

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

Definição de estilo: Título 2;TF-TÍTULO 2

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

Hélio Potelicki

Profª. Andreza Sartori – Orientadora

Gabriel Barreto Albeton – Coorientador 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) ou Transtornos de Déficit de atenção com Hiperatividade (TDAH). 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). O TDAH causa uma série de dificuldades durante o período escolar. A criança apresenta dificuldade de atentar-se a detalhes, seguir instruções, finalizar suas atividades escolares, e envolver-se em atividades que requerem grande esforço mental (SANTOS, 2022).

Em um ambiente escolar, nos momentos em que as crianças estão agitadas, ou em momentos em que o professor necessita ausentar-se da sala de aula, apenas a observação do professor para a avaliação comportamental dos alunos pode ser prejudicada. Os olhos humanos podem, em certos momentos, falhar no quesito de detectar mudanças sutis de comportamento. De fato, estima-se que 90% dos brasileiros com autismo não tenham sido diagnosticados (VADASZ, 2013). Ao utilizar técnicas de IA juntamente ao sistema de monitoramento, os professores podem ter acesso a uma melhor análise do comportamento dos alunos, compreendendo quais são suas dificuldades (GAROFALO, 2019).

Comentado [AFH1]: Frase fora do contexto. Em um parágrafo, você fala sobre comportamento, taxa autismo e tecnicas de ia...

Trabalhos recentes como os apresentados por Kojovic *et al.* (2021) e Sayed *et al.* (2019) obtiveram bons resultados na tentativa de criar modelos preditivos para a detecção de transtornos em crianças pequenas. Ao utilizar 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, os *frameworks* desenvolvidos pelos autores, alcançaram uma acurácia geral média de 80% para a detecção e classificação de crianças sem e com algum transtorno.

No **protótipo** proposto neste trabalho, pretende-se utilizar estimativa de pose, com 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 para monitoramento 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. Com isso, esse trabalho visa auxiliar professores e responsáveis no monitoramento de crianças dentro do ambiente escolar, podendo ajudar na detecção de crianças com problemas de distanciamento social ou com algum transtorno de déficit de atenção.

Formatado: Realce

Comentado [AFH2]: Você precisa padronizar... Protótipo/ferramenta/aplicação/modelo....

Formatado: Realce

Formatado: Realce

Formatado: Realce

Comentado [AFH3]: Esta parte aqui e a primeira frase do parágrafo diz praticamente a mesma coisa

Comentado [AFH4]: Autismo não???

1.1 OBJETIVOS

Este trabalho tem como objetivo disponibilizar um protótipo para o monitoramento do comportamento infantil em ambiente escolar por meio de técnicas de Aprendizado de Máquina e Visão Computacional auxiliar no diagnóstico prematuro de possíveis Transtornos do Espectro do Autismo (TEA) ou Transtornos de Déficit de atenção com Hiperatividade (TDAH) em crianças.

Os objetivos específicos são:

- avaliar se com o uso do aprendizado de máquina é possível detectar Transtornos de Déficit de atenção com Hiperatividade (TDAH) por meio de vídeos da movimentação dos alunos;
- avaliar se é possível identificar padrões de Transtornos do Espectro do Autismo (TEA) dos alunos, por meio da estimativa de pose, durante o período de aula;
- analisar as séries históricas de interação e distanciamento dos alunos e suas implicações

Comentado [AFH5]: Que tipo de padrões a partir das poses. De interação social???

lembrando que o diagnóstico do autismo é multidisciplinar.... e cada autista é um autista diferente...

- (sociabilidade);
d) proporcionar um auxílio no monitoramento de crianças em ambiente escolar.

