

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GEOGRAFIA
CURSO DE ENGENHARIA DE AGRIMENSURA E CARTOGRÁFICA

LAURA CRISTINA MOURA XAVIER

**SOLUÇÃO EM AMBIENTE SIG PARA MAPEAMENTO DA DISTRIBUIÇÃO
ESPACIAL DA PRODUTIVIDADE DE CULTURAS A PARTIR DE RNA E
IMAGENS MULTIESPECTRAIS**

MONTE CARMELO

2021

LAURA CRISTINA MOURA XAVIER

**SOLUÇÃO DE GEOINFORMAÇÃO PARA MAPEAMENTO DA DISTRIBUIÇÃO
ESPACIAL DA PRODUTIVIDADE DE CULTURAS A PARTIR DE RNA E
IMAGENS MULTIESPECTRAIS**

Trabalho apresentado como requisito para obtenção de aprovação na disciplina de Trabalho de Conclusão de Curso II do curso de Engenharia de Agrimensura e Cartográfica da Universidade Federal de Uberlândia campus Monte Carmelo.

Orientador: Prof. Dr. George Deroco Martins.

Coorientador: Vinicius Francisco Rofatto.

MONTE CARMELO

2021

LAURA CRISTINA MOURA XAVIER

**SOLUÇÃO DE GEOINFORMAÇÃO PARA MAPEAMENTO DA DISTRIBUIÇÃO
ESPACIAL DA PRODUTIVIDADE DE CULTURAS A PARTIR DE RNA E
IMAGENS MULTIESPECTRAIS**

Trabalho apresentado como requisito para obtenção de aprovação na disciplina de Trabalho de Conclusão de Curso II do curso de Engenharia de Agrimensura e Cartográfica da Universidade Federal de Uberlândia campus Monte Carmelo.

Aprovado em: ____/____/____

Banca Examinadora:

Prof. Dr. Gorge Deroco Martins (orientador)
Universidade Federal de Uberlândia (UFU)

Prof. Dr. Vinicius Francisco Rofatto (coorientador)
Universidade Federal de Uberlândia (UFU)

Prof. Dr. João Vitor Meza Bravo
Universidade Federal de Uberlândia (UFU)

MONTE CARMELO

2021

AGRADECIMENTOS

Agradeço primeiramente a minha família, em especial a minha mãe Lindalva que sempre me apoiou e incentivou em todas as etapas da vida.

Ao meu companheiro Guilherme que nos últimos anos tem enfrentado a caminhada da vida comigo e tornando-a completa e feliz.

Aos meus grandes amigos que fiz durante a graduação e que vão permanecer eternamente na minha vida. Primeiramente ao Arthur pelo qual tenho um carinho de irmão, e que sempre foi um grande parceiro durante todo esse percurso; Caio por estar comigo desde o primeiro dia, torcendo pelo meu sucesso, e pela oportunidade de termos uma amizade duradoura; Milena por ter cuidado tão bem de mim em todos os anos de república e ser a melhor companhia possível e; Rafael por ser tão presente em minha vida e por compartilhar vários momentos especiais e memoráveis durante a graduação.

Aos demais colegas da 11ª Turma do Curso de Engenharia de Agrimensura e Cartográfica que de alguma forma fizeram parte deste percurso.

Ao meu orientador professor Dr. Geoge Deroco Martins por ser meu principal incentivador na pesquisa e por conseguir executar tão bem seu papel em todos estes anos que trabalhamos juntos.

Ao meu coorientador professor Dr. Vinicius Francisco Rofatto e o professor Dr. João Vitor Meza Bravo, por serem excelentes professores e pela honra de fazerem parte deste trabalho.

A empresa Lallemand Plant Care que forneceu os dados que possibilitaram a execução deste trabalho.

RESUMO

A estimativa da produtividade de culturas é extremamente desafiadora devido à sua relação com vários fatores complexos como as condições ambientais, estrutura física da cultura, composição química do solo e práticas de manejo de campo. Dessa forma, se torna necessária a exploração de técnicas de pré-processamento dos dados coletados, modelagem e avaliação das ferramentas de previsão dessa variável. Recentemente estão sendo desenvolvidos trabalhos que utilizam de técnicas de aprendizado de máquina e dados advindo do sensoriamento remoto para aplicações na agricultura, e mais recentemente para estimativa de variáveis agrícolas. Nesse sentido, o objetivo deste trabalho é desenvolver uma ferramenta em software SIG de código aberto (QGIS) para mapear a distribuição espacial da produtividade das culturas de milho, trigo e girassol utilizando o algoritmo de Redes Neurais Artificiais (RNA) e imagens multiespectrais. Inicialmente foram adquiridos os dados de produtividade das três culturas em estudo, e imagens multiespectrais gratuitas advindas do satélite Sentinel 2. Em seguida, para compor os dados de entrada para o algoritmo de RNA foi realizado o pré-processamento dos dados que compreendeu a extração da reflectância das bandas e os cálculos dos índices de vegetação (NDVI, GNDVI e VARI), além da organização dos arquivos de entrada para treinamento, validação e teste dos modelos de RNA. Por fim, foi desenvolvido o código de RNA, em que os modelos que apresentaram menor RMSE foram implementados na ferramenta desenvolvida em ambiente SIG. Os resultados foram satisfatórios para as culturas de trigo e girassol, atingindo o objetivo de terem RMSE de validação interna menores que 40% e validação externa menores que 30%, porém, para a cultura de milho, os modelos testados não foram capazes de realizar generalizações para os dados de teste. Conclui-se que os modelos de RNA implementados na ferramenta atenderam ao objetivo desejado, porém necessitam de constante atualização visando aumentar o poder de generalização, além da necessidade de realizar estudos mais aprofundados em relação a cultura de milho a fim de obter um modelo apto para estimativa da produtividade.

Palavras-chave: produtividade de culturas; RNA; sensoriamento remoto; SIG.

LISTA DE FIGURAS

Figura 1 - Curva espectral de uma folha verde sadia	10
Figura 2 - Modelo de um neurônio em uma RNA.....	13
Figura 3 - Estrutura geral de um SIG	14
Figura 4 - Mapa das áreas de estudo localizadas no Brasil	16
Figura 5 - Mapa das áreas de estudo localizadas na França	17
Figura 6 - Fluxograma da metodologia do trabalho	18
Figura 7 - Modelo lógico da ferramenta auxiliar.....	24
Figura 8 - Interface de execução da ferramenta auxiliar	25
Figura 9 - Exemplo das entradas e saídas geradas pela ferramenta auxiliar	26
Figura 10 - Grafo da arquitetura das RNAs alimentadas diretamente com múltiplas camadas.....	30
Figura 11 - Gráfico das funções de ativação ReLU e linear.....	34
Figura 12 - Modelo lógico da ferramenta.....	41
Figura 13 - Interface de execução da ferramenta: parte 1	42
Figura 14 - Interface de execução da ferramenta: parte 2	43
Figura 15 - Exemplo dos resultados gerados após a execução da ferramenta em ambiente QGIS	44
Figura 16 - Superfícies de produtividade geradas a partir da ferramenta desenvolvida.....	45
Figura 17- Gráficos de dispersão relacionando a produtividade prevista e medida com a correlação e RMSE da previsão.....	46
Figura 18 - Adicionando a ferramenta como ferramenta de procesamento QGIS	49
Figura 19 - Instalação da ferramenta como plugin do QGIS.....	49

LISTA DE TABELAS E QUADROS

Tabela 1 - Caracterização das áreas de estudo	15
Tabela 2 - Especificações das bandas do satélite Sentinel 2	20
Tabela 3 - Informações sobre as imagens Sentinel 2 adquiridas para o trabalho	21
Tabela 4 - Quantidade de pontos nos dados originais e depois da utilização da ferramenta auxiliar	27
Tabela 5 - Valores dos hiperparâmetros testados	36
Tabela 6 - Valores dos hiperparâmetros e erro de treinamento dos modelos de RNA selecionados para cada cultura	38
Tabela 7 - Estatística descritiva das áreas teste	47
Quadro 1 - Detalhamento da ferramenta auxiliar	25
Quadro 2 - Divisão do banco de dados	28

LISTA DE EQUAÇÕES

Equação 1 – NDVI.....	22
Equação 2 – GNDVI.....	22
Equação 3 – VARI.....	22
Equação 4 – Normalização linear.....	30
Equação 5 – Glorot Normal.....	32
Equação 6 – Função de ativação ReLU.....	32
Equação 7 – Função de ativação linear.....	32
Equação 8 – MSE.....	33
Equação 9 – RMSE.....	35

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ARP	Aeronave Remotamente Pilotada
BOA	Reflectância de fundo da atmosfera
GNDVI	Vegetação de Diferença Normalizada Verde
GNSS	Sistema Global de Navegação por Satélite
IA	Inteligência Artificial
IAF	Índice de Área Foliar
IV	Índice de Vegetação
LOO	<i>Leave-One-Out</i>
MLP	<i>Multi Layer Perceptron</i>
MSE	<i>Mean Square Error</i>
NDVI	Índice de Vegetação da Diferença Normalizada
OSGEO	<i>Open Source Geospatial Foundation</i>
PAT	<i>Precision Agriculture Tools</i>
ReLU	<i>Rectified Linear Unit</i>
RMSE	<i>Root Mean Squared Error</i>
RNA	Rede Neural Artificial
S2A	Sentinel 2A
S2B	Sentinel 2B
SIG	Sistema de Informação Geográfica
SR	Sensoriamento Remoto
TOA	Reflectância do topo da atmosférica
VARI	Índice Resistente à Atmosfera na Região Visível

SUMÁRIO

1	INTRODUÇÃO	4
2	OBJETIVOS	6
3	JUSTIFICATIVA	7
4	FUNDAMENTAÇÃO TEÓRICA	9
	4.1 Sensoriamento remoto aplicado a agricultura de precisão	9
	4.2 Estimativa de produtividade de cultivos por meio do sensoriamento remoto	11
	4.3 Redes Neurais Artificiais	12
	4.4 Sistema de Informação Geográfica	13
5	MATERIAL E MÉTODOS	15
	5.1 Áreas de estudo	15
	5.2 Material	17
	5.3 Métodos	18
	5.3.1 Aquisição de dados	18
	5.3.2 Pré-processamento de dados	21
	5.3.3 Construção da ferramenta	28
6	RESULTADOS	38
	6.1 Arquitetura do modelo de RNA	38
	6.2 Ferramenta desenvolvida	41
	6.3 Aplicação da ferramenta	45
	6.4 Disponibilização	48
7	CONCLUSÕES	50
	REFERÊNCIAS	51

1 INTRODUÇÃO

A agricultura desempenha um papel importante na economia do mundo, sendo uma das principais atividades de arrecadação de impostos, geração de renda e empregos. Com o cenário enfrentado devido a pandemia da Covid-19 foi evidenciado que a ciência e tecnologia desempenham um papel importante no setor agrícola (HENRY, 2020). Uma vez que, o investimento em pesquisa e desenvolvimento apoia a segurança alimentar, permitindo fornecer capacidade de produção local segura em futuras pandemias (HENRY, 2020).

Incentivos em pesquisas na agricultura buscam, principalmente, estratégias para aumento de produtividade, visando aumentar o fornecimento de alimentos capaz de atender à demanda mundial adicional (FAO, 2015). As soluções agrícolas exigem disponibilidade e precisão de dados para auxiliar os agricultores, como é o caso da agricultura inteligente. Esta integra um conjunto de tecnologias, dispositivos, protocolos e paradigmas computacionais para melhorar os processos agrícolas utilizando como ferramentas o *big data*, Inteligência Artificial - IA, armazenamento em nuvem e computação avançada (ZANELLA et al., 2020).

Trabalhos ligados a produção de culturas têm proposto o subsídio do Sensoriamento Remoto (SR) com o uso de imagens multiespectrais de alta e média resolução espacial para geração de modelos de estimativa e mapeamento da produtividade (BERTOLIN et al., 2017; HE et al., 2018; LEDA et al., 2019). Dentre estas, Varella et al. (2011) destacam as que fazem uso de relações entre índices de vegetação, reflectância no comprimento de onda do infravermelho próximo e a produtividade da cultura.

Técnicas de aprendizado de máquina, IA e ciência de dados permitem o desenvolvimento de modelos computacionais para tomada de decisão, gerenciamento de doenças e pragas, investigação de nichos agrícolas entre outros (BESTELMEYER et al., 2020). Recentemente, tem-se utilizado algoritmos de aprendizado de máquina em dados de índices de vegetação derivados de imagens de satélite para estimar a produção agrícola (AL-GAADI et al., 2016; SKAKUN et al., 2018; GASO et al., 2019).

No contexto de aprendizado de máquina, as Redes Neurais Artificiais (RNAs) são técnicas computacionais que apresentam um modelo similar a estrutura neural de organismos inteligentes que adquirem conhecimento através da experiência. Braga et al. (2000) ressaltam que, dessa forma, sua aplicação é bastante eficiente na análise espacial, uma vez que a RNA é capaz de identificar padrões a partir de treinamento e generalizar a informação aprendida, gerando um modelo não-linear capaz de discretizar variáveis.

A estimativa da produtividade das culturas é extremamente desafiadora devido a vários fatores complexos como condições ambientais, estrutura física da cultura, composição química do solo e práticas de manejo de campo (GUIMARÃES, 2019). Sendo assim, é necessária a exploração de técnicas de pré-processamento dos dados coletados, modelagem e avaliação das ferramentas de previsão (GUIMARÃES, 2019). Nesse sentido, os Sistemas de Informação Geográfica (SIGs) vem ganhando destaque, uma vez que os softwares possuem ferramentas que permitem gerir, analisar e manipular dados geográficos, além de possibilitar a criação de plugins e modeladores para a automatização de procedimentos (SILVA, 2017).

O sensoriamento remoto aliado as técnicas de aprendizado de máquina têm sido utilizados para estimar variáveis agrícolas (VIRNODKAR et al., 2020); softwares comerciais ligados a agricultura recentemente têm começado a implementar estes conceitos para disponibilização de produtos para o usuário, como o Agriquest®, Coptical® e Farmsat®, desenvolvido pela empresa Geosys, FalkerMap, desenvolvido pela empresa Falker, e aplicativos como SIMA e XARVIO™. Sendo assim, este trabalho verificou a possibilidade destas técnicas serem utilizadas para estimar a variável produtividade de culturas e apresentar uma ferramenta gratuita e de fácil utilização.

2 OBJETIVOS

O objetivo geral do trabalho é propor uma solução gratuita para o estimativa e mapeamento de produtividade de culturas utilizando RNAs e imagens multiespectrais. Sendo assim, nesse trabalho, os objetivos específicos são:

- desenvolver uma ferramenta em ambiente SIG de código aberto para mapear a distribuição espacial da produtividade de culturas;
- avaliar o desempenho do algoritmo de RNAs para estimativa da variável produtividade de culturas;
- modelar a produtividade com RMSE menor que 30%; e
- disponibilizar uma ferramenta que permita verificar a acurácia dos dados gerados.

3 JUSTIFICATIVA

O uso da Inteligência Artificial (IA) para resolução de problemas complexos vem ganhando força nos últimos tempos, e uma das áreas que começou a se beneficiar deste tipo de tecnologia é a agricultura (BANNERJEE et al., 2018). As RNAs têm apresentado elevado desempenho devido a sua estrutura robusta em camadas, eficiência de aprendizado e generalização, o que as tem tornado eficazes na resolução de problemas complexos (HAYKIN, 2001).

Trabalhos ao longo desta última década foram desenvolvidos visando a adequação e parametrização de RNAs para aplicações na agricultura, como a modelagem de diâmetro (BINOTI et al., 2013), caracterização da água do solo (SOARES et al., 2014), modelagem da produtividade de grãos (SILVA et al., 2014), predição da produtividade por meio de atributos de solo (LEAL et al., 2015), estimativa espacial do volume de madeira empregando bandas espectrais, texturas e índices de vegetação (ARTEAGA, 2019).

Por meio de técnicas de sensoriamento remoto, as quais apresentam elevado potencial para utilização na agricultura, é possível adquirir informações agronômicas das culturas, condições fenológicas e nutricionais, conseqüentemente relacionadas com a produtividade (RUDORFF; MOREIRA, 2002). O sensoriamento remoto, gera uma grande quantidade de dados, e eles estão cada vez mais acessíveis e em melhoramento constante no que se refere a resolução espacial, temporal e espectral (HUANG et al., 2018). A utilização de sensoriamento remoto evita a amostragem convencional sistemática de campo, onerosa e lenta, a qual demanda mão de obra, e além disso, realizada nos dias da colheita, o que impede ações prévias de planejamento administrativo (BERTOLIN et al., 2017).

Jung et al. (2020) destacaram, em seu trabalho de revisão, que os avanços tecnológicos em curso representam a oportunidade para enfrentar desafios das próximas décadas ligados a produção agrícolas. Além disso, ressaltam que os avanços na agricultura digital se beneficiarão da integração de dados de sensoriamento remoto, modelos avançados de simulação de safras e IA.

Seelan et al. (2003) revisaram que, a agricultura de precisão requer informações sobre as condições da cultura em alta resolução espacial. Dessa forma, ressaltam que os agricultores

que utilizam a agricultura de precisão estão familiarizados com as tecnologias SIG e *Global Navigation Satellite System* - GNSS, porém não possuem o treinamento necessário para extrair informações de imagens, além do mais, os softwares existentes para esta finalidade são caros e não integrados com outros softwares já utilizados pelos agricultores.

Considerando a complexidade para estimativa da produtividade de culturas como condições ambientais, estrutura física da cultura, composição química do solo e práticas de manejo de campo, ressaltada por Guimarães (2019), a variável produtividade de culturas ainda carece de metodologias para sua previsão e mapeamento, ainda mais no que se refere a uma ferramenta de acesso livre em ambiente SIG. Para superar este problema, neste trabalho, nós propomos uma ferramenta em software SIG de código aberto e de fácil utilização para a previsão da produtividade de culturas. Esta ferramenta conta com um modelo de previsão baseado em RNAs e imagens multiespectrais.

