de sacada secundária. Este modelo possibilita estimar uma taxa de risco do *baseline* dependente das variáveis incluídas no modelo.

3.4 Métricas de avaliação

Um total de 23 artigos (44% dos trabalhos incluídos) apresentou algum tipo de métrica empregada para avaliar os resultados obtidos com o processamento do rastreamento do olhar. Conforme pode ser visto na figura 6, estes 23 estudos que implementam métricas são divididos em cinco categorias, os que apresentam a curva ROC (total de 7 trabalhos, correspondendo a 21%); os que utilizam a *Area Under Curve (AUC)* (6 estudos, resultando em 18%); os que utilizam o Coeficiente de Correlação (CC) (5 trabalhos, 15%); os trabalhos que aplicam a *Normalized Scanpath Saliency (NSS)* (5 trabalhos, 15%); e as demais medidas, que foram empregadas em 10 trabalhos, resultando em 31%.

As principais métricas encontradas nos estudos analisados, para a avaliação do desempenho do classificador são: *Receiver Operating Characteristics (ROC)* e *Area Under The Curve (AUC)*. Segundo [Fawcett \(2006\)](#), a curva ROC é uma técnica para visualizar, organizar e selecionar classificadores com base em seu desempenho. Na curva ROC, os valores de sensibilidade (*true-positive rate - TPR*) são representados no eixo das ordenadas, e no eixo das abscissas, são representados os valores do complemento da especificidade (*false-positive rate - FPR*), ou seja, (1 - especificidade). O traçado da curva ROC é realizado variando-se um limiar que define, para cada nível de TPR, o nível complementar de FPR. A TPR e FPR são definidas na equação 1 e na equação 2.

Figura 6 – Métricas utilizadas nos artigos. Legenda: (ROC = *Receiver Operating Characteristic*, CC = Coeficiente de Correlação, NSS = *Normalized Scanpath Saliency*, AUC = *Area Under the ROC Curve*).

Fonte – Thiago Victor Cardoso, 2020

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

$$FPR = \frac{FP}{TN + FP} \quad (2)$$

Dessa forma, a curva ROC é uma representação de duas dimensões associadas ao desempenho do classificador, de modo que esses classificadores podem ser comparados e reduzidos a um único valor escalar, que representa o desempenho esperado. Geralmente a AUC é utilizada como método para a redução dessas informações. A AUC é a área que a curva ROC forma com o eixo das abcissas; é comumente utilizada como medida de desempenho, sendo que quanto mais próximo da unidade (valor 1), melhor é o desempenho do sistema sob avaliação.

Röhrbein *et al.* (2015) afirmam que essas métricas são as mais comuns para determinar quão bem um mapa de saliência calculado prevê um mapa de fixação humana real.

Em Lee *et al.* (2011), foi utilizada a curva ROC para avaliar o quão bem um mapa de saliência prevê a atenção visual. Em Otero-Millan *et al.* (2014), esta métrica foi adotada para avaliar o desempenho de diferentes métodos de detecção de micro sacadas.

Em Otero-Millan *et al.* (2014), a AUC apontou a medida de distinção de dois sinais, que está diretamente relacionada à sobreposição das duas distribuições da propriedade

comparada. O estudo de Wang *et al.* (2015) utiliza a curva ROC e a AUC para avaliar o modelo de saliência na predição de fixação entre os indivíduos de controle e com TEA.

Em Deng *et al.* (2016), além da curva ROC foram adotadas outras métricas, como a AUC e NSS. A métrica NSS trata-se de uma medida de correspondência entre o mapa de saliência e o *scanpath*. Também o trabalho de Borji *et al.* (2013) considera outras métricas além da AUC, como o CC e o NSS para avaliar a saliência dos movimentos de olhar presentes nos *datasets*. Em Liang, Liang e Sun (2016) são empregadas as mesmas métricas apontadas pelo trabalho anterior, na avaliação quantitativa do modelo de saliência, uma ferramenta de visualização possibilita ajudar os usuários na avaliação de diferentes desempenhos pertinentes aos algoritmos de predição de saliência.

De acordo com Coutrot, Hsiao e Chan (2018), diferentes métricas são mencionadas para comparar os mapas das posições do olhar, processados pelo modelo de saliência. As técnicas referentes à avaliação do olhar são categorizadas na distribuição espacial dos movimentos dos olhos, que incluem: *Kullback-Leibler divergence* (KLD), CC, similaridade (SIM), *earth moving distance* (EMD), AUC, NSS e porcentagem de fixação (PF).

3.5 Áreas de aplicação

Os artigos analisados contemplam quatro principais áreas de aplicação: saúde, segurança, comércio e cultura, conforme apresentado na tabela 2. De acordo com a figura 7, é apontada uma predominância na área da Saúde, com 20 trabalhos (37%); na área da Segurança, com oito trabalhos (15%). Os trabalhos contemplados na categoria Outras, se enquadram nas áreas da Tecnologia, Visão Computacional, Realidade Virtual e Ciência da Computação, resultando em 19 estudos e totalizando 38%.

Figura 7 – Porcentual dos trabalhos realizados por área de aplicação.

Fonte – Thiago Victor Cardoso, 2020

3.5.1 Classificação aplicada à área de saúde

A classificação de dados oriundos de rastreamento de olhar para diagnósticos foi a finalidade mais observada nos trabalhos aplicados à saúde. Em [Jansson, Rosén e Medvedev \(2013\)](#) é proposta uma abordagem para distinguir indivíduos saudáveis e indivíduos com anomalias com base em métodos estatísticos, de modo a encontrar conjunto de dados divergentes.

[Jansson e Medvedev \(2013\)](#) implementaram métodos para a identificação de um modelo do tipo *Wiener*, que tenta distinguir estatisticamente uma significância entre os parâmetros estimados, usando tal método para comparar as trajetórias de ET em pacientes que possuem *Parkinson*, com perfis de indivíduos saudáveis. Já no trabalho de [Orsati Fernanda Tebexreni \(2009\)](#), foi comparado o desempenho entre crianças com transtorno invasivo do desenvolvimento (TID) e com desenvolvimento normal na percepção de faces humanas.

O TEA é investigado em [Wieckowski e White \(2017\)](#), que apresenta um estudo para examinar o papel da atenção visual direcionada para aspectos não sociais da cena, o que reflete em um possível mecanismo subjacente ao reconhecimento e habilidade expressiva na deficiência do TEA. [Lönnqvist et al. \(2017\)](#) também descrevem a investigação de

características sutis da percepção social e interpretação de indivíduos de alto funcionamento assistindo a uma cena e interpretando-a.

O trabalho de [Otero-Millan et al. \(2014\)](#) aplica um novo método para detectar sacadas com a vantagem de não necessitar de configuração de limiar. Mesmo que os autores não indiquem diretamente a aplicação na identificação de transtornos específicos, este trabalho é indicado para pesquisas que envolvem a neurociência clínica. Outro exemplo de estudo no campo da psicologia foi o trabalho de [Dumitriu et al. \(2018\)](#), que aborda a classificação para avaliação do estado afetivo de diferentes cenários analisados.

[Hayes e Henderson \(2018\)](#) utilizaram a *Successor Representation Scanpath Analysis (SRSA)* para encontrar diferenças de padrões no comportamento do olhar entre indivíduos com o transtorno de *deficit* de atenção, distúrbios do espectro autista e de dislexia. [Thibodeaux et al. \(2014\)](#) conduziram um experimento de um sistema que compara os movimentos oculares, intervalos de sacadas e sobreposições de sacada, para entendimento dos movimentos oculomotores e sua relação com algumas doenças neurodegenerativas.

3.5.2 Identificação biométrica para finalidade de segurança

O principal uso de rastreamento de olhar para os trabalhos aplicados à área de segurança apresentou identificação biométrica como a principal finalidade. Foram incluídos cinco artigos nesta categoria. [Josephson e Holmes \(2010\)](#) analisam a semelhança de *scanpaths* de suspeitos em vídeos, definindo *scanpaths* como “sequências repetitivas de fixações e sacadas que ocorrem durante exposição e subsequentemente após reexposição a um estímulo visual, facilitando o reconhecimento”. Em [Venugopal, Amudha e Jyotsna \(2016\)](#), o *scanpath* é definido por uma sequência de fixações e sacadas que podem dar informações sobre o comportamento de pesquisa de um participante.

[Rigas, Economou e Fotopoulos \(2012\)](#) implementaram um sistema de identificação analisando o movimento do olhar, no qual cada fixação é projetada em um espaço multivariado. Em [Saeed \(2014a\)](#) também é apresentado um sistema de reconhecimento biométrico utilizando os dados de ET, extraídos durante a apresentação de uma cena aos participantes. O vetor de características foi constituído por fatores de fixações e sacadas.

O trabalho de [Darwish e Pasquier \(2013\)](#) visou a investigar as propriedades de movimento do olhar testando diversos tipos de estímulos como figuras contendo pontos

para interligar e imagens, e comparando sua viabilidade pela identificação biométrica em termos de taxa do erro alcançada pelo sistema. Em [Kumar et al. \(2016\)](#) é proposta uma ferramenta denominada SmartEye, que é baseada nas medidas quantitativas de fixação do olhar, *smooth pursuit*, e piscar de olhos em resposta aos estímulos visuais dinâmicos e estáticos. A ferramenta captura os índices relacionados ao olhar sincronizados com o tempo, utilizando-os para compreender esses estímulos com o objetivo de mapear os índices relacionados a uma provável disfunção neurológica.

3.5.3 Detecção de saliência

De acordo com [Liang, Liang e Sun \(2016\)](#), detecção de saliência é um mecanismo que visa a expressar a atenção visual em relação aos dados intuitivamente, tornando-a mais compreensível à ampla coleção de diferentes tarefas de análises. [Deng et al. \(2016\)](#) afirmam que a detecção de saliência é uma das funções fundamentais da visão humana, como também uma importante tarefa para aplicações de visão computacional. Embora o sistema visual humano para detectar saliência seja extremamente rápido e confiável, o modelo computacional ainda permanece um grande desafio.

Em geral, os trabalhos que têm a finalidade de detectar saliências usam estímulos estáticos ([DENG et al., 2016](#)). Usualmente o conjunto de fixações é agregado ao tempo de observação, mas podendo ser aplicado por estímulos dinâmicos sumarizando a distribuição de atenção.

Na área de segurança, [Deng et al. \(2016\)](#) construíram um *framework* combinando um modelo de detecção de saliência para verificar a atenção de motoristas em um ambiente de estrada.

Na área de saúde, [Wang et al. \(2015\)](#) aplicaram o algoritmo de classificação estabelecendo cinco fatores diferentes de estímulos a partir de imagens, no modelo de saliência, para identificar padrões em indivíduos com TEA.

Na área de comércio, os trabalhos apresentaram como finalidade a utilização da tecnologia para a obtenção dos movimentos do olhar, resultando na análise dos diversos padrões de clientes em tempo real, para a tomada de decisão ([VENUGOPAL; AMUDHA; JYOTSNA, 2016](#); [SHAHIMIN et al., 2014](#); [BLASCHECK et al., 2016](#)).

3.6 Visualização de dados

Nesta categoria de trabalhos, os estudos têm por finalidade a representação gráfica das informações provenientes dos movimentos oculares, em resposta ao estímulo disponibilizado.

Os trabalhos que têm a finalidade de visualização, em geral não apresentam algoritmos de AM por não implementarem modelos de classificação, ou por não utilizarem métodos tradicionalmente conhecidos na etapa de classificação. No trabalho de Wang, Yuval-Greenberg e Heeger (2016), foi estabelecido um algoritmo que compara a velocidade do rastreamento do olhar com um limiar, no qual o traçado da posição do olhar de cada bloco (depois da remoção do *blink*) é removido para a configuração do limiar de velocidade na detecção da sacada.

Em Ohl e Rolfs (2017) foi verificada a hipótese de que os movimentos de sacadas podem prover um gargalo natural para a transição do conteúdo frágil da memória sensorial para a memória visual de curto prazo. Li, Bai e Zhao (2017) estabelecem uma abordagem para estimular os movimentos de rastreamento do olhar em uma visão de imagens de 360°.

O estudo de Shahimin *et al.* (2014) aborda determinar os fatores que contribuem para selecionar produtos de óculos de sol em compras online. A ferramenta implementada possibilita rastrear onde um pessoa está olhando pelas fixações por meio dos dados de fixação gerados.

Em Zhang *et al.* (2018), é implementado um sistema que possibilita delinear manualmente e também provê um método automático para definir ROIs. As ROIs definidas e os dados de ET são representados graficamente simultaneamente. O objetivo proposto pelo trabalho foi de auxiliar o pesquisador a ampliar o conhecimento em palestras de vídeos.

Cercenelli *et al.* (2017) propõem uma ferramenta desenvolvida para análise dos movimentos de sacada ao usar um rastreador ocular comercial, não necessitando que os usuários tenham habilidades e conhecimento de programação. A ferramenta visa a simplificar o uso do sistema *ViewPoint ET* a fim de incentivar a introdução dessa tecnologia na prática clínica de oftalmologia. Os autores implementaram um software denominado *SacLab*, que integra quatro módulos de processamento, possibilitando ao usuário testar os estímulos visuais, registrar os movimentos de sacadas, analisar os dados registrados para a

extração automática dos parâmetros sacádicos de interesse clínico, e fazer uma análise agregada de múltiplos registros de movimentos.

O trabalho de Blascheck *et al.* (2016) também considera a implementação de técnicas de visualização para análise das sequências de fixação em diferentes estágios. As ferramentas propostas são: um visualizador de *Area of interest* (AOI) ou *Region of interest*, uma árvore de AOI, uma matriz de AOI e um gráfico de AOI, referente às sequências de fixação. Em Liang, Liang e Sun (2016) foi construído um sistema de análise visual incluindo visões múltiplas do espaço temporal.

O estudo de Bodala *et al.* (2014), investigou o aumento da carga de trabalho cognitivo de um indivíduo executando uma tarefa visual devido ao estímulo visual vago. O autor buscou avaliar a aplicação em vigilância e execução cognitiva.

