



UNIVERSIDADE FEDERAL DO TOCANTINS
CAMPUS DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO PROFISSIONAL EM
MODELAGEM COMPUTACIONAL DE SISTEMAS

Rogério Nogueira de Sousa

**COMAIS LABS: IMPLANTAÇÃO E TRANSFORMAÇÃO
DIGITAL POR MEIO DA MODELAGEM DE SOLUÇÕES
DE INTELIGÊNCIA ARTIFICIAL NO ESTADO
DO TOCANTINS**

Palmas/TO
2023

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

- N778c Nogueira de Sousa, Rogério.
COMAIS LABS: IMPLANTAÇÃO E TRANSFORMAÇÃO DIGITAL POR
MEIO DA MODELAGEM DE SOLUÇÕES DE INTELIGÊNCIA ARTIFICIAL
NO ESTADO DO TOCANTINS. / Rogério Nogueira de Sousa. – Palmas, TO,
2023.
100 f.
Tese (Doutorado) - Universidade Federal do Tocantins – Câmpus
Universitário de Palmas - Curso de Pós-Graduação (Doutorado Profissional)
em Governança e Transformação Digital - PPGTD, 2023.
Orientador: David Nadler Prata
1. Inteligência Artificial. 2. Transformação Digital. 3. Processamento de
Linguagem Natural. 4. Aprendizado de Máquina. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer
forma ou por qualquer meio deste documento é autorizado desde que citada a fonte.
A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184
do Código Penal.

**Elaborado pelo sistema de geração automática de ficha catalográfica da
UFT com os dados fornecidos pelo(a) autor(a).**

Rogério Nogueira de Sousa

**COMAIS LABS: IMPLANTAÇÃO E TRANSFORMAÇÃO
DIGITAL POR MEIO DA MODELAGEM DE SOLUÇÕES
DE INTELIGÊNCIA ARTIFICIAL NO ESTADO DO
TOCANTINS**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins (PPGMCS/UFT), como requisito necessário à obtenção de título de Doutor em Modelagem Computacional de Sistemas.

Orientador: Prof. Dr. David Nadler Prata

**PALMAS – TO
2023**

*Dedico este trabalho à Prof. Dra.
Juliana Ricarte Ferraro (in memoriam),
por me encorajar a enfrentar e apoiar
no enfrentamento deste desafio.*

*“Num mundo dominado por algoritmos,
até que ponto o livre-arbítrio será
importante para as futuras gerações?”
(Yuval Noah Harari)*

AGRADECIMENTOS

Agradeço aos meus Pais, Ismael Caetano de Sousa e Maria Lúcia Nogueira de Sousa, que nunca me faltaram, sempre me incentivando e apoiando.

À minha esposa Jacqueline Rodrigues, que pacientemente me apoiou, incondicionalmente.

À minha filha Lavínia Rodrigues Nogueira, que me inspira a ser uma pessoa melhor.

Aos professores do Programa de Pós-graduação de Modelagem Computacional de Sistema da Universidade Federal do Tocantins, principalmente ao meu orientador, o Professor Dr. David N. Prata.

Aos amigos, por torcerem por mim e por me proporcionar momentos de tranquilidade e descontração.

RESUMO

Este trabalho trata da implantação e transformação digital por meio da modelagem de soluções de Inteligência Artificial (IA) no estado do Tocantins, com enfoque na criação de um ambiente de colaboração interdisciplinar com foco em inovação, que busca por meio de tecnologias emergentes, como a IA agregar valor a soluções tecnológicas em diversas áreas. Trata-se de um laboratório de modelagem computacional de soluções de inteligência artificial na Universidade Federal do Tocantins (UFT), denominado *Computational Modeling of Artificial Intelligence Solutions* (COMAIS), cuja finalidade é contribuir na formação de agentes de transformação digital dos alunos do Programa de Pós-graduação em Modelagem Computacional de Sistemas da instituição. São apresentadas três soluções distintas, desenvolvidas pelo laboratório, sendo uma na área de Processamento de Linguagem Natural, voltada ao apoio ao processo de revisão sistemática, por meio de agrupamento de artigos científicos com base nos seus contextos. As outras duas soluções são da área de Visão Computacional, direcionadas ao setor agropecuário, nas quais, aplicando técnicas de reconhecimento de objetos, os sistemas foram capazes de detectar e georreferenciar erva daninha em áreas de cultivo e detectar bezerros em espaços pré-definidos em áreas de pastagem. A partir da seguinte questão: Se a Inteligência Artificial é reconhecida por seu poder disruptivo e potencial transformador, por que o PPGMCS, mesmo sendo um programa de excelência, ainda carece de uma estrutura específica para suporte multidisciplinar em projetos que utilizam IA? Propôs-se um objetivo geral, que consistiu em desenvolver e implementar um programa de suporte multidisciplinar voltado para projetos que empregam Inteligência Artificial, por meio da estruturação e operacionalização do laboratório de pesquisa COMAIS. Concluiu-se que o Laboratório COMAIS demonstrou excelência na integração da inteligência artificial em diversos setores, oferecendo soluções inovadoras e impactantes, por meio de alinhamento com diretrizes nacionais e compromisso com a inovação reforçam seu papel crucial no avanço da IA na UFT e no Tocantins.

Palavras-chave: Transformação Digital. Modelagem de Soluções. Inteligência Artificial. Visão Computacional. Processamento de Linguagem Natural. Aprendizado de Máquina.

ABSTRACT

This work addresses the implementation and digital transformation through the modeling of Artificial Intelligence (AI) solutions in the state of Tocantins, with a focus on creating an interdisciplinary collaboration environment with a focus on innovation. It aims to add value to technological solutions in various areas through emerging technologies like AI. The work describes a Computational Modeling of Artificial Intelligence Solutions (COMAIS) laboratory at the Federal University of Tocantins (UFT), which aims to contribute to the formation of digital transformation agents among students in the Postgraduate Program in Computational Modeling of Systems at the institution. Three distinct solutions developed by the laboratory are presented: one in Natural Language Processing, supporting the systematic review process by grouping scientific articles based on their contexts, and two in Computer Vision, targeted at the agricultural sector. These solutions apply object recognition techniques to detect and georeference weeds in cultivation areas and calves in predefined spaces in pasture areas. The research question, "If Artificial Intelligence is recognized for its disruptive power and transformative potential, why does the PPGMCS, despite being an excellent program, still lack a specific structure to support multidisciplinary projects using AI?" led to the overall objective of developing and implementing a multidisciplinary support program for projects employing AI through the structuring and operationalization of the COMAIS research laboratory. The conclusion highlights the excellence of the COMAIS Laboratory in integrating artificial intelligence into various sectors, providing innovative and impactful solutions. Its effective management, alignment with national guidelines, and commitment to innovation reinforce its crucial role in advancing AI in UFT and Tocantins.

Keywords: Digital Transformation. Solution Modeling. Artificial Intelligence. Computer Vision. Natural Language Processing. Machine Learning.

LISTA DE FIGURAS

Figura 1 - Layout do laboratório	20
Figura 2 - Antes da Reforma	20
Figura 3 - Depois da Reforma	20
Figura 4 - Logotipo COMAIS	21
Figura 5 - Portal COMAIS	22
Figura 6 - Eixos estratégicos	24
Figura 7 - <i>Viable System Model</i> (VSM)	27
Figura 8 - <i>Design VSM</i> do COMAIS	30
Figura 9 - Sequência de limpeza dos dados.....	34
Figura 10 - Entrada “ABC” e gera “WXYZ” na saída.....	36
Figura 11 - Algoritmos de tradução antes e depois dos <i>Attentions</i>	37
Figura 12 - Visualização do mecanismo de <i>attention</i>	39
Figura 13 - Relação entre sentenças	39
Figura 14 - Representação livre de contexto	40
Figura 15 - Representação profundamente contextualizada.....	40
Figura 16 - Detecção de vacas	46
Figura 17 - Segmentação por instância	47
Figura 18 - Detecção de pontos-chave nas vacas	49
Figura 19 - Arquitetura API.....	56
Figura 20 - <i>Interface Smart Review</i>	57
Figura 21 - Aplicação da Análise de Componentes Principais (PCA) para Visualização gráfica da disposição dos resumos em 3 dimensões.....	58
Figura 22 - Marcação dos polígonos com as daninhas na plantação.....	61
Figura 23 - Saída processada da rede neural, com a identificação das daninhas	62
Figura 24 - Gráfico com a distribuição das daninhas observadas no ensaio	63
Figura 25 - Fluxo de processamento das imagens pela IA	66
Figura 26 - Sensibilidade do modelo proposto.....	68
Figura 28 - Precisão média do modelo proposto (mAP)	68
Figura 29 - Ortofotos: (a) segmentada; (b) com geojson das detecções.....	70
Figura 30 - Página de exibição detalhada do voo, com visão das buvas detectadas pela IA	71

Figura 31 - Página de exibição detalhada, com visão de capim amargoso detectado pela IA	72
Figura 32 - Página de exibição detalhada, com visão de capim carrapicho detectado pela IA 72	
Figura 33 - Taxa de acerto e velocidade em FPS para diferentes <i>stacks</i>	77
Figura 34 - Curva de treinamento relacionando precisão e perda	79
Figura 35 - Detecção de monta.....	81
Figura 36 - Detecção de Vacas	86
Figura 37 - Análise de objeto de tamanho conhecido	87
Figura 38 - Detecção com falso positivo.....	88
Figura 39 - Bezerro detectado corretamente.....	90

LISTAS DE SIGLAS E ABREVIATURAS

API	<i>Application Programming Interface</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
COMAIS LAB	<i>Computational Modeling of Artificial Intelligence Solutions Lab</i>
CNNs	<i>Convolutional Neural Networks</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CuDNN	<i>CUDA Deep Neural Network</i>
DL	<i>Deep Learning</i>
EBIA	Estratégia Brasileira de Inteligência Artificial
FAPTO	Fundação de Apoio Científico e Tecnológico do Tocantins
GDAL	<i>Geospatial Data Abstraction Library</i>
SGD	Gradiente Descendente Estocástico
IA	Inteligência Artificial
IDF	<i>Inverse Document Frequency</i>
LM	<i>Language Model</i>
LSTM	<i>Long Short-Term Memory</i>
MCTIC	Ministério da Ciência, Tecnologia, Inovações e Comunicações
VSM	Modelo de Sistema Viável
NVR	<i>Network Video Recorder</i>
OpenCV	<i>Open Computer Vision</i>
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PTZ	<i>Pan-Tilt-Zoom</i>
PCA	<i>Principal Component Analysis</i>
PLN	Processamento de Linguagem Natural
PPGMCS	Programa de Pós-Graduação em Modelagem Computacional de Sistemas
Propesq	Pró-Reitoria de Pesquisa e Pós-Graduação
RANSAC	<i>RANdom SAmples Consensus</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
UFT	Universidade Federal do Tocantins
YOLO	<i>You Only Look Once</i>

SUMÁRIO

1	INTRODUÇÃO.....	11
2	COMAIS	18
2.1	Modelo de Gestão	24
2.2	Resultados e Discussão da seção	31
3	PROCESSAMENTO DE LINGUAGEM NATURAL.....	34
3.1	<i>Seq2seq</i>	36
3.2	<i>Attentions</i>	36
3.3	<i>Transformers</i>	37
3.4	Modelos de linguagem.....	38
3.5	BERT	38
3.6	<i>Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (SBERT)</i>	40
3.7	<i>K-means</i>	41
4	VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL.....	43
4.1	Técnicas e Algoritmos em Visão Computacional	44
4.2	Detecção de Objetos.....	45
4.2.1	Detecção de objetos baseada em caixas delimitadoras (bounding boxes)....	46
4.2.2	Detecção de objetos baseada em segmentação	47
4.2.3	Detecção de objetos baseada em pontos-chave (keypoints).....	48
4.3	Dataset.....	49
4.4	<i>Yolo</i>	50
4.4.1	Treinamento.....	51
5	SMARTREVIEW: UMA SOLUÇÃO DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA AGRUPAMENTO DE ARTIGOS CIENTÍFICOS NA REVISÃO SISTEMÁTICA	54
5.1	Introdução	54
5.2	Metodologia.....	55
5.3	Resultados e Discussão	58
5.4	Conclusão.....	59
6	DETECÇÃO E LOCALIZAÇÃO DE ERVAS DANINHAS PARA O USO INTELIGENTE DE DEFENSIVOS NA AGRICULTURA	60
6.1	Introdução	60

6.2	Métodos	60
6.2.1	Treinamento do modelo inicial.....	64
6.2.2	Treinamento do modelo final.....	65
6.3	Resultados	71
7	DETECÇÃO DE COMPORTAMENTO ANIMAL EM VÍDEOS UTILIZANDO VISÃO COMPUTACIONAL E APRENDIZADO SUPERVISIONADO	75
7.1	Introdução	75
7.2	Método	75
7.3	Resultados e Discussão	80
7.4	Conclusão	81
8	DETECÇÃO DE BEZERROS EM VÍDEOS UTILIZANDO MEDIÇÃO DE TAMANHO E DETECÇÃO DE POSE	84
8.1	Introdução	84
8.2	Métodos	85
8.3	Resultados	89
8.4	Conclusão	90
	CONCLUSÃO GERAL	93
	REFERÊNCIAS	95

1 INTRODUÇÃO

Em um mundo que se transforma cada vez mais em digital, onde as tecnologias digitais estão progressivamente mais entranhadas no dia a dia dos indivíduos e das empresas, as organizações e as pessoas são impulsionadas pela ampliação da capacidade de transmissão de dados e difusão da computação móvel, por meio de smartphones, tablet e notebooks, gerando grandes volumes de dados de todos os tipos. O digital passa a ser parte integrante no modo de vida das pessoas, alterando barreiras de espaço e tempo, a forma de comunicação das pessoas, suas relações entre si e com as organizações.

Nesse mundo digital, a relação entre empresas e clientes vem sofrendo mudanças constantes. Agora, os clientes formam redes nas quais as tecnologias facilitam a interação, transformando a forma como são influenciados a comprar um determinado produto e como dialogam com as empresas (Rogers, 2017). Esse é apenas um exemplo das incontáveis transformações profundas que estamos vivendo, levando as organizações a repensarem suas estratégias e se adaptarem a esse novo paradigma.

Nesse sentido, o processo que usa das tecnologias e dados digitais para transformar o negócio em digital, por meio da criação de novas atividades ou alteração das já existentes, é conhecido como digitização¹, que diferentemente do termo digitalização², que é o ato de converter dados e processos físicos em formato digital. Quando a digitização e a digitalização promovem alterações sociais e econômicas, ocorre a transformação digital (OECD, 2019).

Embora imprescindíveis, a digitização e a digitalização, por si só, não são transformacionais. A transformação digital tem um espectro mais amplo, provoca uma mudança cultural nas organizações. É disruptiva para diversos setores que observamos, trazendo mudanças imprevisíveis e emergentes.

¹ A digitização é um fenômeno tecnológico contemporâneo. Diz respeito à transformação do negócio em um modelo digital, pautado pelas novas demandas e comportamentos dos consumidores (Sydle, 2022). Disponível em: <https://www.sydle.com/br/blog/digitizacao-digitalizacao-627d6d7b73e6bf3e54b069da#:~:text=A%20digitiza%C3%A7%C3%A3o%20%C3%A9%20um%20fen%C3%B4meno,um%20objeto%20f%C3%ADsico%20para%20digital>. Acesso em: 5 ago.2023.

² A digitalização é um termo parecido, mas, na verdade, é a conversão de um objeto físico para digital (Sydle, 2022). Disponível em: <https://www.sydle.com/br/blog/digitizacao-digitalizacao-627d6d7b73e6bf3e54b069da#:~:text=A%20digitiza%C3%A7%C3%A3o%20%C3%A9%20um%20fen%C3%B4meno,um%20objeto%20f%C3%ADsico%20para%20digital>. Acesso em: 5 ago.2023.

A transformação digital representa uma revolução multifacetada que engloba mudanças tecnológicas, empresariais e sociais. Essa revolução tem o potencial de reformular os mercados existentes ou até mesmo gerar novos, fortalecendo eficiências nos processos de produção. Este fenômeno é possibilitado pela análise inovadora de dados, um processo que tem a capacidade única de gerar valor a partir do aparentemente ordinário ou até mesmo do nada. Além disso, a transformação digital pode transfigurar esses dados em serviços inéditos ou até mesmo estabelecer mercados completamente novos, refletindo seu impacto disruptivo e abrangente na economia global.

Trata-se de uma transformação em que as mudanças não podem ser previstas com antecedência por causa de sua característica dinâmica evolutiva. Logo, variações recebem um novo enfoque, passando a ser um ativo em si ao receber destaque onde os agentes de transformação precisam de uma visão sistêmica da estratégia do negócio para criar táticas flexíveis o suficiente para absorver mudanças adequadamente.

Sobre o exposto, o Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) dita que o ritmo global da Transformação Digital tende a se acelerar e nesse contexto, o Brasil deve encarar essas evoluções como uma oportunidade, pois as tecnologias podem reformular ações do governo e incrementar a produtividade e competitividade das empresas, bem como a capacitação e inclusão para que todos possam se desenvolver de forma a promover a inovação com vistas ao bem social (MCTI, 2018).

Dessa forma, entende-se que a única fonte de vantagem competitiva enraizada no mercado continua sendo a Inovação, isso porque a velocidade e complexidade estão em franco crescimento, exigindo novas estratégias para atingir o consumidor de forma eficaz. Assim, tanto o setor público quanto o privado necessitam ampliar os esforços de inovação e realinhar a forma para alcançar seus objetivos (Muro *et al.*, 2015).

De forma explicativa, inovação é um novo produto ou processo novo que foi aprimorado e que foi disponibilizado para usuários em potencial (produto) ou colocado em uso pela unidade (processo). Nesse sentido, pode-se dizer que o conhecimento é a base para inovação, criação e novidade, enquanto a inovação deve ser implementada, ou seja, usada por outros, sendo fundamental para melhoria nos padrões de vida ao ter o poder de afetar indivíduos, instituições, setores econômicos inteiros e países de várias maneiras. Assim, essa política pode contribuir direta e indiretamente para definir a direção da inovação e moldar como seus efeitos são distribuídos (OECD; EUROSTAT, 2018).

De forma direta, então, a Transformação Digital vem alterando o processo de inovação nas organizações que tinham foco no produto final, baseado em pesquisas complexas, caras e lentas ou na intuição dos gestores, mitigando experimentações, pois o custo desse processo era elevado. Assim, reduzindo a velocidade e a capacidade de inovação das organizações, colocando-as em risco, pois com a aceleração das mudanças, a duração das janelas de oportunidade vem diminuindo, tornando a competição mais agressiva.

Os agentes de transformação procuram promover uma disrupção no modelo de inovação por meio do uso de todas as tecnologias digitais disponíveis nos diversos campos possíveis (Martins; Moser, 2021), propiciando um aprendizado contínuo e agilidade na experimentação, com retornos imediatos do mercado durante todo o processo de inovação, maximizando o aprendizado organizacional, mitigando o impacto das falhas e, concomitantemente, reduzindo custos.

Nesse norte, e visando analisar o cenário brasileiro diante da Transformação Digital, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) realizou uma avaliação, publicada em 2020, denominada Caminho da Era Digital no Brasil, que apontou consideráveis avanços em diversas áreas, mas ponderou quanto a possíveis atrasos.

Entre os desafios apontados pela OCDE, citamos a priorização do financiamento público de Pesquisa e Desenvolvimento, que vem decrescendo nos últimos anos, e a necessidade de ampliar o capital humano por ser o principal ativo para a fundação de uma área de tecnologia da informação e comunicação pujante (OECD, 2020).

Dessarte, diante da escassez de recursos humanos nas áreas relacionadas às tecnologias digitais, alguns países ampliaram investimentos na educação superior voltadas a tais áreas. Ainda conforme a Organização para a Cooperação e Desenvolvimento Econômico (OECD, 2020), outra ação que pode ser adotada é a implantação de programas interdisciplinares de dupla aprendizagem.

Programas interdisciplinares de dupla aprendizagem são iniciativas experimentais que combinam aprendizado teórico em sala de aula com experiência prática de trabalho, abrangendo duas disciplinas ou áreas de estudo distintas. Esses programas são projetados para fornecer aos alunos uma educação abrangente, enfatizando a aplicação prática de conhecimentos teóricos. No contexto de escassez de recursos humanos em relação às tecnologias digitais, esses programas podem combinar, por exemplo, o estudo da computação com outra disciplina, como negócios ou design. Isso permite que os alunos

desenvolvam habilidades técnicas valiosas, enquanto também aprendem a praticar essas habilidades em contextos interdisciplinares (European Commission, 2018)

Conforme o exposto, pode-se afirmar que as universidades são de fundamental importância nesse processo formativo desses discentes, pois tem o papel de formadora de cidadãos, promovendo conhecimentos técnicos, científicos e sociais, orientados para atender um mundo em constante transformação. As instituições de ensino superior impactam diretamente no desenvolvimento tecnológico, produzindo e disseminando conhecimento e novas tecnologias, sendo um instrumento de transformação e um catalisador para formação de entes inovadores.

Contudo, não podemos limitar o papel da Universidade a simples geração de mão de obra especializada, logicamente essa instituição tem que se orientar e focar em seu tripé orientador, que é pesquisa, ensino e extensão, especialmente na esfera pública, que se pauta também nesse tripé, só que pode e deve democratizar o saber, podendo assim levar conhecimento tecnológico avançado que agregue valor para a sociedade como um todo. Ao se colocar nesse papel, em prol da sociedade, temos um reforço descomunal que reforça a cidadania.

Voltando ao escopo das empresas competitivas no mundo digital, estas estão em busca de recursos humanos capacitados, escassos, como mencionado, e a universidade tem o papel de formar cidadãos capazes de atuar nessa nova economia. No entanto, a Transformação exige mais, não apenas das empresas, mas também das instituições de ensino que devem fomentar, além das tradicionais habilidades, novas aptidões como a análise de dados, gerenciamento de mudanças, resiliência e agilidade para se moldar às constantes mudanças, facilidade em operar ferramentas tecnológicas e atualização contínua de seu conhecimento (Martins; Moser, 2021).

Diante dessa demanda, o Ministério da Educação (MEC) instituiu a portaria de número 7, de junho de 2009, que buscando reduzir a distância e fortalecer os elos entre as universidades e setor produtivo, concebeu a possibilidade de criação, por parte das instituições de ensino superior, de programas de pós-graduação *stricto sensu*, na modalidade mestrado e doutorado profissional (Brasil, 2009).

Os programas de pós-graduação profissional surgem como uma resposta aos anseios de profissionais e entidades, que buscam uma formação equilibrada entre o científico e o tecnológico, entre o teórico e o prático, voltada a soluções de problemas e entrega de produtos aderentes, que atendam demandas sociais, organizacionais ou profissionais.

Tais programas devem ter um olhar especial ao desenvolvimento da competitividade das organizações ou arranjos produtivos, nacionais, regionais e locais, sendo elas públicas ou privadas. Democratizar o conhecimento e a sensibilidade social estão no cerne da sua natureza, tendo a inovação como mola propulsora para melhorar a eficácia e a eficiência das organizações e, assim, garantir um vínculo de confiança, dinâmico, integrado e duradouro entre a universidade e o setor produtivo.

Na esteira dessa oportunidade, em 2012 a Universidade Federal do Tocantins (UFT), por meio da sua Pró-Reitoria de Pesquisa e Pós-Graduação (Propesq), lançou o primeiro edital para seleção de alunos para o Programa de Pós-Graduação em Modelagem Computacional de Sistemas (PPGMCS).

O PPGMCS apresenta uma abordagem interdisciplinar contendo duas linhas de pesquisa: Modelagem e Otimização de Sistemas e Governança Digital, ambas alicerçadas e alinhadas aos princípios que regem o processo de Transformação Digital. Estabelecendo uma ótica interdisciplinar, o programa joga luz à importância estratégica dos dados para o processo de inovação ao fazer uso de um modelo de aprendizagem com uma abordagem prática e flexível, com disciplinas balizadas no contexto atual, operando na fronteira do conhecimento e apoiadas em tecnologias emergentes.

Nesse sentido, o PPGMCS contribui diretamente para formação de agentes de transformação, capazes de impactar profundamente todos os aspectos das organizações em que estão inseridos (UFT, 2019).

Em 2022, uma década após sua fundação, o PPGMCS da UFT alcançou uma marca distintiva ao receber o conceito máximo para um programa de pós-graduação *stricto sensu* profissional: nota 5 (cinco) na avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Esse reconhecimento consolida o PPGMCS como um programa de excelência no cenário educacional brasileiro.

Entretanto, mesmo com este reconhecimento e a trajetória ascendente da UFT, persistiram lacunas tecnológicas notáveis. Uma delas, em evidência, foi a ausência de um ambiente adequado dentro do PPGMCS que pudesse oferecer suporte multidisciplinar aos projetos que empregavam Inteligência Artificial (IA). Este cenário suscita uma indagação crucial: Se a Inteligência Artificial é reconhecida por seu poder disruptivo e potencial transformador, por que o PPGMCS, mesmo sendo um programa de excelência, ainda carece de uma estrutura específica para suporte multidisciplinar em projetos que utilizam IA?

Esta carência não é apenas uma questão teórica, mas possui implicações práticas profundas. A falta de um ambiente otimizado para a IA pode restringir significativamente a capacidade do programa de conduzir pesquisas de vanguarda e, conseqüentemente, formar profissionais aptos a responderem aos desafios e demandas contemporâneas do mercado, cada vez mais imerso na era digital.

O Tocantins, enquanto mais novo estado brasileiro, representa um território fértil para inovações e para a introdução de tecnologias disruptivas que podem acelerar seu desenvolvimento e afirmar sua posição no cenário nacional. A implantação do laboratório de Inteligência Artificial (IA) no Programa de Pós-Graduação em Modelagem Computacional de Sistemas (PPGMCS) da UFT nasceu de múltiplas motivações.

Primeiramente, vislumbrou-se a oportunidade de promover o uso de tais tecnologias voltadas especificamente para os desafios e oportunidades singulares de Tocantins. A ideia de que soluções concebidas dentro do próprio estado possam servir de exemplo para o restante do país é poderosa, demonstrando que é plenamente viável desenvolver soluções avançadas em território tocantinense.

Diante dessa contextualização, balizada na problemática que se expôs, propôs-se um objetivo geral, que consistiu em desenvolver e implementar um programa de suporte multidisciplinar voltado para projetos que empregam Inteligência Artificial, por meio da estruturação e operacionalização do laboratório de pesquisa – CONAIS.

De forma específica, os objetivos estes seguiram as seguintes premissas: i) Definir diretrizes claras para impulsionar a inovação e colaboração multidisciplinar no Laboratório COMAIS; ii) Incorporar o Modelo de Sistemas Viáveis (VSM) para otimizar a gestão organizacional do laboratório; iii) Destacar quatro projetos emblemáticos desenvolvidos pelo COMAIS que demonstram o impacto da Inteligência Artificial em diversos setores; iv) Desenvolver e aplicar o *SmartReview*, uma ferramenta de Processamento de Linguagem Natural para agrupamento eficaz de artigos científicos; v) Implementar técnicas de visão computacional para auxiliar agrônomos na determinação de áreas de pulverização e reconhecimento de ervas daninhas; vi) Usar técnicas de IA para monitoramento de gado, incluindo detecção, segmentação e aplicativos móveis para acompanhamento.

