

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA GRADUAÇÃO
CURSO DE ENGENHARIA ELÉTRICA

MOACYR JÚNIOR TROMBETTA

**MODELO DE DETECÇÃO POR IMAGENS DO ATAQUE DA SPODOPTERA
FRUGIPERDA EM PLANTAÇÕES DE MILHO**

São Leopoldo
2022

Moacyr Júnior Trombetta

**MODELO DE DETECÇÃO POR IMAGENS DO ATAQUE DA SPODOPTERA
FRUGIPERDA EM PLANTAÇÕES DE MILHO**

Trabalho de Conclusão de Curso apresentado
como requisito parcial para a obtenção do título
de Engenheiro Eletricista pela Universidade do
Vale do Rio dos Sinos — UNISINOS

Orientador:
Prof. Dr. Rodrigo de Marques Figueiredo

São Leopoldo
2022

RESUMO

Os métodos tradicionais de monitoramento de pragas que utilizam técnicas baseadas em avaliação visual humana e aferição manual de plantas, são métodos geralmente custosos, lentos e tolerantes a erros. Identificar a presença de pragas no início da sua invasão facilita o seu controle, diminuindo custos com pesticidas e preservando a produtividade da plantação. No ramo da inteligência artificial existem ferramentas robustas de análise visual que são empregadas com sucesso em diversas áreas, tornando-se tendência para a agricultura. Sendo assim, este estudo propôs a utilização de redes neurais convolucionais para identificar e classificar o dano do ataque da *Spodoptera frugiperda* na folha do milho. O modelo utilizado foi o *Faster R-CNN*, juntamente com o extrator de características *VGG16*. O modelo foi treinado para identificar e classificar quatro classes. Uma classe de folhas saudáveis e três classes de folhas que apresentam dano, sendo eles: mínimo, moderado e grave. Para isso, foi criado um *dataset* composto por 1263 imagens e 2637 âncoras que representam cada classe. Para analisar o desempenho do modelo foram criados três cenários de testes. Os cenários se diferenciam pela forma de que as classes estão organizadas. Para medir o desempenho do modelo foi utilizada a métrica *Mean Average Precision* (mAP). No cenário 1, composto pelas classes saudável, mínimo, moderado e grave, o modelo resultou na mAP de 43,41%. No cenário 2, formado pelas classes saudável e dano, onde a classe dano passa a ser o agrupamento da mínimo, moderado e grave, o modelo atingiu a mAP de 73,09%. Por fim, no cenário 3, composto também pelas classes saudável e dano, porém desta vez, agrupando o dano mínimo com a classe saudável, a formação obteve a mAP de 74,77%.

Palavras-chave: *Spodoptera frugiperda*. redes neurais convolucionais. *Faster R-CNN*. detecção de danos por peste.

LISTA DE FIGURAS

Figura 1 – <i>Spodoptera frugiperda</i> no sexto instar	8
Figura 2 – Propagação da <i>Spodoptera frugiperda</i> no mundo	9
Figura 3 – Escala de Davis	11
Figura 4 – Diagrama do neurônio artificial	12
Figura 5 – Diagrama de uma rede neural convolucional	13
Figura 6 – Diagrama da Faster R-CNN	15
Figura 7 – Diagrama geral do estudo	28
Figura 8 – Ferramenta labelImg	30
Figura 9 – Exemplo do resultado da técnica de rotação	31
Figura 10 – Estrutura do extrator <i>VGG16</i>	32
Figura 11 – Fluxograma representando a calibração do modelo	34
Figura 12 – Variação da mAP no cenário 1	37
Figura 13 – Função de perda do RPN no cenário 1	38
Figura 14 – Função de perda do ROI no cenário 1	38
Figura 15 – Curva do somatório de perdas no cenário 1	39
Figura 16 – Predição 1 do cenário 1	40
Figura 17 – Âncoras reais da Figura 16	40
Figura 18 – Predição 2 do cenário 1	41
Figura 19 – Âncoras reais da Figura 18	41
Figura 20 – Variação da mAP no cenário 2	42
Figura 21 – Função de perda do RPN no cenário 2	43
Figura 22 – Função de perda do ROI no cenário 2	43
Figura 23 – Curva do somatório de perdas no cenário 2	44
Figura 24 – Predição 1 do cenário 2	45
Figura 25 – Âncoras reais da Figura 24	45
Figura 26 – Predição 2 do cenário 2	46
Figura 27 – Âncoras reais da Figura 26	46
Figura 28 – Variação da mAP no cenário 3	47
Figura 29 – Função de perda do RPN no cenário 3	48
Figura 30 – Função de perda do ROI no cenário 3	48
Figura 31 – Curva do somatório de perdas do cenário 3	49
Figura 32 – Predição 1 do cenário 3	49
Figura 33 – Âncoras reais da Figura 32	50
Figura 34 – Predição 2 do cenário 3	50
Figura 35 – Âncoras reais da Figura 34	51

LISTA DE TABELAS

Tabela 1 – Tabela de precisão média de cada modelo para as dez classes.	21
Tabela 2 – Artigos correlatos	24
Tabela 3 – Artigos correlatos	25
Tabela 4 – Artigos correlatos	26
Tabela 5 – Hiper-parâmetros dos cenário 1, 2 e 3	36
Tabela 6 – Dataset do cenário 1	37
Tabela 7 – AP do cenário 1	37
Tabela 8 – Dataset do cenário 2	42
Tabela 9 – AP do cenário 2	42
Tabela 10 – Dataset do cenário 3	47
Tabela 11 – AP do cenário 3	47
Tabela 12 – Matriz de confusão do cenário 4	52
Tabela 13 – Comparativo dos cenários com o estudo de Du et al. (2022)	52

SUMÁRIO

1	INTRODUÇÃO	6
2	REFERENCIAL TEÓRICO	7
2.1	Conceitos Básicos	7
2.1.1	Fall Armyworm (Lagarta Militar)	7
2.1.2	Escala de Davis	10
2.1.3	Redes Neurais Artificiais	12
2.1.4	Redes Neurais Convolucionais	13
2.1.5	Modelo Faster R-CNN	15
2.2	Estado da Arte	16
3	METODOLOGIA	28
3.1	Visão Geral	28
3.2	Configuração proposta	29
3.2.1	Banco de Dados de Imagens	29
3.2.2	Anotação das Imagens	29
3.2.3	Enriquecimento do Dataset	30
3.2.4	VGG16	31
3.2.5	Faster R-CNN	32
3.2.6	Métrica de Avaliação	35
4	ANÁLISES DE RESULTADOS	36
4.1	Cenário 1	36
4.2	Cenário 2	41
4.3	Cenário 3	46
4.4	Cenário 4	51
4.5	Análise Geral	52
5	CONCLUSÃO	54
5.1	Trabalhos Futuros	54
	REFERÊNCIAS	56

1 INTRODUÇÃO

A *Spodoptera frugiperda* é uma das principais pragas do milho. Essa larva se alimenta de mais de oitenta tipos de plantas como milho, trigo, arroz e sorgo. Está presente em mais de setenta países, nos continentes americano, africano, asiático e recentemente na Oceania. Seus estragos atingem as grandes plantações e a agricultura de subsistência (IPPC SECRETARIAT, 2021). Além disso, segundo Eschen et al. (2021) a invasão de biomas por novas espécies de seres vivos, consequentes da ação humana, causa um custo anual estimado de 65,58 bilhões de dólares para a agricultura africana, sendo a *Spodoptera frugiperda* a praga mais causadora de prejuízo, atingindo 9,4 bilhões de dólares.

Uma forma de monitorar a presença da *Spodoptera frugiperda* em plantações de milho é através da aferição das folhas por amostragem. É verificado em determinados pontos da plantação, se as folhas do milho apresentam dano causado pela larva (SCHNEIDER; BAUMGRATZ, 2015). Entretanto, esse método de avaliação visual feita por trabalhadores do campo é limitado, lento e propenso a erros devida a subjetividade humana ao avaliar. A utilização de um sistema de detecção automática pode tornar esse trabalho menos oneroso, mais rápido e preciso, possibilitando cobrir áreas maiores de plantação. Esses benefícios proporcionam uma resposta mais rápida e adequada aos ataques da *Spodoptera frugiperda*.

Portanto, com o objetivo de auxiliar no monitoramento de ataques da *Spodoptera frugiperda*, este estudo propõe o uso de um modelo de detecção por imagens do dano causado pela larva na folha do milho. Através de imagens das folhas do milho, o modelo será capaz de identificar se existe algum tipo de dano, e caso exista, fará a classificação dentro de três classes: mínimo, moderado e grave.

Como critério para escolha do modelo foi pesquisado por técnicas e algoritmos robustos e que apresentam os melhores resultados atualmente. No domínio do aprendizado profundo, existem as redes neurais convolucionais, que são a base do modelo utilizado. Os modelos de redes neurais convolucionais são consolidados, presentes em aplicações comerciais e amplamente utilizados em competições relacionadas a reconhecimento de imagens, detecção de objetos e segmentação semântica (ZHANG et al., 2021).

As seções deste estudo estão organizados da seguinte forma: no referencial teórico são explorados os conceitos básicos e essenciais para a compreensão do estudo, como fundamentos sobre a *Spodoptera frugiperda* e de redes neurais convolucionais. No referencial teórico também são exploradas publicações que empregaram redes neurais convolucionais como ferramentas no auxílio de problemas que envolvam pestes na agricultura. Na seção referente a metodologia são abordadas as técnicas e etapas para a configuração do modelo. Na quarta seção é feita a análise dos resultados obtidos. Já na última seção, encontra-se a conclusão, que faz considerações finais sobre o modelo e apresenta propostas de melhorias futuras.

2 REFERENCIAL TEÓRICO

Neste capítulo são abordados os conceitos básicos e a revisão do estado da arte. Nos conceitos básicos é explorada a base teórica necessária para o entendimento sobre: a lagarta militar, escala de Davis, redes neurais artificiais, redes neurais convolucionais e o modelo *Faster R-CNN*. Por fim, na revisão do estado da arte são apresentados estudos recentes que utilizam redes neurais convolucionais como parte de soluções na detecção de ataques de pragas e de doenças na agricultura.

2.1 Conceitos Básicos

Os conceitos básicos foram divididos em cinco principais assuntos. Na primeira subseção é contextualizada a *Spodoptera frugiperda*. Nessa subseção serão descritas as suas características físicas, suas etapas de desenvolvimento, o aspecto do seu dano na plantação de milho, as formas mais tradicionais de monitoramento e a sua propagação no mundo. Na segunda subseção é abordada a escala de Davis. Trata-se de uma escala consolidada na agricultura e que tem o objetivo de classificar em dez níveis o dano causado pela *Spodoptera frugiperda* na folha do milho. Na terceira subseção são apresentadas as redes neurais artificiais. Essa área de estudo é responsável pelo desenvolvimento de modelos computacionais inspirados no cérebro humano. Esses modelos são capazes de reconhecer padrões através de aprendizado. As redes neurais artificiais foram essenciais para a criação das redes neurais convolucionais, que serão tratadas na quarta subseção. Por sua vez, essa classe de rede neural é específica para o processamento de imagens. Na última seção é exposto o modelo *Faster R-CNN*, que juntamente com as redes neurais convolucionais, é uma ferramenta fundamental para o desenvolvimento deste estudo.

2.1.1 Fall Armyworm (Lagarta Militar)

É chamada de *fall armyworm* (FAW), a larva da mariposa de nome científico *Spodoptera frugiperda*. A larva também é popularmente conhecida no Brasil como lagarta militar ou lagarta do cartucho. O nome *frugiperda* vem do latim, onde *frugi* remete-se a fruta e *perda* refere-se aos danos econômicos causados pelos hábitos alimentares da lagarta (MCLEOD, 2006).

A aparência física da FAW depende do seu instar (SARKOWI; MOKHTAR, 2021). Instares são etapas do desenvolvimento larval divididos pela alteração na proporção do corpo, cores, padrões e etc. Na FAW, seu ciclo larval é formado por seis instares (CAPINERA, 2002). Nos primeiros estágios sua cor é verde claro e no decorrer dos instares vai se tornando marrom com listras longitudinais. Possuem nove *prolegs* (estruturas que ajudam na movimentação), oito são presente no abdômen e um último par no segmento final. Pontos pretos e elevados são aparentes nas áreas dorsais, sendo que, o último segmento é característico por possuir quatro deles. Outra característica marcante fica na cabeça da FAW, onde é possível identificar um "Y" invertido na

cor amarela. No final do seu desenvolvimento, a larva atinge de 3,8 a 5,1 centímetros de comprimento (SARKOWI; MOKHTAR, 2021). Na Figura 1 é possível identificar as características mencionadas de uma FAW no sexto instar.

Figura 1 – *Spodoptera frugiperda* no sexto instar



Fonte: Capinera (2002)

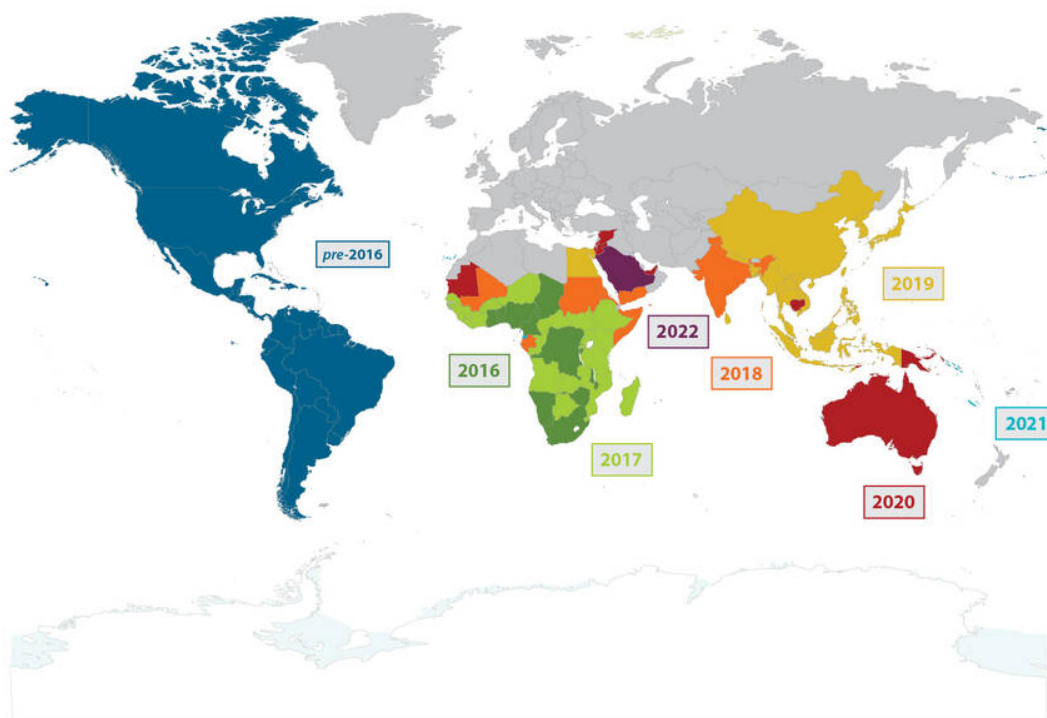
A *Spodoptera frugiperda* adulta, na forma de mariposa, pode viver de sete a vinte e um dias. São insetos que tem um ciclo de vida relativamente curto, porém, apresentam alta fecundidade. As fêmeas iniciam o processo de pré-oviposição após quatro ou cinco dias do início da fase adulta. Esse processo dura de três a quatro dias, e posteriormente a ele, a *Spodoptera frugiperda* fica apta a depositar seus ovos. O processo de oviposição pode demorar até três semanas. A eclosão dos ovos leva de dois a três dias em clima quente (20-30°C), em condições frias esse tempo pode aumentar para até onze dias. O período larval é dividido em seis instares que podem totalizar quatorze até trinta dias, também dependendo da temperatura ambiente (SARKOWI; MOKHTAR, 2021).

O dano causado pela FAW é identificado em folhagens. Inicialmente, no primeiro instar, as larvas apenas consomem o tecido foliar de um lado da folha, deixando a camada epidérmica oposta sem danos. A partir do segundo e terceiro instares, os danos são maiores e as folhas passam a sofrer com buracos, que são geralmente feitos da borda para o centro. Já nos instares mais avançados, a FAW causa grave desfoliação, deixando a planta do milho completamente esfarrapada, sobrando apenas as costelas e os caules. Além das folhas, nestes instares a lagarta também ataca o cartucho e a espiga do milho, reduzindo drasticamente o crescimento da planta (CAPINERA, 2002). Algumas variáveis como tamanho da população, estação do ano, saúde da plantação atacada, presença de inimigos biológicos, controle com inimigos naturais e patogênicos, devem ser considerados para determinar a gravidade do ataque da FAW. Plantas saudáveis e nutridas podem se recuperar rapidamente de ataques foliares. Porém, se não controlados, os ataques possivelmente vão prejudicar a produtividade da plantação (SARKOWI; MOKHTAR, 2021). O estudo de Marengo, Foster e Sanchez (1992) indica que o milho doce nas últimas fases

vegetativas é mais sensível aos ataques da FAW, diminuindo sua produção de 5 a 20% quando exposto a densidades de 0,2 a 0,8 larvas por planta.