Bodala *et al.* (2017) estudaram a possibilidade de desenvolvimento de uma nova visão computacional assistida pela análise do rastreamento do olhar e pelo uso de fixações.

3.7 Dispositivos

Os dispositivos para rastreamento do olhar são equipamentos que captam reflexos das posições dos olhos e armazenam essas informações para posteriormente enviá-las ao computador, de maneira que é possível saber quais foram as posições observadas ao longo do tempo. Existem vários fabricantes que fornecem a solução do *eye tracking*, com diversos modelos e para inúmeras finalidades. A tabela 2 aponta os dispositivos utilizados nos estudos incluídos.

Os estudos primários utilizaram em sua maioria os dispositivos *EyeLink 1000* e *Tobii T60XL*. Três estudos utilizaram o modelo *EyeLink II*. Alguns modelos do Tobii, como o *Tobii TX300*, foram utilizados em três trabalhos, o *Tobii T120* em dois trabalhos, e o *Tobii X120* em dois trabalhos.

Kumar *et al.* (2016) afirma que o modelo *EyeTribe* é um dos *eye trackers* comerciais de baixo custo, quando comparados a outras marcas e modelos, como *EyeLink* e *Tobii*, que é tecnicamente equivalente ao modelo *EyeLink 1000*. Alguns fabricantes, como o Tobii, oferecem captura com frequência média de três milissegundos por fixação. Enquanto que o *EyeLink 1000* e *EyeLink 1000 Plus*, respectivamente apresentam frequências de 3 ms e 2,2 ms. Quanto aos trabalhos analisados por esta RS, nenhum deles mencionou problemas

relacionados à resolução destes equipamentos, afetando diretamente nos resultados obtidos, o que indica, desta forma, que tais equipamentos foram adequados para a utilização nos experimentos.

Em [Lee Barry B. \(2011\)](#) foi utilizada uma câmera infravermelha com captura de 1000 *frames* por segundo, o que consequentemente aumenta a precisão em comparação a outros equipamentos referidos em trabalhos, todavia o autor não informa o modelo e o fabricante do equipamento.

Os custos dos dispositivos de rastreamento do olhar têm diminuído ao longo dos anos devido à própria comercialização de dispositivos com baixo custo, além da aderência de *eye trackers open-source* ([SOGO, 2017](#)), que resultam na possibilidade de um aumento nos estudos com a utilização dessa tecnologia.

3.8 Discussão

A utilização dos dispositivos de ET têm se mostrado importantes instrumentos para avaliar os sinais do olhar do comportamento humano, principalmente quando os estudos direcionam para a área da Saúde. Neste caso, adotar equipamentos não invasivos e que não causem riscos à saúde, constituem-se de um grande avanço científico. Também, o uso do ET tem sido amplamente utilizado nos últimos anos, indicando grande relevância de tal instrumento em diversas áreas e finalidades.

No geral, esta RS analisou trabalhos na área da Saúde, contemplando estudos para auxiliar no diagnóstico de transtornos como o TEA, e também de doenças de *Parkinson*; na área da Segurança, mencionando o uso do dispositivo como recurso biométrico na identificação de indivíduos; na área do Comércio, Cultura e outras áreas, propondo o uso da ferramenta de ET para atividades comerciais, de marketing, visualizar obras de artes, e na exposição de palestras em vídeos.

Com relação às técnicas de aprendizado de máquina, apresentaram resultados satisfatórios na classificação de dados de *eye tracking*. A tabela 2 indica os algoritmos utilizados e referidos na literatura. Dos 52 trabalhos analisados, 22 artigos utilizam algoritmos para a classificação dos dados. É possível observar uma predominância por algoritmos SVM (apresentados em 7 estudos).

As principais métricas observadas nos artigos foram AUC e curva ROC. Essas medidas condizem com a utilização de algoritmos que implementam aprendizado de máquinas na avaliação dos resultados da classificação. A curva ROC e AUC são importantes instrumentos quando se pretende comparar os resultados dos classificadores, por meio do limiar obtido. Apesar da métrica facilitar na escolha do limiar, de maneira gráfica, outras medidas também são de mesma importância, como a métrica de acurácia e F1-score, que são propostas por este presente trabalho.

Algumas limitações dos trabalhos em geral, estão associadas à alta quantidade e complexidade dos dados, elementos necessários ao processo de classificação e visualização dos dados que acarretam dificuldades na execução em tempo real. Embora alguns artigos mencionem o uso de tempo real em alguma parte do processo ou em trabalhos futuros, ainda a implementação desse recurso no monitoramento e predição permanece um grande desafio. Uma possibilidade para atenuar essa dificuldade é a presença de filtros, que extraem as características apenas mediante determinada condição e enquadramento, resultando na redução de ruídos, melhora na qualidade dos dados, e consequentemente em menor processamento. Alguns trabalhos relatam como pesquisas futuras o desenvolvimento de sistemas para dispositivos móveis. Mas, de igual modo, devido à grande densidade dos dados, a implementação autônoma em *smartphones* ou *tablets* impactaria no desempenho do sistema, tornando essa questão objeto de estudo e reforçando a adoção do filtro inteligente. Outra contribuição pode ser implementada com o emprego de algoritmos mais eficientes de Aprendizado de Máquina.

De modo geral, os trabalhos incluídos indicaram dificuldades referentes à quantidade dos dados de exemplos e número reduzido de participantes nos experimentos, exercendo influência no desempenho dos resultados, além do tempo de execução dos algoritmos devido à ampla quantidade de elementos dos sinais de rastreamento e imagens de estímulo. Um fator que carece de estudo seria um aprofundamento no processo de simplificar e diminuir os dados de entrada, focando na informação de maior relevância, de modo que os atributos utilizados no classificador já estejam previamente filtrados, diminuindo no tempo de resposta na tomada de decisão.

Além da ausência de dados, alguns trabalhos relatam problemas com relação ao próprio dispositivo de rastreamento do olhar ([WIECKOWSKI; WHITE, 2017](#); [LÖNNQVIST et al., 2017](#)). Adicionalmente, alguns trabalhos relatam fatores externos, como céu nublado, pedestres, acidentes de trânsito, entre outros, que podem afetar a atenção do motorista e

a saliência visual de trabalhos na área de segurança (DENG *et al.*, 2016), impactando na acurácia dos resultados (BODALA *et al.*, 2017).

Alguns estudos propõem aprofundar a investigação por padrões mais complexos de rastreamento do olhar (KUMAR *et al.*, 2016), além de combinar os registros do *eye tracking* abordando (HOSSEINKHANI; JOSLIN, 2018) as características sociais no entendimento de indivíduos de alto funcionamento com TEA (LÖNNQVIST *et al.*, 2017).

3.9 Conclusão

Os trabalhos incluídos indicaram que há necessidade de maior quantidade de instâncias na base de dados (BORJI *et al.*, 2013; SAEED, 2014a), analisar novos aspectos de características extraídas do vídeo (MARIGHETTO *et al.*, 2017; RIGAS; ECONOMOU; FOTOPOULOS, 2012), além de explorar novos padrões de rastreamento do olhar (DARWISH; PASQUIER, 2013; ALAM; JIANU, 2017). Além disso, foi indicada a necessidade de novos algoritmos mais eficientes e que combinam mais características para otimizar no desempenho do classificador (LARSSON; NYSTROM; STRIDH, 2014). Também a implementação de paralelismo foi indicada para melhorar o desempenho no contexto de aumento da velocidade de processamento (PFLÜGER *et al.*, 2015).

No geral, nota-se uma tendência para a utilização de algoritmos de Aprendizado de Máquina, associados com métodos de extração de características e dispositivos de rastreamento do olhar precisos para a tarefa de classificação e visualização dos dados. Os métodos apresentados na presente revisão podem ser adaptados para diversas aplicações e vários contextos, não se limitando a determinada área.

4 Metodologia

Este capítulo apresenta os procedimentos e etapas que integram o desenvolvimento desta pesquisa que, como já mencionado, visa propor um método computacional para classificar indivíduos como pertencentes ao grupo TEA ou ao grupo DT, a partir do processamento de características da trajetória percorrida pelo rastreamento do olhar. A seção 4.1 descreve os materiais utilizados para a preparação dos dados necessários às etapas subsequentes. Para a concepção deste trabalho houve necessidade de embasamento técnico, a partir da Revisão Sistemática (capítulo 3). Foram construídos métodos e técnicas para a implementação de uma ferramenta responsável por visualizar os movimentos de trajetória (seção 4.2). Por fim, foi implementado um modelo de classificação para auxílio ao diagnóstico de TEA (seção 4.3).

A abordagem desenvolvida foi construída a partir de dados de ET coletados de indivíduos TEA e DT durante uma tarefa de observação de vídeos especialmente construídos para analisar o paradigma de AC, conforme detalhado na seção 4.1. Uma técnica, aqui denominada de identificação de Regiões de Interesse (ROIs) flutuantes, foi definida para acompanhar a trajetória de objetos nos vídeos. Características foram extraídas dessas ROIs e dos sinais de ET, sendo processadas com um algoritmo de aprendizado de máquina, conforme apresentado na seção 4.3.

4.1 Materiais

4.1.1 Participantes

Participaram desta pesquisa 86 pacientes, sendo 30 indivíduos do grupo DT (10 do sexo feminino e 20 do sexo masculino, com idades variando de 3 a 15 anos e média de idade de 10 anos) e 56 indivíduos do grupo TEA (18 do sexo feminino e 38 do sexo masculino, com idades variando de 5 a 18 anos e média de idade de 10 anos).

Os indivíduos do grupo TEA foram avaliados por especialistas do IPq-USP, por meio do histórico de avaliação clínica, utilizando o critério DSM-V e a escala CARS, apresentando valores entre 33 e 45 pontos, com média de 34,5 pontos. Foi aplicado também CARS aos indivíduos com DT, apresentando valores inferiores a 30 pontos.

4.1.2 Estímulos

Foram utilizados três vídeos como estímulos (denominados vídeo A, vídeo B e vídeo C), com aproximadamente 10 segundos de duração cada, e resolução espacial de 1920 x 1080 *pixels* por *frame*.

Os vídeos foram divididos em nove segmentos, sendo que cada segmento considera o mesmo conjunto de ROIs. Neste trabalho, considera-se uma ROI não apenas uma área específica mas sim o objeto que ela representa, que pode mover-se ao longo do vídeo. Cada vez que há mudança neste conjunto de ROIs, isto é, cada vez que há adição ou eliminação de uma ROI que represente um objeto diferente, o vídeo é segmentado. No vídeo A (figura 8) apenas o olhar da terapeuta se movimenta. Por isso, há apenas uma ROI em todo o vídeo, resultando em um único segmento com 191 *frames*. No vídeo B, o olhar da terapeuta permanece estacionado no início. Em seguida, se movimenta em direção ao objeto de interesse, retornando ao repouso nos *frames* finais. Assim, em alguns *frames* apenas a região dos olhos da terapeuta constitui ROI. Em outros, há duas ROIs: olhos da terapeuta e objeto de interesse. Considerando a variação entre ROIs, o vídeo é subdividido em três segmentos com, respectivamente, 54, 115 e 90 *frames*. No vídeo C, a terapeuta movimenta o olhar e aponta para o objeto de interesse, adicionando uma terceira ROI que é o apontar da terapeuta ao objeto de interesse. Utilizando o mesmo conceito, este vídeo é subdividido em cinco segmentos com, respectivamente, 134, 2, 97, 3, 56 *frames*. Os vídeos foram processados *frame a frame* a fim de viabilizar a técnica de identificação de ROIs flutuantes (seção 4.3.1).

4.1.3 Protocolo

O equipamento Tobii Pro TX300 foi utilizado para a captura dos dados de ET. Este dispositivo é composto por unidade de rastreamento ocular, monitor removível de 23 polegadas e frequência máxima de captura de 300Hz. Possui a capacidade de capturar os movimentos da cabeça que apresentem velocidade máxima de 50cm/s.

Para a captura dos dados, os participantes do experimento foram orientados acerca dos procedimentos e convidados a ler e assinar o Termo de Consentimento Livre e Esclarecido aprovado pelo Comitê de Ética da Universidade de São Paulo (anexo B).

Figura 8 – Exemplos de vídeos utilizados como estímulo: no vídeo A, apenas o olhar da terapeuta se movimenta; no vídeo B, o olhar se movimenta em relação ao objeto de interesse; no vídeo C, a mão aponta para o objeto de interesse.

Fonte – Thiago Victor Cardoso, 2020

Para efetuar a captura dos dados, cada participante sentou-se em frente ao monitor do rastreamento ocular, com distância entre 50cm a 70cm, favorecendo a devida calibração do equipamento. A calibração consiste em projetar algumas imagens no monitor a fim de que o usuário siga esses objetos e o dispositivo configure a localização do olhar do indivíduo corretamente. Em seguida, foi iniciada a coleta, com acompanhamento do terapeuta e/ou do cuidador.

O presente trabalho utiliza parte dos dados de rastreamento do olhar e estímulos visuais em formato de vídeos, fornecido pela coleta em (OLIVEIRA, 2017).

Os dados obtidos na coleta são fornecidos pelo software que é integrado ao dispositivo Tobii, sendo mais de 60 atributos. O equipamento possibilita cálculos baseados nas fixações e visitas. A fixação indica quando os olhos permaneceram em uma ROI por determinado tempo, que geralmente é compreendido entre 225 a 250 milissegundos (SAEED, 2014a). Os atributos utilizados neste trabalho são descritos na tabela 3. A visita é definida pela ocorrência em que o indivíduo visualiza a ROI.

Tabela 3 – Características utilizadas provenientes do rastreamento de olhar.

