CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC	
(X) PRÉ-PROJETO ( ) PROJETO	ANO/SEMESTRE: 2022/2

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA E VISÃO COMPUTACIONAL PARA ANÁLISE DO COMPORTAMENTO INFANTIL EM AMBIENTE ESCOLAR

Hélio Potelicki

Prof. Andreza Sartori – Orientador(a)

Gabriel Barreto Alberton – Coorientador(a) Externo

1 INTRODUÇÃO

A utilização do poder computacional para uma melhor compreensão dos comportamentos humanos tem um grande potencial em aplicações de diversos domínios, incluindo a saúde. No entanto, analisar de maneira autônoma o comportamento humano é desafiador, uma vez que estes são contextuais e na maior parte das vezes sociais, ou seja, em relação com outras pessoas, tornando necessária uma maior compreensão da interação humana para entender o comportamento de um indivíduo (WEI *et al.,* 2022). Uma área importante na aplicação da análise comportamental computacional utilizando Inteligência Artificial (IA), é a caracterização do comportamento e das mudanças de desenvolvimento em crianças diagnosticadas com Transtornos do Espectro do Autismo (TEA). Os TEA são um grupo de transtornos do neurodesenvolvimento ao longo da vida, caracterizados por prejuízos na comunicação e interações sociais (KOJOVIC *et al*., 2021).

Em um ambiente escolar, apenas a observação de um professor para avaliar o comportamento individual pode ser demorada, já com o uso da IA, o professor pode fazer uma melhor análise do comportamento do aluno, compreendendo quais são suas dificuldades (GAROFALO, 2019). Em comparação às máquinas, os olhos humanos não detectam mudanças sutis de comportamento e microexpressões. Estima-se que 90% dos brasileiros com autismo não tenham sido diagnosticados (VADASZ, 2013).

Trabalhos recentes como os apresentados por Kojovic *et al*. (2021) e Sayed *et al*. (2019) obtiveram excelentes resultados utilizando técnicas de aprendizado de máquina e visão computacional no desenvolvimento de protótipos para a coleta e tratamento de dados, e na utilização de estimativa de pose baseada em vídeo 2D para uma previsão autônoma de transtornos do espectro do autismo em crianças pequenas. No protótipo proposto neste trabalho, pretende-se avançar mais o uso de estimativa de pose, através de treinamentos e aprimoramentos nas Redes Neurais Artificiais utilizando dados coletados de câmeras em ambientes escolares, posicionadas nas salas de aula com a parceria de escolas dispostas a contribuir com o projeto. Deste modo, este trabalho propõe o desenvolvimento de uma ferramenta que, a partir de imagens de câmeras em um ambiente escolar, compute os dados de cada criança analisando a série temporal, gerada da estimativa de pose, buscando por anomalias comportamentais, através de algoritmos de aprendizado de máquina. Esta extensão do estado da arte possibilita que crianças neurodivergentes consigam ser diagnosticadas o quanto antes.

1.1 OBJETIVOS

Este trabalho tem como objetivo disponibilizar um protótipo para a análise comportamental infantil em ambiente escolar por meio de técnicas de Aprendizado de Máquina e Visão Computacional para auxiliar no diagnóstico prematuro de possíveis transtornos em crianças.

Os objetivos específicos são:

a) coletar por meio de vídeos a movimentação dos alunos para a construção da base de dados;
b) identificar padrões de Transtornos do Espectro do Autismo dos alunos, por meio da estimativa de pose, durante o período de aula;
c) analisar as séries históricas de interação e distanciamento de cada aluno e suas implicações (sociabilidade);
d) disponibilizar os dados dos alunos de maneira que ajude o professor a entender suas dificuldades.