Como desdobramento desses objetivos, a hipótese que circundou a temática da pesquisa previu que com a criação do CONAIS e a aplicação do Modelo de Sistemas Viáveis (VSM) espera-se um impacto significativo nas pesquisas relacionadas à Inteligência Artificial no Tocantins. Essa suposição guiou a investigação, buscando

evidenciar a efetividade da iniciativa em ampliar e qualificar a produção científica e tecnológica em IA na região.

Assim, a Inteligência Artificial, em particular, é vista como um divisor de águas na era moderna. No entanto, muitos ainda veem essa tecnologia com mistério ou até mesmo receio. Assim, torna-se essencial desmistificar a IA e torná-la mais acessível, impulsionando a transformação digital no estado do Tocantins. Este laboratório, portanto, visa também educar, inspirar e estimular a adoção dessas tecnologias, enfrentando desafios e aproveitando oportunidades emergentes no contexto tocantinense.

Para este pesquisador, enquanto cidadão, percebe-se a importância deste estudo porque o mesmo transcende a mera curiosidade acadêmica, justamente por entender que a IA uma ferramenta poderosa para moldar um futuro inovador para o Tocantins, reconhecendo que a universidade, como principal instituição de ensino e pesquisa do estado, tem um papel imperativo nesse processo de transformação digital, atuando como um farol que ilumina os caminhos do progresso tecnológico e do desenvolvimento sustentável.

Dessarte, o contexto que deu origem a esta pesquisa foi a percepção da lacuna entre o potencial de crescimento e inovação na UFT e a infraestrutura tecnológica e de pesquisa atualmente disponível. E nessa perspectiva, esta pesquisa apresenta importante relevância nas esferas Científica e Social.

No âmbito Científico, a Inteligência Artificial (IA) é, indubitavelmente, uma das tecnologias mais transformadoras de nossa era, tendo o potencial de reformular setores inteiros da ciência e tecnologia. Entende-se, por isso, que a criação de um laboratório focado na IA dentro do PPGMCS servirá como um polo de conhecimento avançado, contribuindo para a literatura científica, propiciando pesquisas de ponta e consolidando o Tocantins como um centro referencial em estudos relacionados à Inteligência Artificial no Brasil.

No que se refere à relevância Social, a transformação digital, potencializada pela IA, não é apenas uma evolução tecnológica; é, sobretudo, uma revolução social. A implementação efetiva dessa tecnologia no Tocantins não apenas potencializará o desenvolvimento econômico, mas também democratizará o acesso à informação e capacitará os cidadãos para o mundo digital. O laboratório se configura, assim, como um espaço de educação e sensibilização, um ambiente onde a IA deixa de ser uma entidade abstrata e se torna uma ferramenta tangível e acessível para todos. Dessa forma, esse

espaço se propõe a ser uma alavanca social, preparando a população tocantinense para os desafios e oportunidades da era digital.

Ademais, a relevância deste estudo é profundamente pessoal e engajada. O compromisso ultrapassa as fronteiras da academia, pois se alicerça na convicção de que a IA é um instrumento determinante na construção de um Tocantins mais inovador e próspero. Entende-se que a universidade, enquanto principal baluarte do conhecimento no estado, é imprescindível nesse processo de evolução. A UFT deve, portanto, se posicionar na vanguarda dessa transformação, guiando e inspirando a sociedade tocantinense rumo a um futuro mais digital e sustentável.

Tal compromisso é fruto de uma observação atenta: a disparidade entre o vasto potencial do Tocantins e a infraestrutura tecnológica e de pesquisa presente. Esta pesquisa, assim, nasce da necessidade de colmar essa lacuna, vislumbrando um estado que efetivamente harmonize seu potencial natural e cultural com os avanços da era digital.

Diante da aspiração da Universidade Federal do Tocantins (UFT) de consolidar a inovação, o desenvolvimento tecnológico e promover a transferência de tecnologias para a sociedade através do ensino, pesquisa e extensão, conforme delineado no Plano de Desenvolvimento Institucional (UFT, 2021), surgiu a necessidade de estabelecer mecanismos que efetivamente materializassem esses objetivos.

2 COMAIS

O *Computational Modeling of Artificial Intelligence Solutions* – COMAIS ou Laboratório de Modelagem Computacional de Soluções de Inteligência Artificial da UFT, encontra-se sob a coordenação do PPGMCS, liderado pelo Professor Dr. David Nadler Prata, designou ao doutorando Rogério Nogueira de Sousa a missão de criar um laboratório de pesquisa, voltado para a modelagem de soluções tecnológicas por meio da utilização de técnicas de inteligência artificial.

Nesse contexto, é importante notar que a falta de um ambiente otimizado nos laboratórios para a IA pode restringir significativamente a capacidade do programa de conduzir pesquisas de vanguarda. A infraestrutura e o ambiente de pesquisa, portanto, desempenham um papel crítico no desenvolvimento e avanço da inteligência artificial.

Em pesquisas de Inteligência Artificial (IA), a importância de laboratórios bem estruturados e otimizados não pode ser subestimada, por isso, Goodfellow *et al.* (2016) ressaltam que, especialmente quando se aborda o aprendizado profundo e se trabalha com

grandes conjuntos de dados, os laboratórios precisam contar com hardware especializado, como GPUs e TPUs. A falta desses recursos pode cercear a capacidade de desenvolver modelos de IA de alta complexidade e também diminuir a velocidade do treinamento dos modelos.

Nesse sentido, Russell e Norvig (2010) destacam que os laboratórios devem estar equipados com ambientes de *software* adequadamente otimizados. A disponibilidade e a configuração correta de bibliotecas e *frameworks*, como *TensorFlow*, *PyTorch* e *Keras*, são essenciais para evitar incompatibilidades e restrições ao aplicar algoritmos avançados em tais laboratórios.

Para além do *hardware* e *software*, a atmosfera colaborativa em um laboratório é de suma importância. McKinney (2017) aponta o papel crucial de plataformas como GitHub e Jupyter para promover a colaboração e garantir a reprodutibilidade das pesquisas em IA. Chollet (2017), na mesma direção, aborda a essência do *back-end* e seu significado para os laboratórios de IA, visto que uma infraestrutura de *back-end* otimizada é fundamental para garantir eficiência em processamento, acesso rápido a dados e capacidade de implementação de algoritmos de forma ágil.

Em geral, inadequação da otimização pode ter implicações diretas na competitividade e relevância de um laboratório de IA e nessa lógica, Sutton e Barto (2018) alertam sobre os riscos associados a laboratórios com infraestruturas inadequadas, que podem acabar ficando atrás no cenário global de inovações. Ademais, pode haver dificuldades em atrair talentos de ponta e em impulsionar a inovação.

No entanto, LeCun *et al.* (2015) sugerem que é possível encontrar soluções, enfatizando a importância de estabelecer parcerias com a indústria e investir em treinamento. Em resumo, a otimização de laboratórios é fundamental não apenas como uma “comodidade”, mas como uma exigência intrínseca para a condução de pesquisas de vanguarda em IA.

Desta feita, foram estabelecidas diretrizes para o laboratório COMAIS, incluindo a promoção de inovação, a criação de um ambiente que fomente a colaboração multidisciplinar na formação de equipes, o fornecimento de suporte tecnológico ao programa e a promoção do uso de tecnologias inteligentes para a solução de problemas concretos, além da tarefa de desmistificar tais tecnologias.

Posto o desafio, o discente estabeleceu uma parceria entre a Fundação de Apoio Científico e Tecnológico do Tocantins (FAPTO), UFT e coordenadores de projetos pesquisa, utilizando fonte própria de recursos para financiar a reforma de um espaço físico

dentro da universidade para construir o laboratório. Por meio de uma equipe de arquitetura especializada, a prefeitura da UFT criou um *layout*, apresentado na Figura 1.

Figura 1 - *Layout* do laboratório



Fonte: Acervo do autor.

A diretoria da FAPTO localizou fontes de recursos financeiros e intermediou parcerias com projetos de pesquisa para viabilizar a aquisição de equipamentos e custear a reforma planejada pela UFT, com aval dos professores do PPGMCS. A maior parte dos recursos aportados na fase inicial da implantação foram advindos dos projetos capitaneados pelo professor Doutor Humberto Xavier de Araújo em parceria com a empresa Huawei.

Figura 2 - Antes da reforma



Figura 3 - Depois da reforma

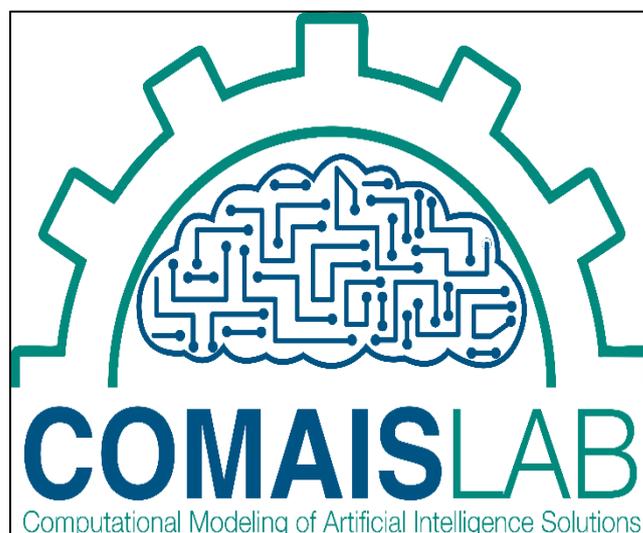


Fonte: Acervo do autor.

Na Figura 2 pode-se observar a sala que foi utilizada para implantação do laboratório antes da reforma, já a Figura 3 apresenta o laboratório depois dos investimentos em manutenção.

Concomitantemente, foram definidos um nome e uma identidade visual para o laboratório que refletissem as premissas colocadas pelo professor, ficando o mesmo denominado como *Computational Modeling of Artificial Intelligence Solutions Lab*³ (COMAIS LAB). A equipe de *designers* gráfico da UFT enviou várias sugestões, optando-se, ao final, pela imagem que consta na Figura 4, a qual remete à ideia de uma ferramenta inteligente.

Figura 4 - Logotipo COMAIS



Fonte: Equipe de Comunicação da UFT (2020).

A imagem apresentada na Figura 4 foi selecionada por encapsular de maneira clara a ideia de uma ferramenta não apenas funcional, mas também dotada de características inteligentes.

A missão do laboratório COMAIS é catalisar a colaboração entre pesquisadores, empresas e governo para formulação de soluções tecnológicas que possuem seus valores agregados e potencializados pelo uso de inteligência artificial.

Dentre outros aspectos, o COMAIS visa:

- Ser um ambiente que estimule a integração;
- Trabalhar multidisciplinaridade de forma colaborativa e criativa;

³ Laboratório de Modelagem Computacional de Soluções de Inteligência Artificial.

- Propiciar a geração de projetos inovadores.

Durante seu processo de criação foram desenvolvidos diversos projetos que usam técnicas de Inteligência Artificial que passaram a integrar seu portfólio. A concepção e modelagem desses projetos requereram uma equipe multidisciplinar, uma vez que são desenvolvidos para atuarem em diversas áreas de conhecimento, tais como: judicial, segurança pública, ambiental e social.

Concomitante, e como escopo de contribuir com a comunidade científica, a intenção é validar o aprendizado dos acadêmicos envolvidos, agregando de forma qualitativa e quantitativa em sua produção científica.

Para dar maior visibilidade às ações realizadas pelo COMAIS, foi criado um portal web⁴, apresentado na Figura 5. O site traz informações como notícias relacionadas ao laboratório, publicações de autoria dos seus colaboradores, dados para contatos com os coparticipantes, informações sobre projetos e iniciativas apoiadas pelo COMAIS, também acesso a serviços providos, formas de contato e localização física. Além das informações públicas, o portal conta com uma área restrita aos colaboradores, os quais podem utilizar ferramentas de apoio a gerenciamento de projetos.

Figura 5 - Portal COMAIS



Fonte: O autor.

⁴ Link de acesso ao portal pelo endereço <http://comais.uft.edu.br>.

Entre os serviços providos pelo COMAIS, destaca-se o *RedCap*, que é uma plataforma *Web* utilizada por pesquisadores para construção e gerenciamento de bancos de dados de pesquisas⁵.

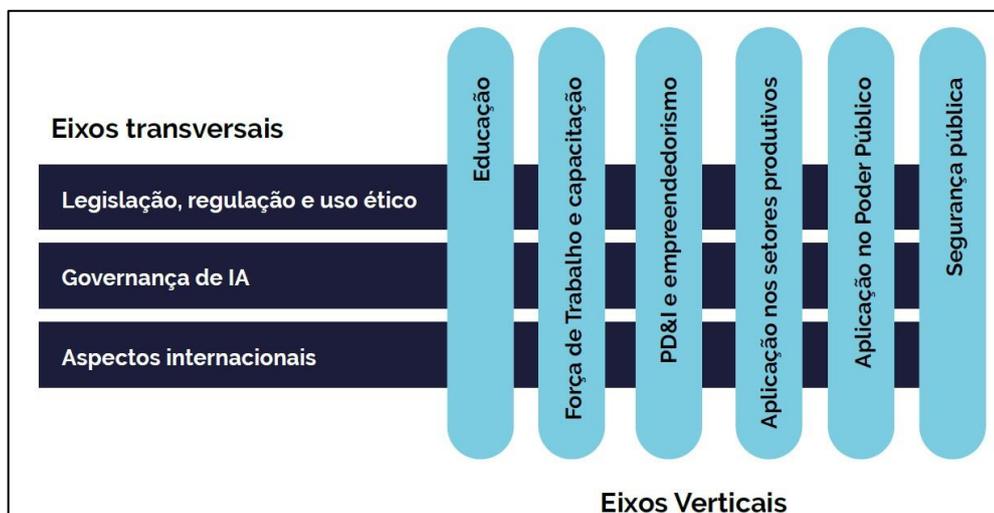
A ferramenta traz uma gama de benefícios como a substituição do uso de formulários de papel, melhoria na segurança das informações, armazenagem e compartilhamento de dados, exportação e importação de dados, entre outras soluções que poderão ser utilizadas por projetos das mais variadas áreas e disciplinas. Mas, atualmente, esse serviço é ofertado apenas aos alunos do programa, com possibilidades futuras de ser expandido para a comunidade em geral.

Por assim entender, o incremento vertiginoso do volume de dados digitais e com o avanço da capacidade computacional, a extração de informações confiáveis, inteligentes e relevantes, de forma ágil, é imprescindível para um mundo cada vez mais competitivo. Diante dessa necessidade, o MCTI, por meio da portaria nº 1.120, de 19 de março de 2020, inseriu o setor de Inteligência Artificial (IA) em seus objetivos como forma de contribuir para a base de inovação em produtos intensivos de conhecimento científico e tecnológico (MCTI, 2020).

De acordo com a OCDE (2020), vários países-membros adotaram estratégias para fomentar o uso da IA e desenvolver talentos nessa área. O Brasil lançou a Estratégia Brasileira de Inteligência Artificial (EBIA), em 2021, como pode ser observado na Figura 6, e que se organiza em 9 (nove) eixos temáticos, com objetivo de impulsionar o desenvolvimento e o uso desta tecnologia disruptiva (Brasil, 2021).

⁵ O *RedCap* pode ser acessada no endereço <https://redcap.comais.uft.edu.br/>.

Figura 6 - Eixos estratégicos



Fonte: Brasil (2021)

Assim, a inteligência artificial (IA) transcende a função de simples ferramenta de processamento de informações, tornando-se também uma estratégia eficaz para a geração de novos dados. No momento atual, os progressos teóricos na IA resultam em aplicações que estão transformando nossa existência.

Essas aplicações processam dados que geramos constantemente por meio de sites, aplicativos e transações comerciais, modificando nossos métodos de trabalho e até eliminando certos cargos profissionais. Este uso abrangente da IA carrega consigo tanto potenciais riscos quanto oportunidades promissoras. Portanto, é imperativo que estejamos preparados para ambos os cenários (Lee, 2019).

2.1 Modelo de Gestão

A inteligência artificial (IA) tem se firmado como uma área tecnológica de crescente importância no cenário contemporâneo, com uma vasta gama de aplicações em variados setores, como saúde, finanças e transportes, entre outros. Nesse contexto, laboratórios especializados em IA assumem um papel fundamental para o avanço e a aplicação desta tecnologia inovadora.

Um desses laboratórios é o COMAIS, que atua com projetos em múltiplas áreas de conhecimento, abrangendo agricultura, saúde, justiça, entre outras. Dada essa amplitude, é indispensável a identificação de modelos e ferramentas eficazes que possam

orientar eficientemente o trabalho de equipes multidisciplinares em sua busca por soluções inovadoras para problemas concretos.

Esses modelos precisam ser capazes de acomodar uma série de complexidades advindas da diversidade de colaboradores e projetos, garantindo uma gestão eficiente em um ambiente caracterizado por sua dinamicidade e complexidade, como o Modelo de Sistema Viável (VSM).

O VSM (do inglês *Viable Systems Model*) é um exemplo de gestão teórica baseado na teoria da complexidade e no conceito de sistemas viáveis, desenvolvido pelo cientista da computação e teórico da administração Stafford Beer, em 1979 (Beer, 1979). Tal modelo permite identificar e compreender os sistemas presentes em uma organização e estabelecer processos e mecanismos de comunicação e coordenação entre os diferentes sistemas (Schwaninger; Scheef, 2016). Além disso, permite visualizar a interação entre as diferentes partes do sistema e identificar pontos críticos que precisam ser melhorados.

Na gestão de departamentos ou de órgãos públicos o VSM pode ser uma ferramenta valiosa para a melhoria da eficiência, eficácia e das operações governamentais, sendo que o modelo pode ajudar a identificar oportunidades de melhoria e a otimizar o uso de recursos. Assim, o Modelo de Sistema Viável pode ajudar gestores de organizações públicas ou privadas a desenvolverem conceitos de cibernética para melhorar a comunicação e estratégias no desenvolvimento de suas atividades finalísticas (Alves *et al.*, 2018).

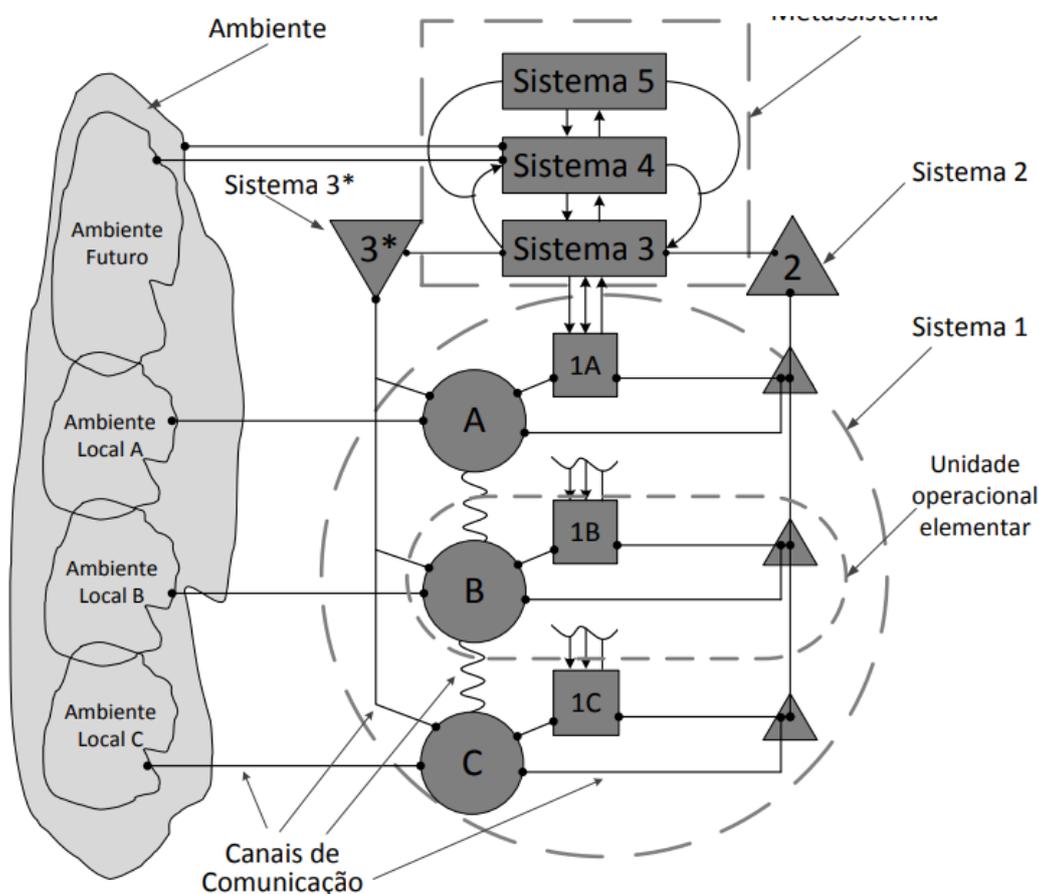
Além disso, a aplicação do VSM pode ajudar a melhorar a transparência e a responsabilidade dos órgãos governamentais, uma vez que permite que os colaboradores compreendam melhor como as decisões são tomadas e quais são as consequências de suas ações. O modelo apresenta evidências de que se trata de uma ferramenta confiável para o diagnóstico e *design* de organizações (Schwaninger; Scheef, 2016) .

O Modelo de Sistema Viável pode fornecer uma estrutura para a tomada de decisões e a coordenação entre os diferentes sistemas, garantindo que os projetos sejam desenvolvidos de maneira alinhada às estratégias. Lançar mão do VSM para atingir os objetivos do laboratório é um dos componentes transformadores deste trabalho.

Contudo, propor o *design* de um modelo de sistema viável, que tem como foco os processos de governança e gestão integrada de projetos em um Laboratório de Inteligência Artificial, implica uma compreensão dos desafios específicos enfrentados nesse tipo de ambiente, assim como uma adaptação da abordagem proposta às suas particularidades.

Em face da complexidade inerente à gestão de um laboratório multidisciplinar, cujo foco recai sobre uma tecnologia emergente e disruptiva como a IA, este estudo adotou o Modelo de Sistemas Viáveis (VSM) como base para propor um design organizacional para a gestão do laboratório, além de avaliar a capacidade do mesmo de se estabelecer como uma unidade viável.

O VSM delinea os elementos necessários para garantir a viabilidade de uma organização, estabelecendo-se sobre cinco sistemas que colaboram mutuamente para atingir os objetivos organizacionais: os Sistemas 1, 2, 3, 4 e 5. Cada sistema desempenha uma função específica e trabalha de forma integrada com os demais para assegurar a integridade e a continuidade da organização (Beer, 1984). A Figura 7 ilustra a disposição desses sistemas.

Figura 7 - *Viable System Model (VSM)*

Fonte: Rizzoli *et al.* (2014).

O Sistema 1, apresentado na figura 7, consiste na dimensão que trata dos recursos humanos. Ele se refere às operações diárias ou ao nível operacional de uma unidade administrativa, ou organização. Além disso, o Sistema 1 é composto pelas unidades primárias que realizam as atividades básicas da organização, como produção, vendas, atendimento ao cliente e desenvolvimento de soluções de tecnologia como no caso em questão.

No laboratório, as operações diárias são conduzidas pelos projetos em andamento e em fase de prospecção, como o laboratório tem como função primeira o desenvolvimento de soluções tecnológicas por meio dos projetos, o sistema 1 é a “espinha dorsal” do mesmo.

No Sistema 2 está o nível tático da organização, que é responsável por monitorar e controlar as operações diárias (Schwaninger; Scheef, 2016), garantindo que as metas e objetivos definidos sejam atingidos e que os recursos sejam utilizados de forma hábil. Por

meio de uma padronização e integração dos projetos, o Sistema 2 garante um amortecimento das oscilações e aumenta a autorregulação.

No caso dos projetos (Sistema 1) tutelados pelo laboratório, o principal artefato que é exigido é o Plano de Trabalho (Universidade federal do Tocantins, 2019). Trata-se de um documento que estabelece as atividades, metas, prazos e recursos necessários para alcançar os objetivos de um projeto. Ele fornece uma visão geral do projeto, descrevendo suas etapas e atividades específicas, bem como os papéis e responsabilidades dos membros da equipe envolvidos no trabalho.

Assim, o plano de trabalho do projeto deve apresentar uma descrição precisa e meticulosa dos objetivos e dos resultados esperados. Além disso, precisa fornecer um inventário dos recursos necessários para a sua conclusão, incluindo materiais de pesquisa, equipamentos e *softwares*, entre outros elementos indispensáveis. O plano deve estabelecer os prazos para cada fase do trabalho e especificar as atividades necessárias para atingir esses marcos.

Além disso, o plano de trabalho se configura como uma ferramenta essencial para o controle e o monitoramento das atividades do projeto, além de gerenciar recursos, o sistema 2 visa propiciar o mínimo de padronização e integração dos projetos.

Nessa vertente, o Sistema 3 é responsável pela gestão operacional da unidade como um todo (Schwaninger; Scheef, 2016), sendo responsável por promover a coesão, buscando garantir o alinhamento das atividades com as metas e objetivos estratégicos da organização. Outra responsabilidade, é coordenar os Sistemas 1 e 2, sendo que no laboratório a função do Sistema 3 é desempenhada pelo coordenador.

Esse profissional tem a responsabilidade de alinhar os objetivos dos projetos com as metas do laboratório como um todo. Para alcançar esse alinhamento, o coordenador realiza a gestão dos recursos, especialmente os de infraestrutura, como servidores e sistemas de armazenamento, que são escassos e têm um custo considerável. O coordenador deve garantir que esses expedientes sejam alocados de forma competente e eficaz, de modo que os projetos possam ser executados com sucesso e os objetivos do laboratório sejam alcançados.

Em se tratando do Sistema 4, este tem como função prover orientação a longo prazo, com foco na exploração estratégica de espaços de atuação potenciais, visando expandir a abrangência da organização (Schwaninger; Scheef, 2016), que é composto por uma unidade conhecida como “sistema de inteligência”. Essa unidade tem a responsabilidade de coletar, analisar e disseminar informações sobre o ambiente externo

à organização. Entre as informações coletadas estão tendências de mercado, alterações regulatórias, atividades da concorrência, avanços tecnológicos e demais dados relevantes para a tomada de decisões estratégicas.

Dessarte, o sistema de inteligência fornece informações para o Sistema 3 (nível gerencial), permitindo que ele tome decisões informadas sobre como a organização deve se adaptar às mudanças do ambiente externo. Também fornece informações para o Sistema 5 (nível estratégico), que é responsável por definir a direção estratégica de longo prazo da organização.