As primeiras populações de FAW foram identificadas nas regiões tropicais do oeste dos Estados Unidos, percorrendo o continente americano até chegar na Argentina. No continente africano, nas regiões oeste e central, foram reportados as primeiras populações em 2016. Já em 2020, a FAW estava presente em outras regiões da África e também na Europa (CAPINERA, 2002). Os continentes africano e asiático possuem muitas regiões com clima e plantações favoráveis para o crescimento da população da FAW. Devido a capacidade da lagarta em se adaptar em diferentes condições climáticas, isso pode torná-la uma peste permanente e não apenas migratória (SARKOWI; MOKHTAR, 2021). Na Figura 2 é possível visualizar a propagação da FAW pelo mundo.

Figura 2 – Propagação da *Spodoptera frugiperda* no mundo



Fonte: FAO (2022)

O monitoramento de populações de *Spodoptera frugiperda* pode ser realizado através de armadilhas com luz negra ou feromônio, sendo a de feromônio mais eficiente. É recomendado utilizar esse tipo de monitoramento preferencialmente na fase da formação do cartucho do milho. A instalação das armadilhas deve ser abaixo das folhas superiores da planta. A captura de mariposas não é um indicador preciso de densidade, porém, indica a presença da peste na área. Após, identificada a presença, é necessário procurar por larvas e ovos nas folhas (CAPINERA,

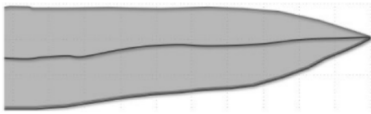
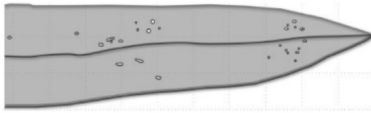
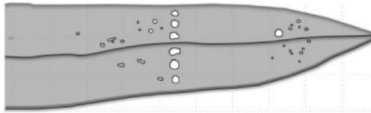
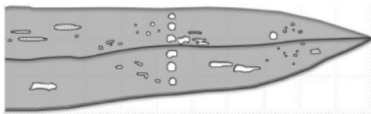
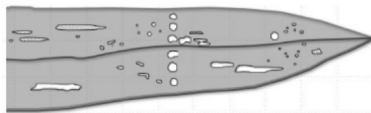
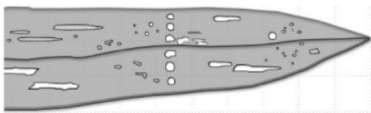
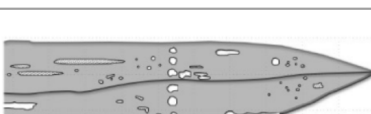
2002).

Outra forma de monitorar a presença e o grau de desenvolvimento da FAW na plantação de milho é através de aferição da plantação por amostragem. Essa aferição tem o objetivo de identificar a gravidade dos danos causados na folha. Cada amostragem deve ser feita para áreas de no máximo 225 hectares. Deve-se percorrer a plantação de forma com que o caminho traçado forme a letra "W", se observado de uma visão superior do terreno. Cinco estações de amostras são criadas, uma para cada extremidade do percurso em "W". Nas estações devem ser inspecionadas 20 plantas em sequência e registrada para cada uma delas a nota do dano, de acordo com a escala de Davis. Após, são somadas a quantidade de plantas que apresentaram notas maiores ou iguais a 3, e calculada a proporção que elas representam na amostra total. Devem ser feitas aplicações de inseticidas na área quando a proporção resultante for igual ou maior a 20% (SCHNEIDER; BAUMGRATZ, 2015). Na próxima subseção será exposta a escala de Davis, assim como, a descrição de cada um dos seus níveis de dano. Nesse estudo a escala de Davis foi escolhida como a escala referência para classificar o ataque da FAW na folha do milho.

2.1.2 Escala de Davis

Para avaliar o impacto das pragas na agricultura, as escalas de danos são amplamente utilizadas. Além disso, elas auxiliam na avaliação da eficácia dos métodos de manejo contra as pragas. Para mensurar o dano causado pela FAW no milho, existem alguns principais estudos utilizados como referência. Alguns desses estudos consideram apenas se a planta possui dano ou não, já outros, possuem escalas que ordenam a severidade do dano. Geralmente o primeiro nível é referente a uma planta sem nenhum tipo de lesão, e o último a uma planta completamente destruída (TOEPFER et al., 2021). Segundo Toepfer et al. (2021), a escala mais utilizada para estimar o dano da FAW no milho é a escala de Davis. Essa escala possui dez níveis que estão no intervalo de 0 a 9. A Figura 3 apresenta os níveis da escala de Davis e suas respectivas características.

Figura 3 – Escala de Davis

Nível	Descrição	
0	Sem dano visível	
1	Somente buracos na folha do cartucho	
2	Buracos e pequenas lesões circulares na folha do cartucho	
3	Buracos, pequenas lesões circulares e algumas pequenas (<1,3cm) lesões alongadas	
4	Pequenas lesões alongadas e algumas médias (1,3~2,5cm) lesões alongadas	
5	Pequenas lesões alongadas e várias médias (1,3~2,5cm) lesões alongadas	
6	Pequenas e médias lesões alongadas e algumas lesões grandes (>2,5cm)	
7	Muitas pequenas e médias lesões alongadas e várias lesões grandes (>2,5cm)	
8	Muitas pequenas e médias lesões alongadas e muitas lesões grandes (>2,5cm)	
9	Folhas do cartucho quase que totalmente destruídas	

Fonte: Adaptado de Toepfer et al. (2021)

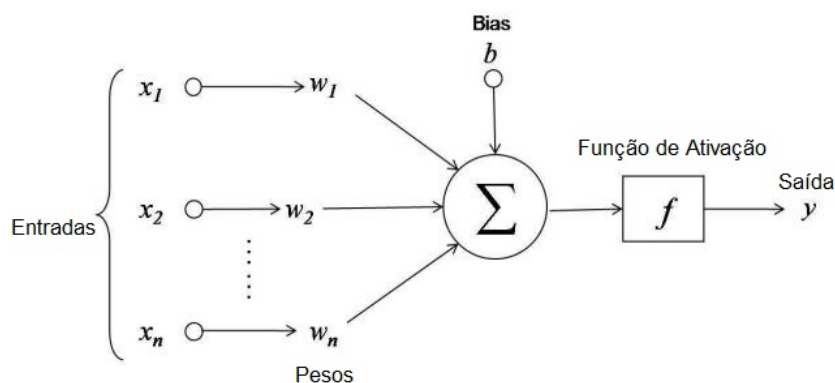
Devido a popularidade da escala Davis na agricultura e por ser uma referência na tomada de decisões contra ataques da FAW, esse estudo propôs utilizá-la como parâmetro para classificar o dano do ataque. Entretanto, os níveis atuais da escala serão agrupadas de acordo com suas variações, agrupando níveis que possuam menor variação de característica, com o propósito de diminuir a quantidade total de níveis da escala original. O objetivo dessa nova divisão é de criar classes com menor complexidade, permitindo que as redes neurais convolucionais tenham uma maior precisão. Na próxima subseção será abordado o tema de redes neurais artificiais, as quais

são fundamentais para o desenvolvimento das redes neurais convolucionais.

2.1.3 Redes Neurais Artificiais

Diferentemente de um computador convencional, o cérebro humano processa informações de uma forma altamente complexa, não-linear e paralela. São através de estruturas compostas por neurônios que o cérebro é capaz de realizar processos como reconhecimento de padrões, percepção e controle. Essas características tem sido motivadoras do desenvolvimento de estudos sobre redes neurais artificiais (RNA). A criação de neurônios artificiais (Figura 4) permite que as redes neurais sejam capazes de realizar tarefas específicas de uma maneira próxima ao do cérebro humano. Essas redes são paralelamente formada por neurônios. Os neurônios por sua vez, se caracterizam por armazenar conhecimento adquirido a partir do processo de aprendizagem que a rede neural artificial é submetida (HAYKIN, 2003).

Figura 4 – Diagrama do neurônio artificial



Fonte: Adaptado de Haykin (2003)

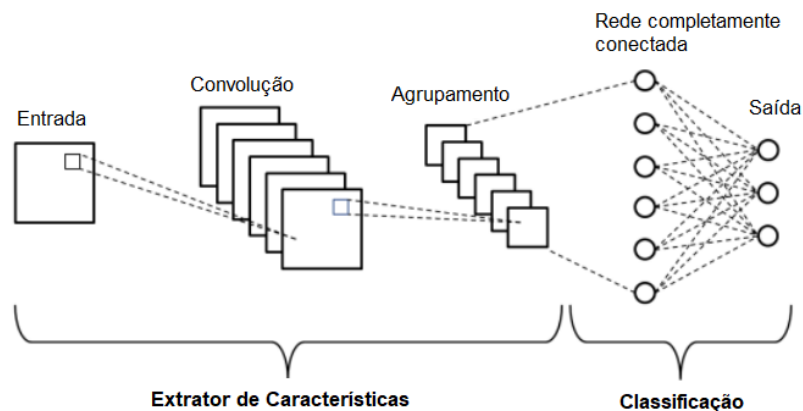
O processo de aprendizado tem a função de modificar os pesos sinápticos de cada neurônio com o objetivo de atingir o sucesso da tarefa proposta à rede. O neurônio artificial basicamente é formado por três elementos: o conjunto de sinapses ou elos de conexão, um somador e uma função de ativação. Os elos de conexão são característicos por armazenar os pesos sináptico W_k . Cada elo contém um peso. Quando submetido a um sinal, o elo multiplica esse sinal pelo seu peso W_k . Após, o somador faz o somatório de todas as sinapses resultantes no neurônio. Esse somatório tem seu valor normalizado pela função de ativação, geralmente definido dentro do intervalo unitário fechado $[0,1]$ ou $[-1,1]$ (HAYKIN, 2003). Por fim, quando o processo de aprendizado é executado, cada iteração desse processo retorna um valor resultante na função de ativação, esse valor resultante é comparado com o valor esperado para aquele neurônio, já que no aprendizado a rede possui a resposta ideal como referência. O objetivo da rede é obter o valor resultante mais próximo possível em relação ao esperado. Sendo assim, cada iteração tende a tornar os pesos das sinapses mais próximos aos ideais, para que a rede neural artificial

seja capaz de executar a tarefa treinada de forma correta, de acordo com o sinal de entrada aplicado.

2.1.4 Redes Neurais Convolucionais

Graças aos estudos aplicados em redes neurais artificiais, foi possível fazer o reconhecimento de imagens através de modelos. Entretanto, devido a limitação desse tipo de rede para problemas complexos onde a estrutura espacial de cada imagem é muito distinta e possuindo *pixels* dos seus elementos de interesse em posições diferentes no plano bi-dimensional, foi necessário desenvolver modelos mais eficientes para aprender com imagens (ZHANG et al., 2021). As redes neurais convolucionais (RNC) foram desenvolvidas para resolver os problemas enfrentados pelas RNA. Atualmente arquiteturas baseadas em RNC são as principais ferramentas no campo da visão computacional. A estrutura básica de uma RNC é formada pelo módulo extrator de características e pelo módulo classificador. Na Figura 5 é exibido o diagrama de uma RNC formada pelos seus dois principais módulos.

Figura 5 – Diagrama de uma rede neural convolucional



Fonte: Adaptado de Phung e Rhee (2019)

O módulo extrator de características é o principal módulo de uma RNC. Esse módulo é responsável por extrair as características das imagens através de sucessivas operações de convolução. O módulo extrator de características é dividido em três principais camadas que são chamadas de: camada de entrada, camada de convolução e camada agrupamento. Na camada de entrada são inseridos dados de imagens. Imagens podem ser coloridas, compostas pelos canais vermelho, verde e azul, possuir altura e largura. Portanto, é possível representá-las a partir de um tensor de terceira ordem. Ou seja, os dados de entrada formam uma matriz tridimensional, onde cada dimensão é chamada de canal. Posteriormente, após sofrerem convoluções, esses canais também podem ser chamados de mapas de características. As camadas de convolução tem o objetivo de resolver o problema de invariância espacial. Esse problema, como já mencionado anteriormente, consiste na dificuldade de uma rede neural artificial em identificar

elementos em imagens complexas. Basicamente, as camadas de convolução são capazes de varrer imagens e atribuir pontuações para características que tenham probabilidade de ajudar na identificação do elemento alvo, sendo assim, permitindo com que a rede neural processe apenas informações possivelmente úteis e com menos parâmetros. Nas primeiras convoluções, independentemente da posição da imagem onde haja um possível elemento alvo, as respostas para todos esses elementos serão as mesmas e passados para o novo mapa de características gerado. Esse princípio é chamado de invariância de tradução. O outro princípio consiste no da localidade. Esse princípio afirma que é também nas primeiras camadas que a rede deve se concentrar nos prováveis elementos alvos, desconsiderando o resto do contexto. Porém, isso não impede que a rede posteriormente utilize os demais contextos da imagem para fazer previsões. Após convoluções, a camada de agrupamento geralmente é ativada. Essa camada tem a função de aplicar técnicas chamadas de *pooling*. Essas técnicas tem como resultado na sua saída, a criação de um novo mapa de características. A proposta dessa camada é de reduzir gradualmente a resolução do mapa, juntando informações. Após sucessivas convoluções seguidas de *pooling*, é esperado mapas de características cada vez mais grosseiros. Esse tipo de agrupamento tem propósito de aliviar a sensibilidade das camadas convolutivas à localização e diminuição da taxa de amostragem espacial (ZHANG et al., 2021).

Nos últimos anos diversos módulos extratores de características foram desenvolvidos, porém, nem todos apresentam a mesma proposta. Existem diversos fatores a serem considerados na definição do uso de um determinado módulo extrator. Extratores compostos por quantidades maiores de camadas e parâmetros podem atingir melhores acurácias, porém, geralmente exigem maior poder computacional e tempo de resposta. Ou seja, a escolha do extrator deve levar em consideração fatores como o poder computacional disponível, tempo de resposta e acurácia desejados (FUENTES et al., 2017). Alguns exemplos consolidados de extratores de características são: *AlexNet*, *ResNet*, *GoogLeNet*, *VGG*, *LeNet*, *ResNet* e *InceptionV3*.

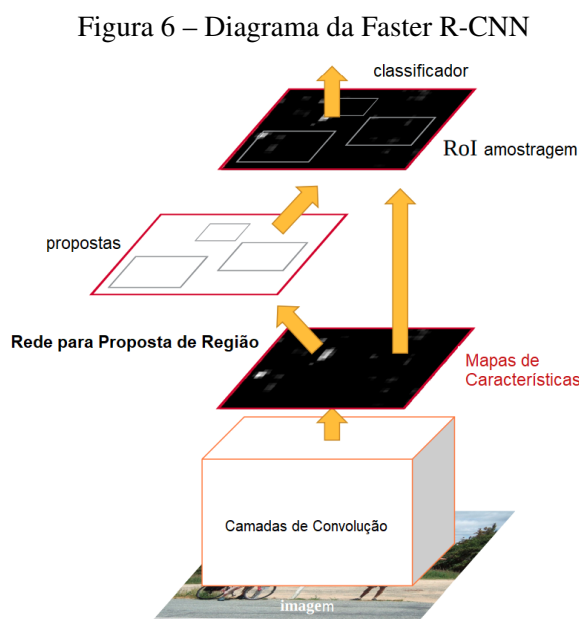
Já o segundo módulo, chamado de classificador, é formado por uma camada completamente conectada. Trata-se de uma rede neural artificial com uma ou mais camadas ocultas. Essas camadas de neurônios ocultos entre a entrada e saída da rede, torna a rede capaz de extrair estatísticas de ordem elevada. Considerando que as saídas das convoluções apresentam muitos sinais, a utilização de camadas ocultas é de grande valor. Os neurônios de cada camada oculta tem como entrada apenas os sinais de saída da camada que os precede, assim sucessivamente, até que a última camada oculta seja interligada com a camada de saída. A resposta na saída da última camada consiste da resposta global da rede, que é consequência da ativação dos nós da camada de entrada da rede totalmente conectada. Portanto, é no módulo classificador que ocorre a classificação das imagens. Os mapas de características gerados na saída do módulo extrator são analisados pelo classificador, e quando é encontrada uma provável classe, ela é associada ao mapa analisado. Sendo que, as classes são previamente definidas durante o processo de aprendizagem (HAYKIN, 2003).