4 FUNDAMENTAÇÃO TEÓRICA

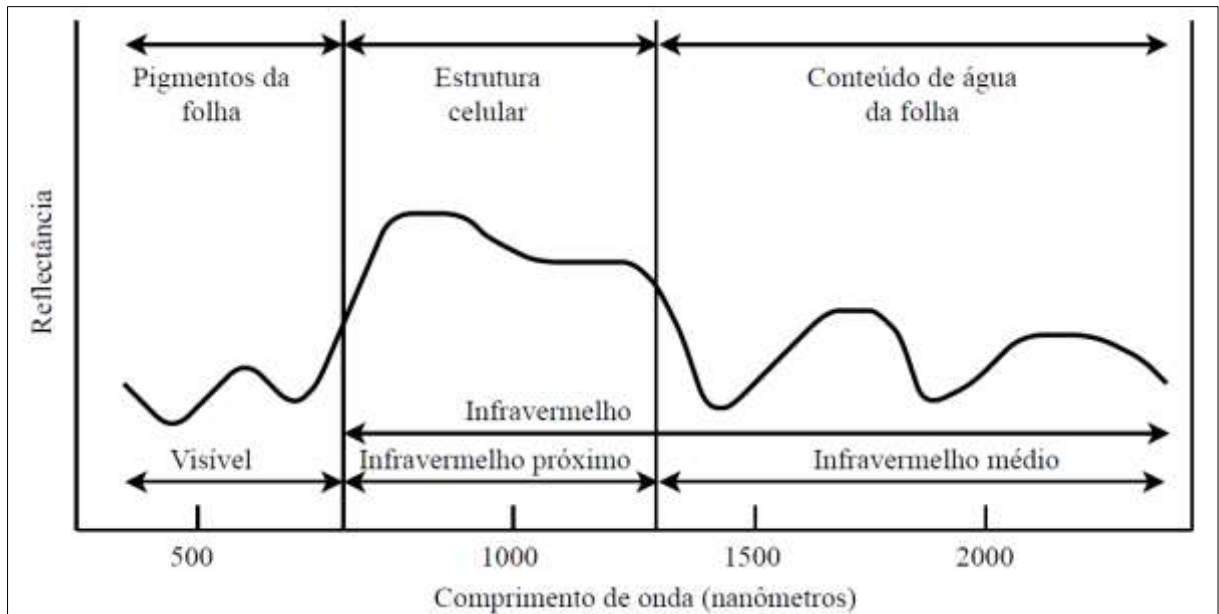
O trabalho realizado trata-se do desenvolvimento de uma ferramenta em ambiente SIG de código aberto para previsão da produtividade de culturas utilizando RNA e imagens multiespectrais. Desta forma, neste tópico são apresentados de forma geral conceitos e aplicações sobre os principais temas deste trabalho. Na seção **4.1 Sensoriamento remoto aplicado a agricultura de precisão** é apresentada a relação do sensoriamento remoto com a vegetação e suporte à agricultura de precisão, plataformas utilizadas e aplicações; na seção **4.2 Estimativa de produtividade de cultivos através do sensoriamento remoto** são apresentadas as vantagens da utilização do sensoriamento remoto para previsão da produtividade de culturas, e a larga utilização de índices de vegetação para este fim; na seção **4.3 Redes Neurais Artificiais** são apresentadas definições, aplicações e características das RNAs e; na seção **4.4 Sistema de Informação Geográfica** é apresentada sua definição, estrutura, recursos e ferramentas já implementadas relacionadas à agricultura.

4.1 Sensoriamento remoto aplicado a agricultura de precisão

As aplicações de sensoriamento remoto na agricultura são baseadas na interação da radiação eletromagnética com o solo ou material vegetal, e para realizar sua medição podem ser utilizadas plataformas como satélites, aeronaves e sensores portáteis (MULLA, 2013). Estas plataformas e seus sistemas de imagem são caracterizados baseando-se na altitude da plataforma, resolução espacial e temporal da imagem.

A quantidade de radiação refletida das plantas está inversamente relacionada à radiação absorvida pelos pigmentos das plantas e varia com o comprimento de onda da radiação incidente (PINTER et al., 2003). Especificamente na vegetação, a reflectância é alta na região do infravermelho próximo (700 – 1300 nanômetros) devido a densidade da folha e dos efeitos da estrutura do dossel (JENSEN, 2009). A Figura 1 apresenta a curva espectral de uma folha verde sadia.

Figura 1 - Curva espectral de uma folha verde sadia



Fonte: adaptado de Novo (2010).

Esta característica específica, alta reflectância entre o comprimento de onda do espectro vermelho e infravermelho próximo, motivou pesquisadores no desenvolvimento de índices espectrais de vegetação que são baseados em razões de valores de reflectância nas regiões visível e infravermelho próximo (SRIPADA et al., 2006).

Geralmente, as aplicações de sensoriamento remoto na agricultura são realizadas para fins como a verificação do rendimento de culturas e biomassa (SHANAHAM et al., 2001), nutrientes da cultura e estresse hídrico (CALERA et al., 2017), infecções de ervas daninhas (THORP; TIAN, 2004; MAHLEIN, 2016) entre outros.

Especificamente se tratando da agricultura de precisão, ela pode ser amplamente definida como um sistema agrícola no qual o manejo da cultura é realizado no lugar, intensidade e momento certos (MULLA, 2013). Este ramo da agricultura é baseado em novas ferramentas e fontes de informação ligadas as tecnologias modernas. Dessa forma, segundo Zhang et al. (2005), dentre estas tecnologias estão incluídos os sistemas GNSS, os SIGSs dispositivos de monitoramento de produção, sensores de solo, plantas e pragas, sensoriamento remoto e tecnologias de taxa variável para aplicadores de insumos.

No que se refere as plataformas de sensoriamento remoto que oferecem produtos de maior resolução espacial, as Aeronaves Remotamente Pilotadas (ARPs) aplicadas a agriculturas são ferramentas poderosas. As ARPs apresentam baixos custos operacionais, porém, o custo de processamento de dados aumenta à medida que o volume de dados aumenta exponencialmente para cobrir áreas maiores (WEISS et al., 2020). Se tratando de plataformas orbitais (satélites), há uma quantidade significativa de pesquisas indicando a popularidade dos dados de satélite para a agricultura de precisão (WEISS et al., 2020). No entanto, pouca atenção tem sido dada em como adaptá-los para aplicações de agricultura de precisão de escala apropriada (DUVEILLER; CESCATTI, 2016).

Existe um potencial significativo na agricultura de precisão para combinar dados de sensoriamento remoto com dados de culturas visando o melhor gerenciamento agrícola (THENKABAIL, 2003). Uma vez que as aplicações de agricultura de precisão requerem informações em uma escala precisa, há um desafio significativo na adaptação de metodologias em diferentes escalas (DUVEILLER; CESCATTI, 2016). Sendo assim, a aquisição, processamento, armazenamento, análise e visualização de produtos agrícolas gerados pelo *big data* de sensoriamento remoto são essenciais para o sucesso da agricultura de precisão inteligente (WOLFERT et al., 2017).

4.2 Estimativa de produtividade de cultivos por meio do sensoriamento remoto

A produtividade de um cultivo agrícola é a relação entre a produção da cultura e a unidade de área. Dessa forma, mapas de produtividade apresentam, especialmente, as áreas de maior e menor intensidade de produção, além de serem úteis para prever lucros e realizar investigações relacionadas ao manejo preciso da área. O sensoriamento remoto apresenta alguns benefícios em relação aos métodos baseados no solo para predição da produção de culturas, como por exemplo, uma maior cobertura espacial e temporal (JUDGE, 2007), redução de custos (BASSO et al., 2013), e minimização de erros grosseiros (relacionados a humanos) (CHIPANSHI et al., 2015).

O sensoriamento remoto para predição da produtividade depende das propriedades espectrais da cultura que variam de acordo com o estágio de crescimento da cultura, tipo de cultura e seu estado de saúde (KARTHIKEYAN et al., 2020). Os Índices de Área Foliar (IAF), parâmetros meteorológicos como temperatura, precipitação e evapotranspiração e vários Índices de Vegetação (IV), são comumente derivados do sensoriamento remoto e utilizados em modelos de predição de produtividade (KOGAN et al., 2013).

Destes, segundo Peng et al. (2014), os IV são mais utilizados, uma vez que também fornecem informações sobre a saúde da cultura e condições de seca. IV derivados de dados de sensoriamento remoto visível e infravermelho próximo utilizam apenas informações de uma pequena parte do espectro eletromagnético, enquanto outras bandas espectrais disponíveis foram comparativamente menos estudadas e podem fornecer informações únicas e/ou complementares para avaliações de safra (GUAN et al.,2017).

Alguns trabalhos vêm relacionando IV derivados do sensoriamento remoto e informações adicionais para predição da produtividade. Skakun et al. (2017) usaram IV juntamente com informações de temperatura do ar para realizar o mapeamento da safra de inverno no Kansas – EUA e na Ucrânia. Noland et al. (2018) estimaram a produção de alfafa usando índices espectrais derivados do sensoriamento remoto e informações da temperatura do ar. Tuvdendorj et al. (2019) descobriram uma regressão linear para produtividade de trigo na primavera combinando IV e índices de seca.

4.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são modelos matemáticos inteligentes semelhantes ao sistema nervoso humano, que permitem adquirir conhecimento através do processamento de informações. Elas podem ser definidas como estruturas formadas por elementos de processamento simples adaptativos e interconectados, os chamados neurônios ou nós artificiais, capazes de realizar cálculos para processamento de dados e representação do conhecimento (HAYKIN, 2001).

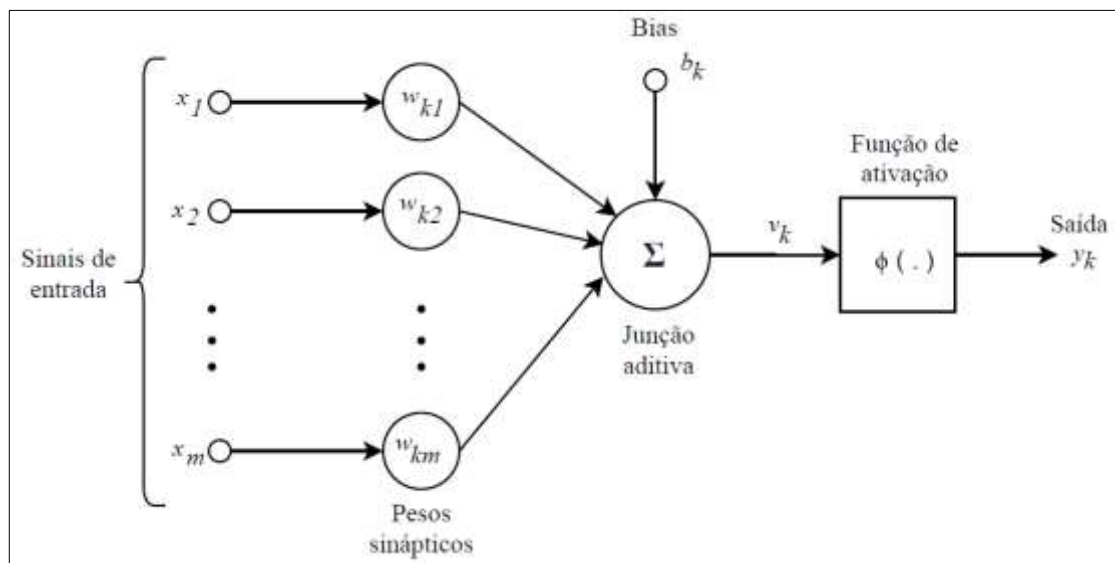
Basheer e Hajmeer (2000), ressaltam que a ideia das RNAs não é replicar os sistemas biológicos, mas utilizar o conhecimento sobre suas funcionalidades para resolver problemas complexos. Elas se assemelham ao cérebro humano em dois aspectos: o conhecimento é adquirido através de um processo de aprendizado e as forças de conexões entre neurônios, os pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido (HAYKIN, 2001).

Jain et al. (1996) apontaram que as RNAs são atrativas por diversos motivos como não linearidade, alto paralelismo, robustez, tolerância a falhas, capacidade de aprender e generalizar entre outras. Estas características são atrativas porque permitem um melhor ajuste aos dados, já que não são lineares, a insensibilidade ao ruído fornece uma previsão precisa na presença de dados incertos ou erros de medição, a aprendizagem e adaptabilidade permitem que o sistema atualize sua estrutura interna em resposta a mudanças no ambiente e a generalização permite a aplicação de modelos a dados não aprendidos (BASHEER; HAJMEER, 2000). Nos últimos

anos houve cada vez mais avanços nas aplicações das RNAs para resolver problemas de classificação, reconhecimento de padrões, regressão e previsão de problemas (CJATTERJEE et al., 2017).

Um neurônio é uma unidade de processamento de informação fundamental para a operação de uma rede neural formada por três elementos básicos: (i) um conjunto de sinapses (elos de conexão) caracterizadas por pesos; (ii) um somador para os sinais de entrada ponderados pelas respectivas sinapses de um neurônio e; (iii) uma função de ativação para restringir a amplitude da saída de um neurônio (HAYKIN, 2001). O modelo neural inclui também um *bias* aplicado externamente, para aumentar ou diminuir a entrada líquida da função de ativação, dependendo do sinal positivo ou negativo, respectivamente (HAYKIN, 2001). A Figura 2 apresenta o modelo de um neurônio em uma RNA.

Figura 2 - Modelo de um neurônio em uma RNA



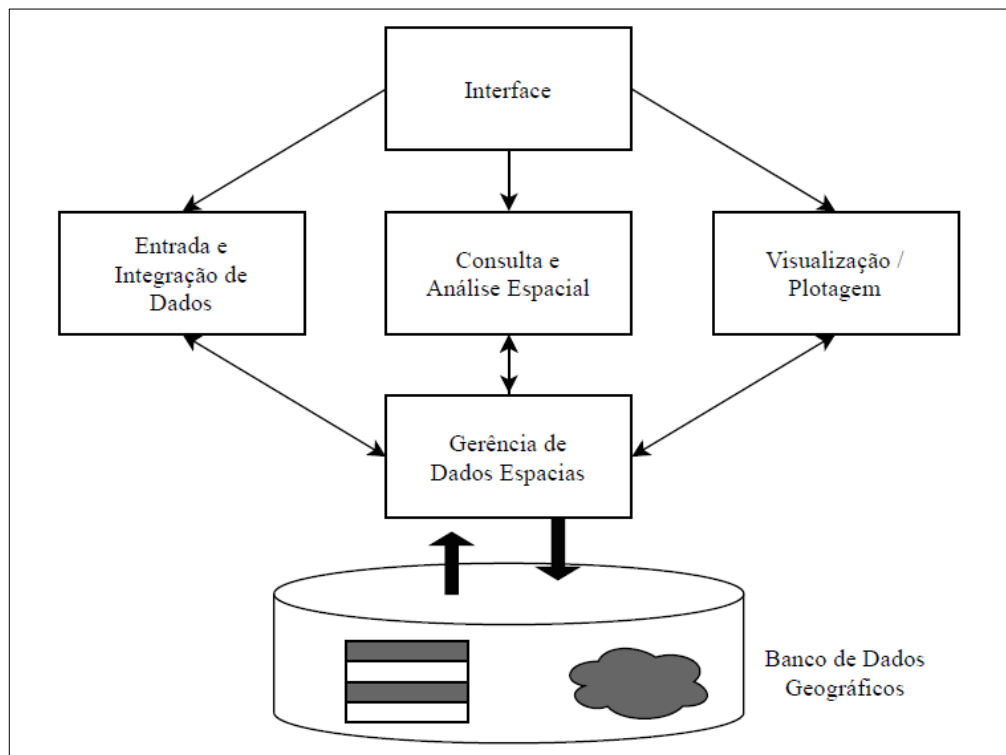
Fonte: adaptado de Haykin (2001).

4.4 Sistema de Informação Geográfica

Dos diversos autores que definiram os Sistemas de Informação Geográfica (SIGs), Bourrough (1986) definiu os SIGs como um conjunto poderoso de ferramentas destinadas a coletar, armazenar, recuperar, transformar e visualizar dados sobre o mundo real. Aranoff (1989) complementou que a localização geográfica é uma característica indispensável para análise das informações de um objeto ou fenômeno espacial em um ambiente SIG.

De acordo com Câmara e Freitas (1995), a estrutura técnica de um SIG possui três níveis: (i) interface com o usuário, (ii) entrada e integração de dados, funções de processamento gráfico de imagens, visualização e plotagem e (iii) armazenamento e recuperação de dados, organizados sob forma de um banco de dados geográficos. Os SIGs fornecem análises além da entrada de dados, gerenciamento, manipulação e consulta de dados. A Figura 3 ilustra a estrutura geral de um SIG.

Figura 3 - Estrutura geral de um SIG



Fonte: adaptado de Câmara e Freitas (1995).

Considerando o mercado de *softwares* livres e de distribuição gratuita, o QGIS é o SIG mais utilizado. Ele possui complementos disponíveis para geoprocessamento, possibilidade de implementação de algoritmos pelos próprios usuários e integração com outros SIGs e plataformas que possuem diversas bibliotecas relacionadas à geoestatística, geoprocessamento e mineração de dados. No *software* QGIS, um *plugin* é uma ferramenta que permite adicionar funcionalidades ao *software* ou automatizar rotinas com algoritmos existentes e pode ser implementado de forma gratuita utilizando as linguagens Python e C++ (SILVA, 2017).

O QGIS é frequentemente usado para análises espaciais estratégicas na agricultura, utilizando informações geográficas para tomada de decisões. Um exemplo de *plugin* do QGIS é o *Precision Agriculture Tools – PAT* desenvolvido para processamento de dados de agricultura de precisão. Ele inclui ferramentas para limpeza de dados e interpolação de mapas

para geração de superfícies a partir de dados de sensor de alta densidade espacial, como monitor de produtividade ou dados de sensor de solo EM38 (QGIS, 2020). Este *plugin* também possui ferramentas para análise de imagens de sensoriamento remoto e para agrupamento de vários conjuntos de dados (QGIS, 2020).

5 MATERIAL E MÉTODOS

Neste tópico são apresentadas as seguintes seções: **5.1 Área de Estudo** seção que apresenta as características das áreas de estudo e mapas de localização; **5.2 Material** seção que apresenta os dados e softwares utilizados e; **5.3 Métodos** seção que apresenta os procedimentos e parâmetros utilizados para desenvolvimento da ferramenta proposta por este trabalho.

5.1 Áreas de estudo

As áreas de estudo estão distribuídas no Brasil e na França. No Brasil são 2 áreas com cultura de milho, concentradas na região sudeste do país, no Estado de Minas Gerais, nos municípios de Patos de Minas e Presidente Olegário. Conforme a classificação de Köppen, o clima destes municípios é o Tropical de Savana – Aw, ou seja, possui clima tropical, invernos secos, verões quentes e a temperatura média do mês mais frio superior a 18°C.