Esse sistema, portanto, é fundamental para a sobrevivência e sucesso da organização, posto que o sistema 4 fornece informações valiosas sobre o ambiente externo, permitindo que a organização se adapte e se antecipe às mudanças do mercado.

No caso deste laboratório, a função de prospectar novas oportunidades no ambiente externo é realizada pelos membros do colegiado do Programa de Pós-Graduação em Modelagem Computacional de Sistemas (PPGMCS). Eles fornecem informações e estudos relevantes ao coordenador do laboratório e ao coordenador do programa para ajudá-los a tomar decisões mais fundamentadas e coesas com a realidade do ambiente externo, em níveis gerencial e estratégico. O colegiado é responsável por nutrir os coordenadores com as informações necessárias para que possam realizar suas funções de forma a realizar um trabalho de excelência.

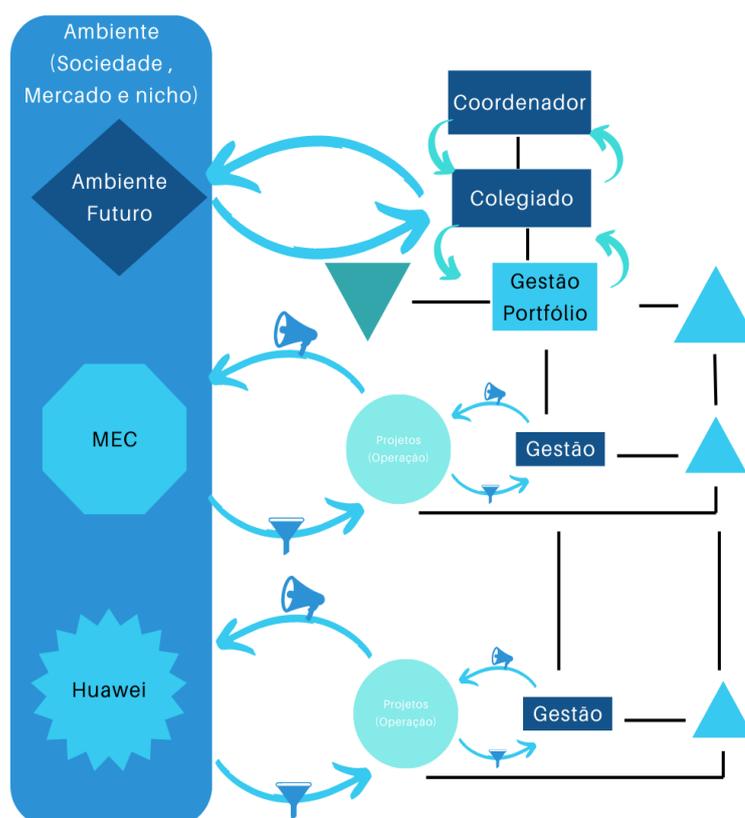
O sistema 5, por sua vez, o Sistema 5 desempenha o papel de assegurar um ambiente interno propício ao debate, permitindo a definição dos valores e propósitos da organização (Alves *et al.*, 2018). Esse sistema é o responsável por determinar a identidade da organização e sua função dentro do sistema de referência mais amplo. Além disso, o Sistema 5 estabelece um equilíbrio entre as orientações presentes e futuras, contemplando um horizonte de tempo de longuíssimo prazo (Schwaninger; Scheef, 2016).

No caso em questão, o sistema 5 é representado pelo coordenador do PPGMCS, que concentra as políticas que orientam o desenvolvimento de estratégias para a inserção e adaptação do laboratório ao ambiente externo. O coordenador lida com a imagem e a reputação do laboratório, avaliando o desempenho dos sistemas anteriores e identificando quaisquer áreas que precisem de melhoria ou ajuste.

Esse profissional é responsável por garantir que o laboratório atinja seus objetivos estratégicos de curto, médio e longo prazo e que o mesmo opere, dentro do possível, de maneira eficiente e com qualidade. Como tal, é importante que o coordenador do PPGMCS esteja integrado e alinhado com os demais sistemas do VSM, a fim de garantir

que o laboratório funcione como um todo, ou seja, de forma coesa e integrado. Com vistas a simplificar a visualização do VSM, aplicada ao laboratório, apresenta-se a Figura 8 a contendo o *design* VSM do laboratório, cujo sistema 1 é formado pelos projetos que interagem com instituições do ambiente externo, como Ministério da Educação e a multinacional Huawei.

Figura 8 – Design VSM do COMAIS



Fonte: O autor.

Na Figura 8 pode-se observar que cada projeto apresenta uma unidade de Gestão, que é responsável por liderar e gerenciar o projeto em todas as suas etapas e que tem a função de definir os objetivos e metas do projeto, elaborar o plano de trabalho, incluindo a seleção da metodologia e dos métodos a serem utilizados.

Ademais, esse sistema coordena a equipe de pesquisadores, fornecendo orientações e assegurando que o projeto está sendo desenvolvido conforme o cronograma preestabelecido. Ele também tem a incumbência de administrar o orçamento do projeto e garantir que os recursos estejam sendo empregados de maneira eficaz.

A unidade de Gestão, dentro de um projeto, tem as atribuições do coordenador do projeto, que além das responsabilidades citadas possui outra importante função, que é comunicar os resultados da pesquisa para a comunidade acadêmica e outros públicos relevantes e deve garantir que todas as etapas da pesquisa estejam em conformidade com as normas éticas e legais aplicáveis, e resolver quaisquer problemas ou desafios que possam surgir durante o processo de investigação.

Em se tratando do Sistema 2, este é representado pelos triângulos (Figura 8), sendo o conjunto de regras que garantem os padrões a serem seguidos pelos projetos, como o já citado plano de trabalho do projeto. Em consonância, o Sistema 3 é a unidade que faz a gestão do portfólio de projetos, que no caso em tela são as atribuições do coordenador do laboratório, que tem função de gerir os recursos, de forma a promover o melhor suporte aos projetos tutelados pelo setor.

O Sistema 4 é a unidade Colegiada do PPGMCS, que se comunica com o Ambiente Futuro para a prospecção de novos projetos e, por último, encontra-se o Sistema 5, no qual o Coordenador é responsável pela estratégia macro do laboratório.

Adicionalmente, é importante destacar que o VSM consolida os elementos necessários para garantir a sustentabilidade de uma organização, utilizando a interação dos cinco sistemas para mapear a operação de um sistema funcional. Com base na análise realizada até este ponto, podemos concluir que o Laboratório incorpora os cinco sistemas preconizados no VSM. Portanto, estabelece-se como uma unidade administrativa efetiva.

Nessa situação, ao utilizar o Modelo de Sistema Viável (VSM) para investigar o conjunto de atividades que visa o aprimoramento acadêmico e socioeconômico a que o Laboratório COMAIS se propõe, emerge um conhecimento que objetiva analisar como essas atividades estão integradas e promove a possibilidade de identificar oportunidades para garantir a excelência do sistema.

Dado o exposto, surgem oportunidades de aperfeiçoamento, como aprimorar o conjunto de regras que compõem o Sistema 2. Por exemplo, padronizar e informatizar o processo de elaboração do plano de trabalho pode assegurar maior aderência e conformidade aos objetivos do Laboratório COMAIS.

2.2 Resultados e Discussão da Seção

De forma assertiva, o laboratório COMAIS tem desempenhado um papel fundamental no desenvolvimento de soluções tecnológicas baseadas em inteligência

artificial, cumprindo as premissas direcionadas pelo professor doutor David Nadler Prata. Por meio de parcerias estratégicas e fontes de recursos financeiros, o laboratório foi equipado e reformado, proporcionando um ambiente propício para a colaboração multidisciplinar e a inovação.

A missão do COMAIS em catalisar a colaboração entre pesquisadores, empresas e governo, aliada à aplicação de técnicas de inteligência artificial, tem gerado resultados significativos. Por meio de projetos desenvolvidos em diversas áreas do conhecimento, o laboratório tem contribuído para a comunidade científica, validando o aprendizado dos acadêmicos envolvidos e enriquecendo a produção científica de forma qualitativa e quantitativa.

A criação de um portal web dedicado ao COMAIS, com informações públicas e uma área restrita aos colaboradores, fortalece a visibilidade das atividades do laboratório e facilita o gerenciamento de projetos. Além disso, o oferecimento de serviços como o *RedCap*, uma plataforma web para construção e gerenciamento de bancos de dados de pesquisa, demonstra o compromisso do laboratório em fornecer soluções tecnológicas relevantes.

A aplicação do Modelo de Sistema Viável (VSM) no COMAIS tem se mostrado decisiva na gestão do laboratório, possibilitando a compreensão dos sistemas que compõem a organização e fornecendo uma estrutura para tomada de decisões estratégicas. Os cinco sistemas do VSM estão presentes e interagem de maneira integrada, garantindo assim, a viabilidade e o bom funcionamento do laboratório.

No entanto, ao aplicar o VSM, identificaram-se oportunidades para aprimoramentos, como a melhoria das regras e processos do Sistema 2, particularmente na padronização e informatização do processo de elaboração do plano de trabalho. Essas melhorias potencializarão a eficiência e promoverão uma maior aderência aos objetivos do laboratório.

Em um cenário em que a inteligência artificial se torna cada vez mais estratégica e impactante em diversas áreas, o COMAIS desempenha um papel relevante no desenvolvimento e aplicação dessa tecnologia. Nesse sentido, o laboratório está alinhado com as diretrizes nacionais de inteligência artificial, como descritas na Estratégia Brasileira de Inteligência Artificial - EBIA (EBIA, 2021), e tem potencial para continuar contribuindo para a inovação e o avanço científico.

Em suma, o laboratório COMAIS se estabeleceu como uma unidade viável e bem-sucedida na modelagem de soluções tecnológicas com base em inteligência artificial. Por

meio de parcerias, recursos financeiros, uma gestão eficiente baseada no VSM e o compromisso em desmistificar e promover o uso de tecnologias inteligentes, o laboratório tem impulsionado a colaboração multidisciplinar e gerado projetos inovadores. Com um futuro promissor, o COMAIS continuará a desempenhar um papel fundamental na transformação tecnológica e na geração de conhecimento.

Assim, nesta tese apresenta-se uma revisão literária sobre o Processamento de Linguagem Natural e Visão Computacional, contextualizando a importância dessas áreas no desenvolvimento de projetos com tecnologias emergentes. Em seguida, são destacados quatro projetos tecnológicos desenvolvidos em colaboração com o Laboratório COMAIS, demonstrando como a aplicação de técnicas de Inteligência Artificial tem impulsionado avanços em diferentes setores.

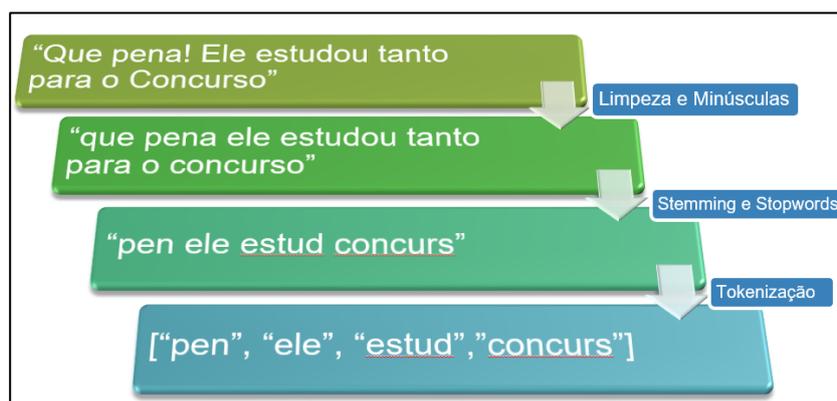
3 PROCESSAMENTO DE LINGUAGEM NATURAL

O Processamento de Linguagem Natural (PLN), conhecido também na academia como linguística computacional, vem crescendo rapidamente, pois suas teorias e métodos estão sendo aplicados em uma gama de novas tecnologias (Bird; Klein; Loper, 2009). Essa área de estudo objetiva fornecer ferramentas para que um sistema computacional seja capaz de lidar com linguagens naturais em diversos níveis, como morfológico, sintático e semântico (Coppin, 2017).

Antes da adoção generalizada do *Deep Learning* (DL) na área de Processamento de Linguagem Natural (PLN), os modelos se fundamentavam na contagem de ocorrências de palavras no texto, sendo elaborados para resolver problemas específicos, como a classificação de texto. Neste contexto, o texto era segmentado em pequenas unidades, conhecidas como *tokens*, que provinham das palavras presentes no texto (Perkins, 2010).

Para normalizar os *tokens* e reduzir a quantidade de dimensões, o texto passa por um processo conhecido como limpeza dos dados (*Data Cleaning*) (Figura 9) e para se reduzir a complexidade dos algoritmos, temos que diminuir o universo de diferentes palavras que ocorrem em um conjunto de textos (vocabulário).

Figura 9 - Sequência de limpeza dos dados



Fonte: O autor.

O processo apresentado na Figura 9 realiza a retirada das pontuações, acentos e números, e converte todas as palavras restantes para minúsculo. Em seguida, são subtraídas as palavras pouco relevantes para o contexto e que ocorrem com muita frequência, as chamadas *stopwords*, e para dar mais generalidade às palavras, é adotada a técnica de *stemming*, que busca remover o sufixo e prefixo das palavras (Lane; Howard; Hapke, 2017) para, ao final da limpeza, termos uma lista de *tokens*.

A ação mais importante no Processamento de Linguagem Natural (PLN) é a conversão do texto em uma representação legível aos algoritmos computacionais, possibilitando a extração de informação contida nos textos por meio de ferramentas tecnológicas (Faria, 2018).

Uma forma de representar o texto é por meio do Modelo de Espaço Vetorial, proposto por Salton, Wong e Yang (1975), que consiste em representar um documento por meio de um vetor formado por um ou mais índices de *tokens* (termos), sendo que cada um é representado por uma posição no vetor (índice) e o conteúdo da posição é preenchido com pesos, ou somente 0 e 1, onde zero significa que o termo não está presente no documento. A dimensionalidade do vetor de representação do documento se dá pela quantidade de termos, logo o tamanho dos vetores é o mesmo da quantidade de palavras contidas no vocabulário. Esse modelo também é conhecido como BOW (*Bag of Words*).

Outra abordagem envolve a utilização de um fator que diminui o peso de relevância de um *token*, baseado na análise da frequência do termo em uma coleção de documentos. Este fator, conhecido como Frequência Inversa de Documento (*Inverse Document Frequency - IDF*), é definido de maneira inversa à quantidade de documentos que contêm o termo, em relação ao total de documentos na coleção (Salton; Buckley, 1988; Jones, 1972).

Diante dos problemas advindos da alta dimensionalidade dos modelos citados, em 2003 foi criado o modelo que usa redes neurais em conjunto com modelos probabilísticos, que gera representações densas, conhecidas como “*embeddings*” (Bengio *et al.*, 2003).

Com o aumento da capacidade computacional, o uso das redes neurais com mais camadas começará a ser mais efetivo, como, por exemplo, as redes neurais recorrentes, que levam a sequência temporal em consideração, possuem a limitação de não serem paralelizáveis, diferentemente das redes convulsionais que podem ser paralelizáveis, mas não guardam informações temporais.

A representação textual também sofreu avanços por meio das redes neurais, em que a representação textual passou a ser gerada por meio do algoritmo *backpropagation*, usando a camada oculta de uma rede treinada. Essa representação é conhecida como *words embeddings* (Mikolov *et al.*, 2013).

Desse modo, os *words embeddings* são vetores pré-treinados, não supervisionados, que funcionam como bons extratores de recursos "universais" e podem ser utilizados em conjuntos de dados, sendo um importante recurso no uso de DL para Processamento de Linguagem Natural (Kim, 2014).

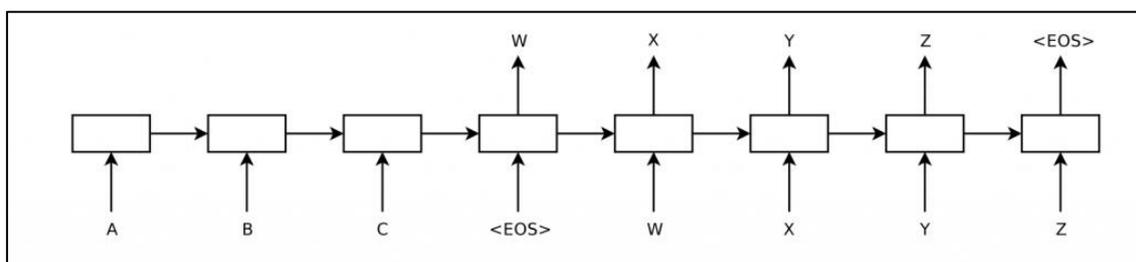
As representações textuais usando *word embedding* apresentam uma maior gama de informações quando comparadas com representações que fazem uso de contagem de frequência de palavras, sendo verdade até mesmo em comparação com modelos que utilizam parâmetros de compensação para efeitos de frequência (Schnabel *et al.*, 2015).

3.1 Seq2seq

Em 2014, surgiu uma metodologia que emprega dois modelos de algoritmos de aprendizado, processando sequências de entrada em uma área ou linguagem específica e gerando representações. Essas representações, por sua vez, alimentam o segundo algoritmo que produz uma sequência em outra área ou linguagem, resultando na criação de uma distribuição de probabilidade.

O nome dessa abordagem é *sequence to sequence* (seq2seq), que busca mapear relações entre sequências. O primeiro modelo baseado em seq2seq utilizou dois algoritmos *Long Short-Term Memory* (LSTM) de apenas 4 camadas, um para entrada de dados, chamada de *Encoder* e outro na saída, denominado *Decoder*. A estrutura *Encoder-Decoder*, apresentada na Figura 10, foi utilizada inicialmente para tradução de texto (*Neural Machine Translations*) (Sutskever; Vinyals; Le, 2014).

Figura 10 - Entrada “ABC” e gera “WXYZ” na saída



Fonte: Sutskever, Vinyals e Le (2014).

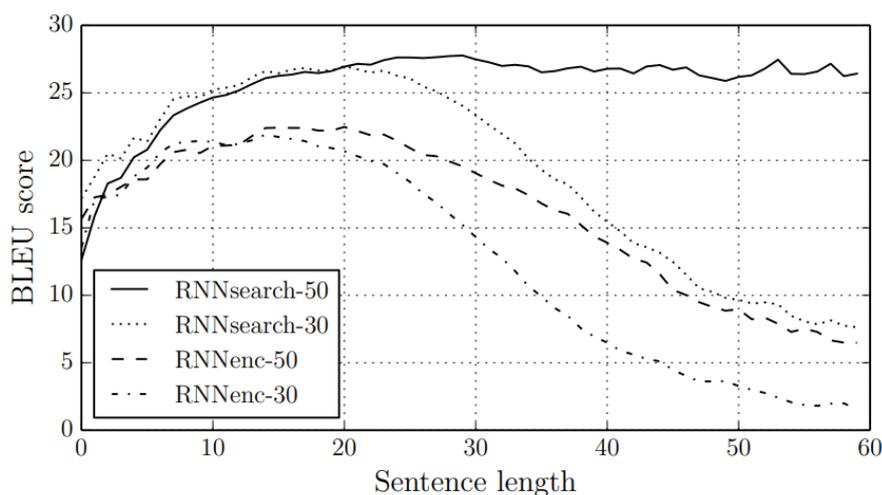
3.2 Attentions

Os modelos iniciais, baseados em *Encoder-Decoder*, apresentaram uma certa fragilidade ao trabalhar com sentenças muito longas, pois só aceitavam entradas de tamanho fixo, logo podendo perder informações de entrada relevantes (Bahdanau; Cho;

Bengio, 2016). Para contornar esse problema, foi criado o mecanismo de *Attention*, que tem a função de guardar informações relevantes da sequência de entrada.

Tais informações são guardadas para auxiliar a predição com base no contexto atual, assim não é necessário tentar guardar toda sequência de entrada em um vetor de tamanho fixo.

Figura 11 - Algoritmos de tradução antes e depois dos *Attentions*



Fonte: Bahdanau, Cho e Bengio (2016).

Na Figura 11 pode-se observar o desempenho da rede *neural RNNsearch-50*, que usa os mecanismos de *Attentions*, mantendo o desempenho em sentenças mais longas.

3.3 Transformers

Transformers é uma arquitetura de modelo de aprendizado baseada na estrutura seq2seq, que visa aprimorar o desempenho, eliminando recorrências e incorporando um mecanismo de atenção conhecido como *self-attention*. Essa arquitetura cria dependências globais, potencializando significativamente a capacidade de paralelização (Vaswani *et al.*, 2017).

Em consonância, a *self-attention* tem a função de mapear relações entre as palavras de uma sentença, não importando a sua posição e guardar a sua relevância para aquela frase. O vetor responsável pela *attention* é adicionado ao vetor de saída do *encoder* e passado para o *decoder*, melhorando assim a qualidade da resposta de saída.

3.4 Modelos de linguagem

Anteriormente, havia uma escassez de dados rotulados para treinamento de algoritmos de aprendizado de máquina, pois para cada tipo de problema a ser resolvido com PLN era necessário um volume considerável de dados, rotulados especificamente para aquele tipo de problema. Essa demanda por dados vem aumentando, devido aos benefícios do uso de *Deep Learning*, que requer grandes volumes de dados na casa dos milhões ou bilhões de dados anotados (Google, 2018).

Para mitigar a falta de dados treinados para um problema específico, foram propostos os Modelos de Linguagem (*Language Model*) - LM, que são representações numéricas do domínio da linguagem. É utilizado para representar palavras ou sentenças no PLN, que são formadas muitas vezes buscando representar a probabilidade estatística para sequência de palavras (Jurafsky; Martin, 2009).

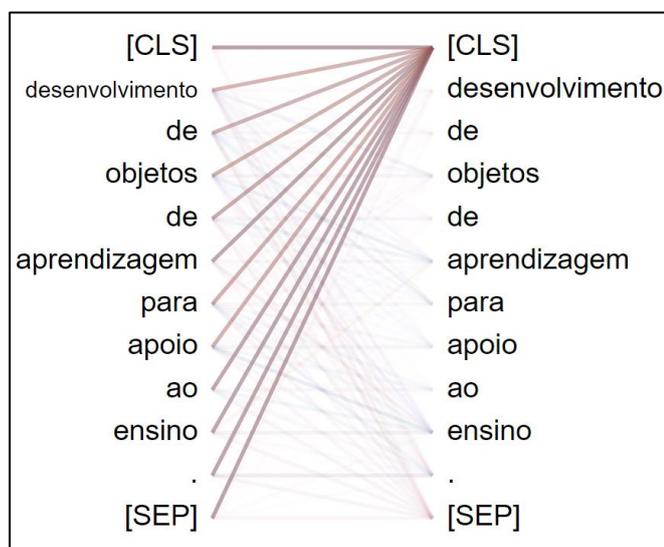
Tais modelos são gerados para serem usados como base em diversos problemas de PLN, como análise de sentimentos, classificação de textos, perguntas e repostas, e outras demandas. Como são criados previamente, são conhecidos como modelos pré-treinados, e trazem consigo a capacidade de relacionar palavras ou sentenças, sem a necessidade de anotação.

Os pré-treinamentos podem ser realizados de forma não supervisionada, usando *autoencoders*, em que as redes usam as sequências de entrada na saída para realizar o pré-treinamento (Dai; Le, 2015). Esses modelos pré-treinados facilitam o processo de aprendizado, pois mapeiam a relação entre as palavras, reduzindo a necessidade de exemplos para problemas específicos (Howard; Ruder, 2018).

3.5 BERT

A plataforma Google, em 2018, juntou na época diversas técnicas emergentes como as citadas *attentions*, *seq2seq*, *transformes*, aprendizado não supervisionado, *autoencoders* e outros, para criar o *Bidirectional Encoder Representations from Transformers* (BERT). Trata-se de um modelo de pré-treinamento de PLN que usa *encoders* bidirecionais para gerar representações profundamente contextualizadas de textos. Por meio de máscaras nas sentenças de entrada é possível realizar a tarefa de predição da palavra que a máscara escondeu nos dois sentidos (bidirecional) da sentença (Google, 2018). A Figura 12 apresenta a relação entre as palavras.

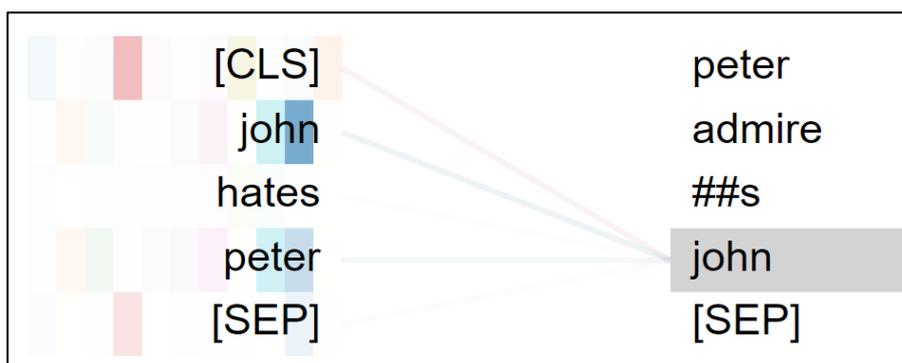
Figura 12 - Visualização do mecanismo de *attention*



Fonte: Vig (2019).

Outra técnica também usada no treinamento do modelo é a previsão da próxima sentença, sendo que diante de uma sentença o modelo tenta aferir se a outra é sua subsequente, assim sendo, o modelo passa criar relações entre sentenças. Os modelos pré-treinados foram criados a partir de textos simples retirados da Wikipedia, quase sem qualquer tratamento (Google, 2018).

Figura 13 - Relação entre sentenças



Fonte: (VIG, 2019)

Os modelos pré-treinados são utilizados para transferência de aprendizado (*transfer learning*) de modelos mais genéricos (BERT), para modelos especialistas. Para isso, ele usa parte do *encoder* para gerar uma representação multidimensional, mantendo

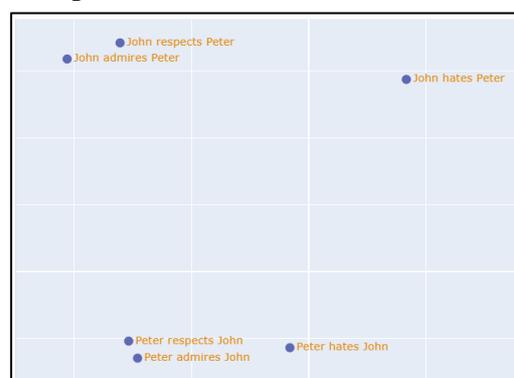
informações das relações linguísticas, como apresentada na Figura 13. Essa representação é usada por um modelo que atente a uma tarefa específica como um classificador de texto.