Neste estudo, as classes definidas representam os danos causados pela FAW na folha do

milho. Serão criadas classes que se diferenciam pela gravidade do dano causado. Ou seja, o módulo extrator de características fará a busca por danos nas folhas, e posteriormente, cada dano será classificado. O objetivo da atual subseção foi de contextualizar as redes neurais convolucionais. Na próxima subseção serão apresentadas as principais características e funções do modelo *Faster R-CNN*, o qual possui uma rede neural convolucional como componente presente na sua estrutura.

2.1.5 Modelo Faster R-CNN

O modelo *Faster R-CNN* foi desenvolvido com a proposta de ser mais rápido que o modelo anterior *Fast R-CNN* proposto pelo autor Girshick em 2015. Esses modelos tem como principal característica classificar e localizar elementos de interesse dentro de imagens, ou seja, cada elemento alvo encontrado tem sua região delimitada na saída do modelo. Para isso, esse modelo é composto por dois principais módulos (REN et al., 2015).



Fonte: Adaptado de Ren et al. (2015)

O primeiro módulo consiste no módulo extrator de características (REN et al., 2015), representado pelo bloco *conv layers* da Figura 6. Como mencionado na subseção anterior, esse módulo é responsável por varrer imagens e extrair características. Esses módulos podem ser os mais diversos, como por exemplo, *ResNet*, *VGG* e *InceptionV3*. Portanto, na saída deste módulo é esperado um conjunto de mapas de características.

O segundo módulo é chamado de *Region Proposal Network* (RPN) ou redes para proposta de região, também presente na Figura 6. Sua função é de avaliar os mapas de características gerados pelo módulo extrator. Na avaliação os mapas são varridos por âncoras de tamanho definidos. Cada região do mapa que corresponde a um conjunto de âncoras, é submetida aos

métodos de classificação e regressão. A classificação deve indicar se aquele conjunto de âncoras é um objeto de interesse ou apenas fundo. Quando o conjunto de âncoras for um objeto de interesse, a regressão deve ajustar o posicionamento desse conjunto para que ele se encaixe da melhor forma no contorno do objeto. Após todos os conjuntos de âncoras serem classificadas, a técnica de supressão de não máximos é aplicada para filtrar os conjuntos com maiores probabilidades de serem regiões de interesse. Portanto, é neste módulo que a arquitetura *Faster R-CNN* identifica os elementos alvos das imagens, porém, ainda sem classificá-los (REN et al., 2015).

Obtidos os conjuntos de âncoras candidatos a objetos, é necessário classificá-los. Para isso, um novo módulo da arquitetura é consumido. Trata-se da *Region of Interest Pooling* (ROI). Através de técnicas de amostragem, esse módulo faz o redimensionamento dos mapas de características das âncoras, ajustando-os para que tenham a mesma altura e largura. Em seguida, a camada *Flatten* recebe os mapas e faz um processo de "achatamento", ou seja, os mapas de característica de dimensão N são transformados em um vetor de uma dimensão. Por fim, cada vetor unidimensional é transmitido para camadas completamente conectadas. Assim como na *Region Proposal Network*, nessa etapa os métodos de classificação e regressão também são utilizados, porém, agora para classificar os conjuntos de âncoras dentro das classes previamente determinadas e não apenas entre objeto e fundo. Depois da classificação, o método de regressão determinará a posição e tamanho do contorno da classe dentro da imagem (REN et al., 2015).

O modelo *Faster R-CNN* surge como uma ferramenta robusta e rápida para soluções que buscam não apenas classificar imagens em classes, mas além disso, buscam identificar em cada imagem as classes encontradas. O objetivo da seção foi explorar os conceitos básicos para o entendimento deste estudo. Na próxima seção serão citadas publicações que utilizaram redes neurais convolucionais como ferramentas para auxiliar em problemas de identificação de danos, pestes ou doenças na agricultura.

2.2 Estado da Arte

Devido a evolução das redes neurais convolucionais nos últimos anos, sua aplicabilidade tornou-se ampla e capaz de auxiliar em tarefas cada vez mais específicas. Nesta seção foram selecionados estudos que empregam RNC para identificação de pestes ou doenças em plantações. Nestes estudos foram analisadas as técnicas utilizadas para o tratamento das imagens, os modelos RNC implementados e os resultados finais de cada configuração. A ordem das publicações nesta seção leva em consideração sua abordagem, iniciando pelos estudos mais específicos sobre a identificação do ataque da FAW no milho, seguido de estudos generalizados na identificação de outras pestes e doenças em culturas.

No estudo de Du et al. (2022), os autores afirmam que a detecção rápida e precisa de pestes na agricultura aumenta significativamente a eficiência no seu controle, consequentemente, reduzindo os custos de produção. A lagarta militar é uma peste presente em diversas partes do mundo e gera graves danos em culturas como a do milho e do arroz. Dentro do seu ciclo de

vida, a lagarta vai se desenvolvendo e seu dano também vai se tornando mais severo, ou seja, de pequenos buracos até a destruição completa da folha. Portanto, para detectar o ataque da lagarta militar, os autores desenvolveram um novo modelo chamado *Pest Region-CNN*, baseado no modelo existente *Faster Region-CNN*. Esse novo modelo tem o objetivo de identificar os danos na folha do milho e classificá-lo em quatro classes: juvenil, menor, moderado e severo. O dano juvenil é característico por não causar perfuração na folha, apenas tornando a região atacada com o aspecto de transparência. O dano menor apresenta uma pequena quantidade de buracos na folha, sendo uma área menor que 10% de uma unidade de área. O dano moderado ocorre quando vários buracos são presentes, representando de 10% a 30% de uma unidade de área. Já o dano severo é o ataque que consome parte da folha, sendo equivalente a mais de 30% de uma unidade de área. O estudo utiliza *ortho-images* com resolução de 3000x4000 *pixels*, capturadas por veículo aéreo não-tripulado (VANT). Inicialmente cada imagem do *dataset* de treinamento passou pelo processo de identificação das âncoras. Esse processo consiste em identificar com retângulos e nomear de acordo com as quatro classes, as diferentes regiões da imagem onde existe algum tipo de dano na folha. Portanto, os quatro tipos de danos classificados, sempre que presentes, foram identificadas nas imagens. Quando identificado de forma isolada, apenas aquele buraco ou raspagem recebe o retângulo nas suas bordas. Entretanto, quando é presente mais de um dano da mesma classe, todos recebem a cobertura do mesmo retângulo. Para enriquecer o *dataset* foram utilizadas técnicas de *data augmentation*. Por fim, sua divisão foi na proporção de 8:2 para treinamento e validação. A principal mudança proposta pelos autores sobre a *Faster Region-CNN*, foi no módulo de extração de características. Devido os danos nas folhas corresponderem a uma proporção muito pequena da área total das imagens, após várias convoluções, os danos desaparecem. Isso ocorre principalmente para as classes juvenil e menor. Deste modo, para contornar esse problema, foi implementado nesse módulo o algoritmo *ResNet-50*. Diferentemente do módulo original da *Faster Region-CNN*, que utiliza apenas a última camada de cada um dos 512 mapas de características gerados, o *ResNet-50* nesse caso, faz o cascadeamento do resultado de cada convolução de forma a preservar contextos de alto e baixo nível com o mínimo de esforço computacional necessário. Outra modificação nesse módulo foi da função de ativação, onde a função *Relu* foi substituída pela função *Mish*. Já no módulo formado pelas regiões propostas, também houveram alterações recorrente à necessidade de identificar as pequenas áreas de danos, como a alteração das escalas das âncoras. Por fim, o *Pest Region-CNN* superou em precisão o *Faster R-CNN* e o *YOLOv5*, em respectivamente, 12% e 19%.

Para Ishengoma, Rai e Said (2021), a utilização de imagens para detecção automática do estado das folhas em plantações de milho, eleva sua produção tornando-a mais lucrativa devido a redução de custos e tempos operacionais. Portanto, nesse estudo os autores propuseram a utilização de inteligência artificial para identificar ataques da lagarta militar em plantações de milho. Para isso, foram utilizados quatro diferentes modelos de redes neurais convolucionais chamadas *VGG16*, *VGG19*, *InceptionV3* e *MobileNetV2*. Esses modelos foram treinados

para classificar, através de imagens capturadas por VANT (Veículo Aéreo Não-Tripulado), a existência de dano na folha da planta. A acurácia dos modelos foi de respectivamente 96%, 93,08%, 96,75% e 98,25%. Entretanto, com o objetivo de obter melhores resultados, os autores buscaram outra alternativa para treinar os modelos com as mesmas imagens, dessa vez, utilizando recortes obtidos pela técnica *Shi-Tomas* de detecção de cantos. Essa técnica, que está disponível na biblioteca OpenCV (*Open Source Computer Vision*), foi aplicada para gerar recortes nas dimensões de 150x150 *pixels*. As novas imagens geradas foram divididas em três grupos: infectado, não infectado e *background*. Porém, os modelos foram apenas treinados com imagens do grupo infectado e não infectado, dividindo o *dataset* na proporção de 70% para treinamento, 15% validação e 15% para testes. Esse novo treinamento possibilitou os modelos *VGG16*, *VGG19*, *InceptionV3* e *MobileNetV2* a elevarem suas acurácias para, respectivamente, 99,92%, 99,67%, 100% e 100%.

Segundo os autores Feng et al. (2022), a utilização de imagens de plantações obtidas por VANT (Veículo Aéreo Não-Tripulado), juntamente com técnicas de *deep learning*, tornam a agricultura mais precisa e ágil na tomada de decisões. Isso ocorre a partir da substituição do trabalho humano em fazer avaliações visuais, que são lentas e subjetivas, por modelos de inteligência artificial específicos para essas funções. Sendo assim, os autores treinaram modelos para detectar uma das principais pestes das plantações do milho, a lagarta militar. A lagarta tem como característica apresentar dois tipos de danos na folha. O primeiro, no início do seu estado larval, considerado como raspagem da folha, é característico por deixar vários pequenos pontos transparentes. Já o segundo, é causado no estado larval mais avançado e se caracteriza por cortes nas folhas ou até mesmo sua destruição completa. Nesse estudo, os autores fizeram o uso de dois *datasets*, onde cada um contém apenas um tipo de estado de desenvolvimento da planta do milho. Eles são chamados de estado vegetativo e reprodutivo. Portanto, o objetivo dos modelos para cada um dos *datasets*, foi classificar as seguintes classes: folha saudável, folha raspada, folha cortada e imagens irrelevantes. Os algoritmos utilizados foram *ResNeSt50*, *ResNet50*, *EfficientNet* e *RegNet*. Os modelos foram treinados com imagens da classe saudável, folha raspada e folha cortada, dividindo os *datasets* nas proporções de 90% para treinamento e 10% para validação. Os algoritmos *ResNeSt50*, *ResNet50*, *EfficientNet* e *RegNet*, respectivamente, obtiveram acurácia no estado vegetativo de 98,77%, 97,59%, 97,89%, 98,07% e de 89,39%, 81,88%, 86,21% e 84,21% no estado reprodutivo.

Os autores Ishengoma, Rai e Ngoga (2022), afirmam em seu estudo que a detecção visual de doenças em grandes plantações consome tempo e também são propensas a erros devida a subjetividade humana em avaliar. Mesmo com a existência de técnicas para a identificação automática de doenças em plantações, essas técnicas ainda não permitem a resposta imediata da detecção. Portanto, os autores propuseram um modelo híbrido de redes neurais convolucionais para acelerar a detecção do ataque da lagarta militar em plantações de milho. O modelo híbrido é baseado numa estrutura paralela, desenvolvida para extrair ao máximo os benefícios de dois modelos de redes neurais convolucionais chamados *VGG16* e *InceptionV3*. Sua estrutura possui

duas camadas de entrada que trabalham com imagens de 150x150x3. Após, as duas saídas dos modelos são concatenadas e ligadas na camada totalmente conectada de tamanho 128, para que finalmente seja classificada em quatro classes utilizando a função de ativação *Softmax*. A utilização de uma estrutura paralela tende a classificar mais rápido que estruturas em séries. Além disso, essa estrutura permite que cada modelo individual use uma quantidade menor de pesos que são responsáveis para extrair características. Isso ocorre devido a concatenação dessas características para a camada completamente conectada, no que reduz o tempo de treinamento. O modelo tem como saída as seguintes classes: saudável, infectado, ervas daninhas e irrelevante. O *dataset* foi dividido na proporção de 70:15:15, sendo respectivamente, para treinamento, validação e teste. Para o teste, o VANT (Veículo Aéreo Não-Tripulado) fez a coleta de 50 imagens de uma plantação. Após, as imagens foram enviadas para o servidor em *cloud*. O servidor realizou o tratamento que resultou em 10.800 imagens de 150x150 *pixels*. As imagens foram divididas em 60 lotes para serem carregadas no modelo. Por fim, a execução levou 29,2 minutos e o resultado da classificação foi enviado por mensagem de texto (*SMS*) e *email* para o operador. Em termos de acurácia e tempo de treinamento, os resultados finais obtidos foram comparados com os dos modelos *VGG16*, *InceptionV3*, *XceptionNet* e *Resnet50*. Sendo que, o modelo de estrutura paralela reduziu entre 16% a 44% o tempo de treinamento se comparado aos modelos treinados de forma individual, chegando a atingir uma acurácia máxima de 96,98%.

Segundo os autores Li et al. (2020), as doenças e pestes do milho são um grande problema no mundo inteiro e causam perdas significantes nas produções. Os tradicionais métodos de reconhecimento de pestes são limitados, ineficientes e consomem tempo. Esse estudo tem o objetivo de identificar de forma precisa dez tipos de pestes do milho. Para isso, alguns modelos existentes de redes neurais convolucionais foram testados e validados. O *dataset* foi criado a partir de imagens obtidas por buscadores como Google, Baidu, Yahoo, e Bing e também por fotos tiradas em plantações. Técnicas de *data augmentation* como rotações, translações e espelhamento foram utilizadas resultando num total de 14.475 imagens que foram divididas na proporção de 9:1 para treinamento e validação. Com a proposta de remover das imagens fundos complexos e manter apenas as pestes que são as áreas de interesse, foram aplicadas técnicas de *thresholding*, *watershed* e detecção de cantos. Os melhores resultados obtidos pelos modelos *GoogLeNet*, *VGG-19*, *ResNet152*, *VGG-19* e *ResNet50*, foram, respectivamente, 94,61%, 93,65%, 93,02%, 91,94% e 91,89%. No artigo, os autores também propuseram utilizar a técnica de transferência de aprendizado no modelo *GoogLeNet* com a finalidade de comparar seu resultado com o de outro estudo que aplicou a mesma técnica no modelo *ResNet101*. Portanto, o modelo *GoogLeNet* foi treinado com o *dataset* da *ImageNet*, e posteriormente, assim como o modelo *ResNet101*, foi treinado com 400 imagens para classificar 10 classes. Por fim, os resultados foram de 96,67% de acurácia em 3.765 segundos de treinamento para o *GoogLeNet* e de 90,45% de acurácia em 4.938 segundos para o *ResNet101*.

Para Zhang et al. (2018), a detecção automática de doenças foliares em plantações é algo muito desejado pela agricultura. Com essa motivação os autores devolveram alterações espe-

cíficas em dois modelos de redes neurais convolucionais existentes, o *GoogLeNet* e o *Cifar10*. Em ambas as redes foram ajustados os hiper-parâmetros de taxa base de aprendizado e do tamanho dos lotes de imagens por época. A estrutura do modelo *GoogLeNet* apresenta três classificadores no total, sendo dois intermediários. No entanto, nesse estudo, com o objetivo de reduzir o número total de parâmetros e, consequentemente, o tempo de treinamento, os autores optaram por utilizar apenas o primeiro classificador. No modelo *Cifar10*, os autores modificaram sua estrutura adicionando entre as camadas totalmente conectadas a operação *dropout* e a função de ativação *Relu*. Com a adição do *dropout* é esperado que a rede não sofra com *overfitting* e com a função *Relu* é esperado que a rede aprenda características relativamente escassas no *dataset*. Os modelos foram treinados para identificar nove classes, sendo oito doenças foliares do milho e uma classe de folhas saudáveis. Inicialmente foram coletadas 500 imagens e utilizando técnicas de *data augmentation* como rotação em ângulos de 90, 180, 270 graus e técnicas de espelhamento, foi possível obter o total de 3.060 imagens. O *dataset* foi dividido na proporção de 80% para treinamento e 20% para validação. Por fim, as modificações dos modelos *GoogLeNet* e *Cifar10* apresentaram, respectivamente, a acurácia de 98,9% e 98,8%.

Para Li et al. (2019), pestes são as maiores causadoras de danos em plantações. Em seu estudo os autores desenvolveram uma arquitetura baseada em *deep learning* para localizar e quantificar pestes em imagens. A arquitetura utilizou o modelo de rede neural convolucional de Zeiler e Fergus (ZF), juntamente com uma *Region Proposal network* (RPN). Na rede ZF, a principal modificação foi na retirada das camadas de *pooling*, com o objetivo de preservar ao máximo os pequenos detalhes das imagens. Já na rede RPN, com o propósito de reduzir o tempo de treinamento, foi implementada a técnica de *Non-Maximum Suppression* que diminui a quantidade de regiões propostas que serão candidatas a serem um objeto de interesse. Após o treinamento, a arquitetura desenvolvida apresentou uma acurácia média de 87,4% para um tempo de treinamento de 10 minutos.