2 TRABALHOS CORRELATOS

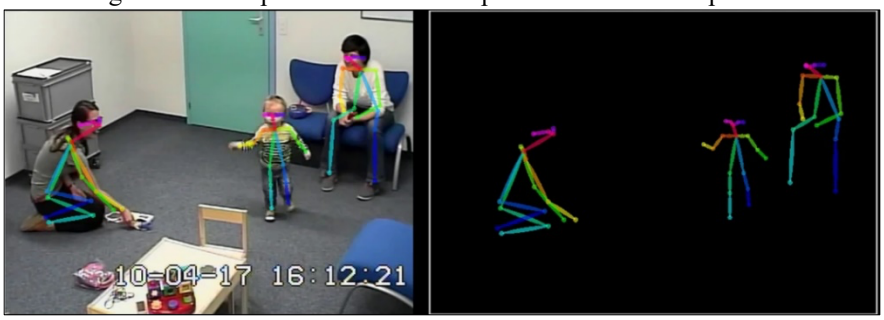
Esta seção contém trabalhos com características semelhantes aos principais objetivos do estudo proposto. Kojovic *et al.* (2021) utilizou um modelo de Aprendizado de Máquina (AM) para estimar a pose de uma pessoa e utilizar na previsão de TEA em crianças. O trabalho de Sayed *et al.* (2019) propõe o uso de técnicas de visão computacional e a captura de movimentos para avaliação cognitiva em crianças. Por fim, o terceiro correlato busca classificar por meio de Redes Neurais Artificiais, os comportamentos relacionados ao autismo analisando as relações entre quadros de um vídeo (WEI *et al.*, 2022).

2.1 USING 2D VIDEO-BASED POSE ESTIMATION FOR AUTOMATED PREDICTION OF AUTISM SPECTRUM DISORDERS IN YOUNG CHILDREN

O trabalho de Kojovic *et al*. (2021), tem como objetivo ajudar na identificação dos Transtornos do Espectro do Autismo (TEA), que é um grupo de transtornos do neurodesenvolvimento ao longo da vida. Os autores utilizaram o rastreamento de movimento para medir o comportamento de aproximação, evitação e o direcionamento do afeto facial das crianças, durante as avaliações.

Foram utilizados dois algoritmos de aprendizado de máquina, a união entre uma Rede Neural Convolucional (RNC) e uma *Long Short Term Memory* (LSTM) para discriminar entre Desordens do Espectro Autista (DEA) e Desenvolvimento Típico (DT), a partir de vídeos de interações sociais entre uma criança (DEA ou DT) e um adulto. A dimensionalidade dos vídeos de entrada foi reduzida utilizando a tecnologia de estimativa de pose do *framework* OpenPose, para que fosse possível extrair pontos esqueléticos de todas as pessoas presentes nos vídeos, como demonstra a Figura 1.

Figura 1 - Exemplo de estimativa de pose 2D utilizando OpenPose



Fonte: Kojovic *et al.* (2021).

A arquitetura proposta por Kojovic *et al.* (2021), utilizou uma RNC para a extração de características de cada amostra. A saída desta extração passa por uma LSTM sensível ao reconhecimento de ações, com o objetivo de explorar o potencial de interações sociais puramente não verbais para informar a atribuição de classe de diagnóstico automatizado. O conjunto de dados utilizado nos estudos inclui: um conjunto de treinamento de 34 crianças com DT e 34 crianças com TEA, em uma faixa etária de 1 a 5 anos; e um conjunto de validação com 34 crianças com DT e 135 crianças com TEA, com idades de 1 a 7 anos.

O modelo de aprendizado profundo treinado utilizando uma CNN VGG-16 pré treinada distinguiu crianças com TEA de crianças com DT, com uma precisão superior a 80% e com F1-score de 0,818. Essa abordagem traz promessas razoáveis de que uma triagem confiável de TEA baseada em aprendizado de máquina, que pode se tornar uma realidade não muito distante no futuro (KOJOVIC *et al.,* 2021).