Na França são 9 áreas com culturas variando entre milho, trigo e girassol, concentradas nas regiões administrativas de Centre, Rays de la Lone e Bretagne, e departamentos Loir-et-Cher, Loiret, Maine-et-Loire e Morbihan. Conforme a classificação de Köppen, o clima destes departamentos é o Oceânico Temperado – Cfb, ou seja, possui clima temperado, com verões amenos, chuvas uniformemente distribuídas, sem estação seca e temperatura média do mês mais quente inferior a 22° C.

A Tabela 1 apresenta uma breve caracterização das áreas de estudo deste trabalho.

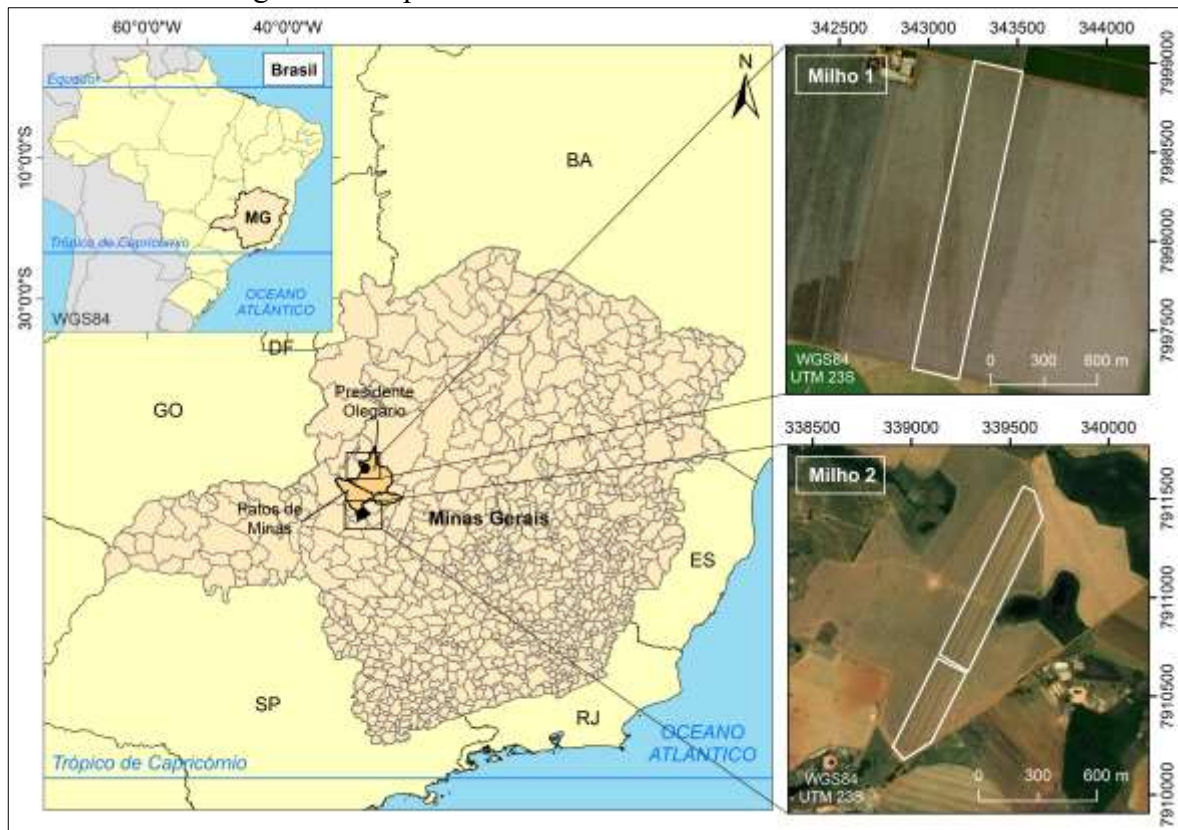
Tabela 1 - Caracterização das áreas de estudo

Cultura	Nome	Localização	Área (ha)
milho	Milho 1	Brasil, Estado de Minas Gerais, Município de Presidente Olegário	46
	Milho 2	Brasil, Estado de Minas Gerais, Município de Patos de Minas	25
	Milho 3	França, Região Centre, departamento de Loir-et-cher	12
	Milho 4	França, Região Centre, departamento de Loiret	6
	Milho 5	França, Região Centre, departamento de Loir-et-cher	12
	Milho 6	França, Região Centre, departamento de Loir-et-cher	11

trigo	Trigo 1	França, Região Bretagne, departamento de Morbihan	9
	Trigo 2	França, Região Centre, departamento de Loir-et-cher	17
	Trigo 3	França, Região Pays de la Lone, departamento de Maine-et-Loire	4
girassol	Girassol 1	França, Região Centre, departamento de Loiret	6
	Girassol 2	França, Região Pays de la Lone, departamento de Maine-et-Loire	5

A Figura 4 apresenta o mapa das áreas localizadas no Brasil.

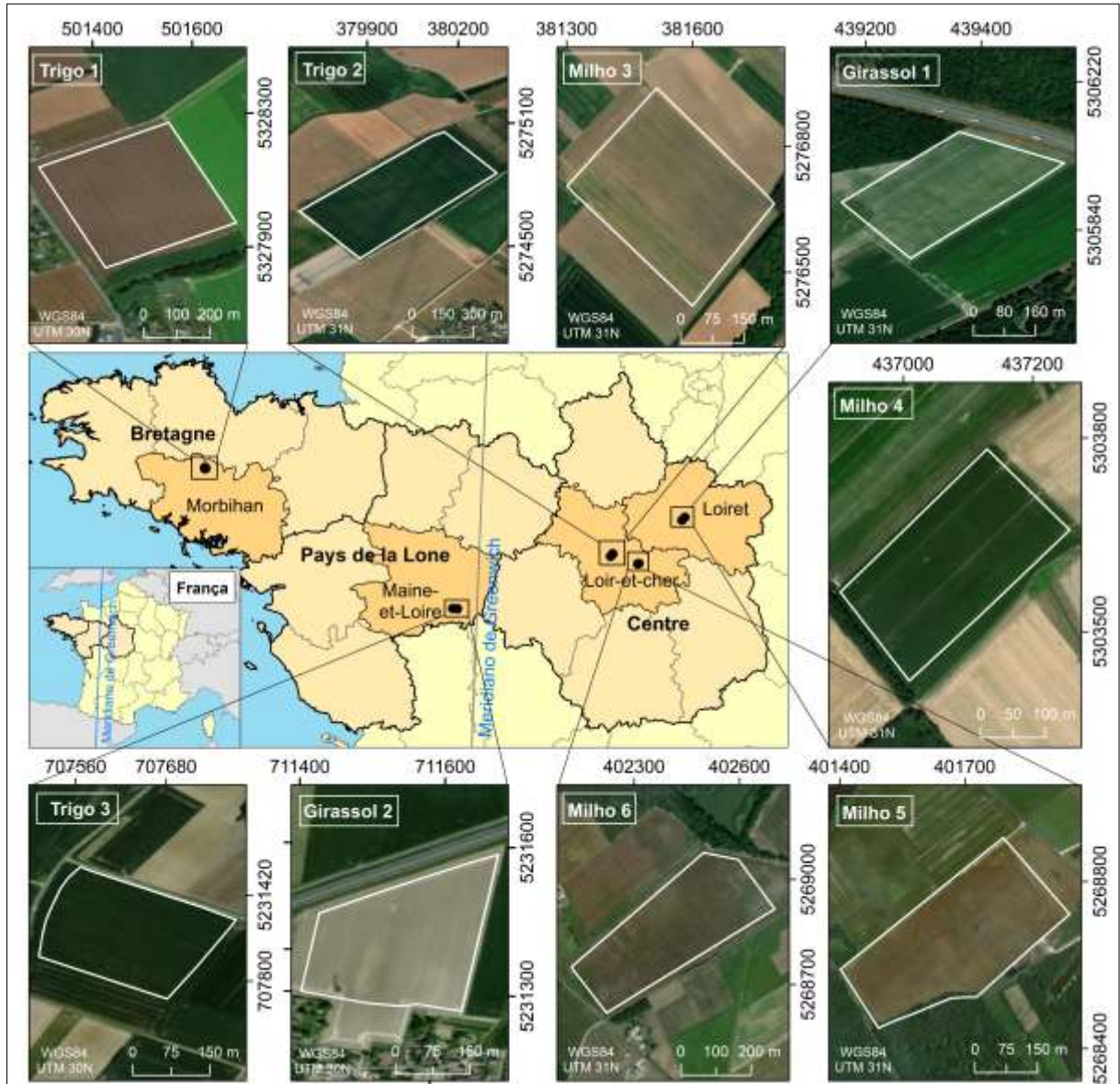
Figura 4 - Mapa das áreas de estudo localizadas no Brasil



Elaboração: a autora.

A Figura 5 apresenta o mapa das áreas de estudo localizadas na França.

Figura 5 - Mapa das áreas de estudo localizadas na França



Elaboração: a autora.

5.2 Material

O material para desenvolvimento deste estudo consistiu em dados para construção da RNA e softwares para manipulação destes dados e implementação da ferramenta.

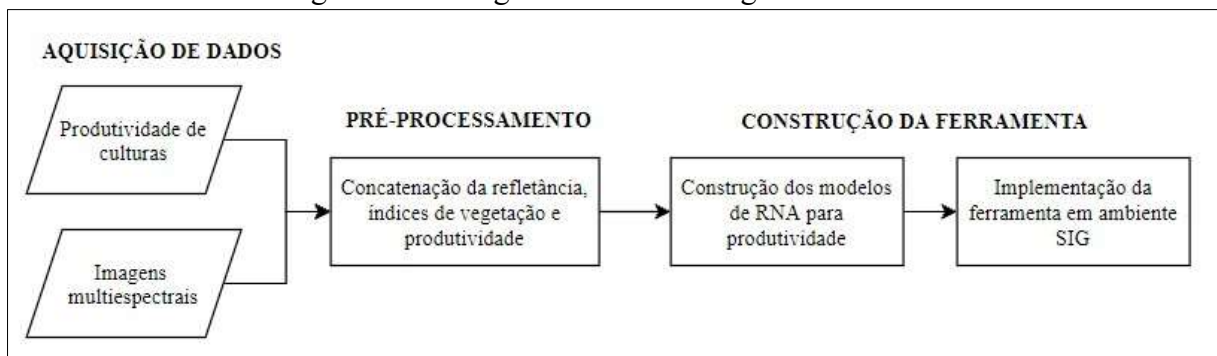
Os dados para construção do modelo de RNA para previsão foram informações da produtividade de culturas em forma de um banco de dados (tabelas), referentes as culturas de milho, trigo e girassol, fornecidas pela empresa Lallemand Plant Care e; imagens multiespectrais do satélite Sentinel 2, obtidas de forma gratuita.

Os softwares utilizados para procedimentos e/ou processamentos específicos ligados a manipulação dos dados e desenvolvimento da ferramenta foram: editor de planilhas Excel desenvolvido pela Microsoft utilizado para organização inicial do banco de dados; serviço de armazenamento em nuvem de notebook Google Collaboratory utilizado para criação e execução de códigos Python a fim de construir o código da RNA e gerar os modelos de previsão e; SIG de código-fonte aberto QGIS versão 3.14.1 utilizado para extração de informações das imagens multiespectrais, e implementação da ferramenta.

5.3 Métodos

A metodologia que este trabalho seguiu está descrita na Figura 6. O fluxograma apresenta a divisão dos procedimentos metodológicos em três etapas sendo elas: (i) Aquisição de dados; (ii) Pré-processamento de dados e; (iii) Construção da ferramenta.

Figura 6 - Fluxograma da metodologia do trabalho



Elaboração: a autora.

5.3.1 Aquisição de dados

Modelos desenvolvidos a partir de dados geralmente dependem do tamanho do banco de dados, assim, as RNAs, como outros modelos empíricos, podem ser obtidos de bancos de dados de qualquer tamanho, no entanto, a generalização desses modelos para dados fora do domínio de desenvolvimento do modelo será afetado adversamente (BASHEER; HAJMMER,

2000). Assim, os dados usados no treinamento da RNA foram suficientemente grandes para cobrir a possível variação conhecida no domínio do problema. Nos subtópicos 5.3.1.1 e 5.3.1.2 são apresentados, respectivamente, os dados de produtividade de culturas e imagens multiespectrais utilizadas neste trabalho.

5.3.1.1 Produtividade de culturas

As informações de produtividade de culturas foram utilizadas para o desenvolvimento da ferramenta especificamente para o treinamento, verificação e extração dos modelos de previsão de culturas. Estas informações foram fornecidas pela empresa Lallemand Plant Care em forma de um banco de dados composto por arquivos vetoriais no formato shapefile (extensões .shp,.shx e.dbf) de geometria do tipo ponto, contendo os valores de produtividade mensurada por máquina do tipo colheitadeira agrícola. Também, em formato shapefile com geometria do tipo polígono foram fornecidos os limites das áreas da colheita. Além disso, foi informado o sistema de coordenadas e sistema de referência, bem como a unidade de medida quilogramas, referente a informação de produtividade.

5.3.1.2 Imagens multiespectrais

As bandas multiespectrais foram utilizadas como dado de entrada para o modelo de previsão da produtividade. Neste estudo foram utilizadas imagens Sentinel 2, devido às suas características como a disponibilização de forma gratuita, resolução espacial de 10 metros, resolução temporal de 5 dias e sensor multiespectral com 13 bandas. Na Tabela 2 são apresentadas algumas especificações das bandas do satélite Sentinel 2.

Tabela 2 - Especificações das bandas do satélite Sentinel 2

Resolução espacial (m)	Número	Nome	Comprimento de onda médio (nm)
10	B02	Azul	490
	B03	Verde	560
	B04	Vermelho	665
	B08	Infravermelho próximo - IVP	842
20	B05	Vermelho limítrofe 1	705
	B06	Vermelho limítrofe 2	740
	B07	Vermelho limítrofe 3	783
	B08A	Vermelho limítrofe 4	865
	B11	Infravermelho de ondas curtas 1 - SWIR 1	1610
	B12	Infravermelho de ondas curtas 2 - SWIR 2	2190
60	B01	Aerossol	443
	B09	Vapor de água	940
	B10	Cirrus	1375

Fonte: adaptado de Engesat (2021).

Para este trabalho, foram utilizadas as bandas de resolução espacial de 10 e 20 metros, sendo as de número B02, B03, B04, B05, B06, B07, B08, B08A, B11 e B12. As imagens do satélite Sentinel 2 foram obtidas pelo site do Copernicus (<https://scihub.copernicus.eu/dhus/#/home>) em formato JPEG 2000 (.jp2), que é um padrão de compressão de imagens de alta definição. O projeto Sentinel foi criado com o objetivo de fornecer serviços baseados em dados de satélite e *in situ* a fim de observar o estado do planeta, para isso, foram planejadas um conjunto de missões satelitárias, chamadas Sentinel, coordenadas pela Agência Espacial Europeia (COPERNICUS, 2021).

No site do Copernicus, para realizar o download das imagens, é necessário selecionar a missão desejada, a plataforma, tipo de produto e data de interesse. A missão Sentinel 2 disponibiliza para os usuários os produtos de Level-1C, o qual fornece a reflectância do topo da atmosfera (TOA) ortorretificada, com registro multispectral sub-pixel e de Level-2A, o qual fornece a reflectância de fundo da atmosfera (BOA) ortorretificada, com registro multispectral sub-pixel (ESA, 2021).

Para este trabalho foi utilizada a missão Sentinel 2, plataformas S2A e S2B e produtos do tipo Level-2A, uma vez que este produto é disponibilizado para os usuários em reflectância de superfície, com ortorrefecção e registro. Quanto a data de interesse, foram utilizadas imagens de aproximadamente 2 meses anteriores a data da colheita, conforme disponibilidade e porcentagem de cobertura de nuvens. No estudo de Ma et al (2021) visando a previsão da produção de milho com rede neural bayesiana foi constatado o desempenho quase ótimo com dados de sensoriamento remoto de 2 meses antes da colheita, devindo ao estágio fenológico da cultura. Na Tabela 3 são apresentas informações sobre as imagens obtidas.

Tabela 3 - Informações sobre as imagens Sentinel 2 adquiridas para o trabalho

Área	Data da colheita	Data da imagem	Plataforma
Milho 1	04/08/2020	08/06/2020	S2A
Milho 2	21/08/2020	28/06/2020	S2A
Milho 3	14/10/2020	12/09/2020	S2B
Milho 4	08/10/2020	12/09/2020	S2B
Milho 5	06/10/2020	08/08/2020	S2A
Milho 6	10/07/2020	15/05/2020	S2B
Trigo 1	20/07/2020	21/05/2020	S2B
Trigo 2	13/07/2020	18/05/2020	S2B
Trigo 3	08/07/2020	18/05/2020	S2B
Girassol 1	06/09/2020	29/07/2020	S2A
Girassol 2	12/09/2020	12/07/2020	S2A

Elaboração: a autora.

5.3.2 Pré-processamento de dados

A etapa de organização e pré-processamento de dados consistiu na manipulação dos dados obtidos na etapa de aquisição a fim de gerar arquivos padronizados para treinamento da RNA, bem como sua verificação, a fim de obter os modelos de previsão de produtividade. O subtópico 5.3.2.1 apresenta o procedimento adotado para extração da reflectância das bandas e cálculo de índices de vegetação.

5.3.2.1 Concatenação da reflectância, índices de vegetação e produtividade

Para a formação do arquivo de entrada do algoritmo de RNA foi necessária a concatenação das informações de produtividade com a reflectância das bandas. Além disso,

considerando a relação dos índices de vegetação com a produtividade, também foram associados três IVs aos dados de entrada. Os IVs indicam atividade da vegetação verde, teor de clorofila, biomassa verde e radiação fotossinteticamente ativa (POVH et al., 2008).

