



Programa de Pós-Graduação em

**Computação Aplicada**

Mestrado/Doutorado Acadêmico

Carlos Eduardo de Moura Apoitia

Um modelo de rede neural para estimar a produtividade de milho a partir de imagens de satélite

São Leopoldo, 2022

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA  
NÍVEL MESTRADO

CARLOS EDUARDO DE MOURA APOITIA

**UM MODELO DE REDE NEURAL PARA ESTIMAR A PRODUTIVIDADE DE  
MILHO A PARTIR DE IMAGENS DE SATÉLITE**

São Leopoldo  
2022

A643m Apoitia, Carlos Eduardo de Moura.  
Um modelo de rede neural para estimar a produtividade de milho a partir de imagens de satélite / Carlos Eduardo de Moura Apoitia – 2022.  
91 f. : il. color. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2022.

“Orientador: Prof. Dr. Gabriel de Oliveira Ramos.”

1. Agricultura. 2. Aprendizado de máquina. 3. Desmatamento. 4. Inteligência artificial. 5. Redes neurais (Computação). I. Título.

CDU 004.8

## RESUMO

A moratória da soja tem sido um importante instrumento que visa reduzir o desmatamento no bioma amazônico. No entanto, a existência do desmatamento ilegal das florestas crescer proporcionalmente com os números de terras embargadas. Visto que na moratória existem algumas brechas que vêm sendo exploradas por produtores ligados ao desmatamento e práticas como a *lavagem de grãos*, o que dificulta aos analistas e órgãos controle na identificação da produtividade real de um imóvel. Este trabalho propõe um modelo de redes neurais artificiais, denominado *Deep Yield Prediction* (DYP), capaz de decodificar as informações aprendidas usando redes neurais convolucionais (CNN) em previsões de produtividade de safras de milho, de forma que auxilie na identificação de quanto um imóvel produzirá. Colaborando assim, para a quebra do ciclo de *lavagem de grãos* e desmatamento ilegal no bioma Amazônico.

**Palavras-chaves:** Agricultura Inteligente. Aprendizado de Máquina. Desmatamento. Inteligência Artificial. Estimativa de Produtividade de safras. Redes Neurais Convolucionais.

## **ABSTRACT**

The soy moratorium has been an important instrument aimed at reducing deforestation in the Amazon biome. However, the existence of illegal deforestation of forests grows proportionally with the numbers of embargoed lands. Since in the moratorium there are some loopholes that have been exploited by producers linked to deforestation and practices such as grain laundering, which makes it difficult for analysts and control agencies to identify the real productivity of a property. This work proposes a model of artificial neural networks, called Deep Yield Prediction (DYP), capable of decoding the information learned using convolutional neural networks (CNN) in corn crop yield predictions, in a way that helps in the identification of how much a property will produce. Therefore collaborating to break the cycle of grain laundering and illegal deforestation in the Amazon biome.

**Keywords:** Artificial Intelligence. Convolutional Neural NetWork (CNN). Crop Yield Prediction. Deforestation. Machine Learning. Smart Agriculture.

## LISTA DE FIGURAS

Figura 1 – Cenário I - comercialização regular . . . . .	17
Figura 2 – Cenário II - comercialização irregular . . . . .	18
Figura 3 – Organização do trabalho em tópicos . . . . .	20
Figura 4 – Região com presença de imóveis embargos . . . . .	26
Figura 5 – Cálculo do método 1 . . . . .	31
Figura 6 – Cálculo NDVI . . . . .	33
Figura 7 – Contextualização da hierarquia da área de estudo . . . . .	35
Figura 8 – Arquitetura de uma CNN . . . . .	39
Figura 9 – Modelo de um neurônio artificial "Perceptron . . . . .	39
Figura 10 – Função ReLu . . . . .	41
Figura 11 – Resultado das buscas . . . . .	46
Figura 12 – Resultado após a etapa de cortes . . . . .	47
Figura 13 – Modelo de estimativa de produtividade de uma unidade produtora . . . . .	65
Figura 14 – Relatório integrado a colheitadeira . . . . .	66
Figura 15 – Método criado para calcular a média de produtividade . . . . .	68
Figura 16 – <i>Merging</i> das imagens do relatório com a imagem pré-colheita . . . . .	68
Figura 17 – Diferenças de características do RGB em um quadrante . . . . .	69
Figura 18 – Relação de algumas das imagens do conjunto de dados . . . . .	70
Figura 19 – Rótulos e valores do conjunto de dados . . . . .	71
Figura 20 – Representação da CNN . . . . .	71
Figura 21 – Utilização de camadas do Keras . . . . .	73
Figura 22 – Divisão dos dados do conjunto de dados . . . . .	74
Figura 23 – Função de regressão . . . . .	74
Figura 24 – Imagem pré-colheita <i>Sentinel-2</i> e Imagem extraída do relatório . . . . .	78
Figura 25 – Recorte da imagem 10x10 px, coluna 2, grid 4, coluna 1, linha 1, produtividade 154 . . . . .	79
Figura 26 – Gráfico execução da métrica - <i>Mean absolute percentage error</i> . . . . .	81
Figura 27 – Gráfico execução da métrica - <i>Mean absolute error</i> . . . . .	81
Figura 28 – Gráfico execução da métrica - <i>Mean squared error</i> . . . . .	82
Figura 29 – Resultados do modelo de estimativa DYP . . . . .	82

## LISTA DE TABELAS

Tabela 1 – Desmatamento em $km^2$ . . . . .	23
Tabela 2 – Contagem por comprimento . . . . .	31
Tabela 3 – Pontuação utilizada para avaliar a qualidade dos artigos . . . . .	48
Tabela 4 – Trabalhos mais relevantes . . . . .	49
Tabela 5 – Tecnologias empregadas nos trabalhos . . . . .	61
Tabela 6 – Relação de cores com a produtividade . . . . .	67

## LISTA DE SIGLAS

ALB	Amazônia Legal Brasileira,
CNN	Convolution Neural Network,
DYP	Deep Yield Prediction,
IMEA	Instituto Mato-Grossense de Economia Agropecuária,
INPE	Instituto Nacional de Pesquisas Espaciais,
IOT	Internet of Things,
MLP	Multilayer Perceptron,
NDVI	Normalized Difference Vegetation Index,
PRODES	Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite,



## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>15</b>
1.1 <b>Motivação</b>	16
1.2 <b>Questão de pesquisa e Hipóteses</b>	18
1.3 <b>Objetivos e contribuições</b>	19
1.4 <b>Organização do texto</b>	19
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>21</b>
2.1 <b>Desmatamento: suas ameaças e seus impactos</b>	21
2.1.1 Projeto de Monitoramento do Desmatamento na Amazônia Legal - PRODES	23
2.2 <b>A Moratória da soja</b>	24
2.2.1 O grão pirata	27
2.2.2 Metodologia para compra de grãos	27
2.3 <b>Agricultura de precisão</b>	28
2.3.1 A estimativa de safras	29
2.3.2 O índice de vegetação com diferença normalizada (NDVI)	32
2.4 <b>A inteligência artificial no agro</b>	33
2.4.1 O aprendizado de máquina	34
2.4.2 Redes neurais convolucionais	38
2.5 <b>Discussão</b>	42
<b>3 REVISÃO DA LITERATURA</b>	<b>43</b>
3.1 <b>Metodologia</b>	43
3.1.1 Questionamentos a serem verificados	43
3.1.2 Processo de pesquisa	44
3.1.3 Critérios e procedimentos de seleção dos artigos	45
3.1.4 Critérios de exclusão	45
3.1.5 Avaliação da qualidade	45
3.2 <b>Resultados</b>	46
3.2.1 Exclusão de trabalhos da pesquisa inicial	47
3.2.2 Avaliação da qualidade dos artigos relevantes	48
3.3 <b>Resultados das questões e discussão</b>	49
3.3.1 Quais tecnologias computacionais foram usadas para previsão de produção?	50
3.3.2 Como a inteligência artificial tem sido usada para a e previsão de produção?	53
3.3.3 Como o aprendizado de máquina e a visão computacional têm sido usados para a previsão de rendimento?	55
3.4 <b>Conclusão</b>	59
<b>4 MODELO PARA ESTIMATIVA DE SAFRAS</b>	<b>63</b>
4.1 <b>Definições do projeto</b>	63
4.2 <b>Decisões de projeto</b>	64
4.3 <b>Arquitetura</b>	65
4.3.1 Conjunto de dados - <i>Dataset</i>	66
4.3.2 Pré-Processamento da imagem	70
4.3.3 Convolucional neural network - CNN	71
4.4 <b>Discussão</b>	74

<b>5</b>	<b>RESULTADOS</b>	<b>77</b>
<b>5.1</b>	<b>Metodologia dos experimentos</b>	<b>77</b>
<b>5.2</b>	<b>Resultados numéricos</b>	<b>80</b>
<b>5.3</b>	<b>Discussão</b>	<b>81</b>
<b>6</b>	<b>CONCLUSÃO</b>	<b>85</b>
<b>6.1</b>	<b>Limitações</b>	<b>86</b>
<b>6.2</b>	<b>Trabalhos futuros</b>	<b>86</b>
	<b>REFERÊNCIAS</b>	<b>87</b>

## 1 INTRODUÇÃO

O crescimento populacional no mundo traz enormes desafios, como aumentar a produção de alimentos sem aumentar as áreas plantadas e ao mesmo tempo garantir a preservação do meio ambiente melhorando a qualidade de vida (N GANDHI, 2016). A produtividade em grande escala é fundamental para o mundo e para muitos países é o principal recurso econômico. Mas é dever enquanto seres humanos tentarmos encontrar ferramentas que possam auxiliar a produzir alimentos de maneira sustentável.

O impacto causado ao meio ambiente estimulado pela conversão do uso e cobertura da terra para atender às crescentes demandas de consumo da humanidade, vem provocando diversas modificações nos ciclos naturais do sistema terrestre. De acordo com FOLEY et al. (2005) as consequências dessas modificações para a manutenção da vida no planeta atribuiu um novo enfoque aos estudos das mudanças de uso e cobertura da terra. Já em GALFORD et al. (2010) é revelado que a conversão de florestas para o uso agropecuário tornou-se um dos principais temas no debate atual das mudanças globais.

No Brasil, o Instituto Nacional de Pesquisas Espaciais (INPE) é o órgão responsável pelo monitoramento e mapeamento do desmatamento na Amazônia legal brasileira desde o ano de 1988. Através desse mapeamento constante é possível acompanhar um crescimento de áreas desmatadas nos últimos anos e conforme BROWN; KOEPPE (2014) as elevadas taxas de desmatamento no bioma Amazônico e a sua relação com a produção de soja foram evidenciadas na campanha “Eating up the Amazon” do Greenpeace em 2006. Através deste movimento, organizações da sociedade civil em conjunto com empresas ligadas ao comércio da *commodity* de soja e demais grãos, criaram um movimento em prol do fim do desmatamento no bioma da Amazônia do Brasil, uma vez que é péssimo para imagem de uma empresa multinacional estar relacionada ao desmatamento da Amazônia.

Diante disso, em julho de 2006 a Associação Brasileira das Indústrias de Óleos Vegetais (ABIOVE) e a Associação Nacional dos Exportadores de Cereais (ANEC), em conjunto com órgãos públicos e da sociedade civil, anunciaram a Moratória da Soja, que inicialmente teria a duração de dois anos. Segundo BRANDAO (2020), ficou acordado que as empresas associadas à ABIOVE e ANEC não comprariam soja e demais grãos produzidos a partir de áreas desmatadas do bioma Amazônico. Entretanto, no decorrer dos anos o acordo já foi renovado por três vezes e desde julho de 2009 vem utilizando imagens de satélites para o sensoriamento remoto visando auxiliar no monitoramento das áreas desflorestadas para identificação de possíveis cultivos de grãos (MELLO et al., 2011). Estima-se que o acordo tenha evitado o desmatamento de cerca de 1,8 milhão de hectares em 10 anos, extensão equivalente ao território de Israel, conforme afirmou CAMPOS ANDRÉ (2021).

No entanto, percebeu-se no decorrer deste estudo a existência de diversas brechas na checagem da moratória, permitindo assim que as empresas continuem comprando grãos de produtores ligados ao desmatamento ilegal no bioma amazônico. Sendo assim, este estudo visa contribuir

com os órgãos controladores e empresas que comercializam grãos para que estes possam identificar possíveis fraudes na origem dos grãos. A tecnologia criada a partir do presente estudo, denominada *Deep Yield Prediction* (DYP), conta com recursos inteligentes capazes de fazer a estimativa da produtividade de safra, possibilitando identificar um imóvel com suspeitas por conta de seu excesso de produção.

Para validação do modelo que visa estimar a produtividade decidiu-se pela cultura de milho, uma vez que esta ocorre no período de estiagem, uma vez que, há pouca incidência de nuvens, facilitando a aquisição de imagens de satélite para o conjunto de dados. Também, fora desenvolvido um método para extrair a média de produção por hectare a partir do relatório gerado pelo software, *Climate-Fieldview*<sup>1</sup>, por meio das colheitadeiras, sendo uma fonte importante de dados devido a sua precisão.

Em seguida, realizou-se o processo de recorte da imagem do imóvel antes da colheita, por meio de imagens do satélite Sentinel-2, para então fazer o *merging* da imagem extraída pré-colheita com o relatório de média de produtividade. Disponibilizando assim, as imagens necessárias para o conjunto de dados e elaboração dos testes e treinamento da rede neural convolucional, que por sua vez é considerada por grande parte da literatura com a mais indicada para reconhecer padrões em imagens. Por fim, será feita a regressão para estimar a média de produtividade e exibição dos resultados.

## 1.1 Motivação

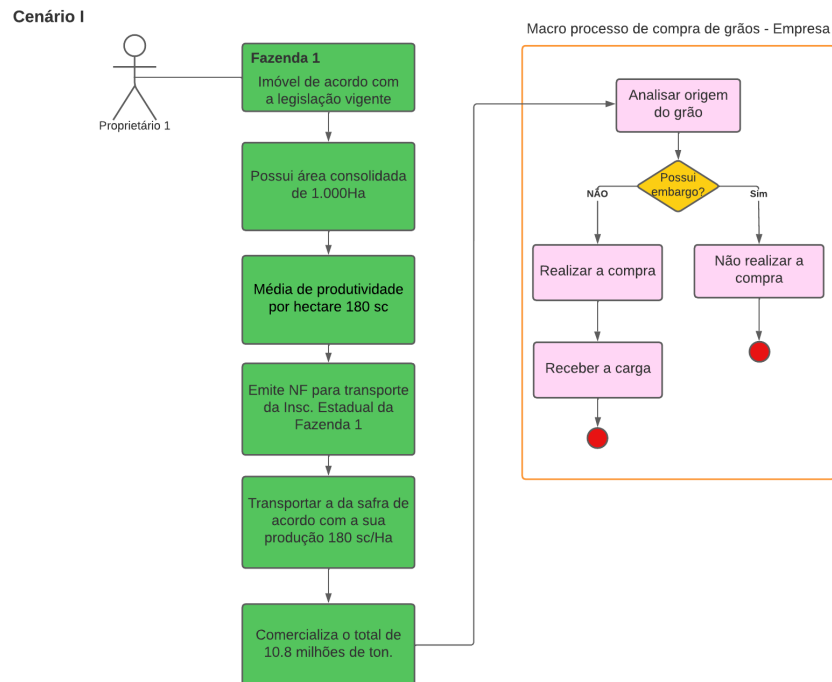
Os índices oficiais apresentados pelo INPE (2021), apontam para uma diminuição de áreas de florestas do bioma Amazônico. Somente no ano de 2021, obteve-se o valor consolidado da área desmatada por corte raso entre o período de 1 de agosto de 2020 e 31 de julho de 2021 é de 13.038 km<sup>2</sup>. Este valor representou um aumento de quase 17% em relação à taxa de desmatamento apurada pelo PRODES em 2020 que foi de 10.851 km<sup>2</sup> para os nove estados da Amazônia Legal Brasileira (ALB). Portanto, o desmatamento desenfreado deste bioma e sua perda de espaço para a agricultura fica cada vez mais evidente.

Pensando nisso, verificou-se a possibilidade de, através de soluções baseadas em análise de imagens e por meio de recursos de aprendizado de máquina, a prever a produção de uma propriedade e alertar para uma *possível* lavagem de grãos. Desta forma, no que tange à produção, é possível identificar dois cenários. No primeiro, temos o agricultor honesto que busca estar regular perante a legislação, conforme a Figura 1. No segundo cenário, temos o produtor que age de forma irregular, buscando comercializar grãos oriundos de áreas de desmate ilegal através de um processo de lavagem de grãos utilizando imóveis sem restrições. Conforme está apresentado na Figura 2.

---

<sup>1</sup>Software Climate Fieldview - <https://www.climatefieldview.com.br/>

Figura 1 – Cenário I - comercialização regular



Fonte: Elaborado pelo autor

Desta forma, este trabalho parte da premissa de que a lavagem de grãos pode ser detectada ao analisar a produtividade de cada propriedade. Em particular, utilizando-se imagens de satélite, é possível fornecer uma estimativa da produtividade real de cada propriedade. Assim, comparando a produtividade estimada de uma propriedade com o montante efetivamente vendido pela mesma, é plausível encontrar evidências da lavagem de grãos.

Haja vista a necessidade de explicar na prática como é possível checar um imóvel, torna-se necessário a criação de alguns indicadores para facilitar a demonstração dessa prática. Uma vez que o custo de uma produção de milho varia de acordo com nível tecnológico adotado para o cultivo, pois a quantidade aplicada de insumos irá interferir nos custos. Além disso, os preços apresentam variação de acordo com a localização no país.

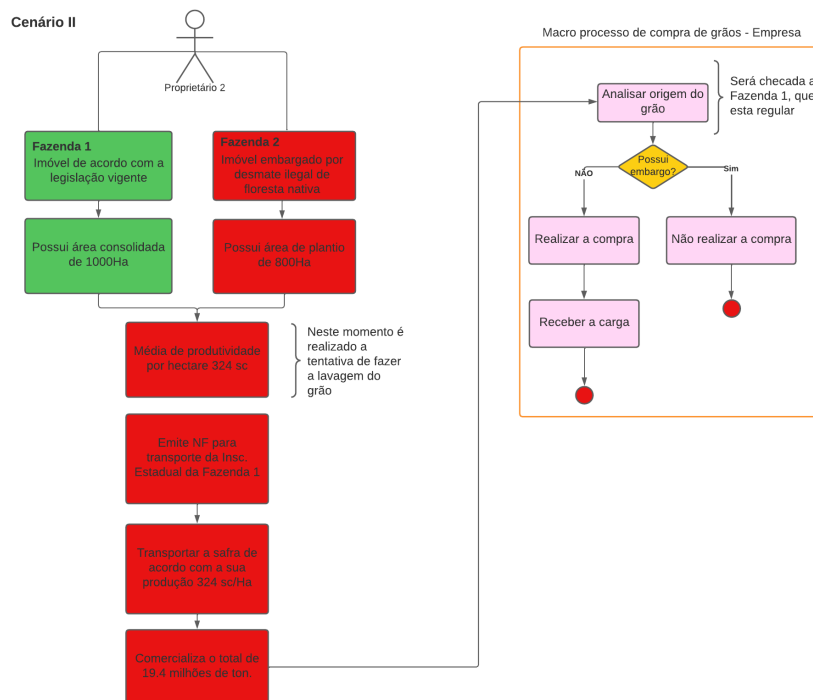
Conforme foi disponibilizado pela Empresa de Pesquisa Agropecuária e Extensão Rural EPAGRI/CEPA (2021) em Santa Catarina, mostram que uma lavoura de milho de média tecnologia, com investimento realizado para uma expectativa de rendimento de 140 sacas/ha, possui um custo estimado R\$ 5.708,67. Para nível alto de tecnologia e expectativa de rendimento de 180 sacas/ha, o custo total é estimado em R\$ 8.210,31.

Mas além da interferência do nível tecnológico adotado na lavoura, a produtividade de milho é influenciada principalmente pelo genótipo utilizado, doses de fertilizantes, época de semeadura, arranjo populacional de plantas e manejo fitossanitário SANGOI et al. (2010). Os insumos, que compreendem sementes, fertilizantes e agrotóxicos, são responsáveis pela maior parte dos custos de produção, cerca de 53 e 60% de todo o investimento nos níveis médio e alto

de tecnologia, respectivamente.

Sendo assim, apresentamos o cenário como base: a média de produção nos imóveis rurais do município de Sorriso/MT. Onde hoje a média de produção está em 180 sacas de milho por hectare, embora ainda não tenhamos disponível os dados oficiais da presente safra, que são apresentados pelo Instituto Mato-Grossense de Economia Agropecuária (IMEA). O máximo de variação que poderia acontecer de super produção, seria de aproximadamente 20%, alcançando a média de produção por hectare em 220 sacas.

Figura 2 – Cenário II - comercialização irregular



Fonte: Elaborado pelo autor

Embora seja um tema de grande interesse, não é sentida uma vontade por parte do poder público em resolver de fato esse problema. Diante disso, criou-se um fator motivacional muito grande na busca de evidências dessas operações.

## 1.2 Questão de pesquisa e Hipóteses

Motivado pelo contexto supracitado, este trabalho busca responder a seguinte questão: como desenvolver uma arquitetura de rede neural convolucional capaz de, a partir de imagens de satélite, estimar a produtividade de uma determinada propriedade rural que pratica a cultura do milho, viabilizando assim a validação da produtividade estimada com a efetiva reportada por um produtor?

A partir desse questionamento, a seguinte hipótese foi levantada: o uso de imagens de satélites em um modelo de estimativa de produtividade de safras, que aprende por meio de redes

neurais convolucionais é capaz de identificar imóveis rurais que estão comercializando grãos além da sua capacidade de produção.

### 1.3 Objetivos e contribuições

O objetivo geral deste estudo é apresentar um pipeline de aprendizado de máquina que seja capaz de fazer a estimativa da safra de milho, tendo um desempenho convergente para a identificar a média de produção por hectares de um imóvel.

Para que o objetivo geral seja atingido, este trabalho se concentrará nos seguintes objetivos específicos: (i) criar uma solução que utilizará imagens de satélite *Sentinel-2*; (ii) utilizar redes neurais convolucionais para fazer a estimativa da produção de milho de um imóvel rural; (iii) investigar a literatura relacionada à estimativa de safras que faz uso de paradigmas da inteligência artificial; (iv) avaliar o modelo de aprendizado de máquina proposto, por meio de um conjunto de dados de produtividade extraído de um relatório de produção de uma colheitadeira.

As principais contribuições deste trabalho podem ser resumidas da seguinte forma:

- A identificação de possibilidades e limitações do uso dos acervos de imagens de satélite com baixa resolução disponibilizadas gratuitamente, no caso do presente estudo o uso do satélite *Sentinel-2*, podendo assim colaborar para a criação de um conjunto de dados poderoso;
- Criação de um modelo de aprendizado de máquina capaz de fazer a estimativa da produtividade de milho a partir de imagens do satélite *Sentinel-2*;
- Apresentação de uma revisão da literatura com suas respectivas contribuições e uso da inteligência artificial na estimativa de safras.

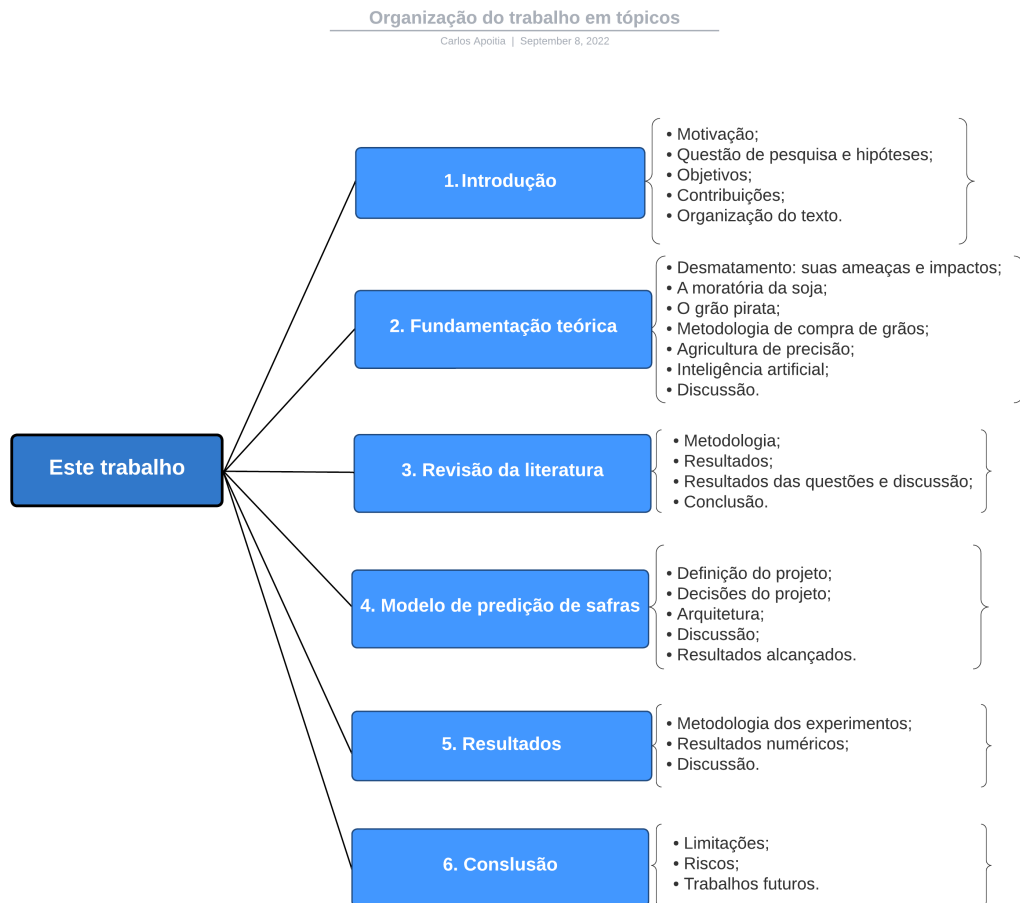
### 1.4 Organização do texto

Para uma melhor compreensão e separação dos conteúdos expostos neste trabalho, a Figura 3 exibirá a divisão dos capítulos e suas principais sessões, além da introdução do trabalho no Capítulo 1, tem-se:

- Capítulo 2: apresenta a **fundamentação teórica**, com uma visão geral dos problemas envolvendo o desmatamento, a moratória da soja, o grão pirata, metodologia de compra de grãos, agricultura de precisão e inteligência artificial.
- Capítulo 3: descreve uma **revisão da literatura** e de temas relacionados à estimativa de safras. Além de destacar as abordagens usadas para a estimativa.
- Capítulo 4: visão geral do **modelo de estimativa de safras**, definição e modelagem do algoritmo de aprendizado de máquina.