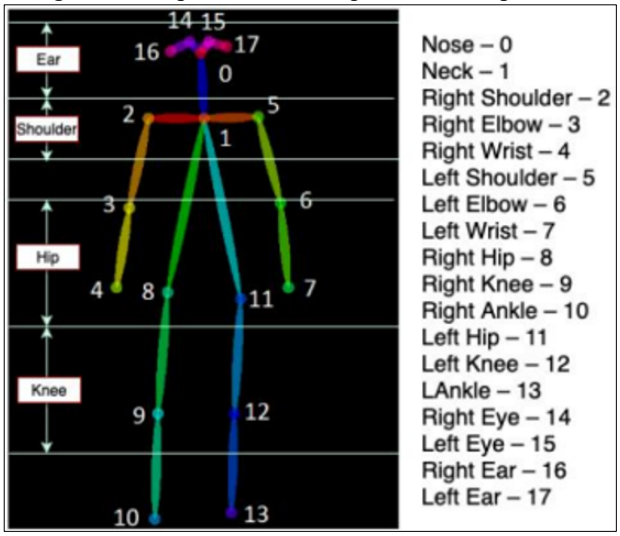
2.2 COGNITIVE ASSESSMENT IN CHILDREN THROUGH MOTION CAPTURE AND COMPUTER VISION: THE CROSS-YOUR-BODY TASK

O protótipo sugerido por Sayed *et al*. (2019), utilizou um método de reconhecimento de atividade humana, baseado em vídeo, para a implantação de um sistema automatizado de avaliação cognitiva infantil. O *Activate Test for Embedded Cognition* (ATEC) busca avaliar crianças executando tarefas físicas e cognitivas, concentrando-se em reconhecimento de atividade onipresente e não intrusivo para movimentos da parte superior do corpo.

Os dados de entrada são baseados em crianças realizando a tarefa *cross-your-body* – técnica projetada para avaliar o ritmo da criança, executando movimentos seguidos alternando os lados de seu corpo (WEI *et al.,* 2022). O conjunto de dados inclui 15 crianças realizando 8 tipos de atividades, resultando em 1900 amostras de vídeo anotadas. Foi utilizado para a localização e mapeamento de articulações nas imagens em RGB, o *framework* OpenPose, onde cada articulação é representada por um vetor 2D em um espaço de coordenadas cartesianas, como demonstra a Figura 2.

O sistema desenvolvido por Sayed *et al*. (2019), utilizando vídeos de crianças executando a tarefa *cross-your-body*, é capaz de reconhecer a mão ativa que executa os movimentos, estima posições espaciais específicas da mão com uma precisão geral de 89,95%, tornando a extração de recursos mais eficiente.

Figura 2 - Mapeamento do esqueleto com OpenPose

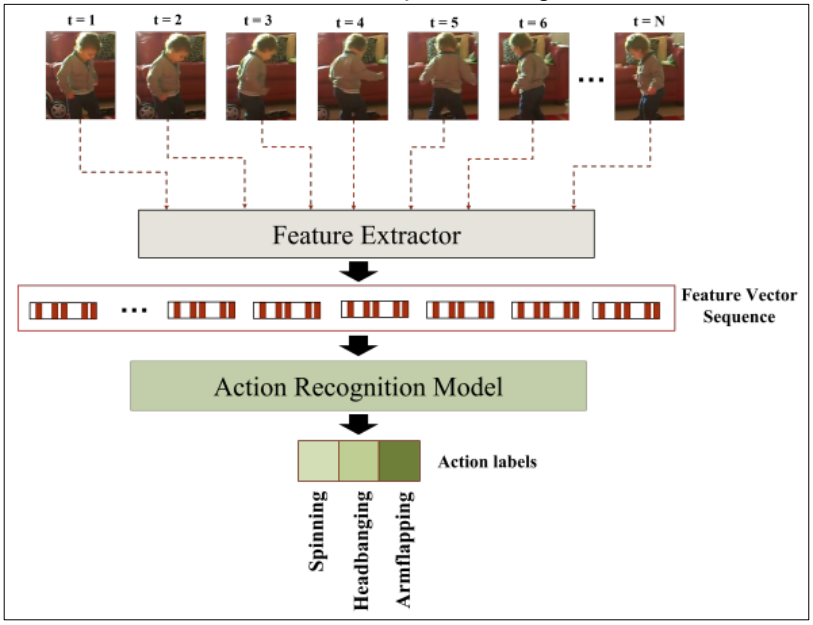


Fonte: Sayed *et al.* (2019).