Características	Descrição
RecordingTimestamp	Tempo registrado da ocorrência do evento. Segundo o fabricante, cada evento ocorrido pelo dispositivo tem cerca de três milissegundos.
GazePointX (MCSPx)	Posição da coordenada horizontal que o indivíduo olhou (visitou) na tela.
GazePointY (MCSPx)	Posição da coordenada vertical que o indivíduo olhou (visitou) na tela.
FixationPointX (MCSPx)	Posição da coordenada horizontal que o indivíduo fixou o olhar na tela
FixationPointY (MCSPx)	Posição da coordenada vertical que o indivíduo fixou o olhar na tela

Fonte – Thiago Victor Cardoso, 2020

4.2 Visualização de trajetórias

Como meio para compreender os dados presentes do rastreamento do olhar, atrelados ao estudo da trajetória sobre tais sinais, foi desenvolvida uma aplicação que estabelece uma forma de visualizar estes dados. Esta ferramenta foi implementada em C++ com a biblioteca Qt 5.13 (CHAMBE-ENG; NORD, 1994), que possibilita exibir o *frame*, os dados de ET relacionados ao *frame*, as ROIs definidas, além de cada condição lógica de determinada ROI (seção 4.3.1). Uma das vantagens desta ferramenta é a possibilidade de exibir os dados de rastreamento do olhar separadamente por indivíduo e por grupo TEA e DT, de modo a proporcionar uma análise mais detalhada do comportamento do olhar mediante as condições presentes nos estímulos e as influências das ROIs. Os elementos a seguir são abordados na ferramenta construída.

Coordenadas de ET. Esta ferramenta representa graficamente os movimentos de trajetória e consolida as coordenadas de rastreamento com os *frames* reproduzidos, exibindo graficamente as posições provenientes do rastreamento do olhar, conforme pode ser visto na figura 9. As coordenadas de rastreamento do olhar são representadas graficamente por uma cor para cada indivíduo. Indivíduos com TEA, são representados em tons de vermelho e indivíduos DT, em tons de azul. Cada registro capturado pelo dispositivo corresponde às coordenadas X, Y do *pixel* visitado (atributos *GazePointX [MCSPx]* e *GazePointY [MCSPx]*), em um instante de tempo (atributo *RecordingTimestamp*), em que cada círculo representa uma posição de coordenada X,Y. A cada coordenada, a área do

Figura 9 – Exemplo da ferramenta de visualização de trajetórias exibindo o rastreamento do olhar de 15 indivíduos DT (lado esquerdo) e TEA (lado direito).

Fonte – Thiago Victor Cardoso, 2020

círculo correspondente é incrementada, indicando assim qual foi a trajetória percorrida por determinado *frame*, de modo a representar qual foi a primeira coordenada (menor área do círculo) e a última coordenada (maior área do círculo). Estes atributos podem ser vistos na tabela 10. A duração de cada coordenada X,Y corresponde ao período de tempo de aproximadamente 3,3 milissegundos.

Frames reproduzidos. Representam os *frames* utilizados como estímulo. Tais vídeos apresentam resolução de 30 *frames*/segundo. De acordo com a figura 9, as coordenadas apresentadas consideram cerca de 10 registros por *frame*, correspondendo a 10 traçados de círculos. A figura 10 indica três *frames* em movimento, com a modelo apontando a mão para o objeto panda. Cada *frame* apresentado indica cerca de 10 registros (10 círculos) por indivíduo, mostrando a trajetória percorrida por estes *frames*.

ROIs demarcadas. Além de compreender o comportamento do olhar mediante à trajetória efetuada pelo indivíduo e os paradigmas envolvidos neste estudo, foi elaborado na própria ferramenta um modo de visualizar as ROI's flutuantes (seção 4.3.1). A visualização consiste em representar visualmente as posições da ROI de acordo com o *frame* relacionado, conforme é exposto na figura 10.

Figura 10 – Exemplo da visualização de trajetórias e as ROIs em movimento (retângulos demarcados em vermelho nas regiões do rosto, panda e mão). Rastreamento do olhar, por meio de variações das cores em azul (lado esquerdo) e vermelho (lado direito). Círculos em tons de azul do lado esquerdo correspondem a indivíduos do grupo DT e círculos em tons de vermelho do lado direito correspondem a indivíduos do grupo TEA.

Fonte – Thiago Victor Cardoso, 2020

4.3 Classificação com base em trajetórias

A figura 11 apresenta a visão geral da abordagem para a construção do modelo de classificação, baseado em trajetórias do rastreamento de olhar.

Figura 11 – Visão geral da abordagem proposta.

Fonte – Thiago Victor Cardoso, 2020

Dados de rastreamento de olhar e os próprios estímulos constituem os dados de entrada. O primeiro passo é a identificação das regiões flutuantes, necessária para que a fase de extração de características obtenha dados adequados referentes ao paradigma de AC. Essas duas etapas constituem o principal diferencial da abordagem proposta. O treinamento do modelo e a classificação usando Comitê de Máquinas (CM) constituem os pontos finais da abordagem necessários para classificação de um indivíduo em TEA ou DT. A seguir, cada uma das fases é apresentada com mais detalhes.

4.3.1 Identificação das ROIs flutuantes

Em cada vídeo, as ROIs são definidas por especialistas e evidenciam regiões para as quais o indivíduo deveria direcionar o seu olhar, considerando o paradigma de AC. Assim, por exemplo, no primeiro vídeo (figura 8) espera-se que o olhar seja direcionado ao rosto da terapeuta, enquanto que no terceiro vídeo em alguns *frames* o olhar pode ser direcionado a três ROIs diferentes (rosto da terapeuta, mão da terapeuta que faz apontamento, e objeto sendo apontado).

As demarcações do vídeo são realizadas de forma semiautomática. Durante o processo são desenhados retângulos nas regiões das imagens que são pertinentes aos parâmetros deste estudo, conforme apresentado na figura 12. O especialista delimita apenas o primeiro *frame* do vídeo que define a região de interesse. Um procedimento automatizado se encarrega de propagar o retângulo demarcado para os demais *frames* até que uma ROI diferente da anterior seja identificada no vídeo.

Figura 12 – Exemplo de demarcação da ROI na imagem.

Fonte – Thiago Victor Cardoso, 2020

Uma das dificuldades em utilizar as ROIs para analisar trajetórias considerando AC é identificar um mesmo objeto que se movimenta e assume características distintas ao longo do vídeo. A esta ROI atribuímos o nome de ROI flutuante. Embora a ROI se movimente ao longo do vídeo, é importante que a extração de características considere a semântica do objeto, identificando-o como pertencente à mesma ROI. Com o objetivo de solucionar essa dificuldade foi desenvolvido um método que considera o movimento completo da trajetória do mesmo objeto de maneira única. Assim, é possível que na extração de características o conjunto de *frames* que possuem as ROIs de mesma semântica sejam considerados como um único grupo, conforme descrito na seção 4.3.2.

Como exemplo, na figura 13 observa-se quatro *frames* e uma ROI em amarelo demarcando a região da mão da terapeuta. A mão da terapeuta se movimenta em *frames* contínuos. Em cada um deles, o “objeto mão” assume tamanho e posição diferentes dos demais *frames*. No entanto, o paradigma de AC acoplado ao ET tem a intenção de identificar se o indivíduo direcionou seu olhar para a mão, independentemente da posição exata que o “objeto mão” assume na imagem. Desta maneira, este método é capaz de acompanhar as ROIs pelo seu significado semântico e não apenas pelas posições dessas regiões.

Figura 13 – Exemplo de ROIs flutuantes: o mesmo objeto apresenta diferentes características ao longo do vídeo.

Fonte – Thiago Victor Cardoso, 2020

Para identificar as ROIs flutuantes, a partir das ROIs demarcadas, os *frames* são processados por métodos de processamento de imagens, visando à obtenção das posições das coordenadas cartesianas dessas áreas e os seus respectivos centros. Se houver mais de uma ROI no mesmo *frame*, cada uma delas é delimitada por uma cor (figura 13). Este passo é fundamental para a realização da fase de extração de características, pois cada cálculo realizado nesta fase depende das posições das áreas e centros destas ROIs. Para processar as posições e centros de cada ROI, foi desenvolvido um algoritmo que possibilita identificar os objetos que são exibidos durante a cena, independentemente de apresentarem movimento.

Como critério de entrada para a execução deste algoritmo é necessário definir em um arquivo de cabeçalho informações relacionadas à identificação do *frame*, às posições das áreas e centros das ROIs, e às condições das ROIs. A identificação do *frame* são os números que perfazem a sequência dos vídeos. As posições das áreas e centros são obtidos pela implementação de um método computacional que detecta os contornos dos objetos demarcados, além das posições de seus respectivos centros. Este método foi construído em C++, utilizando a biblioteca OpenCV (BRADSKI, 2000), para processar os objetos presentes no *frame*. As condições das ROIs são identificadores que representam o estado

lógico destes objetos de maneira visual, relacionando cada condição lógica por meio de uma cor.

A figura 14 exemplifica a construção de três ROIs e suas respectivas condições: a região da face, a mão da terapeuta e um brinquedo. A cor azul indica que o *frame* em questão corresponde ao primeiro instante que objeto apareceu; a cor vermelha representa que o objeto se movimentou em relação ao *frame* anterior; a cor amarela indica que não houve movimento; e a cor verde indica a última presença da ROI no *frame*. Considerando estes aspectos já mencionados, por meio desta técnica é possível determinar todo o movimento de cada objeto demarcado pela ROI, desde a origem ao final da sua exibição no vídeo, deste modo, resultando na concepção das ROIs flutuantes.

Figura 14 – Processo para identificar objetos presentes nas ROIs flutuantes.

Fonte – Thiago Victor Cardoso, 2020

4.3.2 Extração de características

A fase de extração de características é responsável por identificar, para cada indivíduo, um conjunto de posições de sinais de ET referentes às ROIs flutuantes obtidas no passo anterior e realizar os cálculos das características. O resultado deste processo é um conjunto de nove *datasets* (um para cada segmento de vídeo), cada um deles com 86 instâncias referentes aos indivíduos cujos sinais de ET foram coletados. Em cada *dataset*, cada instância tem um total de características calculada em relação ao total de ROIs que foram identificadas (15 características por ROI). Assim, por exemplo, o primeiro *dataset* (cujo segmento de vídeo tem uma única ROI, conforme apresentado na figura 8), tem 15 características em cada instância e os os datasets com três ROIs têm 45 características.

Visando a obter dados adequados dentro do paradigma de AC, algumas das características definidas na abordagem proposta tiveram por base o trabalho de R-Tavakoli *et al.* (2015). Outras foram definidas de forma inovativa a partir da observação dos especialistas envolvidos no projeto. O princípio básico para o processamento dos *frames* visando à obtenção das características é a definição das ROIs flutuantes, descritas na seção 4.3.1. Assim, para cada *frame* são identificados os registros de ET correspondentes. Considerando a quantidade de *frames* por segundo dos vídeos e a frequência de aquisição do equipamento de ET, em média há dez registros de ET para cada *frame*. A identificação dos registros de ET referente a uma ROI usa o conceito de ROI flutuante: uma ROI é identificada como única considerando a sua semântica (seção 4.3.1). Assim, para cada ROI flutuante j identificada, por exemplo, durante 10 *frames* do vídeo, são considerados os registros de ET referentes aos 10 *frames* em questão (em torno de 100 registros de ET) para calcular as 15 características apresentadas a seguir.

1. Média das distâncias ao centro da ROI: equivale à média aritmética das distâncias entre os dados de rastreamento até o centro da ROI, de cada indivíduo. Em média são considerados 10 registros de ET por *frame*, e a quantidade total de *frames* depende da identificação da ROI flutuante. Primeiramente é calculada a distância de cada registro de visita identificada no ET ao centro da ROI respectiva, conforme Equação 3. Em seguida, é obtida a média das distâncias usando a Equação 4.

$$d_i^j = \sqrt{(x_i - ROI.center(x)_j)^2 + (y_i - ROI.center(y)_j)^2} \quad (3)$$

$$m_d^j = \frac{1}{N} \sum_{i=1}^N d_i^j \quad (4)$$

Na equação 3, x_i e y_i são as posições das coordenadas expressas pelos atributos *GazePointX* (*MCSpx*) e *GazePointY* (*MCSpy*) no instante i . O parâmetro *ROI.center* indica as posições x e y dos centros da ROI; j refere-se ao índice de uma ROI ativa no instante i . Na equação 4, N é quantidade de distâncias obtidas durante o rastreamento do olhar do primeiro até o último registro do segmento de vídeo que contém a ROI em questão. Desta forma, essa medida mede o quão distante, na média, o olhar de indivíduo está do centro da ROI j .

2. Desvio padrão das distâncias: esta característica é obtida a partir das distâncias obtidas no processo de cálculo da característica anterior, conforme a Equação 5, em que d_i^j é elemento de distância e m_d^j é a média aritmética calculada pela equação 4.

$$\sigma Dist^j = \sqrt{\sum \frac{(d_i^j - m_d^j)^2}{N}} \quad (5)$$

3. Coeficiente de variação das distâncias: é calculado pela equação 6, em que $\sigma Dist^j$ é o desvio padrão e m_d^j corresponde à média aritmética das distâncias ao centro da ROI j . Esta fórmula indica o percentual de homogeneidade destas distâncias, de modo que, quanto mais homogêneas forem as distâncias, menor será o valor medido por este coeficiente de variação.

$$coef_Dist^j = \frac{\sigma Dist^j * 100}{m_d^j} \quad (6)$$

4. Identificador de visita: é uma variável lógica que assume os valores um ou zero, respectivamente, para indicar se o indivíduo visitou ou não visitou a ROI j .

5. Identificador de fixação: é uma variável lógica que assume os valores um ou zero, respectivamente, para indicar se o indivíduo fixou ou não fixou a ROI j .

6. Tempo de início da primeira visita: determina o primeiro instante de tempo no qual o olhar do indivíduo visitou a ROI j . Esta característica considera a abordagem de antecipação do olhar, conforme detalhado na seção 4.3.3.

7. Tempo de início da primeira fixação: determina o primeiro instante de tempo que o indivíduo fixou o olhar na ROI j . Esta característica considera a abordagem de antecipação do olhar, conforme detalhado na seção 4.3.3.

