

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

GUILHERME ANTUNES

CLASSIFICAÇÃO DE ESPÉCIES DE PLANTAS A PARTIR DE IMAGENS DAS
COMPONENTES FOLHA E FLOR

PONTA GROSSA
2021

GUILHERME ANTUNES

CLASSIFICAÇÃO DE ESPÉCIES DE PLANTAS A PARTIR DE IMAGENS DAS
COMPONENTES FOLHA E FLOR

Dissertação apresentada para obtenção do
título de Mestre na Universidade Estadual de
Ponta Grossa ao Programa de Pós-
Graduação em Computação Aplicada.

Orientador: Prof. Dr. Alceu de Souza Britto Jr.

PONTA GROSSA

2021

A636 Antunes, Guilherme
Classificação de espécies de plantas a partir de imagens das componentes
folha e flor / Guilherme Antunes. Ponta Grossa, 2021.
67 f.

Dissertação (Mestrado em Computação Aplicada - Área de Concentração:
Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta
Grossa.

Orientador: Prof. Dr. Alceu de Souza Britto Júnior.

1. Classificação de plantas. 2. Aprendizado profundo. 3. Rede neural
convolucional. 4. Combinação de classificadores. I. Britto Júnior, Alceu de Souza.
II. Universidade Estadual de Ponta Grossa. Computação para Tecnologias em
Agricultura. III.T.

CDD: 004



UNIVERSIDADE ESTADUAL DE PONTA GROSSA
Av. General Carlos Cavalcanti, 4748 - Bairro Uvaranas - CEP 84030-900 - Ponta Grossa - PR - <https://uepg.br>

TERMO

TERMO DE APROVAÇÃO

Guilherme Antunes

CLASSIFICAÇÃO DE ESPÉCIES DE PLANTAS A PARTIR DE IMAGENS DAS COMPONENTES FOLHA E FLOR

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Ponta Grossa, 09 de dezembro de 2021.

Prof. Dr. Alceu de Souza Britto Júnior (UEPG, Presidente)

Prof. Dr. José Carlos Ferreira da Rocha (UEPG)

Prof. Dr. André Hochuli (PUC/PR)



Documento assinado eletronicamente por **Alceu de Souza Britto Junior, Professor(a)**, em 19/01/2022, às 14:48, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Jose Carlos Ferreira da Rocha, Coordenador(a) do Programa de Pós-Graduação em Computação Aplicada - Mestrado**, em 21/01/2022, às 15:42, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **ANDRE GUSTAVO HOCHULI, Usuário Externo**, em 25/01/2022, às 09:39, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <https://sei.uepg.br/autenticidade> informando o código verificador **0856258** e o código CRC **387EEAE9**.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que me deu força e sabedoria para que pudesse concluir com êxito este curso.

A minha esposa e filha, Daniele e Manuela Antunes, que durante todo esse período me incentivaram, acompanhando de perto todo processo para que eu pudesse avançar na pesquisa. Aos meus pais Luiz Carlos e Gilce e minhas irmãs Liziane e Izabelle, que me apoiaram e incentivaram desde o começo.

Ao meu orientador, Prof. Dr. Alceu de Souza Britto Jr, pelo empenho dedicado à elaboração deste trabalho, me fazendo trilhar uma jornada que até então não conhecia.

A Universidade Estadual de Ponta Grossa, pela oportunidade de fazer o curso, disponibilizando um corpo de professores com muita qualidade.

A todos que direta ou indiretamente fizeram parte de minha formação, o meu muito obrigado.

RESUMO

A classificação de espécies de plantas é um desafio devido a biodiversidade de nosso planeta, e nossos múltiplos ecossistemas tendem a exponencializar esse problema, para isso, soluções ao longo dos anos têm sido desenvolvidas de modo que possamos automatizar o trabalho de classificação de plantas, que antes estava reservado ao corpo técnico específico da área. As Redes Neurais Convolucionais e o Aprendizado Profundo, do inglês *deep learning*, têm viabilizado soluções cada vez mais atrativas para a tarefa de classificação. Fornecendo a seleção de atributos e extração de características de maneira autônoma. Para esta pesquisa foi utilizada uma base de imagens pertencente ao desafio mundial de reconhecimento de plantas 2015, LifeCLEF. Esta base possui 113.205 imagens de 1000 espécies diferentes. Neste trabalho foi abordado a classificação de espécies de plantas utilizando dois componentes de uma mesma planta, sendo eles a folha e a flor. Para tal, redes neurais profundas pré-treinadas na base Imagenet foram utilizadas para classificação de cada componente. Em seguida, diferentes regras de combinação de classificadores foram avaliadas, da soma à criação de um meta-classificador responsável pela fusão. Resultados experimentais permitiram um aumento de até 23 pontos percentuais (de 68% para 91%) quando realizada a fusão dos classificadores das componentes folha e flor.

Palavras-chave: Classificação de Plantas, Aprendizado Profundo, Rede Neural Convolucional, Combinação de classificadores.

ABSTRACT

The classification of plant species is very challenging due to the biodiversity of our planet and our multiple ecosystems exponents this problem, for this, solutions over the years have been developed so that we can automate the work of plant classification, which before it was exclusive for the technical staff of the area. Convolutional Neural Networks and Deep Learning have been increasing the possibility of most suitable solutions for the task of classification making possible autonomous feature extraction. For this research, an image database belonging to the 2015 World Plant Recognition Challenge, LifeCLEF, this database has 113205 images of one thousand different species. In this work the classification of plant species was addressed using two components of the same plant, the leaf and the flower. To this end, pre-trained deep neural networks in the Imagenet database were used to classify each component, then different classifier combination rules were evaluated, creating a meta classifier responsible for merging. Experimental results allowed an increase of up to 23 percentage points (from 68% to 91 %) when the fusion of the leaf and flower components classifiers was performed.

Keywords: Plant Classification, Deep Learning, Convolutional Neural Network, Classifier Combination.

LISTA DE FIGURAS

Figura 1. Semelhanças entre diferentes espécies de plantas (inter-classe).....	13
Figura 2. Semelhanças entre diferentes espécies de plantas (inter-classe).....	14
Figura 3. Variações sobre a mesma espécie (intra-classe) – Folha.....	14
Figura 4. Variações sobre a mesma espécie (intra-classe) – Flor.....	14
Figura 5. Divisões na área de Inteligência Artificial.....	18
Figura 6. CNN x Camada completamente conectada.....	19
Figura 7. Matriz de entrada multiplicada por detector de características.....	20
Figura 8. Pooling usando operador Máximo, com filtro 2x2.....	22
Figura 9. Problema de sobreajuste (overfitting) em RNA.....	23
Figura 10. Etapa Flattening - transformando matriz em um vetor.....	24
Figura 11. Rede Neural Convolutacional.....	25
Figura 12. Estrutura básica - Algoritmo genético.....	26
Figura 13. Arquitetura padrão da rede VGG-16.....	33
Figura 14. Arquitetura padrão da rede NasNetMobile.....	34
Figura 15. Arquitetura padrão da rede EfficientNet B0.....	35
Figura 16. Arquitetura padrão da rede ResNet50.....	35
Figura 17. Arquitetura padrão da rede InceptionV3.....	36
Figura 18. Arquitetura padrão da rede InceptionResnetV2.....	36
Figura 19. Arquitetura padrão da rede Xception.....	37
Figura 20. Método de classificação proposto.....	39
Figura 21. Flores, XML metadados e Folhas.....	41
Figura 22. Imagens por classe - LeafScan e Flower.....	42
Figura 23. Aumento de dados utilizando Augmentor.....	43
Figura 24. Fluxo de uma CNN.....	45
Figura 25. Gráfico de evolução de acurácia leafscan (treino e validação).....	49
Figura 26. Matriz LeafScan - Melhor resultado individual (InceptionResnetV2).....	50
Figura 27. Acurácia LeafScan todas redes pré-treinadas.....	51
Figura 28. Acurácia LeafScan - Aumento de Dados.....	51
Figura 29. Gráfico de evolução de acurácia flower (treino e validação).....	52
Figura 30. Matriz - Flower - Melhor resultado individual (Resnet50).....	54
Figura 31. Acurácia Flower todas as redes pré-treinadas.....	55
Figura 32. Acurácia Flower - Aumento de Dados.....	55

Figura 33. Fluxo da fusão.....	56
Figura 34. Acurácia na Combinação dos modelos.....	57
Figura 35. Matriz - Melhor Combinação (Produto + Algoritmos Genéticos).....	58
Figura 36. Imagens de entrada na SCNN.....	59
Figura 37. Imagens Acer Campestre (flor e folha).....	60
Figura 38. Imagens Viburnum Opulus L (flor e folha).....	60
Figura 39. Imagens Fraxinus Angustifolia Vahl (flor e folha).....	60
Figura 40. Buddleja Davidii Franch predita pela Fraxinus Angustifolia Vahl.....	61
Figura 41. Fraxinus excelsior L. predita pela Fraxinus Angustifolia Vahl.....	61
Figura 42. Fraxinus ornus L. predita pela Fraxinus Angustifolia Vahl.....	61
Figura 43. Melhores resultados (Individuais, combinação, top-3 e top-5).....	62

LISTA DE QUADROS

Quadro 1 - Comparação de resultados de levantamento bibliográfico.....	32
Quadro 2 - Divisão da base, treino, validação e teste.....	42
Quadro 3 - Configuração de Hiperparâmetros.....	44
Quadro 4. Treinamento Leafscan - Melhor result. indiv. (InceptionResnetV2).....	47
Quadro 5. Treinamento Flower - Melhor resultado individual (Resnet50).....	50
Quadro 6. Parâmetros Algoritmos Genéticos.....	54
Quadro 7. Comparação com trabalhos que realizaram combinações.....	59