- Capítulo 5: apresentação do **experimento**, do conjunto de dados e resultados.
- Capítulo 6: apresenta as principais contribuições e **conclusões** obtidas, além dos trabalhos futuros.

Figura 3 – Organização do trabalho em tópicos



Fonte: Elaborado pelo autor



## 2 FUNDAMENTAÇÃO TEÓRICA

Conforme afirmou N GANDHI (2016), uma vez que o crescimento populacional no mundo traz enormes desafios como: o de aumentar a produção de alimentos sem aumentar as áreas plantadas e ao mesmo tempo garantir a preservação do meio ambiente. No decorrer desse capítulo, buscar-se-á apresentar uma melhor compreensão dos problemas abordados tais como os grandes desafios de aliar conservação ambiental e desenvolvimento, nas pautas das políticas públicas e na esfera das organizações privadas.

A seção 2.1 tem o objetivo de contextualizar as ameaças e os impactos gerados pelo desmatamento. Na seção 2.2 são apresentadas as iniciativas que buscam frear o desmatamento no bioma amazônico. Enquanto na seção 2.3 versa-se sobre os avanços na agricultura de precisão. A inteligência artificial e seus paradigmas são apresentados na seção 2.4. Por fim, uma breve discussão sobre os conceitos são discorridos na seção 2.5.

### 2.1 Desmatamento: suas ameaças e seus impactos

Tradicionalmente, as escolhas políticas e econômicas influenciaram diretamente as ações que refletiram em mudanças na cobertura e uso da terra na Amazônia. E de acordo com escritos de (ABRAMOVAY, 2010), nas últimas décadas, decisões na esfera político-econômica proporcionaram uma crescente expansão agropecuária, na qual atingiu diretamente a preservação das florestas. O Desmatamento no Brasil foi, e ainda é, decorrente de tendências de mercado, como: a exploração de madeiras, projetos de colonização em regiões ainda florestais, processos demográficos e migratórios, grilagem de terra, obras de infraestrutura e a expansão da fronteira agropecuária (ALENCAR et al., 2004).

As principais variações nas taxas de desmatamento associadas aos fatores macroeconômicos, como disponibilidade financeira e a taxa de inflação, foram indicadores importantes para entender as tendências nos desmatamentos no Brasil, e assim caracterizar este tipo de desmatamento, em geral, como fruto da exploração da terra vinculada a investimentos em propriedades rurais de grande porte (FEARNSIDE, 2005).

No entanto, o crescimento da agricultura sobre as áreas de florestas tropicais, não tem sido um fenômeno exclusivo do Brasil, persiste também em outras regiões do mundo, por diversos fatores como o aumento populacional, a demanda por maiores quantidades de alimentos, tanto de proteína na base nutricional quanto fibras e a demanda por biocombustíveis (SAYER et al., 2013). O setor público e privado vem sendo responsáveis por facilitar os investimentos, por meio do fluxo de capital internacional, atendendo esse aumento na demanda global de alimentos, através de um modelo de agricultura de larga escala, altamente técnica e produtora de *commodities* (IORIS, 2015).

Evidenciando que a soja e a carne vem sendo as *commodities* que se destacaram na expansão da fronteira agrícola no Brasil. No início dos anos 2000 o avanço da soja foi uma ameaça sig-

nificativa à manutenção da floresta Amazônia, e se intensificou com o acentuado investimento governamental em infraestrutura como rodovias, ferrovias, hidrovias e portos. Nesse contexto, o processo de desenvolvimento e a ampliação das obras de infraestrutura facilitaram o surgimento de uma cadeia de investimentos indiretos e de especulações que geraram, em algumas situações, desmatamentos maiores do que a própria conversão da floresta na produção agrícola ou pecuária (FEARNSIDE, 2005).

A expansão da soja na Amazônia se concentrou em áreas de topografia plana, com solos e climas favoráveis, com infraestrutura de transporte e terras baratas, por meio desses fatores ligados a não internalização de custos sociais e ambientais no processo produtivo impulsionaram este fenômeno (CIVIL et al., 2004). Onde em 2004, o Brasil tornou-se o maior exportador de carne bovina, onde 38% da exportação era destinada para União Europeia, 12% para o Oriente Médio e 10% para Rússia (NEPSTAD; STICKLER; ALMEIDA, 2006). Esse lugar de destaque no mercado de carne foi acompanhado do crescimento significativo do rebanho, com início na década de 1990 e que se estendeu até 2006, e alcançou uma taxa média de crescimento de 6,74% ao ano na Amazônia legal, o que significou um aumento médio do rebanho, maior do que em outras regiões do Brasil.

De acordo com o IBGE, em 2012, o Estado do Mato Grosso respondia por 32% da produção de soja e 12% da produção de carne bovina no Brasil, e continua sendo em 2022 dez anos após tal dado, o estado com maior rebanho bovino, consolidando um modelo de agricultura exportadora de grãos e carne, que contribuiu com 35% do produto interno bruto brasileiro naquele período, sendo o Estado de destaque agrícola no país.

Conforme aborda BRANDAO (2020) O crescimento do rebanho bovino e da área plantada de soja no Mato Grosso, aumentaram. O rebanho obteve um aumento de 46% em relação aos anos 2000 a 2005, passando de 18,9 milhões para 26,6 milhões de cabeças de gado. Este crescimento foi acompanhado pelo incremento de 3,1 milhões de hectares de pastagem, um crescimento de 19% em área de pasto, enquanto a área destinada a agricultura teve um aumento de 111%, em valores absolutos foram 3,2 milhões de hectares. De forma comparativa, no mesmo período, as taxas anuais de desmatamento no Mato Grosso somaram 5,1 milhões de hectares. Confirmando a expansão da agricultura e pecuária sob a floresta no Estado do Mato Grosso.

Pelo ponto de vista ambiental, esse processo desenfreado na redução de extensas áreas dos ecossistemas naturais, tem causado a degradação e perda dos solos e impactou os ciclos hidrológicos e climáticos. Conforme BALLESTER et al. (2003), esses processos de desmatamentos estão contribuindo para as mudanças climáticas, visto que pela ótica global o desmatamento impacta diretamente o clima, pois é um importante emissor de gases de efeito estufa.

Deste modo, ao entender que as mudanças ambientais são fruto das relações humano-ambiente-sociedade, e dos processos políticos em curso (ROBBINS, 2011). O presente estudo cujo o qual, gerará um modelo de predição de produtividade de milho, é de suma relevância. Uma vez que, imóveis embargados por conta de desmate ilegal ficarão impedidos de continuar

vendendo seus grãos. Já que o modelo de predição irá detectar qual é a capacidade de produção da unidade produtora (imóvel rural), impedindo que o responsável por este imóvel faça a "lavagem" dos grãos que tenham origem em terras embargadas por conta do desmate ilegal.

### 2.1.1 Projeto de Monitoramento do Desmatamento na Amazônia Legal - PRODES

O projeto PRODES faz o monitoramento do desmatamento por corte raso na Amazônia legal por meio de satélites e produz, desde 1988, as taxas anuais de desmatamento na região. Com o objetivo de serem utilizadas pelo governo brasileiro para estabelecimento de políticas públicas.

O PRODES faz uso das imagens dos satélites da classe LANDSAT, CBERS 4 e IRS-2.

Os dados espaciais do PRODES estão sendo utilizados em: (a) Certificação de cadeias produtivas do agronegócio como a Moratória da Soja e o Termo de Ajustamento de Conduta da Pecuária-TAC da Carne; (b) Acordos intergovernamentais como a Conferência das Nações Unidas sobre Mudanças Climáticas (COP 21) e os Relatórios de Inventário Nacional de Emissões de Gases de Efeito Estufa e (c) doações monetárias pelo fundo amazônia, que usam o PRODES como dado de referência à atividade de desmatamento na Amazônia Legal PESQUISAS ESPACIAIS (2022).

A política de transparência dos dados do monitoramento do estado da floresta adotada pelo INPE e pelo governo federal desde 2004 permite o acesso completo a todos os dados gerados pelos sistemas de monitoramento, possibilitando realizar avaliações independentes pela comunidade usuária, incluindo o governo em suas esferas e instâncias, a academia, o cidadão e a sociedade civil brasileira em seus diversos arranjos institucionais. PESQUISAS ESPACIAIS (2022)

Em seus relatórios divulgados após a consolidação o que ocorre até o fim do primeiro semestre no ano seguinte, tem apresentado uma constante evolução do desmatamento no bioma amazônico. Conforme esta apresentado na tabela 1, é possível verificar o total de km<sup>2</sup> desmatados nos dois últimos anos, 2020 e 2021.

Tabela 1 – Desmatamento em  $km^2$

ANO	$K M^2$ DESMATADOS
2020	10.851 $km^2$
2021	13.038 $km^2$

## 2.2 A Moratória da soja

Com a iniciativa de enfrentar o problema do desmatamento, a moratória contribuiu para inovações institucionais e de governanças, conforme BRANDAO (2020). Uma vez que grupos da sociedade civil passaram a realizavam campanhas e divulgar trabalhos para chamar atenção sobre os impactos do desmatamento e sua a correlação com a produção de alimentos.

O Greenpeace lançou no ano de 2006, um relatório com a seguinte definição: “*Eating up the Amazon*” – Comendo a Amazônia. Onde divulgou dados sobre a expansão de grãos em áreas do Bioma Amazônico. Este estudo apresentou a corresponsabilidade da indústria que processa e exporta grãos no desmatamento ao comprar grãos oriundos de área desmatada. Por meio desse relatório mobilizou um alerta internacional para a proteção da floresta Amazônia e registrou a conexão da perda da floresta com os mercados consumidores, principalmente o europeu.

De acordo com o relatório, as empresas que comercializavam grãos deveriam ser responsáveis pelo desmatamento, uma vez que o modelo de financiamento adotado pelas empresas para o plantio de grãos, não exigiam nenhuma restrição para o plantio em áreas recém desmatadas. Não obstante, o estudo apontou também para a dinâmica de expansão da soja na Amazônia, que levou a instalação de projetos de infraestrutura para secagem e armazenamento de grãos, como portos graneleiros. E esse movimento estimulou a grilagem de terra e conseqüentemente o desmatamento (GREENPEACE, 2006).

Ainda neste relatório foi demonstrado que a comercialização da soja, pela empresa Cargill, era advinda de áreas recém desmatadas, em seguida essa soja teve como destino o mercado europeu, a qual passou pela indústria alimentícia e chegou até a rede de *fast food McDonald's*, onde foi consumida pelo cidadão europeu na forma de *Chicken McNuggets*. Essa campanha evidenciou como o consumo na Europa impactava diretamente a dinâmica de cobertura e uso da terra na Amazônia, pressionando a floresta para a produção de commodities.

Em decorrência dessa campanha ficou evidente que a cadeia produtiva não conhecia a origem da soja produzida, e não era capaz de rastrear a soja comercializada, portanto, não existia uma política setorial, naquele momento, que pudesse evitar esse risco e excluir a soja proveniente de áreas desmatadas do Bioma Amazônico.

Por conta dessa campanha, foi proposto um diálogo e acordos foram criados entre as principais e maiores *traders* de grãos e organizações da sociedade civil. Buscando soluções para redução do desmatamento da Amazônia no Brasil. Em 24 de julho de 2006 a Associação Brasileira das Indústrias de Óleo Vegetais (ABIOVE) e a Associação Nacional dos Exportadores de Cereais (ANEC) assinaram a Moratória da Soja, que é um compromisso cujo o qual as empresas associadas se comprometeram a não originar soja e demais grãos de áreas desflorestadas do bioma Amazônico, a partir dessa data.

No entanto dois anos depois, a data de corte foi alterada para o ano de 2008. A inovação desse processo é que o arranjo institucional estabelecido para viabilizar a Moratória da Soja estruturou um processo de governança da cadeia da soja que contou com a participação de

organizações da sociedade civil, conforme afirmou (PIATTO; SOUZA, 2017).

Conforme relato de CASHORE (2002), o acordo é possivelmente um dos principais compromissos ambientais privados. A moratória é o reflexo de como a globalização exerceu influência direta, e propiciou a estruturação de processos de governança com arranjos institucionais inovadores, em que atores sociais de campos distintos de ação pudessem criar um mecanismo capaz de ampliar e aprimorar a atuação do campo econômico no fornecimento mais sustentável na sua cadeia de suprimento.

Ainda sobre a moratória da soja, em CARDOSO (2008) argumenta que o acordo significou a participação efetiva de campos distintos no debate e construção de parâmetros de sustentabilidade para a cadeia produtiva. Nesse estudo foi discutido o processo pelo qual a indústria da soja, caracterizada como o campo organizacional dominante, em um momento de crise institucional, gerada pela campanha de combate ao desmatamento da Amazônia liderada pelo Greenpeace, decidiu responder à pressão dos ambientalistas e negociou uma agenda de conservação da floresta dentro do campo organizacional da produção de soja.

A efetividade da moratória da soja, e conseqüentemente desse modelo de governança, é amplamente avaliada e discutida. A governança da moratória disponibiliza publicamente o relatório de monitoramento anual, o qual reportam os desmatamentos vinculados diretamente à soja na Amazônia. Os últimos relatórios mostraram significativas taxas de conformidade com o requisito central do acordo, desmatamento zero de floresta para plantio de soja. O último relatório relativo à safra 2017/2018 mostrou 64,3 mil hectares de soja em áreas desmatadas que estavam em desacordo com a Moratória, essa área representa apenas 1,4% da área de soja plantada no bioma (ABIOVE, 2017).

Conforme estudos de GIBBS et al. (2015) sobre os resultados de caráter prático dessa governança, pesquisadores afirmam que a moratória influenciou a tomada de decisão acerca da mudança na cobertura e uso da terra e conseqüentemente, a redução no desmatamento, pois ela estabeleceu um acordo mais rigoroso que o código florestal para as propriedades privadas. Ainda nesse estudo os autores supracitados apresentam que apenas 115, em meio a milhares de produtores de soja, violaram a moratória desde que ela foi criada. Entretanto, mais de 600 destes produtores violaram o Código Florestal no mesmo período GIBBS et al. (2015).

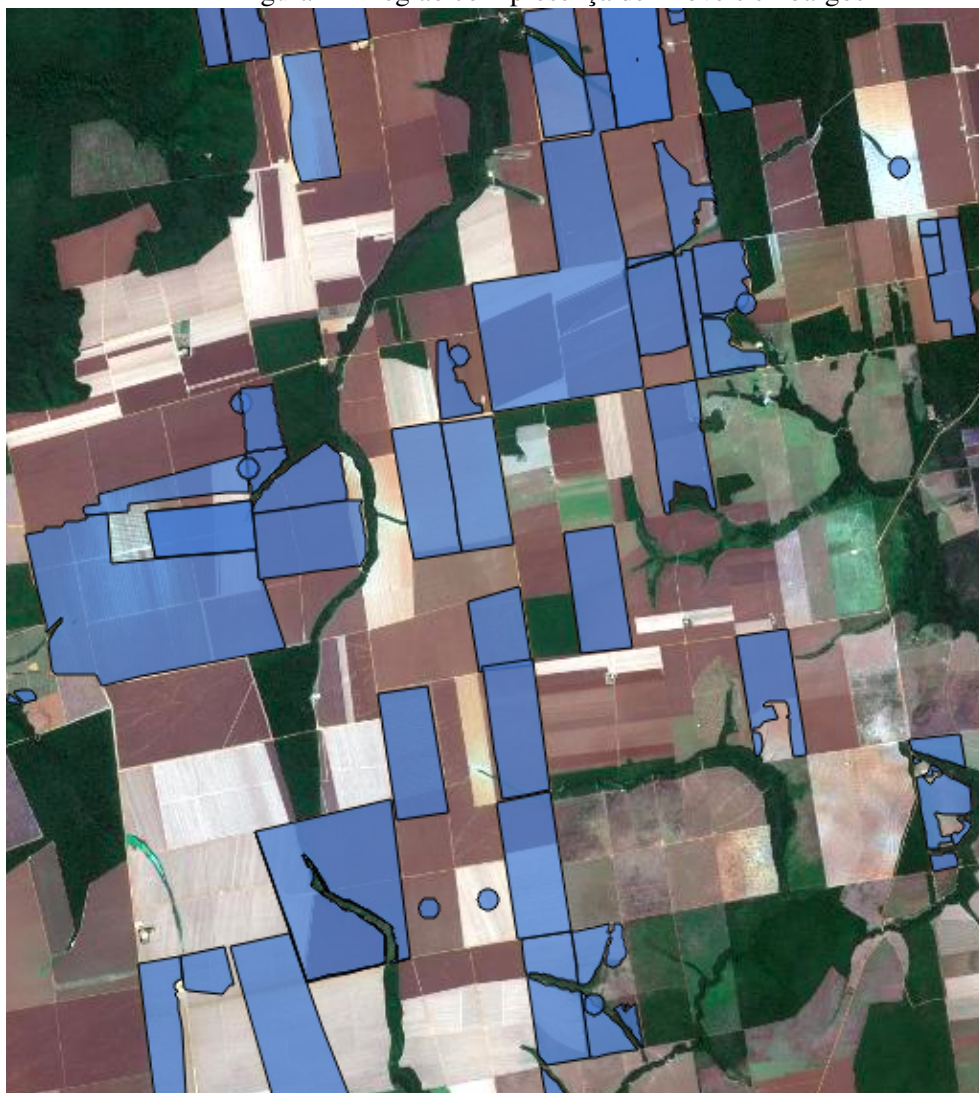
Por outro lado, apesar de ser um dado positivo é importante ressaltar que o cumprimento da moratória não significa que a propriedade rural esteja cumprindo com o código florestal. Assim sendo, a moratória não se distende à regularidade ambiental da propriedade. Somente ao avanço da soja em áreas desmatadas após 2008, em outros termos, se o proprietário continuar desmatando outras áreas da propriedade e não plantar soja nessa área, ele não será enquadrado pela moratória. É o código florestal que regulamenta a porcentagem de área que deve ser preservada na propriedade rural nos diferentes biomas do Brasil e estabelece as regras para regularizar os passivos ambientais (CIVIL et al., 2004).

Para a supressão de vegetação nativa de acordo com a normativa do código florestal é necessário autorização do órgão ambiental estadual. Entretanto, mesmo amparado em cima de

resultados positivos da moratória da soja, ao se analisar os históricos das safras por meio de imagens de satélites e cruzar informações oriundas do IBAMA, ICMBio ou órgãos estaduais, iremos constatar que existem imóveis rurais embargados que continuam produzindo grãos e outras áreas na mesma situação com pastagens, criando cabeças de gado. Com isso, esses imóveis através de seus proprietários estão conseguindo burlar a moratória da soja, informando que a procedência de seus produtos é de um imóvel regular.

Notem que na Figura 4, estão apresentados diversos imóveis de uma região localizada ao norte do Estado do Mato Grosso, cuja a imagem é datada do dia 25 de junho de 2022, safra de milho.

Figura 4 – Região com presença de imóveis embargos



Fonte: Imagem Sentinel-2 (25/06/2022) - Camada pública de embargos do IBAMA

Porém a informação complementar que está sendo demonstrada através desses detalhamentos na cor azul, constitui as áreas embargadas pelo IBAMA. Analisando mais a fundo é possível verificar através da transparência que estes imóveis por ora embargados, estavam plantando.

Todavia, são através destes imóveis que originam o comércio do grão pirata. A destarte, o

presente trabalho intuiu auxiliar na identificação desses produtores, através de um modelo de estimativa de safra alertando assim que determinada unidade produtora (imóvel rural) não seria capaz de produzir tamanha produtividade, combatendo assim a "lavagem" de grãos.

### 2.2.1 O grão pirata

O termo grão pirata, é dado por conta de uma manobra feita por maus agricultores, que plantam de maneira ilegal grãos em terras embargadas pelo IBAMA ou qualquer outro órgão competente. Geralmente essas terras estão embargadas por terem sido desmatadas, o que torna proibida a produção, segundo a legislação ambiental e amparada pela moratória da soja. Mas esses maus agricultores acabam produzindo e comercializando esses grãos através de um outro imóvel que esteja legal, quanto a moratória da soja e ao código florestal. Realizando com isso a denominada lavagem dos grãos.

O sistema de monitoramento da moratória da soja possui brechas. Uma delas se refere ao fato da moratória considerar somente a propriedade rural onde o desmatamento ocorreu, ignorando, por exemplo, fazendas vizinhas do mesmo proprietário. Ponto onde os ativistas discutem que isso traria uma facilidade para a chamada "lavagem de grãos" ou "triangulação", que é quando um agricultor tem várias fazendas e tenta lavar o grão "pirata" por meio de fazendas "limpas", misturando os grãos colhidos em diferentes imóveis.

Com isso, esses grãos "lavados" podem entrar em cadeias de abastecimentos internacionais supostamente "limpos", apesar da existência da moratória.

### 2.2.2 Metodologia para compra de grãos

Hoje existem diversas empresas que comercializam commodities no Brasil, em especial grandes multinacionais. Estas empresas estão comprometidas por meio da moratória da soja a fazerem a aquisição de grãos de uma unidade produtora, somente se esta atender os requisitos pré-estabelecidos na moratória, como não adquirir grãos provenientes de unidades embargadas.

São várias as empresas que comercializam *commodities* no Brasil em especial grandes multinacionais. Estas empresas estão comprometidas por meio da moratória da soja a fazerem a aquisição de grãos de uma unidade produtora, somente se esta atender aos requisitos pré-estabelecidos na moratória, dentre eles o de não adquirir grãos provenientes de unidades embargadas.

Diante disso, criou-se nessas empresas um ritual, onde um analista de commodity realiza uma série de verificações, que vão desde uma análise documental do imóvel à uma checagem sócioambiental que em síntese atende a moratória da soja. Também existem diferentes formas desse contrato de compra de grão acontecerem, como aquele onde o produtor apenas vende sua produção e o contrato onde o produtor antecipa o valor de venda do grão como forma de financiar sua produção em troca de um percentual que for produzido. Sendo assim, serão

elencados a seguir o processo de análise que é efetivado antes de qualquer contrato.

- Análise documental - nesta análise será checado quem é de fato o responsável pelo imóvel, se este existe efetivamente, bem como outras informações referentes ao mesmo, que pode vir a ser um futuro impedimento. Também pode ser solicitada algumas informações relevantes para empresa, como a média de produção da região, ou alguma outra informação que pode vir a ser necessária em conformidades com o tipo de operação que será realizada. Como no caso de um financiamento da produção, o imóvel ainda deverá passar por um crivo de valor de mercado, onde geralmente o futuro devedor, precisará apresentar uma garantia de 2,5 vezes do valor a ser financiado.
- Análise Sócio-ambiental – neste tipo de análise será verificado qualquer item que possa em algum momento tornar a área de produção embargada é levado em conta, posto isto estes itens visam atender a moratória da soja tais como: se a terra sobrepõe uma terra indígena, unidades de conservação, áreas de reserva legal, áreas de preservação permanente, se o imóvel possui alerta de trabalho escravo, se a área produtora é consolidada de acordo com o marco florestal vigente, embargos, multas ou infrações.

Após as análises, então vai para a ultima etapa do processo que visa elaborar o parecer, contendo todas essas checagens, auxiliando o analista a prosseguir com parecer favorável para compra ou não. Enfatizando que deve sempre estar de acordo com a moratória da soja.