8. Ordem da maior visita: refere-se à ordem em que ocorreu a visita com maior duração em relação à ROI j . Por exemplo: se a maior duração foi identificada na primeira visita, este atributo recebe o valor um.

9. Ordem da maior fixação: a exemplo da característica anterior, esta características refere-se à ordem em que ocorreu a fixação de maior duração do olhar na ROI j .

10. Duração da primeira visita: indica o montante de tempo decorrido na primeira visita do olhar na ROI j .

11. Duração da primeira fixação: indica o montante de tempo decorrido na primeira fixação do olhar na ROI j .

12. Maior duração da visita: indica a quantidade de tempo da maior duração de visita do olhar na ROI j .

13. Maior duração da fixação: indica a quantidade de tempo da maior duração de fixação do olhar na ROI j .

14. Porcentual de visitas: representa a divisão entre a soma das durações das visitas na ROI j e a duração total dos *frames* presentes no segmento. A proporção desta característica é compreendida pelo intervalo de 0 a 100 %.

15. Porcentual de fixação: representa a divisão entre a soma das durações das fixações na ROI j e a duração total dos *frames* presentes no segmento. A proporção deste porcentual é compreendido pelo intervalo de 0 a 100 %.

4.3.3 Antecipação do olhar

A antecipação de um fato que está por acontecer é uma das premissas para verificação do paradigma de AC. As características “Tempo de início da primeira visita” e “Tempo de início da primeira fixação” verificam o quanto o olhar do indivíduo está sincronizado com o estímulo apresentado no vídeo, medindo a antecipação ou atraso do olhar em relação à ocorrência da ROI.

A figura 15 mostra um exemplo de extração das duas características citadas, as quais resultam na quantidade de tempo de antecipação ou atraso. Os círculos azuis representam os registros de ET, sendo que em cada parte da figura são representados 10 registros numerados com -10 a -1 (região à esquerda), 1 a 10 (região central) e 11 a 20 (região à direita). O retângulo interno representa a ROI. Para uma dada ocorrência de uma ROI, há três conjuntos de registros de ET: (1) registros que ocorreram imediatamente antes da ocorrência da ROI e que portanto pertencem ao segmento anterior (região esquerda da figura 15); (2) registros que ocorreram no primeiro *frame* em que aquela ROI tornou-se ativa (região central da figura 15) e (3) registros que ocorreram nos *frames* seguinte à ativação daquela ROI (região direita da figura 15).

O cálculo dessas características analisa os conjuntos de registros de ET observados antes, durante e após a ocorrência da ROI, a fim de determinar a primeira ocorrência observada de visita ou fixação do olhar na ROI. Quando os valores se apresentam negativos, entende-se que houve antecipação do olhar para determinada ROI, ou seja, o participante

já estava com o olhar naquela região momentos antes dela se tornar uma ROI. Porém, se apresentar valores positivos, existe a possibilidade de haver um atraso no olhar para a região de interesse. Quanto maior o valor, maior o atraso observado.

Figura 15 – Exemplo de cálculo para obter a antecipação ou atraso dos dados de ET a partir da ROI.

Fonte – Thiago Victor Cardoso, 2020

4.3.4 Construção do modelo

Conforme abordado nas seções 4.1.2 e 4.3.2, as características foram extraídas de cada segmento, compondo nove *datasets*. Uma vez que as características foram devidamente extraídas, formando um vetor de características, foi implementado um modelo computacional capaz de prever a classe de um indivíduo (TEA ou DT) a partir das trajetórias percorridas capturadas pelo ET. A implementação foi realizada em linguagem Python (ROSSUM; JR, 1995).

O modelo aqui proposto consiste em um comitê de nove RFs, um para cada *dataset*. *Random Forest* é um método que realiza a predição por meio de várias árvores de decisão, de modo que a classe com mais votos se torna a previsão do modelo (YIU, 2020). Este método é relevante devido à sua capacidade de utilizar tantos dados numéricos quanto categóricos.

Para a realização do treinamento desse comitê foi implementado um software que consiste em, para cada *dataset*, realizar seu pré-processamento (durante a qual foram executadas adequações como normalização e correção de *missing values*), aplicação de diferentes algoritmos de balanceamento de classes e calibração de parâmetros da RF. Após o treinamento do modelo para cada *dataset*, três estratégias de combinação dos mesmos foram testadas.

Pré-processamento dos dados

Quando uma ROI não é visitada, não há uma primeira ou maior visita daquela ROI. Consequentemente, alguns atributos apresentarão valores ausentes (*missing values*). Nesses casos, foram adotados os seguintes procedimentos para os respectivos atributos:

- **Tempo de início da primeira visita:** o valor ausente foi substituído pela multiplicação do tempo em que ocorreu o último *frame* de determinada ROI, por uma constante w (empiricamente neste trabalho foi definido o valor 4 para esta constante).
- **Tempo de início da primeira fixação:** foi adotado o mesmo procedimento descrito para o item anterior.
- **Ordem da maior visita:** o valor ausente foi substituído pelo valor da ordem do último *frame*, referente a determinada ROI, multiplicada pela constante w .
- **Ordem da maior fixação:** o valor ausente foi substituído pelo valor da ordem do último *frame*, referente a determinada ROI, multiplicada pela constante w .

Normalização dos dados

Outra atividade importante durante o pré-processamento dos dados é a normalização.

Em cada um dos nove *datasets* obtidos no final da fase de extração de características (seção 4.3.2), foi aplicado o método *Min-Max* (SARANYA; MANIKANDAN, 2013), que possibilitou a padronização dos valores presentes em suas colunas no intervalo [0-1] para todos os atributos numéricos.

Balanceamento dos dados

Os dados disponíveis para este experimento são ligeiramente desbalanceados, pois o *dataset* apresenta uma relação de indivíduos com TEA quase duas vezes maior do que indivíduos com DT, o que pode afetar o desempenho dos classificadores e, consequentemente, os resultados obtidos.

Para minimizar este problema foi adotada a técnica *Synthetic Minority Over-sampling Technique* (SMOTE) (CHAWLA *et al.*, 2002), que efetua o balanceamento dos

dados. Esse método gera dados sintéticos contidos na classe minoritária (CHAWLA *et al.*, 2002), (no caso do presente projeto, registros do grupo DT) ou combina tal estratégia com subamostragem da classe majoritária (IEEE, 1976) Na classe minoritária, são introduzidos exemplos sintéticos sobre os dados da vizinhança mais próxima, de forma aleatória para aumentar a frequência desta classe e equilibrar a amostragem destes dados. Assim, além do algoritmo original SMOTE, foram testadas também suas variações conhecidas como *Borderline*, *KMeans*, *SVM* e *Tomek*, implementadas na biblioteca *scikit-learn* do Python. Também foi testado o não balanceamento das classes.

Calibração

No processo de treinamento foi necessário efetuar a calibração dos parâmetros de treinamento da RF de cada um dos nove *datasets*. A calibração dos parâmetros de cada RF foi realizada por meio da função `GridSearchCV` do Python (biblioteca *scikit-learn*) (PEDREGOSA *et al.*, 2011), para maximizar a acurácia. Também foram variados os parâmetros dos métodos de balanceamento, de acordo com a tabela 4, na qual a primeira linha disponibiliza os parâmetros utilizados pelo método *Smote* e na segunda linha são apresentados os parâmetros adotados na utilização do método RF.

Tabela 4 – Parâmetros usados no processo de calibração do modelo.

Algoritmo	Parâmetros usados
Balanceamento	'Borderline', 'KMeans', 'Smote', 'SVM', 'Tomek', 'nenhum'
RF	max_features: números gerados aleatoriamente de 1 a 100; min_samples_leaf: 1, 2, 5, 10, 15, 20; n_estimators: 500, 1000, 1500, 2000.

Os parâmetros que sofreram variações foram: **a) max_features:** que determina o número máximo de características consideradas para dividir um nó; **b) min_samples_leaf:** número mínimo de pontos de dados permitidos em um nó folha; **c) n_estimators:** número de árvores da floresta (KOEHRSEN, 2018). Os parâmetros *bootstrap*, *class_weight*, *criterion*, *max_depth*, *n_jobs*, *random_state* tiveram valores fixos.

Estimação de métricas de desempenho

Conforme descrito um pouco mais à frente, para a definição do comitê de máquinas será levada em consideração a acurácia de cada classificador individual (em uma das estratégias) e também do comitê como um todo para a escolha de uma limiar de votação, que é um parâmetro do comitê proposto. Para estimar essas acurácias e também outras métricas de desempenho foi adotado o mecanismo de validação cruzada ($k = 10$). Esta técnica separa o conjunto de dados dos indivíduos em k partes, entre 9 partes usadas para treinamento e uma parte usada para teste, de modo que o método de validação é repetido k vezes. Cada parte do subconjunto de teste possui entre oito a nove indivíduos.

Ao finalizar o processamento do algoritmo, é estabelecida a melhor acurácia do conjunto e seus respectivos parâmetros. Estes resultados serão discutidos na seção 5.

Para avaliar o desempenho do classificador foram utilizadas as métricas de acurácia (ACC), sensibilidade (TPR), precisão (PPV), especificidade (TNR) e F1-score (F1), definidas nas Equações 7 a 11, onde TP, TN, FP e FN representam, respectivamente, os valores verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

$$PPV = \frac{TP}{TP + FP} \quad (9)$$

$$TNR = \frac{TN}{TN + FP} \quad (10)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (11)$$

Definição do Comitê de Máquinas

Para definir o classificador final foi adotado o conceito de Comitê de Máquinas (CM). Cada um dos CMs testados consiste de um conjunto de nove classificadores RF

(um para cada segmento de vídeo) que, combinados, decidem se determinado indivíduo pertence ao grupo TEA ou DT. Assim, o critério de decisão baseia-se na combinação dos resultados de predição de cada classificador, tendo sido avaliadas três abordagens:

1. **Classificação por configuração e pesos homogêneos:** considerando-se os nove classificadores, cada um contribuindo com um voto de igual peso, um indivíduo é classificado como TEA se ele receber l ou mais votos TEA, sendo l um limiar com valor compreendido entre 1 e 9. De acordo com o Algoritmo 1, o voto é somado como pertencente ao grupo TEA (linha 12) ou DT (linha 14), de acordo com o teste de votação (linha 11). As métricas de desempenho, pertencente a cada *fold*, são calculadas na linha 18, que referem-se à acurácia, sensibilidade, especificidade, precisão e *F1-score*. A métrica de desempenho do conjunto de *folds*, é obtida na linha 20, e referente a cada limiar, é indicado na linha 22. O limiar é variado entre os valores 1 a 9, a fim de que seja obtido a melhor predição (linhas 1, 11 e 22).
2. **Classificação por configuração homogênea e pesos ponderados:** ainda considerando sempre os nove classificadores, considera um peso para a predição de cada classificador. O peso de cada classificador corresponde à sua acurácia média obtida com os parâmetros ótimos (calculada apenas para aquele segmento). Conforme o Algoritmo 2, a acurácia média de cada *foldI* é obtida pela linha 18, e a acurácia resultante de cada comitê é atribuída na linha 20. A principal diferença com a abordagem anterior é que os votos TEA são somados pelas médias da acurácia de determinado segmento. O limiar foi variado entre 1.0 até 5.9, sendo 5.9 o valor correspondente à soma das acurácias dos nove classificadores. A variação do limiar ocorreu em incrementos de 0,1.
3. **Classificação por configuração variável e pesos homogêneos:** são testadas diferentes combinações de um a nove classificadores (cada um com igual peso de voto) e para cada uma delas são testados todos os possíveis valores de limiares. Ou seja, são realizadas 512 combinações, que correspondem às diferentes possibilidades de combinar um a nove classificadores, para obter a melhor acurácia, conforme definido no Algoritmo 3. As linhas 1 e 2, correspondem à criação do mapa de combinações, que consiste em uma matriz de dimensões 512×9 que possui na posição (i, j) o valor 1 se o classificador j participa da configuração i e 0 caso contrário, sendo portanto cada linha dessa matriz correspondente a um comitê distinto. Na linha 3, é definida

a quantidade de segmentos (classificadores) a ser usada na combinação em questão. Na linha 4, o limiar é variado de acordo com o número de segmentos para cada combinação.

Algoritmo 1 Classificação por configuração e pesos homogêneos.

Result: Limiar de classificação e valores das métricas de desempenho do comitê baseado na estratégia de configuração e pesos homogêneos correspondente à maior acurácia.

```

1 for limiar: 1 até 9 do
2   for foldI: 1 até 10 do
3     Inicializa contadores de TP, TN, FP, FN
4     for each indivíduo do foldI do
5       Inicializa votosTEA com 0
6       for segmento: 1 até 9 do
7         if predição da RF deste segmento para esse indivíduo == TEA then
8           | Incrementa a variável votosTEA
9         end
10      end
11      if votosTEA >= limiar then
12        | classe_predita = TEA
13      else
14        | classe_predita = DT
15      end
16      Incrementa TP,FP,TN ou FN de acordo com a classe real do indivíduo e sua
        classe_predita
17    end
18    Calcula as métricas de desempenho deste foldI (de acordo com as variáveis
      TP,FP,TN,FN)
19  end
20  Calcula a média de cada métrica de desempenho
21 end
22 Retorna limiar e métricas de desempenho correspondente à maior acurácia média

```

Fonte – Thiago Victor Cardoso, 2020

Algoritmo 2 Classificação por configuração homogênea e pesos ponderados.**Result:** Limiar de classificação e valores das métricas de desempenho do comitê baseado na estratégia de configuração homogênea e pesos ponderados correspondente à maior acurácia.

```
1 for limiar: 1.0 até 5.9 do
2   for foldI: 1 até 10 do
3     Inicializa contadores de TP, TN, FP, FN
4     for each indivíduo do foldI do
5       Inicializa votosTEA com 0
6       for segmento: 1 até 9 do
7         if predição da RF deste segmento para esse indivíduo == TEA then
8           Incrementa a variável votosTEA
9         end
10      end
11      if votosTEA >= limiar then
12        classe_preditada = TEA
13      else
14        classe_preditada = DT
15      end
16      Incrementa TP,FP,TN ou FN de acordo com a classe real do indivíduo e sua
        classe_preditada
17    end
18    Calcula as métricas de desempenho deste foldI (de acordo com as variáveis
      TP,FP,TN,FN)
19  end
20  Calcula a média de cada métrica de desempenho
21 end
22 Retorna limiar e métricas de desempenho correspondente à maior acurácia média
```

Fonte – Thiago Victor Cardoso, 2020

Algoritmo 3 Classificação por configuração variável e pesos homogêneos.