LISTA DE SIGLAS

AG - Algoritmo Genético
CNN - Rede Neural Convolutiva
DL - Deep Learning
FNC - Rede Totalmente Conectada
GD - Descida do Gradiente
HOG - Histograma de Gradientes Orientada
LBP - Padrão Binário Local
MCS - Sistema de Classificador Múltiplo
ML - Machine Learning
RNC - Rede Neural Convolutiva
RNN - Rede Neural Recorrente
SCNN - Rede Neural Convolutiva Siamesa
SGD - Descida do Gradiente Estocástica
SURF - Recursos Robustos Acelerados
ZM - Momentos de Zernike

SUMÁRIO

1 INTRODUÇÃO	12
1.1 MOTIVAÇÃO	12
1.2 DESAFIOS	13
1.3 OBJETIVOS	15
1.3.1 Objetivo Geral	15
1.3.2 Objetivos Específicos	15
1.4 ESTRUTURA	15
2 FUNDAMENTAÇÃO TEÓRICA	17
2.1 APRENDIZADO DE MÁQUINA	17
2.2 APRENDIZADO SUPERVISIONADO	17
2.3 APRENDIZADO PROFUNDO	17
2.4 REDE NEURAL CONVOLUCIONAL	18
2.4.1 Operador de Convolução	19
2.4.2 Camada de Pooling	21
2.4.3 Camada Dropout	22
2.4.4 Achatamento (Flattening)	23
2.4.5 Camada Totalmente Conectada	24
2.5 ALGORITMOS GENÉTICOS	25
2.6 CONSIDERAÇÕES FINAIS	26
3 ESTADO DA ARTE	27
3.1 CNN PARA CLASSIFICAÇÃO DE IMAGENS	32
3.1.1 Visual Geometry Group (VGG-16)	33
3.1.2 NasNetMobile	34
3.1.3 EfficientNet	34
3.1.4 Resnet	35
3.1.5 Inception	35
3.1.6 InceptionResnet	36
3.1.7 Xception	37
3.2 CONSIDERAÇÕES FINAIS	37
4 MATERIAIS E MÉTODOS	39
4.1 SISTEMA COMPUTACIONAL	39
4.2 PRÉ-PROCESSAMENTO - ETAPA A	40
4.2.1 Preparo das Imagens	41
4.2.2 Aumento dos Dados	42
4.3 DEFINIÇÃO DOS MODELOS PROFUNDOS - ETAPA B	43
4.4 TREINAMENTO DOS MODELOS	44
4.5 AVALIAÇÃO INDIVIDUAL DOS MODELOS - ETAPA D	45
4.6 TÉCNICAS DE FUSÃO - ETAPA E	46
5 RESULTADOS E DISCUSSÕES	47
5.1 LEAFSCAN - FOLHA	47
5.2 FLOWER - FLOR	51
5.3 FUSÃO DO MODELOS	54
5.3.1 Seleção Estática de Classificadores usando Algoritmo Genético	54
5.4 COMPARAÇÃO DOS RESULTADOS OBTIDOS	57

6 CONCLUSÕES	61
REFERÊNCIAS.....	63

1 INTRODUÇÃO

Na natureza existe um vasto número de espécies de plantas, as quais muitas delas já estão catalogadas e referenciadas. Existe uma demanda pela necessidade de conhecer e classificar várias outras espécies, primeiro pelo fato de que muitas estão entrando em extinção, afetando assim a biodiversidade e múltiplos ecossistemas. Outro motivo é que a cada dia novas espécies são descobertas, no entanto, para que uma planta seja identificada e catalogada de maneira adequada, é necessário um corpo técnico que seja especialista para que ela possa ser referenciada de maneira correta, tornando o trabalho de classificação de espécies muito custoso, logo, pesquisas como essa são substanciais para nossa sociedade (RZANNY et al. 2019; MURPHY et al. 2014).

Encontram-se no estado da arte pesquisas que buscam classificar os vegetais através da folha da planta, utilizando como recursos a cor, a forma (GUYER et al. 1986), estrutura de textura (YANIKOGLU et al. 2014), morfologia e venação. Alguns pesquisadores utilizam a forma da folha, com a justificativa de que embora a planta possa sofrer alterações seja pela idade ou por insetos, essa estrutura consiga ser preservada (ARAÚJO et al. 2017).

1.1 MOTIVAÇÃO

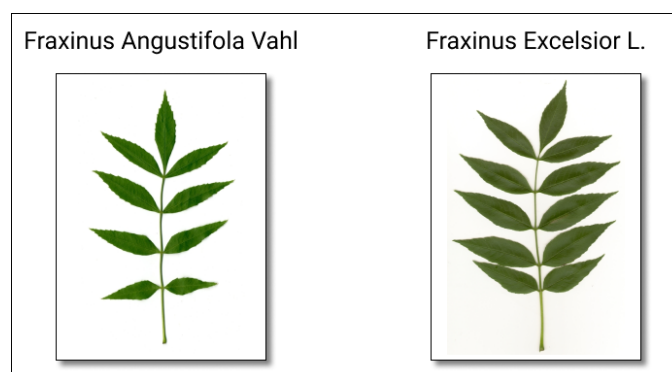
De forma consonante, este trabalho de pesquisa busca contribuir para a automação da classificação de espécies explorando a combinação de componentes da planta. Sendo assim, a forma de análise de multicomponentes é proposta neste trabalho, de modo que possam ser utilizados para classificação não apenas a estrutura da folha do vegetal, mas também a flor, obtendo-se assim uma maior quantidade de informações do objeto para realizar a classificação. Essa maior gama de possibilidades de informações respalda e justifica o trabalho, pois em determinadas situações a degradação da própria planta já impede que ela seja classificada, mas a partir do momento em que mudamos o cenário, podemos possibilitar a classificação.

1.2 DESAFIOS

A diversidade das plantas, aliado às peculiaridades que cada uma possui, torna a tarefa de identificação de plantas um tanto desafiadora. Para que se consiga executar existem alguns conjuntos de dados, podendo destacar a Flavia (WU et. al 2007), criado para que a Visão Computacional pudesse ser avaliada. A problemática desse conjunto é que pela falta de diversidade, tanto na forma como as imagens foram captadas quanto do ponto de vista da diversidade de espécies de plantas, acaba limitando a generalização do modelo para aplicação em um mundo real (ARAÚJO 2016). Logo, a base que será utilizada contempla uma diversidade de espécie de plantas, e não existe um protocolo fixo na forma como as imagens foram captadas, isso traz um ganho na generalização do problema e aplicação, mas também tende a exponencializar o nível de dificuldade para que se alcance bons resultados, pois é um conjunto de dados mais complexo, LifeCLEF 2015.

Um dos desafios prioritários na classificação de espécie de plantas, concentra-se na baixa variabilidade inter-classe, ou seja, plantas que são semelhantes mas pertencem a espécies diferentes, a Figura 1 referencia visualmente um exemplo de semelhanças entre diferentes espécies de plantas.

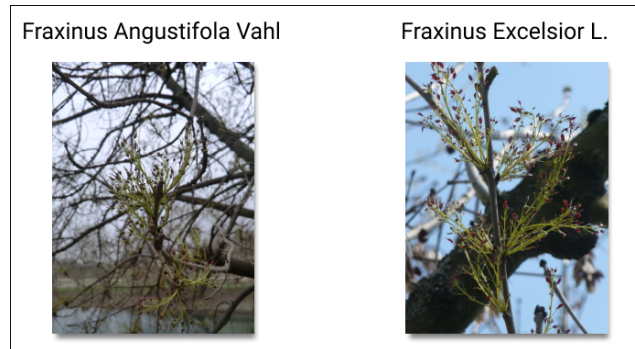
Figura 1. Exemplo de semelhanças entre diferentes espécies de plantas (inter-classe) - Folha



Fonte: O autor

A Figura 2 mostra a semelhança entre diferentes espécies a partir da vista da flor, a qual também será utilizada neste projeto de pesquisa.

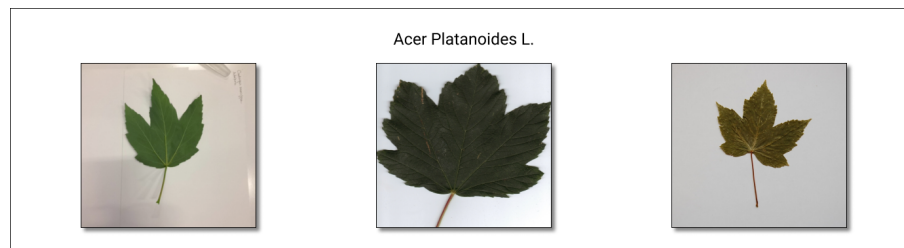
Figura 2. Exemplo de semelhanças entre diferentes espécies de plantas (inter-classe) – Flor



Fonte: O autor

A alta variabilidade intra-classe, ou seja, plantas diferentes mas que pertencem à mesma classe, também é um grande desafio na tarefa de classificação, pois essa grande diversidade vai tornando cada vez mais complexa o aprendizado durante o treinamento para a rede neural. A Figura 3 e 4, mostram essa diversidade a partir da vista da folha e da flor, respectivamente.