No artigo criado pelos autores Wu et al. (2019) é destacada a importância em detectar e quantificar doenças em plantações de forma automática, considerando que quando esse processo é realizado por humanos se torna lento e pode ser impreciso. Motivados a desenvolver uma solução para esse problema, os autores propuseram a criação de um modelo que utiliza redes neurais convolucionais (RNC) para analisar imagens capturadas por VANT (Veículo Aéreo Não-Tripulado). O modelo foi desenvolvido para identificar folhas infectadas pela doença popularmente chamada de ferrugem do milho. O *dataset* foi composto por 3.741 imagens de folhas doentes e 2.526 de folhas saudáveis. Sendo dividido na proporção de 70:15:15, para respectivamente, treinamento, validação e testes. O algoritmo criado teve base nos dois primeiros estágios do modelo de DeChant. No primeiro estágio as imagens contendo folhas doentes são recortadas em novas imagens de tamanho 500x500 *pixels*. Após, são redimensionadas para 224x224 *pixels* e passam pelo processo de *data augmentation*. Em seguida, essas imagens são usadas para treinar a primeira rede *ResNet-34*. No segundo estágio, uma segunda rede *ResNet-34* é utilizada. Entretanto, nessa RNC, com o objetivo de aumentar sua acurácia, é aplicada a

técnica de transferência de aprendizado com o *dataset ImageNet*. Por fim, na saída do segundo estágio são gerados em tons de cinza, mapas de calor indicando as lesões nas folhas nas imagens infectadas. A acurácia final atingido pelo modelo proposto foi de 95,1%.

Segundo os autores Fuentes et al. (2017), o maior desafio para a agricultura são as pragas e doenças nas plantações. Além de que, a detecção precisa e rápida dessas doenças e pragas permite a utilização de técnicas de tratamento apropriadas, que consequentemente, reduzem perdas econômicas. O objetivo do estudo foi de desenvolver um modelo de identificação em tempo real de pragas e doenças do tomate. Para este propósito foi utilizado *deep learning*, mais precisamente três modelos da família de redes neurais convolucionais, que são: *Faster Region-based Convolutional Neural Network (Faster R-CNN)*, *Region-based Fully Convolutional Network (R-FCN)* e *Single Shot Multibox Detector (SSD)*. Com a proposta de explorar ainda mais possibilidades, o modelo do tipo *Faster R-CNN* foi treinado com cinco diferentes componentes extratores de características (*Feature Extractor*). Foram utilizados os algoritmos *VGG-16*, *ResNet-50*, *ResNet-101*, *ResNet-152* e *ResNeXt-50*. Para os modelos *R-FCN* e *SSD* apenas o algoritmo extrator *ResNet-50* foi testado. Foi usado um *dataset* composto por 5.000 imagens da folha da planta do tomate, que foram etiquetadas com dez tipos de classes, onde nove são doenças e uma corresponde ao fundo das imagens. Técnicas de *data augmentation* como rotações, espelhamento e mudança de contraste também foram aplicadas para enriquecê-lo. Por fim, sua divisão foi de 80% para treinamento, 10% validação e 10% para testes. Na tabela 1 é possível visualizar os resultados obtidos representados pela precisão média (AP) de cada modelo para determinada classe, onde seu intervalo varia de 0 até 1.

Tabela 1 – Tabela de precisão média de cada modelo para as dez classes.

Class/Feature Extractor	Meta-Architectures						
	Faster R-CNN					R-FCN	SSD
	VGG-16	ResNet-50	ResNet-101	ResNet-152	ResNeXt-50	ResNet-50	ResNet-50
Leaf mold	0.9060	0.8827	0.803	0.8273	0.840	0.8820	0.8510
Gray mold	0.7968	0.6684	0.449	0.4499	0.620	0.7960	0.7620
Canker	0.8569	0.7580	0.660	0.7154	0.738	0.8638	0.8326
Plague	0.8762	0.7588	0.613	0.6809	0.742	0.8732	0.8409
Miner	0.8046	0.7884	0.756	0.7793	0.767	0.8812	0.7963
Low temperature	0.7824	0.6733	0.468	0.5221	0.623	0.7545	0.7892
Powdery mildew	0.6556	0.5982	0.413	0.4928	0.505	0.7950	0.8014
Whitefly	0.8301	0.8125	0.637	0.7001	0.720	0.9492	0.8402
Nutritional excess	0.8971	0.7637	0.547	0.8109	0.814	0.9290	0.8553
Background	0.9005	0.8331	0.624	0.7049	0.745	0.8644	0.8841
Total mean AP	0.8306	0.7537	0.590	0.6683	0.711	0.8598	0.8253

Fonte: Fuentes et al. (2017)

Os autores Thenmozhi e Srinivasulu Reddy (2019) afirmam que os ataques de pragas são causadores de danos e afetam diretamente na produção grãos em plantações. A classificação e identificação desses ataques em folhas é uma tarefa muito difícil para os trabalhadores do campo, pelo fato de que a aparência dos danos são muito similares na fase inicial. Portanto, os autores propuseram a utilização de redes neurais convolucionais para classificar espécies de insetos presentes nas folhas das plantações. O estudo contou com três diferentes *datasets*, dois

contendo 40 e um contendo 24 classes de insetos. Além do modelo proposto pelos autores, para fins de comparação, foram também treinadas os algoritmos já conhecidos como *AlexNet*, *Resnet-50*, *Resnet-101*, *GoogLeNet*, *VGG-16*, *VGG-19*, *VGGNet*. Nos treinamentos foram aplicadas técnicas de transferência de aprendizado para pré-treinar as camadas e também de *data augmentation* como rotações, translações e espelhamento para prevenir *overfitting*. No modelo proposto, os hiper-parâmetros como taxa de aprendizado, quantidade de épocas e tamanho dos mini-lotes foram variados para encontrar o melhor resultado. A estrutura desse modelo é composta por 6 camadas de convolução, 5 camadas de *max pooling*, uma camada completamente conectada e a função de saída sendo a *Softmax*. Dentro de todos os modelos treinados, o que apresentou a maior acurácia foi o proposto pelo estudo. Sua acurácia foi de, respectivamente, 96,75%, 97,47% e 95,97% para os *datasets* contendo 40, 24 e 40 classes de insetos.

No artigo escrito por Lu et al. (2017), os autores afirmam que a identificação e diagnóstico de doenças nas plantações de arroz é altamente desejado pelo fato de ser considerado um alimento muito importante para o mundo todo. Atualmente *deep learning* é um assunto em alta quando se trata de reconhecimento de padrões e é considerada uma área que pode contribuir efetivamente na resolução de problemas relacionados a doenças de vegetais. Motivados por isso, os autores propuseram a utilização de redes neurais convolucionais para identificar doenças no arroz. Inspirados nos modelos clássicos como *LeNet-5* e *AlexNet*, foi desenvolvido um modelo composto por camadas de convolução 5x5, 9x9, 16x16 e 32x32, camadas de *stochastic pooling* entre as convoluções e a função de ativação *Softmax* na saída. O *dataset* foi composto por 500 imagens com 10 diferentes classes de doenças. Inicialmente as imagens passaram pelo processo de redimensionamento através do *software Matlab R2012a* e depois convertidas em tons de cinza para serem normalizadas. Por fim, o modelo proposto foi treinado com diferentes técnicas de *pooling*. O melhor resultado obtido foi na configuração com as camadas de amostragem do tipo *stochastic pooling* que atingiu 95,48% de acurácia.

Os autores Zhao et al. (2021), afirmam que o diagnóstico de doenças em plantações é de grande importância para o setor da agricultura e técnicas de *deep learning* se tornaram a principal área de pesquisa para auxiliar nesses processos. Esse estudo tem como propósito construir um modelo de rede neural convolucional que integra um mecanismo capaz de se adaptar melhor no diagnóstico de variadas doenças na folha do tomate. A arquitetura é formada por dois principais módulos chamados de *Feature Extraction Network* e *Attention*. A *Feature Extraction Network* é baseada no modelo clássico *ResNet50* e foi escolhido por ser uma rede neural convolucional com menos parâmetros e cálculos, além de ter numa boa performance. No módulo *Attention* é utilizada a dependência de canais. Isso significa que após pré-treinados e calculadas suas funções de custo, os canais podem ser aprimorados ou desativados dependendo do coeficiente de cada um. Essa técnica é aplicada para melhorar o desempenho da RNC. Um teste comparativo foi realizado entre um modelo *ResNet50* simples com outro modelo *ResNet50* acompanhando do módulo *Attention*. A combinação entre o modelo *ResNet50* e o módulo *Attention* resultou numa acurácia de 96,81% com o tempo de detecção de 31,68ms. Já no modelo

simples foi de 92,56% e 33,85ms.

No artigo escrito pelos autores Too et al. (2019), é destacado que *deep learning* tem chamado atenção pelos seus sistemas rápidos e precisos para identificar e classificar imagens. O trabalho proposto pelos autores tem como objetivo avaliar o estado da arte sobre redes neurais convolucionais. Sendo assim, os modelos *VGG16*, *InceptionV4*, *ResNet* com 50, 101 e 152 camadas e o *DenseNets* com 121 camadas, foram treinados para classificar doenças em plantas. O *dataset* utilizado foi o *PlantVillage* (Hughes and Salathe, 2015) que é composto por 54.306 imagens, possuindo 26 tipos distintos de doenças e 14 classes de plantas de culturas. Buscando extrair o máximo de cada modelo, foi aplicada a técnica de ajuste fino. O conceito por trás dessa técnica consiste em transmitir aprendizado através de camadas pré-treinadas por *datasets* robustos. Nesse artigo foi utilizado o *ImageNet*, *dataset* que contém 1,2 milhões de imagens e 1.000 categorias. Portanto, após a utilização dessa técnica, as camadas passam a possuir pesos mais ajustados, tornando o modelo mais preciso na classificação. Os resultados de precisão para para os modelos *VGG 16*, *Inception V4*, *ResNet* com 50, 101 e 152 camadas e o *DenseNets*, foi de respectivamente, 98,02%, 81,92%, 99,67%, 99,66%, 99,68% e 99,76% para 30 épocas.

No estudo de Militante, Gerardo e Dionisio (2019), os autores apresentaram um sistema de detecção de doenças em folhas de plantas através da utilização de redes neurais convolucionais. O modelo é capaz de identificar 27 tipos de doenças para as folhas do tomate, uva, milho, maçã e cana-de-açúcar. O *dataset* utilizado foi retirado do repositório *Plant Village* e é composto por 35.000 imagens. Técnicas de *data augmentation* como rotações, translações e espelhamento também foram usados para aprimorá-lo. A RNC utilizada no estudo é composta por camadas de *max pooling* e com a função de ativação *ReLU* entre as convoluções. O modelo treinado apresentou uma acurácia média de 96,5%.

Para Kamilaris e Prenafeta-Boldú (2018), *deep learning* possui técnicas modernas de processamento e análise de dados de grande potencial e com resultados promissores. Essas técnicas já são utilizadas em vários setores, mas tem sido agregada na agricultura recentemente. Esse estudo tem como proposta a análise de 40 artigos que empregam *deep learning* como soluções para problemas na agricultura e na produção de alimentos. Alguns exemplos de abordagem desses artigos são: classificação e identificação de plantas de cultura, frutos, obstáculos e erva daninhas, previsões de produção e de clima e análise da qualidade do solo. Cada artigo foi examinado para entender quais modelos, *frameworks*, tipos de pré-processamento de dados, *data augmentation* e métricas que são presentes nas suas soluções. No tratamento de imagens, destaca-se que 17 artigos utilizam redes neurais convolucionais tradicionais, como *AlexNet*, *VGG16* e *Inception-ResNet*. Já outros 14 desenvolveram suas próprias RNC. Quanto aos *frameworks*, 13 implementaram *Caffe*, 2 *Tensor Flow*, 1 *deeplearning4j*. Dez artigos desenvolveram seu próprio código. Outros implementaram modificações em cima dos tradicionais, 5 em *Caffe*, 5 em *Keras/Theano*, 4 em *Keras/TensorFlow* e 1 em *Pylearn2*. Tratando de *dataset*, as duas proporções mais utilizada foram de 8:2 e 9:1 para treinamento e validação.

As tabelas 2, 3 e 4 apresentam as principais referências utilizadas no desenvolvimento deste

estudo. Nelas são listadas as informações dos autores, ano de publicação, título do artigo e sua principal proposta. A ordenação dos artigos está de acordo com a proximidade de seus objetivos propostos com o objetivo desta monografia, deste modo, primeiramente são apresentados estudos relacionados diretamente com a identificação de ataques da lagarta militar no milho, e posteriormente, estudos que buscam detectar outras pragas e doenças na agricultura.

Tabela 2 – Artigos correlatos

Autor	Ano	Título	Resumo
Du et al.	2022	A novel object detection model based on faster R-CNN for <i>Spodoptera frugiperda</i> according to feeding trace of corn leaves	Desenvolvimento de um novo modelo chamado <i>Pest Region-CNN</i> , baseado no modelo existente <i>Faster Region CNN</i> para identificar quatro classes de ataques da FAW na planta do milho.
Ishengoma et al.	2021	Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks	Comparativo dos resultados dos modelos <i>VGG16</i> , <i>VGG19</i> , <i>InceptionV3</i> e <i>MobileNetV2</i> , juntamente com a técnica de detecção de cantos <i>Shi-Tomas</i> na identificação de ataques da FAW na folha do milho.
Feng et al.	2022	Autonomous detection of <i>Spodoptera frugiperda</i> by feeding symptoms directly from UAV RGB imagery	Comparativo dos resultados dos modelos <i>ResNeSt50</i> , <i>ResNet50</i> , <i>EfficientNet</i> e <i>RegNet</i> na identificação de ataques da FAW em dois estados de desenvolvimento da planta do milho.
Ishengoma et al.	2021	Hybrid convolution neural network model for a quicker detection of infested maize plants with fall armyworms using UAV-based images	Implementação de uma arquitetura híbrida utilizando os modelos <i>VGG16</i> e <i>InceptionV3</i> , com o objetivo de diminuir o tempo de treinamento na identificação dos ataques da FAW na folha do milho.
Li et al.	2020	Crop pest recognition in natural scenes using convolutional neural networks	Esse estudo tem o objetivo de identificar de forma precisa dez tipos de pestes do milho. A identificação é realizada por redes neurais convolucionais treinadas com imagens tratadas com técnicas de <i>thresholding</i> , <i>watershed</i> e detecção de cantos.

Tabela 3 – Artigos correlatos

Autor	Ano	Título	Resumo
Zhang et al.	2018	Identification of maize leaf diseases using improved deep convolutional neural networks	Comparativo dos resultados dos modelos <i>GoogLeNet</i> e <i>Cifar10</i> que foram modificados para otimizar sua performance na detecção de oito doenças foliares do milho.
Li et al.	2019	Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline	Criação de um modelo de rede neural convolucional baseado no modelo de <i>Zeiler</i> e <i>Fergus</i> , <i>Region</i> juntamente com uma <i>Proposal network</i> , com o objetivo de identificar pestes em plantações.
Wu et al.	2019	Autonomous detection of plant disease symptoms directly from aerial imagery	Desenvolvimento de um modelo de rede neural convolucional para identificar a doença popularmente chamada de ferrugem do milho. O algoritmo criado teve base nos dois primeiros estágios do modelo de DeChant.
Fuentes et al.	2017	A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition	Comparativo dos resultados dos modelo <i>Faster Region-based Convolutional, Neural Network (Faster R-CNN)</i> , <i>Region-based Fully Convolutional Network (R-FCN)</i> e <i>Single Shot Multibox Detector (SSD)</i> na detecção de nove doenças no tomate.
Thenmozhi e Reddy	2019	Crop pest classification based on deep convolutional neural network and transfer learning	Criação de um novo modelo de rede neural convolucional para classificar espécies de insetos presentes em folhas. Os resultados foram comparados com os resultados de modelos já conhecidos como <i>AlexNet</i> , <i>Resnet-50</i> , <i>Resnet-101</i> , <i>GoogLeNet</i> , <i>VGG-16</i> , <i>VGG-19</i> , <i>VGGNet</i> .