Neste trabalho foram utilizados três índices de vegetação sendo eles: (i) índice de vegetação da diferença normalizada – NDVI (ROUSE et al., 1947) índice que indica o verde, densidade e saúde da vegetação; (ii) índice de vegetação da diferença normalizada utilizando a banda do verde – GNDVI (GITELSON et al., 1996) índice de clorofila que indica a atividade fotossintética, utilizado para determinar a absorção de água e nitrogênio no dossel da cultura e; (iii) índice de vegetação resistente a atmosfera na região do visível – VARI (GITELSON et al., 2003) índice que enfatiza a vegetação na porção visível do espectro ao mesmo tempo que mitiga as diferenças de iluminação e os efeitos atmosféricos medindo o quão verde é a imagem (EOS, 2021). As Equações 1, 2 e 3 apresentam os índices de vegetação NDVI, GNDVI e VARI, respectivamente:

$$NDVI = \frac{\text{Infravermelho próximo (IVP)} - \text{Vermelho}}{\text{Infravermelho próximo (IVP)} + \text{Vermelho}} \quad (1)$$

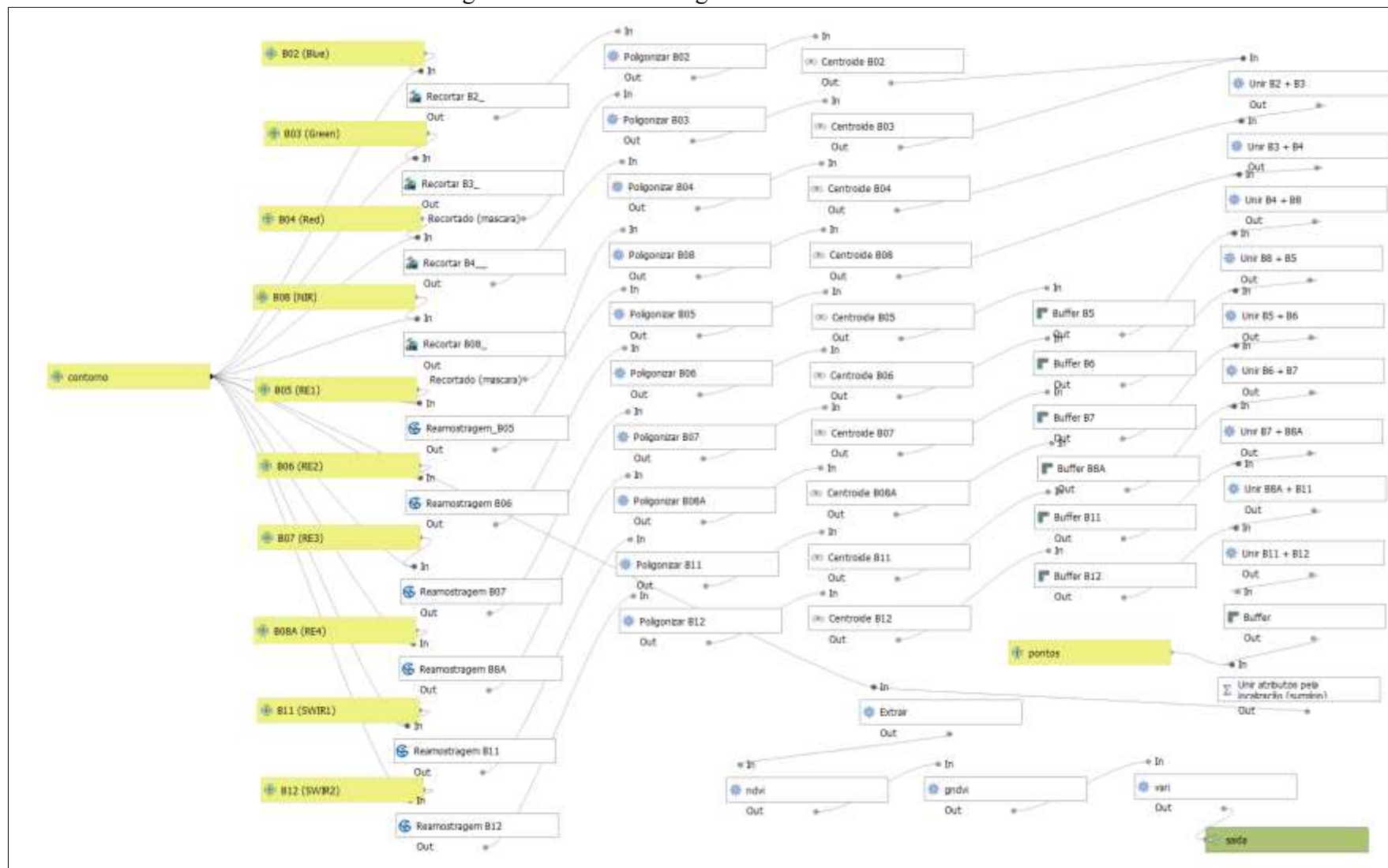
$$GNDVI = \frac{\text{Infravermelho próximo (IVP)} - \text{Verde}}{\text{Infravermelho próximo (IVP)} + \text{Verde}} \quad (2)$$

$$VARI = \frac{\text{Verde} - \text{Vermelho}}{\text{Verde} + \text{Vermelho} - \text{Azul}} \quad (3)$$

Para realizar a associação entre o valor medido de produtividade de cada cultura, os valores de reflectância das bandas da imagem e os IVs, foi desenvolvida, utilizando o software QGIS, uma ferramenta auxiliar com o recurso de modelador gráfico. Esta ferramenta tem como dado de entrada as bandas multiespectrais, o arquivo vetorial do tipo ponto que contém os valores de produtividade e o arquivo vetorial do tipo polígono que delimita a área de estudo. Como resultado, a ferramenta gera um vetor do tipo polígono com a delimitação de cada pixel, contendo em sua tabela de atributos a concatenação dos pixels homólogos da reflectância das bandas B02, B03, B04, B05, B06, B07, B08, B08A, B11, B12, NDVI, GNDVI, VARI e a produtividade média dos pontos que intersectam determinado pixel.

A Figura 7 apresenta o modelo lógico da ferramenta auxiliar desenvolvida. No Quadro 1 segue uma explicação detalhada sobre o desenvolvimento da ferramenta auxiliar, contendo seus dados de entrada, algoritmo, ferramentas utilizadas no QGIS e saída. A Figura 8 apresenta a interface de execução da ferramenta auxiliar. E a Figura 9 apresenta um exemplo das entradas e saída gerada pela ferramenta auxiliar.

Figura 7 - Modelo lógico da ferramenta auxiliar



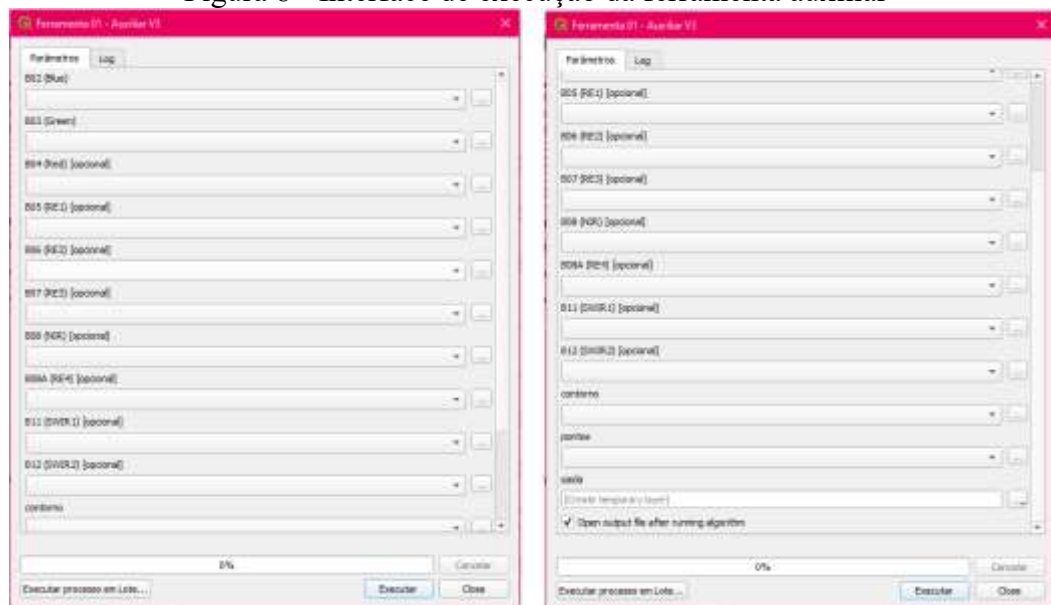
Elaboração: a autora.

Quadro 1 - Detalhamento da ferramenta auxiliar

Desenvolvimento da ferramenta auxiliar		Análises espaciais no QGIS
Entrada	Arquivos rasters das bandas B02, B03, B04, B05, B06, B07, B08, B08A, B11 e B12 Vetor do tipo polígono com o contorno da área Vetor do tipo ponto com atributo de produtividade	-
Algoritmo	1. Recortar bandas B02, B03, B04 e B08 com o contorno da área	Recortar raster pela camada da máscara
	2. Reamostrar bandas de 20 metros para 10 metros e recortar com o contorno da área as bandas B05, B06, B07, B08A, B11 e B12, pelo método de vizinho mais próximo	Reamostragem
	3. Poligonizar bandas B02, B03, B04, B05, B06, B07, B08, B08A, B11 e B12	Pixels de raster para polígonos
	4. Gerar centroide dos polígonos B02, B03, B04, B05, B06, B07, B08, B08A, B11 e B12	Centroides
	5. Fazer buffer de 5 metros nos centroides dos polígonos B05, B06, B07, B08A, B11 e B12	Buffer
	6. Unir centroides dos polígonos B02, B03, B4, B08, e buffers dos centroides de B05, B06, B07, B08A, B11 e B12	Associar atributos por local
	7. Fazer buffer de 5 metros nos centroides unidos	Buffer
	8. Unir buffer dos centroides unidos com vetor de pontos de produtividade, utilizando predicado geométrico intersecta e sumarizar pela produtividade média	Unir atributos pela localização (sumário)
	9. Recortar camada unida com o contorno da área	Extrair por localização
	10. Calcular NDVI, GNDVI e VARI	Calculadora de campo
Saída	Vetor de geometria polígono com atributos de reflectância, IVs e produtividade média	-

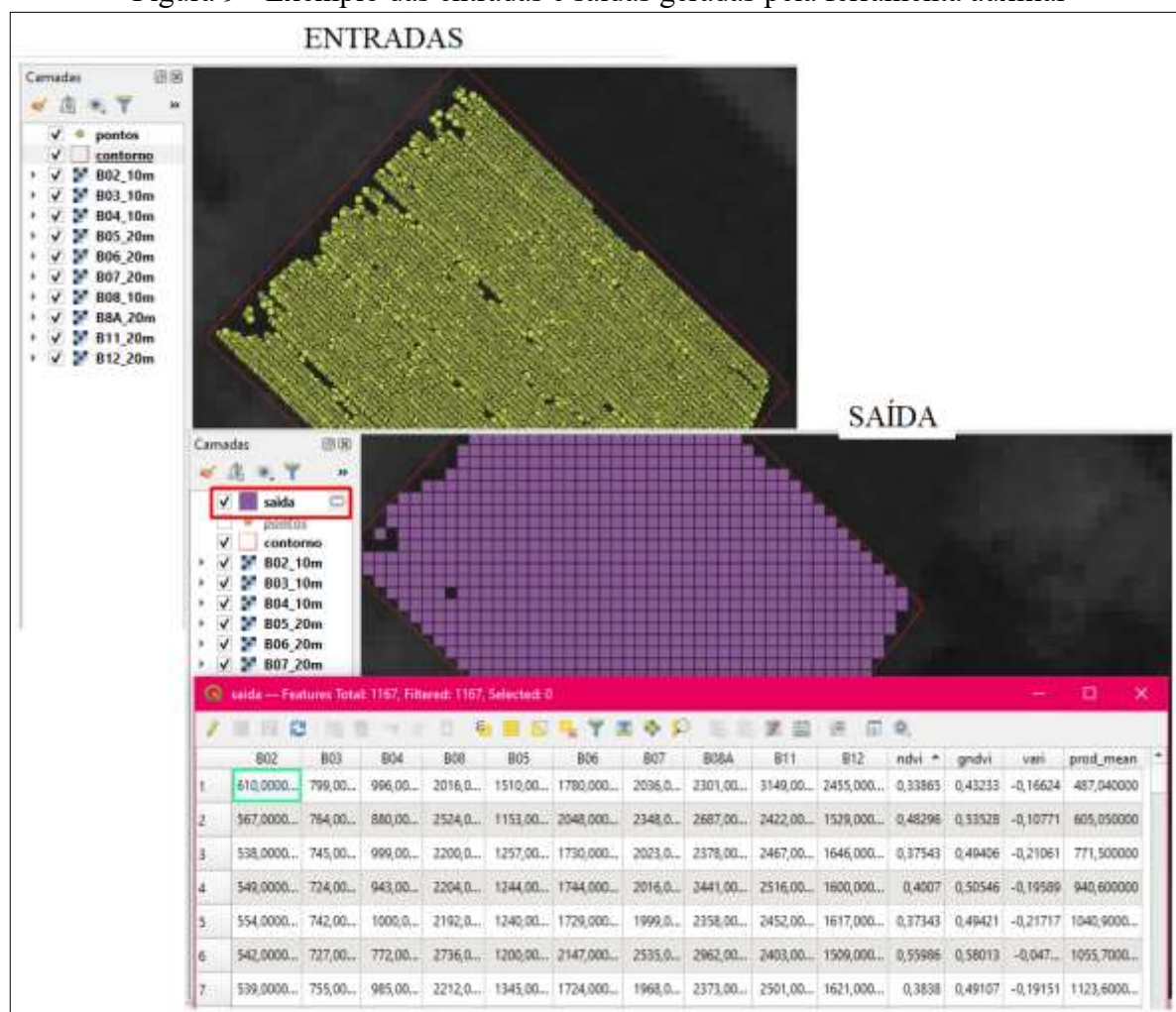
Elaboração: a autora.

Figura 8 - Interface de execução da ferramenta auxiliar



Elaboração: a autora.

Figura 9 - Exemplo das entradas e saídas geradas pela ferramenta auxiliar



Elaboração: a autora.

A ferramenta auxiliar desenvolvida fez com que os pontos de produtividade originais fossem reamostrados para a média de produtividade em uma resolução espacial de 10 metros, reduzindo a quantidade de pontos para compatibilizar com a resolução das bandas de melhor resolução espacial do sensor. A Tabela 4 apresenta a quantidade de pontos iniciais e a quantidade de pontos depois da utilização da ferramenta auxiliar.

Tabela 4 - Quantidade de pontos nos dados originais e depois da utilização da ferramenta auxiliar

Área	Quantidade de pontos de produtividade	
	Dados originais	Ferramenta auxiliar
Milho 1	26190	3747
Milho 2	8425	1348
Milho 3	9854	1167
Milho 4	5906	563
Milho 5	10768	1115
Milho 6	8964	1081
Trigo 1	8241	925
Trigo 2	10405	1659
Trigo 3	4285	392
Girassol 1	4471	576
Girassol 2	6839	524
Total	104348	13097

Elaboração: a autora.

O desenvolvimento de um modelo baseado em RNA envolve a divisão do banco de dados em três subconjuntos, sendo eles treinamento, validação e teste. O subconjunto de treinamento deve incluir todos os dados pertencentes ao domínio do problema e é usado na base de treinamento para atualizar os pesos da RNA. O subconjunto de validação é usado durante o processo de aprendizagem para verificar a resposta da rede para dados não treinados funciona como uma validação interna da rede. E a terceira parte, o subconjunto de teste deve incluir exemplos diferentes dos outros dois subconjuntos, e funciona como uma validação externa da rede.

Objetivando organizar os arquivos de entrada para modelo de RNA foi necessário realizar a divisão do banco de dados. Neste trabalho foi realizada inicialmente a divisão do banco de dados em: (i) treinamento e validação interna e (ii) teste/validação externa. Para isso as extensões dos arquivos shapefiles .dbf (extensão que contém os dados da tabela de atributos) dos vetores gerados na etapa anterior foram utilizadas para formar arquivos do tipo Texto Separado por Vírgula (extensão .csv). Estes arquivos contém os dados de treinamento e validação interna e teste/validação externa referentes a cada cultura e foram manipulados utilizando o software de planilhas Excel.

O Quadro 2 apresenta a divisão do banco de dados contendo as áreas que formaram o conjunto de treinamento e validação interna e as áreas que formaram o conjunto de

teste/validação externa de cada cultura, bem como a quantidade de pontos em cada arquivo .csv gerado.

Quadro 2 - Divisão do banco de dados

Cultura	Área	Treinamento e validação interna	Teste/Validação externa
Milho	Milho 1	x	
	Milho 2		x
	Milho 3	x	
	Milho 4	x	
Milho	Milho 5		x
	Milho 6	x	
	Arquivo .csv	milho_treinamento.csv	milho_teste.csv
	Total de pontos	7381	2463
Trigo	Trigo 1		x
	Trigo 2	x	
	Trigo 3	x	
	Arquivo .csv	trigo_treinamento.csv	trigo_teste.csv
	Total de pontos	2051	925
Girassol	Girassol 1	x	
	Girassol 2		x
	Arquivo .csv	girassol_treinamento.csv	girassol_teste.csv
	Total de pontos	576	524

Elaboração: a autora.

5.3.3 Construção da ferramenta

A etapa de desenvolvimento da ferramenta seguiu as etapas descritas nos subtópicos:

5.3.3.1 Construção dos modelos de RNA para produtividade, etapa em que foi realizada codificação de uma RNA para regressão em linguagem de programação Python para extração dos modelos de RNA para previsão da produtividade das culturas em estudo; *5.3.3.2 Treinamento, verificação e extração dos modelos de previsão*, etapa em que o RNA foi treinada, avaliada internamente e externamente, e foram extraídos os parâmetros dos modelos que apresentaram melhor ajuste para o problema e; *5.3.3.3 Implementação da ferramenta em ambiente SIG* etapa em que foi desenvolvida a ferramenta de previsão em ambiente SIG de acordo com os modelos gerados.

5.3.3.1 Construção dos modelos de RNA para produtividade

O desenvolvimento de uma RNA, segundo Basheer e Hajmmer (2000), passa por algumas fases como: (i) formulação do problema; (ii) projeto do sistema; (iii) realização do sistema; (iv) verificação/validação; (v) implementação e; (vi) manutenção. Estas etapas foram adaptadas para o desenvolvimento do algoritmo de RNA deste trabalho e o código foi desenvolvido utilizando o Google Colaboratory e linguagem de programação Python. As bibliotecas utilizadas foram Numpy, biblioteca que suporta o processamento de grandes arranjos e matrizes juntamente com uma coleção de funções matemáticas, e Scikit-Learn, biblioteca destinada a trabalhos com Aprendizado de Máquina - AM.