Result: Limiar de classificação e valores das métricas de desempenho do comitê baseado na estratégia de configuração variável e pesos homogêneos, correspondente à maior acurácia.

```

1  Calcula a matriz M de configurações
2  for config: 1 até 512 do
3      num_segmentos = soma das colunas da linha config da matriz M
4      for limiar: 1 até num_segmentos do
5          for folds: 1 até 10 do
6              Inicializa contadores de TP, TN, FP, FN
7              for each indivíduo do foldI do
8                  Inicializa votosTEA com 0
9                  for segmento: 1 até 9 do
10                     if M[config][segmento] then
11                         if predição da RF deste segmento para esse indivíduo == TEA then
12                             | Incrementa a variável votosTEA
13                         end
14                     end
15                 end
16                 if votosTEA >= limiar then
17                     | classe_predita = TEA
18                 else
19                     | classe_predita = DT
20                 end
21                 Incrementa TP,FP,TN ou FN de acordo com a classe real do indivíduo e
                    sua classe_predita
22             end
23             Calcula as métricas de desempenho deste foldI (de acordo com as variáveis
                TP,FP,TN,FN)
24         end
25     end
26     Calcula a média de cada métrica de desempenho
27 end
28 Retorna limiar e métricas de desempenho correspondente à maior acurácia média

```

5 Resultados e discussões

Este capítulo apresenta os resultados e discussões do projeto desenvolvido. Inicialmente são apresentados resultados das primeiras etapas realizadas referente aos componentes integrantes do projeto (seção 5.1). Em seguida, são explanados os resultados das abordagens dos Comitês de Máquinas (seção 5.2). Na seção 5.3, são comparados classificadores de outros trabalhos presentes na literatura com o presente modelo. E finalizando este capítulo, na seção 5.4, são mencionadas as limitações contidas neste projeto.

5.1 Ferramentas e técnicas implementadas

A ferramenta de visualização de trajetórias, implementada neste projeto (seção 4.2), que representa graficamente os sinais do olhar delineando os movimentos de trajetórias, foram essenciais para o entendimento do paradigma abordado, pois observar o comportamento dos indivíduos sem uma ferramenta interativa dificulta sobremaneira na percepção dos momentos que retratam uma maior possibilidade de diferenciação entre as classes. Assim, as tarefas de analisar e propor novos atributos são facilitados com a utilização desta ferramenta de visualização de trajetórias. Também, segundo a literatura, a AC é verificada comumente por especialistas sem auxílio de ferramentas (PARISH-MORRIS *et al.*, 2019; BEDFORD *et al.*, 2012), de forma que a construção desta aplicação corresponde a uma contribuição nesta área, pois auxilia na observação dos eventos provenientes do rastreamento do olhar que estão presentes na AC.

Durante a fase de seleção de características foi possível estabelecer os atributos que se relacionaram com o paradigma de AC, adotado neste projeto. Cada atributo implementado reflete as condições presentes nos movimentos de trajetória. Embora todas as características existentes neste trabalho são de grande importância para a composição do *dataset*, é relevante reiterar que o “Tempo de início da primeira visita” e “Tempo de início da primeira fixação” consistem de marcadores do comportamento da antecipação ou atraso do olhar (seção 4.3.3). Este comportamento é intrínseco ao estudo da AC e, por conseguinte, contribuiu para construir a técnica de ROIs flutuantes (seção 4.3.1).

5.2 Comparação das abordagens dos Comitês de Máquinas

A tabela 5 apresenta os resultados das melhores acurácias obtidas na etapa de calibração, assim como os parâmetros empregados para cada um dos nove segmentos de vídeo.

Os parâmetros utilizados, pertencente ao método *SMOTE* foram variados de acordo com os atributos explanados na seção 4.3.4. Para cada segmento utilizado foram efetuadas 10 rodadas de teste, sobre todos os dados presentes em determinado *dataset*, resultando em acurácia média variando entre 0,727 a 0,839.

De acordo com a tabela 5 é possível verificar que existe uma predominância pelo parâmetro *Tomek* aplicado ao método. O *Tomek*, assim como os demais parâmetros, são algoritmos utilizados para a geração dos dados sintéticos referentes à classe minoritária.

Os parâmetros obtidos na etapa de calibração foram utilizados como entrada para as três diferentes abordagens de CM (seção 4.3.4). A soma dos votos TEA de cada comitê é comparada com o respectivo limiar. Por exemplo, na abordagem da classificação por configuração e pesos homogêneos, o limiar é variado de 1 a 9. Na abordagem de classificação por configuração homogênea e pesos ponderados, o limiar é compreendido entre 1,0 a 5,9. E na abordagem de classificação por configuração variável e pesos homogêneos, o limiar é variado de 1 até o número de segmentos de cada configuração.

A estratégia da classificação por configuração e pesos homogêneos é apresentada na figura 16. Para cada valor de limiar, obteve-se uma medida de acurácia. Limiares menores que 5 apresentaram acurácias em média de 66%, alcançando a acurácia máxima de 71% com limiar igual a 5. A partir do valor 6, a acurácia diminuiu para 67% e conforme o limiar foi aumentando, consequentemente a acurácia diminuiu consideravelmente, o que indicou uma queda acentuada na sensibilidade, mas um aumento na especificidade, explicando a diminuição na acurácia. Considerando o limiar igual a 5, o classificador dessa estratégia obteve como demais métricas de desempenho, a sensibilidade (93%), a precisão (72%), a especificidade (30%) e o *F1-score* (81%).

A abordagem da classificação por configuração homogênea e pesos ponderados consiste também de uma importante técnica na otimização do classificador, mas diferentemente da abordagem anterior, adota diferentes pesos no sistema de votação. Estes pesos correspondem às acurácias de cada um dos 9 classificadores. O limiar foi variado de 1,0

a 5.9, conforme pode ser observado na figura 17. As melhores acurácias obtidas nesta abordagem, foram com o limiar entre 2,4 a 3,6. A partir do valor 4,3, a acurácia diminuiu drasticamente. Também, é possível perceber que de igual modo a abordagem anterior concentrou a maior acurácia no ponto médio do limiar (de 1 a 9), nesta abordagem a melhor acurácia foi também próximo no ponto médio (3,45), correspondendo ao limiar de 3,6 e acurácia de 71%. Considerando o limiar 3,6, esta estratégia apresentou como demais métricas de desempenho a sensibilidade (77%), a precisão (79%), especificidade (60%) e o *F1-score* (76%).

A classificação por configuração variável e pesos homogêneos estabelece diferentes combinações de 1 a 9 classificadores que fizeram a predição sobre cada segmento. De acordo com a figura 18, é possível observar valores de acurácia entre 56% a 75%. Esta abordagem, diferentemente das anteriores, apresentou uma faixa estreita e tênue das acurácias obtidas pelas combinações do comitê, o que demonstra ser um modelo adequado para otimizar o desempenho do classificador, mediante um conjunto algoritmos e dados, de modo a combinar os preditores e obter a melhor acurácia. O melhor limiar obtido nesta abordagem foi de 322, apresentando acurácia de 75%. As demais métricas para o limiar de 322 foram: sensibilidade (87%), precisão (78%), especificidade (53%) e *F1-score* (82%).

Figura 16 – Classificação por configuração e pesos homogêneos. Resultado da acurácia variando o limiar de 1 a 9.

Fonte – Thiago Victor Cardoso, 2020

Figura 17 – Classificação por configuração homogênea e pesos ponderados. Resultado da acurácia variando o limiar de 1,0 a 5,9.

Fonte – Thiago Victor Cardoso, 2020

Figura 18 – Classificação por configuração variável e pesos homogêneos. Resultado da melhor acurácia considerando cada uma das 512 configurações (eixo das abscissas).

Fonte – Thiago Victor Cardoso, 2020

Na tabela 6 são apresentados os resultados do comitê de melhor acurácia para as três abordagens, sobressaindo-se a abordagem da classificação por configuração variável e pesos homogêneos. As abordagens da classificação por configuração e pesos homogêneos e a classificação por configuração homogênea e pesos ponderados indicaram acurácia média de 71%, enquanto que o uso da classificação por configuração variável e pesos homogêneos indicou acurácia média de 75%, usando os parâmetros que derivaram da combinação 322 dos segmentos. Essa configuração é composta por três classificadores: dos segmentos 1, 3 e 8. O segmento 1 refere-se à terapeuta olhando em direção à tela do monitor; o segmento 3,

indica a terapeuta olhando para o objeto do panda, considerando duas ROIs; e o segmento 8, referencia-se à terapeuta voltando o olhar para a tela do monitor.

Tabela 5 – Parâmetros das melhores acurácias pelo método de calibração.

Dataset	Smote	Parâmetros	Acc
Segmento 1	Borderline	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:8, min_samples_leaf:1, n_estimators:1000, n_jobs:-1, random_state:707878	0,807
Segmento 2	SVM	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:1, min_samples_leaf:10, n_estimators:1000, n_jobs:-1, random_state:707878	0,804
Segmento 3	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:5, min_samples_leaf:1, n_estimators:500, n_jobs:-1, random_state:707878	0,839
Segmento 4	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:7, min_samples_leaf:1, n_estimators:1000, n_jobs:-1, random_state:707878	0,773
Segmento 5	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:1, min_samples_leaf:1, n_estimators:1000, n_jobs:-1, random_state:707878	0,795
Segmento 6	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:5, min_samples_leaf:5, n_estimators:1000, n_jobs:-1, random_state:707878	0,727
Segmento 7	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:3, min_samples_leaf:1, n_estimators:1000, n_jobs:-1, random_state:707878	0,811
Segmento 8	KMeans	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:10, min_samples_leaf:5, n_estimators:1000, n_jobs:-1, random_state:707878	0,753
Segmento 9	Tomek	bootstrap:True, class_weight:None, criterion:gini, max_depth:None, max_features:2, min_samples_leaf:1, n_estimators:1000, n_jobs:-1, random_state:707878	0,757

Fonte – Thiago Victor Cardoso, 2020

Assim, a melhor acurácia média foi obtida usando os segmentos de vídeo 1, 3 e 8, os quais apresentam três estágios fundamentais da atenção compartilhada: início do olhar da terapeuta na tela do monitor (segmento 1), movimento do olhar da terapeuta ao objeto de interesse (segmento 3), e retorno do olhar em direção à tela do monitor (segmento 8). Há indícios de que, nos grupos TEA e DT considerados neste estudo, estas três perspectivas em conjunto contribuem significativamente para a distinção entre os indivíduos. Adicionalmente, o processamento de cada segmento de vídeo de maneira isolada não obteve acurácia superior, indicando que o processo da ação de forma completa é necessário. Destaca-se, ainda, que ROIs que englobam os olhos se mostraram mais capazes

Tabela 6 – Parâmetros referentes aos melhores resultados do CM.

Abordagem		ACC	TPR	PPV	TNR	F1	Params
Classificação por configuração e pesos homogêneos		0,71	0,93	0,72	0,30	0,81	Limiar 5 (limiar de 1 a 9)
Classificação por configuração homogênea e pesos ponderados		0,71	0,77	0,79	0,60	0,76	Limiar 3.6 (limiar de 1.0; 1.1;1.2 até 5.9)
Classificação por configuração variável e pesos homogêneos		0,75	0,87	0,78	0,53	0,82	Combinacao:322 valor combinacao:[1, 0, 1, 0, 0, 0, 0, 1, 0]; limiar:2; num_segs:3

Fonte – Thiago Victor Cardoso, 2020

de distinguir indivíduos dos dois grupos do que ROIs que consideram a mão da terapeuta. A melhoria dos *datasets*, tanto em termos de quantidade de instâncias, quando em termos de balanceamento, assim como a disponibilização de vídeos adicionais que pudessem ser utilizados como estímulos para avaliar o comportamento de atenção compartilhada, provavelmente contribuiriam para obtenção melhores resultados na classificação dos grupos de indivíduos.

Os resultados das médias de acurácias obtidas pelas 3 abordagens, são apresentadas na figura 19. Cada cor no gráfico indica uma abordagem utilizada. Conforme pode ser visto pela figura, os valores médios de acurácia foram de 0,60 (para a classificação por configuração e pesos homogêneos; e classificação por configuração variável e pesos homogêneos) e 0,57 (na classificação por configuração homogênea e pesos ponderados). A maior acurácia obtida é indicada pela abordagem da classificação por configuração variável e pesos homogêneos com valor de 0,75.

Figura 19 – Gráfico de *boxplot* exibindo as distribuições das acurácias em cada uma das abordagens.

Fonte – Thiago Victor Cardoso, 2020

A avaliação da sensibilidade das três abordagens (classificação por configuração e pesos homogêneos; classificação por configuração homogênea e pesos ponderados; e classificação por configuração variável e pesos homogêneos), apresentaram respectivamente os valores de 93%, 77% e 87%, indica um elevado número de acertos para os indivíduos da classe TEA, ou seja, os verdadeiros positivos. De igual modo, os resultados de precisão também foram adequados, indicando valores de 72%, 79% e 78%.

O modelo é capaz de distinguir melhor os indivíduos da classe TEA, mas tem seu desempenho comprometido para indicar indivíduos da classe DT. No entanto, é importante ressaltar que o número de indivíduos DT correspondem a cerca de 35% do total da base, o que pode ter contribuído para diminuir o desempenho da especificidade do modelo e, conseqüentemente, das métricas que dependem dessa variável, como a acurácia, por exemplo. Embora, a precisão e sensibilidade das três abordagens apresentaram valores

médios significativos, contribuindo com o *F1-score*, e sobressaindo assim a classificação baseada em configuração variável.