2.3 VISION-BASED ACTIVITY RECOGNITION IN CHILDREN WITH AUTISM-RELATED BEHAVIORS

O *framework* proposto por Wei *et al*. (2022) apresenta um sistema de visão computacional baseado em região, que visa ajudar os médicos e pais a analisarem o comportamento de crianças. O *framework* é separado em duas partes, como demonstrado na arquitetura da Figura 3. A primeira parte funciona como um extrator de recursos semânticos (*Feature Extractor*) relacionados à ação do quadro apresentado no vídeo. Essa sequência de características é armazenada em um vetor (*Feature Vector Sequence*) que será passado para a segunda parte, onde um modelo de reconhecimento de ações (*Action Labels*), treinado para identificar comportamentos relacionados ao autismo, rotula as ações da criança.

Figura 3 - Reconhecimento de atividade em crianças com comportamentos relacionados ao autismo



Fonte: Wei *et al.* (2022).

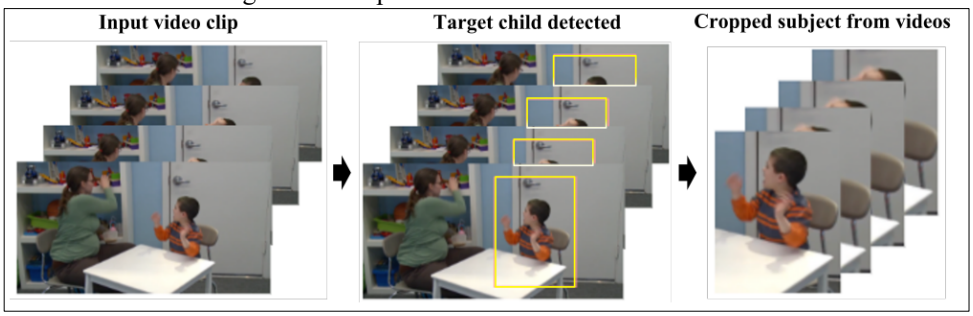
O artigo comparou empiricamente o desempenho de quatro arquiteturas de aprendizado profundo utilizadas no *backbone*, termo para se referir à rede que se encarrega da extração de recursos. Foi realizada uma análise entre as diferentes combinações de *backbone* e modelo de reconhecimento de ação, para a identificação de comportamentos relacionados ao autismo de forma eficiente.

A extração de recursos é uma etapa importante no fluxo de trabalho para o reconhecimento de ação proposto, que permite a extração de recursos informativos dos dados brutos. Também ajuda a reduzir a quantidade de dados redundantes alimentados ao componente de reconhecimento de ação, o que aprimora ainda mais o processo de aprendizado de ação (WEI *et al.,* 2022).

Todos os testes e comparações dos modelos foram realizados utilizando uma versão modificada da base de dados *Self-Stimulatory Behavior Dataset* (SSBD), que se trata de um conjunto de dados de vídeos infantis que exibem comportamentos autoestimulatórios (*stimming*), movimentos que a criança realiza porque ela gosta e sente prazer em realizá-los. Este conjunto de dados está publicamente disponível para uso. Os vídeos desse conjunto de dados são de natureza diversa e gravados em ambientes não controlados. Os vídeos foram retirados de diferentes portais online, como YouTube, Vimeo e Dailymotion. Este conjunto completo conta com 75 vídeos, mas por questões de privacidade somente 60 deles podem ser utilizados. Os vídeos possuem duração média de 90 segundos e são agrupados em três categorias de comportamento de *stimming,* sendo elas: braços batendo; batendo a cabeça; movimentos giratórios.

Antes de passar para o processo de extração de recursos, os vídeos são submetidos a um pré-processamento, utilizando uma biblioteca de visão computacional desenvolvida pelo Facebook, a Detectron2. Esta biblioteca permite criar facilmente modelos de detecção de objetos. O modelo desenvolvido detecta onde a criança está no vídeo separando essa parte do quadro, como demonstra a Figura 4.

Figura 4 - Pré-processamento do vídeo de entrada



Fonte: Wei *et al.* (2022).