Tabela 4 – Artigos correlatos

Autor	Ano	Título	Resumo
Lu et al.	2017	Identification of rice diseases using deep convolutional neural	Inspirados nos modelos clássicos como <i>LeNet-5</i> e <i>AlexNet</i> , foi desenvolvido um modelo com diferentes técnicas de <i>pooling</i> para identificar doenças no arroz.
Zhao et al.	2021	Tomato leaf disease diagnosis based on improved convolution neural network by attention module	Desenvolvimento de um novo modelo de rede neural convolucional para identificar doenças na folha do tomate. Para isso, foi criado uma arquitetura formada por dois principais módulos chamados de <i>Feature Extraction Network</i> e <i>Attention</i> .
Too et al.	2018	A comparative study of fine-tuning deep learning models for plant disease identification	Utilização da técnica de transferência de aprendizado nos modelos <i>VGG16</i> , <i>Inception V4</i> , <i>ResNet</i> e <i>DenseNets</i> , para classificar 26 tipos distintos de doenças em 14 classes de plantas de culturas.
Militante et al.	2019	Plant leaf detection and disease recognition using deep learning	Com o objetivo de identificar 27 tipos de doenças para as folhas do tomate, uva, milho, maçã e cana-de-açúcar, foi desenvolvido um modelo baseado em redes neurais convolucionais.
Kamilaris e Prenafeta-Boldú	2018	Deep learning in agriculture: A survey	Trata-se de um estudo que faz a análise de 40 artigos que empregam <i>deep learning</i> como soluções para problemas na agricultura e na produção de alimentos.

Fonte: Elaborado pelo autor

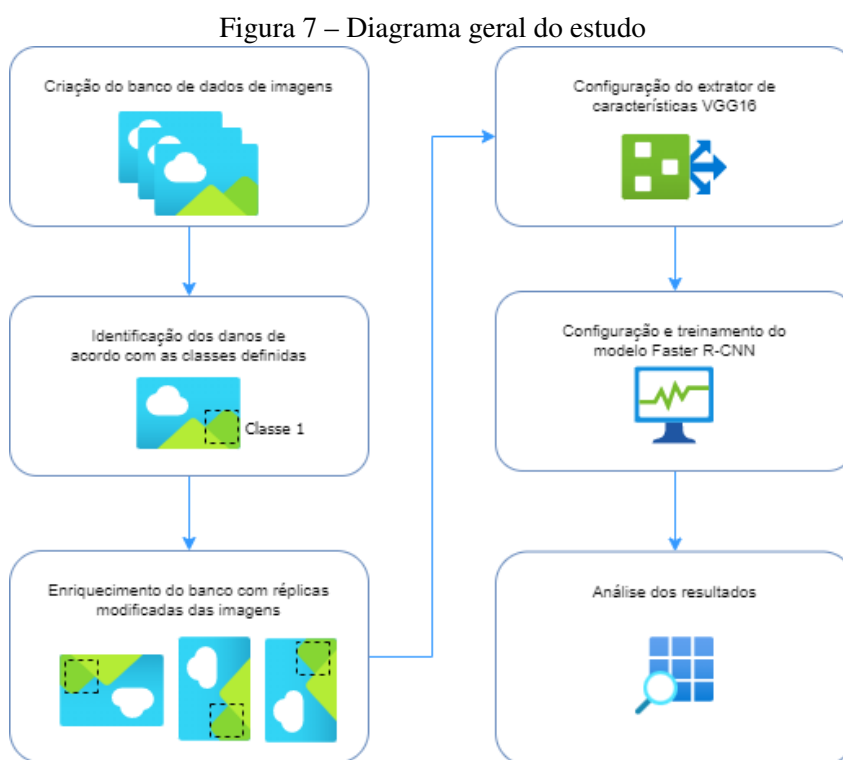
Analisando os artigos correlatos, é possível identificar que seus resultados são afetados com a mudança de alguns principais aspectos. Os principais aspectos que estão relacionados a performance dos modelos propostos são: a forma como os dados são tratados, a quantidade de imagens utilizadas no processo de aprendizagem, os modelos escolhidos e a configuração dos hiper-parâmetros de cada modelo. Os autores Du et al. (2022), Ishengoma, Rai e Said (2021), Feng et al. (2022) e Ishengoma, Rai e Ngoga (2022), utilizaram diferentes modelos de redes neurais convolucionais para identificar os ataques da FAW. O estudo de Du et al. (2022) foi capaz de classificar quatro classes de dano, utilizando o algoritmo *ResNet-50* como módulo extrator de características dentro da arquitetura *Faster R-CNN*. Sua melhor precisão foi obtida aplicando técnicas de *data augmentation* como rotações, mudanças de contraste e coloração. Nesse estudo o extrator *YOLOv5* também foi avaliado, entretanto, apresentou resultados inferiores por ser um extrator não tão eficiente com detalhes. Li et al. (2019) também obteve boa precisão através do uso de redes para proposta de região que é presente na *Faster R-CNN*, porém, combinando com os módulos extratores de características Zeiler e Fergus e *VGG16*. Li et al. (2019) por possuir um conjunto de imagens relativamente pequeno, com 850 imagens, utilizou técnicas de *data augmentation* para enriquecer o conjunto e obter um modelo mais eficiente. Os autores Ishengoma, Rai e Said (2021) e Ishengoma, Rai e Ngoga (2022), empregaram o extrator *InceptionV3* em dois tipos de configurações distintas, que atingiram níveis de precisão acima dos 90%, superando o extrator *VGG* também testado. Zhang et al. (2018) criou vários cenários modificando os hiper-parâmetros dos modelos *GoogLeNet* e *Cifar10*, como taxa de aprendizado base e tamanhos de lotes, e também fez diferentes combinações de *pollings* e *dropouts* até atingir as melhores respostas. Fuentes et al. (2017) buscou identificar doenças e pestes em imagens, através dos modelos *Faster R-CNN*, (*R-FCN*) e (*SSD*). Os três modelos performaram com uma precisão média acima de 82% quando técnicas de *data augmentation* foram empregadas. Em razão de técnicas de *data augmentation* estarem presente na maioria dos estudos e sua implementação beneficiar modelos para atinjam melhores resultados, esse estudo contará com a técnica de rotação das imagens. Esta monografia tem o objetivo de classificar e também identificar os danos da lagarta militar nas imagens, para isso, deve-se utilizar um modelo que tenha essas funções. Foi definido que modelo empregado será o *Faster R-CNN*. O modelo esteve presente em todos os artigos relacionados que propuseram delimitar elementos nas imagens, além de, apresentar bons resultados. Como módulo extrator de características para o *Faster R-CNN*, será utilizado o *VGG16*. Esse extrator também apareceu de maneira positiva nos artigos relacionados. Na próxima seção será apresentada a metodologia deste estudo, ou seja, será exposto todo o desenvolvimento realizado para a configuração completa do modelo de identificação do dano da FAW na folha do milho.

3 METODOLOGIA

O objetivo desta seção é descrever os métodos utilizados para criação do *dataset* e configuração do modelo de identificação. São descritas as abordagens e técnicas aplicadas em cada uma das etapas do estudo. Para isso, a seção foi organizada em duas partes. Na primeira subseção é apresentado o diagrama geral do estudo. Seu objetivo é expor através de um diagrama, a divisão das etapas. Na segunda subseção, cada etapa é detalhadamente descrita, ou seja, todo sistema é apresentado, assim como, sua métrica de avaliação.

3.1 Visão Geral

Este trabalho foi dividido em seis principais etapas. Cada etapa corresponde a um bloco do diagrama da Figura 7.



Fonte: Elaborado pelo autor

Representada pelo primeiro bloco do diagrama, a primeira etapa é a busca por imagens de ataques da lagarta militar na folha do milho, em repositórios *online*. Também chamado de *dataset*, o banco de imagens deve possuir imagens de folhas saudáveis e com diferentes formas de dano, além de, possuir a maior quantidade possível, de forma a proporcionar um bom aprendizado para a rede neural convolucional. Na segunda etapa, representada pelo segundo bloco do diagrama, todas as imagens passam pelo processo de anotação. Esse processo consiste na identificação dos danos na folha e também na sua classificação. Sendo assim, dentro das classes

definidas, as imagens tem a posição dos danos gravados e classificados. As anotações são necessárias para que o modelo possa utilizá-las como referência no seu processo de aprendizado e de validação. Na terceira etapa é feito o enriquecimento do *dataset*. A técnica de rotação é aplicada para gerar imagens que são variações das originais. Na quarta etapa, é realizada a configuração do extrator de características *VGG16*. Na quinta etapa, os hiper-parâmetros do modelo *Faster R-CNN* são ajustados a fim de obter os melhores resultados. Após, é feito seu treinamento. Nesse momento também são definidas as proporções do *dataset* que serão utilizadas nos processos de aprendizado e validação. Por fim, na sexta e última etapa, é definida a métrica a ser utilizada para avaliar a precisão final do modelo. Nesta subseção foram expostas as principais etapas do estudo. Na próxima subseção, cada uma destas etapas será detalhadamente abordada.

3.2 Configuração proposta

As próximas subseções foram organizadas de forma a dividir cada etapa do desenvolvimento do estudo. Cada subseção será explorada de acordo com a ordem do diagrama da Figura 7.

3.2.1 Banco de Dados de Imagens

As imagens utilizadas nesse estudo foram retiradas de um repositório público do site *kaggle.com* e de autoria de Acharya (2020). O repositório é composto por imagens capturadas via *smartphone* de uma plantação de milho localizada no Nepal. As imagens foram registradas durante os meses de junho e julho de 2020, após a identificação de ataques da lagarta militar. A coleção é composta no total de 4223 imagens e apresenta diversos estados de integridade das folhas. Nas imagens que apresentam dano, segundo a escala de Davis, é possível identificar diferentes níveis de ataque. É importante que o *dataset* seja composto por todas essas variações, para que no processo de aprendizado o modelo tenha imagens suficientes para entender como diferenciar as classes de dano.

3.2.2 Anotação das Imagens

Apenas obter um *dataset* composto por diferentes classes de danos não é suficiente para que o modelo *Faster R-CNN* possa aprender. É necessário indicar em cada uma das imagens onde está o dano e qual sua classe. Neste estudo, foram definidas quatro classes utilizando a escala de Davis como referência. A primeira classe é chamada de saudável, sendo correspondente ao nível 0. Ela não apresenta nenhum tipo de ataque. A segunda classe é chamada de dano mínimo e equivale aos níveis 1 e 2 da escala. A terceira classe é chamada de dano moderado e equivale aos níveis 3, 4 e 5. Já a classe chamada de dano grave, corresponde aos níveis 6, 7, 8 e 9. O critério de criação das classes buscou agrupar os níveis de Davis de acordo

com suas principais características. O objetivo é diminuir a complexidade em diferenciar os níveis vizinhos que apresentam diferenças sucintas. A classe de dano mínimo se caracteriza por possuir apenas buracos e pequenas lesões. Na classe de dano moderado surgem buracos maiores e lesões alongadas. Na classe de dano grave as lesões são grandes e a planta pode apresentar desfoliação.

Antes de iniciar o processo de anotação, foi verificado que no *dataset* haviam imagens com diferentes resoluções. Para padronizar, todas foram redimensionadas para possuírem a resolução máxima de 500 *pixels* em cada dimensão, porém, mantendo suas proporções. Esse processo foi feito a partir de codificação com a linguagem de programação *python* e o auxílio da biblioteca *OpenCV*.

Figura 8 – Ferramenta labelImg



Fonte: Tela da aplicação labelImg

Para o processo de anotação foi utilizada a ferramenta *labelImg*. Trata-se de uma ferramenta muito popular e de fácil utilização. Na Figura 8 é possível entender como funciona o processo de anotação. As classes saudável, dano mínimo, moderado e grave foram anotadas a partir de retângulos que cobrem suas regiões. No final desse processo, a ferramenta gera um arquivo com as informações de localização de cada classe e seu tipo. Em seguida, esse arquivo será utilizado para o treinamento do modelo *Faster R-CNN*.

3.2.3 Enriquecimento do Dataset

Como verificado nos trabalhos correlatos, a utilização de técnicas de enriquecimento de *dataset*, ou também chamadas de técnicas de *data augmentation*, ajudam no resultado final de modelos baseados em redes neurais convolucionais. Essas técnicas consistem em gerar modificações das imagens do *dataset*, a fim de possibilitar o modelo a aprender com ainda mais

imagens. Nesse estudo, a técnica de *data augmentation* utilizada foi a de rotação.

Figura 9 – Exemplo do resultado da técnica de rotação

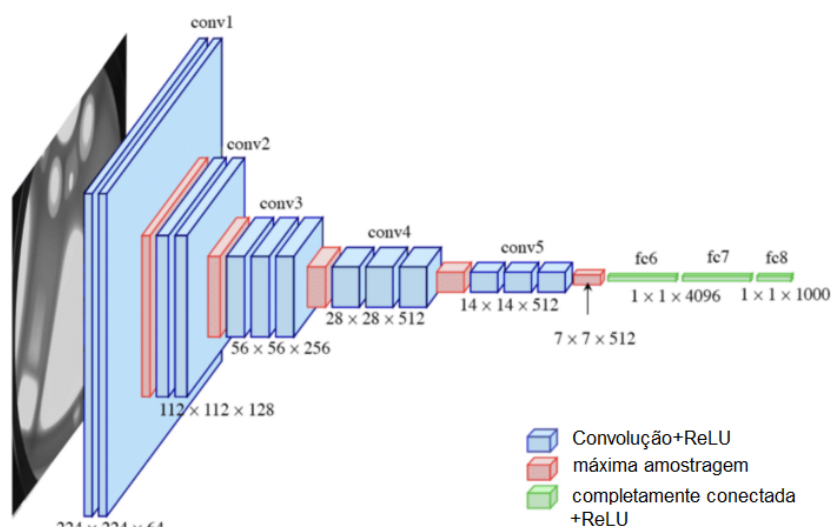


Fonte: Elaborado pelo autor

A Figura 9 exibe uma variação originada pela técnica de rotação, gerada a partir de uma imagem. Na próxima seção será exposta a configuração do extrator de características *VGG16*.

3.2.4 VGG16

O extrator de características *VGG16* é um modelo baseado em redes neurais convolucionais que atingiu 92,70% de precisão no conjunto de dados do *ImageNet*. O conjunto de dados *ImageNet* é composto por 14 milhões de imagens e cerca de 1000 classes de objetos. Conforme a Figura 10, este modelo é composto por camadas de convoluções, *poolings* máximos, concatenações e camadas completamente conectadas. Neste estudo, o *VGG16* foi escolhido como o extrator de características responsável por gerar os mapas de atributos que serão utilizados pelo *Faster R-CNN*.

Figura 10 – Estrutura do extrator *VGG16*

Fonte: Adaptado de Sugata e Yang (2017)

O modelo *VGG16* é presente nas principais bibliotecas de aprendizado de máquina como *PyTorch* e *TensorFlow*. Essas bibliotecas são escritas na linguagem *python* e permitem a implementação de modelos de forma mais simples, com menos linhas de código e algumas configurações. Na próxima subseção será apresentada a configuração do modelo *Faster R-CNN*, o qual utiliza os mapas de atributos gerados pelo *VGG16* para identificar e classificar os elementos de interesse.

3.2.5 Faster R-CNN

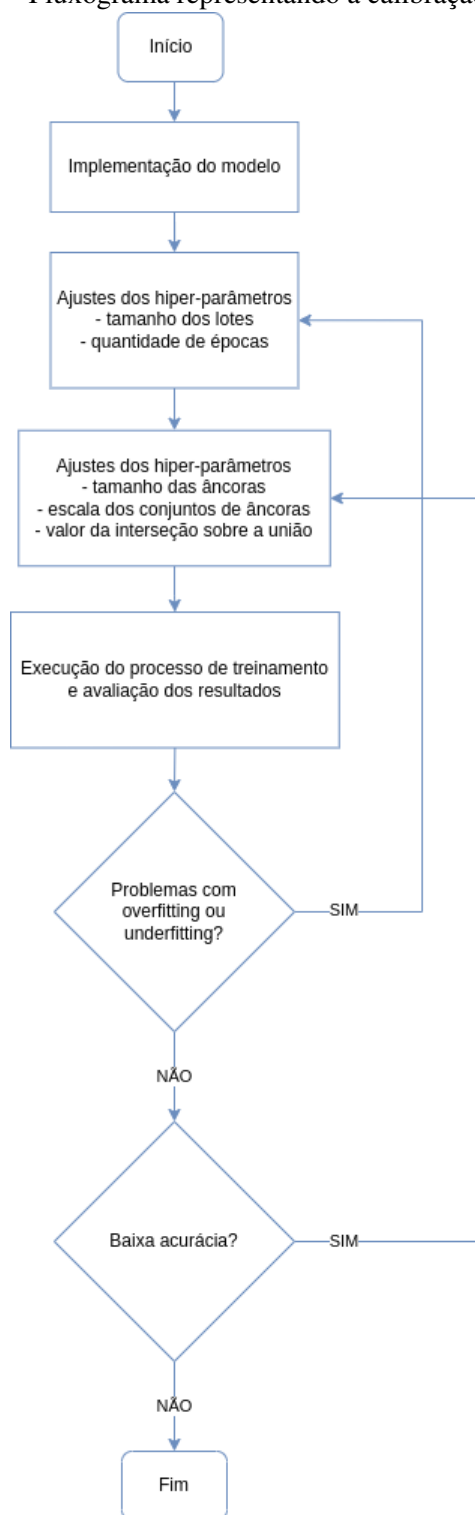
O modelo *Faster R-CNN* é o componente responsável para que as imagens sejam analisadas e tenham os objetos de interesse classificados. Este estudo utilizou a implementação do modelo *Faster R-CNN* escrita por Trzynadlowski (2021) e disponibilizada em seu repositório no *github*. A implementação utiliza tem como base a biblioteca *PyTorch*. Como já mencionado, os dados de entrada desse modelo devem ser mapas de atributos que são processados pelo extrator de características *VGG16*.