Comentado [AFH6]: Acho que este item é um resultado indireto do teu trabalho

2 TRABALHOS CORRELATOS

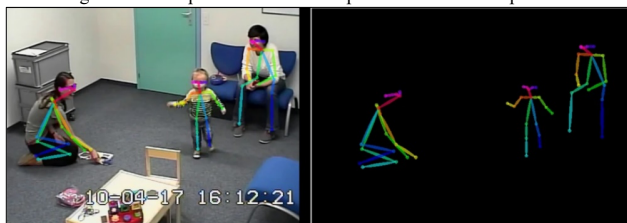
Esta seção contém trabalhos com características semelhantes aos principais objetivos do estudo proposto. Kojovic *et al.* (2021) utilizou um modelo de Aprendizado de Máquina (AM) para estimar a pose de uma pessoa e utilizar na previsão de TEA em crianças. O trabalho de Sayed *et al.* (2019) propõe o 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 Convolutiva (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 utilizou uma VGG-16 pré treinada, como *encoder* para a extração das características de cada um dos vídeos, divididos em segmentos de 5 segundos. A saída desta extração de características foi passada para o *decoder*, uma LSTM operando com 512 camadas LSTM. Em seguida estas saídas passam por 512 camadas totalmente conectadas e ativadas com a função de ativação não linear ReLu. Por fim, as saídas passam por uma camada softmax, que produz duas saídas de previsão para o segmento dado. As classificações por segmento então, são agregadas para a duração do vídeo, para obter um valor final de previsão, que varia entre 0 e 1. Se o vídeo obtiver uma média de probabilidade superior a 0,5 este será classificado como pertencente a uma criança com TEA. O modelo proposto por Kojovic *et al.* (2021), 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, pode se tornar uma realidade não muito distante no futuro (KOJOVIC *et al.*, 2021).

Comentado [AFH7]: Acho que você pode remover

A conclusão dada por Kojovic *et al.* (2021), foi que o modelo de RNA operando em uma representação postural de baixa dimensão, utilizando vídeos de interação social entre uma criança e um adulto, conseguiu distinguir se a criança possui autismo, com uma precisão de previsão do modelo de 80,9%. Os autores optaram por utilizar o modelo que opera sobre um conjunto de dados relativamente reduzido de características, visando utilizar mais características não-verbais, por serem mais relevantes em crianças muito pequenas. Kojovic *et al.* (2021) complementam informando que a precisão da classificação de cerca de 70% foi alcançada com base em

vídeos de somente 10 minutos da interação social entre a criança e o adulto, o que possibilita abrir caminhos para o desenvolvimento de ferramentas de triagem escalável usando vídeos com menor duração.

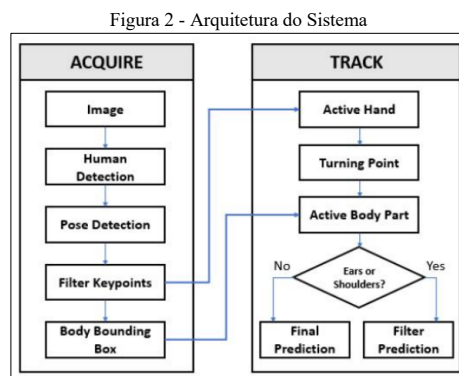
2.2 COGNITIVE ASSESSMENT IN CHILDREN THROUGH MOTION CAPTURE AND COMPUTER VISION: THE CROSS-YOUR-BODY TASK

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 (Sayed *et al.*, 2019). O conjunto de dados inclui 15 crianças realizando 8 tipos de atividades, resultando em 1900 amostras de vídeo anotadas.

O principal objetivo do sistema proposto por Sayed *et al.* (2019), é reconhecer de forma confiável o tipo de atividade realizada em um segmento de vídeo. O *framework* realiza inicialmente uma detecção de corpo, utilizando o *framework* Yolo v3, em seguida, é realizado um rastreo das mãos para reconhecer como e quando a atividade foi realizada. Como ilustrado na Figura 2, o sistema recebe os dados de movimento corporal através do Kinect, sensor de movimentos desenvolvido pela Microsoft, então é produzido o conjunto de características espaço-temporais utilizadas para prever a atividade da criança.