Figura 3. Exemplo de variações sobre a mesma espécie (intra-classe) - Folha



Fonte: O autor

Figura 4. Exemplo de variações sobre a mesma espécie (intra-classe) - Flor



Fonte: O autor

É também destacável a dificuldade de que algumas espécies de plantas possuem um número reduzido de exemplares, tornando mais difícil a criação de um modelo que consiga representar, visto que tarefas de aprendizado de máquinas, e o uso de redes neurais convolucionais necessitam de uma grande quantidade de dados para alcançar um resultado satisfatório.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo geral é desenvolver um método de classificação de espécies de plantas a partir da imagem de seus componentes, sendo eles: folha e flor. Para tal, pretende-se utilizar aprendizagem de máquina, em especial, modelos profundos para aprendizado da representação (extração de características) e classificação propriamente dita de cada componente. Além disso, pretende-se investigar técnicas de fusão dos modelos gerados para cada componente da planta a fim de obter uma classificação final.

1.3.2 Objetivos Específicos

Como objetivos específicos, têm-se:

- Desenvolver modelos profundos para classificação da planta usando os componentes folha e flor;
- Avaliar o desempenho dos classificadores criados individualmente;
- Implementar estratégias de fusão dos classificadores baseados em folha e flor;
- Avaliar as diferentes estratégias de fusão implementadas e comparar com o estado da arte.
- Melhorar o patamar de acurácia presente no estado da arte.

1.4 ESTRUTURA

Este trabalho está dividido em 7 capítulos. O segundo capítulo mostra a Fundamentação Teórica. No capítulo 3 teremos o Estado da Arte. No capítulo 4 os

Materiais e os Métodos para a execução do projeto. No capítulo 5 discorre os Resultados e as Discussões, seguido das Conclusões. Por fim, no Capítulo 7 as Referências.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 APRENDIZADO DE MÁQUINA

As técnicas da Aprendizagem de Máquina permitem a criação de modelos cognitivos a partir dos dados (exemplos) do problema que se pretende resolver. Através de métodos estatísticos e algoritmos computacionais, a busca de padrões para resultar em previsões, ou seja, realizar o treinamento com determinada base de dados e partir disso realizar inferências (MJOLSNESS, 2001). Dentre as tarefas de ML destacam-se a regressão, a classificação, clustering, entre outros.

As técnicas de ML estão organizadas em quatro possíveis abordagens, a saber: supervisionada, não supervisionada, semi-supervisionada e por reforço. Neste trabalho focamos na primeira abordagem, pois o objetivo é a criação de modelos preditivos para a tarefa de classificação.

2.2 APRENDIZADO SUPERVISIONADO

De acordo com Pellucci et. al. (2011), o aprendizado supervisionado consiste em obter conclusões a partir de um conjunto de dados fornecidos, ou seja, as informações são extraídas gerando uma base de conhecimento. Posteriormente é mostrado algum item novo que não estivesse no aprendizado, mas que possuísse características fundamentais já aprendidas anteriormente e que produzisse uma saída correta. É possível calcular um erro do aprendizado supervisionado que seria a diferença da saída gerada, com a saída esperada (MITCHELL, 1997).

2.3 APRENDIZADO PROFUNDO

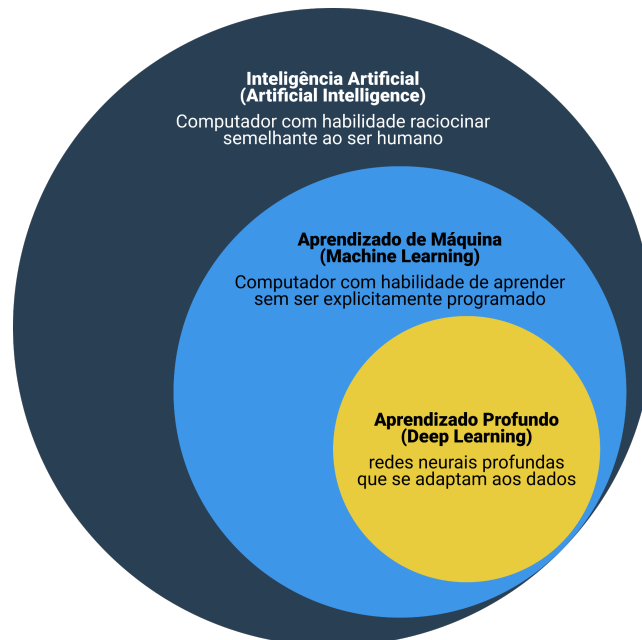
De acordo com Yann LeCunn et. al. (2015) por muito tempo o reconhecimento de padrões dependeu de uma engenharia específica com amplo domínio sobre o ambiente para projetar um extrator de características transformando os dados de entrada em uma representação de vetor de características.

Aprendizagem profunda ou Deep Learning (DL) é uma subárea do Machine Learning (ML), a qual é uma subárea de inteligência artificial como mostra a Figura 5.

Ela aprimora a capacidade de um computador em reconhecer, classificar, detectar. O Deep learning treina o computador que a partir de alguns parâmetros configurados, possa através do reconhecimento de padrões aprender sozinho, utilizando para isso várias camadas de processamento.

É factível que existe uma melhora considerável nos resultados, mas é importante destacar também que os recursos utilizados crescem exponencialmente, à medida que você utiliza uma maior quantidade de camadas, também pela grande quantidade dos dados que geralmente são necessários para realizar o treinamento, portanto, a utilização de GPUs com CUDA, são a resposta para isso.

Figura 5. Divisões na área de Inteligência Artificial.



Fonte: O autor

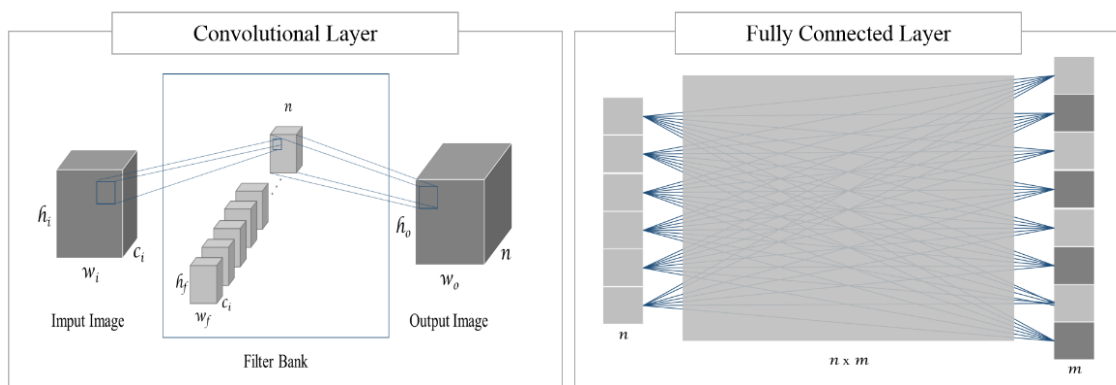
2.4 REDE NEURAL CONVOLUCIONAL

Rede Neural Convolutiva (RNC) também conhecida como ConvNet é uma rede que é utilizada em análises de imagens digitais. De acordo com os trabalhos correlatos, ela tem sido uma das responsáveis para que a Deep Learning tenha sido bem-sucedida no campo da IA. A diferença entre uma rede neural tradicional, para uma rede neural convolutiva já se apresenta na camada de entrada, onde na rede tradicional todos os pixels seriam utilizados para processamento, já uma rede neural

convolucional seleciona automaticamente os pixels que são mais importantes para a representação das características. Os dados na camada de entrada são transformados, uma espécie de pré-processamento para selecionar as principais características. Uma Rede Neural Convolucional é basicamente dividida em dois tipos de redes, uma com camadas de convolução e outra totalmente conectada (REYES et. al. 2015).

Uma camada convolucional, tem como objetivo receber uma imagem como entrada, transformando em uma outra imagem conforme a Figura 6, dando ênfase aos elementos que servem como característica para definir e representar a imagem, já uma camada completamente conectada recebe um vetor como entrada e fornece um outro vetor como saída (REYES et. al. 2015).

Figura 6. CNN x Camada completamente conectada.



Fonte: REYES et. al. 2015

As características de uma CNN que são abordadas nos próximos itens, se referem ao operador de convolução, ao pooling, ao flattening e a camada de rede neural densa.

2.4.1 Operador de Convolução

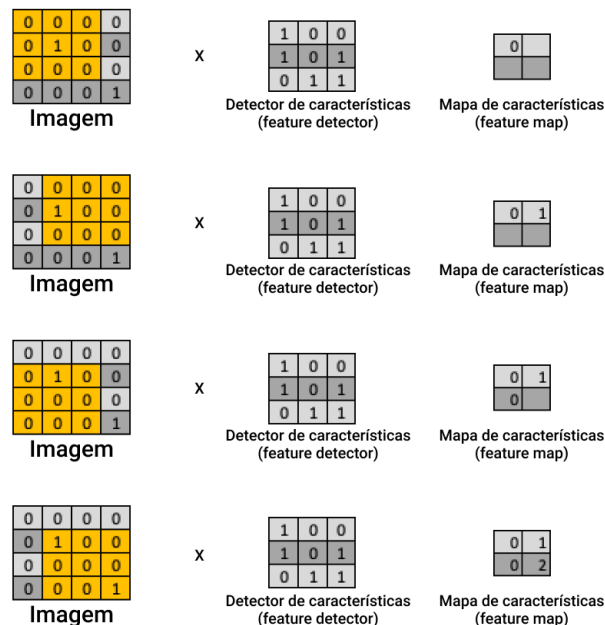
Convolução é adicionar cada pixel da imagem aos seus vizinhos, ponderando por um kernel, uma imagem é na verdade uma matriz de linhas e colunas e é multiplicada por um kernel que também é uma matriz. A convolução por sua vez,

poderia ser sumariamente resumida por um somatório dessa multiplicação de matrizes, onde $f(x,y)$ é a imagem de entrada em forma de função e o $g(x,y)$ é o filtro também chamado de kernel, o qual será utilizado como operador de convolução, gerando outra imagem, representada na Equação 1:

$$(f * g)(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(i,j)g(x-i,y-j) \quad (1)$$

A Equação 1 é constituída por dois somatórios, pois quando nos referimos ao processamento de imagens, é necessário considerar altura e largura. Nesse caso a convolução tem o papel de realizar uma filtragem na imagem, de modo que consiga destacar características importantes para identificação. Durante a aplicação do kernel a matriz vai se deslocando por toda a imagem através do conjunto de pixels e realizando a multiplicação e o somatório dos valores, iniciando em uma extremidade superior e finalizando na outra inferior (DE FARIA 2018). Para ilustrar melhor é possível ver através da Figura 7 o processo de convolução multiplicando a matriz da imagem pela matriz do kernel e percorrendo toda a imagem.