Considerando o desbalanceamento entre as classes, foram destacadas duas métricas como principais indicadores quanto ao desempenho do classificador: a acurácia e o F1-score. Foi obtida a acurácia máxima de 75% e F1-score máximo de 82%. É relevante salientar que o F1-score é uma métrica fundamental para casos em que o conjunto das classes são desproporcionais, visto que relaciona duas medidas (sensibilidade e precisão).

5.3 Comparação dos classificadores

Na tabela 7 são apresentados os resultados das acurácias obtidas pelos métodos citados na literatura para a classificação dos indivíduos no contexto do presente trabalho. Os trabalhos (YANEVA *et al.*, 2018; LIU; LI; YI, 2016; LI *et al.*, 2020) empregaram o paradigma de atenção visual, sendo que dois deles (YANEVA *et al.*, 2018; LIU; LI; YI, 2016) processaram os dados de rastreamento do olhar usando imagens estáticas como estímulos, enquanto que no estudo (LI *et al.*, 2020) foram usados vídeos como estímulos.

Embora as acurácias de algumas dessas abordagens indicaram índices melhores, chegando a 0.88 (LIU; LI; YI, 2016) e 0.84 (LI *et al.*, 2020), é importante salientar que a comparação é limitada, visto que trata-se de dados diferentes em termos de estímulos, pacientes e paradigmas considerados. O presente trabalho considera um grupo mais variável de indivíduos em termos de idade do que a amostra considerada em todos os trabalhos citados, indicando que a abordagem aqui proposta pode apresentar maior possibilidade de generalização dos resultados.

Adicionalmente, o paradigma de AC é um diferencial que pode contribuir para uma análise mais completa do indivíduo, visto que é analisada a trajetória do olhar. Ao considerar a atenção do indivíduo sobre áreas de interesse durante instantes de tempo contínuos, é possível analisar se em algum momento a pessoa manteve o foco na região de interesse, o que é dificultada quando somente o paradigma de atenção visual é empregado, como é o caso dos trabalhos citados. De modo que nenhum dos trabalhos analisados propõe a adoção das ROIs flutuante, estratégia utilizada para estabelecer o armazenamento e processamento dos eventos orientados ao foco da região de interesse e que são pertinentes à observação dos movimentos de trajetórias.

Além disso, tais estudos analisados não consideraram abordagens que propõem aperfeiçoar o desempenho do classificador, como as abordagens adotadas por esta presente pesquisa, que integra três técnicas de comitês de máquinas e prevalecendo a abordagem ganhadora que é a classificação por configuração variável e pesos homogêneos.

Tabela 7 – Comparação dos resultados entre o trabalhos relacionados. Legenda: AV - Atenção Visual; AC - Atenção Compartilhada

Referência	<i>Dataset</i>	Estímulo	Paradigma	Acurácia
Yaneva et al. (2018)	30 adultos	Imagens	AV	0,75 e 0,71
Liu, Li e Yi (2016)	87 crianças (4-11 anos)	Imagens	AV	0,88
Li et al. (2020)	64 crianças (16,3 meses)	Vídeos	AV	0,84
Este trabalho	86 crianças (3-15 anos)	Vídeos	AC	0,75

Fonte – Thiago Victor Cardoso, 2020

5.4 Limitações

Embora a base de dados utilizada neste projeto tenha apresentado dados desbalanceados e a duração dos vídeos foram de 29 segundos, o resultado obtido no classificador indica que a abordagem é promissora, revelando 75% de acurácia e 82% de F1-score. Também, é importante ressaltar que somente alguns vídeos culminaram em um resultado significativo de acurácia. Tais vídeos apresentam uma quantidade substancialmente reduzida de estímulos que retratam a situação da AC. Nos exemplos presentes deste projeto, foram utilizados cenas curtas indicando até três ROIs movimentando-se simultaneamente. Assim, adicionar mais informações de ROIs e destacando eventos relacionados à AC, é possível que o desempenho seja melhorado. Também, para contornar o problema das classes desbalanceadas, foi adotado o método *SMOTE* que baseia-se na criação de dados sintéticos da classe minoritária.

Outro aspecto relacionado ao desempenho do classificador, que também pode ser considerado, seria adotar outros métodos de AM e comparar com a presente abordagem. Algoritmos de Redes Neurais Artificiais, por exemplo, poderiam ser utilizados. Ainda, avaliar os resultados com algoritmos de redução de dimensionalidade sobre as características envolvidas, constituem de uma possibilidade de otimização desses resultados.

6 Considerações finais

O objetivo deste projeto foi propor um método computacional capaz de extrair características relevantes em relação à trajetória percorrida pelo rastreamento do olhar, considerando cada objeto contido em estímulos de vídeos, como ROIs flutuantes, em que a semântica é identificada mesmo se o objeto apresente características distintas ao longo do estímulo, possibilitando a classificação e estabelecendo a distinção entre as classes de indivíduos TEA e DT.

Conforme inicialmente mencionado (seção 1.2), este projeto apresentou a seguinte questão de pesquisa: é possível distinguir indivíduos dos grupos TEA e DT, por meio da análise dos dados de *eye tracking* referentes a acompanhamento de trajetórias usando o paradigma de AC?

De acordo com os resultados explanados no capítulo 5, o projeto atendeu à afirmação apresentada e confirmou a hipótese, implementando o modelo de trajetórias, que contempla a detecção de movimentos em atraso e antecipados em indivíduos, por meio da técnica das ROIs flutuantes, e estabeleceu a classificação dos indivíduos, alcançando acurácia de 75% e *F1-score* de 82%.

A ferramenta de visualização de trajetórias implementada contribuiu significativamente para os resultados deste projeto, e consequentemente para o meio de observação das trajetórias e das representações das ROIs. A importância de estudar a trajetória reflete na observação apresentada pelo paradigma de AC, que a partir deste tema, a ferramenta possibilitou fundamentar as ROIs flutuantes mediante às trajetórias relacionadas. Também, esta aplicação de visualização permite a utilização em outros paradigmas, outros dados de indivíduos e estímulos de vídeos, além da possibilidade de uso para outros propósitos, de modo que a implementação de tal ferramenta permite realizar diversos estudos dos dados de ET com ou sem a presença de determinada ROI.

Os modelos de CM empregados nesta pesquisa, revelaram diversas óticas de avaliações abordando aspectos de classificação por configuração e pesos homogêneos, de classificação por configuração homogênea e pesos ponderados, e de classificação por configuração variável e pesos homogêneos, alcançando melhor resultado na classificação por configuração variável e pesos homogêneos. Assim, estas abordagens podem ser aplicadas em outros modelos computacionais e diversos contextos que considerem AM.

6.1 Contribuições

Do nosso conhecimento, o presente trabalho é o primeiro que define o conceito de ROIs flutuantes para analisar o paradigma de AC. A AC fornece informações relevantes para o diagnóstico, além daquelas fornecidas pelos trabalhos citados. Assim, os índices de acurácia e de *F1_Score* obtidos na abordagem aqui proposta indicam que o modelo tem potencial para ser aplicado na rotina diagnóstica de TEA. Os resultados aqui obtidos, quando combinados com resultados de outros paradigmas, podem contribuir para aumentar a precisão diagnóstica e para diminuir o esforço dispendido pelo profissional de saúde. Além disso, é possível que métodos objetivos como o proposto neste trabalho contribuam para complementar os resultados dos instrumentos atuais usados para diagnóstico, os quais atualmente têm sua aplicação dependente da observação do profissional e/ou cuidador, o que pode variar em função de aspectos relacionados à experiência do primeiro ou à cultura do último.

Ainda em relação às características desenvolvidas no presente trabalho, é importante ressaltar que seus cálculos constituem um avanço na análise do paradigma de AC. Neste aspecto, foi proposto um modelo de trajetória que considera um estudo aprofundado nos movimentos ocasionados por objetos de interesse, mesmo que esses mudem seus aspectos visuais nos vídeos usados como estímulos. Como um recurso substancial para mapear estes objetos que se movimentam em determinadas ROIs, foi desenvolvido o conceito de ROIs flutuantes, que interpretam semanticamente o objeto de interesse em todo o seu movimento na tela.

Outro fator importante na definição do modelo proposto, foi adotar uma medida para antever os movimentos do olhar de um indivíduo em relação ao uma ROI. Para isso foi desenvolvido um método que possibilita a identificação de sinais de ET correspondentes a tempos anteriores e posteriores à presença da ROI, que propiciou detecção de adiamentos ou atrasos do olhar em relação ao comportamento esperado.

Salienta-se, por fim, que a construção de um comitê de máquinas para determinar a melhor acurácia com base na sua votação, avaliando-se todas as combinações possíveis de *datasets* possibilitou o aprimoramento no desempenho do classificador. E, ainda, a abordagem construída pode ser aplicada para quaisquer vídeos de AC disponíveis, bastando que o terapeuta indique o *frame* inicial e final de cada ROI.

Como contribuição à área científica e tecnológica, foi disponibilizado este projeto, com os códigos fontes e os *datasets* gerados a partir dos cálculos das características, além dos comitês de máquinas, métricas e resultados deste experimento. O material pode ser obtido em <https://github.com/ThiagoVCardoso/EyeTrackingTrajectoryModel>.

Parte das contribuições elencadas foram reunidas no artigo intitulado “*Autism Spectrum Disorder diagnosis based on trajectories of eye tracking data*”, submetido para publicação na conferência “*The 2021 ACM/SIGAPP Symposium on Applied Computing*”.

6.2 Trabalhos Futuros

Apesar deste trabalho apresentar bons resultados por meio do modelo implementado, algumas abordagens poderiam ser aprimoradas, como aumentar a integração da ferramenta de visualização de trajetórias com o modelo de classificação, tornando-se em uma aplicação única, e otimizando o processo de demarcação das ROI's flutuantes. É importante salientar que alguns dos processos semiautomáticos, como a identificação das ROIs iniciais, podem ser melhorados com métodos mais automatizados com o uso de técnicas de processamento de imagens. Consequentemente, com a integração e ajustes na interface, a ferramenta pode ser utilizada para auxiliar os profissionais de saúde na tomada de decisão.

Considerando o modelo concebido e implementado, o modo de classificação dos indivíduos, poderia possibilitar que a configuração dos parâmetros do método RF fosse realizada na própria ferramenta. Além disso, é possível implementar outros métodos de AM e expandir o comitê de máquinas com a seleção desses outros classificadores. Por fim, poderiam ser implementados algoritmos para seleção e fusão de características a serem acoplados ao modelo, de modo a incrementar a acurácia e outras métricas pertinentes.

Classes homogêneas de dados de entrada favorecem o desempenho do classificador e facilitam nas fases de treinamento e predição do modelo por não necessitarem de procedimentos artificiais para melhorar os dados. Assim, um trabalho futuro seria adicionar mais vídeos contendo outras condições que destacam a AC, aumentar o número de indivíduos no experimento e incluir novas características para a fase de extração de características, que estejam alinhados ao estudo deste paradigma utilizado.

Referências

- ACM. ACM Digital Library. 2019. Acesso em: 01/10/2019. Disponível em: <https://dl.acm.org>. Citado na página 27.
- AHISSAR, M.; HOCHSTEIN, S. The reverse hierarchy theory of visual perceptual learning. Trends in cognitive sciences, Elsevier, v. 8, n. 10, p. 457–464, 2004. Citado na página 17.
- ALAM, S. S.; JIANU, R. Analyzing eye-tracking information in visualization and data space: From where on the screen to what on the screen. IEEE Transactions on Visualization and Computer Graphics, v. 23, n. 5, p. 1492–1505, May 2017. ISSN 1077-2626. Citado 2 vezes nas páginas 33 e 51.
- AMERICAN PSYCHIATRIC ASSOCIATION. Manual Diagnóstico e Estatístico de Transtornos Mentais. 2014. ISBN 9788582711835. Disponível em: <https://books.google.com.br/books?id=QL4rDAAAQBAJ>. Citado 2 vezes nas páginas 17 e 22.
- ANDERSON, N. C.; ANDERSON, F.; KINGSTONE, A.; BISCHOF, W. F. A comparison of scanpath comparison methods. Behav Res Methods, v. 47, n. 4, p. 1377–1392, dez. 2015. Citado na página 33.
- ASSOCIATION, A. P. Manual diagnóstico e estatístico de transtornos mentais: DSM-5. Artmed, 2014. ISBN 978-85-8271-089-0. Disponível em: <https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596>. Citado na página 17.
- BARANES, A.; OUDEYER, P.-Y.; GOTTLIEB, J. Eye movements reveal epistemic curiosity in human observers. Vision Res., v. 117, p. 81–90, dez. 2015. Citado 2 vezes nas páginas 31 e 39.
- BARRETO, A. M. EYE TRACKING COMO MÉTODO DE INVESTIGAÇÃO APLICADO ÀS CIÊNCIAS DA COMUNICAÇÃO. Dissertação (Mestrado) — Universidade Nova de Lisboa, 2012. Citado 2 vezes nas páginas 19 e 24.
- BEDFORD, R.; ELSABBAGH, M.; GLIGA, T.; PICKLES, A.; SENJU, A.; CHARMAN, T.; JOHNSON, M. H.; TEAM the B. Precursors to social and communication difficulties in infants at-risk for autism: Gaze following and attentional engagement. Journal of Autism and Developmental Disorders, v. 42, n. 10, p. 2208–2218, out. 2012. ISSN 1573-3432. Disponível em: <https://doi.org/10.1007/s10803-012-1450-y>. Citado 3 vezes nas páginas 19, 23 e 73.
- BILL, G.; WHYTE, E.; GRIFFIN, J. W.; SCHERF, K. S. Measuring sensitivity to eye gaze cues in naturalistic scenes: Presenting the eye gaze focus database. International Journal of Methods in Psychiatric Research, Wiley Online Library, p. e1833, 2020. Citado na página 17.
- BILLECI, L.; NARZISI, A.; CAMPATELLI, G.; CRIFACI, G.; CALDERONI, S.; GAGLIANO, A.; CALZONE, C.; COLOMBI, C.; PIOGGIA, G.; MURATORI, F.; RASO, R.; RUTA, L.; ROSSI, I.; BALLARANI, A.; FULCERI, F.; DARINI, A.; MAROSCIA, E.; LATTARULO, C.; TORTORELLA, G.; SIRACUSANO, R.; COMMINIELLO, V.; GROUP, A. L. E. R. T. Disentangling the initiation from the response in joint attention: an eye-tracking study in toddlers with autism spectrum disorders. Translational

Psychiatry, v. 6, n. 5, p. e808–e808, maio 2016. ISSN 2158-3188. Disponível em: <https://doi.org/10.1038/tp.2016.75>. Citado 3 vezes nas páginas 18, 19 e 23.