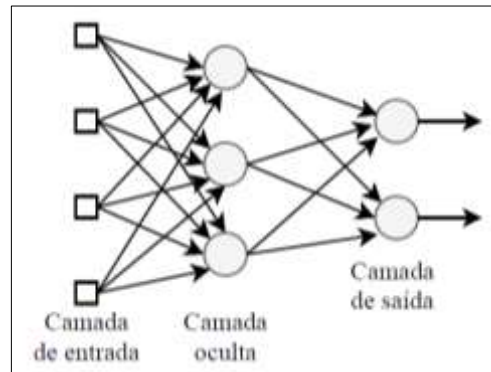
A fase **(i) formulação do problema** está intimamente ligada ao entendimento adequado do problema. Neste caso, o problema foi a previsão da variável produtividade através dos valores de reflectância das bandas de imagens multiespectrais, sendo caracterizado como um problema de regressão.

Na fase **(ii) projeto do sistema** são determinados o tipo de arquitetura da RNA, a regra de aprendizagem que se ajusta ao problema, o pré-processamento de dados e a divisão do banco de dados em subconjuntos.

Arquitetura da RNA

Em uma RNA, os neurônios estão organizados em camadas, assim basicamente tem-se uma camada de neurônios, chamadas de camada de entrada que se projeta em uma camada de neurônios, chamada de camada de saída. Para este trabalho a arquitetura da RNA foi do tipo alimentada diretamente com múltiplas camadas, a qual resolve problemas de natureza não linear e se distinguem pela presença de uma ou mais camadas de neurônios, sendo denominada (s) de camada (s) oculta (s). A Figura 10 apresenta o grafo da arquitetura das RNAs do tipo alimentada diretamente com múltiplas camadas (no inglês *Multilayer Perceptrons*).

Figura 10 - Grafo da arquitetura das RNAs alimentadas diretamente com múltiplas camadas



Fonte: adaptado de Haykin (2001).

Regra de aprendizagem

Em uma RNA o processo de aprendizagem implica em uma sequência de eventos, que segundo Haykin (2001) são: (i) a RNA é estimulada por um ambiente; (ii) a rede neural sofre modificações nos seus parâmetros como resultado desta estimulação e; (iii) a rede neural responde de uma maneira nova ao ambiente, devido às modificações ocorridas na sua estrutura interna. Sendo assim, no que se refere ao aprendizado de uma RNA, não há um algoritmo de aprendizado único, e eles se diferem pela forma em que ocorre o ajuste de um peso sináptico de um neurônio.

Neste trabalho foi utilizada a regra de aprendizado por correção do erro e algoritmo de retropropagação (do inglês *backpropagation*) que utiliza o gradiente descendente para o processo de aprendizagem. No aprendizado um sinal do neurônio de saída é comparado com a resposta desejada, produzindo um sinal de erro. Este sinal de erro tem o propósito de aplicar uma sequência de ajustes corretivos aos pesos sinápticos do neurônio que aproximarão passo a passo o sinal de saída da resposta desejada. Este objetivo é alcançado minimizando-se uma função de custo e com os ajustes dos pesos sinápticos do neurônio, até o processo ser encerrado (HAYKIN, 2001).

O algoritmo de *backpropagation* é composto por duas etapas, sendo o passo para frente (*forward pass*), onde as entradas são passadas através da rede e as previsões de saída são obtidas, e o passo para trás (*backward pass*), onde é calculado o gradiente da função de custo na camada de saída e ele é aplicado recursivamente a regra da cadeia para atualizar os pesos da RNA (HAYKIN, 2001).

Pré-processamento dos dados

Várias técnicas de pré-processamento são geralmente aplicadas antes que os dados possam ser usados para treinamento de uma RNA visando acelerar a convergência. Entre estas técnicas estão a remoção do ruído, tratamento de dados não normalmente distribuídos, inspeção de dados e exclusão de *outliers* (MASTERS, 1994). É desejado, na construção de uma RNA, que os dados de treinamento sejam distribuídos quase uniformemente entre as várias classes, assim evita-se que a rede neural seja tendenciosa para as classes super-representadas (SWINGLER, 1996).

Neste trabalho foi aplicada a normalização linear de dados dentro de uma faixa uniforme (-1 até 1), sendo um procedimento essencial para evitar que números maiores substituam os menores e para evitar a saturação prematura de neurônios ocultos, o que impede o processo de aprendizagem (MASTERS, 1994). A Equação 4 apresenta a normalização linear aplicada nos dados:

$$x_{normalizado} = -1 - 2 \cdot \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

onde: $x_{normalizado}$ é o valor normalizado de x ; e x_{max} e x_{min} são os valores máximo e mínimo, respectivamente, no banco de dados.

Divisão do banco de dados em subconjuntos

Treinar uma RNA para aprender hiperparâmetros de uma função de previsão e testá-la com os mesmos dados é um erro que pode ocasionar sobreajuste (do inglês *overfitting*), já que o modelo estará muito bem ajustado ao conjunto de dados, porém é ineficaz para dados diferentes dos utilizados no treinamento (JAMES et al., 2013). Para evitar este sobreajuste é comum manter parte dos dados disponíveis como um conjunto de validação.

Ao avaliar configurações diferentes para a RNA, ainda há risco de sobreajuste no conjunto de validação já que os parâmetros podem ser ajustados até que a rede tenha um desempenho ideal, ocasionando “vazamento” para o modelo e métricas de avaliação (JAMES, et al., 2013). Para resolver este problema, outra parte do banco de dados pode ser retirada, denominada teste, dessa forma, o conjunto treinamento é utilizado para treinamento de rede, após o qual é realizada uma avaliação interna com o conjunto de validação, e por fim, quando o modelo é satisfatório no conjunto de validação é realizada uma avaliação final com o conjunto de teste, denominada validação externa (FRIEDMAN, 2017).

Contudo, ao particionar os dados em três subconjuntos (i) treinamento, (ii) validação e (iii) teste, o número de amostras para treinamento do modelo é reduzido, e os resultados ficam sujeitos a uma escolha aleatória dos conjuntos de treinamento e validação. Uma solução para este problema é um procedimento denominado validação cruzada, em que um conjunto de teste deve ser separado para uma avaliação externa final, porém o conjunto de validação não é necessário (GOUTTE, 1997).

A abordagem mais básica de validação cruzada é chamada de K-fold, em que o conjunto de treinamento é dividido em k conjuntos menores, e para cada um dos k subconjuntos são realizados os procedimentos de: (i) k-1 dos dados são usados para treinamento, e (ii) o modelo resultante é validado por uma medida de precisão na parte restante dos dados (PLUTOWSKI et al., 1994). A medida de desempenho final é a média dos valores de erro calculados durante a validação de cada modelo treinado.

Neste trabalho optamos por utilizar a técnica de validação cruzada, e dessa forma, a divisão do banco de dados ocorreu na etapa de pré-processamento de dados, conforme apresentou o subtópico de 5.3.2.2 *Organização dos arquivos de entrada*, onde os dados foram divididos em (i) treinamento e validação interna e (ii) teste/validação externa. Objetivando avaliar com rigor os modelos aprendidos pela rede, a técnica de validação cruzada utilizada foi a *Leave One Out (LOO)*, onde cada conjunto de treinamento é gerado utilizando todas as amostras, exceto uma, que é deixada de fora como um conjunto de validação. Assim, para n amostras, temos n diferentes conjuntos de treinamento e n diferentes conjuntos de validação, que em termos de precisão, geralmente resulta em alta variância como um estimador para o erro de validação (RAO et al., 2008).

Na fase **(iii) realização do sistema** foi realizado o treinamento da RNA e simultaneamente foi avaliado o desempenho da rede analisando o erro de previsão da produtividade utilizando o conjunto de dados de treinamento e validação interna. Vale ressaltar que para cada cultura em estudo (milho, trigo e girassol) foram treinados e avaliados modelos de RNA diferentes.

No treinamento de modelos matemáticos, os hiperparâmetros são variáveis definidas antes do treino que permitem controlar o processo de treinamento do modelo, estes afetam diretamente o desempenho do modelo. Neste trabalho os hiperparâmetros foram divididos em dois tipos (i) hiperparâmetros pré-determinados que são: inicialização de pesos, otimizador gradiente descendente, função de ativação e função de custo; e (ii) hiperparâmetros empíricos que são: taxa de aprendizagem, coeficiente de momento, número de camadas ocultas, quantidade de neurônios nas camadas ocultas e número de épocas.

Hiperparâmetros pré-determinados

- A inicialização dos pesos de uma RNA é a atribuição de valores iniciais para os pesos de todas as conexões de neurônios. Segundo Basheer e Hajmmer (2000), alguns pesquisadores, indicam que a inicialização de pesos poder ter efeito na convergência da RNA. A função de inicialização dos pesos dos modelos de RNA deste trabalho foi a Glorot Normal que utiliza a função de distribuição normal, com inicialização de pesos em um intervalo de $[-\sigma, +\sigma]$, sendo o desvio padrão σ , calculado pela Equação 5:

$$\sigma = \sqrt{\frac{2}{in+out}} \quad (5)$$

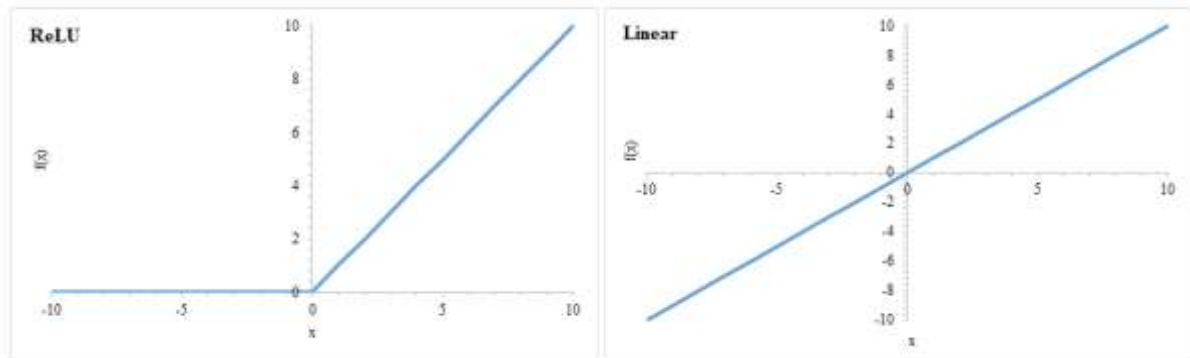
onde: *in* é a quantidade de camadas na camada anterior e *out* é a quantidade de camadas na camada atual. A Glorot Normal é uma das melhores funções de inicialização de pesos, pois considera o tamanho das camadas e ajuda a quebrar a simetria do aprendizado da rede tornando a convergência mais rápida e eficiente (DOSHI, 2018).

- O gradiente descendente é um algoritmo de otimização usado para minimizar a função de custo movendo-se iterativamente na direção da descida mais íngreme, forçando o ajuste dos parâmetros da RNA para atingir a variância mínima nos valores previstos. Neste trabalho foi utilizado algoritmo *Mini-batch Gradient Descent*, em que os dados são divididos em lotes para treinamento, neste caso de tamanho 32, e em geral, quando se tem um lote representativo, ele aponta para a direção que desce e tem convergência e execução rápida (BOTTOU, 1998).
- As funções de ativação são utilizadas na saída de um neurônio tendo como principal objetivo a transformação de uma entrada em um valor de saída (chamada de ativação) dentro de um intervalo determinado de forma que ajude na solução do problema (FRIEDMAN, 2017). Neste trabalho foram utilizadas duas funções de ativação, a função ReLU (unidade linear retificada) definida pela Equação 6, nas camadas escondidas, e a função de ativação Linear, definida pela Equação 7, nas camadas de saída. Os gráficos destas funções são apresentados na Figura 11.

$$f(x) = \max(0, x) \quad (6)$$

$$f(x) = x \quad (7)$$

Figura 11 - Gráfico das funções de ativação ReLU e linear



Elaboração: a autora.

ReLU é a função de ativação mais amplamente utilizada, não linear e permite que apenas alguns neurônios são ativados, tornando a rede esparsa e eficiente (AVINASH, 2017). E a função linear é uma função de ativação que realiza uma transformação linear da entrada, é ideal para tarefas simples ou utilização na camada de saída para problemas de regressão (AVINASH, 2017).

- Os métodos de treinamento de redes supervisionadas requerem uma medida de discrepância entre o valor de saída da rede e o valor desejado. Esta discrepância calculada para uma amostra ou conjunto de amostras é denominada perda, e o somatório das perdas para todo o banco de dados é denominado custo (FRIEDMAN, 2017). Neste trabalho a função de custo utilizada foi a do Erro Quadrático Médio – MSE (do inglês *Mean Squared Error*), apresentada na Equação 8:

$$J = \frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

sendo: J o custo; n número de amostras no banco de dados; y_i saída desejada e; \hat{y}_i saída predita da RNA.

Hiperparâmetros empíricos

- A taxa de aprendizagem é um hiperparâmetro definido no intervalo entre 0 e 1 e determina a magnitude do ajuste dos pesos. Assim, quanto mais próximo de 0 for a taxa de aprendizagem, menor serão as variações dos pesos sinápticos da rede de uma iteração para outra, e mais suave será a trajetória no espaço de pesos (HAYKIN, 2001). Por outro lado, se a taxa de aprendizagem for mais próxima de 1, visando acelerar a taxa de

aprendizagem, as grandes modificações nos pesos sinápticos resultantes podem fazer com que a rede oscile e nunca convirja para os valores corretos dos pesos (HAYKIN, 2001).

- O coeficiente de momento é um hiperparâmetro definido no intervalo entre 0 e 1, utilizado como método para evitar o perigo de instabilidade, ele acelera as atualizações de pesos quando há necessidade de reduzir a taxa de aprendizado para evitar oscilações. Um coeficiente de momento próximo a 1 reduz o risco de a rede ficar presa em mínimos locais, porém aumenta o risco de ultrapassar a solução, assim como uma taxa de aprendizado próxima a 1 combinada a um coeficiente de momento extremamente pequeno leva a um treinamento lento (HAYKIN, 2001).
- Quanto ao número de camadas ocultas em um RNA, na maioria dos problemas de aproximação de função ou previsão/modelagem, uma é o suficiente para aproximar funções contínuas, já para funções de aprendizagem com descontinuidade podem ser necessárias duas camadas ocultas (HAYKIN, 2001).
- A quantidade de neurônios nas camadas ocultas deve ser definida objetivando uma boa generalização. Uma vez que, uma rede com poucos neurônios ocultos pode ser incapaz de diferenciar padrões complexos, levando a uma estimativa linear, e uma rede com muitos nós ocultos, seguirá o ruído nos dados devido a parametrização excessiva, levando a uma generalização ruim para dados não treinados (BASHEER; HAJMMER, 2000).
- O número de épocas em uma RNA é um hiperparâmetro que representa a quantidade de passagens completas do conjunto de dados para treinamento, também chamada como número de ciclos. Se a quantidade de épocas for pequena, pode ocorrer subajuste (do inglês *underfitting*), e se a quantidade de épocas for alta, ou seja, treinamento excessivo, pode ocorrer sobreajuste.

A Tabela 5 apresenta os valores testados para os hiperparâmetros de taxa de aprendizagem, coeficiente de momento, número de camadas escondidas, quantidade de neurônios nas camadas escondidas e número de épocas.

Tabela 5 - Valores dos hiperparâmetros testados

Hiperparâmetro	Valores para teste
Taxa de aprendizagem	0.1 ; 0.01 ; 0.001 ; 0.0001
Coefficiente de momento	0.5 ; 0.9 ; 0.99
Camadas ocultas	1 ; 2
Neurônios	16; 20 ; 32; 50
Épocas	100 ; 200 ; 300 ; 500; 1000 ; 2000 ; 3000 ; 5000

Elaboração: a autora.

Na fase **(iv) verificação/validação**, embora a construção da RNA já incluiu a validação interna enquanto o treinamento estiver em andamento (validação cruzada LOO), as melhores RNA para cada cultura em estudo foram examinadas para sua capacidade de generalização utilizando o subconjunto de teste. Essa verificação teve como objetivo a confirmação da habilidade dos modelos, avaliando a capacidade da RNA responder com precisão a exemplos não utilizados no desenvolvimento, neste caso áreas diferentes com as mesmas culturas.

Ambas validações, interna e externa, foram realizadas a partir da análise da medida de desempenho raiz do erro quadrático médio – RMSE (do inglês *root mean square error*) em porcentagem, apresentado na Equação 9.

$$RMSE (\%) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sum_n \frac{y_i}{n}} * 100 \quad (9)$$

onde: y_i é a saída desejada; \hat{y}_i é a saída predita RNA e n é o número de saídas da RNA.

Neste trabalho o objetivo foi obter modelos que apresentassem medida de desempenho RMSE da validação interna até 40% e da validação externa até 30%. Dessa forma, para implementação na ferramenta, foi extraído o modelo matemático da melhor RNA para cada cultura, sendo arquitetura que apresentasse a menor o RMSE de validação interna e RMSE de validação externa.

A fase de **(v) implementação** e a fase de **(vi) manutenção** dos modelos de previsão de produtividade para cada cultura são apresentadas em seguida no tópico 5.3.3.2 *Implementação da ferramenta em ambiente SIG*.

5.3.3.2 Implementação da ferramenta em ambiente SIG

Depois de gerados os modelos matemáticos da RNA para previsão da produtividade se dá a fase de implementação da RNA, que neste trabalho é uma ferramenta em ambiente SIG com as seguintes características:

- Interface padrão do ambiente SIG do software QGIS;
- Disponibilização de informações de ajuda ao usuário contendo todas as características da ferramenta;
- Opção para cálculo do erro de previsão do modelo a partir de dados de produtividade inseridos pelo usuário;
- Dados de entrada obrigatórios: cultura a ser mapeada, camada vetorial do tipo polígono com a delimitação da área, camadas rasters das bandas multiespectrais do satélite Sentinel 2 em valores de reflectância.
- Dados de entrada opcionais: camada vetorial do tipo ponto contendo em sua tabela de atributos campo com valores de produtividade;
- Dados de saída: camada raster recortada na área de interesse que representa a superfície referente a previsão da produtividade e estatística descritiva desta camada (mínimo, máximo, intervalo, soma, média e desvio padrão);
- Dados de saída alternativos: camada vetorial do tipo ponto contendo a produtividade mensurada, produtividade prevista e o erro.