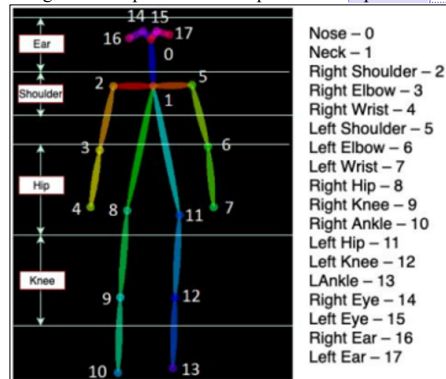
O *framework* é dividido em dois módulos, *Acquire* e *Track*. O módulo *Acquire* é encarregado por capturar e analisar o vídeo quadro a quadro para criar um vetor esqueleto para todo o vídeo, buscando pelos quadros RGB do Kinect e detectando o assunto de interesse, com o intuito de detectar e filtrar a pose 2d, e dividindo o corpo em regiões com base na sua altura. Dessa forma, é gerado um vetor de articulações de esqueleto, filtrado para o reconhecimento de atividade. Logo após o pré-processamento, o vetor de articulações é passado para o módulo *Track* para que seja feito o reconhecimento de gestos.



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

O Kinect oferece seu próprio SDK que pode ser utilizado para obter a pose do corpo em 3D, mas o problema é que o rastreamento do esqueleto pelo Kinect não funciona tão bem sob oclusões. Por este motivo Sayed *et al.* (2019), utilizam do Kinect somente para retirar a cor e canais de profundidade do ambiente. Foi utilizado para a localização e mapeamento de articulações nas imagens em RGB, o *framework* OpenPose por fornecer estimativas mais precisas da pose do corpo, onde cada articulação é representada por um vetor 2D em um espaço de coordenadas cartesianas, como demonstra a Figura 3. 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-se posições espaciais específicas da mão com uma precisão geral de 89,95%, tornando a extração de recursos mais eficiente.

Figura 3 - Mapeamento do esqueleto com OpenPose



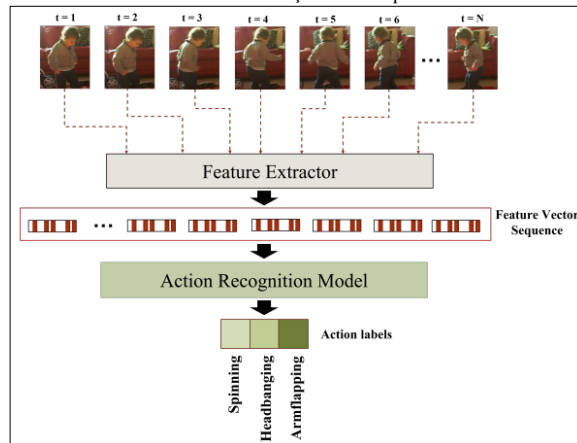
Fonte: Sayed *et al.* (2019).

Comentado [AFH8]: Acho que você poderia explicar o esta na figura..

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 4. 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 4 - Reconhecimento de atividade em crianças com comportamentos relacionados ao autismo



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

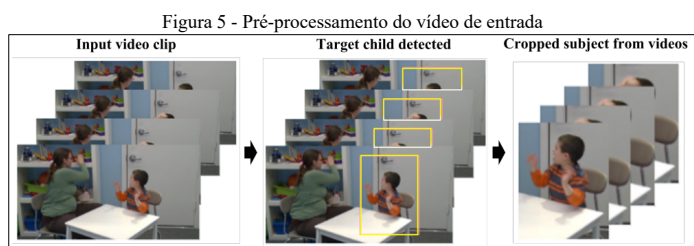
Wei *et al.* (2022) comparam empiricamente o desempenho de quatro arquiteturas de aprendizado profundo, a Long Short-Term Memory (LSTM), Temporal Convolutional Networks (TCN), Multi-Stage Temporal Convolutional Network (MS-TCN) e uma Extended Multi-Stage Temporal Convolutional Network (MS-TCN++). Estas foram utilizadas no *decoder*, termo para se referir à rede que se encarrega pela classificação das características. 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 5.