Figura 7. Matriz de entrada multiplicada por detector de características.



Fonte: O autor

Com a aplicação a matriz sofre uma redução de dimensionalidade comparada à matriz original, isso facilitará o processamento e é bem provável que alguma informação seja perdida, no entanto, o propósito principal é selecionar as principais, esse mapa preserva as características mais importantes. Em consonância com Faria (2018) o deslocamento de um pixel por vez é denominado de *stride*, o qual é comumente utilizado em reconhecimento de imagem. Outro parâmetro utilizado é chamado de *wide convolution*, onde a aplicação do filtro é feita nos elementos que não possuem vizinhos, ou seja, nas bordas da matriz. Geralmente é utilizado a técnica de preenchimento também conhecida como *padding*, que expande a matriz de entrada, para o caso da *wide convolution*, o preenchimento é realizado com número zero. O método contrário é quando não se utiliza o preenchimento, que é chamado de *narrow convolution*. Logo após a geração do mapa de características, temos a aplicação da função de ativação.

De acordo com Jarret et al. (2009) e Glorot et al. (2011), a função ReLu é a ativação ideal para se utilizar em redes neurais. Essa função tem por finalidade retirar todos os possíveis valores negativos e substituir por zero, isso acaba trazendo uma atenuação na imagem, facilitando o processamento para identificação das características, do ponto de vista do custo computacional e do ponto de vista matemático. Por envolver processos de substituir os valores negativos por zero e não utilizar, multiplicação, divisão ou exponenciação torna muito mais rápido o processo de treinamento das redes neurais.

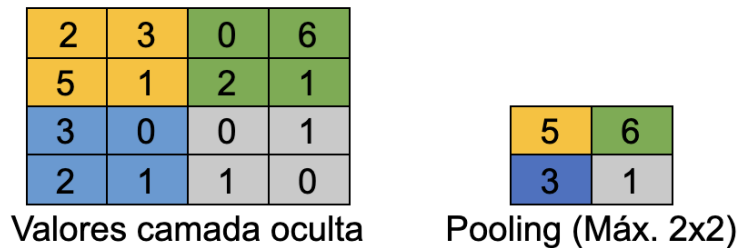
Por fim, a camada convolucional fornece um mapa de características, sendo que os pesos dos filtros são aprendidos durante o treinamento da rede.

2.4.2 Camada de *Pooling*

A operação de *pooling* geralmente ocorre logo após as camadas convolucionais (FARIA, 2018). Através de cálculos estatísticos substitui a saída convolucional por uma determinada região (GOODFELLOW et al., 2016). Dentre os operadores possíveis, citamos o *max* que reporta o valor máximo da região escolhida, o que culmina ainda mais o destaque das características, fora esse operador, ainda podemos citar outros, tais como: a média, a média ponderada da região, ou a norma L2. O *pooling* tem também como objetivo a redução da dimensionalidade. No processamento de imagens o *pooling* não varia quando utilizada a rotação ou

translação, logo quando o operador max é utilizado, sempre fornecerá o valor máximo da região analisada. Na Figura 8 é possível visualizar a operação de *pooling* usando o operador *max*, gerando seu respectivo max pooling.

Figura 8. *Pooling* usando operador Máximo, com filtro 2x2.



Fonte: O autor

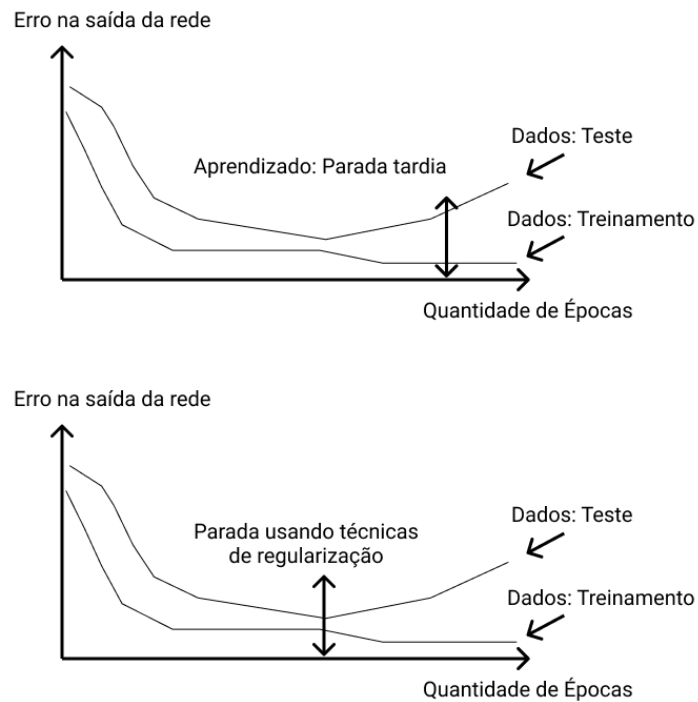
Ainda na Figura 8, a matriz de característica (também chamado de mapa de características) tem como entrada 4 x 4 neurônios, e após a operação gerou uma max pooling de 2 x 2 neurônios, respaldando a redução de dimensionalidade considerável e também a ênfase nas informações mais importantes.

2.4.3 Camada *Dropout*

Um dos problemas conhecidos da CNN é o problema de sobreajuste ou o problema de *overfitting*, que é quando o modelo se ajusta muito aos dados apresentados de treinamento e acaba não generalizando bem. Na figura 9, temos esse problema onde no gráfico superior parou muito tarde, o erro de treinamento ainda vinha reduzindo, mas os erros dos dados de teste começaram a aumentar, quando aplicamos técnicas de regularização, podemos reduzir isso, como é o caso do segundo gráfico da figura, que a partir de técnicas de regularização aplicadas, conseguiu parar o treinamento no melhor momento (FARIA 2018).

Entre as técnicas conhecidas de regularização, destacamos a *dropout* que foi proposta por HINTON et al. (2012) e tem como finalidade de maneira estocástica (estatisticamente aleatória) definir zeros em alguns neurônios de saída, os demais neurônios que não sofreram alteração, são treinados pelo algoritmo *backpropagation*. Neste modelo, os neurônios que são zerados ou desativados, não interferem nos demais, tornando assim o modelo mais robusto.

Figura 9. Problema de sobreajuste (overfitting) em RNA

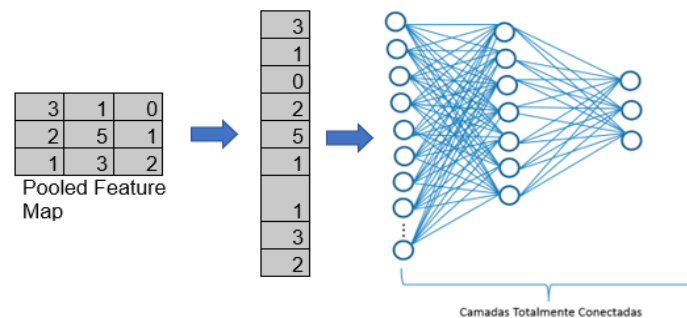


Fonte: Adaptada de FARIA 2018

2.4.4 Achatamento (*Flattening*)

A etapa do *flattening* transforma os mapas de features em um vetor de características para que pudesse dar entrada na rede neural densa. Na Figura 10 é possível visualizar cada elemento da matriz se tornando em um elemento do vetor e posteriormente dando entrada na rede neural densa.

Figura 10. Etapa Flattening - transformando matriz em um vetor.



Fonte: Adaptada de FARIA 2018

2.4.5 Camada Totalmente Conectada

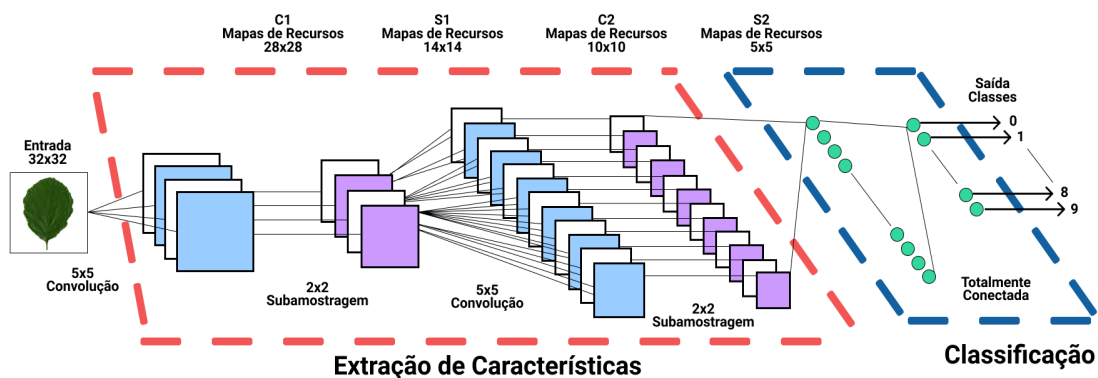
As CNNs podem apresentar uma ou mais camadas completamente conectadas, também chamadas de fully connected, na camada convolucional e na camada de pooling foram gerados mapas de características, nesta etapa todos os neurônios se conectam à camada totalmente conectada, utilizando esses atributos para gerar a classificação, existe ainda uma outra camada, que seria a camada de saída, camada com as classes para a classificação propriamente dita, logo a quantidade de neurônios na camada de saída é correspondente ao número de classes.