BLASCHECK, T.; KURZHALS, K.; RASCHKE, M.; STROHMAIER, S.; WEISKOPF, D.; ERTL, T. Aoi hierarchies for visual exploration of fixation sequences. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA: ACM, 2016. (ETRA '16), p. 111–118. ISBN 978-1-4503-4125-7. Disponível em: <http://doi.acm.org/10.1145/2857491.2857524>. Citado 4 vezes nas páginas 33, 36, 46 e 48.

BODALA, I. P.; ABBASI, N. I.; SUN, Y.; BEZERIANOS, A.; AL-NASHASH, H.; THAKOR, N. V. Measuring vigilance decrement using computer vision assisted eye tracking in dynamic naturalistic environments. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S.l.: s.n.], 2017. p. 2478–2481. ISSN 1558-4615. Citado 4 vezes nas páginas 32, 37, 48 e 51.

BODALA, I. P.; KE, Y.; MIR, H.; THAKOR, N. V.; AL-NASHASH, H. Cognitive workload estimation due to vague visual stimuli using saccadic eye movements. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. [S.l.: s.n.], 2014. p. 2993–2996. ISSN 1094-687X. Citado 4 vezes nas páginas 25, 32, 40 e 48.

BORJI, A.; TAVAKOLI, H. R.; SIHITE, D. N.; ITTI, L. Analysis of scores, datasets, and models in visual saliency prediction. In: 2013 IEEE International Conference on Computer Vision. [S.l.: s.n.], 2013. p. 921–928. ISSN 1550-5499. Citado 4 vezes nas páginas 33, 40, 43 e 51.

BRADSKI, G. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000. Citado na página 60.

CARPENTER, M.; NAGELL, K.; TOMASELLO, M. F. The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data. Monogr Soc Res Child Dev., v. 63(4), p. i–vi – 1–143, 1998. ISSN 0037-976X. Citado na página 23.

CERCENELLI, L.; TIBERI, G.; CORAZZA, I.; GIANNACCARE, G.; FRESINA, M.; MARCELLI, E. SacLab: A toolbox for saccade analysis to increase usability of eye tracking systems in clinical ophthalmology practice. Comput. Biol. Med., v. 80, p. 45–55, 01 2017. Citado 4 vezes nas páginas 31, 37, 40 e 47.

CHAKRABARTI, B.; HAFHEY, A.; CANZANO, L.; TAYLOR, C. P.; MCSORLEY, E. Individual differences in responsivity to social rewards: Insights from two eye-tracking tasks. PLoS ONE, v. 12, n. 10, p. e0185146, 2017. Citado na página 31.

CHAMBE-ENG, E.; NORD, H. https://wiki.qt.io/About_Qt. 1994. Citado na página 55.

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 23.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: Synthetic minority over-sampling technique. J. Artif. Int. Res., AI Access Foundation, El

Segundo, CA, USA, v. 16, n. 1, p. 321–357, jun. 2002. ISSN 1076-9757. Citado 2 vezes nas páginas 66 e 67.

CHOI, L. K.; CORMACK, L. K.; BOVIK, A. C. On the visibility of flicker distortions in naturalistic videos. In: 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). [S.l.: s.n.], 2013. p. 164–169. Citado na página 33.

COUTROT, A.; HSIAO, J. H.; CHAN, A. B. Scanpath modeling and classification with hidden markov models. Behav Res Methods, v. 50, n. 1, p. 362–379, 02 2018. Disponível em: http://antoinecoutrot.magix.net/public/index.htmlhttp://antoinecoutrot.magix.net/public/assets/phd_machine_learning_trajectories.pdf. Citado 5 vezes nas páginas 33, 38, 39, 40 e 43.

DARWISH, A.; PASQUIER, M. Biometric identification using the dynamic features of the eyes. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). [S.l.: s.n.], 2013. p. 1–6. Citado 7 vezes nas páginas 32, 34, 37, 38, 39, 45 e 51.

DENG, T.; YANG, K.; LI, Y.; YAN, H. Where does the driver look? top-down-based saliency detection in a traffic driving environment. IEEE Transactions on Intelligent Transportation Systems, v. 17, n. 7, p. 2051–2062, July 2016. ISSN 1524-9050. Citado 4 vezes nas páginas 32, 43, 46 e 51.

DUMITRIU, T.; CÎMPANU, C.; UNGUREANU, F.; MANTA, V. Experimental analysis of emotion classification techniques. In: 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP). [S.l.: s.n.], 2018. p. 63–70. Citado 2 vezes nas páginas 31 e 45.

FALCK-YTTER, T.; HOFSTEN, C. von. Chapter 12 - how special is social looking in asd: A review. Elsevier, v. 189, p. 209 – 222, 2011. ISSN 0079-6123. Disponível em: <http://www.sciencedirect.com/science/article/pii/B9780444538840000269>. Citado na página 18.

FAWCETT, T. An introduction to roc analysis. Pattern Recognition Letters, v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S016786550500303X>. Citado na página 41.

FROST-KARLSSON, M.; GALAZKA, M. A.; GILLBERG, C.; GILLBERG, C.; MINISCALCO, C.; BILLSTEDT, E.; HADJIKHANI, N.; JOHNELS, J. Å. Social scene perception in autism spectrum disorder: An eye-tracking and pupillometric study. Journal of clinical and experimental neuropsychology, Taylor & Francis, v. 41, n. 10, p. 1024–1032, 2019. Citado na página 17.

GOMES, P. T.; LIMA, L. H.; BUENO, M. K.; ARAÚJO, L. A.; SOUZA, N. M. Autism in brazil: a systematic review of family challenges and coping strategies. Jornal de Pediatria, v. 91, n. 2, p. 111 – 121, 2015. ISSN 0021-7557. Disponível em: <http://www.sciencedirect.com/science/article/pii/S002175571400165X>. Citado na página 17.

GOSCHY, H.; KOCH, A. I.; MÜLLER, H. J.; ZEHETLEITNER, M. Early top-down control over saccadic target selection: Evidence from a systematic salience difference

manipulation. Atten Percept Psychophys, v. 76, n. 2, p. 367–82, fev. 2014. Citado na página 33.

HAJI-ABOLHASSANI, A.; CLARK, J. J. An inverse yarbush process: predicting observers' task from eye movement patterns. Vision Res., v. 103, p. 127–42, out. 2014. Citado 3 vezes nas páginas 34, 37 e 39.

HAYES, T. R.; HENDERSON, J. M. Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. PLoS ONE, v. 13, n. 5, p. e0196654, 2018. Citado 3 vezes nas páginas 32, 38 e 45.

HEPACH, R.; HEDLEY, D.; NUSKE, H. J. Prosocial attention in children with and without autism spectrum disorder: Dissociation between anticipatory gaze and internal arousal. Journal of abnormal child psychology, Springer, v. 48, n. 4, p. 589–605, 2020. Citado na página 19.

HOFSTEN, C. von; UHLIG, H.; ADELL, M.; KOCHUKHOVA, O. How children with autism look at events. Research in Autism Spectrum Disorders, Elsevier, v. 3, n. 2, p. 556–569, 2009. Citado 2 vezes nas páginas 19 e 23.

HOSSEINKHANI, J.; JOSLIN, C. Significance of bottom-up attributes in video saliency detection without cognitive bias. In: 2018 IEEE 17th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC). [S.l.: s.n.], 2018. p. 606–613. Citado 2 vezes nas páginas 32 e 51.

HOU, Y.; XIAO, T.; ZHANG, S.; JIANG, X.; LI, X.; HU, X.; HAN, J.; GUO, L.; MILLER, L. S.; NEUPERT, R.; LIU, T. Predicting movie trailer viewer's "like/dislike" via learned shot editing patterns. IEEE Transactions on Affective Computing, v. 7, n. 1, p. 29–44, Jan 2016. ISSN 1949-3045. Citado 2 vezes nas páginas 34 e 40.

HUDSON, M.; BURNETT, H. G.; JELLEMA, T. Anticipation of action intentions in autism spectrum disorder. Journal of autism and developmental disorders, Springer, v. 42, n. 8, p. 1684–1693, 2012. Citado na página 19.

IEEE. Two modifications of cnn. IEEE Transactions on Systems, Man, and Cybernetics, SMC-6, n. 11, p. 769–772, 1976. Citado na página 67.

IEEE. IEEE. 2019. Acesso em: 01/10/2019. Disponível em: <https://ieeexplore.ieee.org>. Citado na página 27.

ITTI, L.; KOCH, C. Computational modelling of visual attention. Nature Reviews Neuroscience, v. 2, n. 3, p. 194–203, mar. 2001. ISSN 1471-0048. Disponível em: <https://doi.org/10.1038/35058500>. Citado na página 26.

JANSSON, D.; MEDVEDEV, A. Parametric and non-parametric stochastic anomaly detection in analysis of eye-tracking data. In: 52nd IEEE Conference on Decision and Control. [S.l.: s.n.], 2013. p. 2532–2537. ISSN 0191-2216. Citado 2 vezes nas páginas 31 e 44.

JANSSON, D.; ROSÉN, O.; MEDVEDEV, A. Non-parametric analysis of eye-tracking data by anomaly detection. In: 2013 European Control Conference (ECC). [S.l.: s.n.], 2013. p. 632–637. Citado 2 vezes nas páginas 31 e 44.

JOSEPHSON, S.; HOLMES, M. E. Have you seen any of these men?: Looking at whether eyewitnesses use scanpaths to recognize suspects in photo lineups. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. New York, NY, USA: ACM, 2010. (ETRA '10), p. 49–52. ISBN 978-1-60558-994-7. Disponível em: <http://doi.acm.org/10.1145/1743666.1743677>. Citado 2 vezes nas páginas 32 e 45.

JUST, M. A.; CARPENTER, P. A. Eye fixations and cognitive processes. Cognitive Psychology, v. 8, n. 4, p. 441 – 480, 1976. ISSN 0010-0285. Disponível em: <http://www.sciencedirect.com/science/article/pii/0010028576900153>. Citado na página 24.

JUST, M. A.; CARPENTER, P. A. The role of eye-fixation research in cognitive psychology. Behavior Research Methods & Instrumentation, v. 8, n. 2, p. 139–143, Mar 1976. ISSN 1554-3528. Disponível em: <https://doi.org/10.3758/BF03201761>. Citado na página 24.

KLIN, A. Autismo e síndrome de Asperger: uma visão geral. Brazilian Journal of Psychiatry, scielo, v. 28, p. s3 – s11, 05 2006. ISSN 1516-4446. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-44462006000500002&nrm=iso. Citado na página 22.

KLIN, A.; LIN, D. J.; GORRINDO, P.; RAMSAY, G.; JONES, W. 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 18.

KOEHRSEN, W. Hyperparameter Tuning the Random Forest in Python. 2018. Disponível em: <https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>. Citado na página 67.

KROGH-JESPERSEN, S.; KALDY, Z.; VALADEZ, A. G.; CARTER, A. S.; WOODWARD, A. L. Goal prediction in 2-year-old children with and without autism spectrum disorder: An eye-tracking study. Autism Research, Wiley Online Library, v. 11, n. 6, p. 870–882, 2018. Citado na página 24.

KUMAR, D.; DUTTA, A.; DAS, A.; LAHIRI, U. Smarteye: Developing a novel eye tracking system for quantitative assessment of oculomotor abnormalities. IEEE Transactions on Neural Systems and Rehabilitation Engineering, v. 24, n. 10, p. 1051–1059, Oct 2016. ISSN 1534-4320. Citado 5 vezes nas páginas 25, 31, 46, 48 e 51.

KURZHALS, K.; BOPP, C. F.; BÄSSLER, J.; EBINGER, F.; WEISKOPF, D. Benchmark data for evaluating visualization and analysis techniques for eye tracking for video stimuli. In: Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization. New York, NY, USA: ACM, 2014. (BELIV '14), p. 54–60. ISBN 978-1-4503-3209-5. Disponível em: <http://doi.acm.org/10.1145/2669557.2669558>. Citado na página 34.

KURZHALS, K.; HEIMERL, F.; WEISKOPF, D. Iseecube: Visual analysis of gaze data for video. ACM, New York, NY, USA, p. 351–358, 2014. Disponível em: <http://doi.acm.org/10.1145/2578153.2628812>. Citado 2 vezes nas páginas 34 e 41.