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

Os resultados experimentais mostram que a estrutura proposta pode reconhecer comportamentos relacionados ao autismo, 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. Ao utilizar o modelo MS-TCN com o *backbone Efficient Symmetric Network* (ESNet) o modelo alcançou uma acurácia geral de 81,1% de acerto, e 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 um 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, separando-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

| Trabalhos Correlatos | Kojovic <i>et al.</i> (2021) | Sayed <i>et al.</i> (2019) | Wei <i>et al.</i> (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 ou responsáveis, acompanhar os dados de movimentação e atenção das crianças de cada turma, por meio de uma interface gráfica contendo dados de movimento das crianças, auxiliando no monitoramento e no entendimento de cada turma. Como demonstra a Figura 6, a disponibilização desses dados em uma interface ainda não foi implementada pelos autores dos correlatos descritos acima. O projeto tem como público-alvo, escolas parceiras, interessadas em melhorar ou implementar um sistema de monitoramento dentro das salas de aula, e que estejam dispostas a ceder a base de dados de vídeos em sala de aula e ambiente para testes. 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 poderá contribuir para o monitoramento infantil em sala de aula, e ajudar na identificação prematura de possíveis transtornos de déficit de atenção e transtornos do espectro autista.

Comentado [AFH9]: Acho que aqui cabe descrever um pouco sobre a técnica(s) a serem utilizada(s).. Aprendizado/visão computacional

Comentado [AFH10]: A figura em si, não contribui muito...

Figura 6 - Interface gráfica com dados dos alunos



Fonte: Yujie (2019).

3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

O protótipo proposto deve:

Aplicação:

- permitir o usuário responsável verificar a série histórica de movimentação dos alunos em sala de aula (Requisito Funcional – RF);
- permitir o usuário responsável acompanhar movimentação e atenção dos alunos em sala de aula (RF);
- disponibilizar painéis visuais com informações das crianças em sala de aula para os professores e responsáveis (Requisito Não-Funcional – RNF);
- receber dados em tempo real das câmeras em sala de aula (RNF);
- ser implementado na linguagem Python (RNF);
- armazenar os dados coletados em um banco de dados (RNF).

Aprendizado de máquina:

- utilizar técnicas de aprendizado de máquina e visão computacional, para que a partir de vídeos retirados das câmeras nas salas de aula, seja possível modelar o classificador de TDAH e TEA (RF);
- ser capaz de informar se a criança possui comportamentos relacionados à transtornos de déficit de atenção através das imagens das câmeras da sala (RF);
- ser capaz de informar se a criança possui comportamentos relacionados à transtornos do espectro

Excluído: N

Comentado [AFH11]: Estes requisitos são praticamente os teus objetivos específicos..

- autista através dos dados da estimativa de pose (RF).
d) ser implementado na linguagem Python (RNF).

3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- 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), Transtornos de Déficit de Atenção com Hiperatividade (TDAH) e Redes Neurais Artificiais;
- submissão ao conselho de ética: submissão para o conselho de ética da FURB;
- entendimento do negócio: serão feitas reuniões com profissionais da área de saúde que trabalham com TDH e TEA visando entender as características dos comportamentos e da movimentação de crianças com transtornos;
- construção da bases dados: através dos dados de movimentação dos alunos coletados das imagens de câmeras dentro das salas, os dados serão anonimizados conforme estabelecido na Lei Geral de Proteção de Dados (LGPD);
- análise e refinamento dos dados: analisar e refinar as características e informações existentes nas bases de dados e validar com os profissionais da área e com os órgãos que disponibilizaram os dados;
- elicitação de requisitos: reavaliar os requisitos da seção anterior, e especificar outros, mediante a necessidade identificada na revisão bibliográfica;
- especificação: formalizar a diagramação das classes e dos casos de uso do protótipo;
- implementação: utilizar a linguagem de programação Python para o desenvolvimento do protótipo;
- 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