Para a construção da ferramenta foi utilizada o componente de modelador gráfico disponível no software QGIS que permite a criação de rotinas, visando a automatização de processos, além de permitir projetar uma interface gráfica para as ferramentas desenvolvidas. Para uma posterior disponibilização no repositório oficial do QGIS em forma de plugin, será necessário exportar o código Python e manipulá-lo com o plugin PluginBluider.

A fase de **(vi) manutenção** do sistema envolve a atualização do modelo e da ferramenta conforme ocorrerem mudanças nas variáveis do sistema, como por exemplo, a obtenção de novos dados, o que pode acarretar em um novo ciclo de desenvolvimento.

6 RESULTADOS

Neste tópico são apresentados os resultados alcançados com este trabalho. Para melhor compreensão este tópico foi dividido em **6.1 Arquitetura do modelo de RNA**, onde são apresentados os hiperparâmetros definidos para cada cultura; **6.2 Ferramenta desenvolvida** onde é apresentada a interface da ferramenta e sua estrutura lógica; **6.4 Aplicação da ferramenta** onde é apresentada a aplicação da ferramenta para estimativa da produtividade nas culturas de milho, trigo e girassol nos dados de teste e; **6.4 Disponibilização**, onde é apresentada a forma de disponibilização e instalação da ferramenta.

6.1 Arquitetura do modelo de RNA

As melhores arquiteturas de RNA para cada cultura foram obtidas por meio do monitoramento do RMSE com a variação de alguns hiperparâmetros, sendo eles a taxa de aprendizagem, coeficiente de momento, número de camadas ocultas, quantidade de neurônios nas camadas ocultas e número de épocas de treinamento. A Tabela 6 apresenta os valores dos hiperparâmetros e erro de validação dos modelos de RNA selecionados para cada cultura.

Tabela 6 - Valores dos hiperparâmetros e erro de treinamento dos modelos de RNA selecionados para cada cultura

Hiperparâmetro	Cultura		
	Milho	Trigo	Girassol
Taxa de aprendizagem	0,1	0,1	0,1
Coeficiente de momento	0,50	0,90	0,99
Número de camadas ocultas	1	1	1
Quantidade de neurônios na camada oculta	16	16	16
Número de épocas	100	100	100
RMSE(%) da validação interna	101,62%	34,23%	20,19%
Desvio padrão do RMSE (%)	858,32	70,15	26,11

Elaboração: a autora.

Analisando a Tabela 6 é possível notar que os hiperparâmetros das RNAs para previsão da produtividade foram, em sua maioria os mesmos para as três culturas em estudo, isso pode estar ligado ao fato de que mesmo sendo culturas diferentes, a variável de estimativa foi a

mesma. O coeficiente de momento foi o único hiperparâmetros que apresentou diferença entre as culturas, variando de 0,50 a 0,99.

Quanto as arquiteturas para as três culturas, foram definidas RNAs de 1 camada oculta com 16 neurônios, camada de entrada com 13 neurônios (B02, B03, B04, B05, B06, B07, B08, B08A, B11, B12, NDVI, GNDVI, VARI) e 1 neurônio na camada de saída, arquitetura de configuração 13 – 16 – 1. Trabalhos que utilizam as mais diversas variáveis de entrada para previsão de produtividade de culturas utilizando RNA do tipo MLP, apresentam arquiteturas similares para estimativas, ou seja, 1 camada oculta com grande quantidade de neurônios nestas camadas. Soares et al. (2015), estimaram a produtividade do milho a partir de variáveis morfológicas, utilizando a arquitetura 4 – 35 – 1; Silva (2015) obteve os melhores resultados para previsão da produtividade de madeira utilizando rede MLP de arquitetura 30 – 14 – 1.

Hara et al. (2021) na revisão sobre os avanços do sensoriamento remoto para o monitoramento de culturas e estimativa de rendimentos, destacam que modelos baseados em RNA são uma alternativa adequada, e que embora as RNA sejam amplamente utilizadas na previsão de produtividade, sua aplicação prática ainda enfrenta algumas dificuldades. Sendo elas a seleção de um número apropriado de camadas de rede ocultas, velocidade de treinamento do modelo e a aplicação de uma quantidade suficiente de dados na forma de variáveis independentes.

Na Tabela 6 estão apresentados os RMSEs da validação interna dos modelos de RNA para cada cultura, valor que representa a média dos RMSEs obtidos ao realizar a validação pelo método de validação cruzada *LOO*, e os desvios padrão desses valores. Para cultura de milho a melhor arquitetura apresentou RMSE de 101,62% e desvio padrão de 858,32, o que pode ser considerado como um modelo incapaz de realizar generalizações; já as culturas de trigo e girassol apresentaram RMSE de 34,23% e 20,19% e desvios padrão de 70,15% e 26,15%, respectivamente, sendo modelos capazes de generalizar a produtividade dessas culturas e dentro do objetivo deste estudo.

Vale ressaltar que o método de validação cruzada *LOO*, o qual o treinamento é realizado com $n - 1$ dados e a validação com apenas com 1 dos registros, é um método simples, porém rigoroso de validação a fim de não desperdiçar muitos dados, e geralmente resulta em alta variância como um estimador para o erro de validação. Dessa forma, o alto valor de RMSE e desvio padrão da cultura de milho além de estarem ligados a natureza dos dados, também estão ligados ao método de validação cruzada utilizado.

Outro fator possivelmente relacionado aos valores obtidos de RMSE e desvio padrão da validação interna é a quantidade de dados utilizados para treinamento e validação interna. Para

a cultura de milho, a qual o modelo apresentou maior erro (101,62%) e variabilidade, foram utilizadas 7981 amostras, e os modelos de trigo e girassol, utilizaram 2051 e 576 amostras, respectivamente, o que provavelmente configurou menores valores de RMSE (34,23% e 20,19%, respectivamente) e desvios padrão.

Alguns trabalhos têm mostrado as vantagens da utilização das RNAs sobre outros métodos na previsão de produtividade de culturas. Zaefizadeh et al. (2011) analisaram o desempenho de uma RNA MLP de configuração 3 – 15 – 1, e um modelo de regressão linear múltipla para previsão do rendimento de cevada, obtendo melhores resultados de erro médio absoluto de 0,21 e 0,22 toneladas por hectare com o modelo de RNA. Khaki et al. (2020) analisaram a eficácia de um modelo híbrido de redes neurais convolucionais e recorrentes em comparação com outros métodos como, florestas aleatórias, redes neurais profundas e regressão LASSO, para previsão da produtividade de milho e soja, o modelo obtido apresentou um coeficiente de correlação da validação interna variando entre 0,86 e 0,88, bem como, RMSE de treinamento variando entre 11,48% e 13,26%.

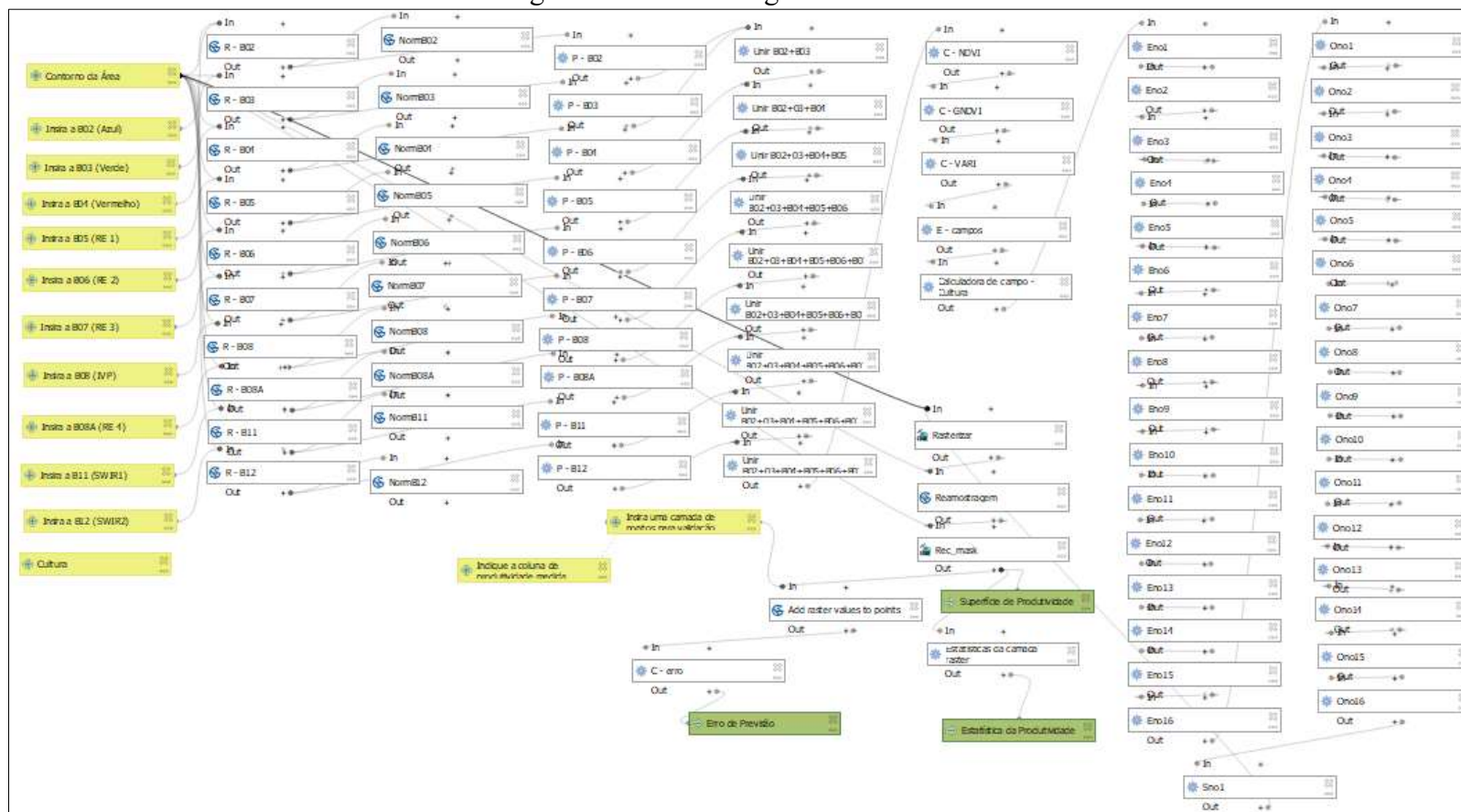
O número de fatores que influenciam o rendimento da cultura e os parâmetros que descrevem a condição do dossel dificultam a seleção dos parâmetros para previsão de produtividade. Alguns trabalhos têm relacionado à produtividade, além de dados de sensoriamento remoto, dados meteorológicos e informações sobre tecnologia agrícola. Cai et al. (2019), propuseram o uso da fluorescência da clorofila induzida pelo sol, advinda de uma banda IVP, combinada a dados climáticos e mostraram que houve uma melhora da previsão de trigo em relação a utilização de apenas os dados de sensoriamento remoto.

No trabalho de Kim et al. (2019), além dos dados de sensoriamento remoto foram utilizadas informações meteorológicas para avaliar o rendimento das safras de milho e soja, e os resultados evidenciaram que o modelo de RNA apresentou menor acurácia de previsão em comparação a modelos de RNA profundas, com erro de previsão médio de 7,60% para o milho e 7,80% para soja. Além disso, obtiveram coeficiente de correlação para os modelos de milho e soja, de 0,95 e 0,90, respectivamente. Posteriormente, Kross et al. (2020), estudaram a relação entre a topografia da área, índices de vegetação e a produtividade do milho e da soja, e comprovaram que a rede MLP desenvolvida foi mais eficiente na previsão de produtividade de milho, com erro menor que 15%.

6.2 Ferramenta desenvolvida

O esboço do modelo lógico da ferramenta desenvolvida está apresentado na Figura 12.

Figura 12 - Modelo lógico da ferramenta



Elaboração: a autora.

No modelo lógico da ferramenta desenvolvida neste estudo (Figura 12), estão representadas as entradas (na cor amarela), os processos e aplicação dos modelos (na cor branca), e as saídas (na cor verde). Dentro os processos estão: (i) a reamostragem dos pixels para resolução de 10 metros; (ii) a normalização dos dados de entrada; (iii) a poligonização; (iv) o cálculo dos índices de vegetação NDVI, GNDVI e VARI; (v) a união dos dados de entrada para rede; (vi) o cálculo da produtividade pelos modelos de RNA; (vii) a rasterização; (viii) a reamostragem para 10 metros; (ix) o recorte da superfície na área de interesse e; (x) o cálculo da estatística da superfície de produtividade. Além destes, quando o usuário entrar com a camada de pontos de validação, o cálculo do erro de previsão.

A interface de execução da ferramenta para estimativa da produtividade de culturas está apresentada nas Figuras 13 e 14.

Figura 13 - Interface de execução da ferramenta: parte 1

The screenshot displays the 'Produtividade' software interface. The window title is 'Produtividade' with a close button. It features two tabs: 'Parâmetros' (selected) and 'Log'. The 'Parâmetros' tab contains several input fields, each with a dropdown menu and a browse button (...):

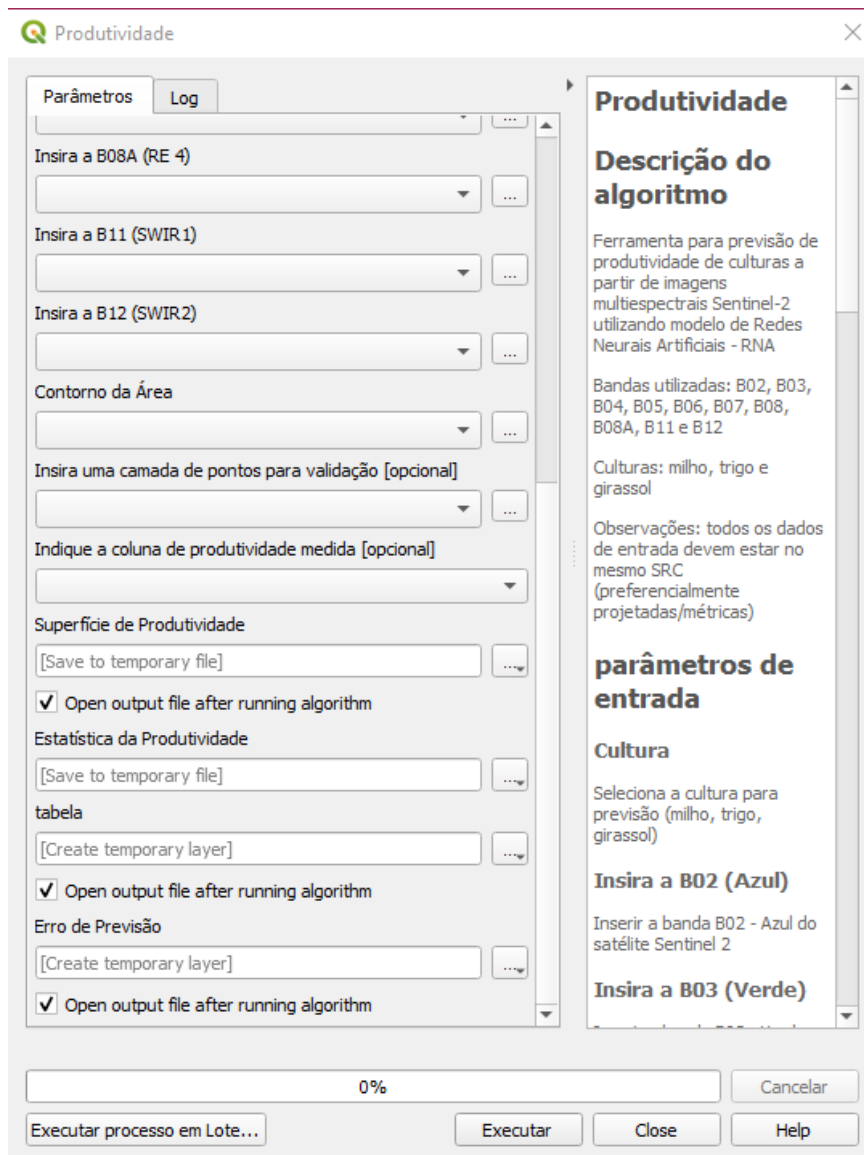
- Cultura
- Insira a B02 (Azul)
- Insira a B03 (Verde)
- Insira a B04 (Vermelho)
- Insira a B05 (RE 1)
- Insira a B06 (RE 2)
- Insira a B07 (RE 3)
- Insira a B08 (IVP)
- Insira a B08A (RE 4)
- Insira a B11 (SWIR1)
- Insira a B12 (SWIR2)
- Contorno da Área

At the bottom of the parameter section is a progress bar showing '0%' and a 'Cancelar' button. Below the progress bar are four buttons: 'Executar processo em Lote...', 'Executar', 'Close', and 'Help'.

The right-hand side of the interface is a scrollable panel titled 'Produtividade' containing:

- Descrição do algoritmo**: Ferramenta para previsão de produtividade de culturas a partir de imagens multiespectrais Sentinel-2 utilizando modelo de Redes Neurais Artificiais - RNA. Bandas utilizadas: B02, B03, B04, B05, B06, B07, B08, B08A, B11 e B12. Culturas: milho, trigo e girassol. Observações: todos os dados de entrada devem estar no mesmo SRC (preferencialmente projetadas/métricas).
- parâmetros de entrada**:
 - Cultura**: Seleciona a cultura para previsão (milho, trigo, girassol).
 - Insira a B02 (Azul)**: Inserir a banda B02 - Azul do satélite Sentinel 2.
 - Insira a B03 (Verde)**

Figura 14 - Interface de execução da ferramenta: parte 2



A interface da ferramenta seguiu o padrão das ferramentas tradicionais do software QGIS. Ao lado direito da ferramenta encontram-se todas as informações sobre a ferramenta, como a descrição do algoritmo, culturas aptas para previsão, parâmetros de entrada, arquivos de saída, características dos modelos e RMSE de validação interna dos modelos, bem como as informações sobre os desenvolvedores. Na seção Parâmetros são requisitadas as entradas, sendo elas:

- (i) Menu suspenso com as opções Milho/Trigo/Girassol para seleção da cultura a ser estimada (obrigatório);
- (ii) Campos para seleção das camadas em formato raster para as bandas do satélite Sentinel 2 (obrigatórios);

- (iii) Campo para seleção da camada vetorial do tipo polígono para o contorno da área de interesse (obrigatório);
- (iv) Campo para seleção de camada vetorial do tipo ponto para os pontos de validação (opcional); e
- (v) Campo para indicação do campo da tabela de atributos da camada de validação que contém a variável produtividade (opcional).