Nas Figuras 14 e 15 utilizaram-se a técnica de redução de dimensionalidade, o *Principal Component Analysis* (PCA) para visualizar as representações numéricas das sentenças em duas dimensões. A Figura 14 mostra a distribuição espacial de 6 sentenças geradas a partir de um modelo baseado no BERT, com aprendizagem profundamente contextualizada e na figura 15 temos a distribuição das mesmas sentenças utilizando uma representação livre de contexto, no caso *words embedding*.

Figura 14 - Representação livre de contexto



Figura 15- Representação profundamente contextualizada



Fonte: O autor.

Podemos observar que Figura 14, ao usar uma distribuição livre de contexto, as frases que variam apenas os pronomes pessoais ficaram sobrepostas, quer dizer, o modelo desconsiderou os pronomes para representação do contexto. Enquanto Figura 15 além de considerar os pronomes pessoais, considerou-se a ordem em que eles aparecem, e aferiu relevância à ação praticada (verbo), que pode ser observada pela aproximação espacial entre as sentenças com contexto similar.

Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (SBERT). O SBERT é um *framework* capaz de extrair um vetor numérico denso (*embeddings*) a partir de frases ou parágrafos. Tal vetor é utilizado computacionalmente para representar o texto que lhe deu origem (Reimers; Gurevych, 2019).

Adicionando operações de *pooling* (Agrupamento), na saída do BERT (Devlin *et al.*, 2018) é extraído um vetor de tamanho fixo, que tem seus pesos atualizados, fazendo uso de redes siamesas e *triplet network*, construindo vetores de representação semanticamente significativos (Reimers; Gurevych, 2019).

Trate-se de uma técnica considerada computacionalmente eficiente (Reimers; Gurevych, 2019), na qual, ao usar GPU, ela apresenta um incremento de 9% de velocidade em relação ao *InferSent* (Conneau *et al.*, 2018) e é 55% mais rápida que o *Universal Sentence Encoder* (Cer *et al.*, 2018).

Os vetores de representação, gerados pelo SBERT, podem ser usados em diversas tarefas básicas de Processamento de Linguagem Natural, como comparação entre frases, com intuito de checar o quanto são semanticamente parecidas, agrupamento de textos contextualmente próximos e recuperação de informação por meio de busca semântica (Reimers; Gurevych, 2019).

Existem vários modelos pré-treinados capazes de gerar espaços vetoriais usando o SBERT, entre eles temos o modelo multilíngue XLM-R, habilitado com 100 idiomas diferentes e que está dividido em outros dois modelos, um com 768 dimensões XLM-R-Base e o XLM-R e o *Large*, com 1024 dimensões. Neste trabalho usa-se o XLM-R-Base, que supera significativamente, em diversas tarefas de processamento de linguagem natural, o modelo multilíngue BERT (mBert), além de ser bastante competitivo em relação a outros modelos monolíngue (Conneau *et al.*, 2020).

3.6 *K-means*

O *K-means* consiste em um algoritmo de aprendizagem não supervisionada iterativo, com baixa complexidade computacional (Luo; Li; Chung, 2009), em que o número de *clusters* (agrupamentos) é atribuído de forma arbitrária por meio de uma constante. Cada *cluster* se forma em torno de um centroide, que é reposicionado a cada interação, visando se tornar o ponto mais central do *cluster*.

Assim, os artigos são rotulados com base em sua relação com os centroides, ou seja, se o elemento é mais próximo de um centroide em relação aos demais, ele passa a pertencer ao *cluster* do centroide mais próximo e recebe o rótulo desse *cluster*. Isto porque os elementos de um *cluster* tendem a ser similares e diferentes dos não pertencentes ao grupo. O *K-means* itera até que não haja mais movimentações de elementos entre os *cluster*, ou até que o número máximo de iterações tenha sido atingido (Jain, 1988).

Neste capítulo, mergulhamos profundamente nas diversas técnicas e abordagens que moldaram e continuam a evoluir o campo do Processamento de Linguagem Natural (PLN). Desde os primeiros métodos baseados em contagens de ocorrências até as técnicas avançadas como o BERT, vimos a rica tapeçaria de desenvolvimentos que impulsionam

a capacidade dos computadores de entender e interagir com a linguagem humana. O PLN, como evidenciado, não é apenas um domínio acadêmico, mas uma ferramenta prática que está moldando o presente e o futuro da tecnologia.

Conforme o Processamento de Linguagem Natural (PLN) avança, surge a imperatividade de organizar grandes quantidades de dados. Entender a linguagem é apenas o começo; a verdadeira questão é como organizá-la e utilizá-la. No Capítulo 4, focamos na Visão Computacional, uma subárea da Inteligência Artificial que interpreta imagens e vídeos. Essencial em campos como medicina, robótica e segurança, a visão computacional envolve etapas como aquisição de imagens, pré-processamento, segmentação, extração de características e, finalmente, reconhecimento. As técnicas e algoritmos, detecção de objetos por diferentes abordagens e a exploração de *datasets* e ferramentas como *YOLO*, compõem este fascinante domínio que intersecta a análise de imagens e a inteligência artificial.

4 VISÃO COMPUTACIONAL E INTELIGÊNCIA ARTIFICIAL

A visão computacional é uma subárea da inteligência artificial que tem como objetivo desenvolver algoritmos e sistemas que possam interpretar e compreender imagens e vídeos. Com a crescente disponibilidade de dados visuais em diferentes áreas, como medicina, robótica, segurança, entre outras, a visão computacional se tornou uma área fundamental para diversas aplicações.

Nesse contexto, a aplicação de técnicas de visão computacional tem sido amplamente utilizada em diversos campos, tais como na medicina para diagnóstico de anomalias em imagens médicas, na robótica para navegação e interação de robôs em ambientes desconhecidos e na segurança para identificação de comportamentos suspeitos e reconhecimento de objetos, entre outras.

Um sistema de visão computacional é composto por diversas etapas interdependentes e sequenciais que permitem a análise de imagens digitais (Long; Zhang; Feng, 2003). Na primeira etapa, a aquisição de imagens, consiste na captura de imagens em tempo real ou no carregamento de imagens já existentes em um dispositivo de armazenamento.

Em seguida, o pré-processamento é aplicado às imagens, o que pode incluir operações como correção de iluminação, remoção de ruído e normalização de cores para melhorar a qualidade da imagem e prepará-la para as etapas subsequentes. A etapa de segmentação é responsável por identificar e separar objetos ou regiões de interesse dentro da imagem, a fim de facilitar a análise, sendo que extração de características é então realizada para identificar e quantificar características específicas dos objetos ou regiões segmentadas, que podem ser usadas para fins de classificação ou reconhecimento.

Por fim, o reconhecimento envolve a identificação de padrões ou características em uma imagem que correspondem a objetos conhecidos ou classes específicas de objetos. Em conjunto, essas etapas permitem que um sistema de visão computacional analise e entenda imagens digitais para uma ampla variedade de aplicações em diferentes setores.

4.1 Técnicas e Algoritmos em Visão Computacional

As técnicas e algoritmos em visão computacional são divididas em quatro principais categorias: pré-processamento, detecção de características, correspondência de características e reconhecimento de objetos.

O pré-processamento é responsável por transformar a imagem bruta em uma representação mais adequada para o processamento posterior. Essa etapa inclui técnicas como suavização de ruído, equalização de histograma e segmentação de imagens.

A detecção de características, por sua vez, é utilizada para identificar padrões ou objetos específicos em uma imagem. Alguns dos algoritmos mais comuns nessa categoria são o detector de bordas *Canny* (Canny, 1986) e o detector de características *Scale-Invariant Feature Transform* (SIFT) (Lowe, 2004).

A correspondência de características é utilizada para encontrar pontos de correspondência entre duas ou mais imagens. Alguns dos algoritmos mais comuns nessa categoria são o algoritmo de *RANdom SAMple Consensus* (RANSAC) (Fischler; Bolles, 1981) e o algoritmo de casamento de características.

O reconhecimento de objetos é uma tarefa essencial na visão computacional, envolvendo a identificação e a classificação de objetos em imagens ou vídeos. Dada a evolução das tecnologias de captura de imagem e o incremento na disponibilidade de dados, o reconhecimento de objetos tem ganhado relevância e aplicabilidade em diversas áreas.

Nesse cenário, o *Deep Learning* surge como uma técnica poderosa ao utilizar redes neurais profundas, capazes de extrair características complexas de imagens e vídeos, as quais são então utilizadas para realizar com precisão as tarefas de reconhecimento de objetos. A superioridade do *Deep Learning* no reconhecimento de objetos reside em sua habilidade inerente de aprendizagem automática de características complexas.

Essa metodologia, por utilizar redes neurais profundas, aprimora a eficiência na classificação de objetos em larga escala. As redes neurais empregadas são capazes de aprender com enormes volumes de dados, identificando padrões em diferentes níveis de abstração. Este aspecto possibilita a detecção de objetos mesmo em circunstâncias desafiadoras, como variações de iluminação, rotação e escala.

As redes neurais convolucionais (*Convolutional Neural Networks - CNNs*) são redes neurais que têm a capacidade de ler e processar dados advindos de imagens por

meio de diversos processos internos de mapeamento de matrizes e aplicação de funções matemáticas (Feltrin, 2020). Sua arquitetura é composta por camadas de convolução, *pooling* e outras camadas adicionais. A camada de convolução é responsável por aplicar um conjunto de filtros em cada região da imagem, permitindo que a rede identifique padrões específicos em diferentes escalas e rotações. Essa camada é responsável por reduzir a dimensionalidade dos recursos extraídos pelas camadas convolucionais.

Além disso, as CNNs são capazes de lidar com um grande volume de dados de entrada, o que é especialmente importante para a detecção de objetos em imagens de alta resolução. Por isso elas são amplamente utilizadas em aplicações de visão computacional, como a detecção de objetos em imagens médicas, carros autônomos, vigilância por vídeo e reconhecimento facial.

Adicionalmente, a evolução das CNNs levou ao surgimento de diversas arquiteturas eficientes para o reconhecimento de objetos, como a AlexNet (Krizhevsky; Sutskever; Hinton, 2012), VGGNet (Simonyan; Zisserman, 2015), ResNet (He *et al.*, 2015) e InceptionNet (SZEGEDY *et al.*, 2015). Essas arquiteturas possuem diferentes números de camadas e parâmetros e são ajustadas para diferentes conjuntos de dados e tarefas de reconhecimento de objetos.

4.2 Detecção de Objetos

A detecção de objetos, que envolve a identificação da presença e localização de objetos em imagens, representa um pilar fundamental no campo da visão computacional. Há diversas abordagens para essa tarefa que empregam o *Deep Learning*, cada uma com suas peculiaridades e aplicações próprias.

Por exemplo, de acordo com Girshick *et al.* (2014), uma dessas abordagens é o *R-Regions with Convolutional Neural Networks* (CNN), que propõe a utilização de regiões candidatas para a identificação de objetos em uma imagem, combinando-as com uma rede neural convolucional para a classificação.

Ainda, segundo Redmon *et al.* (2016), outra abordagem significativa é o *You Only Look Once* (YOLO), que divide a imagem em uma grade e atribui a cada célula da grade a tarefa de prever objetos baseados no centro do objeto estar na célula. Cada uma dessas abordagens traz suas próprias vantagens e desvantagens em termos de precisão e eficiência, e a escolha de qual utilizar pode depender do contexto e dos requisitos específicos da aplicação.

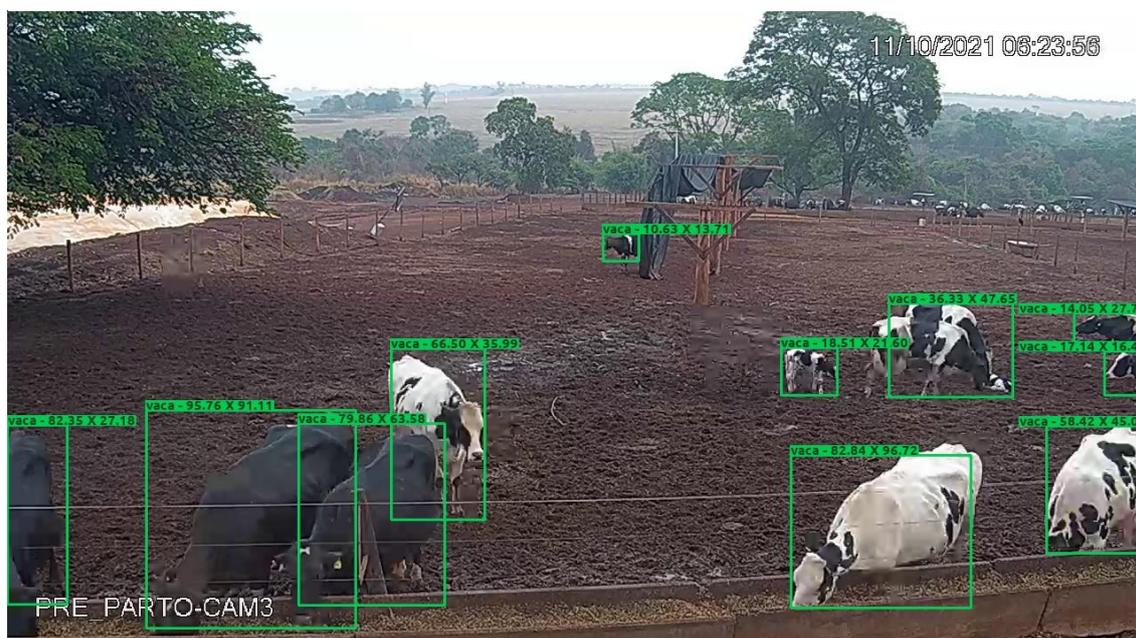
4.2.1 Detecção de objetos baseada em caixas delimitadoras (*bounding boxes*)

A detecção de objetos baseada em caixas delimitadoras (*bounding boxes*) é uma técnica que utiliza redes neurais para detectar objetos em uma imagem e, em seguida, desenhar uma caixa delimitadora em torno de cada objeto.

Esse método não apenas reconhece os objetos, mas também fornece suas localizações espaciais na imagem, desenhando um retângulo delimitador ao redor de cada objeto identificado (Redmon *et al.*, 2016). Tal abordagem é altamente eficaz em tarefas de identificação e localização simultânea de múltiplos objetos em uma imagem.

Ainda, tal técnica é comumente usada em tarefas de detecção de objetos em que é necessário identificar a localização exata dos objetos na imagem. Na Figura 16 apresenta-se um exemplo de detecção de vacas usando *bounding box* que descobre e informa as posições das vacas na imagem.

Figura 16- Detecção de vacas



Fonte: O autor.

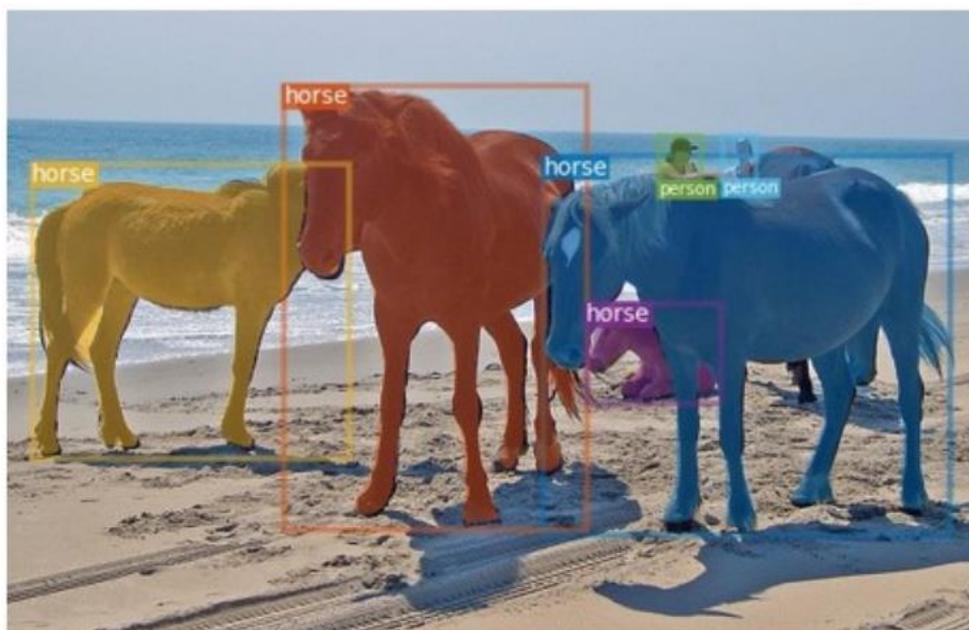
A técnica de detecção baseada em caixas delimitadoras é fundamental quando se necessita especificar a localização exata dos objetos na imagem, o que é particularmente útil em muitas aplicações práticas (Szegedy *et al.*, 2014). Na Figura 16, por exemplo, vê-se o emprego dessa técnica na detecção de vacas.

Desse modo, como delineado por Szegedy *et al.* (2014), a técnica permite não apenas identificar a presença dos objetos (neste caso, as vacas), mas também indicar sua localização precisa na imagem por meio das caixas delimitadoras. Essa capacidade de localizar e identificar múltiplos objetos numa única imagem é crucial para muitas aplicações práticas, como vigilância por vídeo, veículos autônomos, e análise de imagem médica, entre outros.

4.2.2 Detecção de objetos baseada em segmentação

A detecção de objetos baseada em segmentação (*instance segmentation*) é uma técnica que utiliza aprendizado de máquina para segmentar a imagem em regiões que correspondem a diferentes objetos. Como essa técnica busca classificar cada pixel de uma imagem em uma categoria de objeto (Cao *et al.*, 2020), ela é útil para tarefas de detecção de objetos em que é necessário identificar a área exata ocupada por cada objeto na imagem, como em análise de imagens médicas. Na Figura 17 mostram-se três cavalos e duas pessoas detectadas na imagem, em que cada cavalo é identificado como um objeto diferente, embora na mesma categoria.

Figura 17– Segmentação por instância



Fonte: Cao *et al.* (2020).

A Figura 17 ilustra a aplicação dessa técnica, demonstrando a detecção de três cavalos e duas pessoas na imagem. Note que cada cavalo é identificado como um objeto distinto, apesar de pertencerem à mesma categoria. A segmentação de instâncias proporciona um alto nível de detalhe na detecção, permitindo uma interpretação mais precisa e complexa da cena.

Portanto, essa técnica é particularmente útil para tarefas de detecção de objetos em que se precisa identificar a área exata ocupada por cada objeto na imagem. Por exemplo, em análise de imagens médicas, é fundamental distinguir precisamente os tecidos saudáveis dos doentes.

Outra aplicação do uso da técnica de detecção de objetos, baseada em segmentação, é na detecção de ervas daninhas, pois é possível obter uma máscara dos objetos, e assim calcular a área de cobertura da daninha e os pontos precisos da localização dela na imagem. A partir dessa informação o produtor rural pode definir com maior precisão a área de aplicação de herbicida.

4.2.3 Detecção de objetos baseada em pontos-chave (*keypoints*)

A detecção de objetos baseada em pontos-chave (*keypoints*) é uma técnica utilizada para detectar pontos-chave em um objeto e, em seguida, utilizá-los para descrever a forma e as características do objeto.

Essa técnica é comumente usada em tarefas de detecção de objetos em que é necessário identificar a forma e a orientação dos elementos na imagem, como em reconhecimento de gestos, como demonstrado na Figura 18.

Figura 18- Detecção de pontos-chave nas vacas



Fonte: O autor.

Na Figura 18, por exemplo, essa técnica é usada para entender a orientação das vacas na imagem. Os pontos-chave identificados na cabeça e nos pés das vacas permitem inferir que as vacas estão de pé. Isso exemplifica como a detecção de pontos-chave pode fornecer uma percepção mais rica e detalhada do conteúdo de uma imagem, além de simplesmente identificar a presença e a localização de objetos.

Alguns exemplos de algoritmos de detecção de objetos usando *deep learning* incluem o *Faster R-CNN* (Ren *et al.*, 2016) e o *YOLO* (Redmon *et al.*, 2016), entre outros. Cada um desses algoritmos usa uma abordagem diferente para a detecção de objetos e é adequado para diferentes tipos de aplicações.

4.3 Dataset

Em visão computacional, um *dataset* é um conjunto de dados que consiste em um conjunto de imagens digitais e suas respectivas anotações. Esses conjuntos de dados são usados para treinar e validar algoritmos de visão computacional, permitindo que eles aprendam a reconhecer padrões, objetos e características específicas dentro das imagens.

Cada imagem no *dataset* é geralmente acompanhada de anotações, que podem incluir informações como a localização e o contorno de objetos ou regiões de interesse na imagem, bem como outras informações relevantes, como rótulos de classe, descrições e atributos adicionais.

Existem muitos *datasets* de visão computacional disponíveis publicamente, como o *CIFAR-10*, *ImageNet*, *COCO* e muitos outros, que são amplamente usados em pesquisas acadêmicas e desenvolvimento de aplicações comerciais em visão computacional.

As anotações das imagens são realizadas conforme a forma de detecção dos objetos que o sistema utilizará, por exemplo, e se o sistema detectar objeto utilizando caixa delimitadoras (*bounding box*), a marcação das imagens do *dataset* deverá ser realizada por esse meio, que são retângulos que delimitam a área de um objeto de interesse em uma imagem. Mas se o sistema apresenta os objetos detectados com segmentação, a marcação deverá acontecer a partir da criação de uma máscara (polígono) que identifica os pixels pertencentes a um objeto de interesse na imagem, conforme demonstrado na Figura 21.

Os *datasets* de visão computacional são essenciais para o desenvolvimento e aprimoramento de algoritmos de visão computacional, pois fornecem aos pesquisadores e desenvolvedores uma ampla variedade de exemplos para treinar seus modelos e avaliar seu desempenho.

4.4 *Yolo*

O *You Only Look Once (YOLO)* é um *framework* de detecção de objetos em imagens e vídeos desenvolvido para inteligência artificial. Ele se destaca por sua capacidade de realizar a detecção em tempo real com alta precisão (Bochkovskiy; Wang; Liao, 2020).

A abordagem do *YOLO* difere de outros métodos de detecção, pois realiza a detecção de objetos em uma única passada (daí o nome "*You Only Look Once*") (Redmon *et al.*, 2016). Ao contrário de técnicas baseadas em *sliding windows* ou regiões de interesse, o *YOLO* divide a imagem em uma grade e prediz os *bounding boxes* (caixas delimitadoras) e as classes dos objetos diretamente nessas regiões.

O *YOLO* utiliza uma arquitetura de rede neural convolucional (CNN) para realizar a detecção. A rede é treinada em um grande conjunto de dados anotados, em que os

objetos de interesse são marcados com *bounding boxes*. Durante o treinamento, o YOLO aprende a reconhecer padrões e características que são relevantes para a detecção de objetos.

A principal vantagem do *YOLO* é sua eficiência e rapidez. Em vez de avaliar várias regiões da imagem separadamente, como muitos outros métodos, o *YOLO* realiza a detecção em uma única passada pela rede neural. Isso permite que se detecte objetos em tempo real, tornando-o adequado para aplicações que exigem detecção em tempo real, como sistemas de vigilância, veículos autônomos e análise de vídeo em tempo real.

Além disso, o *YOLO* é capaz de detectar múltiplos objetos simultaneamente, mesmo que estejam sobrepostos ou em diferentes escalas. Ele fornece informações sobre as classes dos objetos detectados e as coordenadas precisas dos *bounding boxes*, permitindo uma detecção e localização detalhadas dos objetos em uma cena.

Devido à sua eficiência e precisão, o *YOLO* se tornou amplamente utilizado e é considerado um dos principais *frameworks* para detecção de objetos em tempo real. Diversas versões e variantes do *YOLO* foram desenvolvidas, cada uma com melhorias e otimizações específicas.

4.5 Treinamento

A fase de treinamento na criação de modelos de inteligência artificial, baseados em *You Only Look Once (YOLO)* desempenha um papel fundamental no desenvolvimento de sistemas de detecção de objetos em tempo real. O *YOLO* é reconhecido por sua precisão e eficiência na detecção de objetos. No entanto, para alcançar um desempenho ótimo, é necessário realizar uma fase de treinamento cuidadosa.

Essa etapa começa com a preparação dos dados de treinamento e isso envolve coletar um conjunto representativo de imagens e anotar manualmente os objetos de interesse em cada imagem, criando *bounding boxes* que delimitam esses artefatos. Além disso, é necessário configurar um arquivo de configuração que defina os parâmetros do modelo, como o número de classes de objetos, o tamanho das imagens e os hiperparâmetros de treinamento.

Para iniciar o treinamento, os pesos pré-treinados são utilizados como ponto de partida, que são normalmente obtidos a partir de conjuntos de dados de referência, como o *ImageNet*, e são fundamentais para o ajuste do modelo em relação à tarefa específica de detecção de objetos.

Durante o treinamento, as imagens de treinamento são alimentadas no modelo e comparadas às anotações dos objetos presentes nas imagens. O modelo é otimizado por meio de um algoritmo de otimização, como o Gradiente Descendente Estocástico (SGD), que ajusta os pesos do modelo para minimizar a discrepância entre as saídas do modelo e as anotações. Esse processo iterativo é repetido por várias épocas, refinando continuamente a capacidade do modelo em detectar e localizar objetos.

Avaliações e ajustes são realizados durante o treinamento, utilizando um conjunto de dados de validação separado. Essa etapa permite avaliar a capacidade do modelo em generalizar para novos dados e evitar o *overfitting*. Com base nas avaliações, ajustes nos hiperparâmetros do modelo são realizados, como a taxa de aprendizado e o tamanho das imagens para melhorar o desempenho do modelo.

Após o treinamento, o modelo é testado em um conjunto de dados de teste independente, que contém imagens que não foram vistas durante o processo. Isso permite medir o desempenho final do modelo em termos de detecção e localização de objetos.

Em suma, a fase de treinamento em modelos de inteligência artificial, baseados em *YOLO*, envolve a preparação dos dados, a configuração do modelo, a inicialização dos pesos pré-treinados, o treinamento iterativo, as avaliações e ajustes. Essa fase é essencial para otimizar o desempenho do modelo e aprimorar sua capacidade de detectar e localizar objetos com precisão em imagens.

Depois de explorarmos a Visão Computacional no Capítulo 4, que abrange desde a detecção de objetos até a análise profunda de imagens, o Capítulo 5 nos introduz ao *SmartReview*⁶, ferramenta revolucionária o agrupamento de artigos científicos na revisão sistemática, aproveitando técnicas sofisticadas de processamento de linguagem natural.

Embora a Visão Computacional se concentre na interpretação de imagens, o *SmartReview* se dedica à síntese e análise de textos, especificamente de literatura científica. Em vez de meramente entender a linguagem, o *SmartReview* analisa, categoriza e identifica padrões em artigos, otimizando a revisão sistemática. Entretanto, ao passo que o software avança, é vital reconhecer a essência humana na pesquisa, garantindo que a tecnologia atue como uma ferramenta complementar ao discernimento

⁶ Sediado no Laboratório COMAIS, o *SmartReview* é acessível por meio do *link*: <http://sr.comais.uft.edu.br> e já tem se mostrado eficaz em suas aplicações. Por exemplo, a ferramenta foi crucial na análise e abstração dos dados para a publicação do artigo “e Health-Based Interventions for Older Patients with Prostate Cancer: A Quick Review of the Literature” (Silva Neto *et al.*, 2022), reforçando seu impacto e relevância na pesquisa acadêmica.

e expertise dos pesquisadores. Vamos, portanto, mergulhar na revolução que o *SmartReview* traz para a pesquisa baseada em evidências, seu potencial e suas limitações.