Utiliza-se como função de ativação a *softmax*, que é normalmente encontrada em casos em que possui mais de duas classes. A softmax transforma a saída para cada classe para 0 e 1, dividindo também pela soma das saídas, sendo assim o classificador recebe a probabilidade de uma determinada classe. É ideal que se utilize na camada de saída, para fornecer as probabilidades. Essa função é matematicamente descrita como mostrado na Equação 2 (FARIA, 2018).

$$\text{soft max}(y_i) = \frac{e^{z_i}}{\sum_{j \in Y} e^{z_j}} \quad (2)$$

De maneira sumariada e mais lúdica a Figura 11 aborda a CNN como um todo, da entrada, passando pela convolução, as camadas totalmente conectadas, a classificação e a saída para uma melhor visualização e compreensão.

Figura 11. Rede Neural Convolucional.



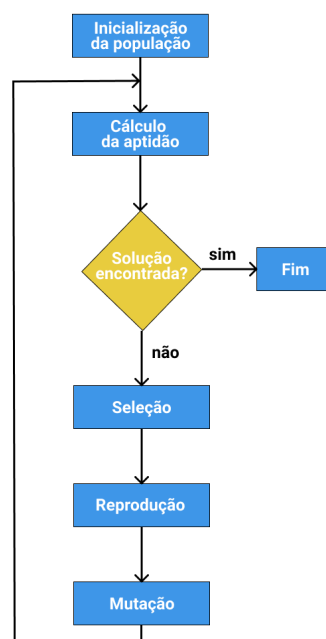
Fonte: O autor

2.5 ALGORITMOS GENÉTICOS

Em 1859 Charles Darwin publicou o seu livro onde denominava-se " A origem das espécies " sendo considerada a base da biologia evolutiva. Sua tese é que as formas de vida evoluem ao longo do tempo através de uma seleção natural, ou seja, indivíduos mais fortes tendem a sobreviver e com uma maior probabilidade de reprodução. (GOTARDE, DALLILO, 2021). De forma semelhante acontece com os algoritmos genéticos - AGs, embora existam uma diversidade de variações algo que eles têm em comum, é o conceito de seleção, mutação e reprodução, aos quais dependem completamente do indivíduo.

Para inicialização, os indivíduos da população são gerados aleatoriamente. Duas formas básicas para a geração de novos indivíduos. O elitismo, considerando que a população foi ordenada com os melhores no começo, alguns desses indivíduos entram para a nova população, uma espécie de clone dos melhores. O cruzamento, a partir da escolha de dois indivíduos e combinação dos cromossomos de cada, um novo indivíduo é obtido. A mutação é o clone modificado, copiando o indivíduo anterior e alterando um pouco o cromossomo (MASSAGO, 2013). A Figura 12 exemplifica uma estrutura básica de um algoritmo genético.

Figura 12. Estrutura básica - Algoritmo genético



2.6 CONSIDERAÇÕES FINAIS

Neste capítulo foram observados os principais conceitos que envolvem o aprendizado de máquina, o funcionamento de uma rede neural convolucional com todos os atributos que compõem e cálculos matemáticos que envolvem o processo de convolução, seguido dos algoritmos genéticos que com inspiração na seleção natural das espécies, auxiliará a escolha dos classificadores para a melhor solução.

3 ESTADO DA ARTE

Na literatura encontramos diversas referências sobre o reconhecimento de padrões e posteriormente sobre as maneiras como as classificações são realizadas, como por exemplo podemos citar um das pesquisas pioneiras em reconhecimento de padrões, realizadas por Ojala et. al. (1996) onde cada pixel C tem um conjunto de vizinhos P , a diferença de intensidade de textura de C e os seus vizinhos definem um histograma, esse é um estudo de caso bem antigo, e embora sirva como sustentação para todas as pesquisas posteriores a essa, tentaremos nos rodear de pesquisas mais recentes.

Em 2017 foi criado um sistema múltiplo de classificação para o reconhecimento de folhas de plantas. Nele é utilizado um conjunto de classificadores de base SVM (máquina de vetor de suporte) e Rede Neural, e através da combinação do conjunto de recursos, entre eles, HOG que é o histograma de gradientes, LBP Padrão binário local, SURF que é recursos robustos acelerados e momentos de Zernike ZM. Com a afirmação de que classificadores com base em diferentes conceitos e treinado em características que correspondam a textura e formato das folhas podem gerar uma abordagem mais robusta. Com isso foi possível avaliar a combinação de textura e forma para diferenciar espécies de vegetais. Aqui a base de imagens utilizada foi retirada do ImageCLEF 2011 e ImageCLEF 2012. Inicialmente as imagens passam por um pré-processamento, retirando estruturas que não farão parte, como o fundo e caule da planta após isso elas são enviadas para a extração de recursos obtendo descritores para cada imagem, o descritor gera um vetor de características que são usadas para treinar um conjunto de imagens. Para manter a heterogeneidade foram utilizados dois algoritmos de aprendizado, a máquina de vetor de suporte e a rede neural. Algo que pode ser observado é que os classificadores treinados em forma e descritores de textura cometem erros diferentes, portanto, quando os classificadores foram combinados o resultado foi uma melhora em termos de precisão em aproximadamente 28% em relação ao classificador único, corroborando para a proposta inicial que era a combinação de múltiplos classificadores treinados com características de forma e textura para identificação da planta, limitado aqui apenas para as imagens da folha (ARAÚJO et. al. 2017).

Segundo ARAÚJO et. al (2018) propôs uma granulação fina de dois níveis de classificação da folha em que as representações são combinadas em cada nível

usando modelos profundos, trazendo como novidade fusão de informações locais e globais, foram utilizadas CNNs pré-treinadas, após a entrada da imagem ela passa por uma CNN que corresponde a toda imagem e outra CNN que vai pegar partes da imagem, após isso é feita uma fusão que determina o gênero da planta, caso aquela planta possua só uma espécie já é possível determinar também a espécie, senão ela passa por uma outra CNN de nível mais alto, após a fusão determinando então a espécie da planta, ou seja, ela trabalha de maneira hierárquica. Para a entrada de imagens são usados alguns métodos de pré-processamento e de aumento de dados, onde o objetivo é tirar possíveis ruídos da imagem, como haste, caule, tronco e para o aumento de dados a partir da imagem original cópias são rotacionadas no sentido horário, como citado além da imagem inteira, também são usadas partes das imagens como as veias presentes nas folhas, em cada nível hierárquico tanto para gênero quanto para espécie foram utilizadas duas CNNs pré-treinadas do GoogleNet do conjunto de dados do ImageNet, as CNNs foram adaptadas. As imagens para o conjunto de dados foram extraídas do ImageCLEF 2015 que apresenta uma diversidade de fotografos que realizaram a captação o que aumenta o grau de dificuldade pela variância de ambientes e pelo desequilíbrio de algumas classes que apresentavam poucas imagens. O protocolo utilizado foi de 70% para treinamento e 30% validação, o objetivo de reconhecimento de plantas é identificado como recuperação, portanto, ao invés de um número como precisão geral é utilizada uma pontuação média. Logo os resultados foram um aumento de aproximadamente 2% em relação ao melhor resultado presente no estado da arte. O próprio autor atribui que são possíveis mais pesquisas para melhorar o resultado, como utilizar redes siamesas, complementando também a possibilidade de utilizar outras partes da imagem da planta.

Lee et. al. (2018) apresentou algumas contribuições para a classificação de imagens com múltiplas vistas. Primeiro uma nova CNN chamada rede neural convolucional híbrida de órgão genérico que especificamente extrai informações sobre órgãos e classifica com base na correlação entre o que foi escolhido e características genéricas. A CNN híbrida trabalha com quatro camadas, a saber: camada compartilhada, camada de órgão, genérico e camada de espécie. As imagens foram redimensionadas para três escalas diferentes, tendo assim, uma maior robustez. A segunda proposta foi uma RNN uma rede neural recorrente, PlantStructNet, que em suma leva um número variável de vistas das plantas das imagens composta por um

ou vários órgãos e otimiza as dependências entre eles para classificação de espécies. As conclusões foram positivas em relação ao que se existe relatado no estado da arte, algo que poderia ser implementado em uma pesquisa futura é a integração da CNN híbrida com a RNN.

Com o avanço dessas tecnologias de reconhecimento e aprendizado profundo, Goeau et. al. (2018) questionaram a eficácia e confiabilidade de sistemas automáticos trabalhando com uma grande quantidade de dados, e em contrapartida um número reduzido de dados nas mãos de especialistas em plantas, pois, aquilo que está presente em uma imagem nem sempre servirá para responder prontamente a que gênero, espécie ela pertence. Para isso, um estudo experimental foi desenvolvido, sendo 9 sistemas de aprendizagem profunda implementados por 3 equipes de pesquisa diferentes e do outro lado, 9 botânicos especialistas da flora francesa. O resultado foi que os sistemas automáticos de aprendizagem profunda estão próximos da experiência humana, sendo maduros o suficiente para várias tarefas que exijam uma mão de obra especializada e bem promissores até para sistemas de vigilância ecológica.