Primeiramente, o *dataset* deve ser organizado. Para isso, foi dividido em treinamento e teste, na proporção de 50% para cada. Cada divisão terá um arquivo no formato Pascal VOC XML, contendo as informações geradas na etapa de anotação das imagens. Após, as classes devem ser informadas como parâmetros ao modelo, neste caso, são quatro: saudável, dano mínimo, moderado e grave.

Na configuração do *Faster R-CNN*, é necessário definir alguns hiper-parâmetros que estarão diretamente relacionados a convergência do modelo. Portanto, de acordo com o fluxograma da Figura 11, alguns hiper-parâmetros podem ser ajustados caso o modelo não consiga aprender. Os principais hiper-parâmetros são:

- Tamanho dos conjuntos de âncoras - O mapa de atributos é preenchido por pequenas âncoras. O conjunto de âncoras está relacionado com o tamanho da região analisada a cada passo. Essa região é avaliada a fim de identificar se pode ou não ser uma possível região de interesse. No modelo original o tamanho de cada conjunto de âncoras é de 8, 16 e 32.
- Escala dos conjuntos de âncoras - Cada conjunto de âncoras pode ter diversas escalas, no modelo original as escalas são de 1:2, 1:1 e 2:1.
- Valor da interseção sobre a união - Representa a proporção da área que o conjunto de âncoras deve ter sobre de uma região anotada (região que representa um objeto de interesse) para ser considerado objeto e não fundo. No artigo original, para ser considerado um conjunto de âncoras válido, o conjunto deve ter mais de 50% da sua área sobre uma região anotada.
- Tamanho dos lotes - Em redes neurais convolucionais, o tamanho dos lotes depende do poder computacional da máquina que treinará o modelo.
- Quantidade de épocas - Deve ser ajustado a fim de obter a convergência da função de erro. Caso esse valor não esteja adequado o modelo sofrerá com um dos seguintes problemas. O *underfitting* é causado pelo baixo aprendizado do modelo em relação ao *dataset* durante o processo de treinamento, ou seja, sua precisão final será baixa. Já o *overfitting* fará com que o modelo seja apenas capaz de prever os dados do *dataset*, sendo pouco assertivo para novas imagens.

Figura 11 – Fluxograma representando a calibração do modelo



Fonte: Elaborado pelo autor

Na atual subseção foi apresentado o método de calibração do modelo, o qual necessita de uma métrica de precisão para avaliar sua performance. Na próxima subseção será apresentada a forma de como será extraída essa métrica de precisão.

3.2.6 Métrica de Avaliação

Para analisar a precisão final do modelo será utilizada a métrica *Mean Average Precision* (mAP), muito popular em modelos detectores de objetos como o *Faster R-CNN*. Entretanto, primeiramente deve-se encontrar a *Average Precision* (AP). A AP resulta em valores dentro do intervalo de 0 a 1, sendo que, quanto mais próximo de 1, maior a precisão do modelo. Inicialmente, para encontrar a AP deve-se utilizar as equações de *precision* (3.1) e *recall* (3.2) para cada imagem do *dataset* e classe do modelo. Para isso, deve ser considerados os seguintes atributos:

- Verdadeiro Positivo (VP): objeto existente e que foi detectado pelo modelo.
- Verdadeiro Negativo (VN): objeto inexistente e que o modelo não detectou.
- Falso Positivo (FP): objeto inexistente mas o modelo considerou como objeto detectado.
- Falso Negativo (FN): objeto existente e que não foi detectado pelo modelo.

$$precision = \frac{VP}{VP + FP} \quad (3.1)$$

$$recall = \frac{VP}{VP + FN} \quad (3.2)$$

A equação *precision* (3.1) mede a porcentagem de acerto que o modelo obteve em relação aos objetos preditos. Já a equação *recall* (3.2) mede o desempenho do modelo em relação a predição de todos os objetos existentes.

Calculados os valores, forma-se o gráfico *precision x recall*. Esse gráfico gera uma curva onde a área abaixo dela representa a *Average Precision*. Para resumir a forma da curva AP, é calculada a interpolação em todos os pontos conforme a equação 3.3, onde r é o valor de *recall* e $\rho_{interp}(r_{n+1})$ é o valor máximo de *precision* nos diferentes intervalos de *recall*.

$$AP = \sum_{r=0}^1 (r_{n+1} - r_n) \rho_{interp}(r_{n+1}) \quad (3.3)$$

Por fim, possuindo o AP de cada classe é possível obter o *Mean Average Precision* do modelo, apenas fazendo a média das AP de todas as classes.

Nesta seção foi verificada a metodologia que será empregada neste estudo. Na próxima seção será apresentada a análise dos resultados.

4 ANÁLISES DE RESULTADOS

A partir de cenários criados para medir o desempenho do modelo, neste capítulo será desenvolvida a análise dos resultados obtidos e também descrita a característica de cada cenário construído. Por fim, é realizada a análise do modelo como um todo. Foram criados três principais cenários. Estes cenários se caracterizam por tratarem de forma diferente cada classe do *dataset*. Porém, todos os cenários foram configurados com os mesmos valores de hiper-parâmetros, seguindo a recomendação da documentação da implementação do modelo *Faster R-CNN*. Os hiper-parâmetros estão apresentados na tabela 5. Destaca-se o hiper-parâmetro taxa de aprendizado que inicia com o valor 0,001 e permanece o mesmo até a época 25, após, é alterado para 0,0001. Outra característica em comum nos três cenários é a proporção de âncoras utilizadas para treinamento e teste que é de 50% para cada.

Tabela 5 – Hiper-parâmetros dos cenário 1, 2 e 3

Hiper-parâmetro	Valor
Épocas (<i>Epoch</i>)	50
Taxa de Aprendizado (<i>Learning rate</i>)	0,001 e 0,0001
Impulso (<i>Momentum</i>)	0,9
Dcaimento dos pesos (<i>Weight decay</i>)	0,0005
Tamanho dos lotes (<i>Batch size</i>)	1
Tamanho das âncoras	128x128, 256x256 e 512x512
Escala do conjunto de âncoras	0,5:1, 1:1 e 2:1
Valor da interseção sobre a união	> 50%

Fonte: Elaborado pelo autor

De forma complementar, um quarto cenário também foi desenvolvido. Entretanto, sem o objetivo de identificar na imagem a localização e a classificação do dano da FAW. Sendo assim, apenas a rede neural convolucional *VGG16* é utilizada, descartando o modelo *Faster R-CNN*. Trata-se de um cenário que apenas classifica a imagem de forma binária, ou seja, uma imagem que contenha qualquer tipo de dano é classificada como dano, ao contrário, caso não haja dano, a imagem é classificada como saudável. Nas próximas seções, cada cenário é abordado de forma específica.

4.1 Cenário 1

Este primeiro cenário foi construído de acordo com a principal proposta deste estudo, ou seja, classificar em três diferentes escalas o dano na folha do milho. Portanto, o modelo foi treinado com as seguintes classes de dano: mínimo, moderado e grave. Além disso, com o objetivo de beneficiar o modelo a aumentar sua precisão, foi adicionada a classe saudável no *dataset*. Essa classe permite ao modelo diferenciar a classe de dano mínimo com a classe saudável, já que são classes que possuem pequenas diferenças.

A tabela 6 apresenta a quantidade total de âncoras por classe.

Tabela 6 – Dataset do cenário 1

Classe	Âncoras
saudável	701
mínimo	105
moderado	160
grave	1671

Fonte: Elaborado pelo autor

Após realizado o treinamento, foi possível obter a métrica mAP de 43,41% sobre o *dataset* de teste. Onde cada classe resultou na AP de acordo com a tabela 7.

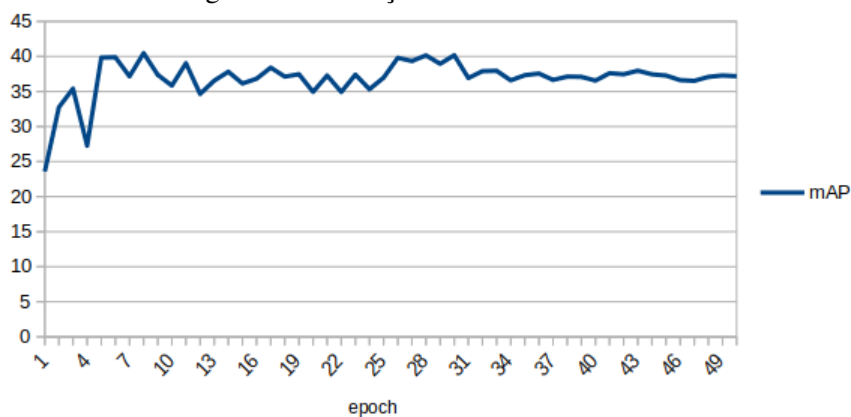
Tabela 7 – AP do cenário 1

Classe	AP
saudável	46,60%
mínimo	30,90%
moderado	23,50%
grave	72,60%

Fonte: Elaborado pelo autor

A variação da mAP em cada época, durante o treinamento, é representada pela Figura 12.

Figura 12 – Variação da mAP no cenário 1

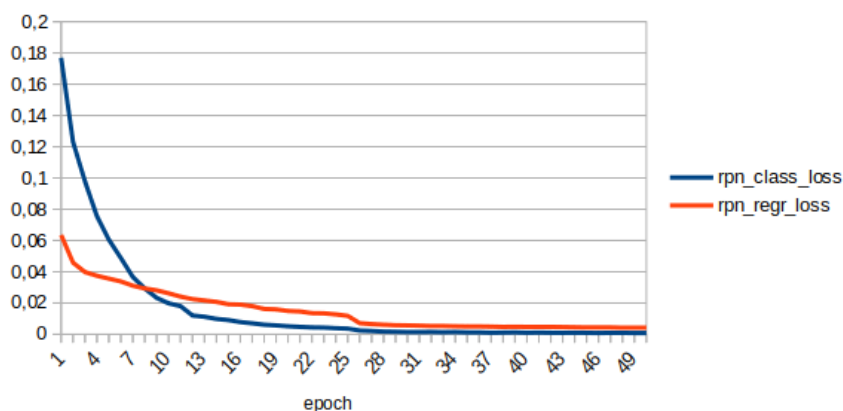


Fonte: Elaborado pelo autor

Como já mencionado anteriormente, o modelo *Faster R-CNN* é composto por dois componentes principais, o componente *Region Proposal Network* (RPN), responsável por identificar nas imagens possíveis âncoras que possam ser classes de interesse e contorná-las, e o componente *Region of Interest Pooling* (ROI), responsável por classificar a classe encontrada e determinar seu tamanho e posição. Portanto, o desempenho de ambos são importantes para a performance final. Sendo assim, é interessante entender a função de perda associado a cada um, pois, deseja-se valores próximos a zero.

Na Figura 13 é exposto o gráfico do resultado da função de perda do componente RPN. A curva azul representa a perda em relação a classificação da âncora em ser um objeto de interesse ou fundo. Já a curva laranja representa a perda da regressão que tem o propósito de contornar da melhor forma a âncora.

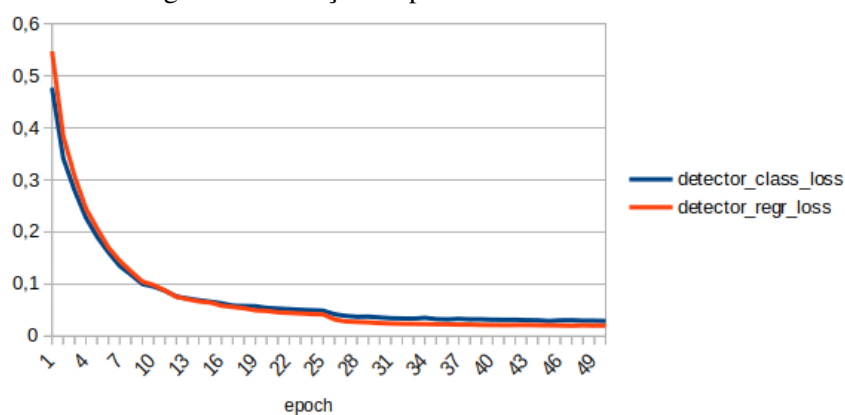
Figura 13 – Função de perda do RPN no cenário 1



Fonte: Elaborado pelo autor

Na Figura 14 é exposto o gráfico do resultado da função de perda do componente ROI. A curva azul também representa a perda em relação a classificação do conjunto de âncoras, porém, o ROI busca classificá-lo entre saudável, mínimo, moderado e grave. Já a curva laranja representa a perda da regressão ao procurar o contorno da classe.

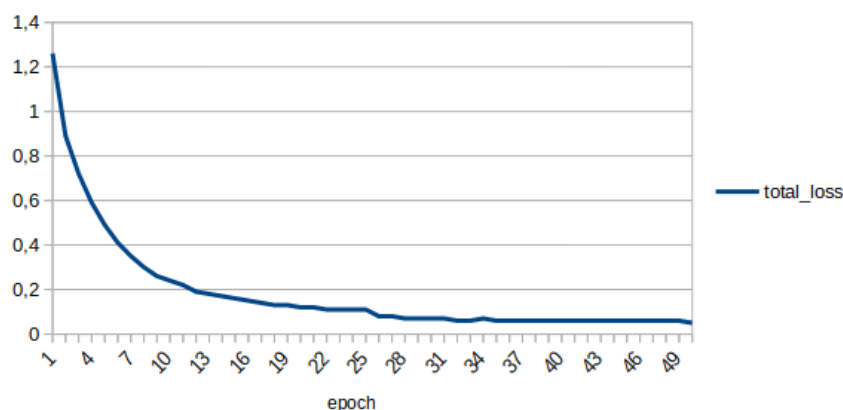
Figura 14 – Função de perda do ROI no cenário 1



Fonte: Elaborado pelo autor

Na Figura 15 todas as perdas em cada época são somadas a fim de obter a curva deste somatório.

Figura 15 – Curva do somatório de perdas no cenário 1



Fonte: Elaborado pelo autor

De acordo com os gráficos podem ser destacados alguns pontos. Primeiramente, conforme o gráfico da Figura 12, o modelo atingiu a mAP próxima a final já a partir de 7 épocas, onde ficou em torno de 40%, mesmo que a função de perda continuasse a cair a uma taxa considerável até 13 épocas. Se observadas as perdas de cada componente, pode-se considerar que ambos atingiram um valor baixo, sendo o componente ROI, responsável por 90% da perda total. A partir destes dados, garante-se que o modelo foi capaz de aprender com o *dataset* de treinamento. Entretanto, quando observadas a AP de cada classe, tabela 7, é visto que a classe grave obteve o melhor resultado, atingindo 72,60%. Já as demais classes, ficaram entre 23,50% e 46,60%. Levando em consideração a função de perda, entende-se que a baixa AP das classes mínimo e moderado pode estar associada a quantidade insuficiente de âncoras disponíveis no treinamento, além de que, são classes que apresentam diferenças visuais sucintas. Portanto, são classes que geraram falsos positivos e falsos negativos, prejudicando a precisão final.

Na Figura 16 é exposta uma predição realizada após o treinamento do modelo. Nela foram detectadas cinco âncoras saudáveis uma âncora da classe grave.

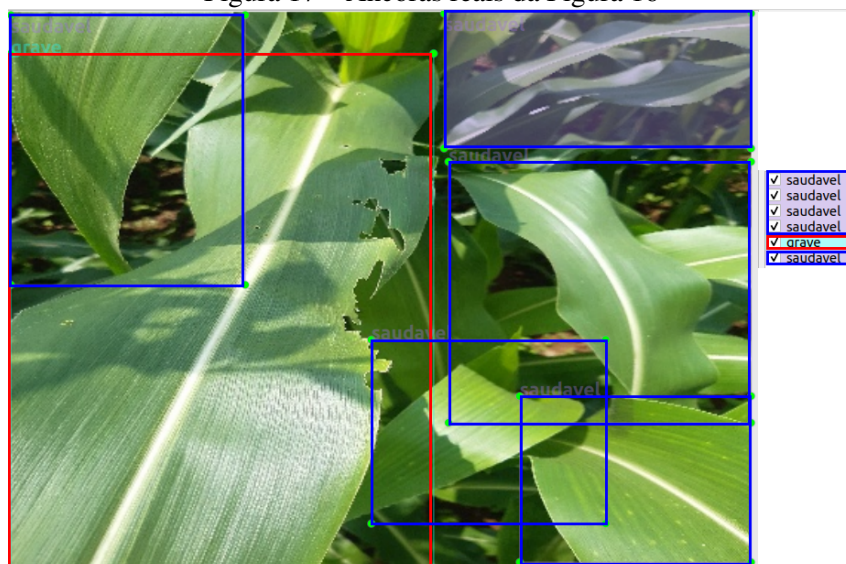
Figura 16 – Predição 1 do cenário 1



Fonte: Elaborado pelo autor

Na Figura 17 são exibidas as âncoras reais da Figura 16. Nota-se que o modelo foi capaz realizar a predição de todas.