5 SMARTREVIEW: UMA SOLUÇÃO DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA AGRUPAMENTO DE ARTIGOS CIENTÍFICOS NA REVISÃO SISTEMÁTICA

5.1 Introdução

As revisões sistemáticas desempenham um papel crucial na pesquisa baseada em evidências, fornecendo resumos abrangentes e confiáveis das provas existentes (Linde; Willich, 2003). O processo de realização de uma revisão sistemática envolve várias etapas, incluindo seleção de artigos, extração de dados e síntese de evidências. Um exemplo de revisão sistemática é a meta-análise, que utiliza métodos estáticos para agrupar e resumir vários resultados (Cook; Mulrow; Haynes, 1997).

A síntese de evidências dentro de uma revisão sistemática busca agregar conjuntos mais variados de evidências e integrar dados quantitativos com qualitativos, demandando o desenvolvimento de técnicas inovadoras e criativas (Booth *et al.*, 2013). A transparência nos relatórios é um aspecto fundamental para uma revisão sistemática de impacto, pois auxilia os leitores na avaliação da validade e aplicabilidade da revisão (Tricco; Tetzlaff; Moher, 2011).

Em resposta a essas necessidades, o laboratório COMAIS desenvolveu o *SmartReview*, uma solução de processamento de linguagem natural, ferramenta que realiza o agrupamento (*clustering*) de artigos científicos com base em seus contextos, auxiliando no processo de síntese de evidências em revisões sistemáticas.

Portanto, a ferramenta se apresenta como uma solução poderosa para enfrentar os desafios da síntese de evidências, proporcionando um melhor entendimento das relações entre os estudos e facilitando a interpretação dos resultados. O *SmartReview* busca se destacar em seu objetivo, oferecendo benefícios significativos em relação a outras soluções disponíveis.

Além de fornecer resumos abrangentes e confiáveis, a ferramenta promove a eficiência e qualidade das revisões sistemáticas. Com uma capacidade de agrupamento eficiente de artigos científicos, com base em contextos específicos, o *SmartReview* permite uma análise mais integrada e abrangente das evidências existentes.

Por meio do processamento de linguagem natural, o *SmartReview* otimiza o processo de síntese de evidências, facilitando a identificação de padrões e tendências entre

os estudos analisados. Essa abordagem inovadora contribui para o avanço científico e para a tomada de decisões embasadas em evidências sólidas.

Dessa forma, o laboratório COMAIS busca demonstrar seu compromisso em promover soluções tecnológicas que impulsionam a pesquisa baseada em evidências. O *SmartReview* é uma ferramenta para pesquisadores e profissionais da área da saúde, fornecendo resumos abrangentes e confiáveis das evidências existentes, facilitando o acesso a informações relevantes para a tomada de decisões informadas.

5.2 Metodologia

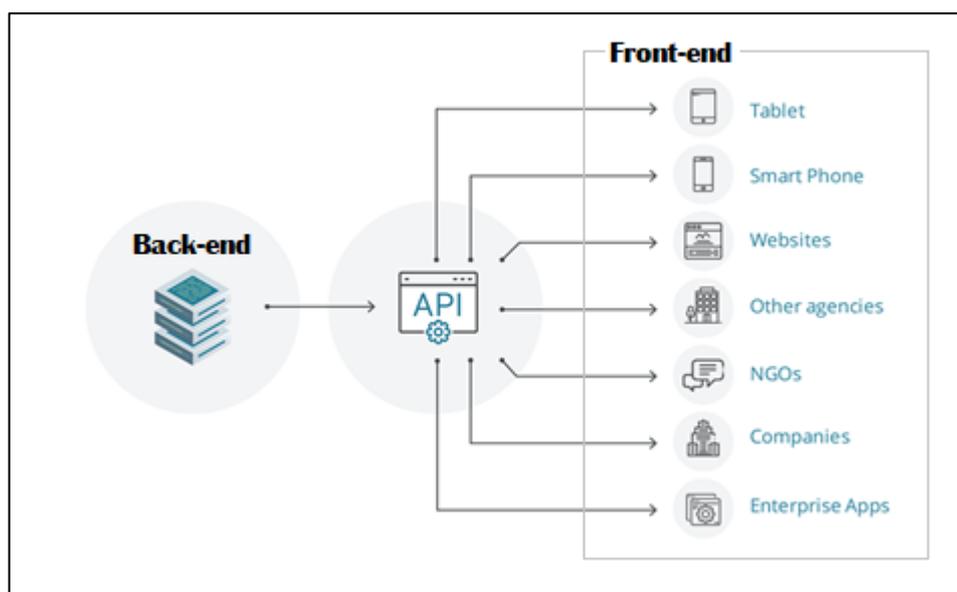
A arquitetura da solução é bifurcada em dois segmentos principais. A primeira é o componente que opera no servidor, responsável pela implementação das regras que constituem o sistema, também conhecido como *back-end*. A segunda componente é uma interface web que opera no cliente, que desempenha o papel crucial de conectar a solução aos usuários, conhecida como *front-end*.

Essas duas partes trabalham em conjunto para garantir um funcionamento fluido e eficiente da solução, proporcionando uma experiência de usuário otimizada. A linguagem utilizada no desenvolvimento das ferramentas, que rodam no servidor e compõe a solução é a *Python*, na versão 3.9. Por ser uma linguagem de alto nível, orientada a objetos, *Python* pode ser utilizado em diversas plataformas pelo fato de ser interpretada (Python.org, 2019).

Essa linguagem tem se mostrado uma boa escolha pela velocidade de desenvolvimento e manutenção e vem se estabelecendo como uma das mais populares da computação científica (Pedregosa *et al.*, 2011). O Python ainda conta com uma comunidade que se orgulha da sua cultura de desenvolvimento, com tradição de APIs bem definidas e uma grande capilaridade de métodos de aprendizagem de máquina (Szymański; Kajdanowicz, 2019).

A interface que roda no cliente (Figura 19) foi desenvolvida por meio de uma biblioteca JavaScript de código aberto, conhecida como *React*, criada pela empresa Facebook em 2013 (Saundariya *et al.*, 2021). Tem como vantagem a capacidade de criação de páginas com componentes dinâmicos e reutilizáveis. Essa interface tem a função de interagir com os usuários, apresentando dados de agrupamentos de artigos científicos conforme o contexto.

Figura 19 - Arquitetura API



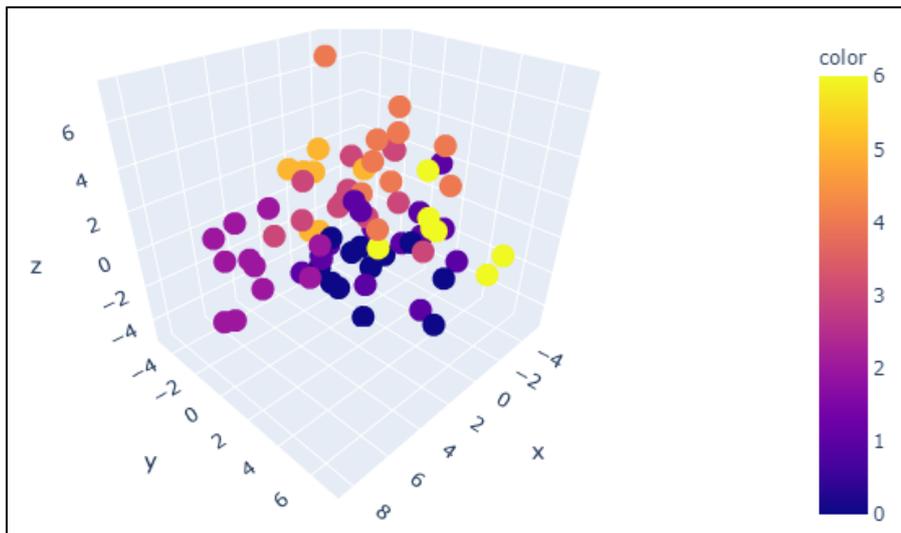
Fonte: O autor.

A interligação entre o *front-end* e o *back-end* se dá por meio de serviços web, que são providos por um servidor por meio de *Application Programming Interface*⁷ (API), como pode ser visualizada na Figura 19, permitindo que diferentes aplicativos, em ambientes distintos, possam requisitar e prover informações independentemente da plataforma ou linguagem em que foram escritos, garantindo assim um baixo acoplamento entre a interface do usuário e o sistema que processa as informações.

A estratégia adotada permite que a solução possa ser integrada em outros sistemas, como, por exemplo, em outros sistemas da UFT ou em aplicativos para dispositivos móveis.

⁷ Interface de Programação de Aplicação.

Figura 21- Aplicação da Análise de Componentes Principais (PCA) para Visualização gráfica da disposição dos resumos em 3 dimensões



Fonte: O autor.

Na Figura 21 é demonstrado como essa alta dimensionalidade é complexa para visualização e interpretação humana, então foi necessário empregar uma técnica de redução de dimensionalidade. Assim, é possível visualizar a disposição dos resumos de maneira mais intuitiva e compreensível em um espaço tridimensional.

Tal redução é uma técnica comumente usada em visualização de dados para tornar possíveis interpretações mais claras e análises mais profundas (Maaten; Hinton, 2008).

5.3 Resultados e Discussão

O *SmartReview* mostra-se uma solução promissora para facilitar o processo de revisão sistemática. Ao automatizar o agrupamento de artigos científicos com base em seus contextos, a ferramenta oferece uma visão mais abrangente e estruturada do conhecimento existente em determinada área de pesquisa. Isso permite que os pesquisadores identifiquem rapidamente padrões, tendências e lacunas no conhecimento, facilitando a síntese de evidências e a tomada de decisões embasadas.

Além disso, o *SmartReview* oferece uma interface amigável que simplifica a interação com a ferramenta. Os usuários podem visualizar os agrupamentos de artigos de forma intuitiva e explorar os detalhes de cada *cluster*, que facilita a compreensão dos resultados e permite uma análise mais aprofundada da pesquisa.

Embora inicialmente desenvolvido para auxiliar na síntese de evidências em revisões sistemáticas, o potencial do *SmartReview* vai além. A aplicação pode ser adaptada para outros contextos, como a seleção de artigos elegíveis para triagem de texto completo ou a análise de tendências e avanços científicos em uma determinada área de pesquisa, ampliando seu escopo de utilidade.

5.4 Conclusão

O *SmartReview* demonstra ser uma solução inovadora para o agrupamento de artigos científicos na revisão sistemática. Por meio de técnicas avançadas de processamento de linguagem natural, a ferramenta automatiza o processo de síntese de evidências, facilitando a identificação de padrões e tendências no conhecimento científico. Com uma arquitetura flexível e interface intuitiva, o *SmartReview* tem o potencial de agilizar o trabalho dos pesquisadores, permitindo uma análise mais eficiente e uma tomada de decisão embasada em evidências confiáveis.

No entanto, é importante ressaltar que o *SmartReview* é uma ferramenta complementar e não substitui o julgamento e a expertise dos pesquisadores. É fundamental que os usuários compreendam os princípios metodológicos e a natureza das evidências utilizadas na revisão sistemática. Além disso, a transparência nos relatórios e a avaliação crítica das evidências continuam sendo elementos essenciais para garantir a qualidade e a confiabilidade das revisões sistemáticas.

Em suma, o *SmartReview* representa um avanço significativo no campo da pesquisa baseada em evidências. Ao automatizar e otimizar o processo de agrupamento de artigos científicos, essa ferramenta contribui para a produção de revisões sistemáticas mais abrangentes e confiáveis. Com seu potencial de reduzir a carga de trabalho e facilitar a identificação de padrões e tendências, o *SmartReview* se torna um aliado valioso para pesquisadores em sua busca por conhecimento e inovação.

Avançando das potencialidades e inovações trazidas pelo *SmartReview* na revisão sistemática no Capítulo 5, o Capítulo 6 nos conduz por uma vertente mais aplicada e prática da tecnologia no mundo real: o setor agrícola. A detecção e localização de ervas daninhas tornam-se críticas no manejo eficiente e sustentável das lavouras. Assim, enquanto o Capítulo 5 enfatiza a análise e agrupamento de textos, o Capítulo 6 foca na aplicação prática dessas técnicas para otimizar e revolucionar o setor agrícola.

6 DETECÇÃO E LOCALIZAÇÃO DE ERVAS DANINHAS PARA O USO INTELIGENTE DE DEFENSIVOS NA AGRICULTURA

6.1 Introdução

A presença de ervas daninhas representa um fator negativo de impacto relevante para a produção agrícola. O controle dessas plantas invasoras, como capim amargoso, carrapicho e buva, geralmente envolve o uso de herbicidas químicos, muitas vezes aplicados de forma indiscriminada em toda a extensão do plantio. Isso resulta em aumento dos custos de produção, possíveis problemas de contaminação ambiental e o surgimento de ervas daninhas mais resistentes.

Nesse contexto, a detecção e localização precisa das ervas daninhas, dentro do campo, são fundamentais para o uso mais inteligente e eficiente desses agentes químicos, visando reduzir o uso desnecessário de defensivos sem comprometer a produtividade. A combinação de informações precisas sobre a presença das ervas daninhas e o uso da tecnologia possibilitam a aplicação de herbicidas de forma mais direcionada e eficaz.

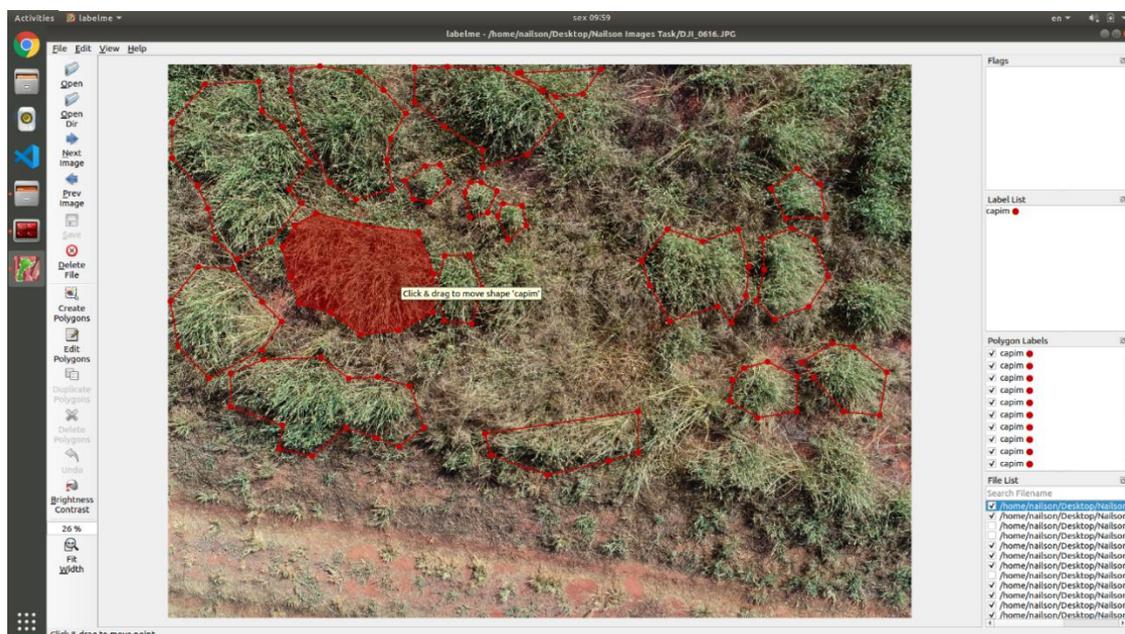
Para realizar a análise e identificação das ervas daninhas, são utilizadas técnicas de visão computacional, que simulam a tarefa realizada por um ser humano ao identificá-las. Desse modo, foi desenvolvida uma *Application Programming Interface* (API), que se baseia nessas técnicas de visão computacional para fornecer informações que auxiliam o agrônomo na tomada de decisão em relação à delimitação da área de pulverização.

Isto posto, o voo de pulverização é inteligentemente planejado pelo módulo de inteligência artificial, levando em consideração a área indicada. Isso permite a seleção adequada do tipo e quantidade de defensivo, além de possibilitar a otimização dos voos para a aplicação dos herbicidas.

6.2 Métodos

O *dataset* para o desenvolvimento do sistema capaz de detectar ervas daninhas utilizou imagens obtidas nos voos de drones, e foram marcadas por meio de polígonos que sobrepõem cada planta, conforme podemos observar na Figura 22.

Figura 22 - Marcação dos polígonos com as daninhas na plantação



Fonte: O autor.

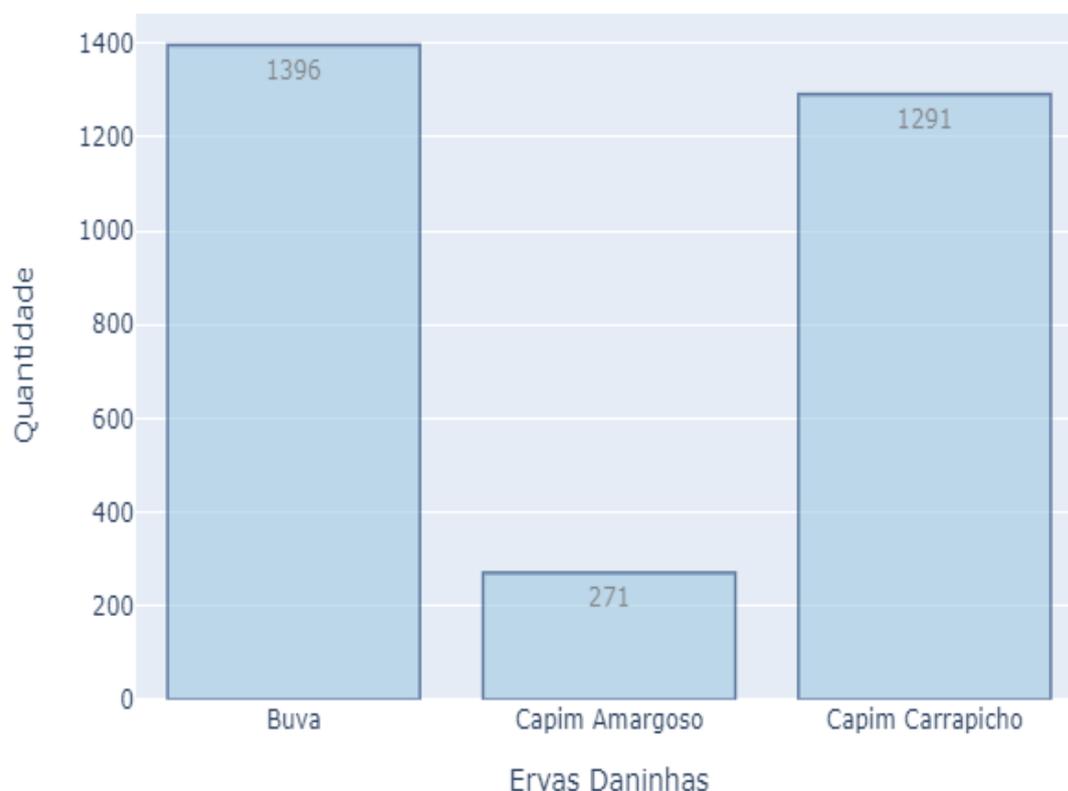
Ao se fazer a análise da Figura 22, nota-se que no processamento dessas imagens esses polígonos informam quais daninhas foram identificadas na imagem, permitindo assim realizar o treinamento da rede neural. A variedade do *dataset*, associado aos treinamentos, podem ser considerados fundamentais para tornar o algoritmo de inteligência artificial mais robusto e, conseqüentemente, mais assertivo.

A etapa de construção do *dataset* requer um grande esforço manual, pois cada erva daninha deve ser marcada manualmente nas imagens capturadas. Esse processo é realizado por um colaborador devidamente treinado por um engenheiro agrônomo, por exemplo, o qual pode explicar as características das plantas e como visualmente identificá-las, possibilitando assim, a marcação de milhares de pontos de referência.

Conforme mencionado, a formação do *dataset* é crucial para que a máquina possa ter os dados para o aprendizado. Para o cenário brasileiro de cultivo de soja, as daninhas que mais causam prejuízos são a buva e o capim amargoso, além do capim carrapicho, que pode causar prejuízos consideráveis nas plantações. Com essa informação definimos que essas seriam as três categorias a serem identificadas, como pode ser visto na Figura 22, onde é possível observar duas das três espécies devidamente marcadas.

No primeiro *dataset* construído, foram identificadas e marcadas cerca de 2958 ervas daninhas, em 150 imagens, distribuídas conforme Figura 24.

Figura 24 - Gráfico com a distribuição das daninhas observadas no ensaio



Fonte: O autor.

Na Figura 24 é possível observar um desbalanceamento na distribuição das espécies de daninha. Isso se deu devido a maior predominância do capim carrapicho e da buva na região sobrevoada.

O balanço entre as três daninhas é almejado para que a rede neural possua melhor capacidade de identificação, independentemente de sua predominância em um determinado talhão. Assim, em condições normais, o algoritmo de inteligência artificial estará preparado para identificar, de forma igual, todos os três tipos de daninhas.

Para a anotação das imagens e montagem do *dataset* foi utilizado o *software open-source LabelMe* (Russell *et al.*, 2008), que disponibiliza uma interface intuitiva de criação de polígonos que correspondem a elementos distintos de cada uma das classes apresentadas, conforme já mostrado na Figura 24.

Esse programa cria arquivos de anotação em formato JSON, o que dá a possibilidade de testar o mesmo *dataset* em algoritmos de *Instance Segmentation* distintos, com o objetivo de escolher aquele que apresenta melhor taxa de acerto dentro de um tempo de execução razoável, para no final obter o resultado desejado, que é a identificação precisa e rápida da daninha.

O processo de treinamento do algoritmo de inteligência artificial, com as imagens que compõem o *dataset*, exige recursos computacionais e tempo de processamento. Assim, na seção seguinte são apresentados alguns modelos de treinamento, os quais devem ser escolhidos em função dos recursos computacionais disponíveis (dispositivo de borda ou nuvem) e de tempo.

6.2.1 Treinamento do modelo inicial

Como a quantidade de amostras de capim amargoso disponível para a criação do *dataset* foi muito inferior em relação às outras classes, o treinamento inicial do modelo só levou em conta as duas classes predominantes: buva e capim carrapicho. Desse modo, como parte do levantamento bibliográfico inicial, foram selecionados alguns algoritmos dentre eles:

Matterport/Mask_RCNN (He *et al.*, 2018): Esse é o algoritmo da Universidade de Cornell que implementa a arquitetura *Mask-RCNN*. Entretanto, no treinamento com as imagens disponíveis foram identificados problemas quando as imagens estavam sendo carregadas na memória. Como a implementação depende do *Tensorflow* 1.0, as otimizações para imagens grandes (cerca de 3000 x 4000 *pixels*) não estavam presentes, causando perdas significativas no processo de treinamento.

NVIDIA/DeepLearningExamples/TensorFlow2/Segmentation/MaskRCNN: Essa é uma implementação da própria NVIDIA do famoso algoritmo *Mask-RCNN*, implementado com *Tensorflow* 2.0 e que possibilita uma diminuição de cerca de 4 vezes no tempo de treinamento, além de ganhos de desempenho na inferência.

O modelo *Mask-RCNN* disponível foi treinado num conjunto de dados *COCO* 2017, disponível em arquivos *TFRecord* (formato otimizado para treinamento) separados entre conjunto de treinamento e conjunto de validação. O projeto tem uma curva de conhecimento muito acentuada que resultou na inviabilidade de se utilizá-lo no tempo hábil de aprendizagem disponível.

PixelLib (Olafenwa, 2021): Esse último provou ter um custo-benefício melhor na relação modelo treinado em função da curva de aprendizagem, uma vez que é possível treinar um modelo de segmentação de instâncias de forma efetiva e com melhor tempo de desenvolvimento para o cronograma do projeto.

Utilizando a ferramenta *PixelLib* para treinar o modelo aqui desenvolvido, um arquivo H5 foi obtido com pesos que configuram uma rede que sabe distinguir a localização de buvas e capim carrapicho. Por meio desse arquivo de pesos, a rede neural pôde ser montada em uma aplicação na nuvem, recebendo as imagens a serem processadas assim que recebidas durante o voo via rede móvel 5G.

A utilização na nuvem se faz necessária em função da alta demanda computacional para processar todas as imagens capturadas, enquanto a tecnologia móvel 5G foi utilizada para o envio de todo o conjunto de imagens capturadas através do voo, agilizando a produção de resultados acionáveis para o agrônomo permitindo, numa mesma visita em campo, a localização exata das ervas daninhas, o tipo de erva daninha e o tipo de herbicida que deveria ser aplicado.

A identificação precisa é crucial dado que a janela de tempo que o agricultor tem entre colheitas e plantio é muito estreita, e que o combate às daninhas, feito de forma correta, aumenta a sua produtividade e elimina a necessidade de aplicação de herbicidas em toda a extensão da lavoura, trazendo economia de tempo e defensivos, sustentabilidade e menor risco de contaminação nas áreas nas quais não há necessidade de correção.

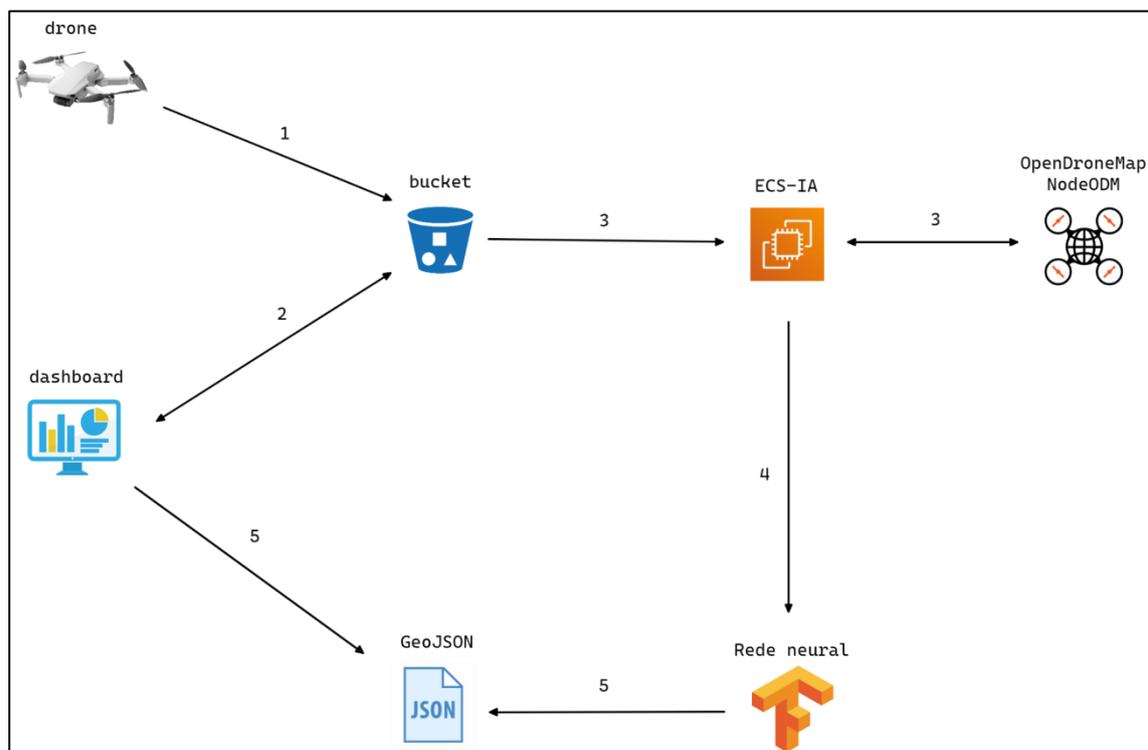
6.2.2 Treinamento do modelo final

Considerando o fato de as imagens em RGB capturadas pelo drone apresentarem uma taxa de sobreposição que garante que toda a área seja efetivamente observada, foi possível notar que se essas imagens fossem diretamente processadas pela IA, várias daninhas seriam detectadas em duplicidade, e filtrar essa duplicidade se tornaria difícil e oneroso.