Os resultados experimentais mostram que a estrutura proposta pode reconhecer comportamentos relacionados ao autismo com precisão e robustez, obtendo os melhores resultados ao utilizar um modelo *Multi-Stage Temporal Convolutional Network* (MS-TCN), que utiliza uma arquitetura de vários estágios, para realizar a tarefa de segmentação de ação temporal. Utilizando o modelo MS-TCN com o *backbone Efficient Symmetric Network* (ESNet) o modelo alcançou 0.71 de pontuação F1-Score.

3 PROPOSTA DO PROTÓTIPO

Esta seção tem como objetivo apresentar a justificativa para a elaboração desse protótipo, bem como, seus principais requisitos e a metodologia adotada.

3.1 JUSTIFICATIVA

O Quadro 1 explana as diferenças entre os trabalhos correlatos, expondo as principais características de cada um. A primeira linha refere-se ao uso do *framework* OpenPose, ferramenta que vem sendo amplamente utilizada atualmente para estimativa de pose baseada em vídeos 2D. O OpenPose é o primeiro sistema que suporta detecção de pontos-chave do corpo humano em tempo real e multipessoal (PAWANGFG, 2021). O *framework* foi o escolhido por Kojovic *et al*. (2021) e Sayed *et al*. (2019), e utilizado como base para coleta de informações de estimativa de pose. Observa-se que apenas o modelo de Wei *et al.* (2022) segue uma abordagem de classificação de ações características do comportamento de *stimming* sem o uso da estimativa de pose.

Em se tratando de aferição da acurácia, somente Sayed *et al*. (2019) utilizam um método de matriz de confusão, que é uma tabela que permite a visualização do desempenho de um algoritmo de classificação. Os demais correlatos optaram pelo uso do F1-Score, que é uma medida harmônica entre duas métricas, precisão e *recall*. Apesar utilizarem diferentes abordagens para extração de dados dos vídeos, Kojovic *et al*. (2021) e Wei *et al.* (2022), utilizaram arquiteturas *encoder*/*decoder*, ou seja, uma rede pré-treinada que somente cuida da extração de características dos vídeos e outra rede para a classificação.

Wei *et al*. (2022) ainda utilizou uma etapa de pré-processamento das imagens, com o uso da plataforma Detectron2, utilizada para a detecção de objetos. No pré-processamento é realizando a detecção da criança, cortando-a do restante do vídeo, deixando a criança em destaque. Nos resultados apresentados a arquitetura de Kojovic *et al.* (2021) alcançou a medição de F1-Score de 0,81 e Wei *et al*. (2022) com F1-Score de 0,83. Destaca-se que os autores utilizaram dados de entradas diferentes.

Quadro 1 - Comparativo entre os trabalhos correlatos

Características \ Trabalhos Correlatos	Kojovic *et al.* (2021)	Sayed *et al.* (2019)	Wei *et al.* (2022)
Framework utilizado	OpenPose	OpenPose	ESNet + MS-TCN
Métrica de aferição da acurácia	F1-Score	Matriz de confusão	F1-Score
Encoder (Extração de características)	VGG 16	Não	ESNet + TCN
Decoder (Classificador)	LSTM	Não	MS-TCN
Pré-processamento de imagem	Não	Não	Detectron2
Acurácia geral (%)	80	89,95	81,1

Fonte: elaborado pelo autor.

Este protótipo torna-se relevante porque irá possibilitar aos professores e responsáveis, observar os dados de cada criança isoladamente ou em comparativo com as demais crianças da turma, através de uma interface gráfica contendo dados de movimento de cada criança, como demonstra a Figura 5, a disponibilização destes dados em uma interface ainda não foi implementada pelos autores dos correlatos descritos acima. Por fim, além de disponibilizar as informações e gerar uma base de dados da série histórica de cada aluno nos períodos de aprendizado, o protótipo irá contribuir para a identificação prematura de transtornos do neurodesenvolvimento infantil.

Figura 5 - Interface gráfica com dados dos alunos

Fonte: Yujie (2019).