Figura 17 – Âncoras reais da Figura 16



Fonte: Elaborado pelo autor

A Figura 18 apresenta outro exemplo de predição, onde uma âncora de dano moderado e uma de dano grave foram identificadas.

Figura 18 – Predição 2 do cenário 1



Fonte: Elaborado pelo autor

Na Figura 19 são exibidas as âncoras reais da Figura 18. Novamente o modelo fez a predição correta.

Figura 19 – Âncoras reais da Figura 18



Fonte: Elaborado pelo autor

Na próxima seção será explorado o segundo cenário, que diferentemente do primeiro, não possui as classes de dano mínimo, moderado e grave.

4.2 Cenário 2

Este cenário foi criado com o objetivo de medir a precisão do modelo em identificar danos de forma geral. Sendo assim, o *dataset* utilizado é composto apenas pelas classes saudável e

dano, sendo a classe dano a união das classes mínimo, moderado e grave. A tabela 8 apresenta a quantidade total de âncoras de cada classe.

Tabela 8 – Dataset do cenário 2

Classe	Âncoras
saudável	701
dano	1936

Fonte: Elaborado pelo autor

Após realizado o treinamento, foi possível obter a mAP de 73,09% sobre o *dataset* de teste. Onde cada classe teve a AP de acordo com a tabela 9.

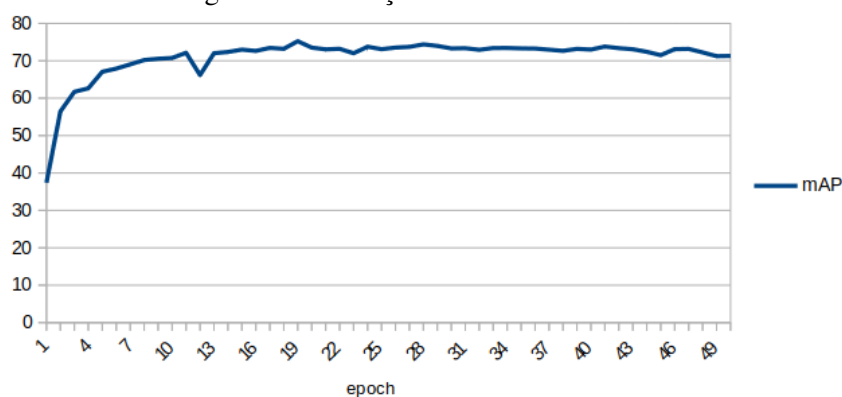
Tabela 9 – AP do cenário 2

Classe	AP
saudável	65,2%
dano	80,9%

Fonte: Elaborado pelo autor

A variação da mAP em cada época, durante o treinamento, é representada pela Figura 20. Nota-se que foram necessárias 13 épocas para que o modelo atingisse cerca de 70% de mAP.

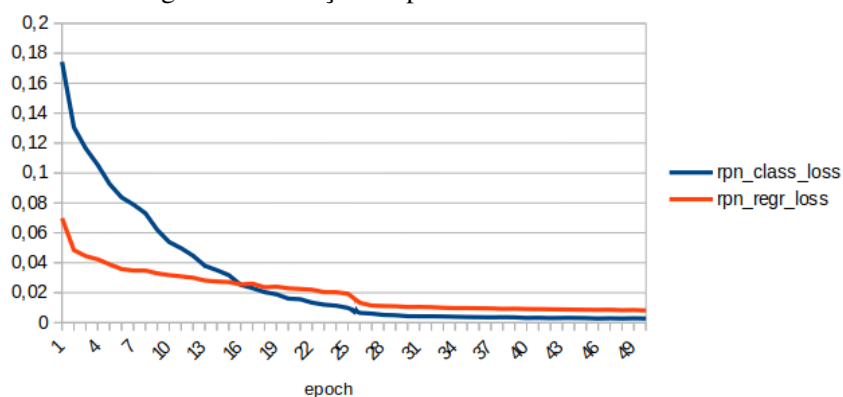
Figura 20 – Variação da mAP no cenário 2



Fonte: Elaborado pelo autor

Já as perdas associadas a classificação e regressão do componente RPN estão representadas na Figura 21. Neste cenário, assim como no cenário 1, o RPN busca diferenciar objetos de interesse e fundo, além de, contorná-los da melhor forma.

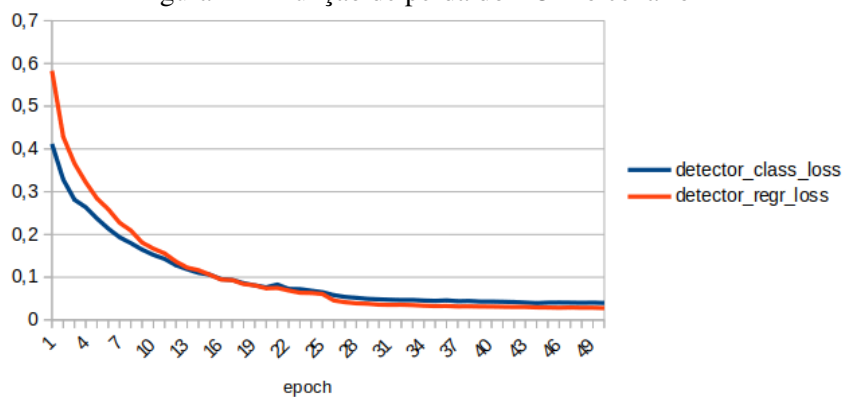
Figura 21 – Função de perda do RPN no cenário 2



Fonte: Elaborado pelo autor

Na Figura 22 é exposto o gráfico de perdas do componente ROI. Neste cenário, o ROI deve definir se o conjunto de âncoras é saudável ou dano, além de identificar sua posição na imagem.

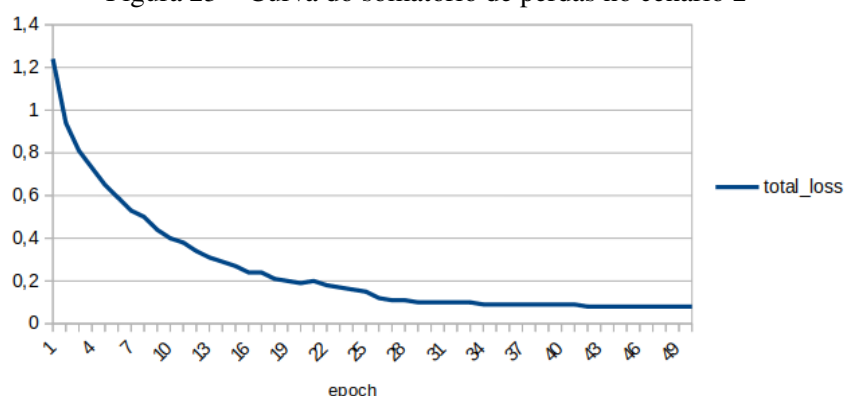
Figura 22 – Função de perda do ROI no cenário 2



Fonte: Elaborado pelo autor

Na Figura 23 todos as perdas em cada época são somados a fim de obter a curva deste somatório.

Figura 23 – Curva do somatório de perdas no cenário 2



Fonte: Elaborado pelo autor

De acordo com as métricas, o cenário 2 atingiu a precisão mAP de 73,09%, sendo muito maior que no cenário 1, o qual atingiu 43,41%. Individualmente a classe saudável atingiu um valor de 65,20% de AP, contra os 46,60% da classe saudável do cenário 1. A classe dano, se também comparada às classes de dano do cenário anterior, resultou numa precisão maior, sendo de 80,90%. Os gráficos das funções de perda dos componentes RPN e ROI convergiram para próximo de zero, com curvas muito similares, indicando que o modelo aprendeu com êxito com o *dataset* de treinamento.

O principal motivo para o modelo performar melhor neste cenário pode estar relacionado à quantidade de âncoras que a classe dano mínimo e moderado possuem. Pelo fato de totalizarem uma proporção pequena de âncoras, quando comparadas ao *dataset* como um todo, seus falsos positivos e falsos negativos são menos relevantes, ou seja prejudicam menos a AP das classes, consequentemente, a mAP final.

A Figura 24 é uma predição realizada após o treinamento do modelo. Uma âncora da classe saudável e uma âncora da classe dano foram identificadas.

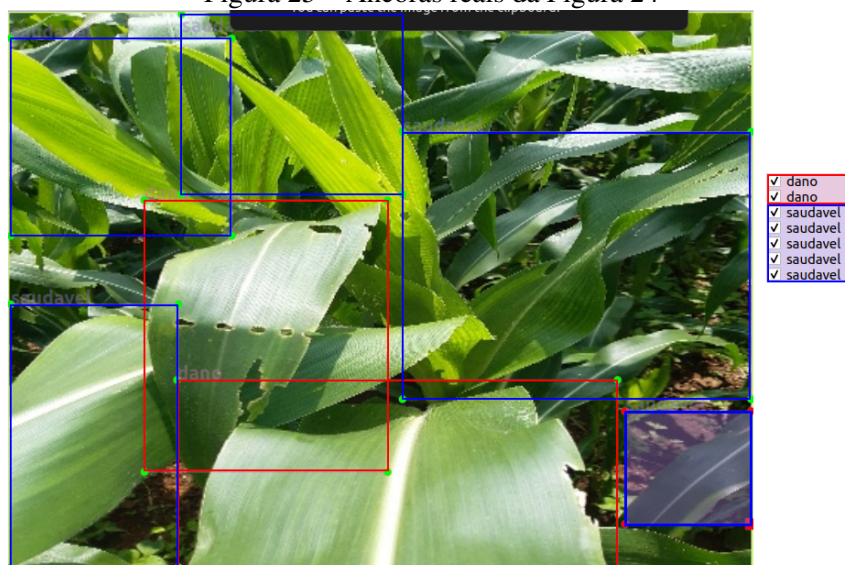
Figura 24 – Predição 1 do cenário 2



Fonte: Elaborado pelo autor

Na Figura 25 são exibidas as âncoras reais da Figura 24. Diferentemente do que foi predito, existem cinco âncoras da classe saudável e duas âncoras da classe dano, ou seja, foram gerados seis falsos negativos e um falso positivo.

Figura 25 – Âncoras reais da Figura 24



Fonte: Elaborado pelo autor

A Figura 26 apresenta outro exemplo de predição. Foram detectadas duas âncoras com dano e uma saudável.

Figura 26 – Predição 2 do cenário 2



Fonte: Elaborado pelo autor

A Figura 27 exibe as âncoras reais da Figura 26. Portanto, observa-se que a predição foi correta.

Figura 27 – Âncoras reais da Figura 26



Fonte: Elaborado pelo autor

Na próxima seção é apresentado o terceiro cenário, o qual, também fará análise de outra variação no *dataset*.

4.3 Cenário 3

O objetivo deste cenário é complementar a análise do cenário 2. Da mesma forma que o cenário 2, este cenário conta com apenas as classes saudável e dano. Porém, a classe de dano

mínimo foi retirada do conjunto de dano e adicionada no conjunto da classe saudável. Dessa forma, busca-se entender a influência da classe dano mínimo no resultado final.

Tabela 10 – Dataset do cenário 3

Classe	Âncoras
saudável + mínimo	806
dano	1831

Fonte: Elaborado pelo autor

Após realizado o treinamento, foi possível obter a mAP de 74,77% sobre o *dataset* de teste. Onde cada classe teve a AP de acordo com a tabela 11.

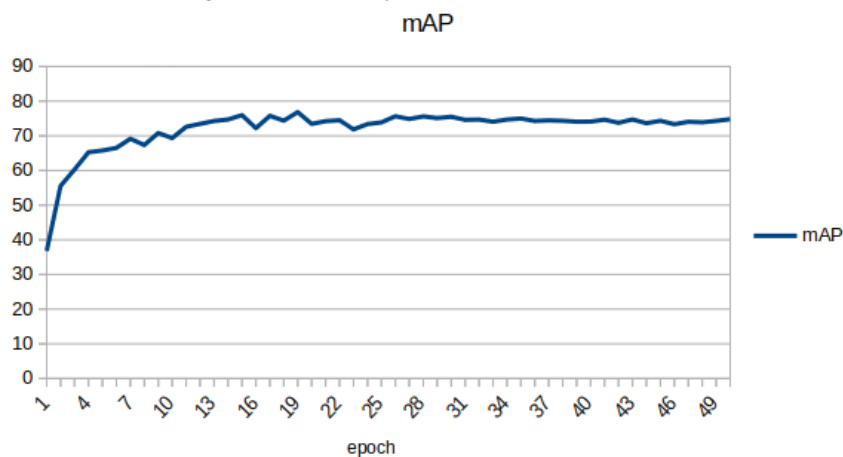
Tabela 11 – AP do cenário 3

Classe	AP
saudável + mínimo	66,10%
dano	83,50%

Fonte: Elaborado pelo autor

A variação de mAP em cada época, durante o treinamento, é representada pela Figura 28. Nota-se que foram necessárias 13 épocas para que o modelo atingisse cerca de 70% de mAP.

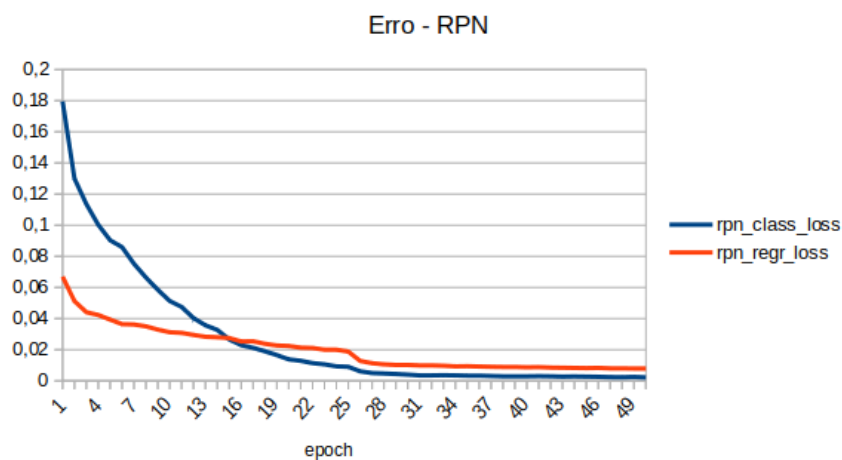
Figura 28 – Variação da mAP no cenário 3



Fonte: Elaborado pelo autor

As funções de perda associada à classificação e a regressão do componente RPN estão representados na Figura 29.

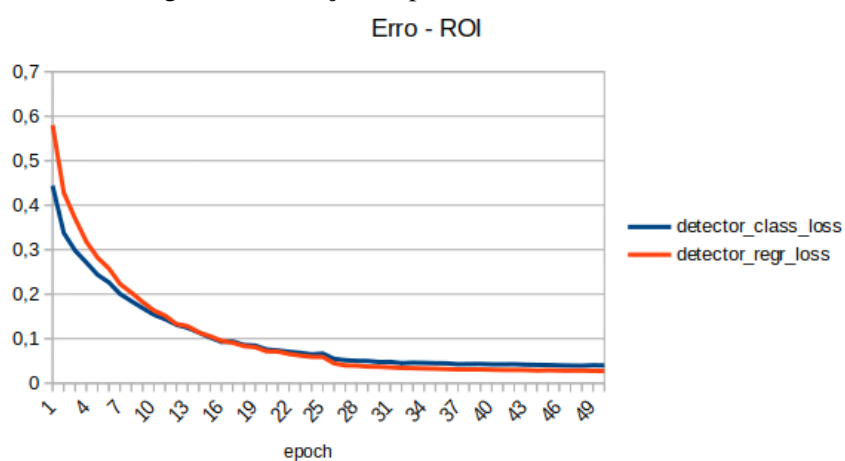
Figura 29 – Função de perda do RPN no cenário 3



Fonte: Elaborado pelo autor

Na Figura 30 é exposto o gráfico de perda do componente ROI.