- KURZHALS, K.; HLAWATSCH, M.; BURCH, M.; WEISKOPF, D. Fixation-image charts. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA: ACM, 2016. (ETRA '16), p. 11–18. ISBN 978-1-4503-4125-7. Disponível em: <http://doi.acm.org/10.1145/2857491.2857507>. Citado 2 vezes nas páginas 34 e 36.
- KURZHALS, K.; HLAWATSCH, M.; HEIMERL, F.; BURCH, M.; ERTL, T.; WEISKOPF, D. Gaze stripes: Image-based visualization of eye tracking data. IEEE Transactions on Visualization and Computer Graphics, v. 22, n. 1, p. 1005–1014, Jan 2016. ISSN 1077-2626. Citado 3 vezes nas páginas 34, 36 e 41.
- LARSSON, L.; NYSTROM, M.; STRIDH, M. Discrimination of fixations and smooth pursuit movements in high-speed eye-tracking data. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. [S.l.: s.n.], 2014. p. 3797–3800. ISSN 1094-687X. Citado 7 vezes nas páginas 25, 33, 34, 37, 38, 40 e 51.
- Larsson, L.; Nyström, M.; Stridh, M. Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. IEEE Transactions on Biomedical Engineering, v. 60, n. 9, p. 2484–2493, Sep. 2013. ISSN 1558-2531. Citado na página 25.
- LEE BARRY B., S. H. Contrast sensitivity and retinal ganglion cell responses in the primate. Psychology & Neuroscience, scielo, v. 4, p. 11–18, 06 2011. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1983-32882011000100003&lang=pt. Citado 3 vezes nas páginas 33, 40 e 49.
- LEE, W.; HUANG, T.; YEH, S.; CHEN, H. H. Learning-based prediction of visual attention for video signals. IEEE Transactions on Image Processing, v. 20, n. 11, p. 3028–3038, Nov 2011. ISSN 1057-7149. Citado 2 vezes nas páginas 37 e 42.
- LI, B.; BARNEY, E.; HUDAC, C.; NUECHTERLEIN, N.; VENTOLA, P.; SHAPIRO, L.; SHIC, F. Selection of eye-tracking stimuli for prediction by sparsely grouped input variables for neural networks: Towards biomarker refinement for autism. In: ACM Symposium on Eye Tracking Research and Applications. New York, NY, USA: Association for Computing Machinery, 2020. (ETRA '20 Full Papers). ISBN 9781450371339. Disponível em: <https://doi.org/10.1145/3379155.3391334>. Citado 2 vezes nas páginas 80 e 81.
- LI, F.; BAI, H.; ZHAO, Y. Visual attention guided eye movements for 360 degree images. In: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). [S.l.: s.n.], 2017. p. 506–511. Citado 2 vezes nas páginas 33 e 47.
- LIANG, H.; LIANG, R.; SUN, G. Looking into saliency model via space-time visualization. IEEE Transactions on Multimedia, v. 18, n. 11, p. 2271–2281, Nov 2016. ISSN 1520-9210. Citado 7 vezes nas páginas 33, 36, 37, 41, 43, 46 e 48.
- Little, G. E.; Bonnar, L.; Kelly, S. W.; Lohan, K. S.; Rajendran, G. Gaze contingent joint attention with an avatar in children with and without asd. In: 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). [S.l.: s.n.], 2016. p. 15–20. Citado na página 19.
- LIU, W.; LI, M.; YI, L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism research : official

journal of the International Society for Autism Research, v. 9, 04 2016. Citado 2 vezes nas páginas 80 e 81.

LORD, C.; BRUGHA, T. S.; CHARMAN, T.; CUSACK, J.; DUMAS, G.; FRAZIER, T.; JONES, E. J.; JONES, R. M.; PICKLES, A.; STATE, M. W. *et al.* Autism spectrum disorder. Nature reviews Disease primers, Nature Publishing Group, v. 6, n. 1, p. 1–23, 2020. Citado na página 17.

LÖNNQVIST, L.; LOUKUSA, S.; HURTIG, T.; MÄKINEN, L.; SIPO, A.; VÄYRYNEN, E.; PALO, P.; LAUKKA, S.; MÄMMELÄ, L.; MATTILA, M.-L.; EBELING, H. How young adults with autism spectrum disorder watch and interpret pragmatically complex scenes. Q J Exp Psychol (Hove), v. 70, n. 11, p. 2331–2346, nov. 2017. Citado 4 vezes nas páginas 32, 44, 50 e 51.

MANUAL de Orientação: Departamento científico de pediatria do desenvolvimento e comportamento sbp. Rio de Janeiro: [s.n.], 2019. Citado 3 vezes nas páginas 18, 22 e 23.

MARIGHETTO, P.; COUTROT, A.; RICHE, N.; GUYADER, N.; MANCAS, M.; GOSSELIN, B.; LAGANIERE, R. Audio-visual attention: Eye-tracking dataset and analysis toolbox. In: 2017 IEEE International Conference on Image Processing (ICIP). [S.l.: s.n.], 2017. p. 1802–1806. ISSN 2381-8549. Citado 2 vezes nas páginas 33 e 51.

OHL, S.; KLIEGL, R. Revealing the time course of signals influencing the generation of secondary saccades using Aalen’s additive hazards model. Vision Res., v. 124, p. 52–8, 07 2016. Citado 2 vezes nas páginas 31 e 41.

OHL, S.; ROLFS, M. Saccadic eye movements impose a natural bottleneck on visual short-term memory. J Exp Psychol Learn Mem Cogn, v. 43, n. 5, p. 736–748, maio 2017. Citado 2 vezes nas páginas 31 e 47.

OLIVEIRA, J. d. S. d. Classificador para auxílio ao diagnóstico de TEA baseado em um modelo computacional de atenção visual. Dissertação (Mestrado) — 2018. Dissertação (Mestrado em Sistemas de Informação) - Escola de Artes, Ciências e Humanidades, University of São Paulo, 2017. Citado 2 vezes nas páginas 21 e 54.

ORSATI FERNANDA TEBEXRENI, M. T. S. J. S. M. E. C. d. Percepção de faces em crianças e adolescentes com Transtorno Invasivo do Desenvolvimento. Paidéia (Ribeirão Preto), scielo, v. 19, p. 349–356, 12 2009. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-863X2009000300009&lang=pt. Citado 2 vezes nas páginas 32 e 44.

OTERO-MILLAN, J.; CASTRO, J. L. A.; MACKNIK, S. L.; MARTINEZ-CONDE, S. Unsupervised clustering method to detect microsaccades. J Vis, v. 14, n. 2, fev. 2014. Citado 4 vezes nas páginas 31, 40, 42 e 45.

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 17.

PALUMBO, L.; BURNETT, H. G.; JELLEMA, T. Atypical emotional anticipation in high-functioning autism. Molecular autism, Springer, v. 6, n. 1, p. 47, 2015. Citado na página 19.

PARISH-MORRIS, J.; PALLATHRA, A. A.; FERGUSON, E.; MADDOX, B. B.; POMYKACZ, A.; PEREZ, L. S.; BATEMAN, L.; PANDEY, J.; SCHULTZ, R. T.; BRODKIN, E. S. Adaptation to different communicative contexts: an eye tracking study of autistic adults. *Journal of Neurodevelopmental Disorders*, v. 11, n. 1, p. 5, abr. 2019. ISSN 1866-1955. Disponível em: <https://doi.org/10.1186/s11689-019-9265-1>. Citado 3 vezes nas páginas 17, 19 e 73.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. *et al.* Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011. Citado na página 67.

PFLÜGER, H.; HÖFERLIN, B.; RASCHKE, M.; ERTL, T. Simulating fixations when looking at visual arts. *ACM Trans. Appl. Percept.*, ACM, New York, NY, USA, v. 12, n. 3, p. 9:1–9:20, jun. 2015. ISSN 1544-3558. Disponível em: <http://doi.acm.org/10.1145/2736286>. Citado 3 vezes nas páginas 33, 36 e 51.

PUBMED. PubMed. 2019. Acesso em: 01/10/2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/>. Citado na página 27.

R-TAVAKOLI, H.; ATYABI, A.; RANTANEN, A.; LAUKKA, S. J.; NEFTI-MEZIANI, S.; HEIKKILÄ, J. Predicting the Valence of a Scene from Observers' Eye Movements. *PLoS ONE*, v. 10, n. 9, p. e0138198, 2015. Citado 4 vezes nas páginas 34, 38, 39 e 62.

RICHARDSON, D.; SPIVEY, M. Eye-tracking: Characteristics and methods. 05 2008. Citado na página 25.

RIGAS, I.; ECONOMOU, G.; FOTOPOULOS, S. Human eye movements as a trait for biometrical identification. In: *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. [S.l.: s.n.], 2012. p. 217–222. Citado 4 vezes nas páginas 32, 36, 45 e 51.

ROSSUM, G. V.; JR, F. L. D. Python reference manual. [S.l.]: Centrum voor Wiskunde en Informatica Amsterdam, 1995. Citado na página 65.

RÖHRBEIN, F.; GODDARD, P.; SCHNEIDER, M.; JAMES, G.; GUO, K. How does image noise affect actual and predicted human gaze allocation in assessing image quality? *Vision Res.*, v. 112, p. 11–25, jul. 2015. Citado 3 vezes nas páginas 31, 36 e 42.

SAAVEDRA-PEÑA, G.; LAI, H.; SZE, V.; HELDT, T. Determination of saccade latency distributions using video recordings from consumer-grade devices. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.: s.n.], 2018. p. 953–956. ISSN 1558-4615. Citado na página 31.

SAEED, U. Automatic person recognition using eye movement during scene understanding. In: *17th IEEE International Multi Topic Conference 2014*. [S.l.: s.n.], 2014. p. 240–244. Citado 8 vezes nas páginas 25, 32, 34, 37, 38, 45, 51 e 54.

SAEED, U. Automatic person recognition using eye movement during scene understanding. In: *17th IEEE International Multi Topic Conference 2014*. [S.l.: s.n.], 2014. p. 240–244. Citado na página 25.

- SARANYA, C.; MANIKANDAN, G. A study on normalization techniques for privacy preserving data mining. v. 5, p. 2701–2704, 06 2013. Citado na página 66.
- SCIELO. Scientific Electronic Library Online. 2019. Acesso em: 01/10/2019. Disponível em: <https://scielo.org/>. Citado na página 27.
- SHAHIMIN, M. M.; SAAD, N.; SHARANJEET-KAUR; SIVAJI, A.; TZUAAN, S. S.; KHEAN, C. N. Online sunglasses purchasing: Where do people look? In: 2014 3rd International Conference on User Science and Engineering (i-USEr). [S.l.: s.n.], 2014. p. 248–252. Citado 3 vezes nas páginas 33, 46 e 47.
- SHIHAB, A. I.; DAWOOD, F. A.; KASHMAR, A. H. Data analysis and classification of autism spectrum disorder using principal component analysis. Advances in Bioinformatics, Hindawi, v. 2020, 2020. Citado na página 18.
- SOGO, H. Sgttoolbox: Utility for controlling SimpleGazeTracker from Psychtoolbox. Behav Res Methods, v. 49, n. 4, p. 1323–1332, 08 2017. Citado 2 vezes nas páginas 33 e 49.
- SRIVASTAVA, N.; AGRAWAL, U.; ROY, S. K.; TIWARY, U. S. Human identification using linear multiclass svm and eye movement biometrics. In: 2015 Eighth International Conference on Contemporary Computing (IC3). [S.l.: s.n.], 2015. p. 365–369. Citado 2 vezes nas páginas 32 e 40.
- THIBODEAUX, D.; MUHAMMAD, H.; COSTA, S. L.; ALVAREZ, T. L. Instrumentation to stimulate gap saccades, pro-saccades, and overlap saccades. In: 2014 40th Annual Northeast Bioengineering Conference (NEBEC). [S.l.: s.n.], 2014. p. 1–2. ISSN 2160-6986. Citado 2 vezes nas páginas 31 e 45.
- TOBII. Tobii, What is eye tracking? 2019. <https://www.tobiipro.com/blog/what-is-eye-tracking/>. Accessed: 2019-10-01. Citado na página 24.
- VENUGOPAL, D.; AMUDHA, J.; JYOTSNA, C. Developing an application using eye tracker. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT). [S.l.: s.n.], 2016. p. 1518–1522. Citado 3 vezes nas páginas 32, 45 e 46.
- WANG, H. X.; YUVAL-GREENBERG, S.; HEEGER, D. J. Suppressive interactions underlying visually evoked fixational saccades. Vision Res., v. 118, p. 70–82, jan. 2016. Citado 2 vezes nas páginas 31 e 47.
- WANG, S.; JIANG, M.; DUCHESNE, X. M.; LAUGESON, E. A.; KENNEDY, D. P.; ADOLPHS, R.; ZHAO, Q. Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. Neuron, v. 88, n. 3, p. 604–16, nov. 2015. Citado 4 vezes nas páginas 31, 40, 43 e 46.
- WIECKOWSKI, A. T.; WHITE, S. W. Eye-Gaze Analysis of Facial Emotion Recognition and Expression in Adolescents with ASD. J Clin Child Adolesc Psychol, v. 46, n. 1, p. 110–124, 2017. Citado 3 vezes nas páginas 32, 44 e 50.
- YANEVA, V.; HA, L. A.; ERASLAN, S.; YESILADA, Y.; MITKOV, R. Detecting autism based on eye-tracking data from web searching tasks. In: Proceedings of the Internet of Accessible Things. New York, NY, USA: Association for

Computing Machinery, 2018. (W4A '18). ISBN 9781450356510. Disponível em: <https://doi.org/10.1145/3192714.3192819>. Citado 2 vezes nas páginas 80 e 81.

YIU, T. Understanding Random Forest. How the Algorithm Works and Why it Is So Effective. 2020. Disponível em: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. Citado na página 65.

ZHANG, X.; YUAN, S.; CHEN, M.; LIU, X. A complete system for analysis of video lecture based on eye tracking. IEEE Access, v. 6, p. 49056–49066, 2018. ISSN 2169-3536. Citado 3 vezes nas páginas 33, 35 e 47.

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 Bétio, 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 zozura.

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

Telefone: (11)3091-1046

CEP: 03.828-000

E-mail: cep-each@usp.br

ESCOLA DE ARTES, CIÊNCIAS
E HUMANIDADES - EACH/USP

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 Bétio, nº 1000

Bairro: Ermelino Matarazzo

CEP: 03.828-000

UF: SP

Município: SAO PAULO

Telefone: (11)3091-1046

E-mail: cep-each@usp.br

Anexo B – 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. 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 Bértio, 1000 - Ermelino Matarazzo, São Paulo - SP, 03828-000 - T14 - I1 – tel:(11) 3091-1046; e-mail: 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 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 (.....).....