Destacam-se aqui que o analista não tem nenhuma checagem de quanto a área produtora de fato produz, fica subentendido que a área vai produzir de acordo com a média da região, que é uma informação fornecida pelo Instituto Mato-Grossense de Economia Agropecuária (IMEA), mas que não entra de fato no *checklist* de itens das análises. Abrindo por tanto, uma margem para que produtores mal intencionados consigam embutir na sua produção, grãos oriundos de imóveis embargados.

### **2.3 Agricultura de precisão**

A agricultura tem passado por processo constante de evolução e se transformou nos últimos anos, produzindo muita informação e dados de diversas fontes. A agricultura de precisão é componente da agricultura inteligente, traz uma abordagem de gerenciamento agrícola, de tempo e espaço, sempre no intuito de aumentar a produção agrícola, de forma eficiente e eficaz.

Promovendo a agricultura de precisão para aumentar a produção e qualidade de alimentos, tudo por meio de tecnologias. A agricultura inteligente permite que agricultores monitorem e controlem, de forma remota, podendo fazer até mesmo um passeio, sem estar de corpo presente, nas suas lavouras, atendendo as devidas necessidades.

A agricultura inteligente tem enfrentado os desafios associados ao aumento da produção de alimentos, preservação do meio ambiente e os cenários políticos sem aumentar as áreas cultivadas, trazendo a possibilidade de ser mais sustentável e promover a preservação ambiental.



Desenvolvendo dessa forma um papel importante para o aumento da produtividade agrícola e preservação do meio ambiente. Possibilitando através do uso de diferentes tecnologias e recursos computacionais, tais como automação, captura de dados, processamento de tomada de decisão, transmissão de dados, inteligência artificial, internet das coisas, redes neurais e etc. Proporcionando fazer o controle desde a semeadura até a colheita.

Visando o aumento do retorno econômico e a redução da degradação ao meio ambiente, a agricultura de precisão vem promovendo melhorias na gestão das informações relacionadas a produção, baseando-se na variabilidade espacial e temporal de uma unidade produtiva, auxiliando os gestores na tomada de decisão. De acordo com BRAMLEY (2009), que considera a agricultura de precisão como um conjunto de tecnologias que promovem melhorias na gestão dos sistemas de produção com base no reconhecimento de que o "potencial de resposta" podem variar consideravelmente, mesmo em pequenas distâncias, da ordem de pouco metros.

Diante disso, a agricultura de precisão traz muitos benefícios tanto aos produtores quanto para a população e ao meio ambiente. Pois há um ganho em lucratividade devido às análises que podem ser realizadas de maneira mais precisa, permitindo ao produtor obter análises sobre os riscos e na economia de insumos como fertilizantes com aplicação de somente quantidades exatas e necessárias de acordo com a variabilidade espacial e temporal. Levando ao uso reduzido de agrotóxicos e ocasionando a redução do impacto ao meio ambiente.

### 2.3.1 A estimativa de safras

Hoje existem diversos pesquisadores ao redor do mundo buscando criar um algoritmo de estimativa, visando auxiliar nas tomadas de decisões em diversos tipos de culturas que vem sendo cultivadas. Levando em conta que temos um constante aumento da população e que a agricultura pode ser um fator de peso para a balança comercial de um país. O que se torna necessário que tenha uma produção suficiente.

De acordo com TERLIKSIZ; ALTILAR (2019), a previsão do rendimento das culturas tem impacto direto nas economias nacionais e internacionais e desempenham papel importante na gestão e segurança alimentar. O que acaba sendo muito importante para o desenvolvimento econômico de uma nação.

Quanto aos agricultores, todos necessitam de uma estimativa de produção, uma vez que os mesmos precisam ter previsibilidade sobre o retorno do investimento, servindo também para otimizar os trabalhos com maquinários e implementos, para a preparação da logística de transporte e armazenamento e também por servir de auxílio para ensaios com variabilidade genética da produção e diferentes formas de manejo.

Diante disso, se existem diferentes critérios que são adotados para calcular a estimativa de produção de uma propriedade. No caso do presente estudo, será dado o foco na estimativa de produção de milho, uma vez que foi a cultura escolhida a ser estudada, visto esta cultura se balizar como a mais importante a nível comercial nas Américas.

Dada a essa importância econômica a planta do milho se distingue pelos diversos diversos critérios de sua utilização, que vai desde a alimentação animal à indústria de alta tecnologia. O consumo do milho em grãos, para alimentação animal representa boa parte do consumo do cereal, isto é, cerca de 70% no mundo. Já nos Estados Unidos, aproximadamente 50% está destinado a esse fim, enquanto que no Brasil varia entre 60% a 80%. Mas até o ano de 2030 um novo segmento, o etanol, deverá estar consumindo 20% da produção nacional.

Apesar do consumo humano, não ter uma participação muito elevada, este cereal constitui um fator importante em regiões de baixa renda. Em algumas situações, o milho constitui a ração diária de alimentação. A título de exemplo, no Nordeste do Brasil, o milho é a fonte de energia para muitas pessoas que vivem no Semiárido; citando caso análogo está na população mexicana, que tem no milho o ingrediente básico para sua culinária.

#### 2.3.1.1 Cálculos para estimativas de produção

De acordo com os engenheiros agrônomos Hauagge e Madalóz (2020), os métodos mais comuns para se calcular a estimativa de produção de milho, são:

##### **Método 1:** simples e objetivo (Figura 5)

1. Saber a população de plantas por área;
2. Coletar algumas espigas da área desejada;
3. Calcular o peso médio de grãos de cada uma delas;
4. Após ter o peso médio de grãos de cada espiga, basta multiplicar pelo número total de plantas encontradas no talhão.

Figura 5 – Cálculo do método 1

- Peso médio de grãos das espigas selecionadas: **180 gramas**
- População da área: **70.000 plantas/ha**

$$\text{Cálculo: } \mathbf{0,180 \times 70.000 = 12.600 \text{ kg/ha}}$$

Fonte: Adaptado de (USAMA FAYYAD; SMYTH, 1996).

Onde quanto maior for o número de espigas coletadas e avaliadas, menor será a margem de erro para a produtividade real no momento da colheita.

### **Método 2:** baseado na regra do cálculo *Corn Yield Calculator*

Esta regra foi publicada pela Universidade de Illinois nos Estados Unidos, onde se pode realizar seguindo as etapas:

1. Contar o número de espigas em 4 metros quadrados, conforme está apresentado na Tabela 2;

Tabela 2 – Contagem por comprimento

ESPAÇAMENTO ENTRE LINHAS(Cm)	COMPRIMENTO PARA SE OBTER 4m <sup>2</sup>
50	8,0 m
60	6,6 m
70	5,7 m
80	5,0 m
90	4,4 m
100	4,0 m

Fonte: Adaptado de (USAMA FAYYAD; SMYTH, 1996)

2. No mesmo local, deve selecionar quatro espigas representativas, contar o número de fileiras de grãos e o número de grãos por fileira para cada espiga. Não devendo ser contabilizado os grãos da extremidade que sejam menores que a metade do tamanho normal;
3. Em seguida, será estimado a produção através de cada uma das espigas, da seguinte maneira:

$$\mathbf{(n^\circ \text{ de espigas em } 4\text{m}^2) \times (n^\circ \text{ de fileiras de grãos}) \times (n^\circ \text{ de grãos por fileira}) \times 0,70^{*1} = \text{kg/ha com 15\% de umidade}}$$

<sup>1</sup>Fator de correção do método e transformação de bushels/acre para kg/ha.

4. Calcular a média de produção estimada das quatro espigas;
5. Repetindo as etapas de 1 a 4 em vários pontos do talhão e calcular a média dos resultados para estimar a produtividade final da área.

**Método 3:** indicado pela Empresa de Assistência Técnica e Extensão Rural (EMATER)

Neste método pode-se estimar a produtividade com a seguinte expressão:

$$\text{Produtividade (toneladas/ha a 15,5\% de umidade)} = [ (NE \times P) / EM ] / 1000^2$$

Assim sendo, temos os métodos mais usuais aplicados a cultura do milho. É possível verificar que nenhum método leva em conta, os dados históricos de produção vinculados a outros fatores que são preponderantes para uma boa colheita. Tais como: clima, índices pluviométricos, temperatura, pressão atmosférica, entre outros. A ausência de algum destes fatores vem corroborar com a linha de estudo proposta para este trabalho.

### 2.3.2 O índice de vegetação com diferença normalizada (NDVI)

O *Normalized Difference Vegetation Index* (NDVI), tem ganhado destaque nos últimos anos, sendo objeto de pesquisas e de algumas aplicações na agricultura. Esse índice que tem sido muito utilizado na agricultura de precisão para o gerenciamento de plantações e permite várias aplicações que diminuem as perdas e aumentam a produtividade. Por meio do sensoriamento remoto.

Embora tendo mais de 40 anos desde a sua primeira aparição. O índice é muito utilizado não somente na agricultura, mas também em diversos estudos ambientais. Através dele é possível fazer diferentes tipos de análises nas mais diversas escalas de uma determinada região ou plantação. Um dos usos mais frequentes do NDVI é a utilização para a medição da intensidade da atividade de clorofila nas plantas, podendo inclusive realizar comparações com períodos anteriores.

A energia captada e absorvida pelas plantas possui diversos espectros. O NDVI nada mais é que um cálculo realizado em cima dessas bandas espectrais. Esses espectros são captados por sensores, que na maioria dos casos estão instalados em drones ou satélites, e posteriormente são tratados pela seguinte equação do índice apresentada na Figura 6.

O NDVI é o resultado da razão entre a diferença da reflectância do infravermelho (IVP) e do vermelho (V) pela soma das mesmas variáveis. Esse cálculo resulta em um índice que varia entre -1 e 1. Na prática o valor representa a presença de vegetação, quanto maior ele é, maior é quantidade vegetativa do local. Valores negativos ou próximos a zero indicam áreas de água, ou solo nu, onde existe muito pouca atividade clorofiliana e com isso baixa quantidade de vegetação.

---

<sup>2</sup>NE: número médio de espigas em 10m lineares; P: peso médio de grãos por espiga corrigido para 15,5% de umidade, obtido pela média do peso de grãos de 3 espigas coletadas (gramas); EM: espaçamento entre linhas (m)

Figura 6 – Cálculo NDVI

$$\text{NDVI} = \frac{(\text{Infravermelho (IVP)} - \text{Vermelho (V)})}{(\text{Infravermelho (IVP)} + \text{Vermelho (V)})}$$

Fonte: Adaptado pelo autor.

O NDVI vem sendo utilizado na agricultura para ter um melhor controle das plantações e ganhos de produtividade. Dentre os usos aponta-se: o monitoramento das plantações e culturas, detecção de excessos de água, localização de pragas, mapeamento agrícola e mapeamento hídrico.

Assim o NDVI é um indicativo que pode auxiliar bastante nos levantamentos de sensoriamento remoto, possibilitando ótimos resultados e permitindo uma combinação de inúmeras fontes de informação. Esses dados utilizados em conjunto representam toda a condição de uma plantação, garantindo controle total sobre ela e com isso ganhos de produtividade.

## 2.4 A inteligência artificial no agro

A inteligência artificial é uma das principais áreas de pesquisa da ciência da computação, devido a sua robustez para resolver problemas complexos e em que muitas vezes são impossíveis de serem resolvidas por seres humanos (G. BANNERJEE U. SARKAR, 2018). A inteligência artificial pode perceber um determinado ambiente e tomar as medidas cabíveis pra resolve-las, baseando-se no que foi treinada (ENRIQUE CASTILLO JOSE M. GUTIERREZ, 1997).

Existem diversos paradigmas que compõe o universo da inteligência artificial, tais como: aprendizado de máquina (do inglês, *machine learning* – *ML*), sistemas especialistas, sensores, reconhecimento de padrões, robótica, redes neurais entre outros. Sua aplicabilidade são incalculáveis e a cada dia surgem novas soluções (ENRIQUE CASTILLO JOSE M. GUTIERREZ, 1997).

Uma das áreas que vem se destacando no uso da inteligência artificial, é a agricultura. Em suas mais variadas subáreas (PRADEEP N, 2019). Usada desde a análise de dados, passando pelo combate a pragas e chegando na predição de safras. Dentre as soluções que fazem uso de algum paradigma da inteligência artificial as que mais aparecem, nas pesquisas, quando se fala em *smart farm*, ou fazendas inteligentes, são: o aprendizado de máquina, o aprendizado profundo, as redes neurais artificiais e redes neurais convolucionais.

Na agricultura a inteligência artificial vem sendo utilizada para melhorar a eficiência e eficácia no aumento da produção, evitando o desperdício de tempo, produtos, recursos financeiros e danos ambientais (ENRIQUE CASTILLO JOSE M. GUTIERREZ, 1997).

### 2.4.1 O aprendizado de máquina

Basicamente o *Machine Learning* é a prática de usar algoritmos para analisar dados, aprender com eles e, por conseguinte, determinar ou prever sobre algo no mundo. Portanto, em vez de codificar manualmente as rotinas de software com um conjunto específico de instruções para realizar uma tarefa específica, a máquina é treinada usando grandes quantidades de dados e algoritmos que lhe permitem aprender como executar a tarefa.

Sendo assim, foi criada uma classificação dos algoritmos de aprendizado. Onde temos: aprendizado supervisionado, não supervisionado, por reforço e o aprendizado evolutivo.

A categoria de aprendizado supervisionado é aquela que começa com um conjunto de respostas corretas (conjunto de treinamento) para generalizar e responder corretamente a todas as entradas possíveis. E nos algoritmos de aprendizado não supervisionado comparam entradas entre si para identificar semelhanças e classificá-las em categorias. Enquanto os algoritmos de aprendizado por reforço podem ser definidos como um estágio intermediário entre supervisionado e não supervisionado.

O algoritmo é informado quando responde incorretamente a um questionamento, mas isso não acontece quando responde corretamente. Com isso, é necessário que seja explorado as respostas diferentes até encontrar as respostas corretas. Enfim, o aprendizado evolutivo é uma forma de aprendizado baseada em algoritmos evolutivos que, por sua vez, constituem uma classe de algoritmos utilizados para otimização.

Algoritmos evolutivos fazem parte de uma classe mais geral de algoritmos de otimização, os chamados algoritmos estocásticos adaptativos e o termo “Algoritmo Evolutivo” surgiram para abranger uma série de propostas de algoritmos desenvolvidas de forma independente, mas com conceitos semelhantes como a analogia com evolução natural ou genética natural, verificação de múltiplos pontos paralelamente e utilização de operadores adaptativos (ARAÚJO, 2005).

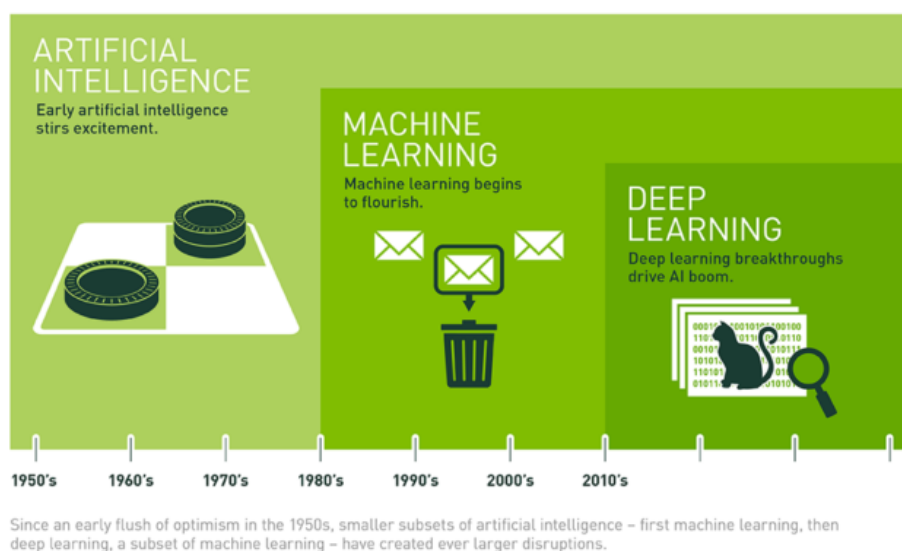
Ao se versar sobre algoritmos de aprendizado é preciso ter em mente que será exigido uma grande quantidade de dados. No entanto só quantidade não é importante, tem que ser considerado uma grande quantidade de dados, preferencialmente sem muito ruído. Uma vez que o custo computacional que aumenta proporcionalmente, a quantidade de dados é um ponto relevante a ser considerado.

- Seleção de características de interesse: é identificar as características mais significativas para o problema que será confrontado;
- Seleção de algoritmo: consiste em selecionar o algoritmo mais apropriado para tratar o problema que está sendo estudado;
- Seleção de Parâmetros: alguns algoritmos precisam ser ajustados por parâmetros que exigem a definição da experimentação;

- **Treinamento:** dado um conjunto de entradas, parâmetros e algoritmo, o treinamento consiste na construção do modelo computacional que será usado na previsão das respostas para novos dados;
- **Classificação:** o sistema precisa ser avaliado quanto a precisão dos dados que foram treinados.

Um fator importante a ser contextualizado, é a hierarquia dessa área de estudo, conforme a Figura 7 são as diferenças entre os termos, que muitas vezes são utilizados de forma indiscriminada e generalizada.

Figura 7 – Contextualização da hierarquia da área de estudo



Fonte: Adaptado de (COPELAND, 2016)

Dentro da inteligência artificial existem diversos paradigmas diferentes que modelam essa “inteligência”. Alguns destes paradigmas podem ser classificados na área de aprendizado de máquinas, do inglês (*machine learning*), como descrito anteriormente, “aprendem” a tomar uma decisão baseada em exemplos de um problema, e não de uma programação específica. E que necessitam de dados para extrair características e aprendizado que poderão ser utilizados numa tomada de decisão futura.

Um subgrupo específico do paradigma de aprendizado de máquina é denominada de aprendizado profundo, do inglês *deep learning (DL)*, este se utiliza redes neurais profundas e dependem de muitos dados para o treinamento. Existem muitos fatores que as diferenciam do paradigma de aprendizado de máquina, como: a necessidade e o efeito de muitos dados, poder computacional e a flexibilidade na modelagem dos problemas. E alguns desses fatores favorecem a utilização dessas técnicas em áreas como visão computacional e processamento de linguagem natural. O aprendizado profundo possui um grande diferencial, que é a capacidade de interligar vários neurônios em estruturas mais complexas para resolver problemas.

Nesse contexto aponta-se que o aprendizado profundo detém uma função de inteligência artificial e um dos seus principais recursos estabelece-se no aprendizado da máquina, para que esta venha a aprimorar sua capacidade de reconhecer dados, gerando *insights* e sobretudo: levando em consideração uma grande massa de dados.

Como um grande case de aplicação destas práticas e conjuntos de tecnologias é o Google Tradutor. Em 2016, uma grande companhia de *Mountain View* anunciou a utilização do *Google Neural Machine Translation* (GNMT), um sistema para melhorar a qualidade das traduções realizadas pelo serviço. Antes da implementação dessa tecnologia baseada em aprendizado de máquina, o serviço realizava traduções utilizando partes independentes de frases. Após o algoritmo de aprendizado profundo, essa tarefa se tornou mais inteligente e agora o sistema busca traduzir frases completas e adaptando-se ao contexto.

Aprendizagem profunda vem sendo amplamente utilizada para processamento de imagens, mas diversos estudos recentes têm aplicado a aprendizagem profunda em outros domínios como texto e mineração de dados. A arquitetura básica em uma rede neural é uma rede não-linear totalmente conectada de nós de processamento organizados em camadas. A primeira faixa é a camada de entrada, a faixa final é a camada de saída e todas as outras camadas estão escondidas.

#### 2.4.1.1 Algoritmos de Aprendizado de Máquina

De acordo com GUIMARÃES (2019), na teoria matemática das redes neurais, o teorema de aproximação universal observa que uma rede avança tal como as redes Perceptron de multicamadas do inglês, *Multilayer Perceptron* com uma única camada oculta contendo um número finito de neurônios é capaz de aproximar qualquer função contínua mensurável com qualquer grau de precisão desejado. As redes *Multilayer Perceptron* são recursos muito poderosos podendo ser utilizada para criar modelos preditivos a partir de um conjunto de dados de entrada contendo um determinado número de variáveis e, como resultado, prever uma ou mais variáveis de destino. A topologia de um *Multilayer Perceptron* é simples e direta. A rede é dividida em camadas, que vêm em três tipos: camada de entrada, camadas ocultas e camadas de saída. Neste trabalho, um único nó de saída foi usado para produzir a estimativa da produtividade do milho por meio da regressão.

Cada nó na camada inferior é conectado a todos os nós na camada acima, formando um gráfico bipartido completo. Cada sinal de entrada propaga-se no sentido para frente, camada a camada através da rede. Um peso está associado à conexão. A saída de um nó na rede é calculada com uma função de ativação cuja entrada é a soma ponderada das conexões de entrada.

##### I - *Multilayer Perceptron* - MLP

Vem a ser uma rede neural com uma ou mais camadas ocultas com um número indetermi-



nado de neurônios. A camada oculta possui esse nome porque não é possível prever a saída desejada nas camadas intermediárias. Para treinar a rede MLP, o algoritmo comumente utilizado é o de retropropagação do inglês, (*Backpropagation*).

O algoritmo de aprendizado da MLP é chamado retropropagação é composto de 4 passos:

- 1º Passo: inicialização
  - Atribuir valores aleatórios para os pesos e limites;
  - Escolha dos valores iniciais influencia o comportamento da rede;
  - Na ausência de conhecimento prévio dos pesos e limites devem ter valores iniciais aleatórios e pequenos uniformemente distribuídos.
  
- 2º Passo: ativação
  - Calcular os valores dos neurônios da camada oculta;
  - Calcular os valores dos neurônios da camada de saída.
  
- 3º Passo: treinar os pesos
  - Calcular os erros dos neurônios das camadas de saída e oculta;
  - Calcular a correção dos pesos;
  - Atualizar os pesos dos neurônios das camadas de saída e oculta.
  
- 4º Passo: iteração
  - Repetir o processo a partir do passo 2 até que satisfaça o critério de erro.

## II - *Support Vector Machine*

Os algoritmos de aprendizagem de máquina (SVM) têm como objetivo a determinação de limites de decisão que produzam uma separação ótima entre classes por meio da minimização dos erros (VAPNIK; GUYON; HASTIE, 1995).

A classificação é o processo que permite relacionar *pixels* das imagens de satélite a uma devida classe do terreno. Os algoritmos normalmente utilizados para a realização deste processo são conhecidos por classificadores de imagens (MATHER et al., 1999).

O SVM consiste em uma técnica computacional de aprendizado para problemas de reconhecimento de padrão. Introduzida por meio da teoria estatística de aprendizagem por isso essa classificação é baseada no princípio de separação ótima entre classes, tal que se as classes são separáveis, a solução é escolhida de forma a separar o máximo as classes. Recentemente, o SVM foi utilizado na área de sensoriamento remoto com relativo sucesso (MELGANI; BRUZZONE, 2004).

#### 2.4.2 Redes neurais convolucionais

As redes neurais convolucionais (CNNs) descrita por LECUN et al. (1998), são redes modeladas de acordo com a arquitetura do córtex visual, em que os neurônios não estão totalmente conectados, mas são espacialmente distintos. Já em OQUAB et al. (2014) relata que as CNNs fornecem excelentes resultados na generalização da classificação de objetos em imagens.

De acordo com o (HAYKIN, 2009), a estrutura de uma rede neural convolucional é dividida em três objetivos principais: Extração de características, Mapeamento de características e Subamostragem.

Onde a extração de características recebe os sinais de entrada de um campo receptível, é possível extrair as características locais. Isso faz com que a posição exata de cada pixel, no caso de uma imagem, seja irrelevante, ou seja, não é importante saber onde um determinado objeto se encontra, desde que os aspectos que o compõem estejam organizados e conectados da mesma forma.

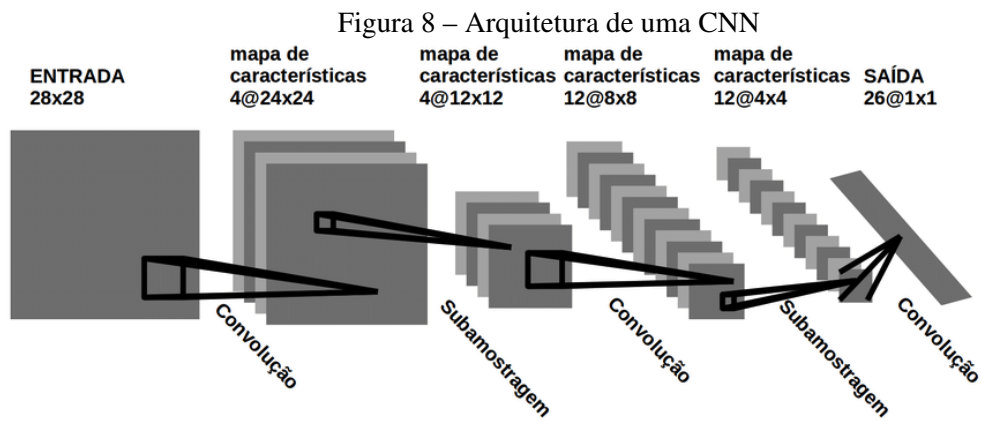
Enquanto no mapeamento de características, Cada camada computacional da rede neural convolucional é composta por diferentes mapas de características, que dão robustez ao modelo, fazendo com que seja possível lidar com variações, como distorção, rotação ou translação na imagem.