Para contornar esse problema, e aumentar a precisão, foi adicionada uma etapa no processamento dessas imagens, em que é gerado um ortomosaico RGB, que elimina essa sobreposição. Além disso, o processo permite gerar, com maior facilidade, as imagens

que são exibidas na *dashboard*. O fluxo de todo o processamento pode ser observado na Figura 25.

Figura 25 - Fluxo de processamento das imagens pela IA



Fonte: O autor.

Esse detalhamento pode ser visualizado a seguir:

1. As imagens são enviadas para a nuvem e armazenadas em um *bucket*;
2. A *dashboard* lista o voo e dá a opção para o usuário iniciar o processamento das imagens;
3. É gerado um ortomosaico georreferenciado a partir das imagens RGB sobrepostas do mapeamento;
4. Esse ortomosaico é dividido em partes e utilizado como *input* para a rede neural que detecta as daninhas;
5. Um arquivo GeoJSON é gerado com os polígonos e posições no mapa das ameaças detectadas, permitindo que a *dashboard* apresente essas informações no mapa, de forma precisa e categorizada, por tipo de daninha.

Conforme ilustrado na Figura 25, a infraestrutura de processamento das imagens é constituída por uma máquina ECS na nuvem da Huawei que utiliza o sistema

operacional Windows Server 2016. O sistema operacional e o ambiente de execução são adequados, principalmente para a geração do ortomosaico. É nessa máquina que é feito todo o processamento no qual os ortomosaicos são construídos, utilizando a ferramenta *NodeODM* da organização *OpenDroneMap* (2020). Com isso, as duplicações foram eliminadas, e então a imagem pôde ser subdividida para a correta e precisa detecção de daninhas.

O processamento das imagens para a detecção de daninhas é feito utilizando-se a rede neural *YOLOv5*. O resultado das detecções das daninhas é então computado aproveitando as informações geodésicas que existem nas imagens, podendo ser inferido o ponto geográfico exato da erva daninha detectada.

Esse processo necessita de uma alta capacidade de processamento paralelo, obtida na *Huawei Cloud* com uma GPU Nvidia T4, utilizando o *framework* CUDA. O resultado do processamento das daninhas é exportado para o formato GeoJSON, que é de fácil integração com a *dashboard* web e com os diversos equipamentos de mercado que empregam informações geodésicas.

O modelo foi treinado com base na rede neural o *yolov5l6*, que possui 76,8 milhões de parâmetros, e sua escolha foi feita com base na sua maior capacidade de trabalhar com imagens de alta resolução, como são as imagens obtidas pelo drone, utilizado nesse projeto.

O treinamento de 200 *epochs* durou 1 dia, 13 horas e 25 minutos, atingindo resultado adequado.

Figura 26 - Sensibilidade do modelo proposto

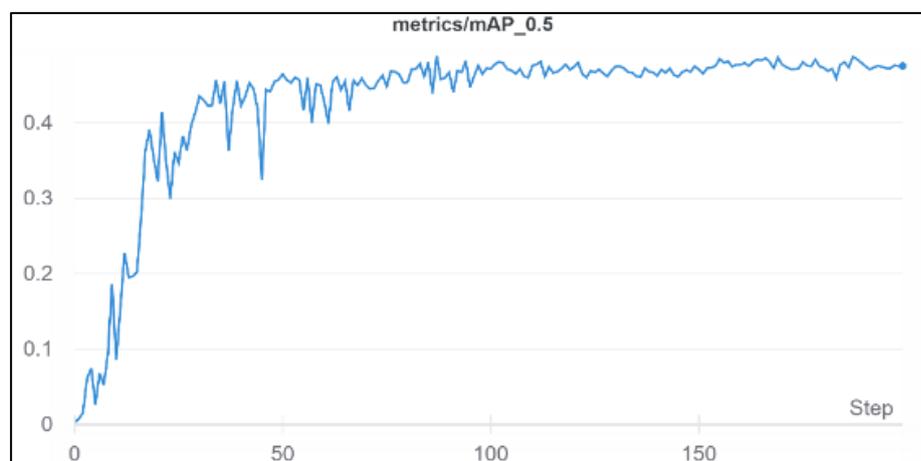


Figura 27- Precisão do modelo proposto

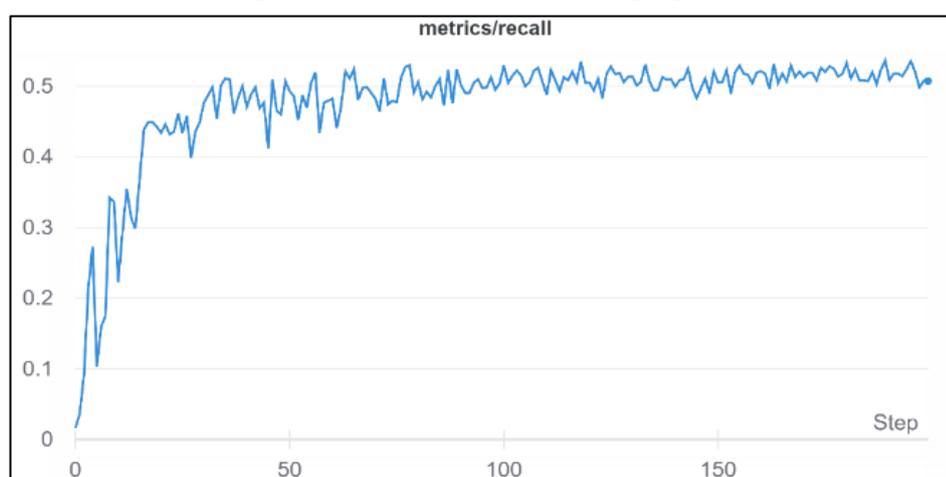
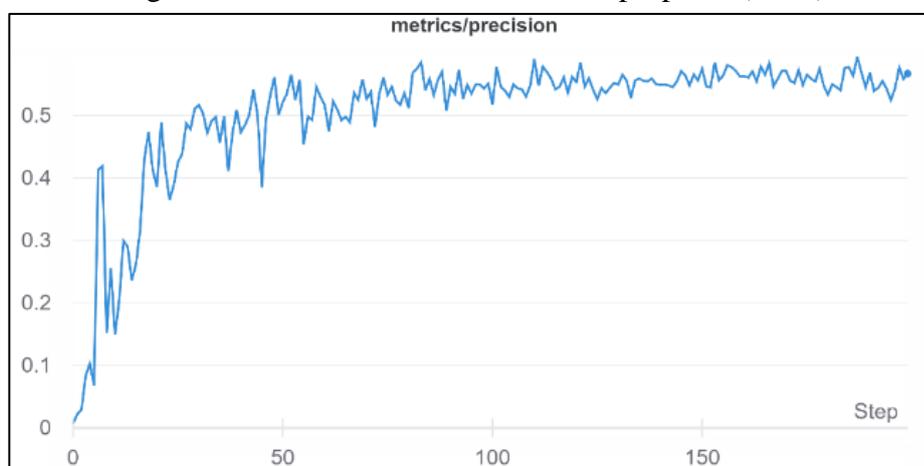


Figura 28 - Precisão média do modelo proposto (mAP)



Fonte: O autor.

Os gráficos no eixo horizontal mostram as *epochs*, que são, de forma simplificada, as etapas de repetidos treinos, até que a IA consiga diferenciar as categorias de daninhas através das imagens.

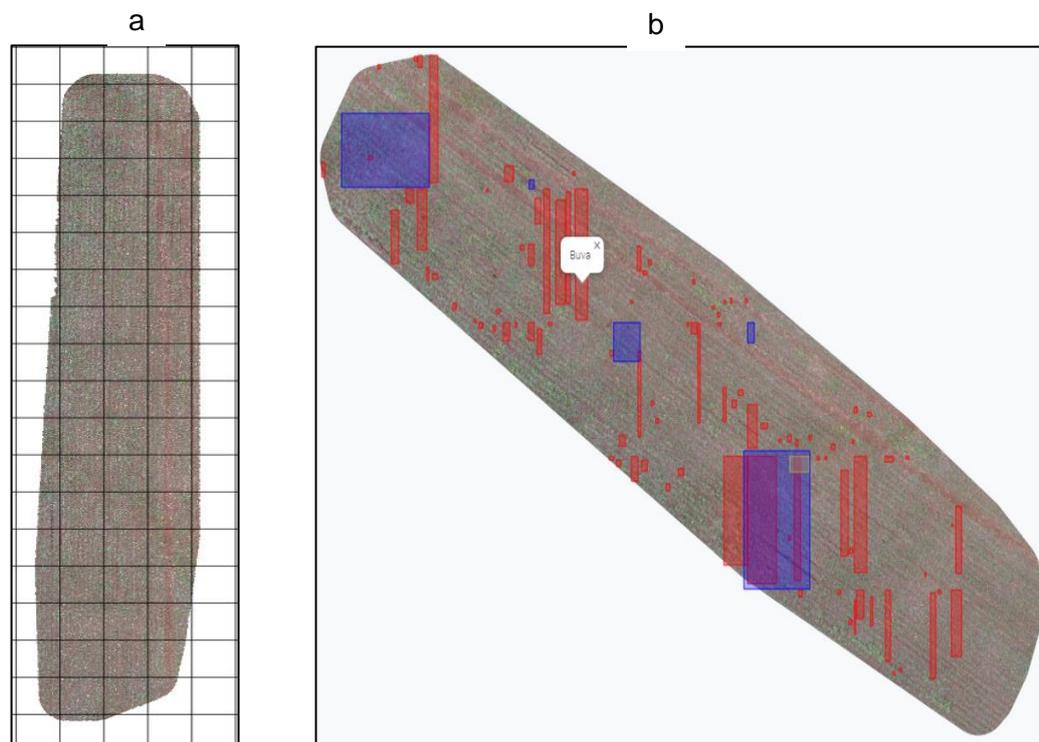
Para realizar a detecção de áreas com infestação de ervas daninha, por meio de imagens RGB, foram utilizadas duas abordagens. A mais simples é aplicar o modelo de detecção foto a foto, uma vez que o modelo foi treinado utilizando marcações em fotos individuais. Logo, submeter as imagens oriundas de registros fotográficos das áreas de plantio requer poucas etapas de pré-processamento. Embora apresente maior agilidade na detecção, tal abordagem não leva em consideração a possibilidade de sobreposição de áreas detectadas, comprometendo assim, de forma significativa, as localizações georreferenciadas das áreas de infestações vistas.

A segunda abordagem visa promover maior qualidade das informações de geolocalização fornecidas pela solução, em comparação à primeira abordagem. Assim, o processo foi realizado utilizando a técnica de fotogrametria, que é a ciência de se obter informações seguras sobre objetos físicos e do meio ambiente, através de processos de registro, medição e interpretação das imagens obtidas por meio de sensores (Coelho; Brito, 2007).

Para garantir que não haja sobreposição de detecções e remover qualquer distorção geométrica (posição e inclinação), a solução tecnológica adotou ortofotos em detrimento do uso de fotos individuais. Ortofotos são objetos cartográficos que descrevem uma região da superfície terrestre, produzidas mediante um conjunto de imagens aéreas, que suprimem deslocamentos da imagem devido à inclinação fotográfica e ao relevo.

Utilizar ortofotos na detecção de ervas daninhas, além de aprimorar a acurácia quanto ao posicionamento das áreas de infestação, fornece ao usuário uma visualização mais compreensível da área de infestação em relação ao plantio como um todo. Para gerar ortofotos sem a necessidade de intervenção humana, a solução acessa as imagens dos voos, que foram depositadas em unidades de armazenamentos que estão na nuvem.

Figura 29 – Ortofotos: (a) segmentada; (b) com geojson das detecções



Fonte: O autor.

Como se pode observar na imagem (Figura 29) contida em uma ortofoto tende a ter dimensões que quase sempre extrapolam as de uma fotografia e, conforme mencionado, pelo fato de o modelo ser treinado utilizando marcações em fotos individuais, se fez necessário realizar a segmentação da imagem, onde cada seguimento possui as dimensões de uma fotografia.

Usando a biblioteca *Geospatial Data Abstraction Library* (GDAL), a solução extrai a matriz que representa a imagem e usa a dimensão esperada para definir a quantidade de linhas e colunas que irá dividir a matriz, conforme apresentado na Figura 29 (a). Cada seguimento é endereçado pela sua linha e coluna e submetido ao modelo de detecção, que por sua vez registra os *bounding box* das plantas invasoras detectadas, o endereço do seguimento e o tipo de planta. Com a intenção de aprimorar a visualização das detecções, os *bounding box* que estão muito próximos ou sobrepostos do mesmo tipo de erva daninha são ordenados e agrupados.

Ao concluir o processamento de todos os segmentos, as detecções são reposicionadas, levando em consideração toda a ortofoto e não apenas o seguimento. Posteriormente, os pontos de posicionamento (x,y) das *bounding box* são convertidos em coordenadas geográficas (latitude e longitude).

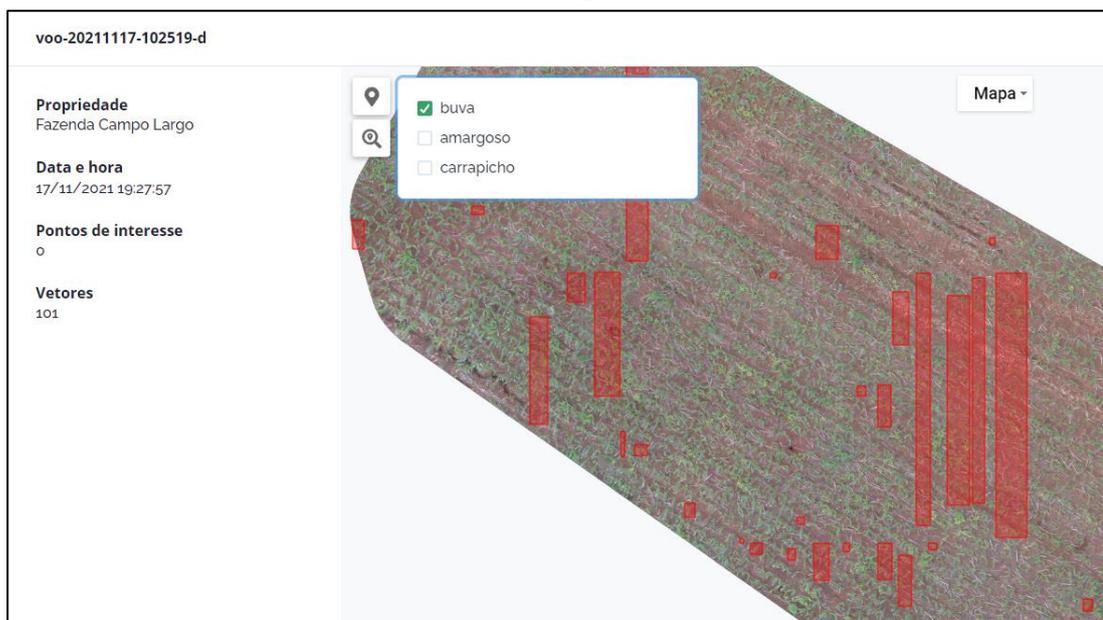
Assim, um arquivo geojson é gerado e encaminhado à *dashboard*, contendo os polígonos formulados com tais coordenadas e cada polígono é marcado e identificado com uma propriedade que informa o tipo de erva daninha observada, conforme apresentado na Figura 29 (b).

6.3 Resultados

Ao final, o usuário pode conferir as imagens georreferenciadas dos pontos de interesse, de forma detalhada. Os orthomosaicos são sobrepostas às imagens obtidas pelo Google Maps, assim se tem uma referência visual e geográfica (vide Figuras 30, 31 e 32), sabendo em qual talhão apresentaram-se as infestações.

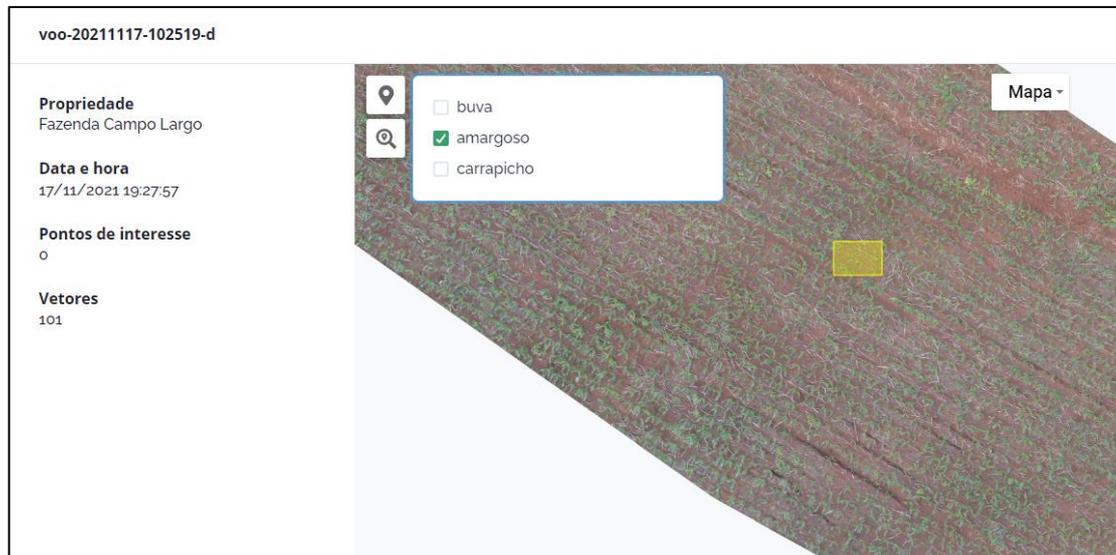
Os pontos de interesse, que são onde há informação de infestação de daninhas, são dispostos em camadas de visualização, permitindo tanto uma visualização geral quanto uma visualização detalhada dessas plantas. Pelo fato de as imagens obtidas para identificação das daninhas terem maior resolução que o próprio Google Maps, é possível mudar a visualização dos mapas usando ferramentas de cartografia como Qgis.

Figura 30 - Página de exibição detalhada do voo, com visão das buvas detectadas pela IA



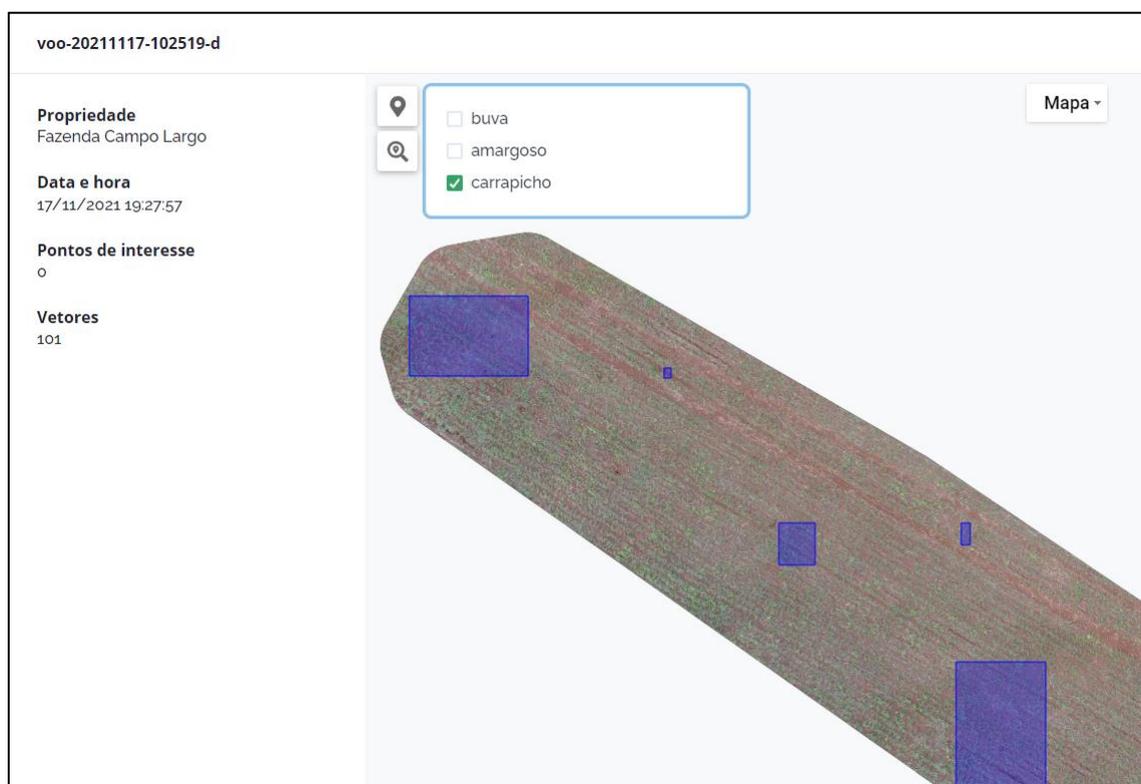
Fonte: O autor.

Figura 31 – Página de exibição detalhada, com visão de capim amargoso detectado pela IA



Fonte: O autor.

Figura 32 - Página de exibição detalhada, com visão de capim carrapicho detectado pela IA



Fonte: O autor

As Figuras 30, 31 e 32 apresentam a página de exibição detalhada do voo, oferecendo uma visão específica das ervas daninhas detectadas pela IA. Nota-se a

presença dos pontos de interesse, que indicam os locais de infestação de ervas daninhas. Esses pontos são organizados em camadas de visualização, proporcionando uma visão abrangente e ao mesmo tempo detalhada dessas plantas.

Interessantemente, devido à resolução superior das imagens obtidas para identificação das ervas daninhas em comparação com as do Google Maps, existe a possibilidade de substituir a visualização dos mapas utilizando ferramentas de cartografia, como Qgis. Isso possibilita uma análise mais precisa e detalhada do local infestado, evidenciando ainda mais a relevância da ferramenta desenvolvida.

6.4 Conclusão

A detecção e localização de ervas daninhas por meio de técnicas de visão computacional são fundamentais para o uso inteligente de defensivos na agricultura, visando reduzir o uso excessivo de herbicidas e minimizar os impactos ambientais. Neste trabalho foi desenvolvida uma solução que utiliza drones para capturar imagens dos campos agrícolas e um algoritmo de inteligência artificial para detectar e localizar as ervas daninhas.

O processo de construção do *dataset* para treinamento do algoritmo envolveu a marcação manual das ervas daninhas em imagens capturadas pelos drones. Foram utilizados polígonos para identificar cada planta invasora, permitindo assim o treinamento da rede neural. Foi dado destaque às três principais espécies de ervas daninhas na região estudada: buva, capim amargoso e capim carrapicho. O desafio foi obter um *dataset* equilibrado, garantindo que a rede neural fosse capaz de identificar todas as espécies de forma igualmente precisa.

Diversos algoritmos foram avaliados, e o modelo final foi treinado utilizando a rede neural *YOLOv5*, que apresentou a melhor relação entre desempenho e precisão para o processamento das imagens de alta resolução obtidas pelos drones. O treinamento do modelo foi realizado com base em um conjunto de 150 imagens, e o resultado obtido foi satisfatório, com as ervas daninhas sendo detectadas e localizadas com precisão.

Para a detecção em escala das ervas daninhas, foram utilizadas duas abordagens. A primeira consiste na detecção foto a foto, que é mais simples e ágil, porém não leva em consideração a possibilidade de sobreposição das áreas detectadas. A segunda abordagem utiliza a técnica de fotogrametria para gerar ortofotos, eliminando a sobreposição e fornecendo informações georreferenciadas mais precisas. Essa abordagem permite uma

visualização mais compreensível das áreas de infestação em relação ao plantio como um todo.

Os resultados obtidos foram promissores, com as ervas daninhas sendo detectadas e localizadas corretamente. As detecções foram registradas em um arquivo GeoJSON, que é facilmente integrado a uma *dashboard* web, permitindo a visualização das áreas de infestação no mapa, categorizadas por tipo de erva daninha. A solução tecnológica desenvolvida fornece ao agricultor informações precisas sobre as áreas com infestação, possibilitando uma aplicação mais inteligente e precisa dos defensivos agrícolas.

Em conclusão, a detecção e localização de ervas daninhas, por meio de visão computacional e algoritmos de inteligência artificial, ofereceram uma abordagem eficiente e sustentável para o uso de defensivos na agricultura. Assim, a combinação de drones, processamento de imagens, técnicas de fotogrametria e treinamento de redes neurais permite uma detecção precisa e uma aplicação mais eficiente de herbicidas, reduzindo custos, minimizando impactos ambientais e aumentando a produtividade agrícola.

No Capítulo 6, abordou-se a aplicação meticulosa de visão computacional e técnicas de inteligência artificial para a detecção e gestão de ervas daninhas em campos agrícolas, visando práticas mais sustentáveis e eficazes. Por outro lado, o Capítulo 7 transita para a análise sistemática do comportamento animal, utilizando-se de abordagens similares em visão computacional e aprendizado supervisionado para monitoramento em contextos pecuários. Embora diferindo em aplicação, ambos os capítulos enfatizam a crescente intersecção da tecnologia com setores tradicionais, ilustrando o potencial transformativo da inteligência artificial na otimização de práticas e no enriquecimento do conhecimento em áreas críticas.

7 DETECÇÃO DE COMPORTAMENTO ANIMAL EM VÍDEOS UTILIZANDO VISÃO COMPUTACIONAL E APRENDIZADO SUPERVISIONADO

7.1 Introdução

Ao analisar e observar o comportamento animal, são utilizadas técnicas de visão computacional que imitam a tarefa feita por um humano em identificar os animais e seus comportamentos. Para observar o comportamento animal e capturar os eventos chave, a forma apropriada é o vídeo.