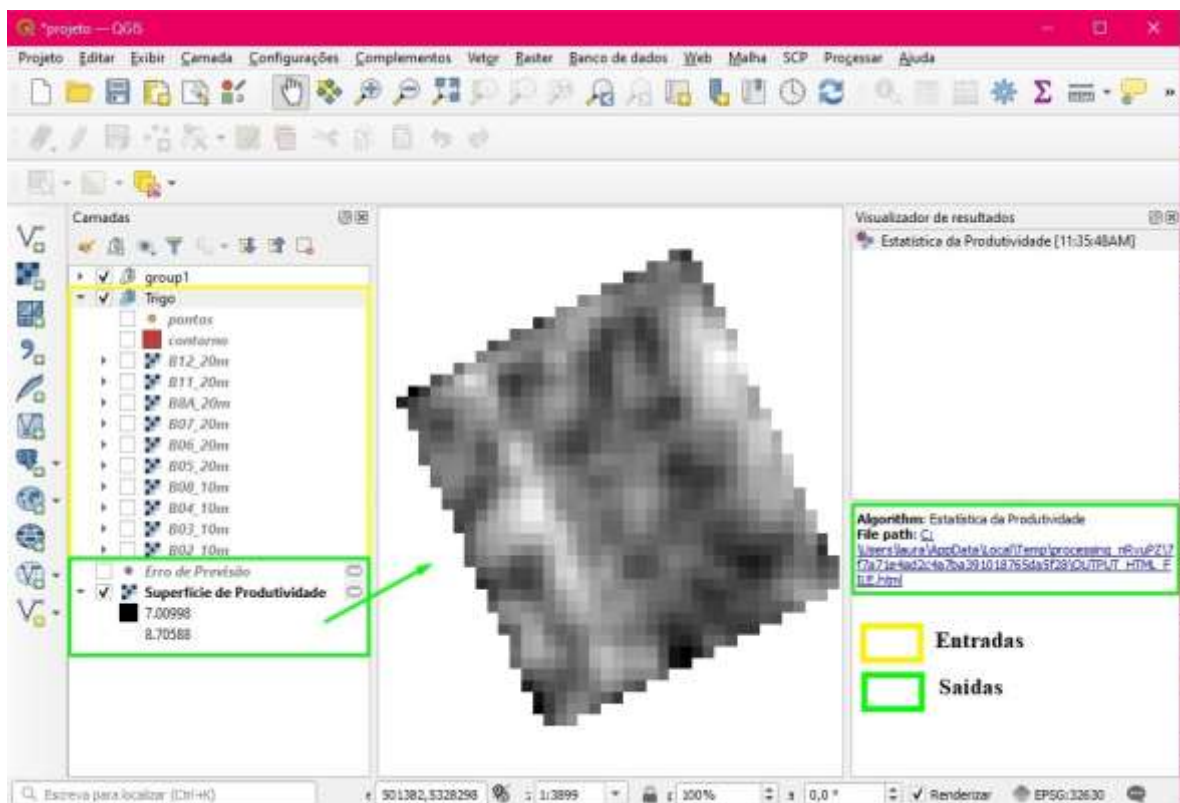
Além das entradas, o usuário necessita indicar o caminho de saída para salvar os resultados, sendo eles:

- (i) Raster da superfície com os valores de produtividade estimada das culturas;
- (ii) Arquivo com extensão .html contendo as estatísticas descritivas da camada;
- (iii) Camada vetorial do tipo ponto com a produtividade mensurada, produtividade prevista e erro.

Caso o usuário não indique o caminho de saída os resultados serão salvos em uma camada temporária a ser aberta assim que o algoritmo terminar a execução.

A Figura 15 apresenta um exemplo dos resultados gerados depois da execução da ferramenta em ambiente QGIS.

Figura 15 - Exemplo dos resultados gerados após a execução da ferramenta em ambiente QGIS



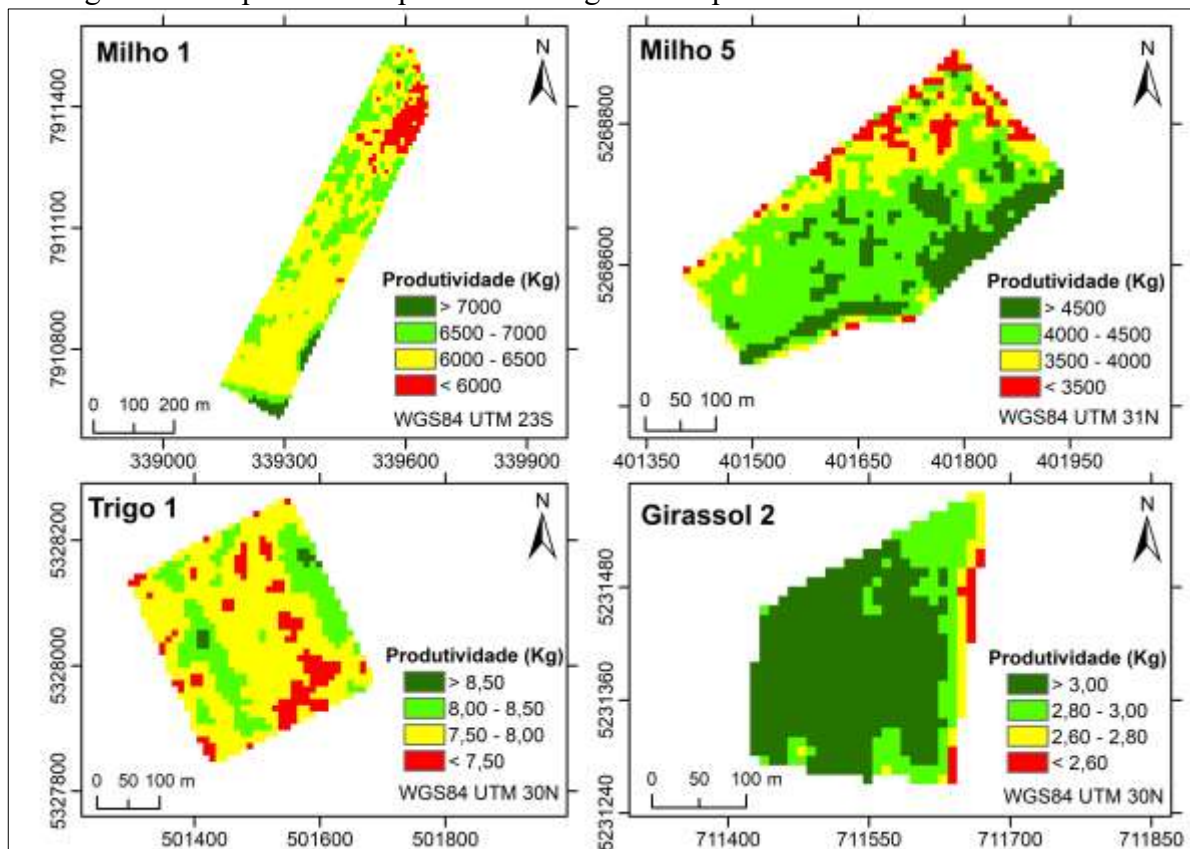
Elaboração: a autora.

Na Figura 15 foram apresentados os dados de entrada (contornados em cor amarela), e os resultados, que são a superfície de produtividade gerada, a estatística da previsão, e o erro de previsão (contornados em cor verde). O código fonte completo da ferramenta desenvolvida pode ser acessado pelo link <https://github.com/LauraMouraXavier/Ferramenta---Trabalho-de-Conclus-o-de-Curso>.

6.3 Aplicação da ferramenta

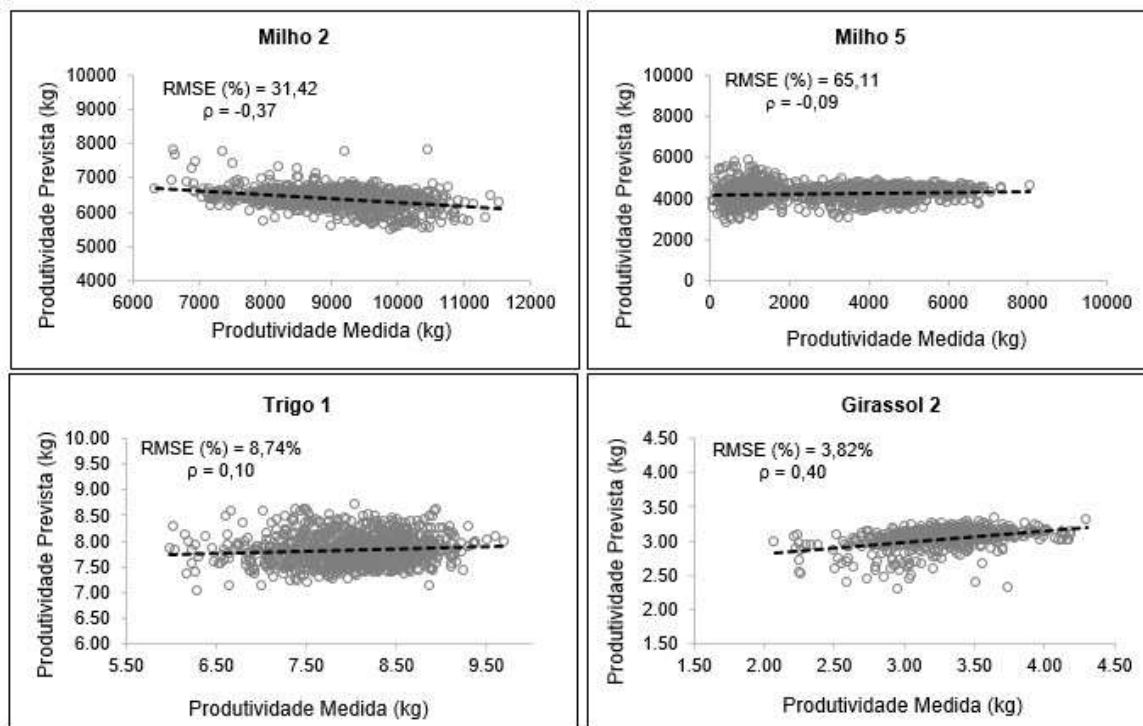
Para teste do modelo de RNA e aplicação da ferramenta desenvolvida foram utilizadas as áreas de teste/validação externa, as denominadas Milho 2, Milho 5, Trigo 1 e Girassol 2. A Figura 17 apresenta as superfícies de produtividade geradas a partir da ferramenta desenvolvida neste trabalho, depois de gerados os *layouts*. Na Figura 17 são apresentados os gráficos de dispersão relacionando a produtividade medida e prevista, bem como o coeficiente de correlação de Person (ρ) e o RMSE (%) da previsão.

Figura 16 - Superfícies de produtividade geradas a partir da ferramenta desenvolvida



Elaboração: a autora.

Figura 17- Gráficos de dispersão relacionando a produtividade prevista e medida com a correlação e RMSE da previsão



Elaboração: a autora.

A partir da análise dos resultados gerados pela ferramenta foi possível gerar as superfícies estimadas de produtividade de cada cultura em formato raster, bem como obter suas estatísticas e erros, esses dados podem ser exportados e importados para qualquer outro software para as diversas finalidades, como por exemplo a geração de mapas temáticos.

Analisando os valores de erros e correlações, como esperado, a cultura de milho, apresentou os maiores valores de RMSE, uma vez que na etapa de validação interna o modelo já havia apresentado valor de 101,62%. Na área de Milho 2 foi constatado um RMSE de 31,42% e ρ igual a -0,37, dessa forma percebe-se que este modelo não apresentou resultados satisfatórios devido à correlação negativa entre a produtividade mensurada e prevista. Já na área de Milho 5 o RMSE apresentou valor de 65,11% e ρ muito baixa de -0,09, constando novamente o resultado insatisfatório do modelo de RNA implementado para a cultura de milho. Além disso, vale ressaltar que as áreas Milho 2 e Milho 5 são localizadas em países diferentes o que pode requisitar modelos de previsão diferentes.

As áreas de cultura de trigo e girassol apresentaram erros menores em relação as áreas de milho. A área Trigo 1, apresentou RMSE de 8,74% e baixa ρ , porém positiva de 0,10, e a área Girassol 2, RMSE de 3,82% e ρ moderada e positiva de 0,40. Percebe-se que apesar dos

erros para estas duas culturas serem abaixo do esperado por este trabalho, 30%, ainda sim a área de Trigo apresentou uma baixa correlação entre a produtividade medida e prevista.

Na Tabela 7 são apresentadas as estatísticas das medidas descritivas das superfícies geradas para cada área, e as estatísticas descritas da produtividade mensurada.

Tabela 7 - Estatística descritiva das áreas teste

Produtividade	Medida descritiva (kg)	Área			
		Milho 2	Milho 5	Trigo 1	Girassol 2
Prevista	Mínimo	5233.19	2105.24	7.01	2.31
	Máximo	8206.15	6135.15	8.71	3.34
	Intervalo	2972.95	4029.92	1.69	1.03
	Soma	9840089.65	5020823.75	7541.11	1661.59
	Média	6389.67	4198.01	7.81	3.03
	Desvio Padrão	317.13	474.16	0.29	0.16
Mensurada	Mínimo	6340.3	103.70	5.98	2.08
	Máximo	11531.3	8069.53	9.69	4.3
	Intervalo	5190.99	7965.83	3.72	2.22
	Soma	18480464.12	3550537.83	7436.1	1714.71
	Média	8923.45	3184.33	8.11	3.27
	Desvio Padrão	793.85	1787.14	0.63	0.37

Elaboração: a autora.

Analisando a Tabela 7 é possível dimensionar o quanto o modelo de RNA foi insatisfatório para previsão da produtividade de milho, já que os valores apresentaram discrepâncias altíssimas entre a produtividade prevista e mensurada. Podemos notar pela estatística de soma, que na área Milho 2 a produção foi reduzida pela metade, e a Milho 5 foi superestimada em cerca de 1,4 mil kg.

No caso das áreas de cultura de trigo e girassol, em que os modelos apresentaram RMSE menores, podemos observar que as estatísticas descritivas apresentaram valores mais próximos da realidade. A diferença entre a produtividade prevista e mensurando em relação à soma e a média foi de 105,01 kg e 0,30 kg, respectivamente, para a área de Trigo 1; e para a área de Girassol 2 a diferença em relação à soma e a média foi de 53,12 kg e 0,03 kg, respectivamente.

A geração das estatísticas descritivas da superfície é um parâmetro direcionado ao usuário que deseja quantificar a produtividade de culturas antecipadamente, cerca de 2 meses antes. Uma vez que, com a utilização da ferramenta, será possível obter estimativas representativas do total da colheita, média de produção por 100 m² e a dispersão em relação à média da produtividade. Ressaltamos que para a obtenção de melhores resultados com a

utilização da ferramenta, dado o seu desenvolvimento, a recomendação é utilização de imagens Sentinel 2, de cerca de 2 meses anteriores a colheita, e secundamente, seu reuso após a colheita, a fim de verificar a precisão dos dados gerados em primeiro momento.

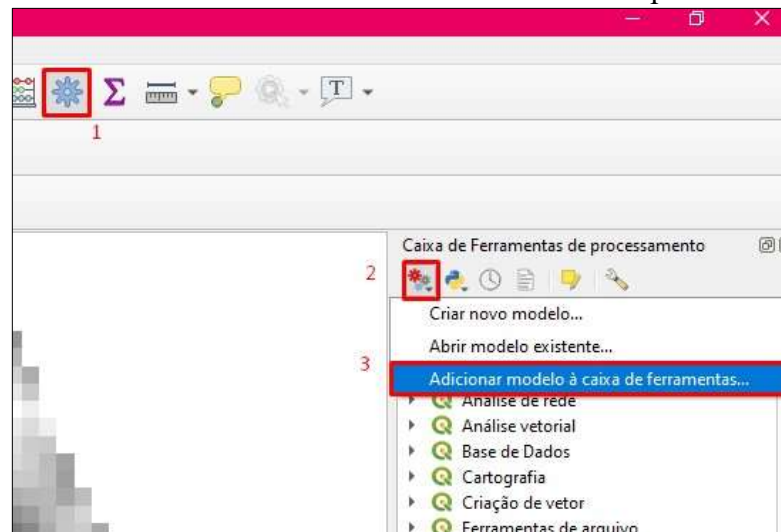
Os dados advindos de modelos de RNA são uma fonte de informações necessárias no gerenciamento da produção agrícola, ainda mais quando é possível obter informações sobre a produtividade alguns meses antes da colheita, o que é extremamente valioso para se adotar uma estratégia adequada na importação e exportação de produtos, além do conhecimento prévio possibilitar a racionalização dos meios de produção (HARA et al., 2021).

6.4 Disponibilização

A disponibilização da ferramenta desenvolvida neste estudo será realizada após as considerações dos membros da banca. Vale ressaltar aqui que a ferramenta é compatível com a versão do QGIS 3.14.1 ou superiores. Além disso, a ferramenta será disponibilizada de duas formas, sendo elas: (i) em forma de ferramenta da caixa de ferramentas do QGIS através da disponibilização do arquivo de extensão .model3 e; (ii) plugin denominado ProdSen2 que poderá ser baixado do repositório do QGIS e instalado, conforme configuração padrão do QGIS.