Já na subamostragem, é aplicada para coletar amostras de cada mapa de característica, produzindo uma sumarização.

As redes neurais de convolução (CNN), de acordo com LECUN et al. (1998), utilizam camadas com filtros de convolução que são aplicados a características locais. Originalmente criada para atender a visão computacional, os modelos CNN foram subsequentemente mostrados como eficazes para processamento da linguagem natural (PNL) e alcançaram excelentes resultados na análise semântica (YIH; HE; MEEK, 2014) e classificação de imagens (OLIVEIRA, 2019).

Em HAYKIN (2001) introduziu arquitetura original da rede neural convolucional, alterando entre camadas convolucionais incluindo não-linearidades e camadas de subamostragem. As camadas convolucionais já incluem não-linearidades e, assim, uma camada convolucional na verdade representa duas camadas. Os mapas de características da camada final de subamostragem são então inseridos no classificador real que consiste em um número arbitrário de camadas totalmente conectadas. A camada de saída geralmente usa as funções de ativação softmax.

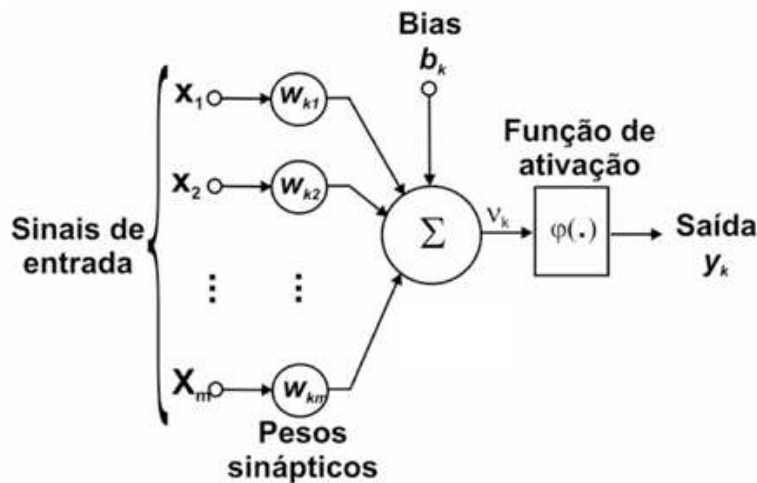
É possível notar na Figura 8, a representação de como a entrada é propagada para avaliar redes neurais profundas discretizadas com uma profundidade arbitrária de camadas ocultas para chegar a uma classificação.



Fonte: Adaptado de pelo autor

Na figura 9, o modelo de neurônio perceptron recebe múltiplos sinais de outros neurônios através de seus detritos, e cada um destes sinais é multiplicado pelo próprio peso da conexão. Em KUBAT (1999) é relatado que estes sinais são adicionados no corpo celular ou função somatória, e quando este sinal composto alcança um valor umbral, um sinal potencial é enviado pelo axônio, o qual é a saída do neurônio. Cada neurônio realiza uma função linearmente dependendo dos valores dos pesos de entrada seguidos por operações não lineares.

Figura 9 – Modelo de um neurônio artificial "Perceptron"



Fonte: Adaptado de (HAYKIN, 2001)

A história da evolução da humanidade é permeada de ferramentas criadas pelo homem para melhorar seu desempenho. E as máquinas mediante suas ações aumentam a capacidade humana de realizar uma série de tarefas na medida em que consegue armazenar muitas informações e buscar dados com rapidez impressionante, processando milhões de informações ao mesmo tempo, levando menos do que um nanosegundo. Por meio desses instrumentos tecnológicos obtém-se maior velocidade, precisão e flexibilidade.

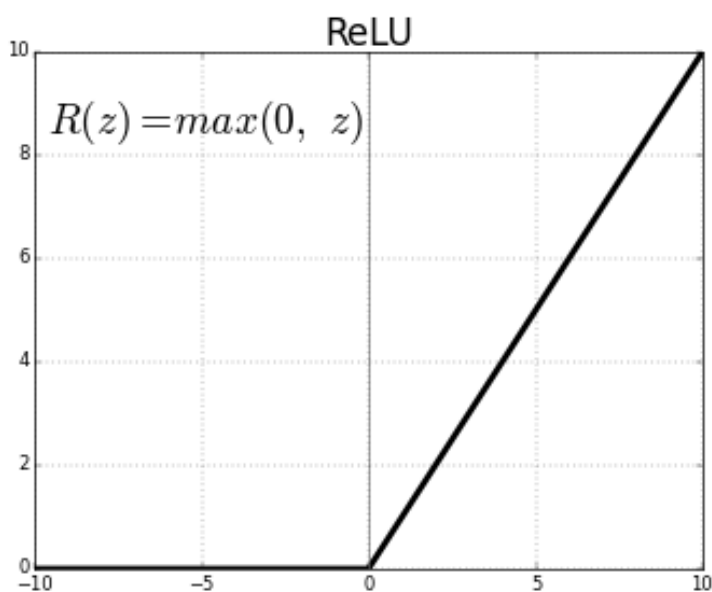
### 2.4.2.1 Função de Ativação

A função de ativação é a transformação não linear que fazemos ao longo do sinal de entrada. Elas basicamente decidem se um neurônio deve ser ativado ou não. Ou seja, se a informação que o neurônio está recebendo é relevante para a informação fornecida ou deve ser ignorada.

Há tempos atrás, a função sigmoide era a função de ativação mais popular devido à sua não linearidade. Com o passar do tempo, as redes neurais avançaram para arquiteturas de rede mais profundas, o que levantou a nós depararmos com outro problema: a dissipação do gradiente. Diante disso a unidade linear retificada (ReLU) acabou sendo a opção padrão para a função de ativação das camadas ocultas, pois inibe o problema de dissipação do gradiente já que possui um gradiente maior que o da função sigmoide.

Assim sendo para o modelo DYP foi utilizado a função *Rectified Linear Unit* (ReLU) na camada de entrada e nas camadas ocultas. Atualmente a ReLU é a função de ativação mais utilizada ao projetar redes neurais. A função ReLU se apresenta conforme com a figura 10.

Figura 10 – Função ReLU



Fonte: Adaptado pelo autor

Em redes neurais, é de fundamental relevância ter uma função de ativação não-linear. A principal vantagem de usar a função ReLU sobre outras funções de ativação é que ela não ativa todos os neurônios simultaneamente. Sendo assim, caso venha a ocorrer uma entrada negativa, ela irá convertê-la para zero e o neurônio não será ativado.

Com isso, busca-se a evitar *overfitting* adicionamos uma camada de *dropout*, que de forma simplificada ignorará alguns neurônios escolhidos aleatoriamente durante a fase de treinamento, ou seja, tais neurônios não serão considerados durante o processo. Entretanto tecnicamente, em cada estágio de treinamento, nós individuais são eliminados da rede com probabilidade  $1 - p$  ou

mantidos com probabilidade  $p$ , reduzindo o tamanho da rede.

## 2.5 Discussão

As técnicas de aprendizado de máquina que vem a ser uma paradigma da inteligência artificial, baseiam-se na ideia de que sistemas podem aprender com os dados, identificando padrões e tomando decisões com o mínimo de intervenção humana.

A rede neural convolucional desponta-se em face de um uso muito popular no reconhecimento e classificação de imagens, onde técnicas de aprendizado profundo tem apresentado ótimos resultados.

Pelo fato de trabalhar numa estrutura hierárquica, cada neurônio na segunda camada convolucional está conectado apenas aos neurônios localizados dentro de um pequeno retângulo na primeira camada. Essa arquitetura permite que a rede se concentre em recursos de baixo nível na primeira camada oculta e em seguida, monte-os em recursos de nível superior na próxima camada oculta.

Portanto, se torna uma das principais razões nas quais CNN's funcionam muito bem no reconhecimento de imagens. Além de criar um modelo capaz de fazer a estimativa da produtividade de uma safra, temos a intenção de testar e comparar os resultados de alguns algoritmos que se demonstram promissores de durante a revisão da literatura. Apesar de haver um foco nesta pesquisa para a cultura do milho, pretende-se deixar em aberto para um futuro, a possibilidade de criar um modelo que seja capaz de realizar a estimativa da produtividade de safras independente da cultura plantada.

### 3 REVISÃO DA LITERATURA

Este capítulo encontram-se apresentados os trabalhos relacionados a estimativa da produtividade de safras, do inglês crop yield prediction, e pesquisas que abordam temas similares na área. Sem dúvida, diversos autores já realizaram e muitos ainda vem realizando trabalhos com a ideia de conseguir um melhor método para a estimativa de safras. No entanto poucos buscam aplicar a capacidade do modelo de oferecer alguma explicação sobre seus resultados.

Nesse sentido, o presente capítulo também apresenta uma breve discussão sobre cada trabalho, considerando suas principais contribuições e limitações. Na seção 3.1 discorre-se sobre a metodologia aplicada para a escolha de cada trabalho. A seção 3.2 estão disponibilizados os resultados das buscas. Já na seção 3.3 são apresentados os resultados baseados nos critérios de exclusão. Por fim, a seção 3.4 versa-se sobre a conclusão e as considerações finais sobre o capítulo.

#### 3.1 Metodologia

O processo de pesquisa realizado neste trabalho é baseado no método de revisão sistemática - SLR. A revisão sistemática é uma revisão rigorosa dos resultados de uma pesquisa, identificando, avaliando e interpretando os trabalhos já publicados (KITCHENHAM, 2004).

Uma revisão sistemática não se destina apenas a agregar evidências existentes, mas também a apoiar pesquisadores e o desenvolvimento de novas pesquisas (B. KITCHENHAM P. BRERETON, 2009). Para a elaboração deste artigo, baseou-se na proposta de SLR de (B. KITCHENHAM P. BRERETON, 2007), com as seguintes etapas:

- As questões de pesquisa foram definidas: as questões nortearam a busca de artigos relevantes para o tema pesquisado;
- O artigo processo de busca: apresenta como a pesquisa foi realizada e as fontes científicas utilizadas;
- Seleção dos estudos relevantes: criação e definição dos critérios utilizados na seleção dos estudos considerados relevantes;
- Avaliação da qualidade dos estudos: análise da qualidade, quantitativamente, dos estudos já selecionados.

##### 3.1.1 Questionamentos a serem verificados

A elaboração de questões de pesquisa é uma das partes principais do SLR, pois fornecerá os meios para compreender e conduzir SLR no tema pesquisado, (B. KITCHENHAM P. BRERETON, 2007). Nesta pesquisa as questões buscam saber como se dá uma aplicação da inteligência

artificial na estimativa de safras. Elaborou-se uma questão central, com base em estudos prévios sobre o tema do trabalho para esta revisão, em busca de soluções existentes.

De acordo com a questão principal, questões específicas foram elaboradas para preencher as possíveis lacunas e identificar como soluções existentes. As questões de pesquisa foram definidas de acordo com a questão principal, a fim de respondê-la:

- QP1 - Quais tecnologias computacionais foram usadas para a estimativa de produtividade?
- QP2 - Como a Inteligência Artificial foi usada para a estimativa de produtividade?
- QP3 - Como aprendizado de máquina e visão computacional foram usadas para a estimativa de produtividade?

### 3.1.2 Processo de pesquisa

A realização do processo de pesquisa foi definida da seguinte forma: leitura de diversos artigos sobre o tema de interesse, definição de termos para busca e seleção de fontes. A seguir foram executados os passos descritos nas subseções 3.2, 3.3 e 3.4 e suas subseções. Os critérios de seleção e estratégias de busca foram criados e selecionados a fim de responder às questões de pesquisa.

As fontes foram selecionadas ou escolhidas a partir da relevância das publicações na área, desde fontes eletrônicas, trabalhos publicados em jornais, periódicos e congressos. As bases eletrônicas selecionadas foram cinco:

- [ieeexplore.ieee.org](http://ieeexplore.ieee.org);
- MDPI;
- [Scholar.google.com](http://Scholar.google.com);
- [ScienceDirect.com](http://ScienceDirect.com);
- [SpringerLink.com](http://SpringerLink.com).

As palavras-chave para a criação da *string* de busca da pesquisa, foram levantadas e elaboradas, a partir das questões de pesquisa e leituras de artigos sobre o tema. A definição da *string* de pesquisa foi baseada nas palavras-chave e sinônimos do tópico de interesse. Diversas strings foram testadas até o seu uso, eliminando o máximo possível de resultados indesejados. A *string* foi adaptada de acordo com o mecanismo de pesquisa de cada fonte e mantendo todas as palavras.



("Agriculture" OR "Intelligent Agriculture" OR "Smart Agriculture" OR "Precision Agriculture") AND ("AI" OR "Artificial Intelligence" OR "Machine Learning" OR "ML" OR "Computer Vision" OR "Reasoning") AND ("Crop" OR "Yield" AND (Prediction" OR "Discrimination"))

### 3.1.3 Critérios e procedimentos de seleção dos artigos

Os critérios de seleção foram elaborados para retirar estudos não relevantes ao objetivo da pesquisa. Os artigos para o estudo foram selecionados de acordo com os critérios de inclusão e exclusão definidos para este artigo (B. KITCHENHAM P. BRERETON, 2007).

### 3.1.4 Critérios de exclusão

Os critérios de exclusão dos artigos utilizados foram:

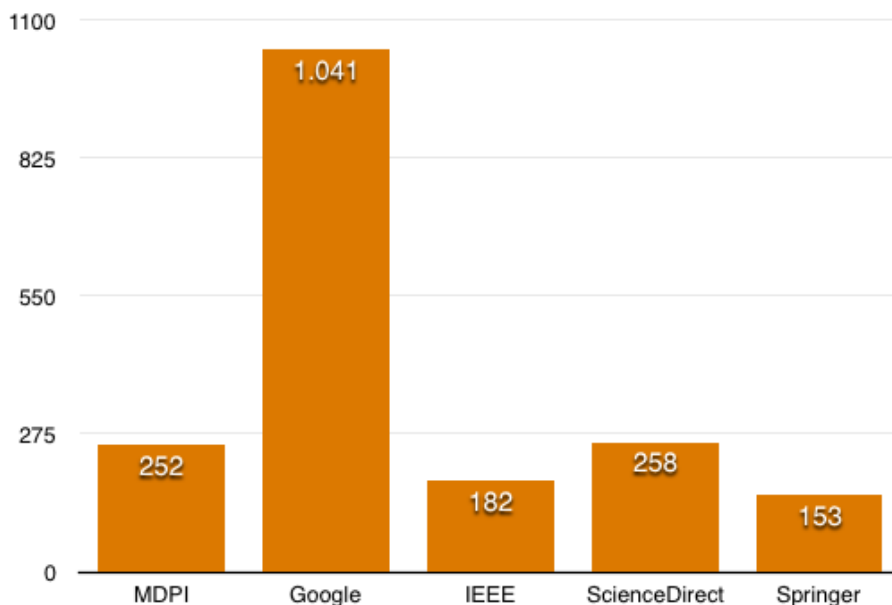
- E1 - Estudos como dissertações, teses, livros e capítulos de livros, tutoriais, relatórios;
- E2 - Artigos que apresentavam apenas o resumo, em outros idiomas, ou que não estavam disponíveis;
- E3 - Artigos que apresentaram formas de estimativa de rendimento que não fossem através da inteligência artificial;
- E4 - Artigos duplicados, publicados em duas fontes;
- E5 - Artigos que possuem resumo e conclusão não condizentes com os objetivos da pesquisa;
- E6 - Artigos que divergem dos o tema pesquisado;
- E7 - Artigos que são resenhas de publicações, independentemente da natureza;

### 3.1.5 Avaliação da qualidade

Os critérios de avaliação da qualidade dos artigos foram definidos de acordo com (B. KITCHENHAM P. BRERETON, 2009). Três questões foram elaboradas para realizar a avaliação dos artigos, com base no interesse da pesquisa.

- QA1: O artigo propõe ou apresenta alguma solução de inteligência artificial desenvolvida para estimativa das safras? Baixo encontram-se várias vezes a falta de concordância de número.
- QA2: O artigo usa modelos de aprendizado de máquina na estimativa das safras?

Figura 11 – Resultado das buscas



Fonte: Adaptado pelo autor

- QA3: O artigo avaliado demonstra que o uso de inteligência artificial na estimativa das safras está contribuindo para o aumento da produção e/ou preservação do meio ambiente?

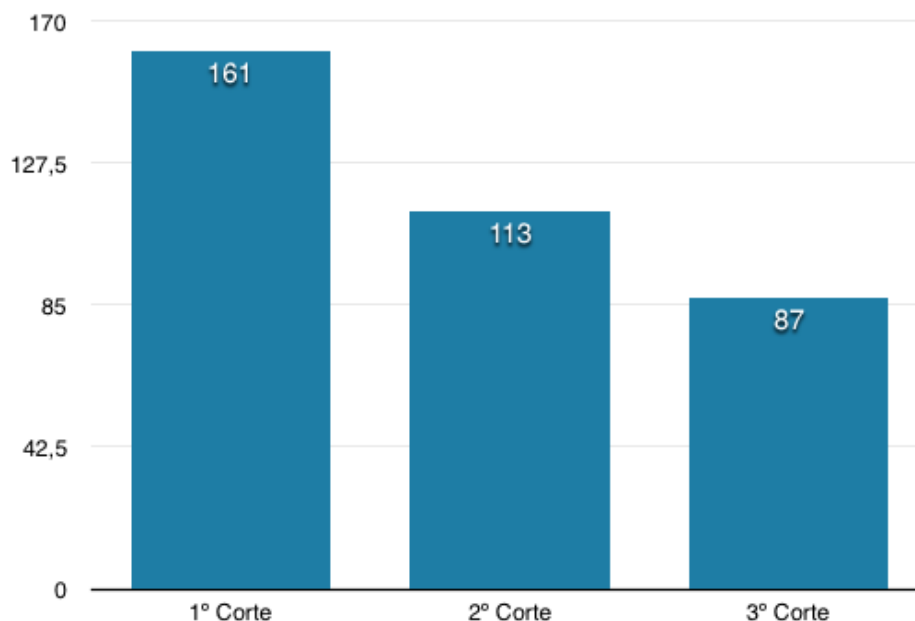
Ainda baseado em na metodologia de (B. KITCHENHAM P. BRERETON, 2009), foram definidas três respostas possíveis para cada questão, sendo: Sim = 1; Em partes = 0,5 e Não = 0. Após a aplicação dos critérios de avaliação de qualidade e análise da pontuação, decidiu-se que os artigos classificados com pontuação inferior a 2 seriam removidos.

### 3.2 Resultados

Nesta seção são apresentados os resultados do processo de pesquisa, seleção e avaliação da qualidade. Os resultados da busca nas fontes selecionadas utilizando a *string* de busca, o que resultou em um total de 1.886 trabalhos, conforme mostrado na figura 11.

Após o resultado de busca por trabalhos, que resultou em um total 1.886, conforme mostrado na figura 13, procedeu-se ao corte por etapas, utilizando os critérios mencionados na subseção 3.1.1.

Figura 12 – Resultado após a etapa de cortes



Fonte: Adaptado pelo autor

### 3.2.1 Exclusão de trabalhos da pesquisa inicial

A aplicação dos cortes, se deram por conta de buscarmos atender as questões de pesquisas, conforme a subseção 3.1.1 e em seguida foi aplicado os filtros de exclusão conforme os critérios mencionados no item 3.1.4. Esses cortes foram realizados em 3 etapas, de acordo com o que é apresentado na figura 12.

Na primeira etapa de corte, foram utilizados os seguintes critérios: atender as questões de pesquisa, retirar os estudos como teses, livros, tutoriais, relatórios, revisões, pesquisas e outros demais critérios como os trabalhos que apresentavam forma de previsão de rendimento que não fossem através da inteligência artificial, trabalhos duplicados, resumos que divergiam com o tema pesquisado ou resenhas e apenas trabalhos que estivessem no formato PDF. Mas sempre buscando atender os critérios de exclusão e as questões de pesquisa.

Na segunda etapa dos cortes, os critérios adotados foram os idiomas, onde foi mantido apenas o português e o inglês. Continuando, trabalhos que não foram identificados pelo primeiro corte mas que se encaixavam no critério de exclusão, resumos não consistentes e trabalhos sem demonstração efetiva de resultados.

Sendo assim, foi realizada uma nova checagem buscando identificar mais alguns trabalhos, que passaram despercebidos pelas etapas anteriores. Onde atingiu-se o número de 87 trabalhos, conforme é mostrado na Figura 13.

Por fim, agora com a relação de trabalhos que passariam pela avaliação de qualidade, estes seriam os artigos considerados relevantes para a revisão e avanço deste estudo. Os artigos

que foram cortados enquadram-se nos critérios de exclusão, o que não significa que não sejam interessantes, mas não contemplaram os indicadores necessários para esta pesquisa.

### 3.2.2 Avaliação da qualidade dos artigos relevantes

Nesta seção, está apresentado o fator de pontuação bem como a avaliação dos trabalhos selecionados para passar nos critérios de avaliação da qualidade definidos na subseção 3.1.5. Os resultados das avaliações dos 87 artigos estão apresentados na Tabela 3. Conforme os critérios de definição de qualidade, foi optado pela escolha dos trabalhos com pontuação entre 2 e 3.

A revisão sistemática mostrou que as tecnologias de inteligência artificial estão fazendo um trabalho muito preciso na estimativa da safras, gerando aumento na produção e qualidade dos alimentos, preservando o meio ambiente (G. BANNERJEE U. SARKAR, 2018). Mesmo com todas as dificuldades que ainda existem, a pesquisa mostrou o progresso e os benefícios do uso de tecnologias inteligentes na agricultura.

A Tabela 3 mostra o fator de pontuação que foi criado para ser atribuída a cada questão de avaliação de qualidade da seção 3.1.5.

Tabela 3 – Pontuação utilizada para avaliar a qualidade dos artigos

Pergunta	Descrição	Resultado
Sim	respondeu a pergunta	1
Em partes	pergunta parcialmente respondida	0,5
Não	não é mencionado no artigo	0

Ao aplicar avaliação de qualidade em 87 artigos, destes, 57 referências foram retiradas por não atenderem ao *score* estabelecido para corte, embora sejam interessantes. Os 30 artigos restantes foram considerados mais relevantes para este trabalho, conforme estão apresentados na Tabela 4.

- QA1: O artigo propõe ou apresenta alguma solução de inteligência artificial desenvolvida para estimativa da safras?
- QA2: O artigo usa modelos de aprendizado de máquina na estimativa da safras?
- QA3: O artigo avaliado demonstra que o uso de inteligência artificial na estimativa da safras está contribuindo para o aumento da produção e / ou preservação do meio ambiente?

Tabela 4 – Trabalhos mais relevantes

Autores	Ano	QA1	QA2	QA3	Total
O.M Adisa et al.	2019	1	1	0,5	2,5
C.C. Junior et al.	2019	1	1	0	2
L.E.O.Aparecido et al.	2019	0,5	0,5	1	2
A. Rogachev et al.	2019	1	1	1	3
M. Mainaitijiang et al.	2018	1	0,5	0,5	2
A. S. Terliksiz et al.	2018	1	1	1	3
J. Cmobarac et al.	2009	1	1	0,5	2,5
R Medar et al.	2019	1	1	1	3
H.S. Abdullahi et al.	2019	1	1	1	3
D Gómez et al.	2019	1	1	1	3
SS Lee et al.	2019	1	1	1	3
VM Purushan et al.	2020	1	1	1	3
M Champaneri et al.	2020	1	1	1	3
FC Eugenio et al.	2020	1	1	1	3
FF Haque et al.	2020	1	1	1	3
Gandhi et al.	2017	1	1	1	3
J. Zhou et al.	2019	1	1	0.5	2.5
L. Kouadio et al.	2018	1	1	0.5	2.5
M. Piekutowska et al.	2018	1	1	1	3
C.Zhang et al.	2019	1	1	1	3
J.G.N. Zannou et al.	2019	1	1	1	3
Nevavuori et al.	2019	1	1	1	3
Shendryk et al.	2021	1	1	1	3
N Efremova et al.	2019	1	1	0.5	2.5
T. Braddock et al.	2019	1	1	0.5	2.5
AÉ Nyéki et al.	2019	1	1	1	3
S. Khaki et al.	2020	1	1	0.5	2.5
O Marko et al.	2016	1	1	0.5	2.5
S Khanal et al.	2018	1	1	0.5	2.5
A Kross et al.	2018	1	0.5	1	2.5

### 3.3 Resultados das questões e discussão

Nesta seção haverá a análise dos trabalhos qualificados, explorando as contribuições de cada estudo. Cada uma das três questões foi transformada em uma subseção. Portanto, tem-se a subseção 3.3.1 que tratará de quais tecnologias computacionais foram usadas para estimativa de safras. Já a subseção 3.3.2 tratará de como a inteligência artificial esta sendo usada para previsão de safras. Por fim a subseção 3.3.3 que traz como o aprendizado de máquina e visão computacional tem sido usados para a estimativa de safras.