3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

O protótipo proposto deve:

a) permitir ao usuário selecionar qual aluno deseja analisar (Requisito Funcional – RF);
b) permitir o usuário verificar a série histórica do aluno (RF);
c) disponibilizar painéis visuais com informações das crianças em sala de aula para os professores e responsáveis (RF);
d) ser capaz de informar se a criança possuí comportamentos relacionados à transtornos (RF);
e) receber dados em tempo real (RF);
f) ser implementado na linguagem Python (Requisito Não-Funcional – RNF);
g) utilizar técnicas de aprendizado de máquina e visão computacional para a modelagem do modelo preditivo (RNF);
h) armazenar os dados coletados em um banco de dados (RNF).

3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

a) levantamento bibliográfico: pesquisar trabalhos relacionados e materiais sobre técnicas de Aprendizado de Máquina e Visão Computacional para estimativa de pose, Transtorno do Espectro Autista (TEA) e Redes Neurais Artificiais;
b) elicitação de requisitos: reavaliar os requisitos da seção anterior, e especificar outros, mediante a necessidade identificada na revisão bibliográfica;

c) especificação: formalizar a diagramação das classes e dos casos de uso do protótipo;
d) implementação: utilizar a linguagem de programação Python para o desenvolvimento do protótipo;
e) testes: em conjunto com a etapa de implementação, realizar testes do protótipo em ambiente real para a validação dos resultados obtidos, confiabilidade dos dados e performance do protótipo.

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 2 – Cronograma de atividades

	2023									
	ago.		set.		out.		nov.		dez.	
etapas / quinzenas	1	2	1	2	1	2	1	2	1	2
Levantamento bibliográfico										
Elicitação de requisitos										
Especificação										
Implementação										
Testes										

Fonte: elaborado pelo autor.

4 REVISÃO BIBLIOGRÁFICA

Esta seção descreve brevemente os assuntos que irão fundamentar o estudo a ser realizado: utilização de técnicas de aprendizado de máquina e visão computacional, para análise do comportamento infantil e de padrões de transtorno do espectro autista através do uso de estimativa de pose.

Visão computacional é uma área do aprendizado de máquina que busca analisar, interpretar e extrair informações relevantes de imagens e/ou vídeos para que decisões possam ser tomadas, ou para gerar dados relevantes para uma aplicação futura (SALLES, 2022). A tecnologia mais utilizada atualmente na área da visão computacional são as Redes Neurais Artificiais (RNA) (SILVA, 2004). As RNAs são modelos matemáticos que representam os princípios de atividades do cérebro com base na neurobiologia e na teoria do comportamento (SILVA, 2004). Os modelos de RNA são mais adequados quando se tem a necessidade de resolver problemas que envolvem classificação ou predição.

Os transtornos do espectro do autismo (TEA) são um grupo de transtornos do neurodesenvolvimento ao longo da vida caracterizados por prejuízos na comunicação, nas interações sociais e pela presença de padrões de interesses e comportamentos restritos e repetitivos (KOJOVIC *et al.*, 2021). Uma das técnicas de inteligência computacional, comumente usada na tentativa de prever séries temporais, é o treinamento de RNA. Estas são baseadas na arquitetura e aprendizagem do cérebro humano (MANO, 2008).

Estimativa de Pose Humana (EPH) é uma forma de identificar e classificar as articulações do corpo humano. Essencialmente, é uma maneira de capturar um conjunto de coordenadas para cada articulação (braço, cabeça, tronco etc.), que são conhecidos como pontos-chave que podem descrever a pose de uma pessoa (BARLA, 2022). O modelo de RNA proposto por KOJOVIC *et al.* (2021) utilizou a EPH derivada de vídeos contendo a interação social entre uma criança e um adulto e distinguiu de forma robusta se a criança tem ou não tem autismo.

REFERÊNCIAS

BARLA, Nilesh. **A Comprehensive Guide to Human Pose Estimation.** [2022]. Disponível em: https://www.v7labs.com/blog/human-pose-estimation-guide. Acesso em: 26 set. 2022.

GAROFALO, Débora. **Como a inteligência artificial pode colaborar com sua aula.** São Paulo, [2022]. Disponível em: https://novaescola.org.br/conteudo/18312/como-a-inteligencia-artificial-pode-colaborar-com-sua-aula. Acesso em: 20 set. 2022.