Nesta referência o estudo de caso foi a análise de imagem para aprimorar a classificação de coberturas de terra, usando redes totalmente convolucionais e imagens com vistas múltiplas, através de um sistema aéreo não tripulado. Segundo Liu et. al. (2018) a área de estudo está localizada no sul da Flórida, após a captação de imagem realizada foi gerado o ortomosaico através do software Agisoft PhotoScan, o teste tinha como objetivo comparar o desempenho de uma rede completamente convolucional em comparação com Random Forest e o SVM – Máquina de vetor de suporte já presentes no estado da arte e podendo ser comparado pois utilizavam a mesma metodologia. A FCN obteve um resultado melhor quando comparado a RF, SVM ou DCNN, no entanto, é importante ressaltar que o tempo para treinamento foi muito maior em relação às demais, ou seja, a utilização é viável, mas acompanhada da ressalva de que o custo computacional será maior, para um resultado melhor.

Em meados de 2019 foi proposto um método para uma classificação refinada de madeira usando redes neurais convolucionais, embora a madeira não seja exclusiva deste projeto de pesquisa, ela se encaixa em nosso escopo, pois o objetivo aqui é tratar a planta com mais de uma visualização, não limitando apenas a folha, e sim podendo usar o caule e / ou tronco, e foi selecionado pelo fato da estratégia que foi utilizada após o uso das CNN (SHUSTROV et. al. 2019).

Shustrov e sua equipe (2019), inicialmente separaram a imagem do plano de fundo usando para fazer isso a binarização como limiar, após isso foi utilizado regra de decisão, para isso cada classificador de partes das imagens produz uma classe com probabilidade para a classe que precisa ser combinada para obter a identificação final das espécies. Como regras de decisão foram utilizadas as seguintes: Regra de votação por maioria, regra de probabilidade média e regra máxima. As arquiteturas de CNNs que foram comparadas: AlexNet, VGG-16, GoogleNet e ResNet. Conseguindo a maior precisão com a estrutura GoogleNet que classificou 94,7% de maneira precisa, pelas três regras que foram estabelecidas, não foi encontrado grandes diferenças, ao se usar qualquer uma das três propostas aliado ao GoogleNet, uma precisão quase perfeita de 99,4%.

Pode-se concluir que utilizando apenas 25 partes das imagens é possível obter uma taxa de acerto de 99% de identificação, sendo viável a utilização para identificar placas em tempo real em uma serraria. A base de imagens se concentrava em apenas madeira, o que torna um pouco simplista a maneira com que os classificadores precisam trabalhar, no entanto, a diferença das outras literaturas é que utilizaram regras de decisão, e quando a combinação das regras com a CNN o resultado era muito promissor, está aqui uma opção viável que pode ser desenvolvida neste projeto, adicionado claro, diferentes vistas de uma planta.

Seeland et. al. (2019) propôs o desempenho de uma rede neural convolucional que foi treinada em 1000 espécies pertencendo a 516 gêneros e 124 famílias. A base de plantas utilizada, continha imagens de diferentes escalas e conteúdos, como: flores, frutas, caule, casca, o que torna esse objeto de pesquisa muito próximo do que queremos analisar neste projeto. As imagens não foram processadas, e no primeiro momento foi investigado se o classificador se confunde ao tratar de uma variabilidade visual, no segundo momento de conjunto de experimentos, se as características visuais de gênero e família são absorvidas de modo que mesmo não participando do conjunto de treinamento, possam ser identificadas. Inicialmente entre as 1000 espécies, a CNN classificou de maneira assertiva 82,2% no top 1 e 92,9% no top 5 que significa que entre cinco opções disponíveis, o classificador escolheu de maneira correta.

Na segunda proposta para saber se as imagens podem ser identificadas mesmo sendo excluídas do treinamento, o resultado geral é reduzido em mais da metade em gênero (55,4% relativo) e família (56,7% relativo), com algumas exceções

em determinadas classes específicas onde encontrou um resultado melhor, provavelmente pela similaridade das plantas. Foi utilizado o conjunto de imagens PlantCLEF 2016.

Em 2019 um grupo de pesquisadores tomou por despertamento o fato de utilizar mais de uma perspectiva para a correta classificação de imagens de plantas. Dentro os questionamentos colocados pelos próprios autores, estão as seguintes: Quais são as mais e menos importantes perspectivas em relação a precisão de previsão? Que ganho pode ser alcançado pela combinação de perspectivas? Como as precisões diferem entre CNNs treinadas por perspectiva de imagem em contraste com uma CNN única treinada em todas as imagens? A especificidade de uma perspectiva ou uma combinação de perspectivas é universal ou depende das espécies? Qual a sensibilidade dos resultados de identificação para o número de imagens de treinamento utilizadas? Para responder essas questões uma rede neural convolucional foi treinada para cada perspectiva para explorar informações de imagens de diferentes órgãos. A aquisição de imagens foi realizada pelo aplicativo gratuito Flora Capture, tendo como referência 5 imagens de cada planta, sendo uma imagem da planta inteira, uma imagem frontal, terceira uma perspectiva lateral da flor, então uma imagem superior e por fim uma imagem inferior, foram trabalhados com 101 espécies e utilizado como arquitetura o Inception-ResNet-V2 e uma rede que foi pré-treinada no ImageNet em larga escala. Foram realizadas duas abordagens, uma onde foi treinado um classificador para cada uma das cinco perspectivas e outra um classificador treinado em todas as imagens, independente da sua perspectiva.

Como resultado principal pode destacar-se que a fusão de todas as cinco perspectivas foi a que trouxe o melhor resultado, alcançando 97,1% de precisão. Foi utilizado 80 imagens para cada espécie, ao se experimentar uma redução para 60 e 40 imagens, não teve um efeito negativo consistente em nenhuma perspectiva, no entanto, quando se reduziu para 20 a precisão caiu muito de acordo Rzanny et. al. (2019). O Quadro 1 resume de maneira mais visual, os artigos e suas aplicações.

Quadro 1 - Comparação de resultados de levantamento bibliográfico.

Referência	Aplicação	Features	Classificadores	Base	Score
Voncarlos (2017)	Folha da planta	HOG, LBP, SURF e Zernique	SVM + Rede Neural	ImageCLEF 2011 ImageCLEF 2012	0.81 0.66
Voncarlos (2018)	Folha da planta	CNN pré-treinadas	Fusão de CNNs Fine tuning	ImageCLEF 2015	0.86
Lee (2018)	Múltipla vista da planta	HGO CNN híbrida	Rede Neural Recorrente	LifeCLEF 2015	0.73
Liu (2018)	Coberturas de terra	FCN	RF, SVM, DCNN e FCN	Captação veículo aéreo não tripulado	0.87
Shustrov (2019)	Madeira	CNN pré-treinada	AlexNet, VGG-16, GoogLeNet e ResNet	Imagens de maquinários de serrarias	0.99
Seeland (2019)	Flor, fruta, caule e casca	CNN pré-treinada	Inception-ResNet-V2	PlantCLEF 2016	0.72
Rzanny (2019)	Planta inteira, flor, folha	CNN pré-treinada	Inception-ResNet-V2	Aplicativo Flora Capture	0.97

Fonte: O autor

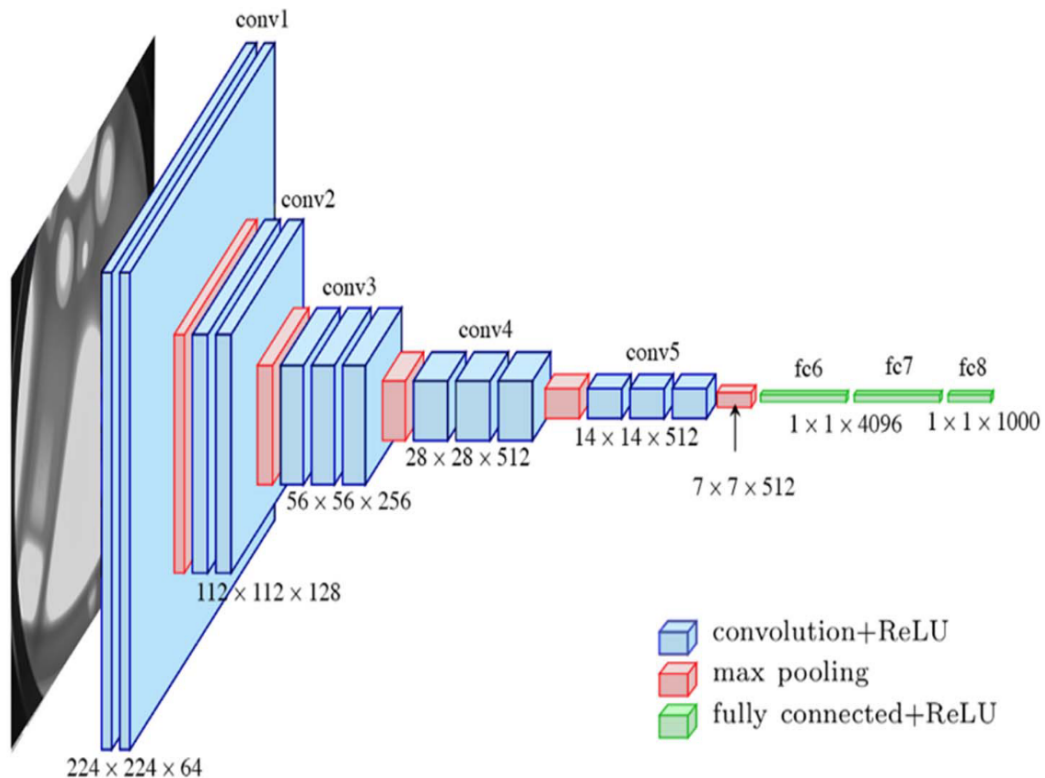
3.1 CNN PARA CLASSIFICAÇÃO DE IMAGENS

Seguem algumas redes neurais convolucionais que são amplamente utilizadas para a tarefa de classificação e que serão utilizadas na metodologia do trabalho.