A tecnologia computacional está presente em toda a cadeia produtiva do agronegócio, com o objetivo de trazer eficiência e eficácia, redução de custos e aumento da produtividade. Onde esta ganhando cada vez mais espaço e ainda mais repercussão em tempos de crise econômica. Algumas tecnologias tem se destacado no segmento, como o uso de imagens de satélite, aere-

ves não tripuladas, sensores, robôs, inteligência artificial, aprendizado de máquina, e uma série de softwares para monitorar o desenvolvimento da cultura.

Alguns exemplos de usos dessas tecnologias são, o uso de aeronaves não tripuladas para fotografar, filmar e até mesmo ser utilizado na aplicação de agrotóxicos, o uso de imagens de satélite para avaliar e monitorar o desenvolvimento de uma cultura, em casos mais avançados aplicar alguns filtros capazes de identificar falhas na plantação, excesso de água e até possíveis pragas. O uso da inteligência artificial vem crescendo a cada dia, possibilitando uma melhor leitura e identificação de padrões em diversas fontes de dados que envolvem a cadeia do agro-negócio.

### 3.3.1 Quais tecnologias computacionais foram usadas para previsão de produção?

No estudo elaborado por ADISA (2019), vem afirmar que os métodos de previsão de rendimento relatados na literatura incluem regressão, simulação, sistemas especialistas e rede neural artificial (RNA), cujo o qual em sua análise optou por trabalhar com algoritmos de aprendizado supervisionado, uma vez que compreendem uma função flexível que treina o perceptron multicamada para um conjunto de dados específico. Para este estudo foi possível conseguir um conjunto de dados de quase 30 anos, onde utilizaram 80% dos dados para treinamento e 20% para o restante dos testes.

Feito isso, (ADISA, 2019) realizou testes com diferentes combinações de dados e alcançou o seu melhor resultado, aproximadamente 86% de precisão, com a combinação dos dados de índices de chuva, temperatura máxima, temperatura mínima, área de produção (fazenda) e terra cultivada. É possível presumir que com a adição de fontes de dados, que envolvam as informações de intervenção ao controle de pragas com o uso de defensivos e dados sobre a características do solo, poderiam ter alcançado melhores resultados.

Já no trabalho elaborado por JUNIOR et al. (2019), que buscou em estimar as áreas de soja e milho no Estado do Paraná. Apesar de ter tido dificuldades por conta da assinatura espectral ao diferenciar uma cultura da outra, a combinação da descoberta de padrões de dados por meio da mineração de dados a rede neural artificial, conseguiu alcançar o indicador de 0,78 no índice Kappa e 89% de precisão geral. Conseguindo um bom indicador de acurácia para a separação das culturas.

O estudo realizado por KROSS et al. (2018), demonstrou o potencial do modelo de rede neural artificial do *Advangeo®Prediction Software* para a cultura de milho e previsões de rendimento de soja no Canadá. Os resultados indicaram que as imagens de satélite podem ser utilizadas para prever o rendimento, desde que os modelos sejam criados para tipos de cultura exclusivos. Onde a cultura do milho atingiu um melhor desempenho que o modelo da soja, apresentando os seguintes resultados para cada cultura: todos os imóveis com a cultura de milho tiveram *the relative mean absolute errors (RMAEs)*, inferiores a 10% (2012) ou 20% (2011). Já na previsão de rendimento da rede neural da cultura da soja entre 42% (2011) e 100% (2012)

dos locais de teste de soja tiveram *RMAEs* inferiores a 20%. O autor ainda relata acreditar que obtendo informações índices de chuvas podem melhorar a performance das predições.

Enquanto em GÓMEZ et al. (2019), em sua pesquisa desenvolveu um sistema de predição com uso de imagens de satélite *Sentinel-2* e criou diversos modelos de aprendizado de máquina com o intuito de elaborar inúmeros testes de pre-processamento para verificar quais obtivessem os melhores resultados de predição. Os modelos que apresentaram os melhores resultados foram o *regresso quantile lasso*, com a raiz do erro quadrático médio de 9,18% e o modelo *leap backwards* com a raiz do erro quadrático médio de 10,94%. Enquanto o *SVM* conseguiu alcançar 11,7% e o *random forest* 11,16%. Demonstrando a dinâmica de seu modelo em prever a produção de batatas.

Na pesquisa de OLIVEIRA APARECIDO et al. (2019) foram desenvolvidos algoritmos de predição através do paradigma de redes neurais artificiais do tipo *multilayer perceptron (MLP)*, propondo estabelecer uma zona agrícola com riscos climáticos (ZARC), para a cultura de Girasol. Neste cenário identificou na região sudeste do Brasil, áreas que apresentavam os melhores índices climáticos para a plantação dessa cultura. Mas ao meu entender, não ficou claro sua opção por esta região.

No entanto, em sua pesquisa identificou locais em potencial de acordo com o clima, alcançando resultados muitos promissores e chegando a identificar 64% de áreas propícias para o cultivo no decorrer da primavera e do verão nessa região. Somente neste período o clima apresenta condições ideais para o cultivo. Mas acredito que inclusão de dados como relevo, possa reduzir esse índice de áreas propícias.

Os estudos de LEE et al. (2019), propõem uma plataforma auto previsível de rendimento baseada em doenças. Por meio do uso do aprendizado de máquina com as fontes de dados de: temperatura, umidade, sol, precipitação, temperatura do solo. Criando assim, previsões de culturas com base em fatores como mudanças climáticas e doenças. Por meio de uma rede neural convolucional. Para avaliar seus resultados, fez uso de alguns experimentos, onde conseguiu alcançar a previsão de cerca de 34% mais precisa ao usar várias doenças do que quando não as usa. Portanto, LEE et al. (2019) articulam que o seu modelo pode prever os rendimentos agrícolas com mais precisão do que os métodos tradicionais.

Já no trabalho realizado por ROGACHEV (2019) buscou na combinação da visão computacional em conjunto com o algoritmo de LSTM, classificar áreas de campos com desenvolvimento uniforme de plantas ou áreas com defeitos (ervas daninhas, semeadura irregular, danos às plantas por pragas e doenças) e identificar a heterogeneidade das culturas. Assim, obteve inicialmente resultados muito promissores, embora não dispusesse de um grande acervo para testes. Nesse estudo foi utilizado apenas Veículos Aéreos Não Tripulados (VANT's) para a captura das imagens aéreas, pois segundo o ROGACHEV (2019) o uso de imagens de satélites podem fornecer imagens distorcidas, uma vez que podem ficar descalibradas por conta das condições atmosféricas.

Embora eu acredite que os VANT's possam fornecer imagens mais precisas e autonomia

para escolha do dia para a captura, é uma opção que pode ser muito mais cara, levando em conta os tipos de sensores que possam vir a ser necessários e o outro fator seria a cobertura de diversos imóveis rurais. Eu não desconsideraria o uso das imagens de satélites combinadas com várias outras fontes de dados.

Os estudos de MAITINIYAZI MAIMAITIJIANGA VASIT SAGANA (2019) comprovam diante de sua pesquisa que a fusão de avançados sensores a bordo de um veículo aéreo não tripulado (VANT), podem impulsionar o auxílio à tomadas de decisão no decorrer da evolução da cultura. O uso deste recurso combinado ao algoritmo de aprendizado profundo, obteve resultados significativos na previsão de rendimento e adaptabilidade para prever o rendimento de três diferentes genótipos de soja. Sedo assim é aceitável admitir que se possa obter resultados ainda mais positivos, caso inclua alguns dados literais em seu conjunto, tais como: clima, índices de chuva e temperatura ao seu modelo.

Enquanto no trabalho de MARKO et al. (2016), teve como objetivo desenhar a estratégia de seleção de variedades de soja que devem ser plantadas na fazenda teste para maximizar a produtividade na safra seguinte, com base no conhecimento adquirido a partir de dados históricos heterogêneos. Através do algoritmo de KNN, avançou no processo de identificação. Onde foi proposto a regressão de histogramas ponderados com o intuito de prever o rendimento de diferentes variedades. Com base no rendimento previsto, realizaram otimizações no portfólio para obter a seleção ideal das variedades de sementes a serem plantadas.

Na pesquisa de TERLIKSIZ; ALTILAR (2019), foram trabalhados com os recursos de *support vector machine (SVM)*, imagens de satélite, aprendizado profundo e os indicadores de produtividade soja dos anos de 2003 à 2016, chegaram à de precisão de 90%, através de seu algoritmo de aprendizado profundo. Esta contribuição poderia ter alcançado melhores resultados se explorassem uma maior variedade de fontes de dados, como dados climáticos e índices pluviométricos.

Nas pesquisas de NYÉKI et al. (2019), optou por trabalhar com imagens hiper-espectrais geradas a partir de uma aeronave não tripulada. Onde apresentou um o modelo de previsão de produtividade de milho de um imóvel específico por cinco anos e foram avaliados diversos algoritmos de aprendizado de máquina, como: *Counter propagation Artificial Neural networks (CP-ANN)*, *XY-fused networks (XY-F)*, *supervised Kohonen networks (SKN)*, *extreme gradient boosting (XGBoost)* e *support-vector machine (SVM)*. Com o objetivo de definir os modelos de aprendizado de máquina entre o rendimento do milho e os fatores que influenciam o rendimento.

Ainda na pesquisa de (NYÉKI et al., 2019), foram utilizados outros conjuntos de dados como os de parâmetros relacionados à solos, meteorológicos, e os parâmetros da cultura de milho. Feito isso, os resultados apresentaram que o melhor método de previsão foi o *XGBoost*. Uma vez, que a acurácia da predição atingiu no conjunto de testes: 95,38% de acurácia, 91,3% de sensibilidade e 97,62% de especificidade. Diante disso, torna o algoritmo *XGBoost* o mais indicado nessa combinação de análise de dados para a previsão de produtividade da cultura de milho.



CRNOBARAC (2009), vem afirmar em decorrência de suas observações que a mineração de dados oferece possibilidades de transformar dados brutos em informações valiosas e que podem ser utilizadas para tomar melhores decisões. E em sua pesquisa, utilizou como fonte de dados os parâmetros meteorológicos básicos no período de vegetação, temperatura máxima, mínima e média umidade do ar, além do nível de precipitação.

Ainda nesta pesquisa foi utilizado o software WEKA, cujo qual, possui uma infinidade de algoritmos de aprendizado e ferramentas de pré-processamento. Para este estudo CRNOBARAC (2009) optou pelo algoritmo M5P, por achar o mais adequado para o conjunto de dados que foi utilizado. Pois o algoritmo M5P é uma combinação de classificação e regressão de dados. Segue o modelo dos métodos da árvore de decisão, mas nas folhas ele possui funções de regressão linear em vez de rótulos de classe. Alcançando resultados significativos em suas regras de regressão.

É plausível deduzir que caso houvesse em sua fonte de dados outros dados como pressão atmosférica, relevo, características de solo e média de produção das safras anteriores. Pois com estes outros dados poderia fechar todos os parâmetros necessários para uma melhor análise de todo o cenário que envolve as culturas.

ABDULLAHI (2017) desenvolveu em sua pesquisa um modelo que pode ser utilizado em diferentes regiões com composição do solo, clima e outros fatores separados de uma região geográfica com o intuito de identificar doenças na cultura de milho. Através do recurso de aprendizado profundo combinado com recursos de visão computacional, chegou ao resultado de 99,58% na predição da saúde da cultura. Foi usado um total de cerca de 1918 fotos e foi aumentado para 4588 imagens para obter melhores resultados e poder realizar os treinamentos necessários. Neste estudo poderiam ter realizado testes em diferentes culturas com o intuito de mapear se uma planta está saudável ou não.

Enquanto no trabalho de ZHOU (2019) que teve como pilares de sua pesquisa a investigação de um método para estimar as datas de maturidade das linhagens de soja usando um sistema de imagem baseado em VANT. Com isso, foi elencado alguns marcos de sua pesquisa como: investigar o melhor momento (estágio de crescimento) para coletar imagens, selecionar recursos de imagem importantes para estimar a data de vencimento relativa e avaliar a precisão da estimativa. Através do software MATLAB e do modelo PLSR (regressão parcial de mínimos quadrados), alcançou resultados superior aos outros trabalhos, porém semelhante em termo de precisão de previsão. Eu acredito que poderia ainda ter resultados melhores se tivessem feito uma combinação com algumas fontes de dados, como: clima, temperatura, índices pluviométricos, pressão e relevo.

### 3.3.2 Como a inteligência artificial tem sido usada para a e previsão de produção?

O uso de inteligência artificial na agricultura inteligente está crescendo a cada dia e sua aplicação ocorre em todos os campos da agricultura. A necessidade de controles e previsões

de safra aumentam a cada dia, uma vez que há uma necessidade sempre crescente de mercado, mas que seja possível de produzir mais sem aumentar as áreas já plantadas.

A previsão de safras vêm ganhando atenção especial, pois reduzem riscos para os agricultores e auxiliam o Governo numa questão de abastecimento, uma vez que esse acompanhamento que pode vir a garantir a balança comercial do país. A inteligência artificial também tem sido utilizada como um recurso que auxilia na identificação de padrões, sejam eles no reconhecimento de tipos de pragas, na identificação de falhas nos recursos minerais terrestres ou até mesmo em notas de correção em campo por meio de processamento de imagens.

No trabalho realizado por VM PURUSHAN (2020), que desenvolveu um algoritmo de regressão linear para prever o rendimento das culturas de batata e arroz. Por meio do uso da mineração de dados, ele buscou identificar padrões nos dados que foram gerados através de sensores em campo, cujo os quais, deveriam extrair os seguintes dados: Umidade do solo, temperatura, umidade do ar e índices de chuvas. Com isso, o autor supracitado em face de suas pesquisas obteve um rendimento medido através da métrica raiz do erro quadrático médio de 0,0748, que é menos de 10 por cento da média do previsto. Já com todos os valores do conjunto de dados de treinamento atingiu 1,0045. Enquanto, o valor R-quadrado foi de 0,994, o que torna este algoritmo segundo (VM PURUSHAN, 2020) preciso para fazer uma boa previsão para a cultura do arroz.

Os pesquisadores GANDHI; ARMSTRONG; NANDAWADEKAR (2017), desenvolveram uma rede neural para prever o rendimento da colheita de arroz em regiões da Índia. Através das fontes de precipitação, temperatura mínima, temperatura média, temperatura máxima e evapotranspiração da cultura de referência, área, produção e rendimento para a estação *Kharif* (junho a novembro) para os anos de 1998 a 2002. Por meio da ferramenta WEKA, a rede neural perceptron multicamadas foi desenvolvida. Com o método de validação cruzada, o resultado alcançou 97,5% de acurácia e sensibilidade de 96,3 e especificidade de 98,1.

Em outro estudo, o autor C ZHANG L DI (2019) utilizou inteligência artificial para identificar padrões baseados em dados de nutrientes das folhas de soja, em três estágios da colheita, visando identificar uma possível necessidade de correção nos padrões encontrados, auxiliando na indecisão do agricultor e obtendo maior produtividade da cultura.

Utilizando outras fontes de dados, o autor L KOUADIO RC DEO (2018) utilizou um algoritmo de aprendizado de máquina para analisar as propriedades de fertilidade do solo, obtendo resultados altamente precisos. Por meio do método de análise de mapas históricos de plantio, utilizando um algoritmo de aprendizado de máquina.

Já o autor R MEDAR (2019) chegou ao resultado de 88% na necessidade de concordância com o futuro da colheita. A estimativa de safra, por meio de algoritmos de aprendizado de máquina, também foi a opção de estudo do autor JGN ZANNOU (2019), onde através do TensorFlow e através de análise de imagens atingiu uma acurácia média de 74,5% para a detecção e 99% para estimativa do peso de sorgo.

### 3.3.3 Como o aprendizado de máquina e a visão computacional têm sido usados para a previsão de rendimento?

Em uma fazenda inteligente, é comum o uso de diversos tipos de algoritmos de aprendizado de máquina e mais recentemente os algoritmos de aprendizado profundo, portanto, esta subseção será focada nos recursos de aprendizado de máquina mais utilizados em conjunto com a visão computacional. A visão computacional tem se mostrado muito eficaz na área de reconhecimento, cujo objetivo é o desenvolvimento e aplicação de algoritmos de detecção, identificação e classificação de objetos em uma imagem.

A pesquisa de HAGHVERDI; WASHINGTON-ALLEN; LEIB (2018) propôs métodos baseados em aprendizado profundo para identificar regiões de interesse em imagens de algodão (características morfológicas). Uma metodologia foi desenvolvida usando uma rede neural convolucional para detectar e contar o número de flores de algodão recém-abertas em imagens aéreas. Foi desenvolvido um algoritmo de identificação e contagem de cápsulas para estimar a produtividade do algodão, a partir de imagens aéreas adquiridas durante a safra em campo experimental.

Quanto ao uso dos paradigmas de aprendizado de máquina, por serem um processo de indução de hipóteses de experiência e também por serem utilizados para tarefas de aprendizagem preditivas ou descritivas, eles se mostraram adequados para a predição da safras. Na predição da safras, na agricultura inteligente, várias técnicas de aprendizado de máquina estão sendo usadas, mas a aplicação não está apenas focada na agricultura.

M CHAMPANERI D CHACHPARA (2020), fez uso da mineração de dados, visando identificar qual o ambiente mais propício para prever o rendimento na agricultura. Tendo em seu conjunto de dados os parâmetros climáticos. No entanto, obteve 75% precisão em suas previsões para todos os tipos de cultura, por meio do algoritmo de aprendizado de máquina *random forest*.

Ainda segundo M CHAMPANERI D CHACHPARA (2020) o *random forest* é o algoritmo mais popular e poderoso. Sendo capaz de executar tarefas de classificação e regressão, que operam construindo uma infinidade de árvores de decisão durante o tempo de treinamento e gerando saída da classe que é o modo do classes (classificação) ou previsão média (regressão) das árvores individuais.

FC EUGENIO M GROHS (2020), buscou trabalhar sobre a previsão de rendimento da cultura de soja em áreas que fazem uso de irrigação. Para isso, fez uso de aeronaves não tripuladas para garantir imagens espectrais com boa resolução e buscou no aprendizado de máquina para fazer o reconhecimento dos padrões nas imagens, através do algoritmo de multicamadas de perceptrões. Por sua vez, a produtividade de grãos foi o parâmetro agrônomico mais complexo para modelagem, com correlações de 0,70 e 0,92 para treinamento/teste, o que se pode considerar como excelentes resultados, reforçando a grande capacidade de combinação entre sensoriamento remoto via veículos aéreos não tripulados e aprendizado de máquina (ML) em

aplicações voltadas para a agricultura de precisão.

A predição de safra tem sido usada por meio de algumas técnicas, tais como: identificação de padrões em uma coleção de dados históricos sobre a produção agrícola em conjunto com dados climáticos e de relevo, e em muitos levantamentos, o processamento de imagem aparece em conjunto com o uso do paradigma da visão computacional. O uso de imagens, seja por meio de aeronaves não tripuladas ou imagens de satélite, tem obtido excelentes resultados de acerto, o que geralmente leva à opção de escolha da fonte, sejam os custos e o fato de haver um histórico de imagens das fases da colheita.

Já em KHANAL et al. (2018), teve como os objetivos do seu estudo: (i) avaliar o papel das imagens de sensoriamento remoto; (ii) comparar o desempenho de vários algoritmos de aprendizado de máquina; e (iii) identificar a importância de variáveis derivadas de imagens de sensoriamento remoto, na predição espacial de propriedades do solo e produtividade de milho.

Ainda em KHANAL et al. (2018), foi utilizado um modelo de dados digital de elevação, com resolução de 1m, que foram utilizados para derivar as características topográficas das lavouras. Imagens multi-espectrais coletadas em condições de solo descoberto, com resolução de 0,30m, foram usadas para derivar os índices de solo e vegetação.

Sendo assim foram desenvolvidos os modelos para previsão de propriedades do solo e rendimento de milho usando regressão linear (LM) e cinco algoritmos de aprendizado de máquina, que foram: *random forest (RF)*; *neural network (NN)*; *support vector machine (SVM)* com as funções de *kernel* radial e linear; *Gradient Boosting Model (GBM)* e *Cubist (CU)* foram avaliados em termos de coeficiente de determinação ( $R^2$ ) e erro quadrático médio (RMSE).

Por fim, KHANAL et al. (2018), identificaram por meio de suas pesquisas que os algoritmos de aprendizado de máquina superam o algoritmo de regressão linear na maioria das vezes e que os índices do solo e vegetação baseados em imagens de solo descoberto desempenharam um papel mais significativo na demonstração da variabilidade em campo da produtividade de milho e propriedades do solo do que as variáveis topográficas. Por conseguinte, a acurácia dos modelos desenvolvidos para predição de propriedades do solo e produtividade de milho observada neste estudo sugere que a abordagem de integração de dados de sensoriamento remoto e algoritmos de aprendizado de máquina é promissora para mapear propriedades do solo e produtividade de milho em escala local, o que pode ser útil na localizar áreas de preocupações potenciais e implementar práticas agrícolas específicas do local.

Conforme aborda R MEDAR (2019), que diante de seus estudos compôs o módulo principal de sua solução, os algoritmos Naive Bayes e o KNN (K - Vizinho Mais Próximo / K - Vizinho Mais Próximo), este para processamento para treinamento dos conjuntos de dados, enquanto Ingênuo Bayes foi usado para a classificação dos conjuntos de dados. Esses algoritmos têm sido amplamente utilizados perante a possibilidade de não exigirem uma grande quantidade de dados para os testes.

Em BRADDOCK et al. (2019), aborda sobre o uso do sistema de visão de máquina que faz uso por meio de uma câmera colorida para adquirir imagens das árvores durante o período

de floração. Um algoritmo de segmentação de imagens faz então o reconhecimento e conta as flores na árvore. Esse algoritmo de segmentação de rede neural superficial usa informações de cor e posição como entrada. A alta correlação resultante entre a contagem de flores e o número de frutos na árvore mostra o potencial deste método.

O sistema proposto por NEVAVUORI; NARRA; LIPPING (2019), utilizou aeronaves não tripuladas e câmeras espectrais como recursos para aquisição de imagens com alta resolução e em conjuntos as redes neurais convolucionais (CNNs), construiu um modelo para previsão de rendimento das culturas de trigo e cevada com base em dados NDVI e RGB. Conseguindo alcançar bons resultados, chegando ao erro percentual absoluto médio (MAPE) de 8,8% durante o período inicial da fase de crescimento < 25% e Ao utilizar dados adquiridos posterior a fase de crescimento > 25%, obteve-se MAPE: 12,6%.

Como objetivo de suas pesquisas EFREMOVA; ZAUSAEV; ANTIPOV (2019), buscaram melhorar a eficiência do uso da água, reduzir a dependência da irrigação, aumentar o rendimento geral das culturas e minimizar o risco de perda de safra devido à seca e condições climáticas extremas. Por meio das imagens de satélite, realizou a segmentação das culturas, classificação de solo e NDVI e previsão de umidade do solo, dados de solo e registros de dados climáticos.

EFREMOVA; ZAUSAEV; ANTIPOV (2019) também afirmam que, ao aplicar o aprendizado de máquina a dados de sensores e dados de solo, os sistemas de gerenciamento de fazendas podem evoluir para uma plataforma habilitada para inteligência artificial em tempo real que pode fornecer recomendações acionáveis e ferramentas de suporte a decisões para os agricultores.

Demonstrou ainda que o tempo como um recurso adicional não apenas melhora o desempenho da modelagem com dados de aeronaves não tripuladas, mas também melhora as capacidades preditivas, seu modelo agrícola foi avaliado por meio de algoritmos supervisionados, como regressão linear (RL) e *support vector machine (SVM)*.

No estudo elaborado por FF HAQUE (2020), encontram-se proposta a validação de modelos agrícolas que foram avaliados com algoritmos supervisionados, como Regressão Linear (LR) e Regressão Vetorial de Suporte (SVR). A estimativa por meio da regressão ajudou a obter um valor contínuo, bem além do discreto valor obtido pelo outro algoritmo. O modelo criado com a ajuda do algoritmo supervisionado mostrou alguns resultados promissores com os quais a rede neural foi iniciada, cujo o modelo foi avaliado de forma semelhante com o mesmo conjunto de dados. Foi observado que os modelos não supervisionados com a ajuda de entradas ponderadas adicionadas e percorrendo cada valor durante a diferença na veracidade da saída estão apresentando melhores resultados.

O aprendizado profundo mostrou resultados promissores na classificação das imagens, sendo o ImageNet uma função que auxilia na classificação das informações fornecidas para as saídas necessárias. Segundo FF HAQUE (2020), esses modelos tendem a ter repugnância e erros de ajuste que são trazidos do projeto do modelo. Uma vez que se tenta manter o aprendizado igual, a menos que as informações sejam alteradas manualmente todas as vezes e, por outro lado, o

aprendizado não supervisionado tende a fazer com que todo o processo aprenda à medida que as informações variam da saída concluída.

Ainda em FF HAQUE (2020), discorre-se que o modelo treinado que foi usado para um conjunto de dados com um parâmetro significativo pode ser usado em outro conjunto de dados que pode diferir na semelhança de parâmetros, mas o resultado é o mesmo se houver conhecimento sobre os atributos presentes.