As imagens, mesmo que em intervalos curtos de tempo, não são suficientes para observar uma monta, a qual costuma durar por volta de dois segundos. Para isso, são utilizadas câmeras de vídeo instaladas em pontos estratégicos de maior movimentação dos animais, e assim é possível adquirir as imagens para o desenvolvimento e execução de um algoritmo de inteligência artificial.

O fluxo envolve elaboração e instalação da estrutura, gravação das imagens, montagem do banco de imagens (*dataset*) a partir das imagens gravadas, treinamento da IA, e execução da IA. A escolha da técnica levou em conta que os animais deveriam ser observados em tempo real, por vídeo de forma ininterrupta, 24hs/dia, 7 dias/semana.

Dentre as categorias de algoritmos de inteligência artificial, existe o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço. A que apresenta os melhores resultados para as tarefas que exigem visão computacional com vídeo em tempo real é o aprendizado supervisionado, considerando que os comportamentos são bem definidos, claramente identificados e categorizados, sendo possível criar categorias em uma imagem, facilitando assim o alerta dos eventos.

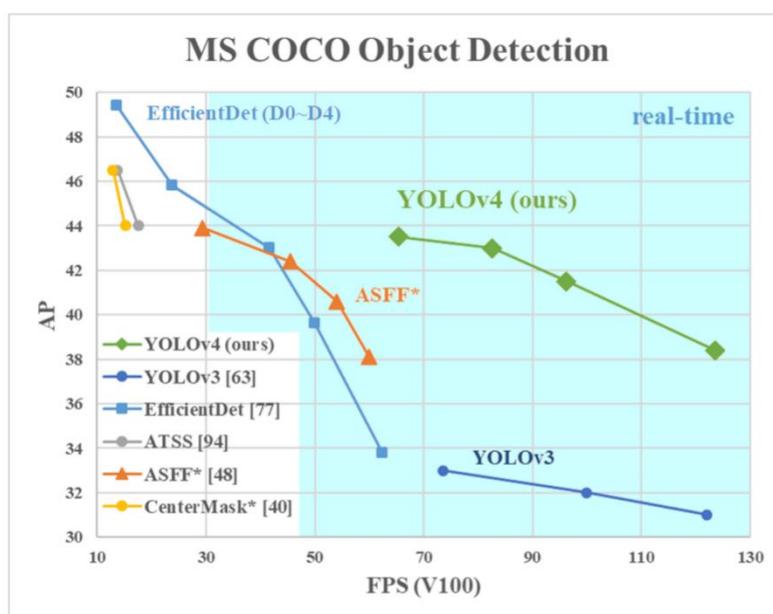
7.2 Método

Para que os eventos fossem propagados desde sua observação pelas câmeras, passando pela IA, e para que fosse possível receber os alertas foi necessário construir um *backend Web*, que recebeu da IA as observações feitas, e um aplicativo mobile para que fosse possível não somente receber uma mensagem de alerta, mas sim receber uma imagem do acontecimento e um vídeo curto. O *backend web* foi desenvolvido em Django, um *framework web* para a linguagem *Python*.

Para o aplicativo mobile optou-se por utilizar o *framework React Native*, que permite desenvolvimento multi-plataforma, podendo atingir usuários de android e iOS. Especialmente para a identificação desses eventos em vídeo, a IA tem um período curto para fazer a inferência de cada um dos *frames* de vídeo. Em um vídeo a 30 *frames* por segundo o tempo máximo é de 33 ms (milissegundos), uma inferência mais longa que isso, por mais precisa que for, não é rápida o suficiente.

E mesmo que seja rápida, se essa detecção não for suficientemente precisa, esta não irá detectar os comportamentos de forma adequada. Assim, a escolha de uma arquitetura de rede neural deve levar em consideração essas duas métricas, precisão e velocidade. No gráfico da Figura 33 é possível comparar no eixo vertical a taxa de acerto e no eixo horizontal a velocidade em FPS (*Frames* por segundos) que a rede é capaz de executar. Vale observar que são valores sintéticos, para que diferentes modelos de rede neurais possam ser comparados entre si, e que possa ser feita a melhor escolha para o problema a ser resolvido.

Figura 33 - Taxa de acerto e velocidade em FPS para diferentes *stacks*



Fonte: Bochkovski (2020).

A opção escolhida, por ter esse melhor balanço entre precisão e velocidade, é a YOLOv4 (Bochkovskiy; Wang; Liao, 2020), a qual no início do projeto era o estado da arte dentre as redes neurais para esse tipo de detecção em tempo real.

Para a sua implementação, optou-se por utilizar a linguagem *Python* e as bibliotecas *Open Computer Vision* (OpenCV), e *CUDA Deep Neural Network* (CuDNN) feita para o *hardware* da NVIDIA. Tal opção se deu pela primeira trabalhar com manipulação de imagens e a segunda por ter um bom suporte, integração e documentação com a primeira, agilizando o desenvolvimento do programa para dar foco na solução como um todo, dada a vasta oferta de *hardware* de NVIDIA no mercado, desde os computadores de desenvolvimento às instâncias de GPU na *Huawei Cloud*. Assim foi delimitado a *stack* a ser utilizada ao longo do projeto.

Para o desenvolvimento da IA que identificasse as vacas e o comportamento de cio, além de detectar bezerros foi necessário primeiro obter um banco de imagens próprio, uma vez que não se tem bancos de imagens públicos à disposição, com quantidade e variedade suficientes para os requisitos do projeto. Por uma fazenda ser um ambiente aberto e que passa por variações tanto de iluminação como de cenário, devido às estações do ano, é necessário garantir aderência do que foi desenvolvido com o ambiente a ser observado.

A formação desse banco de imagens utilizou as mesmas câmeras que iriam observar o comportamento animal. Assim, foi adquirido um conjunto de câmeras de alta resolução *Pan-Tilt-Zoom* (PTZ) *ethernet*, um *Network Video Recorder* (NVR) e equipamentos de rede para permitir conexões remotas via rede 5G.

Com a gravação das imagens a equipe técnica modelou os *datasets*, como são também conhecidos esses bancos de imagens. Esse tipo de câmera pode ajustar a posição a ser observada com grande facilidade, que além de evitar deslocamentos da equipe para ajustes simples, permite que as câmeras sejam ajustadas conforme o ambiente.

Por ser uma câmera de rede, ela já provê os recursos para que um *software* possa consumir as imagens capturadas em tempo real, e o NVR, por poder capturar imagens 24 horas por dia, evita que qualquer comportamento seja perdido, o que é de grande valor para a própria formação e precisão do *dataset*.

Com os vídeos obtidos foi possível capturar os dois tipos de evento (cio e parto), seguindo para a etapa de etiquetagem, na qual os comportamentos eram identificados por marcação, para posterior treinamento. Assim, se obteve o material para o

desenvolvimento do *software* que usa a mesma estrutura para a observação, em tempo real, dos animais.

Para o treinamento foram anotadas um total aproximado de 2000 imagens nesse processo, conforme apresentado na Figura 33, na qual observa-se em verde a marcação de monta e em vermelho a marcação de vaca.

As imagens foram manualmente etiquetadas, com as categorias desejadas. Essa é uma tarefa longa, mas crucial para que a IA possa identificar os comportamentos desejados. Essa formação de *dataset* não é feita somente uma vez, sendo necessária que se repita algumas vezes para se obter imagens nas mais diversas posições do animal, e que para esse cenário inclua variedade climática.

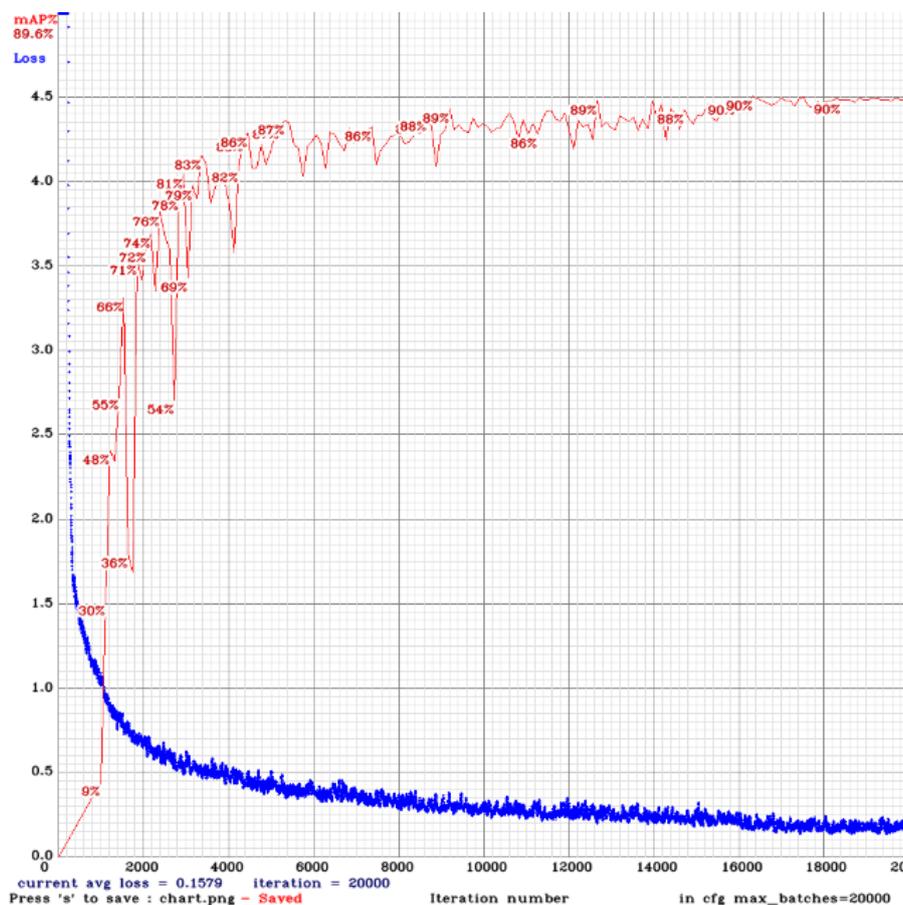
Se esse processo não for seguido, pode acontecer o fenômeno conhecido como *data drift*, onde o *dataset* e o treinamento não condizem mais com a realidade observada, levando à perda na eficiência das detecções. Durante o desenvolvimento do projeto, foi possível observar esse fenômeno no início das chuvas, sendo que uma quantidade significativa de falsos positivos surgiu, pois todo conjunto de *dataset* tinha sido modelado durante o período de seca.

O treinamento, a partir das imagens do *dataset* formado é executado, como mencionado anteriormente, para a rede neural que se deseja utilizar. Para garantir que irá resultar em boas detecções as métricas durante o treinamento: *mean average precision - mAP*, que indica a precisão média e o *loss*, que é a taxa de erro, devem ser exaustivamente analisadas.

O objetivo do treino é que seja atingida a maior mAP e o menor *loss* possível, a precisão chega a 90% e a perda abaixo de 0.5%. Uma vez que se completa o treinamento, os pesos da rede estarão ajustados, podendo então serem testados com o ambiente que se destina observar.

Ao realizar todos os testes, são feitas medições com base em quantas vezes aquele treino acertou e quantas errou, definindo a precisão. A seguir, a linha vermelha, presente na Figura 34 representa a curva de mAP subindo ao longo do treino e em azul a curva de *loss* diminuindo, alcançando assim a melhor combinação possível.

Figura 34 - Curva de treinamento relacionando precisão e perda



Fonte: O autor.

Assim, reitera-se que objetivo principal do processo de treinamento, que é alcançar a mais alta mAP (precisão média) e a menor perda (*loss*) possível. Na prática, busca-se uma precisão de 90% e uma perda abaixo de 0,5%. A Figura 34 ilustra esse processo por meio de duas curvas distintas: a linha vermelha, que representa a evolução da mAP ao longo do treinamento, e a linha azul, que mostra a diminuição do *loss*. Conforme o treinamento avança, observa-se uma tendência de aumento da mAP e redução do *loss*, indicando uma otimização do modelo.

Essa visualização é crucial para entender a eficácia do treinamento. Após a conclusão do treinamento, os pesos da rede neural estarão otimizados, permitindo que o modelo seja testado no ambiente específico para o qual foi projetado. Ao realizar esses testes, é possível medir a precisão do modelo com base no número de acertos e erros, fornecendo uma indicação clara de sua eficiência.

7.3 Resultados e Discussão

Uma taxa de erro de 10% é considerada alta, uma vez que uma taxa de erro desejável é em torno de 1 a 3%. Caso a taxa de erro atinja o limiar alto, as novas observações são analisadas e comparadas com o *dataset* para que se entenda onde o modelo não acertou, e com base nessas análises se ajusta o *dataset* para um novo treino, podendo envolver a inclusão de novas imagens que não foram observadas anteriormente, se rejeitam imagens que possam estar “confundindo” o modelo e podem ser necessárias mais etapas de *data augmentation*. Esse ciclo é feito até que se considere indispensável e que o modelo seja considerado pronto.

Entretanto, isso não impede que no futuro tenha que ser feito novos treinamentos. A vantagem, no entanto, no momento que se tem a infraestrutura tanto física quanto de *software* é que as próprias observações podem retroalimentar os futuros treinos. Apesar de parecer possível, isso ainda não é inteiramente automático, uma vez que a equipe de IA deve avaliar as métricas e saber o que incluir e o que não incluir no novo treinamento, objetivando sempre a maior acurácia.

Para a observação da monta constatou-se que a categorização foi suficiente para a detecção do fenômeno, dado que a monta é um “objeto” bem diferente da vaca, como observado na Figura 35, em que foi possível alertar o produtor com a detecção dessa categoria em um modelo treinado para isso.

Figura 35 - Detecção de monta



Fonte: O autor.

No que diz respeito à observação da monta, verificou-se que a categorização adotada foi eficaz na detecção desse comportamento específico. A monta apresenta características distintas em relação à vaca, tornando sua detecção mais simplificada. A Figura 35 exemplifica essa situação, na qual o sistema pôde notificar o produtor ao identificar essa categoria, graças a um modelo previamente treinado para tal finalidade.

Embora possa parecer uma tarefa simples, a realidade é que esse processo não é totalmente automatizado. A equipe responsável pela inteligência artificial precisa avaliar cuidadosamente as métricas, decidindo o que deve ou não ser incluído no próximo ciclo de treinamento, sempre buscando a máxima precisão possível.

7.4 Conclusão

A detecção e análise do comportamento animal em vídeos utilizando técnicas de visão computacional e aprendizado supervisionado são essenciais para diversas aplicações, como monitoramento de rebanhos, estudos comportamentais e controle de saúde animal.

Neste trabalho foi desenvolvida uma abordagem que utiliza câmeras de vídeo para capturar as imagens dos animais em tempo real e um algoritmo de inteligência artificial,

baseado na arquitetura *YOLOv4* para a detecção e classificação dos comportamentos desejados.

A escolha da arquitetura *YOLOv4* se mostrou adequada, pois ela apresenta um bom equilíbrio entre velocidade e precisão, sendo capaz de realizar inferências em tempo real. Utilizamos a biblioteca OpenCV em conjunto com o *hardware* CUDA da NVIDIA para implementar e executar o algoritmo de detecção. Além disso, desenvolvemos um *backend web* e um aplicativo mobile para receber os alertas e exibir imagens e vídeos dos eventos detectados.

A formação do banco de imagens próprio foi fundamental para garantir a aderência do algoritmo ao ambiente de observação, realizado por meio de câmeras de alta resolução PTZ e um NVR para capturar as imagens em tempo real, permitindo a gravação contínua e evitando a perda de comportamentos importantes. A etiquetagem manual das imagens foi realizada para criar um *dataset* adequado ao treinamento da IA.

Durante o treinamento da rede neural, monitoramos as métricas de precisão média (mAP) e perda (*loss*) para garantir que o modelo fosse capaz de realizar detecções precisas dos comportamentos desejados em que foram realizadas várias iterações de treinamento, ajustando o *dataset* e incluindo técnicas de *data augmentation*, quando necessário, para melhorar a acurácia do modelo.

Os resultados obtidos foram promissores, com uma taxa de acerto de aproximadamente 90% e uma taxa de erro abaixo de 0,5%. No entanto, é importante ressaltar que a melhoria contínua do modelo é necessária, uma vez que o ambiente e os comportamentos dos animais podem mudar ao longo do tempo. Desse modo, o ciclo de retroalimentação, com base nas observações reais, permite aprimorar o modelo e garantir sua eficácia ao longo do tempo.

Em suma, a detecção de comportamento animal em vídeos, utilizando visão computacional e aprendizado supervisionado, oferece uma abordagem eficiente e precisa para monitorar e analisar o comportamento dos animais. A combinação da tecnologia de câmeras de vídeo, algoritmos de inteligência artificial e infraestrutura adequada permite a detecção e classificação de comportamentos importantes para a tomada de decisões no contexto pecuário.

No Capítulo 7, explorou-se a detecção e análise de comportamentos animais por meio de técnicas de visão computacional, com enfoque específico na observação geral do rebanho. Ao avançarmos para o Capítulo 8, a pesquisa adentra uma especialização mais apurada, concentrando-se na identificação discriminativa entre bezerros e vacas. Esta

transição ilustra a evolução da aplicação de técnicas sofisticadas de inteligência artificial para situações específicas e desafiadoras no contexto pecuário. Ao incorporar a detecção de pose e a arquitetura *ResNet-50*, o estudo reitera o valor de abordagens complementares e a integração de múltiplas técnicas para aprimorar a acurácia e aplicabilidade prática dessas soluções tecnológicas em ambientes agrícolas.

8 DETECÇÃO DE BEZERROS EM VÍDEOS UTILIZANDO MEDIÇÃO DE TAMANHO E DETECÇÃO DE POSE

8.1 Introdução

A detecção de bezerros em vídeos é uma tarefa desafiadora devido à semelhança visual entre esses animais e as vacas, sendo que a identificação correta dos bezerros é fundamental em diversas aplicações, como monitoramento de rebanhos, controle de saúde animal e estudos comportamentais. No entanto, modelos preexistentes frequentemente confundem bezerros com vacas e vice-versa, o que impacta negativamente na precisão e confiabilidade dos sistemas de detecção.

Aqui se propôs uma abordagem inovadora, baseada na medição de tamanho e detecção de pose para resolver o problema de detecção de bezerros em vídeos. A ideia principal é aproveitar a diferença de tamanho entre bezerros e vacas, utilizando apenas uma câmera, para discernir corretamente os animais. Para isso, são realizadas medições de altura por meio de prévias calibrações com referências conhecidas presentes nas imagens.

Entretanto, essa abordagem inicial sozinha não resolve totalmente o problema, pois algumas vacas deitadas em ângulos específicos apresentam medidas semelhantes às dos bezerros. Para superar esse desafio, empregou-se uma rede neural treinada para a detecção de pose dos animais. Quando a detecção de uma vaca apresenta medidas semelhantes às de um bezerro, a imagem é submetida a essa rede neural, que identifica as partes do animal detectadas, como a cabeça, patas e outros membros.

Essa detecção de pose é essencial para evitar falsos positivos de bezerros e eliminar problemas de se considerar como bezerro uma vaca que aparece parcialmente no vídeo ou está distante da câmera. No entanto, é importante ressaltar que a detecção de pose dos bezerros só é possível quando eles estão em pé, o que impõe uma limitação nesse aspecto.

Os resultados preliminares obtidos com essa abordagem mostraram a eliminação da maioria dos falsos positivos de bezerros, aumentando significativamente a precisão e confiabilidade da detecção. Assim, a combinação da medição de tamanho e detecção de pose proporcionou uma abordagem mais robusta para distinguir bezerros de vacas em vídeos.

8.2 Métodos

Nesta parte, o primeiro objetivo é detectar o objeto “vaca” nas imagens para depois segmentar entre os objetos selecionados quais deles são bezerros, sendo que para implementar o modelo de detecção de vacas e manipular as imagens e vídeos, utilizou-se a biblioteca *OpenCV*.

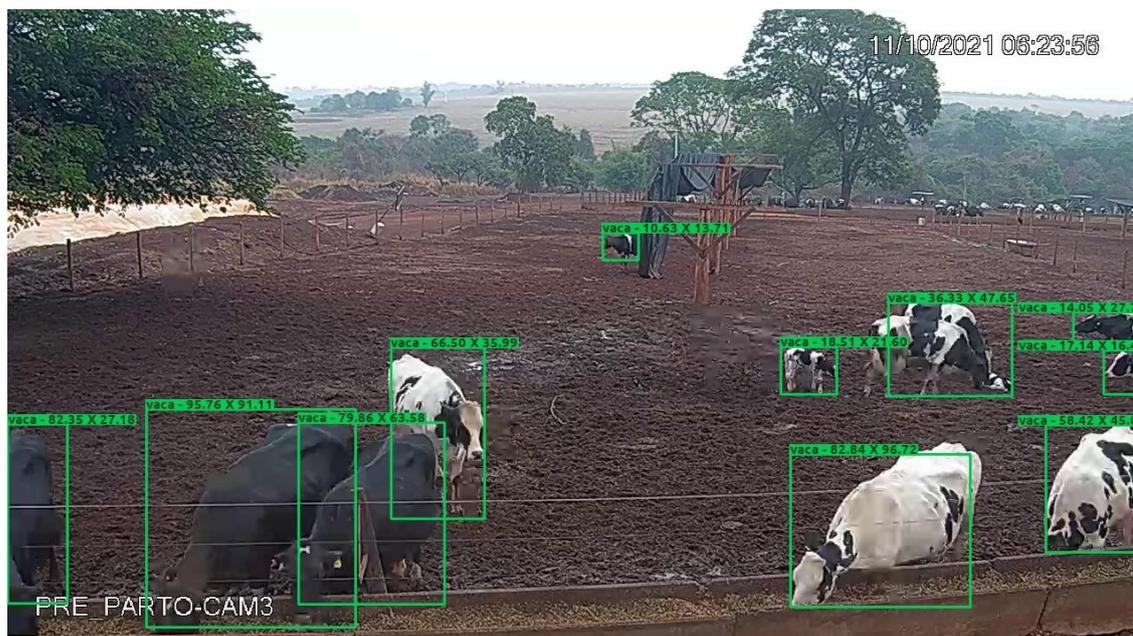
Desenvolvida em *Python*, o *OpenCV* é uma biblioteca amplamente utilizada em todo o mundo, com mais de 2500 algoritmos disponíveis. Perfazendo mais de 2,5 milhões de *downloads*, essa biblioteca oferece uma ampla gama de funcionalidades para processamento de imagens e vídeos, incluindo leitura e escrita de arquivos de mídia, transformações geométricas, detecção de objetos, reconhecimento de padrões, entre outros recursos. Sua facilidade de uso e eficiência tornam o *OpenCV* uma escolha popular para tarefas relacionadas à visão computacional (CuljaK *et al.*, 2012).

O *OpenCV* por meio do módulo *Deep Neural Network* (DNN) é capaz de prover suporte a redes neurais profundas, permitido que se realize a carga de modelos pré-treinados disponíveis na internet e ainda prover acesso à aceleração por *hardware*, como uso de GPUs, que aumenta a velocidade de inferência para dos modelos suportados.

Para detecção de objetos foi realizado a carga no *OpenCV* do modelo *Yolo* na versão 4 (Bochkovski; Wang; Liao, 2020), previamente treinado com o *dataset* Microsoft *COCO* (Lin *et al.*, 2015), que conta com 2.5 milhões de objetos devidamente anotados em 328 mil imagens. No caso, o tipo de objeto de interesse desse trabalho é a vaca (*cow*), que possui mais de 10 mil objetos anotados.

Com um modelo capaz de detectar vacas, como apresentado na Figura 36, o sistema, agora por meio do *OpenCV*, vai verificar imagem por imagem se há a presença de vacas, e ao detectar o objeto em questão, o sistema deve analisar se o objeto se trata de uma vaca ou bezerro, uma vez que o modelo pré-treinado não faz essa distinção.

Figura 36 - Detecção de Vacas



Fonte: O autor.

Uma forma de fazer a distinção entre vaca e bezerro é aferir o tamanho do animal, pois o bezerro é menor que a vaca. No cenário em que temos uma câmera de alta resolução *Pan-Tilt-Zoom* (PTZ) no piquete, sem a capacidade de aferir a profundidade das imagens, o que inviabiliza que o equipamento por si seja capaz de informar o tamanho dos objetos detectados, foi utilizado a técnica de similaridades de triângulos (Swaminathan *et al.*, 2019) para contornar essa deficiência, com intuito de apontar o tamanho relativo dos objetos detectados, no caso “vacas”.

Considerando que a câmara fica fixa no piquete, foi utilizado um objeto de tamanho real conhecido, que foi deslocado em diversas regiões do piquete, como visto na Figura 36. Ao deslocar o objeto pelo piquete durante a gravação de um vídeo que foi utilizado para coletar a relação tamanho real do objeto, tamanho do objeto na imagem e a distância do objeto da câmera, em vários pontos da imagem.

A relação entre tamanho real do objeto e tamanho do objeto na imagem em pixel, em determinado ponto da imagem, foi aplicada em objetos de tamanhos desconhecidos que se encontram em pontos próximos aos já analisados, assim sendo possível aferir o tamanho real aproximado do objeto.

Objetos identificados como vaca são submetidos ao processo de detecção de tamanho real aproximado. Diante dessa informação, é definido um limiar entre tamanho de uma vaca e um bezerro, que pode variar conforme a raça do rebanho analisado,

fornecendo ao sistema a capacidade de reclassificar os objetos classificados como vaca para bezerro, caso o tamanho real aproximado seja inferior ao limiar.

Figura 37 – Análise de objeto de tamanho conhecido



Fonte: O autor.

Sendo assim, na abordagem citada foi capaz de identificar bezerros, pois de acordo com a pose e a posição da vaca pode ocorrer os falsos positivos, conforme visto na Figura 37.

A partir disso, para mitigar a apresentação de falsos positivos, o sistema passou a checar os objetos identificados como bezerros, utilizando a ferramenta de código aberto para estimativa de pose conhecida como MMpose (Jiang *et al.*, 2023).

A análise de pose animal refere-se à estimativa da localização e orientação das diferentes partes do corpo de um animal em uma imagem ou vídeo. Essa informação permite determinar a pose do animal, como checar se ele está deitado ou em pé, e identifica quais partes do bovino estão apresentadas na imagem.

Desta feita, a abordagem utilizada neste estudo foi baseada em modelos 2D, sendo a arquitetura ResNet-50 uma escolha popular nesse contexto. A *ResNet-50* é uma rede neural convolucional profunda composta por 50 camadas, conhecida por sua capacidade de aprender representações complexas e discriminativas das imagens (HE *et al.*, 2015).

Por meio dessa arquitetura, foi possível obter uma detecção precisa das articulações dos animais, contribuindo para a análise e inferência da pose do animal. Essa técnica permite explorar as informações visuais presentes nas imagens e utilizar o poder

das redes neurais convolucionais para realizar uma estimativa confiável das posições das articulações dos animais.