3.1.1 Visual Geometry Group (VGG-16)

Simonyan e Zisserman (2015) investigaram o desafio sobre o reconhecimento de imagens em larga escala sob a ótica das redes neurais convolucionais. Eles foram os vencedores do ImageNet Challenge de 2014, garantindo o primeiro e o segundo lugar na categoria de localização e classificação, respectivamente. Em suma, a VGG-16 possui esse nome, pois ela trabalha com 16 camadas convolucionais, utilizando filtro 3×3 , dentre elas as 13 primeiras camadas trabalham na extração das *features* e as três últimas são camadas completamente conectadas que atuam na classificação, nem todas as camadas possuem o *max pooling*, ao todo são 5 camadas de pooling que a cada uma reduzem pela metade o tamanho da imagem, todas as camadas ocultas são equipadas com ReLU e a última camada é a softmax que possui um saída para cada classe.

Figura 13. Arquitetura padrão da rede VGG-16.

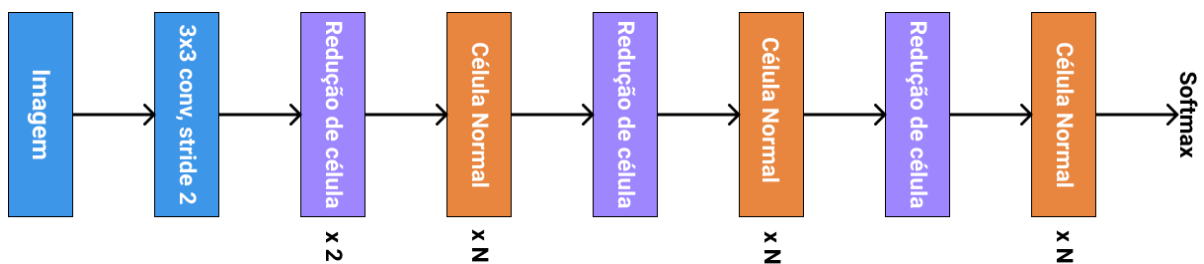


Fonte: Ferguson et al. 2017

3.1.2 NasNetMobile

Nasnet é uma CNN que dispõe de uma arquitetura escalável, operando através de aprendizagem por reforço e possuindo 5,3 milhões de parâmetros, divididos em 12 células e totalizando 769 camadas (ZOPH et al. 2018 e SAXEN et al. 2019). Esta CNN alcançou 74,4% no top-1 e 91,9% top-5 no ImageNet. Através da Figura 14 é possível visualizar a arquitetura básica da CNN NasNetMobile.

Figura 14. Arquitetura padrão da rede NasNetMobile.



Fonte: ZOPH et al. 2018

3.1.3 EfficientNet

Para a CNN EfficientNet foram analisadas o processamento de duas versões, a B0 e a B7. A CNN base B0 alcançou um desempenho de 77.1% no top-1 e 93.3% no top-5 de performance no ImageNet com 5,3 milhões de parâmetros e 238 camadas. A versão B7 com 66 milhões de parâmetros e 814 camadas em sua arquitetura atingiu um patamar maior com resultados de 84.3% e 97.0% de performance para top-1 e top-5, respectivamente (TAN e LE 2020). Os autores compreenderam que a CNNs devem ser ampliadas não apenas em profundidade, como normalmente uma rede profunda é caracterizada por possuir mais camadas, mas sim utilizando a profundidade, largura e resolução de maneira conjunta. A arquitetura base da EfficientNet B0 pode ser visualizada através da Figura 15.

Figura 15. Arquitetura padrão da rede EfficientNet B0.

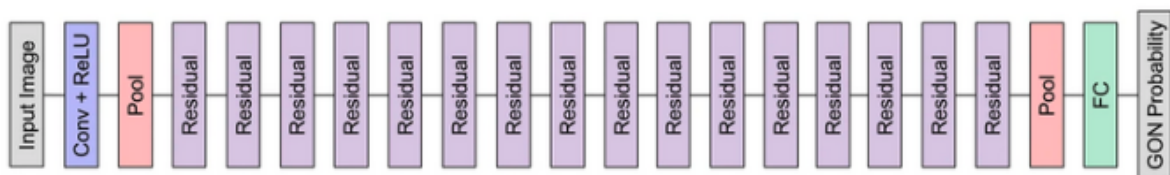
Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBCConv1, k3x3	112×112	16	1
3	MBCConv6, k3x3	112×112	24	2
4	MBCConv6, k5x5	56×56	40	2
5	MBCConv6, k3x3	28×28	80	3
6	MBCConv6, k5x5	14×14	112	3
7	MBCConv6, k5x5	14×14	192	4
8	MBCConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Fonte: Tan e Le 2020

3.1.4 Resnet

A CNN ResNet50 com mais de 25 milhões de parâmetros e 50 camadas profundas, alcançando 74,9% no top-1 e 92,1% no top-5 no desafio do ImageNet, conquistando o primeiro lugar no ano de 2015, tendo sido desenvolvida por HE et al. 2015. A versão ResNet152V2 alcançou 78% no top-1 e 94,2% no top-5, possuindo mais de 60 milhões de parâmetros divididos em 152 camadas, a ResNet50 pode ser visualizada através da Figura 16.

Figura 16. Arquitetura padrão da rede ResNet50.



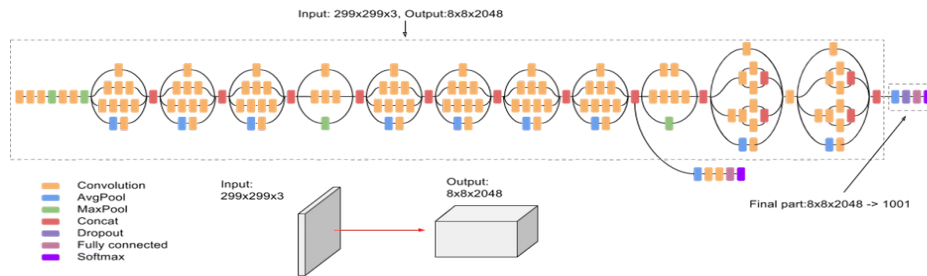
Fonte: Christopher et al. 2018

3.1.5 Inception

Com aproximadamente 24 milhões de parâmetros, a InceptionV3 obteve um desempenho de 77,9% no top-1 e 93,7% no top-5 no ImageNet. Um dos objetivos da Inception é a redução do custo computacional sem perder o desempenho, para isso, as técnicas utilizadas para otimização são convoluções fatoradas, redução de

dimensão, cálculos paralelizados e regularização que é uma pequena CNN inserida entre as camadas durante o treinamento onde a perda incorrida é adicionada à rede principal. Na Figura 17 é possível visualizar a arquitetura da rede InceptionV3.

Figura 17. Arquitetura padrão da rede InceptionV3.

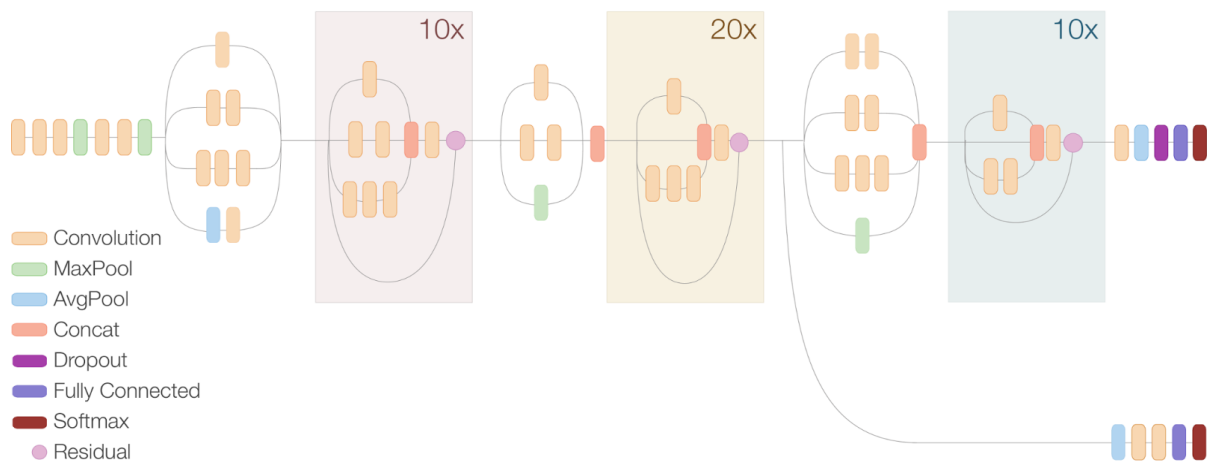


Fonte: Ahn et al. 2018

3.1.6 InceptionResnet

A InceptionResnetV2 possui aproximadamente 56 milhões de parâmetros, alcançando 80% no top-1 do ImageNet e 95% no top-5. A rede Inception que tinha como primazia a redução do custo computacional, começa agora a receber conexões residuais que, de acordo com Szegedy et al. 2016 essa inserção possibilita a aceleração do treinamento significativamente. A partir da Figura 18 é possível visualizar a arquitetura da InceptionResnetV2.