Por fim, de acordo com FF HAQUE (2020), o conjunto de dados que foi obtido por meio de 140 pontos de dados, que descrevem o efeito dos atributos no rendimento das culturas. Obteve uma taxa que foi apresentada com a ajuda do erro quadrático médio (MSE) e o desvio padrão entre os resultados do rendimento com o real também foi mostrado, que resultou em 0,0045 para o MSE, que é cerca de 0,000345 como o desvio padrão. O que demonstram ótimos resultados.

Por meio dos estudos de JGN ZANNOU (2019), foram obtidos os resultados de 74% de acerto na identificação da quantidade de sorgo com o uso de aprendizado de máquina e 99% no peso médio produzido. Onde é importante notar que quanto maior o volume de dados para teste e aprendizagem, mais preciso o algoritmo se torna.

A pesquisa realizada por KHAKI; PHAM; WANG (2020), apresentou resultados numéricos cujo qual, ilustram que uma abordagem capaz de prever simultaneamente a produtividade de milho e soja seja possível e pode alcançar maior precisão do que os modelos individuais.

Ao utilizar o aprendizado de transferência entre milho e rendimento de soja para compartilhar os pesos do extrator de recursos de *backbone*, o *YieldNet* foi capaz de diminuir substancialmente o número de parâmetros de aprendizagem. Onde nessa abordagem de aprendizagem por transferência conseguiu permitir economizar recursos de computação enquanto também maximizava a precisão de previsão.

Além dessa possibilidade, a precisão alcançada usando uma antecipação de quatro meses tem muitas implicações importantes para as decisões que envolvem o manejo das culturas. Conforme KHAKI; PHAM; WANG (2020), à força da estrutura de aprendizado profundo que foi desenvolvida combinada com uma perda generalizada da função, tornou essa abordagem pronta para ser escala.

Portanto, KHAKI; PHAM; WANG (2020) consideram importante que para melhorar a precisão de sua metodologia, mais dados podem ser coletados, uma vez, que seu modelo seria capaz de fazer a predição de qualquer tipo de cultura.

Por fim, SHENDRYK; DAVY; THORBURN (2021) que separaram 80% de seu conjunto de dados para treinamento e os 20% restantes para testes, ele obteve o melhor resultado com o algoritmo de aumento do Gradiente de aprendizado de máquina, onde ele pode apresentar um resultado 4 meses antes do início da colheita. Também em suas pesquisas ele buscou direcionar a aplicação de fertilizantes, agendamento da colheita e até decisões que envolveram a equipe de marketing.

### 3.4 Conclusão

A inteligência artificial e seus paradigmas estão sendo usadas em todos os segmentos e a agricultura não poderia ficar de fora. A agricultura inteligente está crescendo a cada dia e a presente revisão sistemática diante de suas pesquisas exhibe os resultados de artigos científicos que tratam do uso de inteligência artificial na predição da produção de safra do milho.

A pesquisa tem demonstrado um vasto número de estudos e trabalhos existentes, mas com a aplicação dos filtros, restaram mais de 80 artigos para análise. Entre as ferramentas de inteligência artificial mais utilizadas, apontadas pela pesquisa, estão os aprendizado de máquina e redes neurais convolucionais.

O uso tem sido feito de diversas formas e diversas áreas de aplicação, desde: aplicação do produto, infestação de pragas, tratamentos inadequados do solo, irrigação, geração de informações para o agricultor e sugestão de soluções, entre outros. Na predição da safras, o agricultor pode planejar com sua previsão de colheita. Grande parte dos trabalhos analisados apresentam a imagem, seja ela capturada ou de um banco de dados, como fonte para a realização dos processos e análises para fornecer os resultados.

Os benefícios da utilização da inteligência artificial na agricultura são incalculáveis, já que a cada dia surgem novas aplicações e novas formas de utilização. Por meio da predição das safras, o agricultor tem como ferramenta de gestão a previsão da produtividade de sua safra, facilitando deste modo o planejamento da pós-colheita.

A avaliação de métodos de aprendizado de máquina para previsão de rendimento de produtividade baseada em sensoriamento remoto pleiteia a disponibilidade de recursos para mapeamento que ainda não são muito comuns entre os agricultores. Mas estudos que utilizam a CNN e nos quais usam dados de imagem de sensoriamento remoto como os satélites e Veículos Aéreos Não Tripulados (VANT's) para treinar seus modelos de previsão, vem ganhando força em todo mundo. Estes estudos vem tendo como objetivos a detecção de culturas e ervas daninhas, avaliação de biomassa e previsão de rendimento.

Embora a pesquisa de (NEVAVUORI; NARRA; LIPPING, 2019) e o trabalho de (EFREMOVA; ZAUSAEV; ANTIPOV, 2019), fazerem o uso das Redes Neurais Convolucionais (CNN), a comparação com o presente trabalho não faz sentido, uma vez que estas pesquisas dos colegas não tem relação direta com o presente trabalho. Já que a pesquisa de NEVAVUORI; NARRA; LIPPING (2019) optou por trabalhar com aeronaves não tripuladas e uso de câmeras espectrais, capazes de capturar imagens RGB/NIR, possibilitando o uso do Índice de Vegetação por Diferença Normalizada (NDVI) com o foco em fazer a predição da produtividade nas culturas de trigo e cevada. Enquanto na pesquisa de EFREMOVA; ZAUSAEV; ANTIPOV (2019), preferiu fazer o uso de imagens de satélite em conjunto com outras fontes de dados, mas com o foco em identificar o teor de umidade do solo, buscando assim trazer uma economia no uso dos recursos naturais.

Portanto, foi optado por não levar estes trabalhos adiante como experimentos no método

que será criado para o modelo de predição de safras de milho proposto neste presente estudo.

Vale ressaltar que dos mais de 80 artigos avaliados na íntegra, quase todos apresentam a utilização de uma ferramenta de análise de dados ou aprendizado de máquina, proporcionando cada vez mais a capacidade de aprender, treinar e tornar mais preciso e rápida uma tomada decisão. Mas uma coisa que chamou a atenção foram a quantidade de trabalhos que fazem uso da análise de dados por meio do processo de descoberta de conhecimento e uso de técnicas de mineração de dados. Buscando em conjunto com a rede neural trazer mais flexibilidade e facilidades nas decisões e proporcionando uma altíssima precisão nos processos.



Tabela 5 – Tecnologias empregadas nos trabalhos

Autores	RNA	D.minig	UAV	Satélites	Visão computacional	P. de imagens
Gandhi	X	X	-	-	-	-
Purushan	X	X	-	-	-	-
Mainaitijiang	X	-	X	-	-	-
Efremova	X	-	-	X	-	-
Medar	X	X	-	-	-	-
Shendryk	X	-	-	X	-	-
Braddock	X	-	-	-	X	-
Lee	X	-	-	-	-	X
Eugenio	X	-	X	-	-	-
Kross	X	-	-	X	-	-
Khanal	X	-	X	X	-	-
Nevavuori	X	-	X	-	-	-
Nyéki	X	X	-	-	-	-
Marko	X	X	-	-	-	-
Zannou	X	X	-	-	-	-
Adisa	X	X	-	-	-	-
Junior	X	X	-	-	-	-
Aparecido	X	-	-	-	-	-
Rogachev	X	-	X	-	-	-
Terliksiz	X	-	-	X	-	-
Cmobarac	X	-	-	-	-	-
Abdullahi	X	-	-	-	-	X
Gómez	X	-	-	X	-	-
Champaneri	X	X	-	-	-	-
Haque	X	-	-	X	-	-
Kouadio	X	X	-	-	-	-
Piekutowska	X	-	-	-	-	-
Zhou	X	-	X	-	-	-
Zhang	X	-	-	X	-	-

Fonte: Adaptado pelo autor



## 4 MODELO PARA ESTIMATIVA DE SAFRAS

Como já observado, a estimativa da safra corrobora para diversas situações positivas. Dessa forma, modelos de estimativa que possam ser integrados à rotina de analistas em operações de commodities, podem reduzir os riscos nas aquisições de grãos e o mais importante seria contribuir para a redução no desmatamento do bioma Amazônico. Em suma, a metodologia padrão realizada por um analista, detalhada na Seção 2.2.2, consiste na identificação, checagem de passivos ambientais, localização do imóvel e parecer. O objetivo desta pesquisa é adicionar uma etapa paralela à localização do imóvel, com o modelo para estimativa de safras servindo como sistema de apoio ao parecer do analista.

No entanto, mesmo para os sistemas de computador mais modernos, a análise de estimativa de safra é uma tarefa desafiadora no modelo de estimativa. Existem três razões principais:

1. **Acesso a imagens de alta resolução**, uma vez que, se torna um custo elevadíssimo a aquisição de imagens de satélite com alta resolução e também o fato de necessitar a inclusão do imóvel a ser monitorado, na rota do satélite. Outra forma de adquirir imagens de alta resolução seria fazendo o uso de aeronaves remotamente pilotadas, o que também se torna inviável para este estudo, sendo que não haveria cobertura de todos imóveis rurais;
2. **Diferentes critérios de filtros para a normalização**, já que pode depender da qualidade da imagem para ter uma resposta positiva e ao mesmo tempo deve se ater aos limites do alvo na imagem, ou seja, para o presente estudo deve extrair informações somente da plantação;
3. **Utilização de imagens de satélites *Sentinel-2***, embora tenha a disponibilidade gratuita do seu acervo, a qualidade das imagens se torna um desafio, visto que, sua resolução são apenas dez metros por pixel.

O modelo para estimativa de produtividade de safras busca mitigar estes desafios através de um pipeline robusto de aprendizado de máquina projetado especificamente para esta finalidade. Este capítulo visa expor as partes que compõem o modelo de estimativa de produtividade de safras, começando com uma visão geral e explicando cada uma das etapas da arquitetura do *Deep Yield Prediction (DYP)*.

### 4.1 Definições do projeto

Esta seção visa definir as decisões de design para o modelo de estimativa de produtividade de safras. A estratégia proposta é modelada para fornecer informações sobre a média de produção por hectare que influenciarão o parecer e para que o modelo possa ser facilmente adaptado à rotina do analista. Além disso, o modelo de estimativa de produtividade de safra será concebido

para se associar ao processo adotado pelo analista de *commodities* ao mapear os riscos numa identificação da unidade produtora (imóvel rural) ou para fins de fiscalização por meio de uma autoridade fiscalizadora.

Como a verificação da produtividade da unidade produtora não é um requisito no *checklist* do analista, no máximo ele terá uma estimativa de produtividade com base na produção regional, informação esta, que é fornecida pelo Instituto Mato-Grossense de Economia Agropecuária (IMEA). Esse analista reconhecendo uma unidade produtora suspeita de super produção, que nestes casos geralmente estão envolvidas com a lavagem de grãos, poderá impedir o aval da compra.

Com base nessas decisões iniciais, foi definido um pipeline de aprendizado de máquina. A primeira decisão neste ponto é em relação ao tamanho da amplificação da imagem do imóvel, que limitará a capacidade de hardware disponível para treinamento.

No modelo de estimativa de safras, foi proposto o uso de uma rede neural convolucional para a análise das imagens de satélite anterior ao início da colheita. Esta CNN será responsável por extrair *features* das imagens e então utilizar uma rede neural *multilayer perceptron* de regressão, a qual vai fazer as estimativas da produtividade da área com base na imagem fornecida como entrada na rede.

## 4.2 Decisões de projeto

Existem alguns fatores que tornam laboriosa a definição de um modelo de estimativa de safras, tais como: (i) diferentes critérios para o pré-processamento; (ii) diferentes formas de se definir o conjunto de dados; (iii) conseguir acesso a dados confiáveis e precisos; e (iv) definir qual a cultura a ser analisada.

Os critérios adotados no pré-processamento, são de total importância numa rede neural, dependendo da função se pode tanto aumentar o volume dos dados como reduzir e até melhorar. Também neste trabalho foi optado por passar a preencher as imagens que continham os limites da plantação com a média dos valores de produtividade do restante das imagens.

Ao discutir sobre o conjunto de dados, já se tinha como ponto de partida o uso de imagens do satélite *Sentinel-2*, por estar disponível de forma gratuita e por já ter uma facilidade na composição das imagens. Mas também como um estímulo, esta fonte de imagem ser o mesmo satélite utilizado pelo relatório geo localizado da colheitadeira. Diante disso, como o satélite fornece suas imagens com a definição de dez metros por pixel, foi criado um método com a possibilidade de fragmentar o imóvel a ser analisado em partes menores de 10 x 10 *pixels*, o que equivaleria a 1 hectare para cada parte.

Quanto aos dados de produtividade, realizou-se diversas pesquisas de campo de modo a encontrar imóveis rurais que tivessem dados confiáveis e precisos sobre a produtividade de milho. Sendo assim, a partir de um produtor parceiro, se obteve um relatório que foi extraído da colheitadeira, o qual contém a média de produtividade geo localizada, abrindo portas para

resultado que tanto era almejado.

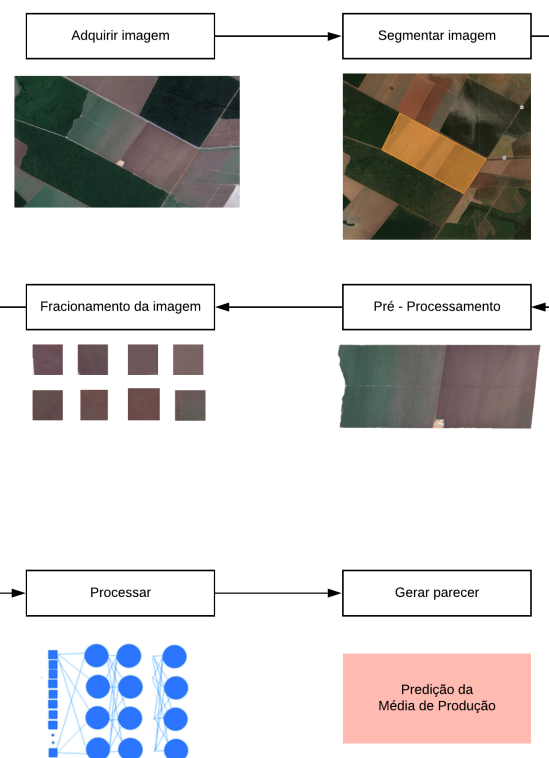
Já ao se decidir pela cultura de milho, foi levado em conta esta cultura possuir as condições meteorológicas ideais para o acompanhamento via satélite, já que ocorre no período de estiagem no Estado do Mato Grosso, local das fontes de dados.

### 4.3 Arquitetura

Uma das bases conceituais do modelo de estimativa de safras é conceder ao analista uma média de produção por hectare da área consolidada do imóvel rural. O núcleo do modelo de estimativa de safras será composto por uma CNN que receberá fragmentos de imagens, uma de cada vez, fornecendo uma estimativa de produtividade de cada área. Posteriormente, o modelo unifica os resultados de cada fragmento para então ter o resultado da imagem inteira. O processo de inferência do modelo de estimativa de safras conforme é apresentado na Figura 13.

Com o intuito de facilitar a compreensão o pipeline foi idealizado conforme os seguintes módulos principais, aquisição da imagem, seleção da área de interesse (unidade produtora), pré-processamento, fracionamento da imagem, processamento, e emissão do parecer. Na Figura 13 é apresentado como foi elaborado o modelo de treinamento e a análise uma área consolidada.

Figura 13 – Modelo de estimativa de produtividade de uma unidade produtora



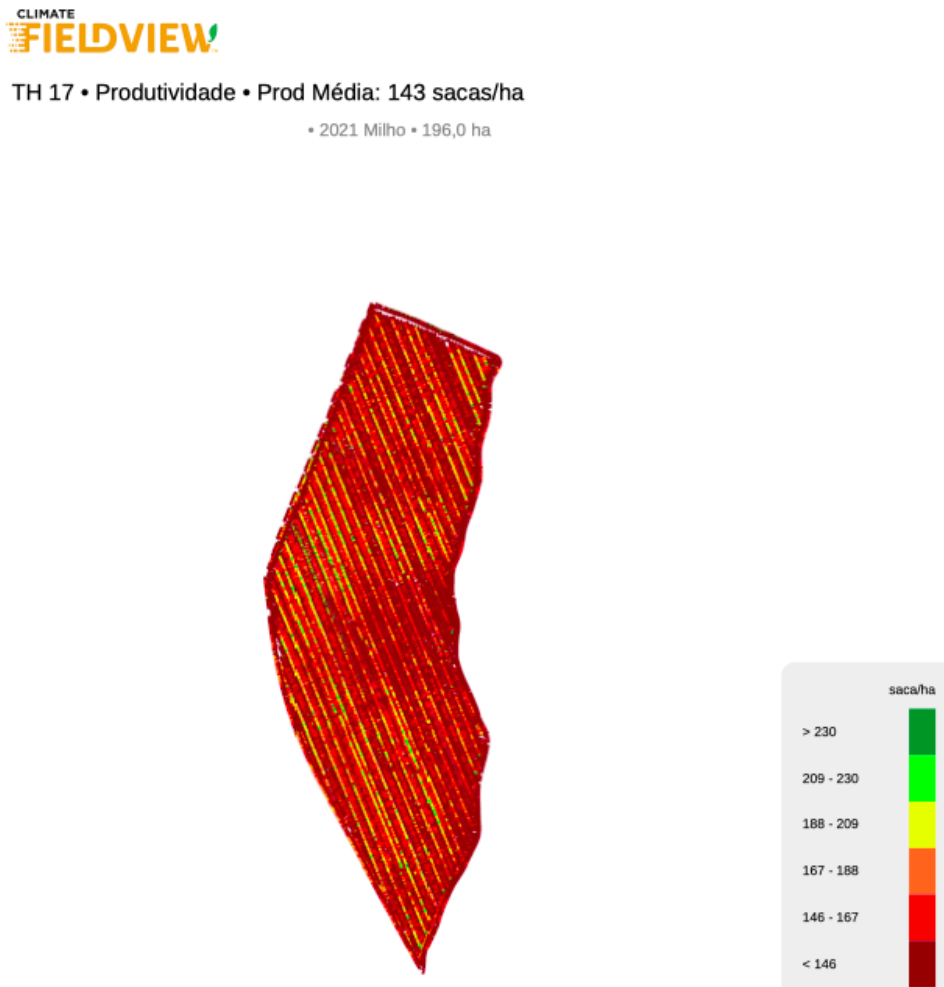
Fonte: Adaptado pelo autor

#### 4.3.1 Conjunto de dados - *Dataset*

Um importante passo no aprendizado de máquina é a definição dos dados adequados para o treinamento e testes da rede neural. Trabalhar com um conjunto de dados fidedigno permite detectar possíveis erros na rede neural e até melhorar os resultados. A decisão de utilizar de um conjunto de dados já disponível é considerada uma opção por muitos trabalhos, pois a obtenção de dados próprios é considerada uma tarefa complexa e demorada em alguns casos, no presente trabalho foi optado por criar o próprio conjunto de dados.

Na definição do conjunto de dados, decidiu-se por ir a campo e conseguir dados de produtividades da cultura de milho, com produtores locais na região norte do Estado do Mato Grosso, mais precisamente na cidade de Sorriso. Neste processo, após diversas interações foi adquirido por meio de um parceiro para este estudo e conquistou-se o acesso à média de produtividade da safra de 2021 de algumas fazendas. Analisando o relatório que foi obtido, identificou-se que foi gerado através de um software, denominado *Climate Fieldview*, integrado à colheitadeira. Conforme está apresentado na Figura 14.

Figura 14 – Relatório integrado a colheitadeira



Através deste relatório, foram obtidos dados extremamente precisos sobre a produtividade de maneira geo localizada da safra de milho. Ao aprofundar a análise do relatório que foi concedido, foi checado junto a empresa fabricante do mesmo, qual o satélite utilizado para fornecer as imagens que são utilizadas como um *background* na impressão de sua média de produção e então foi constatado de se tratar do mesmo satélite que fora adotado para este trabalho e no qual já era possuído o conhecimento e habilidades para localizar uma imagem e compô-la. Esse *matching* foi decisivo para a definição de um método para cálculo da produtividade e também para a definição do tamanho dos arquivos na entrada de processamento da rede neural. Uma vez que, o *Sentinel-2* possui como resolução dez metros por pixel em suas imagens e o relatório nos fornece a produtividade por pixel. Sendo assim, decidiu-se por usar o quadrante de 10x10 *pixels*, ou seja, 1 hectare.

Agora definido o tamanho da entrada dos dados na rede neural, seguiu-se para o desafio de criar o conjunto de dados, onde também foi desenvolvido um método para localizar cada imagem do imóvel rural e sua produtividade, ficando da seguinte forma: c10g4c112p140, ou seja, coluna 10, grid 4, coluna\_grid 1, linha\_grid 2 e produtividade 140. Sendo essa a definição que foi levada para ser o rótulo dos arquivos de entrada no conjunto de dados.

Para definir a produtividade, foi necessário a criação de um cálculo para parametrizar a média de produção com base nas cores do relatório e decidiu-se por não utilizar o intervalo hora definido no relatório, onde ficou fixado com base na média dos primeiros 200 quadrantes calculados. Então ficou definido para entre o verde e o azul 230 saca/ha, verde 215 saca/ha, amarelo 190 saca/ha, entre o amarelo e o vermelho 170 saca/ha, vermelho 160 saca/ha, entre o vermelho e o bege 70 saca/ha e branco 0 saca/ha. Sendo esse processo repetido até ser alcançado as 761 imagens, ou seja, a totalidade do imóvel analisado, conforme está apresentado na Tabela 5.

Tabela 6 – Relação de cores com a produtividade

CORES	PRODUTIVIDADE
BRANCO	0 Sc/Ha
ENTRE VERMELHO E BEGE	70 Sc/Ha
VERMELHO	160 Sc/Ha
ENTRE VERMELHO E AMARELO	170 Sc/Ha
AMARELO	190 Sc/Ha
VERDE	215 Sc/Ha
ENTRE O VERDE E O AZUL	230 Sc/Ha

E assim foi dado início da criação das imagens para o conjunto de dados, contabilizando os 100 *pixels* de cada imagem gerada e definindo quanto dava a média desse fragmento da fazenda, conforme está apresentado na Figura 15, que foi um método criado para calcular uma média de produtividade com base nos dados fornecidos pelo relatório da colheitadeira.

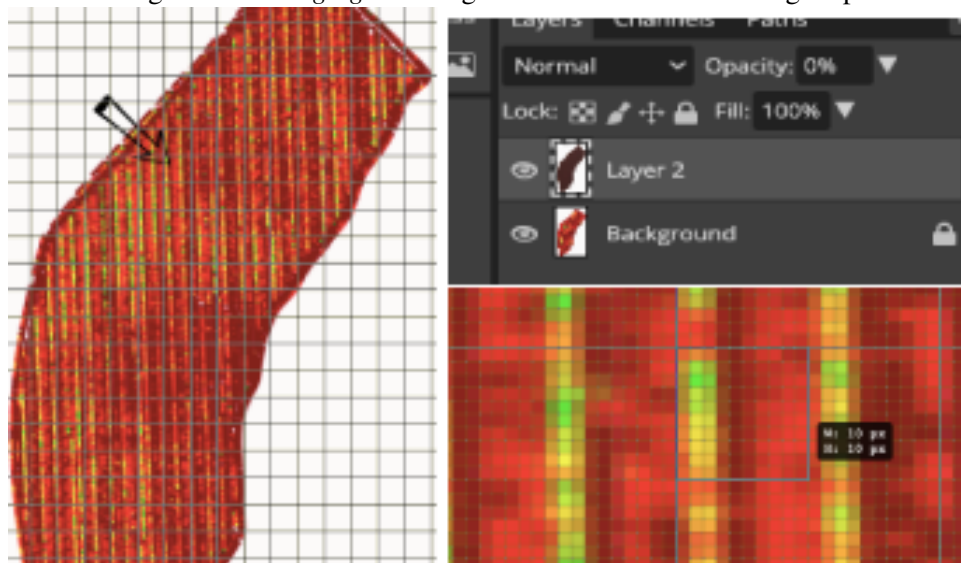
Figura 15 – Método criado para calcular a média de produtividade  
Calculo da média por hectare

	Média relatório	Pixels	Parametrização DPY	Média DPY
verde escuro	>230	0	230	0
verde	209 - 230	8	215	17,2
amarelo	188 - 209	8	190	15,2
laranja	167 - 188	12	170	20,4
vermelho	146 - 167	57	160	91,2
bordo	< 146	15	70	10,5
nada		0		
			1035	154,5
Qtde pixels		100		

Fonte: Adaptado pelo autor

Mesmo tendo uma imagem relativamente pequena, 10x10 *pixels*, é possível verificar no quadrante destacado de tamanho 10x10 *pixels* a presença das tonalidades de cor verde, amarela, vermelha e as cores que ficam no intervalo entre amarelo e vermelho e também entre as cores vermelho e bege. Conforme está apresentado na Figura 16.