| etapas / quinzenas | 2023 | | | | | | | | | |
|---------------------------------|------|---|------|---|------|---|------|---|------|---|
| | ago. | | set. | | out. | | nov. | | dez. | |
| | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| levantamento bibliográfico | | | | | | | | | | |
| submissão ao conselho de ética | | | | | | | | | | |
| entendimento do negócio | | | | | | | | | | |
| construção da bases dados | | | | | | | | | | |
| análise e refinamento dos dados | | | | | | | | | | |
| elicitação de requisitos | | | | | | | | | | |
| especificação | | | | | | | | | | |
| implementação | | | | | | | | | | |
| testes | | | | | | | | | | |

Fonte: elaborado pelo autor.

4 REVISÃO BIBLIOGRÁFICA

Esta seção descreve 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 transtornos do desenvolvimento infantil. A seção 4.1 descreve sobre os conceitos de visão computacional. A seção 4.2 descreve sobre dois dos principais transtornos do desenvolvimento infantil. E por fim, a seção 4.3 descreve sobre uma tecnologia recente para rastreamento de articulações do corpo humano.

4.1 VISÃO COMPUTACIONAL

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. Estas são baseadas na arquitetura e aprendizagem do cérebro humano (MANO, 2008).

Uma RNA geralmente consiste em várias camadas de nós interconectados, chamados de neurônios. Esses neurônios são inspirados pelos neurônios do cérebro, que transmitem informações e processam sinais. Em uma

Comentado [AFH12]: Aqui falta uma etapa sobre o modelo (AM/Visão computacional)

Comentado [AFH13]: Aqui você poderia ter separado melhor aplicação / AM

Comentado [AFH14]: Esta etapa você pode realizar em 2023/1. Vai agilizar o processo

Comentado [AFH15]: Separar em atividades da aplicação / AM

Comentado [AFH16]: Seções curtas. No projeto cada assunto deveria ter no mínimo 1 página

RNA, cada neurônio recebe entrada de outros neurônios, processa essa entrada usando uma função matemática e então transmite o resultado para outros neurônios na próxima camada. No contexto da visão computacional, a entrada para uma RNA pode ser uma imagem ou uma sequência de imagens e a saída pode ser uma classificação ou previsão com base nos dados visuais.

4.2 TRANSTORNOS DO DESENVOLVIMENTO INFANTIL

Esta subseção descreve os transtornos que afetam o desenvolvimento infantil, que serão analisados e estudados para validar se será possível a sua implementação no protótipo de monitoramento. A seção 4.2.1 descreve sobre os transtornos do espectro autista (TEA). E a seção 4.2.2 descreve sobre o transtorno do déficit de atenção com hiperatividade (TDAH).

4.2.1 TRANSTORNOS DO ESPECTRO DO AUTISMO

Os 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). Apesar de ser uma condição permanente, os tratamentos para TEA minimizam as dificuldades enfrentadas pela criança portadora. Os tratamentos favorecem o desenvolvimento de habilidades e o desenvolvimento em geral, dentro das possibilidades de cada criança e a idade em que o transtorno foi diagnosticado, pois quanto mais rápido a criança for encaminhada para o tratamento, melhor será seu desenvolvimento (MAGIEREK, 2019).

Kojovic *et al.* (2021) reconhecem que, embora o processo de diagnóstico compreende a indexação de comportamentos específicos, o TEA também vem com deficiências amplas que muitas vezes são identificadas a partir de atos comportamentais únicos. Como por exemplo, os comportamentos não-verbais atípicos se manifestam por meio de padrões globais de postura e movimentação atípica.

4.2.2 TRANSTORNO DO DÉFICIT DE ATENÇÃO COM HIPERATIVIDADE

O TDAH é definido como um transtorno que se caracteriza por desatenção, impulsividade e inquietação. Indivíduos com esses problemas são desatentos com suas atividades rotineiras, tendo dificuldades em determinar prioridades para o que deve ser feito. O que acaba fazendo com que se sintam sobrecarregados (SANTOS, 2022).