KOJOVIC, Nada; NATRAJ, Shreyasvi; MOHANTY, Sharada P. *et al.* **Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children.** 2021. Sci Rep 11, 15069. Disponível em: https://doi.org/10.1038/s41598-021-94378-z. Acesso em: 29 ago. 2022.

MANO, João Pedro. **Previsão de séries temporais mediante redes neurais.** 2008. Disponível em: https://www.puc-rio.br/ensinopesq/ccpg/pibic/relatorio_resumo2009/relatorio/fis/joao_pedro.pdf. Acesso em: 27 set. 2022.

MUELLER, John Paul; MASSARON, Luca. **Machine Learning for Dummies.** Nova Jersey: John Wiley & Sons, Inc 2016. 399 p.

PAWANGFG. **OpenPose: Human Pose Estimation Method.** Uttar Pradesh, [2021]. Disponível em: https://www.geeksforgeeks.org/openpose-human-pose-estimation-method/. Acesso em: 20 set. 2022.

SALLES, Álvaro. **O que é Visão Computacional e para que serve.** São Paulo, [2022]. Disponível em: https://santodigital.com.br/o-que-e-visao-computacional-e-para-que-serve/. Acesso em: 20 set. 2022.

SAYED, Saif; TSIAKAS, Konstantinos; BELL, Morris *et al.* **Cognitive assessment in children through motion capture and computer vision: the cross-your-body task.** 2019. Disponível em: https://doi.org/10.1145/3361684.3361692. Acesso em: 31 ago. 2022.

SILVA, José Demisio Simões. **Uso de redes neurais em visão computacional e processamento de imagens.** 2004. Relatório de atividades em estágio (Nuclear Engineering Department) – Universidade de Tennessee, Knovville, USA.

VADASZ, Estevão. **Cerca de 90% dos brasileiros com autismo não recebem diagnóstico.** São Paulo, **[**2013]. Disponível em: https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2013/04/02/estima-se-que-90-dos-brasileiros-com-autismo-nao-tenham-sido-diagnosticados.htm. Acesso em: 20 set. 2022.

WEI, Pengbo; ARISTIZABAL, David; GAMMULE, Harshala *et al.* **Vision-Based Activity Recognition in Children with Autism-Related Behaviors.** 2022. Disponível em: https://arxiv.org/pdf/2208.04206.pdf. Acesso em: 1 set. 2022.

YUJIE, Xue. **Camera above the classroom.** Shenzhen, [2019]. Disponível em: https://sixthtone.medium.com/camera-above-the-classroom-532738e23d09. Acesso em: 06 out. 2022.

FORMULÁRIO DE AVALIAÇÃO BCC – PROFESSOR AVALIADOR – PRÉ-PROJETO

Avaliador(a): Aurélio Faustino Hoppe

Atenção: quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.

	ASPECTOS AVALIADOS	Atende	atende parcialmente	não atende
ASPECTOS TÉCNICOS	1. INTRODUÇÃO O tema de pesquisa está devidamente contextualizado/delimitado?			
	O problema está claramente formulado?			
	2. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado?			
	Os objetivos específicos são coerentes com o objetivo principal?			
	3. TRABALHOS CORRELATOS São apresentados trabalhos correlatos, bem como descritas as principais funcionalidades e os pontos fortes e fracos?			
	4. JUSTIFICATIVA Foi apresentado e discutido um quadro relacionando os trabalhos correlatos e suas principais funcionalidades com a proposta apresentada?			
	São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?			
	São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?			
	5. REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO Os requisitos funcionais e não funcionais foram claramente descritos?			
	6. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?			
	Os métodos, recursos e o cronograma estão devidamente apresentados e são compatíveis com a metodologia proposta?			
	7. REVISÃO BIBLIOGRÁFICA Os assuntos apresentados são suficientes e têm relação com o tema do TCC?			
	As referências contemplam adequadamente os assuntos abordados (são indicadas obras atualizadas e as mais importantes da área)?			
ASPECTOS METODOLÓGICOS	8. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?			
	A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?			