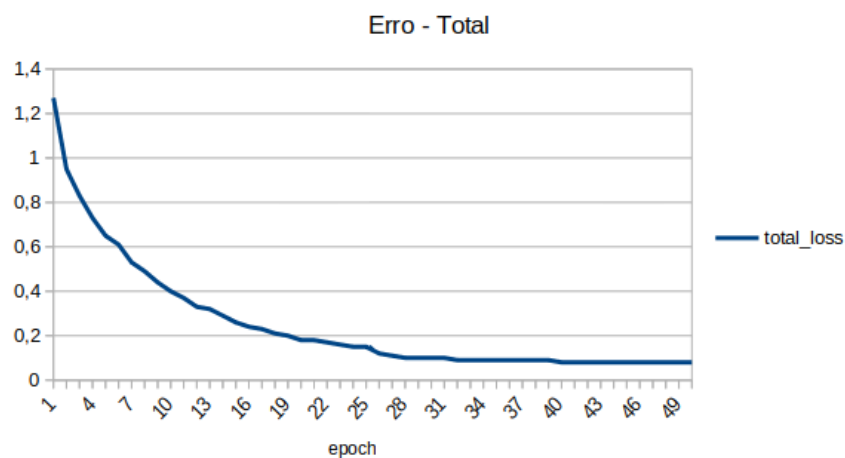
Figura 30 – Função de perda do ROI no cenário 3



Fonte: Elaborado pelo autor

Na Figura 31 a resultante de cada função de perda em cada época são somados a fim de obter a curva deste somatório.

Figura 31 – Curva do somatório de perdas do cenário 3



Fonte: Elaborado pelo autor

A Figura 32 é uma predição realizada após o treinamento do modelo. Estão presentes duas âncoras saudáveis e uma âncora com dano.

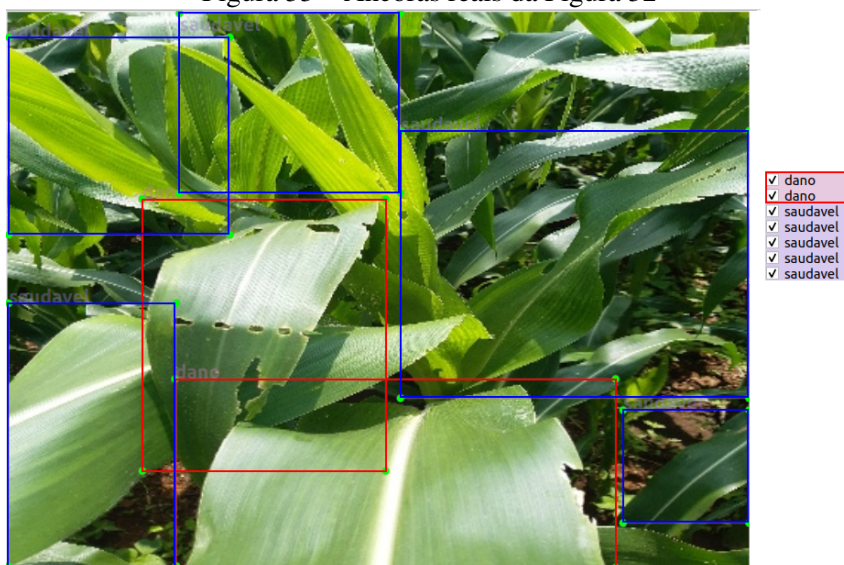
Figura 32 – Predição 1 do cenário 3



Fonte: Elaborado pelo autor

Na Figura 33 são exibidas as âncoras reais da Figura 32. Pode-se observar que o modelo acertou em apenas uma âncora de dano. Logo, foram gerados seis falsos negativos e dois falsos positivos.

Figura 33 – Âncoras reais da Figura 32



Fonte: Elaborado pelo autor

A Figura 34 apresenta outro exemplo de predição que identificou uma âncora saudável e uma âncora de dano.

Figura 34 – Predição 2 do cenário 3



Fonte: Elaborado pelo autor

Na Figura 35 são exibidas as âncoras reais da Figura 34. As duas âncoras preditas estão corretas, porém, existem três que não foram identificadas, ou seja, três falsos negativos.

Figura 35 – Âncoras reais da Figura 34



Fonte: Elaborado pelo autor

Comparando os resultados dos cenários 2 e 3, observa-se que a AP das classes saudável e saudável+mínimo, atingiram valores bem próximos, da mesma forma que as classes de dano. Consequentemente, ambos cenários atingiram a mAP em torno de 73%. Isso novamente aponta que a quantidade de âncoras que representam o dano mínimo geraram poucos falsos positivos e falsos negativos, tendo influencia menor na mAP final.

4.4 Cenário 4

O objetivo deste cenário complementar é de medir o desempenho da rede *VGG16* numa tarefa menos complexa. Diferentemente do propósito principal deste estudo, neste cenário as imagens são apenas classificadas pela presença ou ausência de danos, sem identificá-los ou diferenciá-los.

Para isso, o mesmo *dataset* com 4223 imagens foi dividido em treinamento e teste na proporção de 50% para cada. A implementação da rede *VGG16* foi feita utilizando a aplicação orange data mining nas suas configurações padrões. Após o treinamento, a rede foi submetida ao *dataset* de teste para ter suas métricas calculadas. A precisão e a acurácia resultantes foram, respectivamente, de 94% e 93,40%. Na tabela 12 é apresentada a matriz de confusão do cenário 4.

Tabela 12 – Matriz de confusão do cenário 4

		Predito	
		dano	saudável
Atual	dano	1108	7
	saudável	133	867

Fonte: Elaborado pelo autor

Como já esperado, a tarefa de apenas classificar imagens de forma binária obteve ótimos resultados. Sua complexidade em relação à identificar e classificar objetos de interesse dentro de imagens, é consideravelmente menor. Sendo uma abordagem poderosa para determinados contextos.

4.5 Análise Geral

O modelo *Faster R-CNN* foi capaz de convergir na identificação das três classes de dano propostas neste estudo. Os gráficos de mAP e função de perda dos componentes RPN e ROI, mostram que o modelo aprendeu com o *dataset* de treinamento. Outro ponto importante é referente a avaliação do modelo através do *dataset* de teste, onde o resultado final de mAP foi muito coerente aos resultados gerados nos treinamentos e expostos pelos seus gráficos de mAP por época. Em relação às AP das classes de dano mínimo e moderado, entende-se que o modelo não performou bem devido à limitação do *dataset*.

Por fim, com o objetivo de comparar os resultados deste estudo com o resultado de outro estudo correlato, que também utilizou o modelo *Faster R-CNN* para identificação de danos na folha do milho causados pela FAW, chega-se ao estudo de Du et al. (2022). Du et al. (2022) construiu um *dataset* composto por quatro classes: juvenil, menor, moderado e severo. Cada classe representa a gravidade do dano, sendo a juvenil a menos atacada e a severa a classe que apresenta mais danos na planta. O modelo de Du et al. (2022) obteve a mAP de 31,20%, sendo próximo ao cenário 1 deste estudo, que resultou na mAP de 43,41% e possui classes similares ao trabalho de Du et al. (2022). A tabela 13 exhibe os resultados de cada cenário e também o resultado de Du et al. (2022).

Tabela 13 – Comparativo dos cenários com o estudo de Du et al. (2022)

Cenário	Total de classes	mAP
1	4	43,41%
2	2	73,09%
3	2	74,77%
Du et al. (2022)	4	31,20%

Fonte: Elaborado pelo autor

Utilizando uma abordagem complementar, no cenário 4 foi possível compreender a grande capacidade que a rede neural convolucional *VGG16* sozinha possui em fazer a predição de imagens, mostrando-se uma ferramenta robusta para contextos mais simples de predição. No próximo capítulo será feito o fechamento deste estudo e também propostas de melhorias futuras.

5 CONCLUSÃO

Este estudo explorou a utilização de um modelo baseado em redes neurais convolucionais, para identificar e classificar em quatro classes o dano da FAW na folha do milho. O modelo *Faster R-CNN* mostrou-se promissor na tarefa proposta. Apesar de ter apresentado baixa AP nas classes de dano mínimo e saudável, ainda existem possibilidades de otimizá-lo em trabalhos futuros, ao obter um *dataset* maior e explorar outros hiper-parâmetros.

O estudo inicialmente apresentou a base teórica sobre a FAW, suas características físicas, hábitos alimentares, aspecto do seu dano na planta do milho, suas formas mais tradicionais de monitoramento, sua propagação no mundo e seu impacto econômico na agricultura. Também foi apresentada a escala de Davis, a qual auxilia na definição dos níveis de dano e pode ser usada como referência para o controle da praga. Complementando o referencial teórico, as redes neurais convolucionais foram introduzidas, dando ênfase no extrator de características *VGG16* e no modelo *Faster R-CNN*. Finalizando a base teórica, foram selecionados estudos que utilizaram redes neurais convolucionais como ferramentas de auxílio na agricultura.

De acordo com a metodologia, foi criado o *dataset* contendo imagens das plantas atacadas e feita a configuração da rede neural convolucional. Após, três cenários foram montados e submetidos ao treinamento, teste e extração das métricas de desempenho. O primeiro cenário, composto pela classe saudável e as diferentes classes de dano, obteve um bom valor de AP na classe de dano grave, acima dos 70%. Os demais valores de AP foram abaixo dos 50%, precisamente, 46,60%, 30,90% e 23,50%, respectivamente, para as classes saudável, danos mínimo e moderado. Complementando a análise, os cenários 2 e 3 foram criados com o propósito de entender a influência das classes dano mínimo e moderado, já que são classes com poucas âncoras. Ambos obtiveram valores bem próximos de AP quando comparados, ou seja, a influência dessas classes foi mínima nos cenários 2 e 3. Além disso, essa menor quantidade de classes possibilitou aos cenários 2 e 3 apresentarem valores melhores de mAP, sendo de, 73,09% e 74,77%, contra os 43,41% do cenário 1.

Com os resultados de cada cenário foi possível obter a precisão do modelo *Faster R-CNN* e também entender seus pontos positivos e suas limitações. Destaca-se principalmente a mAP resultante dos treinamentos e do testes, que foram valores próximos para cada cenário, ou seja, sendo um ótimo indicador de que o modelo foi bem calibrado.

Portanto, o modelo de detecção e classificação apresentado por este estudo, mostra-se propício como parte de soluções robustas de monitoramento de pragas em plantações de milho.

5.1 Trabalhos Futuros

Um fator importante para que redes neurais convolucionais possam atingir alta precisão, é em relação à qualidade e quantidade de âncoras utilizadas no seu treinamento. Ou seja, uma melhoria proposta é a da criação de um *dataset* com pelo menos 2000 âncoras para cada classe.

Assim como empregado nos estudos de Du et al. (2022), Ishengoma, Rai e Said (2021), Feng et al. (2022), Ishengoma, Rai e Ngoga (2022) e Wu et al. (2019), outra sugestão é monitorar plantações com o uso de veículo aéreo não-tripulado. O VANT realiza o monitoramento de forma mais eficiente, sendo mais rápido e fácil que a captura terrestre feita por um humano. Outro ponto interessante do uso de VANT é em relação às imagens capturadas, que são capazes de cobrir diversas plantas. Portanto, para que o modelo performe corretamente é necessário treiná-lo a partir deste padrão de imagem.

Outra possibilidade interessante é a de explorar novos modelos de redes neurais convolucionais detectoras de objetos que são evoluções do *Faster R-CNN*, como por exemplo o *Mask R-CNN*.

REFERÊNCIAS

- ACHARYA, R. **Corn leaf infection dataset**. 2020. Disponível em: <<https://www.kaggle.com/datasets/qramkrishna/corn-leaf-infection-dataset>>, Acesso em: 10 de ago. de 2022.
- CAPINERA, J. L. Fall armyworm, *spodoptera frugiperda* (j.e. smith) (insecta: lepidoptera: noctuidae): eeny098/in255, rev. 7/2000. **EDIS**, [S.l.], v. 2020, n. 7, Oct. 2002.
- DU, L. et al. A novel object detection model based on faster r-cnn for *spodoptera frugiperda* according to feeding trace of corn leaves. **Agriculture**, [S.l.], v. 12, n. 2, 2022.
- ESCHEN, R. et al. Towards estimating the economic cost of invasive alien species to african crop and livestock production. **CABI Agriculture and Bioscience**, [S.l.], v. 2, 05 2021.
- FAO. **The global action for fall armyworm control**: a resource mobilization guide. Roma, 2022.
- FENG, J. et al. Autonomous detection of *spodoptera frugiperda* by feeding symptoms directly from uav rgb imagery. **Applied Sciences**, [S.l.], v. 12, n. 5, 2022.
- FUENTES, A. et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. **Sensors**, [S.l.], v. 17, n. 9, 2017.
- HAYKIN, S. **Redes neurais**: princípios e prática. [S.l.]: Bookman, 2003.
- IPPC SECRETARIAT. **Scientific review of the impact of climate change on plant pests - a global challenge to prevent and mitigate plant-pest risks in agriculture, forestry and ecosystems**. Roma, 2021.
- ISHENGOMA, F. S.; RAI, I. A.; NGOGA, S. R. Hybrid convolution neural network model for a quicker detection of infested maize plants with fall armyworms using uav-based images. **Ecological Informatics**, [S.l.], v. 67, p. 101502, 2022.
- ISHENGOMA, F. S.; RAI, I. A.; SAID, R. N. Identification of maize leaves infected by fall armyworms using uav-based imagery and convolutional neural networks. **Computers and Electronics in Agriculture**, [S.l.], v. 184, p. 106124, 2021.
- KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: a survey. **Computers and Electronics in Agriculture**, [S.l.], v. 147, p. 70–90, 2018.
- LI, W. et al. Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline. **Scientific Reports**, [S.l.], v. 9, 05 2019.
- LI, Y. et al. Crop pest recognition in natural scenes using convolutional neural networks. **Computers and Electronics in Agriculture**, [S.l.], v. 169, p. 105174, 2020.
- LU, Y. et al. Identification of rice diseases using deep convolutional neural networks. **Neurocomputing**, [S.l.], v. 267, 07 2017.
- MARENCO, R. J.; FOSTER, R. E.; SANCHEZ, C. A. Sweet corn response to fall armyworm (lepidoptera: noctuidae) damage during vegetative growth. **Journal of Economic Entomology**, [S.l.], v. 85, n. 4, p. 1285–1292, 08 1992.

MCLEOD, R. **Bug guide**. 2006. Disponível em: <<https://bugguide.net/node/view/40787>>, Acesso em: 09 de jun. de 2022.

MILITANTE, S. V.; GERARDO, B. D.; DIONISIO, N. V. Plant leaf detection and disease recognition using deep learning. In: IEEE EURASIA CONFERENCE ON IOT, COMMUNICATION AND ENGINEERING (ECICE), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 579–582.

PHUNG; RHEE. A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. **Applied Sciences**, [S.l.], v. 9, p. 4500, 10 2019.

REN, S. et al. Faster r-cnn: towards real-time object detection with region proposal networks. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2015. **Anais...** Curran Associates: Inc., 2015. v. 28.

SARKOWI, F.; MOKHTAR, A. The fall army worm the fall armyworm (faw) spodoptera frugiperda: a review on biology, life history, invasion, dispersion and control. **Outlooks on Pest Management**, [S.l.], p. 27, 03 2021.

SCHNEIDER, J. P.; BAUMGRATZ, C. **Pioneer sementes**. 2015. Disponível em: <<https://www.pioneersementes.com.br/blog/74/aplicacao-complementar-de-inseticida-monitoramento-e-tomada-de-decisao>>, Acesso em: 09 de jun. de 2022.

SUGATA, T.; YANG, C. Leaf app: leaf recognition with deep convolutional neural networks. **IOP Conference Series: Materials Science and Engineering**, [S.l.], v. 273, p. 012004, 11 2017.

THENMOZHI, K.; Srinivasulu Reddy, U. Crop pest classification based on deep convolutional neural network and transfer learning. **Computers and Electronics in Agriculture**, [S.l.], v. 164, p. 104906, 2019.

TOEPFER, S. et al. Streamlining leaf damage rating scales for the fall armyworm on maize. **Journal of Pest Science**, [S.l.], v. 94, 09 2021.

TOO, E. C. et al. A comparative study of fine-tuning deep learning models for plant disease identification. **Computers and Electronics in Agriculture**, [S.l.], v. 161, p. 272–279, 2019. BigData and DSS in Agriculture.

TRZYNADLOWSKI, B. **Faster r-cnn in pytorch and tensorflow 2 w/ keras**. 2021. Disponível em: <<https://github.com/trzy/FasterRCNN>>, Acesso em: 10 de ago. de 2022.

WU, H. et al. Autonomous detection of plant disease symptoms directly from aerial imagery. **The Plant Phenome Journal**, [S.l.], v. 2, n. 1, p. 190006, 2019.

ZHANG, A. et al. Dive into deep learning. **arXiv preprint arXiv:2106.11342**, [S.l.], 2021.

ZHANG, X. et al. Identification of maize leaf diseases using improved deep convolutional neural networks. **IEEE Access**, [S.l.], v. 6, p. 30370–30377, 2018.

ZHAO, S. et al. Tomato leaf disease diagnosis based on improved convolution neural network by attention module. **Agriculture**, [S.l.], v. 11, n. 7, 2021.