Figura 16 – *Merging* das imagens do relatório com a imagem pré-colheita



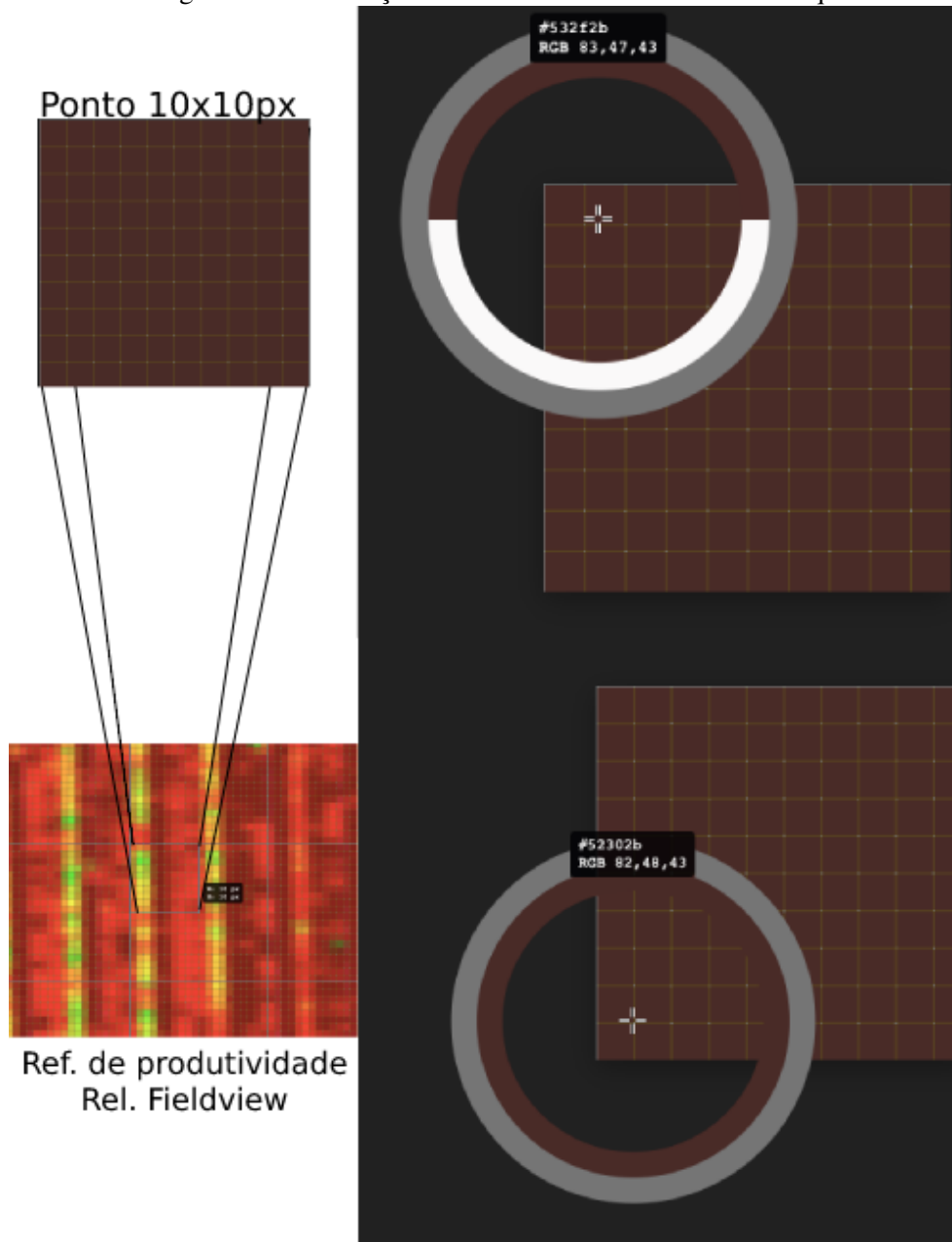
Fonte: Adaptado pelo autor

Outro fator a ser ressaltado foi o critério para o recorte da imagem onde é feito o *merging*



da imagem do relatório com a produtividade com a imagem de satélite do dia 04 de junho de 2021, última imagem antes do início da safra, como está apresentado na Figura 17. Com isso, ao definir a área de recorte visualizando o relatório, automaticamente extraiu-se o recorte da camada onde continha a imagem pré-colheita.

Figura 17 – Diferenças de características do RGB em um quadrante



Fonte: Adaptado pelo autor

Por fim, foi feito o recorte das 761 imagens do imóvel rural da presente pesquisa, todas estas imagens foram rotuladas com sua localização em relação a imagem total do imóvel bem como, a média da produtividade calculada, como está apresentado na Figura 18.

Figura 18 – Relação de algumas das imagens do conjunto de dados

### Conjunto de dados - DPY

	Produtividade 124 sc/Ha	- c15g3c11p124
	Produtividade 110 sc/Ha	- c12g21c21p110
	Produtividade 155 sc/Ha	- c15g4c11p155
	Produtividade 088 sc/Ha	- c12g10c21p088
	Produtividade 060 sc/Ha	- c12g1c21p060
	Produtividade 122 sc/Ha	- c2g2c11p122
	Produtividade 116 sc/Ha	- c2g4c11p116
	Produtividade 090 sc/Ha	- c2g4c112p090

Fonte: Adaptado pelo autor

#### 4.3.2 Pré-Processamento da imagem

Como ponto inicial foi criado uma função para gerar os dados categóricos baseados no nome do arquivo, que foi gerado a partir das análises do relatório da colheitadeira e recorte da imagem pré-colheita, com isso, foi gerada uma tabela dividida com as colunas: *GridColumn*, *Grid*, *Column*, *Line* e *Productivity*. Que são os rótulos hora definidos para as imagens com a localização no imóvel e media de produção. Conforme está apresentado na Figura 19.

Ao definir a função para pré-processar os dados onde, é passado por um quadro gerado através de uma função do pandas, que auxilia a descartar alguns registros do conjunto de dados, por considerar dados com muitos ruídos. Em seguida, foram pegos os valores e usou-se *scikit-learn MinMax Scaler* que reduz os dados dentro de um determinado intervalo, geralmente de 0 a 1. Ele transforma os dados escalando recursos para um determinado intervalo. Dimensionando os valores para um intervalo de valor específico sem alterar a forma da distribuição original.

Figura 19 – Rótulos e valores do conjunto de dados

	GridColumn	Grid	Column	Line	Productivity
0	2	2	1	1	122
1	2	2	2	1	52
2	2	2	1	2	117
3	2	2	2	2	119
4	2	3	1	1	134
..	...	...	...	...	...
754	15	3	2	2	151
755	15	4	1	1	155
756	15	4	2	1	152
757	15	4	1	2	143
758	15	4	2	2	145

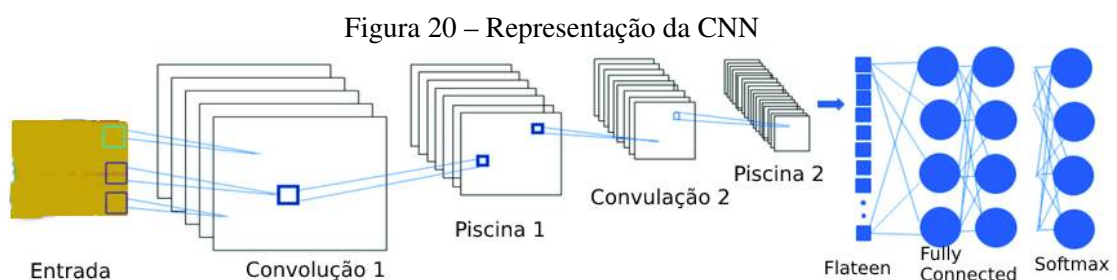
[759 rows x 5 columns]

Fonte: Adaptado pelo autor

#### 4.3.3 Convolutional neural network - CNN

O grande diferencial das CNNs está nas diversas camadas convolucionais, que aplica uma função matemática de convolução nos dados de entrada e depois realizando o agrupamento. A saída da convolução é passada para a próxima camada convolucional até chegar na última camada conhecida como camada densa que será representada por uma rede *multilayer* perceptron, como pode ser visualizada na Figura 20.

Nesta seção é apresentada a implementação da rede neural convolucional, por meio do *Keras*, que vem ser uma biblioteca aberta de aprendizado de máquina e que permite modelar e treinar modelos de redes neurais de maneira mais prática.



Fonte: Adaptado pelo autor

Alguns pontos importantes merecem destaques, como os parâmetros de entrada da rede neural, tais como: (i) *width* e *height*, que são passadas as dimensões da imagem de entrada; (ii) *depth*, que vem a ser o numero de canais para a imagem, onde foi definido o 3 para imagens RGB; (iii) *filters*, que é uma tupla de filtros progressivamente maior para que a rede possa aprender recursos mais detalhados; (iv) *regress*, vem a ser um booleano que indica se uma camada de ativação linear totalmente conectada será ou não anexada à CNN para fins de regressão.

O módulo *Keras* fornece uma camada convolucional para imagens, chamada de *Conv2D*,

onde é realizada a convolução e deve-se especificar a quantidade de filtros, e o tamanho do *kernel*. Mas também disponibiliza outras camadas, para serem utilizadas em combinação sequencial, são elas:

- *MaxPooling2D*, que tem a função de agir reduzindo ou agrupando *pixels* de uma determinada região, com a finalidade de diminuir a variância a pequenas alterações e também de reduzir a quantidade de parâmetros;
- *Dropout*, vem a ser o responsável por regular, onde alguns neurônios são desligados aleatoriamente, juntamente com suas conexões, durante o treinamento apenas;
- *Flatten*, recebe uma matriz resultante das camadas anteriores. Com o objetivo de concatenar todas as linhas de uma matriz para formar um único vetor linear, de uma única dimensão;
- *Dense* é voltada para uso na rede neural totalmente conectada, ou *fully connected*, nela deve-se informar a dimensão da saída e a função de ativação a ser utilizada;
- *Activation*, utiliza-se uma função de ativação, ao final de uma camada *Conv2D* e da camada *Dense*. Baseando-se numa inspiração biológica.

Conforme está demonstrado na Figura 21.

Figura 21 – Utilização de camadas do Keras

```

def create_cnn(width, height, depth, filters=(16, 32, 64), regress=False):
    # initialize the input shape and channel dimension, assuming
    # TensorFlow/channels-last ordering
    inputShape = (height, width, depth)
    chanDim = -1

    # define the model input
    inputs = Input(shape=inputShape)

    # loop over the number of filters
    for (i, f) in enumerate(filters):
        # if this is the first CONV layer then set the input
        # appropriately
        if i == 0:
            x = inputs

        # CONV => RELU => BN => POOL
        x = Conv2D(f, (3, 3), padding="same")(x)
        x = Activation("relu")(x)
        x = BatchNormalization(axis=chanDim)(x)
        x = MaxPooling2D(pool_size=(2, 2))(x)

    # flatten the volume, then FC => RELU => BN => DROPOUT
    x = Flatten()(x)
    x = Dense(100)(x)
    x = Activation("relu")(x)
    x = BatchNormalization(axis=chanDim)(x)
    x = Dropout(0.5)(x)

    # apply another FC layer, this one to match the number of nodes
    # coming out of the MLP
    x = Dense(4)(x)
    x = Activation("relu")(x)

    # check to see if the regression node should be added
    if regress:
        x = Dense(1, activation="linear")(x)

    # construct the CNN
    model = Model(inputs, x)

    # return the CNN
    return model

```

Fonte: Adaptado pelo autor

Foi utilizada a *ReLU* como função de ativação, sendo utilizada apenas nas camadas ocultas. Enquanto na camada de saída, por outro lado, foi utilizada uma função identidade (ou linear), comumente utilizada em tarefas de regressão.

Em seguida, avançou-se para a definição da divisão dos dados do conjunto, onde está sendo

passado o percentual que serão direcionados para treinamento e teste. Para o presente estudo foi definido o mais habitual de acordo com a literatura, ou seja, 75% para treinamento e o restante 25% para testes, conforme está apresentado na Figura 22.

Figura 22 – Divisão dos dados do conjunto de dados

```
# partition the data into training and testing splits using 75% of
# the data for training and the remaining 25% for testing
split = train_test_split(areasInfosTable, areaImages, test_size=0.25, random_state=42)
(trainAttrX, testAttrX, trainImagesX, testImagesX) = split
```

Fonte: Adaptado pelo autor

Já a função de perda, esta definida para o uso do erro percentual absoluto médio. Por fim é ativado a função de regressão, definido que a rede passará por um treino de 300 *epochs*, utilizando um otimizador mais avançado (*Adam*) e definiu-se algumas métricas, a fim de validar os resultados da estimativa. Como pode ser visto na Figura 23.

Figura 23 – Função de regressão

```
# find the largest productivity in the training set and use it to
# scale our productivity to the range [0, 1] (will lead to better
# training and convergence)
maxProductivity = trainAttrX["Productivity"].max()
trainY = trainAttrX["Productivity"] / maxProductivity
testY = testAttrX["Productivity"] / maxProductivity

# create our Convolutional Neural Network and then compile the model
# using mean absolute percentage error as our loss, implying that we
# seek to minimize the absolute percentage difference between our
# price *predictions* and the *actual productivity*
model = create_cnn(20, 20, 3, regress=True)
opt = Adam(lr=1e-3, decay=1e-3 / 300)
model.compile(loss="mae", optimizer=opt, metrics=["mae", "mse", "mape"])

# train the model
print("[INFO] training model...")
history=model.fit(x=trainImagesX, y=trainY,
                 validation_data=(testImagesX, testY),
                 epochs=300, batch_size=8)
```

Fonte: Adaptado pelo autor

#### 4.4 Discussão

No bioma amazônico, de acordo com a matéria publicada por WASLEY (2019), diversas unidades produtoras praticam a lavagem de grãos provenientes de imóveis embargados. Geralmente, essas unidades não caem no radar das empresas ligadas ao comércio de *commodities*,

pois mesmo que estas pratiquem na íntegra o que versa a moratória da soja, elas não são capazes de verificar a capacidade de produção real das unidades produtoras.

Desta forma, visto que no processo atual os analistas responsáveis não possuem a informação de produtividade de uma unidade produtora, tendo no máximo acesso a uma média de produção da região (que é fornecida pelo Instituto Mato-Grossense de Economia Agropecuária IMEA), é preciso conceber formas de estimar esta produtividade. Neste contexto, espera-se que o modelo apresentado neste trabalho seja capaz de trazer ao analista uma estimativa de produtividade da unidade produtora, contribuindo assim para o combate à lavagem de grãos. Vale destacar que isto seria viável ao utilizar o modelo DYP, dado que o analista será capaz de identificar propriedades que vendem além do que é efetivamente produzido.





## 5 RESULTADOS

Neste capítulo são apresentados os resultados do modelo de estimativa da produtividade de milho. A seção 5.1 descreve a metodologia dos experimentos. Já na seção 5.2 são apresentados os resultados numéricos que foram alcançados. Por fim na seção 5.3 é adicionada uma breve discussão sobre os resultados obtidos.

### 5.1 Metodologia dos experimentos

Para os experimentos deste trabalho, não foi selecionada nenhuma *baseline*, uma vez que o trabalhos mais circunjacentes foram dos colegas EFREMOVA; ZAUSAEV; ANTIPOV (2019) e NEVAVUORI; NARRA; LIPPING (2019). Conforme mencionado na seção 3.4, tais comparações seriam injustas considerando que os trabalhos possuem tecnologias diferentes e aplicações diferentes.

A metodologia para definição dos experimentos foi um dos momentos decisivos do presente trabalho, por conta de uma necessidade de ter que ir à campo em busca de informações que pudessem servir de base para criar o escopo metodológico. Esta etapa ocorreu durante a pandemia. Embora o segmento do agronegócio não tenha parado o seu ciclo, os pequenos produtores não tinham informações ou dados que pudessem dar base para definição do modelo de estimativa, enquanto os grandes produtores naquele momento não davam abertura para que pudesse apresentar e explicar a finalidade desta pesquisa.

Apenas neste ano, em maio de 2022, foi dada uma oportunidade de acesso a um grande produtor parceiro, que depois de diversas tratativas forneceu um relatório contendo a média de produção por hectare da safra 2021 de milho. Este relatório possui uma informação muito precisa, pois foi gerado através da própria colheitadeira e de maneira geo-localizada. Na Figura 24 está demonstrado a imagem extraída do satélite *Sentinel-2* pré-colheita e a imagem contida no relatório do *FieldView*.

Figura 24 – Imagem pré-colheita *Sentinel-2* e Imagem extraída do relatório  
Imagem satélite *Sentinel-2* 04/06/2021 Imagem extraída do relatório  
FieldView



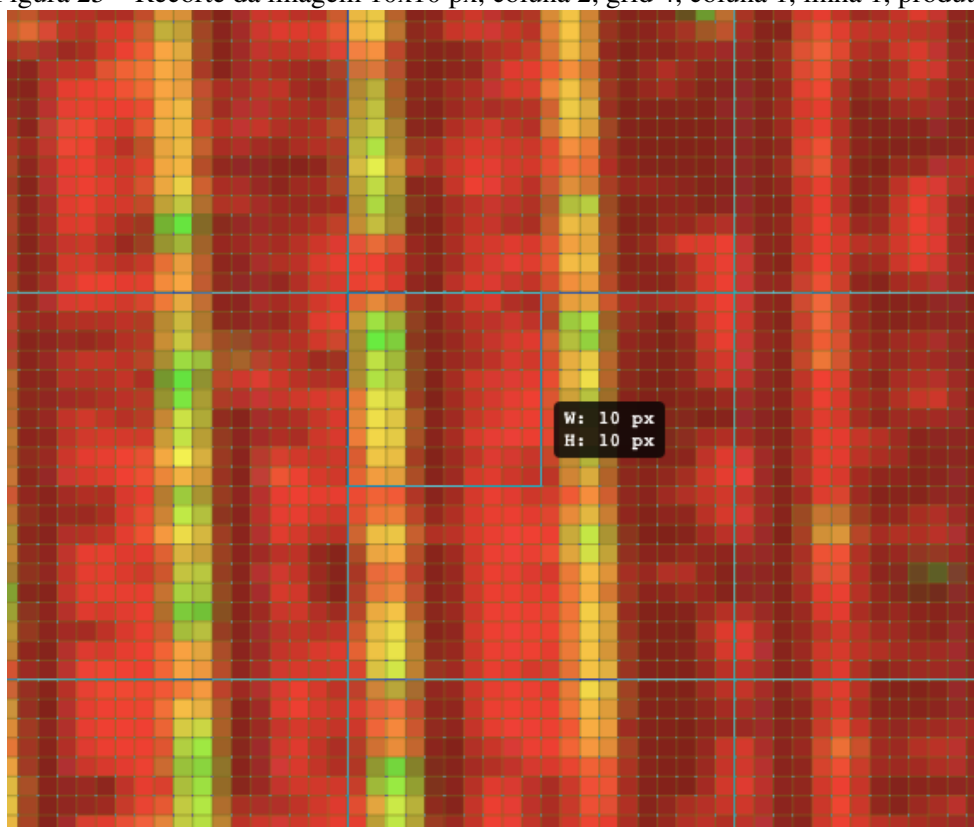
Fonte: Adaptado pelo autor

A partir desse artefato em mãos, se teve um salto na qualidade da informação sobre a produtividade de um imóvel. Onde antes foi herdado uma média de produtividade declaratória, agora passou-se a ter uma média comprovatória do que foi produzido por hectare. Diante disso, foi dado início na definição do critério de cálculo para extrair a produtividade com base na cor do pixel. Então, foi feito o *merging* das imagens da plantação com a imagem extraída do relatório e com isso, foi dado início ao processo de recorte das imagens.

Assim sendo, foi elaborado o recorte das imagens conforme explicado na seção 4.3.1. Onde

foi definido o tamanho do recorte e com o intuito de aumentar o volume do conjunto de dados e por saber que a resolução da imagem de satélite tinha uma definição de dez metros por *pixel*, decidiu-se fazer os recortes em tamanho de 10 x 10 *pixels* o que equivale a 1 hectare, como está apresentado na Figura 25.

Figura 25 – Recorte da imagem 10x10 px, coluna 2, grid 4, coluna 1, linha 1, produtividade 154



Fonte: Adaptado pelo autor

Mas para manter a possibilidade de identificar a localização do recorte foi necessário a criação de um modelo de nomenclatura para os arquivos que seriam gerados com a produtividade, conforme explicado na seção 4.3.1, que ficou como: coluna, *grid*, coluna da *grid*, linha da *grid* e produtividade.

No entanto, foi necessário a criação de um cálculo para parametrização da média de produção com base nas cores do relatório. Assim foram definidos os valores que cada cor representaria de produtividade. Logo, após somar todos os 100 *pixels* foi feito uma média de acordo com cada cor e depois somada, apresentando assim o valor médio do quadrante 10x10 *pixels* ou de 1 hectare, como está apresentado na Figura 26. Sendo esse processo repetido até alcançarmos as 761 imagens, ou seja, a totalidade do imóvel analisado. Assim sendo, era o momento de colocar a rede neural para analisar o conjunto de dados e se avaliado os resultados do modelo de estimativa.

Por fim, para avaliarmos a performance do modelo, buscamos um indicador que represente fielmente os erros que estamos obtendo, quando comparamos o que foi projetado, com o que

realmente esta ocorrendo. Diante disso, optamos pelo uso dos indicadores mais utilizados de acordo com a literatura para se avaliar se um modelo teve boa ou má performance. Sendo assim, verificaremos sua acuracidade por três medidas distintas: *mean absolute error (MAE)*, *mean absolute percentage error (MAPE)* e *mean squared error (MSE)*.

Após o treinamento, a adição de uma nova informação implica em retreinamento da rede para esta informação e para toda a informação anterior. Uma vez que a estratégia de retreinamento consiste em incrementar a estrutura padrão de treinamento do modelo, dando novas informações para a rede aos poucos.

- Diferença entre as produtividades previstos do imóvel e as produtividades reais do imóvel;
- Diferença percentual;
- Diferença percentual absoluta;

Diante disso, foi calculado a média e o desvio padrão da diferença do percentual absoluto. Nesta seção é demonstrado o que foi feito para avaliar os resultados obtidos com o modelo de estimativa de safras de milho.

## 5.2 Resultados numéricos

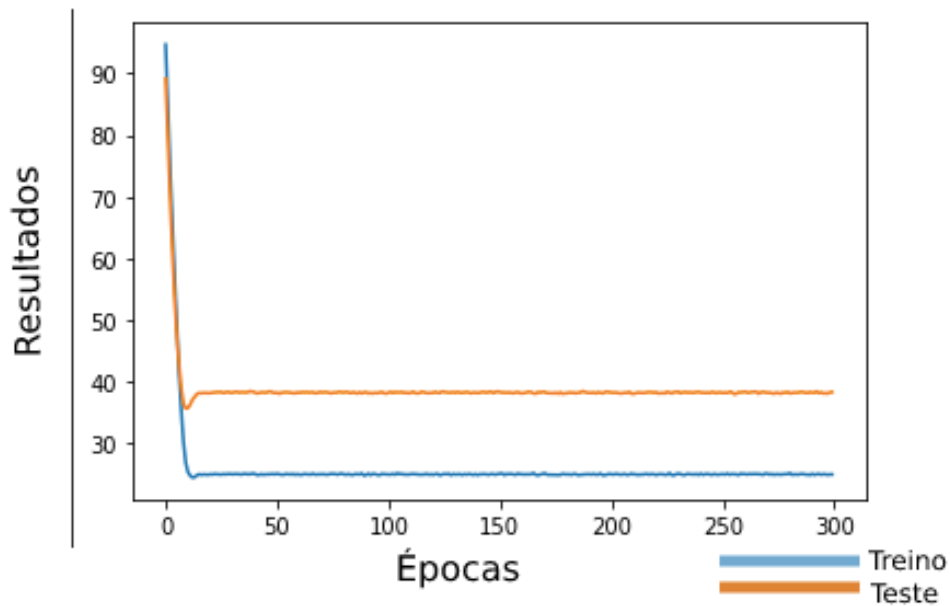
Analisando e avaliando os resultados do modelo, o erro percentual absoluto médio começa extremamente alto, acima de 94%, mas já nas primeiras épocas ele caiu drasticamente, no entanto, ao ter o treinamento concluído depois das 300 épocas, foi alcançada uma perda no treinamento inferior a 25%. Enquanto a perda nos dados de teste chegou a ultrapassar os 89%, finalizou com 38%. Como pode ser observado na Figura 26.

Significando que o erro percentual absoluto médio final implica que, em média, a rede terá 38% de desconto em suas previsões de safras de milho. O que nos demonstra por meio da métrica MAPE, que o modelo ainda está longe do ideal que se almejava.

Já em comparação com a métrica de erro médio absoluto, foi obtido os resultados nas primeiras épocas na ordem de superior à 0,7020 e após as 300 épocas alcançado o resultado de 0,1087. Já nos dados de validação iniciou em suas primeiras épocas com 0,6498 e finalizou com 0,1382. Conforme está demonstrado na Figura 27.