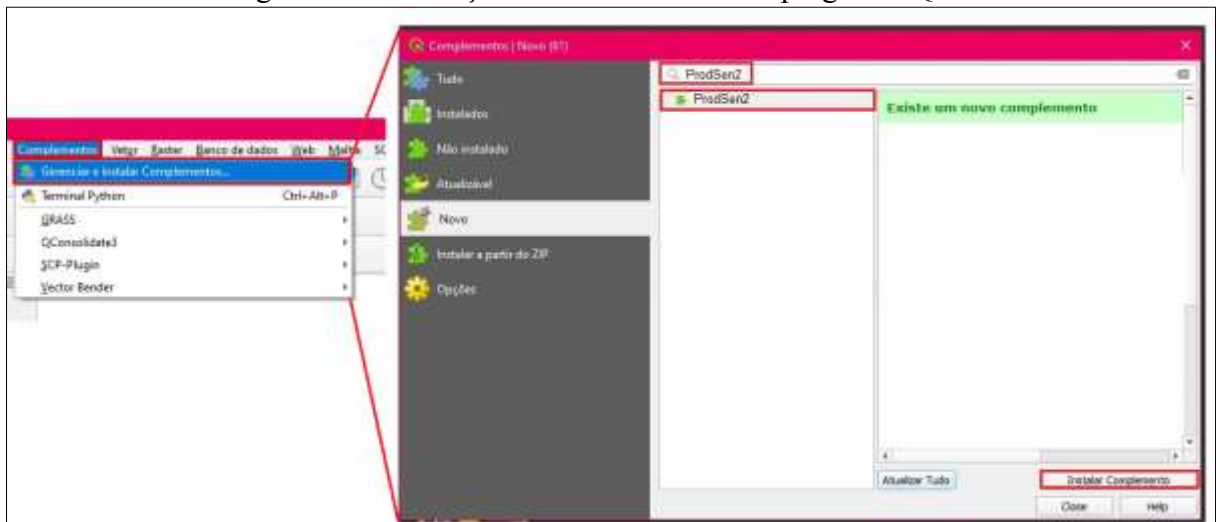
A Figura 18 apresenta como a ferramenta pode ser adicionada a caixa de ferramentas do QGIS, vale ressaltar que uma vez adicionada não é necessário recarregar o arquivo novamente, apenas buscar pelo nome da ferramenta no campo pesquisar da Caixa de ferramentas de processamento. E a Figura 19 apresenta como buscar um plugin e realizar a instalação da ferramenta no QGIS. Destaca-se aqui que para disponibilização no repositório oficial de completos é necessário obter um ID OSGEO, carregar o complemento e esperar a aprovação por um membro da equipe. Cada versão desenvolvida da ferramenta deverá passar pelo processo de aprovação pelos membros da OSGEO.

Figura 18 - Adicionando a ferramenta como ferramenta de procesamento QGIS



Elaboração: adaptado de QGIS.

Figura 19 - Instalação da ferramenta como plugin do QGIS



Elaboração: adaptado de QGIS.

7 CONCLUSÕES

O objetivo deste trabalho foi atingido de forma satisfatória visto que foi desenvolvida uma ferramenta em software SIG de código aberto que estima e mapeia a distribuição espacial da produtividade das culturas de milho, trigo e girassol, utilizando RNAs e imagens multiespectrais Sentinel 2. Além disso, conta com a opção de validação dos dados gerados a fim de obter a acurácia de mapeamento. O modelo de RNA para previsão de produtividade foi satisfatório para as culturas de trigo e girassol, as quais neste estudo foram treinadas com menor quantidade de amostras. Para estas culturas foram obtidos RMSEs do teste/validação externa menores que 30%, valor objetivado neste estudo.

Se tratando da cultura do milho, cultura a qual o modelo implementado é incapaz de prever valores de produtividade, há necessidade de estudos mais aprofundados a fim de obter um modelo apto para estimativas. Dentre estes estudos podemos citar a investigação da correlação dos dados de entrada com a variável produtividade do milho; o estudo da quantidade ideal de amostras para treinamento do modelo e; a influência da localização da área para geração do modelo, visto que neste trabalho foram utilizadas áreas situadas no Brasil e na França.

Os modelos de RNA implementados nesta ferramenta necessitam de constante atualização, já que estão aptos a prever valores de produtividade de acordo com o domínio dos dados utilizados para treinamento. Dessa forma, a medida que forem obtidos novos dados, os processos de treinamento e validação deverão ser realizados novamente, bem como a atualização dos modelos de previsão na ferramenta.

Para trabalhos futuros a serem desenvolvidos em relação a esta ferramenta podemos citar diversas linhas, como por exemplo, técnicas mais avançadas para definição dos hiperparâmetros dos modelos de RNA; a geração de modelos de previsão para outras culturas; a implementação de outros modelos de previsão baseados em AM, como a Regressão Linear, Florestas Aleatórias, Máquina de Vetores de Suporte entre outras; o estudo mais completo em relação as bandas mais correlacionadas a variável produtividade de culturas; e a utilização de imagens multiespectrais advindas de outros satélites para previsões.

Quanto a limitação deste trabalho ressaltamos a não validação com usuário, a fim de identificar a relevância e viabilidade da ferramenta, bem como possíveis adaptações, devido as condições pandêmicas vivenciadas. Além disso, lembramos que, estimar a produtividade de culturas é uma tarefa desafiadora, já que está relacionada com as condições ambientais, estrutura física da cultura, composição do solo e práticas de manejo, dessa forma, recomenda-se a utilização desta metodologia associada a outras já utilizadas.

REFERÊNCIAS

- AL-GAADI, K. A. et al. *Prediction of potato crop yield using precision agriculture techniques. PloS one*, v. 11, n. 9, p. e0162219, 2016.
- ARONOFF, S. *Geographic information systems. A management perspective*, v. 4, n. 4, p. 58, 1989.
- ARTEGA, A. J. M. **Aplicação de imagens orbitais na espacialização do volume de *Eucalyptus* sp. via redes neurais artificiais.** 2019. 92 f. Dissertação (Mestrado) - Curso de Planejamento e Uso de Recursos Renováveis, Centro de Ciências e Tecnologias Para A Sustentabilidade, Universidade Federal de São Carlos, Sorocaba, 2019.
- AVINASH, V. S. *Understanding activation functions in neural networks (Machine Learning Mastery)*. 2017. Disponível em: <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>. Acesso em: set. 2021.
- BANNERJEE, Gouravmoy et al. *Artificial intelligence in agriculture: A literature survey. International Journal of Scientific Research in Computer Science Applications and Management Studies*, v. 7, n. 3, p. 1-6, 2018.
- BASHEER, I. A.; HAJMEER, M. *Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods*, v. 43, n. 1, p. 3-31, 2000.
- BASSO, B. et al. *Review of crop yield forecasting methods and early warning systems. In: Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy.* 2013.
- BERTOLIN, N. O. et al. *Predição da produtividade de milho irrigado com auxílio de imagens de satélite. Revista Brasileira de Agricultura Irrigada*, v. 11, n. 4, p. 1627, 2017.
- BESTELMEYER, B. T. et al. *Scaling Up Agricultural Research With Artificial Intelligence. IT Professional*, v. 22, n. 3, p. 33-38, 2020.
- BINOTI, M. L. M. S. et al. *Aplicação de redes neurais artificiais para estimação da altura. Revista Árvore*, v. 37, n. 4, p. 639-645, 2013.
- BRAGA, A.P. et al. **Redes neurais artificiais: teoria e aplicações.** Rio de Janeiro: LTC, 2000. 250 p.
- BURROUGH, P.A. *Principles of geographical. Information systems for land resource assessment.* Clarendon Press, Oxford, 1986
- BOTTOU, L. *Online algorithms and stochastic approximations. Online Learning*, v. 5, p. 6-3, 1998.
- CAI, Y. et al. *Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and forest meteorology*, v. 274, p. 144-159, 2019.
- CALERA, A. et al. *Remote sensing for crop water management: from ET modelling to services for the end users. Sensors*, v. 17, n. 5, p. 1104, 2017.
- CÂMARA, G.; FREITAS, U. *Perspectivas em sistemas de informação geográfica. Fator GIS*, v. 3, n. 10, p. 31-34, 1995.

- CHANDER, G.; MARKHAM, B. L.; HELDER, D. L. *Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote sensing of environment*, v. 113, n. 5, p. 893-903, 2009.
- CHATTERJEE, S. et al. *Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Computing and Applications*, v. 28, n. 8, p. 2005-2016, 2017.
- CHIPANSHI, A. et al. *Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. Agricultural and Forest Meteorology*, v. 206, p. 137-150, 2015.
- COPERNICUS. *Copernicus Services*. Disponível em: <https://www.copernicus.eu/en>. Acesso em: set.2021.
- DOSHI, N. *Deep Learning Best Practices: weight initialization*. 2018. Disponível em: <https://medium.com/usf-msds/deep-learning-best-practices-1-weight-initialization-14e5c0295b94>. Acesso em: set. 2021.
- DUVEILLER, G.; CESCATTI, A. *Spatially downscaling sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity. Remote Sensing of Environment*, v. 182, p. 72-89, 2016.
- DRUMMOND, S. T. et al. *Statistical and neural methods for site-specific yield prediction. Transactions of the ASAE*, v. 46, n. 1, p. 5, 2003.
- ENGESAT. *Imagens de Satélite*. Disponível em: <http://www.engesat.com.br/imagem-de-satelite/>. Acesso em: nov. 2020.
- EOS. **Índices de vegetação: uma análise de vegetação completa**. Disponível em: <https://eos.com/pt/blog/indices-de-vegetacao/>. Acesso em: set. 2021.
- ESA. *Products and Algorithms*. Disponível em: <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/products-algorithms>. Acesso em: set. 2021.
- FAO. **Agricultural Outlook 2015**. Disponível em: https://www.oecd-ilibrary.org/agriculture-and-food/oecd-fao-agricultural-outlook-2015_agr_outlook-2015-en. 2015. Acesso em: 25 nov. 2020.
- FRIEDMAN, J. H. *The elements of statistical learning: Data mining, inference, and prediction*. Springer Open, 2017.
- GASO, D.V. et al. *Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. Comput. Electron. Agric.* 2019, 159, 75–83.
- GEORGE, T. *A new look at agricultural development and the non-agriculture economy in low-income countries. Global Food Security*, v. 26, p. 100449, 2020.
- GOUTTE, C. *Note on free lunches and cross-validation. Neural Computation*, v. 9, n. 6, p. 1245-1249, 1997.
- GUAN, K. et al. *The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote sensing of environment*, v. 199, p. 333-349, 2017.

- GUIMARÃES, E. S. **Aprendizado de Máquina aplicado à predição da produtividade da cultura da soja utilizando dados de clima e solo**. 2019. 78 f. Dissertação (Mestrado) - Curso de Matemática, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - Usp, São Carlos, 2019.
- HARA, P.; PIEKUTOWSKA, M.; NIEDBAŁA, G.. *Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data*. **Land**, v. 10, n. 6, p. 609, 2021.
- HAYKIN, S. **Redes Neurais: princípios e práticos**. 2. ed. Porto Alegre: Bookman, 2001. 899 p. Tradução: Paulo Martins Engel.
- HE, Mingzhu et al. *Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data*. **Remote Sensing**, v. 10, n. 3, p. 372, 2018.
- HENRY, Robert. *Innovations in agriculture and food supply in response to the COVID-19 pandemic*. **Molecular plant**, v. 13, n. 8, p. 1095, 2020.
- HUANG, Y. et al. *Agricultural remote sensing big data: Management and applications*. **J. Integr. Agric.** v. 17, p. 1915–1931, 2018.
- JAIN, A. K. et al. *Artificial neural networks: A tutorial*. **Computer**, v. 29, n. 3, p. 31-44, 1996.
- JAMES, G. et al. **An introduction to statistical learning**. New York: springer, 2013.
- JENSEN, J. R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Parêntese Editora, 2009. Tradução: José Carlos Epiphânio.
- JUDGE, J. *Microwave remote sensing of soil water: Recent advances and issues*. **Transactions of the ASABE**, v. 50, n. 5, p. 1645-1649, 2007.
- JUNG, J. et al. *The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems*. **Current Opinion in Biotechnology**, v. 70, p. 15-22, 2020.
- KARTHIKEYAN, L. et al. *A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses*. **Journal of Hydrology**, p. 124905, 2020.
- KIM, N. et al. *A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern united states, 2006–2015*. **ISPRS International Journal of Geo-Information**, v. 8, n. 5, p. 240, 2019.
- KOGAN, F. et al. *Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models*. **International Journal of Applied Earth Observation and Geoinformation**, v. 23, p. 192-203, 2013.
- KROSS, A. et al. *Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields*. **Remote Sensing**, v. 12, n. 14, p. 2230, 2020.
- LEAL, A. J.F. et al. *Redes neurais artificiais na predição da produtividade de milho e definição de sítios de manejo diferenciado por meio de atributos do solo*. **Bragantia**, v. 74, n. 4, p. 436-444, 2015.
- LEDA, V. C. et al. *Sensoriamento remoto aplicado a modelagem de produtividade da cultura da cana-de-açúcar*. **Energia na Agricultura**, v. 34, n. 2, p. 263-270, 2019.

- LOONEY, Carl G. *Advances in feedforward neural networks: demystifying knowledge acquiring black boxes*. **IEEE Transactions on Knowledge and Data Engineering**, v. 8, n. 2, p. 211-226, 1996.
- MA, Yuchi et al. *Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach*. **Remote Sensing of Environment**, v. 259, p. 112408, 2021.
- MAHLEIN, A. K. *Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping*. **Plant disease**, v. 100, n. 2, p. 241-251, 2016.
- MASTERS, T. **Practical neural network recipes in C++**. Morgan Kaufmann, 1994.
- MULLA, D. J. *Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps*. **Biosystems engineering**, v. 114, n. 4, p. 358-371, 2013.
- NOLAND, R. L. et al. *Estimating alfalfa yield and nutritive value using remote sensing and air temperature*. **Field Crops Research**, v. 222, p. 189-196, 2018.
- NOVO, E. M.L. **Sensoriamento Remoto: princípios e aplicações**. Editora Blucher, 2010.
- PENG, C. et al. *Relationships between remote-sensing-based agricultural drought indicators and root zone soil moisture: a comparative study of Iowa*. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 11, p. 4572-4580, 2014.
- PINTER, P. J. et al. *Remote sensing for crop management*. **Photogrammetric Engineering & Remote Sensing**, v. 69, n. 6, p. 647-664, 2003.
- PLUTOWSKI, M. et al. **Cross-validation estimates IMSS. Training (as training can be faster on smaller datasets)**, v. 2, p. 4, 1994.
- POVH, F. P. et al. *Comportamento do NDVI obtido por sensor ótico ativo em cereais*. **Pesquisa Agropecuária Brasileira**, v. 43, p. 1075-1083, 2008.
- QGIS. **QGIS Python Plugins Repository: Pat - Precision Agriculture Tools**. Disponível em: <https://plugins.qgis.org/plugins/pat/>. Acesso em: 05 dez. 2020.
- RAO, R. Bharat; FUNG, Glenn; ROSALES, Romer. *On the dangers of cross-validation. An experimental evaluation*. In: **Proceedings of the 2008 SIAM international conference on data mining**. Society for Industrial and Applied Mathematics, 2008. p. 588-596.
- RUDORFF, B. F. T.; MOREIRA, M. A. **Sensoriamento Remoto Aplicado a Agricultura**, Capítulo 9. 2002, INPE, São José dos Campos. INPE-8984-PUD/62.
- SEELAN, S. K. et al. *Remote sensing applications for precision agriculture: A learning community approach*. **Remote sensing of environment**, v. 88, n. 1-2, p. 157-169, 2003.
- SHANAHAN, J. F. et al. *Use of remote-sensing imagery to estimate corn grain yield*. **Agronomy Journal**, v. 93, n. 3, p. 583-589, 2001.
- SILVA, A. A. V. et al. *Estimativa da produtividade de trigo em função da adubação nitrogenada utilizando modelagem neuro fuzzy*. **Revista Brasileira de Engenharia Agrícola e Ambiental**, n. 18, p. 180-187, 2014.
- SILVA, F. **Aplicação de redes neurais artificiais para estimar a produtividade do forwarder na extração de madeira**. 2015. 50f. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa. 2015.

- SILVA, P. F. P. **Desenvolvimento de um plugin em ambiente SIG open source (QGIS) para obter parâmetros e métricas de copas de árvores, através de imagens adquiridas por VANT**. 2017. 91 f. Dissertação (Mestrado) - Curso de Engenharia Geográfica, Faculdade de Ciências, Universidade do Porto, Porto, 2017.
- SKAKUN, S. et al. *Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model*. **Remote Sensing of Environment**, v. 195, p. 244-258, 2017.
- SKAKUN, S. et al. *Winter wheat yield assessment using Landsat 8 and Sentinel-2 data*. In *Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 22–27 July 2018; p. 5964–5967.
- SOARES, F. C. et al. *Redes neurais artificiais na estimativa da retenção de água do solo*. **Ciência Rural**, v. 44, p. 293-300, 2014.
- SOARES, Fátima Cibele et al. *Predição da produtividade da cultura do milho utilizando rede neural artificial*. **Ciência Rural**, v. 45, p. 1987-1993, 2015.
- SRIPADA, R. P. et al. *Aerial color infrared photography for determining early in-season nitrogen requirements in corn*. **Agronomy Journal**, v. 98, n. 4, p. 968-977, 2006.
- SWINGLER, K. **Applying neural networks: a practical guide**. Morgan Kaufmann, 1996.
- THENKABAIL, P. S. *Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images*. **International Journal of Remote Sensing**, v. 24, n. 14, p. 2879-2904, 2003.
- THORP, K. R.; TIAN, L. F. *A review on remote sensing of weeds in agriculture*. **Precision Agriculture**, v. 5, n. 5, p. 477-508, 2004.
- TUVDENDORJ, B. et al. *Determination of Appropriate Remote Sensing Indices for Spring Wheat Yield Estimation in Mongolia*. **Remote Sensing**, v. 11, n. 21, p. 2568, 2019.
- VARELLA, C. A. A.; MEDEIROS, R.; GLERIANI, J. M. **Agricultura de Precisão e Sensoriamento Remoto**. In: SANTOS, F.; BORÉM, A.; CALDAS, C. *Cana-de-açúcar: bioenergia, açúcar e etanol: 58 tecnologias e perspectivas*. 2ed.rev e ampl. Viçosa, MG: Os editores, 2011. Cap. 9, p. 258.
- WEISS, M.; JACOB, F.; D. G. *Remote sensing for agricultural applications: A meta-review*. **Remote Sensing of Environment**, v. 236, p. 111402, 2020.
- WOLFERT, S. et al. *Big data in smart farming—a review*. **Agricultural Systems**, v. 153, p. 69-80, 2017.
- ZANELLA, A. R. A. et al. *Security challenges to Smart Agriculture: Current State, Key Issues, and Future Directions*. **Array**, p. 100048, 2020.
- ZHANG, Q. et al. *Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model*. **Remote Sensing of Environment**, v. 99, n. 3, p. 357-371, 2005.