Figura 38 – Detecção com falso positivo



Fonte: O autor.

Por meio da técnica mencionada, foi possível realizar uma checagem consistente da identificação de partes do animal como nariz, cabeça, ombro, joelhos e etc., em cada objeto detectado e identificado como bezerro.

Com intuito de checar se a imagem continha o animal como um todo ou se tratava de partes dele, o que pode ocasionar falsos positivos, como visto na Figura 38, na qual observaram-se duas detecções, sendo uma corretamente identificada como bezerro e outra uma detecção incorreta (falso positivo), pois trata-se de parte de uma vaca adulta, no caso somente a cabeça do bovino.

Diante da informação que se trata apenas de parte do animal, não tendo informações suficientes para aferir sua altura, as detecções desse formato passam a ser descartadas, visando reduzir as detecções incorretas. Assim o sistema só informa a presença de bezerro se ele estiver em pé, o que possibilita dar maior precisão ao processo de inferência da altura do animal, melhorando a qualidade da informação prestada ao produtor.

8.3 Resultados

A proposta de metodologia inovadora para a detecção de bezerros em vídeos, através da combinação da medição de tamanho e detecção de pose, mostrou resultados promissores na superação dos desafios previamente encontrados no campo de detecção animal. Ao introduzir a detecção de pose junto com a medição de tamanho foi possível aumentar a precisão e confiabilidade dos sistemas de detecção de bezerros, permitindo a distinção precisa entre vacas e bezerros, como apresentado na Figura 39.

Durante a fase de teste, registrou-se uma melhoria significativa na precisão da identificação de bezerros em comparação com os métodos existentes. Isso foi observado através da diminuição acentuada de falsos positivos, sendo que a maioria desses erros estava relacionada à confusão entre bezerros e vacas em determinadas posturas ou orientações.

Foi observado que a maioria dos falsos positivos remanescentes estava associada à classificação incorreta de vacas adultas deitadas como bezerros, uma vez que o tamanho estimado do animal se assemelhava ao de um bezerro. Porém, a inclusão da detecção de pose para examinar a orientação e postura do animal resultou na eliminação quase total desses casos.

Por outro lado, a limitação do nosso sistema é que ele pode não identificar corretamente bezerros que estão deitados, pois isso depende de se aferir a altura total do animal quando ele está em pé. Esse é um aspecto que planejamos abordar em futuros aprimoramentos do sistema. Outra limitação é que o método assume que a câmera está fixa, e caso haja qualquer alteração de posição ou de ângulo, o sistema deve ser recalibrado.

É importante destacar que este trabalho forneceu resultados sólidos e contribuições significativas para a área de visão computacional na pecuária. em vista disso, a implementação de sistemas de detecção mais precisos e confiáveis de bezerros, usando um agregado de técnicas de visão computacional, pode trazer benefícios tangíveis em termos de monitoramento de rebanhos, controle de saúde animal e estudos comportamentais.

Figura 39 - Bezerro detectado corretamente



Fonte: O autor.

Com a integração da detecção de pose à medição de tamanho, conseguiu-se melhorar a precisão e a confiabilidade dos sistemas de identificação de bezerros. Esse avanço permitiu uma distinção mais acurada entre vacas e bezerros, como pode ser observado na Figura 39.

Assim, depreende-se que a introdução de uma metodologia inovadora para detectar bezerros em vídeos, combinando a medição de tamanho com a detecção de pose, ofereceu avanços significativos no desafio persistente da detecção animal.

8.4 Conclusão

No contexto específico da detecção de bezerros, a abordagem adotada foi baseada na diferença de tamanho entre bezerros e vacas. Utilizando apenas uma câmera, foram realizadas medições de altura e largura dos animais, permitindo discernir corretamente bezerros e vacas.

Essa abordagem foi complementada pelo uso da detecção de pose, em que uma rede neural treinada foi utilizada para identificar as partes detectadas do animal, como a cabeça, patas e outros membros. Dessa forma, foi possível evitar falsos positivos de bezerros e eliminaram-se problemas de considerar como bezerro uma vaca parcialmente

visível no vídeo ou distante da câmera. É importante ressaltar, que a detecção de pose dos bezerros só é possível quando eles estão em pé, o que impõe uma limitação nesse aspecto.

Os resultados obtidos com essa abordagem demonstraram uma significativa melhoria na detecção de bezerros, aumentando a precisão e confiabilidade do sistema. A combinação da medição de tamanho e detecção de pose proporcionou uma abordagem mais robusta para distinguir bezerros de vacas em vídeos, o que é fundamental em diversas aplicações, como monitoramento de rebanhos e estudos comportamentais.

Por fim, a utilização da arquitetura *ResNet-50* e a integração das técnicas de medição de tamanho e detecção de pose mostraram-se eficazes e promissoras para a detecção e identificação precisa dos bezerros.

9 CONCLUSÃO GERAL

Em conclusão, os projetos desenvolvidos pelo Laboratório COMAIS demonstraram a relevância e o impacto da aplicação de técnicas de inteligência artificial em diferentes áreas do conhecimento. O laboratório desempenhou um papel fundamental no desenvolvimento de soluções tecnológicas baseadas em IA, cumprindo assim as premissas estabelecidas.

Dentre os projetos destacados, o *SmartReview* apresentou-se como uma solução promissora para facilitar o processo de revisão sistemática. Por meio do processamento de linguagem natural, a ferramenta automatizou o agrupamento de artigos científicos, proporcionando uma visão mais abrangente e estruturada do conhecimento existente em determinada área de pesquisa.

Na área da agricultura, o projeto de Detecção e Localização de Ervas Daninhas para o Uso Inteligente de Defensivos demonstrou a aplicação da visão computacional e do aprendizado de máquina na detecção precisa de ervas daninhas, permitindo um uso mais inteligente e eficiente de defensivos agrícolas. Essa abordagem contribui para a redução de custos, a preservação do meio ambiente e o controle mais eficaz das infestações de plantas invasoras.

Outro projeto de destaque foi a Detecção de Comportamento Animal em Vídeos Utilizando Visão Computacional e Aprendizado Supervisionado. Essa iniciativa se valeu de técnicas de visão computacional e aprendizado supervisionado para analisar e identificar comportamentos animais em vídeos, proporcionando *insights* valiosos para o monitoramento de rebanhos, controle de saúde animal e estudos comportamentais.

Além disso, o projeto de Detecção de Bezerros em Vídeos Utilizando Medição de Tamanho e Detecção de Pose trouxe uma abordagem inovadora para distinguir bezerros de vacas em vídeos. Por meio da medição de tamanho e da detecção de pose, foi possível melhorar a precisão e confiabilidade na detecção de bezerros, auxiliando no monitoramento e controle desses animais.

Esses projetos representam apenas uma amostra do trabalho realizado pelo Laboratório COMAIS, evidenciando seu compromisso em promover a inovação e contribuir para o avanço científico e tecnológico. O laboratório se estabeleceu como uma unidade viável e bem-sucedida na modelagem de soluções tecnológicas baseadas em inteligência artificial, impulsionando a colaboração multidisciplinar e gerando conhecimento de qualidade. No intuito de avançar ainda mais, o COMAIS continuará

focando na transformação tecnológica e na geração de conhecimento de forma assertiva e na busca da excelência.

Com base nas discussões desta tese de doutoramento é possível afirmar que a hipótese original foi confirmada, visto que a criação do Laboratório COMAIS trouxe uma série de contribuições para a pesquisa e o desenvolvimento de soluções em inteligência artificial, desempenhando um papel crucial na inovação e na colaboração multidisciplinar. O laboratório não apenas cumpriu suas premissas iniciais, mas também se alinhou com diretrizes nacionais de inteligência artificial e gerou impacto significativo em várias áreas de conhecimento.

Quanto ao Modelo de Sistema Viável (VSM), ele foi aplicado efetivamente na gestão do COMAIS. O VSM possibilitou uma compreensão clara dos sistemas que compõem o laboratório e forneceu uma estrutura robusta para tomadas de decisões estratégicas. O texto menciona explicitamente que os cinco sistemas do VSM estão atuantes e interagindo de maneira integrada no laboratório, o que confirma sua aplicação bem-sucedida. Além disso, ao adotar o VSM, oportunidades de aprimoramento foram identificadas, mostrando uma abordagem proativa de gestão que busca a excelência contínua.

Dessa forma, considerando o papel fundamental do COMAIS no avanço da inteligência artificial e a implementação bem-sucedida do VSM, a hipótese “A criação do COMAIS e a aplicação do Modelo de Sistemas Viáveis (VSM) espera-se um impacto significativo nas pesquisas relacionadas à Inteligência Artificial no Tocantins” pode ser considerada plenamente confirmada.

A contribuição do COMAIS para a pesquisa, o desenvolvimento e a inovação na área de inteligência artificial na UFT são incontestáveis, e a adoção do Modelo de Sistema Viável demonstrou ser uma estratégia acertada, contribuindo para a eficiência, viabilidade e sucesso do laboratório.

REFERÊNCIAS

- ALVES, J. M.; RODRIGUES, W.; FIGUEROA, F. E. V.; SOUZA, F. N. da S. Da caixa preta ao projeto político de desenvolvimento da piscicultura: um diagnóstico a partir do viable system model. CONGRESSO BRASILEIRO DE SISTEMAS; 14^o CONGRESSO BRASILEIRO DE SISTEMAS, 2018. Disponível em: <http://issbrasil.usp.br/ocs/index.php/cbs/14cbs/paper/view/260>. Acesso em: 1 jan. 2018.
- BAHDANAU, D.; CHO, K.; BENGIO, Y. **Neural Machine Translation by Jointly Learning to Align and Translate**arXiv, 19 maio 2016. Disponível em: <http://arxiv.org/abs/1409.0473>. Acesso em: 23 nov. 2022.
- BEER, S. **The heart of enterprise**. Chichester: John Wiley & Sons Ltd, 1979.
- BEER, S. The Viable System Model: Its Provenance, Development, Methodology and Pathology. **The Journal of the Operational Research Society**, v. 35, n. 1, p. 7, jan. 1984.
- CHOLLET, F. **Deep Learning with Python**. 2017.
- BENGIO, Y.; DUCHARME, R.; VINCENT, P.; JAUVIN, C. **A Neural Probabilistic Language Model**. p. 19, 2003.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. 1. ed. Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472: Julie Steele, 2009.
- BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H.-Y. M. **YOLOv4: Optimal Speed and Accuracy of Object Detection**arXiv, 22 abr. 2020. Disponível em: <http://arxiv.org/abs/2004.10934>. Acesso em: 6 jun. 2023.
- BOOTH, A.; HARRIS, J.; CROOT, E.; SPRINGETT, J.; CAMPBELL, F.; WILKINS, E. Towards a methodology for cluster searching to provide conceptual and contextual “richness” for systematic reviews of complex interventions: case study (CLUSTER). **BMC Medical Research Methodology**, v. 13, n. 1, p. 118, 28 set. 2013.
- BRASIL. **Portaria normativa no-7**, de 22 de junho de 2009. Dispõe sobre o mestrado profissional no âmbito da Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES. Disponível em: chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/http://portal.mec.gov.br/dmdocuments/port_mestrado_profissional1.pdf. Acesso em: 3 ago. 2023.
- BRASIL. **Estratégia Brasileira de Inteligência Artificial**.2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf. Acesso em: 17 nov. 2022.
- CANNY, J. A Computational Approach to Edge Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. PAMI-8, n. 6, p. 679–698, nov. 1986.
- CAO, J.; ANWER, R. M.; CHOLAKKAL, H.; KHAN, F. S.; PANG, Y.; SHAO, L. **SipMask: Spatial Information Preservation for Fast Image and Video Instance Segmentation**arXiv, 29 jul. 2020. Disponível em: <http://arxiv.org/abs/2007.14772>. Acesso em: 5 maio. 2023.

- CER, D.; YANG, Y.; KONG, S.; HUA, N.; LIMTIACO, N.; JOHN, R. S.; CONSTANT, N.; GUAJARDO-CESPEDES, M.; YUAN, S.; TAR, C.; SUNG, Y.-H.; STROPE, B.; KURZWEIL, R. Universal Sentence Encoder. **arXiv:1803.11175 [cs]**, 12 abr. 2018. Disponível em: <http://arxiv.org/abs/1803.11175>. Acesso em: 5 jun. 2021.
- COELHO, L.; BRITO, J. N. **Fotogrametria digital**. Rio de Janeiro: EdUERJ, 2007. 196 p. p. Disponível em: <http://www.efoto.eng.uerj.br/aprenda2/tutoriais/item/25-e-book>. Acesso em: 16 dez. 2022.
- CONNEAU, A.; KHANDELWAL, K.; GOYAL, N.; CHAUDHARY, V.; WENZEK, G.; GUZMÁN, F.; GRAVE, E.; OTT, M.; ZETTMLOYER, L.; STOYANOV, V. Unsupervised Cross-lingual Representation Learning at Scale. **arXiv:1911.02116 [cs]**, 7 abr. 2020. Disponível em: <http://arxiv.org/abs/1911.02116>. Acesso em: 5 jun. 2021.
- CONNEAU, A.; KIELA, D.; SCHWENK, H.; BARRAULT, L.; BORDES, A. Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. **arXiv:1705.02364 [cs]**, 8 jul. 2018. Disponível em: <http://arxiv.org/abs/1705.02364>. Acesso em: 5 jun. 2021.
- COOK, D. J.; MULROW, C. D.; HAYNES, R. B. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. **Annals of Internal Medicine**, v. 126, n. 5, p. 376–380, mar. 1997.
- COPPIN, B. **Inteligência Artificial**. Rio de Janeiro: LTC, 2017.
- CULJAK, I.; ABRAM, D.; PRIBANIC, T.; DZAPO, H.; CIFREK, M. **A brief introduction to OpenCV**. Em: 2012 Proceedings of the 35th International Convention MIPRO, 2012, [...]. 2012. p. 1725–1730.
- DAI, A. M.; LE, Q. V. **Semi-supervised Sequence Learning** arXiv, 4 nov. 2015. Disponível em: <http://arxiv.org/abs/1511.01432>. Acesso em: 30 nov. 2022.
- DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. **arXiv:1810.04805 [cs]**, 10 out. 2018. Disponível em: <http://arxiv.org/abs/1810.04805>. Acesso em: 8 jul. 2019.
- BRASIL. Ministério da Ciência, Tecnologia e Inovações. Estratégia Brasileira de Inteligência Artificial - EBIA. 2021. Disponível em: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf . Acesso em: 3 out. 2023.
- EUROPEAN COMMISSION. Skills for jobs: A new strategic approach to vocational education and training. Luxembourg: Publications Office of the European Union, 2018. 152 p.
- FARIA, E. L. de. **Redes neurais convolucionais e máquinas de aprendizado extremo aplicadas ao mercado financeiro brasileiro**. Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, set. 2018.
- FELTRIN, F. **Visão computacional em python**. [s.l.] Uniorg, 2020.
- FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. **Communications of the ACM**, v. 24, n. 6, p. 381–395, 1 jun. 1981.

GIRSHICK, Ross; DONAHUE, J; DARRELL, T; MALIK J. Rich feature hierarchies for accurate object detection and semantic segmentation. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p. 580-587.

GOODFELLOW, Ian J.; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 2016.

GOOGLE. **Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing**. Disponível em: <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>. Acesso em: 30 nov. 2022.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. **Mask R-CNN** arXiv, 24 jan. 2018. Disponível em: <http://arxiv.org/abs/1703.06870>. Acesso em: 8 maio. 2023.

HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition** arXiv, 10 dez. 2015. Disponível em: <http://arxiv.org/abs/1512.03385>. Acesso em: 5 maio. 2023.

HOWARD, J.; RUDER, S. Universal Language Model Fine-tuning for Text Classification. **arXiv:1801.06146 [cs, stat]**, 18 jan. 2018. Disponível em: <http://arxiv.org/abs/1801.06146>. Acesso em: 6 set. 2019.

JAIN, A. K. **Algorithms for clustering data**. [s.l.] Englewood Cliffs, N.J.: Prentice Hall, 1988. xiv, 320 p. p.

JIANG, T.; LU, P.; ZHANG, L.; MA, N.; HAN, R.; LYU, C.; LI, Y.; CHEN, K. **RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose** arXiv, 13 mar. 2023. Disponível em: <http://arxiv.org/abs/2303.07399>. Acesso em: 4 jul. 2023.

JONES, K. S. A statistical interpretation of term specificity and its application in retrieval. **Journal of Documentation**, v. 28, p. 11–21, 1972.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition**. Upper Saddle River, N.J: Pearson Prentice Hall, 2009. 988 p.

KIM, Y. **Convolutional Neural Networks for Sentence Classification** arXiv, 2 set. 2014. Disponível em: <http://arxiv.org/abs/1408.5882>. Acesso em: 22 nov. 2022.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Em: **Advances in Neural Information Processing Systems**, 2012, [...]. Curran Associates, Inc., 2012. v. 25

LANE, H.; HOWARD, C.; HAPKE, H. M. **Natural Language Processing in Action**. 3. ed. [s.l.] Manning Publications Co., 2017.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep Learning. **Nature**, v. 521, p. 436-444, 2015.

LEE, K. F. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos comunicamos e vivemos**. Tradução M. Barbão. 1. ed. Rio de Janeiro: Globo Livros, 2019. 292 p.

LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; BOURDEV, L.; GIRSHICK, R.; HAYS, J.; PERONA, P.; RAMANAN, D.; ZITNICK, C. L.; DOLLÁR, P. **Microsoft COCO: Common Objects in Context** arXiv, 20 fev. 2015. Disponível em: <http://arxiv.org/abs/1405.0312>. Acesso em: 18 dez. 2022.

- LINDE, K.; WILLICH, S. N. How objective are systematic reviews? Differences between reviews on complementary medicine. **Journal of the Royal Society of Medicine**, v. 96, n. 1, p. 17–22, jan. 2003.
- LONG, F.; ZHANG, H.; FENG, D. D. Fundamentals of Content-Based Image Retrieval. Em: FENG, D. D.; SIU, W.-C.; ZHANG, H.-J. **Multimedia Information Retrieval and Management: Technological Fundamentals and Applications**. Signals and Communication Technology. Berlin, Heidelberg: Springer, 2003. p. 1–26.
- LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, nov. 2004.
- LUO, C.; LI, Y.; CHUNG, S. M. Text document clustering based on neighbors. **Data & Knowledge Engineering**, Including Special Section: Conference on Privacy in Statistical Databases (PSD 2008) – Six selected and extended papers on Database Privacy. v. 68, n. 11, p. 1271–1288, 1 nov. 2009.
- MARTINS, J. L.; MOSER, A. **Transformação digital: o futuro no presente da educação**. Portal de Livros da Editora, v. 1, n. 6, p. 159–159, 6 jul. 2021.
- MAATEN, L. van der; HINTON, G. Visualizing data using t-SNE. **Journal of Machine Learning Research**, v. 9, n. Nov, p. 2579–2605, 2008.
- MCTI. **ESTRATÉGIA BRASILEIRA PARA A TRANSFORMAÇÃO DIGITAL**2018. Disponível em: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>. Acesso em: 13 nov. 2022.
- MCTI, I. Portaria Nº 1.122, de 19 de março de 2020 - **DOU** - Imprensa Nacional. 2020. Disponível em: <https://in.gov.br/web/dou>. Acesso em: 17 nov. 2022.
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. Efficient Estimation of Word Representations in Vector Space. **arXiv:1301.3781 [cs]**, 16 jan. 2013. Disponível em: <http://arxiv.org/abs/1301.3781>. Acesso em: 10 abr. 2019.
- MURO, M.; ROTHWELL, J.; ANDES, S.; FIKRI, K.; KULKARNI, S. **America's Advanced Industries: What They Are, Where They Are, and Why They Matter**. Washington, DC: The Brookings Institution, 2015. . Disponível em: https://www.brookings.edu/wp-content/uploads/2015/02/AdvancedIndustry_FinalFeb2lores.pdf. Acesso em: 13 nov. 2022.
- OECD. **Going Digital: Shaping Policies, Improving Lives**. Paris: Organisation for Economic Co-operation and Development, 2019.
- OECD. **A Caminho da Era Digital no Brasil**. [s.l.] OECD, 2020.
- OECD; EUROSTAT. **Oslo Manual 2018**. [s.l.: s.n.]256 p.
- OLAFENWA, A. **Simplifying Object Segmentation with Pixellib Library**. 1 jan. 2021. Disponível em: <https://vixra.org/abs/2101.0122>. Acesso em: 8 maio. 2023.
- OpenDroneMap. **ODM - A command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images** OpenDroneMap, 2020. Disponível em: <https://github.com/OpenDroneMap/ODM>. Acesso em: 16 dez. 2022.
- PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KOPF, A.; YANG, E.; DEVITO, Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.;

- STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *In: Advances in Neural Information Processing Systems 32*. [s.l.] Curran Associates, Inc., 2019. p. 8024–8035.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011.
- PERKINS, J. *Python Text Processing with NLTK 2.0 Cookbook*. 32 Lincoln Road Olton Birmingham, B27 6PA, UK.: Packt Publishing Ltd., 2010.
- REDMON, Joseph *et al.* You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 779-788.
- REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. **You Only Look Once: Unified, Real-Time Object Detection** arXiv, 9 maio 2016. Disponível em: <http://arxiv.org/abs/1506.02640>. Acesso em: 5 maio. 2023.
- REIMERS, N.; GUREVYCH, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Em: **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, 2019, Hong Kong, China. [...]. Hong Kong, China: Association for Computational Linguistics, 2019. p. 3982–3992.
- REN, S.; HE, K.; GIRSHICK, R.; SUN, J. **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks** arXiv, 6 jan. 2016. Disponível em: <http://arxiv.org/abs/1506.01497>. Acesso em: 5 maio. 2023.
- RIZZOLI, A. L.; BOLFE, C.; CRUZ, J.; SANTOS, C. M. dos; CONCEIÇÃO, M. E. da; SCHLINDWEIN, S. L. **Estruturação do metassistema do modelo do sistema viável em uma empresa pública de Santa Catarina, Brasil** Congresso Brasileiro de Sistemas, 2014. Disponível em: <http://www.issbrasil.usp.br/ocs/index.php/cbs/10cbs/paper/view/61>. Acesso em: 5 maio. 2023.
- ROGERS, D. L. **Transformação digital: Repensando o seu negócio para a era digital**. [s.l.] Autêntica Business, 2017. 262 p.
- RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 2010.
- RUSSELL, B. C.; TORRALBA, A.; MURPHY, K. P.; FREEMAN, W. T. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, v. 77, n. 1–3, p. 157–173, 2008.
- SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, v. 24, n. 5, p. 513–523, 1 jan. 1988.
- SALTON, G.; WONG, A.; YANG, C. S. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, v. 18, n. 11, p. 613–620, 1 nov. 1975.
- SAUNDARIYA, K.; ABIRAMI, M.; SENTHIL, K. R.; PRABAKARAN, D.; SRIMATHI, B.; NAGARAJAN, G. Webapp Service for Booking Handyman Using MongoDB, Express JS, React JS, Node JS, 2021 3RD INTERNATIONAL

CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION (ICPSC), 2021, [...]. 2021. p. 180–183.

SCHNABEL, T.; LABUTOV, I.; MIMNO, D.; JOACHIMS, T. Evaluation methods for unsupervised word embeddings. **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, p. 298–307, 2015.

SCHWANINGER, M.; SCHEEF, C. A Test of the Viable System Model: Theoretical Claim vs. Empirical Evidence. **Cybernetics and Systems**, v. 47, n. 7, p. 544–569, 2 out. 2016.

SILVA NETO, L. S.; DIAS, F. C. F.; OSÓRIO, N. B.; ROLIM, C. L. A. eHealth-Based Interventions for Older Patients with Prostate Cancer: A Quick Review of the Literature. **Telemedicine Reports**, v. 3, n. 1, p. 79–92, nov. 2022.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition** arXiv, 10 abr. 2015. Disponível em: <http://arxiv.org/abs/1409.1556>. Acesso em: 5 maio. 2023.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. 2018.

SUTSKEVER, I.; VINYALS, O.; LE, Q. V. **Sequence to Sequence Learning with Neural Networks** arXiv, 14 dez. 2014. Disponível em: <http://arxiv.org/abs/1409.3215>. Acesso em: 22 nov. 2022.

SZEGEDY, Christian *et al.* Going deeper with convolutions. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p. 1-9.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. **Rethinking the Inception Architecture for Computer Vision** arXiv, 11 dez. 2015. Disponível em: <http://arxiv.org/abs/1512.00567>. Acesso em: 5 maio. 2023.

SWAMINATHAN, V.; ARORA, S.; BANSAL, R.; RAJALAKSHMI, R. Autonomous Driving System with Road Sign Recognition using Convolutional Neural Networks, 2019 INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE IN DATA SCIENCE (ICCIDS), 2019, [...]. 2019. p. 1–4.

SZYMAŃSKI, P.; KAJDANOWICZ, T. scikit-multilearn: A Python library for Multi-Label Classification. **Journal of Machine Learning Research**, v. 20, n. 6, p. 1–22, 2019.

TRICCO, A. C.; TETZLAFF, J.; MOHER, D. The Art and Science of Knowledge Synthesis. **Journal of Clinical Epidemiology**, v. 64, n. 1, p. 11–20, jan. 2011.

TURNEY, P. D.; PANTEL, P. From Frequency to Meaning: Vector Space Models of Semantics. **Journal of Artificial Intelligence Research**, v. 37, p. 141–188, 27 fev. 2010.

UFT, U. F. do T. **Plano de Desenvolvimento Institucional 2021 – 2025** 2021. Disponível em: https://docs.uft.edu.br/share/s/16G29vJbQ1ikIp_eqtOvgw. Acesso em: 12 nov. 2022.

UNIVERSIDADE FEDERAL DO TOCANTINS. **Resolução n° 03**, de 27 de Março de 2019. Disponível em: <https://docs.uft.edu.br/share/s/DfEtPqoKTGSLnjXjiD5D4A>. Acesso em: 2 maio. 2023.

UNIVERSIDADE FEDERAL DO TOCANTINS - UFT. Mestrado e Doutorado em Modelagem Computacional de Sistemas lança edital de seleção. Por Virgínia Magrin |

Publicado: Quarta, 17 de Julho de 2019, 10h54 | Última atualização em Quinta, 18 de Julho de 2019, 08h56. Disponível em: <https://ww2.uft.edu.br/index.php/ultimas-noticias/25825-mestrado-em-modelagem-computacional-de-sistemas-lanca-edital-de-selecao>. Acesso em: 3 ago. 2023.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is All you Need. *Em*: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. **Advances in Neural Information Processing Systems 30**. [s.l.] Curran Associates, Inc., 2017. p. 5998–6008.

VIG, J. **A Multiscale Visualization of Attention in the Transformer Model** arXiv, 12 jun. 2019. Disponível em: <http://arxiv.org/abs/1906.05714>. Acesso em: 1 dez. 2022.