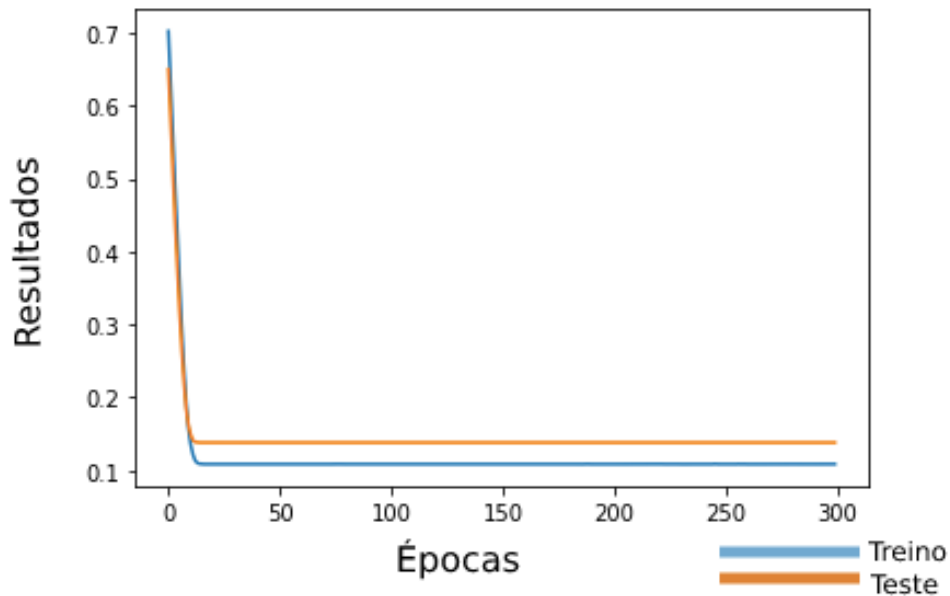
Enquanto os resultados obtidos com a métrica do erro quadrático médio, começou as épocas com 0,5161 e após percorrer as 300 épocas alcançou-se o resultado de 0,0246. Ao verificar os dados de validação que começou com o valor na primeiras épocas de 0,4543 e finalizou com o valor de 0,0358, conforme está apresentado na Figura 28.

Figura 26 – Gráfico execução da métrica - *Mean absolute percentage error*



Fonte: Adaptado pelo autor

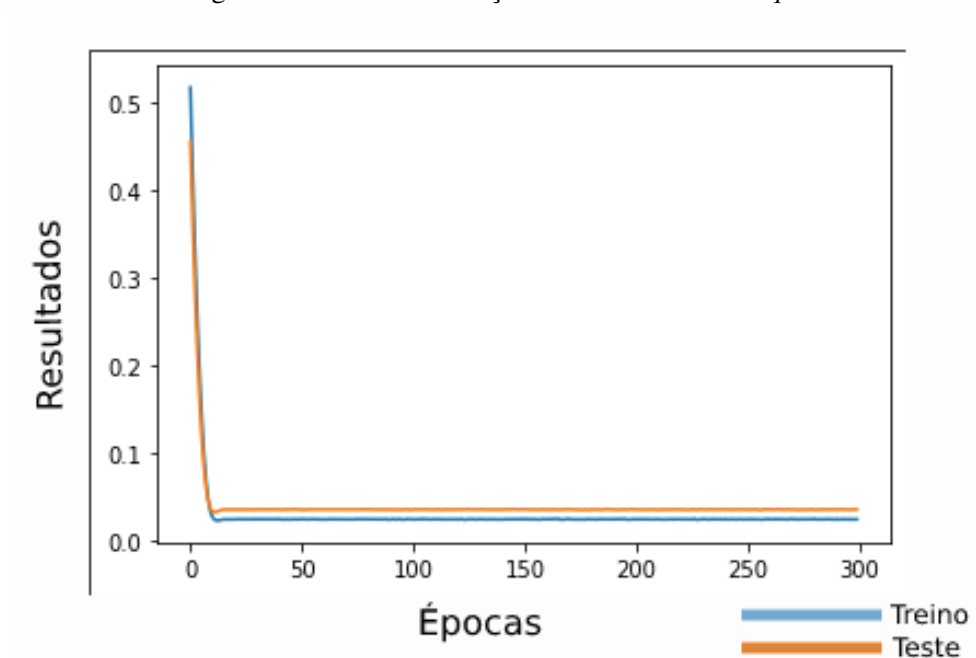
Figura 27 – Gráfico execução da métrica - *Mean absolute error*



Fonte: Adaptado pelo autor

### 5.3 Discussão

É possível verificar que há uma simetria entre as perdas nos dados de treinamento e teste, onde andam separadas por um pequeno intervalo de início até o final das épocas. Mas é o suficiente para identificar que não há *outliers* no conjunto de dados. Analisando a perda através

Figura 28 – Gráfico execução da métrica - *Mean squared error*

Fonte: Adaptado pelo autor

da métrica MAPE, foi verificado que ela se aproxima de 30% próximo da 10ª época, mas em seguida ela sobe um pouco novamente e praticamente estabiliza sua perda em 38%.

Embora tenha sido realizado alguns testes de divisão no modelo de dados para treinamento e teste, não surtiu resultado impactante, apenas reduziu 1% da perda ficando no final com 37%. Diante disso, acredita-se que seja possível melhorar a performance da regressão através de aplicação de outros tipos de filtros, tanto no pré-processamento como direto na imagem através de sensores espectrais, como o infra-vermelho ou através do índice de vegetação da diferença normalizada. Mas é suposto ser de suma importância o aumento do conjunto de dados.

Vale destacar, que embora o modelo DYP não possua um pequeno índice de perda no momento, este modelo pode ser melhorado. Já que obteve uma média de produtividade por hectare de 126,09 sacas por hectare, como pode ser visualizado na Figura 29, enquanto a média apresentada no relatório *FieldView*, apresentou uma média de produtividade de 143 sacas por hectare, anteriormente apresentado na seção 4.3.1 na Figura 17, ou seja, uma diferença de aproximadamente 12%. Diferença esta, que é entendido que seja possível de ser reduzida, por meio de um ajuste no método de cálculo da média.

Figura 29 – Resultados do modelo de estimativa DYP

```
[INFO] predicting area productivity...
[INFO] avg. area productivity: 126.096179183, std area productivity: 27.4530056344
[INFO] mean: 38.32%
```

Fonte: Adaptado pelo autor

O ajuste no método se dá por conta de inicialmente se ter definido uma média como base na primeira coluna analisada, onde acabou-se definindo uma média que ficou muito baixa na

tonalidade que fica entre as cores vermelho e bege, ocasionando uma perda elevada. Outro item que vale frisar são as imagens com a resolução de 10 metros por *pixel* que são acessadas através do satélite *Sentinel-2*, embora haja variação de tonalidade dos *pixels*, uma imagem com uma resolução mais elevada, com certeza daria um melhor resultado. Pois seria possível de identificar maiores variações em um quadrante de recorte maior, como  $32 \times 32$  *pixels*. Lembrando que foi decidido por fazer o recorte em  $10 \times 10$ , objetivando aumentar o volume de dados do conjunto.

Enfim, acredita-se que o modelo DYP seja capaz de auxiliar os trabalhos de checagem de média de produção e identificar possíveis lavagens de grãos, já que, não torna viável para o produtor que faz uso dessa prática, ficar destinando à vários locais partes de sua produção, o que dificultaria a identificação da lavagem de grãos. Como o modelo DYP no presente momento possui uma diferença de 12% à menos em relação a produção real, ainda o torna um importante recurso que pode a vir ser usado para combater o grão pirata.





## 6 CONCLUSÃO

Esse trabalho buscou trazer insumos que visam colaborar nos serviços de análise e fiscalização de uma unidade produtora de milho, auxiliando na identificação da média de produção por hectare de milho. Buscando também incluir no processo de análise de riscos, a validação da produtividade da unidade produtora, o que pode vir a contribuir na identificação de possíveis lavagens de grãos.

Apesar dos testes e treinamentos terem sido realizados em áreas do Estado do Mato Grosso, acreditamos que este modelo seja capaz generalizar sua aplicabilidade em qualquer outro local desde que tenha as imagens de satélite com uma semana de antecedência da cultura de milho, conforme foi adotado para criação dos testes e treinamento atuais. Os motivos pelos quais optamos por estas opções já foram explanadas na subseção 4.3.1.

Dando importância ao grande acervo de imagens disponíveis de satélites. Foi optado por fazer uso dessa fonte, no modelo de estimativa de safras de milho, hora denominado DYP e a escolha da cultura de milho visando justamente ter acesso a diversas imagens, já que a safra de milho ocorre no período de estiagem o que facilita encontrar imagens limpas, ou seja, sem fator meteorológico que possa vir a atrapalhar como nuvens. Viabilizando a estimativa da produtividade de qualquer imóvel rural, que será definido pelo próprio usuário. Proporcionando ao analista ou autoridade fiscalizadora, a checagem da origem dos grãos e colaborar para identificar fraudes ligadas a "lavagem de grãos" ou "grãos piratas".

Existem diversas técnicas inteligentes com potencial para serem utilizadas na estimativa de safras. Porém, estudos utilizando CNN aplicados a imagens apresentaram melhores resultados e reduzem a dependência humana no processo. Nesse contexto, acredita-se que a rede neural convolucional tenha sido a melhor referência para o aprendizado das características da imagem das safras para o modelo de estimativa de safras, *Deep Yield Prediction* (DYP).

O modelo DYP apresentou resultados significativos e que podem ser evoluídos. Por meio do resultado demonstrado pela métrica do erro percentual absoluto médio, MAPE, onde apresentou nos resultados de treinamento nas primeiras épocas superior a 94% e após percorrer as 300 épocas chegou ao resultado inferior a 25%. Enquanto na validação que começou com o valor nas primeiras épocas de ultrapassou os 89% e finalizou com o percentual inferior a 38%.

Portanto, esse resultado traz o significado de que o erro percentual absoluto médio implica que, em média, a rede terá 38% de desconto em suas previsões, ou seja, para o imóvel analisado obteve-se uma média de 126,09 sacas de milho por hectare, enquanto a produtividade real obteve 143 sacas por hectare. O que não vem ser um bom resultado. Embora que essa perda esteja representando um percentual de -12% em relação à média de produtividade real. Com isso é possível verificar que o conjunto de dados, não é de todo mal, uma vez que não dá sinais de *outliers*. Sendo assim, pode-se prosseguir em busca aumentar o conjunto de dados e fazer novos experimentos de filtros, bem como remodelar o método de cálculo.

Diante disso, espera-se alcançar melhores resultados a partir de possíveis evoluções. Bus-

cando assim contribuir, para um maior dinamismo e precisão ao analisar ou fiscalizar uma unidade produtora, possibilitando saber de fato o quanto uma unidade é capaz de produzir. Reduzindo o risco do comércio de grãos provenientes de imóveis embargados por conta de desmatamento ilegal do bioma amazônico e corroborando para redução do desmatamento ilegal na Amazônia.

## 6.1 Limitações

Já para a evolução, vale destacar alguns tipos de ruídos que podem estar presentes nas imagens, tais como: redemunhos de vento, ataques de animais e condições meteorológicas.

Os redemunhos de vento ocorrem quando o solo se aquece até determinado ponto, ocasionando a transferência desse calor à uma parcela de ar que está parada acima dele criando o fenômeno do redemunho, o que leva a ocasionar a perda de diversos pés da cultura.

Os ataques de animais também podem prejudicar uma cultura de milho, principalmente nas divisas com as áreas de preservação, sendo geralmente ocasionadas por porcos selvagens. Porém estes ataques não trazem destaques nas imagens para o presente estudo, pois nesta pesquisa são utilizadas as imagens pré colheita, enquanto estes tipos de ataques ocorrem logo após a plantação.

Já a meteorologia é ruído a ser destacado, mas optamos por trabalhar com a cultura de milho, justamente por ocorrer no período de estiagem na região onde foram adquiridos os relatórios de produtividade. O que leva a manter o foco em culturas que ocorrem nessa janela, uma vez que, facilita a frequente aquisição de imagens de satélite.

## 6.2 Trabalhos futuros

Como trabalhos futuros pretende-se:

- Aumentar o conjunto de dados, para utilizar o modelo em diferentes culturas;
- Utilizar outras fontes de dados para aprimorar a qualidade das informações, visando atender outras culturas;
- Alertar unidades produtoras embargadas que estão persistindo em produzir, podendo auxiliar outros seguimentos da cadeia produtiva, como os órgãos de controle.

## REFERÊNCIAS

- ABDULLAHI, H. S. Convolution neural network in precision agriculture for plant image recognition and classification. **2017 Seventh International Conference on Innovative Computing Technology (INTECH)**, [S.l.], 2017.
- ABIOVE. Gts, inpe moratória da soja. , [S.l.], 2017.
- ABRAMOVAY, R. Desenvolvimento sustentável: qual a estratégia para o brasil? **Novos estudos CEBRAP**, [S.l.], p. 97–113, 2010.
- ADISA, O. M. Application of artificial neural network for predicting maize production in south africa. **mdpi sustainability**, [S.l.], 2019.
- ALENCAR, A. et al. **Desmatamento na amazônia**: indo além da "emergência crônica". [S.l.]: Ipam Belém, 2004. v. 90.
- ARAÚJO, R. M. d. Aprendizado de máquina em sistemas complexos multiagentes: estudo de caso em um ambiente sob racionalidade limitada. , [S.l.], p. 06, 2005.
- B. KITCHENHAM P. BRERETON, D. B. M. T. J. B. S. L. Systematic literature reviews in software engineering – a systematic literature review. **Information and Software Technology**, [S.l.], v. 51, 2009.
- B. KITCHENHAM P. BRERETON, D. B. M. T. S. L. M. J. E. M. G. V. Guidelines for performing systematic literature reviews in software engineering. **Information and Software Technology**, [S.l.], 2007.
- BALLESTER, M. V. R. et al. A remote sensing/gis-based physical template to understand the biogeochemistry of the ji-paraná river basin (western amazônia). **Remote Sensing of Environment**, [S.l.], v. 87, n. 4, p. 429–445, 2003.
- BRADDOCK, T. et al. Fruit yield prediction using artificial intelligence. In: ASABE ANNUAL INTERNATIONAL MEETING, 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 1.
- BRAMLEY, R. G. Lessons from nearly 20 years of precision agriculture research, development, and adoption as a guide to its appropriate application. **Crop and Pasture Science**, [S.l.], v. 60, n. 3, p. 197–217, 2009.
- BRANDAO, J. **Arranjo institucional para o combate ao desmatamento no município de querência/mt**. 2020. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo, 2020.
- BROWN, J. C.; KOEPPE, M. Moratória da soja na amazônia brasileira e governança ambiental: regulação através do mercado ou regulação estatal? **Revista Pós Ciências Sociais**, [S.l.], v. 11, n. 22, 2014.
- C ZHANG L DI, L. L. L. G. Machine-learned prediction of annual crop planting in the us corn belt based on historical crop planting maps. **ELSEVIER**, [S.l.], 2019.
- CAMPOS ANDRÉ, D. C. **‘soja-pirata’ cultivada na amazônia acelera desmatamento e tem participação de gigantes do agro**. 2021.

CARDOSO, F. C. **Do confronto à governança ambiental: uma perspectiva institucional para a moratória da soja na amazônia.** 2008. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo, 2008.

CASHORE, B. Legitimacy and the privatization of environmental governance: how non–state market–driven (nsmd) governance systems gain rule–making authority. **Governance**, [S.l.], v. 15, n. 4, p. 503–529, 2002.

CIVIL, C. et al. Plano de ação para a prevenção e controle do desmatamento na amazônia legal. **Relatório da Presidência da República, Casa Civil**, [S.l.], v. 156, 2004.

COPELAND, M. What’s the difference between artificial intelligence, machine learning and deep learning? , [S.l.], 2016.

CRNOBARAC, J. Data mining approach for predictive modeling of agricultural yield data. **Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment (BioSense09), At Novi Sad, Serbia**, [S.l.], 2009.

EFREMOVA, N.; ZAUSAEV, D.; ANTIPOV, G. Prediction of soil moisture content based on satellite data and sequence-to-sequence networks. **arXiv preprint arXiv:1907.03697**, [S.l.], 2019.

ENRIQUE CASTILLO JOSE M. GUTIERREZ, A. S. H. Expert systems and probabilistic network models. , [S.l.], 1997.

FC EUGENIO M GROHS, L. V. M. S. Estimation of soybean yield from machine learning techniques and multispectral rpas imagery. **ELSEVIER**, [S.l.], v. 0, 2020.

FEARNSIDE, P. M. Deforestation in brazilian amazonia: history, rates, and consequences. **Conservation biology**, [S.l.], v. 19, n. 3, p. 680–688, 2005.

FF HAQUE, A. A. Crop yield prediction using deep neural network. **IEEE**, [S.l.], 2020.

FOLEY, J. A. et al. Global consequences of land use. **science**, [S.l.], v. 309, n. 5734, p. 570–574, 2005.

G. BANNERJEE U. SARKAR, S. D. I. G. Artificial intelligence in agriculture: a literature survey. **International Journal of Scientific Research in Computer Science Application and Management Studies Artificial - IJSRCSAMS**, [S.l.], v. 7, 2018.

GALFORD, G. L. et al. The amazon frontier of land-use change: croplands and consequences for greenhouse gas emissions. **Earth Interactions**, [S.l.], v. 14, n. 15, p. 1–24, 2010.

GANDHI, N.; ARMSTRONG, L. J.; NANDAWADEKAR, M. Application of data mining techniques for predicting rice crop yield in semi-arid climatic zone of india. In: **IEEE TECHNOLOGICAL INNOVATIONS IN ICT FOR AGRICULTURE AND RURAL DEVELOPMENT (TIAR)**, 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 116–120.

GIBBS, H. K. et al. Brazil’s soy moratorium. **Science**, [S.l.], v. 347, n. 6220, p. 377–378, 2015.

GREENPEACE. The eating up the amazon. , [S.l.], 2006.

GUIMARÃES, E. d. S. **Aprendizado de máquina aplicado à predição da produtividade da cultura da soja utilizando dados de clima e solo**. 2019. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo, 2019.

GÓMEZ, D. et al. Potato yield prediction using machine learning techniques and sentinel 2 data. **MDPI**, [S.l.], 2019.

HAGHVERDI, A.; WASHINGTON-ALLEN, R. A.; LEIB, B. G. Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. **Computers and Electronics in Agriculture**, [S.l.], v. 152, p. 186–197, 2018.

HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.: s.n.], 2001.

HAYKIN, S. **Neural networks and learning machines**. [S.l.]: Pearson, 2009, 2009.

INPE. A taxa consolidada de desmatamento por corte raso para os nove estados da amazônia legal em 2020 foi de 10.851 km<sup>2</sup>. **INPE**, [S.l.], 2021.

IORIS, A. A. Cracking the nut of agribusiness and global food insecurity: in search of a critical agenda of research. **Geoforum**, [S.l.], v. 63, p. 1–4, 2015.

JGN ZANNOU, V. H. Sorghum yield prediction using machine learning (redes neurais convolucionais e regressão linear.). **IEEE**, [S.l.], 2019.

JUNIOR, C. C. et al. Artificial neural networks and data mining techniques for summer crop discrimination: a new approach. **Canadian Journal of Remote Sensing**, [S.l.], 2019.

KHAKI, S.; PHAM, H.; WANG, L. Yieldnet: a convolutional neural network for simultaneous corn and soybean yield prediction based on remote sensing data. **arXiv preprint arXiv:2012.03129**, [S.l.], 2020.

KHANAL, S. et al. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. **Computers and electronics in agriculture**, [S.l.], v. 153, p. 213–225, 2018.

KITCHENHAM, B. Procedures for undertaking systematic reviews., **Information and Software Technology**, [S.l.], 2004.

KROSS, A. et al. Evaluation of an artificial neural network approach for prediction of corn and soybean yield. , [S.l.], p. 24–27, 2018.

KUBAT, M. Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7. **The Knowledge Engineering Review**, [S.l.], v. 13, n. 4, p. 409–412, 1999.

L KOUADIO RC DEO, V. B. J. A. Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties (extreme learning machine (elm) ). **ELSEVIER**, [S.l.], 2018.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v. 86, n. 11, p. 2278–2324, 1998.

LEE, S. et al. A self-predictable crop yield platform (scyp) based on crop diseases using deep learning. **MDPI**, [S.l.], 2019.

M CHAMPANERI D CHACHPARA, C. C. Crop yield prediction using machine learning. **International Journal of Science and Research (IJSR) website**, [S.l.], 2020.

MAITINIYAZI MAIMAITIJIANGA VASIT SAGANA, P. S. S. H. F. E. F. B. F. Soybean yield prediction from uav using multimodal data fusion and deep learning. **ELSEVIER Remote Sensing of Environment**, [S.l.], 2019.

MARKO, O. et al. Soybean varieties portfolio optimisation based on yield prediction. **Computers and Electronics in Agriculture**, [S.l.], v. 127, p. 467–474, 2016.

MATHER, J. C. et al. Calibrator design for the cobe\* far infrared absolute spectrophotometer (firas). **The Astrophysical Journal**, [S.l.], v. 512, n. 2, p. 511, 1999.

MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. **IEEE Transactions on geoscience and remote sensing**, [S.l.], v. 42, n. 8, p. 1778–1790, 2004.

MELLO, M. P. et al. Rede bayesiana aplicada ao monitoramento da moratória da soja. **XV Simpósio Brasileiro de Sensoriamento Remoto-SBSR, Curitiba, PR-Brazil**, [S.l.], 2011.

N GANDHI, O. P. Rice crop yield prediction using artificial neural networks. **IEEE**, [S.l.], 2016.

NEPSTAD, D. C.; STICKLER, C. M.; ALMEIDA, O. T. Globalization of the amazon soy and beef industries: opportunities for conservation. **Conservation biology**, [S.l.], v. 20, n. 6, p. 1595–1603, 2006.

NEVAVUORI, P.; NARRA, N.; LIPPING, T. Crop yield prediction with deep convolutional neural networks. **Computers and electronics in agriculture**, [S.l.], v. 163, p. 104859, 2019.

NYÉKI et al. Maize yield prediction based on artificial intelligence using spatio-temporal data. , [S.l.], 2019.

OLIVEIRA APARECIDO, L. E. de et al. Neural networks in climate spatialization and their application in the agricultural zoning of climate risk for sunflower in different sowing dates. **Archives of Agronomy and Soil Science**, [S.l.], v. 17, p. 470–473, 2019.

OLIVEIRA, D. T. d. Um sistema inteligente que prevê a produtividade do algodão em imagens de lavouras comerciais. , [S.l.], 2019.

OQUAB, M. et al. Learning and transferring mid-level image representations using convolutional neural networks. In: **IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION**, 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 1717–1724.

PESQUISAS ESPACIAIS, I. N. de. **Projeto de monitoramento do desmatamento na amazônia legal**. 2022.

PIATTO, M.; SOUZA, L. I. de. 10 anos da moratória da soja na amazônia: história, impactos e a expansão para o cerrado. , [S.l.], 2017.

PRADEEP N, K. S. N. C. G. V. A. S. Modern techniques for agricultural disease management and crop yield prediction. **IGI Global BOOKS**, [S.l.], 2019.

R MEDAR, V. R. Crop yield prediction using machine learning techniques. **IEEE**, [S.l.], 2019.

- ROBBINS, P. **Political ecology: a critical introduction**. [S.l.]: John Wiley & Sons, 2011. v. 16.
- ROGACHEV, A. Monitoring of agricultural land productivity using unmanned aerial vehicles and artificial neural networks. **IOP Conference Series: Earth and Environmental Science**, [S.l.], 2019.
- SANGOI, L. et al. Perfilamento e prolificidade como características estabilizadoras do rendimento de grãos do milho, em diferentes densidades. **Revista Brasileira de Milho e Sorgo**, [S.l.], v. 9, n. 3, p. 254–265, 2010.
- SAYER, J. et al. Ten principles for a landscape approach to reconciling agriculture, conservation, and other competing land uses. **Proceedings of the national academy of sciences**, [S.l.], v. 110, n. 21, p. 8349–8356, 2013.
- SHENDRYK, Y.; DAVY, R.; THORBURN, P. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in australia using machine learning. **Field Crops Research**, [S.l.], v. 260, p. 107984, 2021.
- TERLIKSIZ, A. S.; ALTILAR, D. T. Use of deep neural networks for crop yield prediction: a case study of soybean yield in lauderdale county, alabama, usa. **IEEE, Workshop on Applications of Computer Vision**, [S.l.], 2019.
- USAMA FAYYAD, G. P.-S.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, [S.l.], v. 17, p. 37–54, 1996.
- VAPNIK, V.; GUYON, I.; HASTIE, T. Support vector machines. **Mach. Learn.**, [S.l.], v. 20, n. 3, p. 273–297, 1995.
- VM PURUSHAN, A. S. Effective data analysis and crop yield prediction using machine learning algorithm. **Google Scholar**, [S.l.], 2020.
- WASLEY, A. **Gigantes do agro compram "soja pirata" que acelera desmatamento da amazônia**. 2019.
- YIH, W.-t.; HE, X.; MEEK, C. Semantic parsing for single-relation question answering. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (VOLUME 2: SHORT PAPERS), 52., 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 643–648.
- ZHOU, J. Estimation of the maturity date of soybean breeding lines using uav-based multispectral imagery. **Remote Sensing**, [S.l.], 2019.