O TDAH é um transtorno psiquiátrico do neurodesenvolvimento encontrado em crianças e adolescentes jovens, podendo ser detectado a partir dos 6 anos de idade (Sayed *et al.*, 2019). Deficiências cognitivas nas funções executivas podem não apenas causar mau desempenho dentro do ambiente escolar, mas também pode causar repercussões na família, no emprego e na comunidade como um todo, o que resulta em baixa autoestima e autoaceitação.

De acordo com Sayed *et al.* (2019) os recentes avanços em aprendizado profundo levaram ao uso de redes neurais convolucionais para extração de padrões de aceleração incorporados favorecendo medidas objetivas para ajudar no diagnóstico do TDAH, contudo, essas abordagens podem ser intrusivas, pois o paciente precisa utilizar de dispositivos e sensores vestíveis. Mas com configurações baseadas em câmeras podem favorecer um ambiente discreto para coleta de dados, podendo utilizar visão computacional e métodos de aprendizado profundo para realizar a extração de recursos espaço-temporais importantes para o reconhecimento de padrões de interesse.

4.3 ESTIMATIVA DE POSE HUMANA

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 objetivo da EPH é permitir que um computador entenda a pose de uma pessoa em uma imagem ou vídeo, que pode ser usado para aplicações como reconhecimento de ação, reconhecimento de gestos e análise de atividade. 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.

MAGIEREK, Valeska. **Autismo não tem cura, mas tem tratamento**. Barbacena, [2019]. Disponível em: <https://centroamadesenvolvimento.com.br/autismo-nao-tem-cura-mas-tem-tratamento/>. Acesso em: 04 dez. 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.

SANTOS, Vanessa Sardinha dos. **Transtorno de Deficit de Atenção com Hiperatividade (TDAH)**. 2022. Disponível em: <https://brasilescola.uol.com.br/psicologia/transtorno-deficit-atencao.htm>. Acesso em: 24 de nov. 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, Knoxville, 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 – 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? | X | | |
| | O problema está claramente formulado? | X | | |
| | 2. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado? | X | | |
| | Os objetivos específicos são coerentes com o objetivo principal? | X | | |
| | 3. TRABALHOS CORRELATOS São apresentados trabalhos correlatos, bem como descritas as principais funcionalidades e os pontos fortes e fracos? | X | | |
| | 4. JUSTIFICATIVA Foi apresentado e discutido um quadro relacionando os trabalhos correlatos e suas principais funcionalidades com a proposta apresentada? | X | | |
| | São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta? | | X | |
| | São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta? | | X | |
| | 5. REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO Os requisitos funcionais e não funcionais foram claramente descritos? | X | | |
| | 6. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC? | X | | |
| | Os métodos, recursos e o cronograma estão devidamente apresentados e são compatíveis com a metodologia proposta? | X | | |
| | 7. REVISÃO BIBLIOGRÁFICA (atenção para a diferença de conteúdo entre projeto e pré-projeto) Os assuntos apresentados são suficientes e têm relação com o tema do TCC? | | X | |
| ASPECTOS METODOLÓGICOS | As referências contemplam adequadamente os assuntos abordados (são indicadas obras atualizadas e as mais importantes da área)? | X | | |
| | 8. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica? | X | | |
| | A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)? | | X | |

Comentado [AFH17]: Acho que no decorrer do TCC II será ajustado. OK

Comentado [AFH18]: Rever pois está faltando etapas..

O projeto de TCC será reprovado se:

- qualquer um dos itens tiver resposta NÃO ATENDE;
- pelo menos 4 (quatro) itens dos **ASPECTOS TÉCNICOS** tiverem resposta ATENDE PARCIALMENTE; ou
- pelo menos 4 (quatro) itens dos **ASPECTOS METODOLÓGICOS** tiverem resposta ATENDE PARCIALMENTE.

PARECER: (X) APROVADO () REPROVADO