Figura 18. Arquitetura padrão da rede InceptionResnetV2.

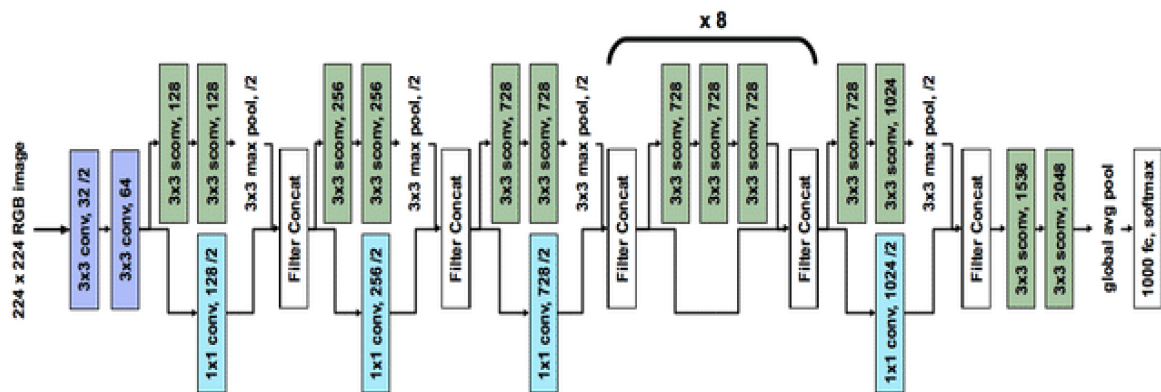


Fonte: Mehta et al. 2018

3.1.7 Xception

Contando com aproximadamente 23 milhões de parâmetros e obtendo um desempenho no ImageNet de 79% e 94,5% no top-1 e top-5, respectivamente. A rede contempla 36 camadas divididas em 14 módulos, sendo sumariamente uma pilha linear de camadas de convolução. A Figura 19 demonstra a arquitetura padrão da Xception.

Figura 19. Arquitetura padrão da rede Xception



Fonte: Leonardo et al. 2018

3.2 CONSIDERAÇÕES FINAIS

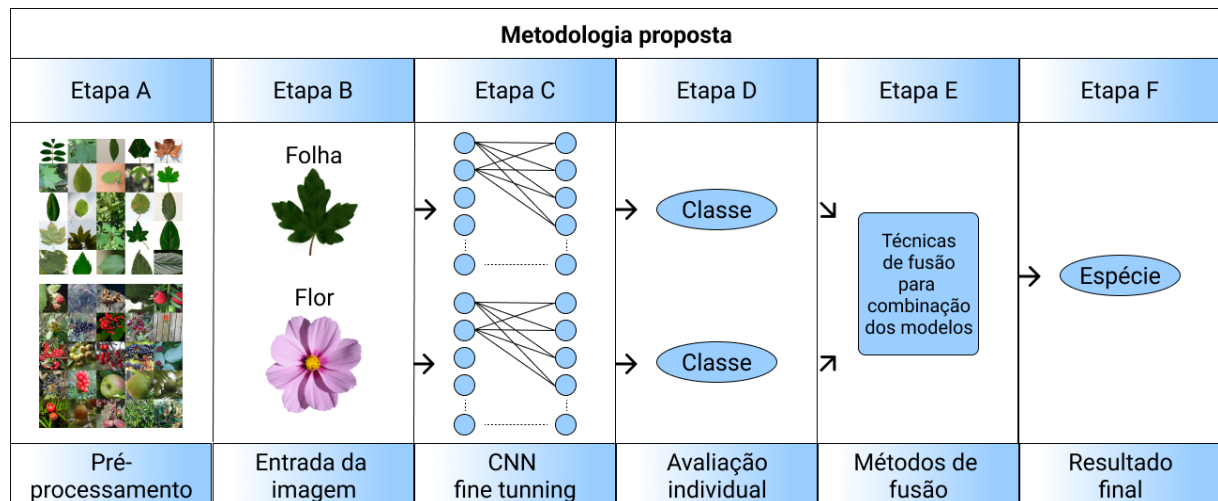
A partir dessa revisão de literatura, conclui-se que é necessário a pesquisa que será realizada para o crescimento e engajamento das ferramentas de classificação de vegetais, tanto para o objetivo de pulverizar a agilizar a maneira como o reconhecimento e classificação é realizado hoje, quanto do ponto de vista do custo computacional, onde se busca melhores resultados com um menor custo, e o fato de utilizar vistas diferentes da imagem da planta, que embora apareça em algumas literaturas, ainda existem lacunas que podem ser exploradas, trazendo essa como uma abordagem atualizada para a problemática da identificação de plantas. A Tabela 1 contemplou resultados alcançados pelo diferentes experimentos, destaca-se aqueles que utilizaram a base de dados do LifeCLEF pois se assemelha muito ao mundo real, contendo diferentes tipos de espécies, dos mais variados autores que não necessariamente utilizaram o mesmo protocolo na captação, embora todo trabalho tenha uma contribuição, os que utilizam uma base de dados do estado da arte trazem

uma probabilidade maior de generalização do modelo, pelo fato de pôr à prova, quando utilizado uma base de imagens própria existe uma sensibilidade para generalizar os modelos para outros cenários. A grande quantidade de redes neurais que já foram treinadas no Imagenet e citadas neste capítulo, serão utilizadas em nossa metodologia e nos auxiliará na redução do tempo de treinamento.

4 MATERIAIS E MÉTODOS

A Figura 20 demonstra brevemente o funcionamento da metodologia abordada neste projeto de pesquisa. No primeiro módulo denominado ETAPA A, os dados são pré-processados, subitamente submetidos à entrada nos modelos de CNN pré-treinada com fine-tuning de maneira individual, na Etapa D uma avaliação individual de cada modelo, após isso, na Etapa E são avaliadas técnicas para fusão dos modelos, em seguida teremos o resultado final com a avaliação final e a predição da espécie.

Figura 20. Método de classificação proposto.



Fonte: O autor

4.1 SISTEMA COMPUTACIONAL

Os algoritmos desenvolvidos para a classificação de espécies de plantas através da folha e flor, foram implementados na linguagem de programação orientada a objetos Python, que se encontra na versão 3.7.3.

O software adotado para que o algoritmo possa rodar na linguagem Python, foi o PyCharm CE na versão para uso da comunidade, a qual não incide nenhum custo e onde já existia um prévio conhecimento, a máquina utilizada para rodar foi um macbook versão 2012 com processador i7, sem placa de vídeo disponível. Pela grande quantidade de dados a serem processados, foi necessário um hardware mais

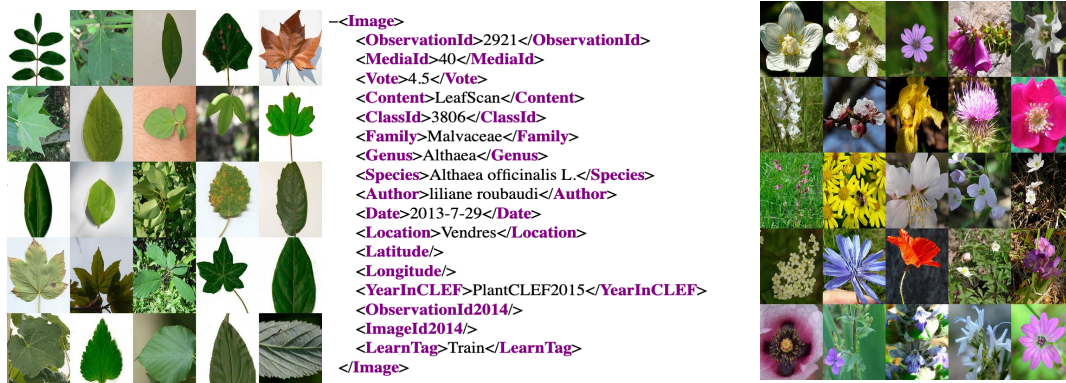
robusto, para isso foi utilizado o Google Colab em sua versão Pro, que possibilita o uso de GPU's e TPU's para aprendizagem de máquina na nuvem.

4.2 PRÉ-PROCESSAMENTO - ETAPA A

Neste trabalho utiliza-se a base de imagens pública LifeCLEF 2015, essa é uma base comumente conhecida no meio acadêmico e servirá para os fins da pesquisa do projeto. A base completa do LifeCLEF 2015 é composta de 113.205 fotos pertencendo a 41.794 observações de 1000 espécies de plantas, dados esses que foram coletados por 8.960 colaboradores diferentes, sendo que cada imagem pertence a apenas uma visualização dos 7 tipos presentes relatados nos metadados (planta inteira, fruto, folha, flor, caule, galho, folhas escaneadas). A forma de captação do LifeCLEF se aproxima muito do mundo real, pois as fotos são tiradas por diferentes usuários que não necessariamente utilizam o mesmo protocolo de aquisição e são tiradas em épocas diferentes do ano. A base é dividida em treino e teste, e elas foram separadas por amostragem aleatória (IMAGECLEF, 2015).

Cada imagem está acompanhada de um arquivo xml, com os seus respectivos metadados, informações como autor, o gênero, a espécie, entre outros. Para esta pesquisa, limitaremos o uso da base de dados às imagens pertencentes à folha e à flor da planta, de modo que serão utilizadas as espécies onde exista a intersecção entre as duas vistas da planta, para fins de combinação, logo serão utilizadas 43 espécies de cada. Seguem abaixo, exemplo de xml e de imagens retiradas da base LifeCLEF 2015.

Figura 21. Flores, XML metadados e Folhas.

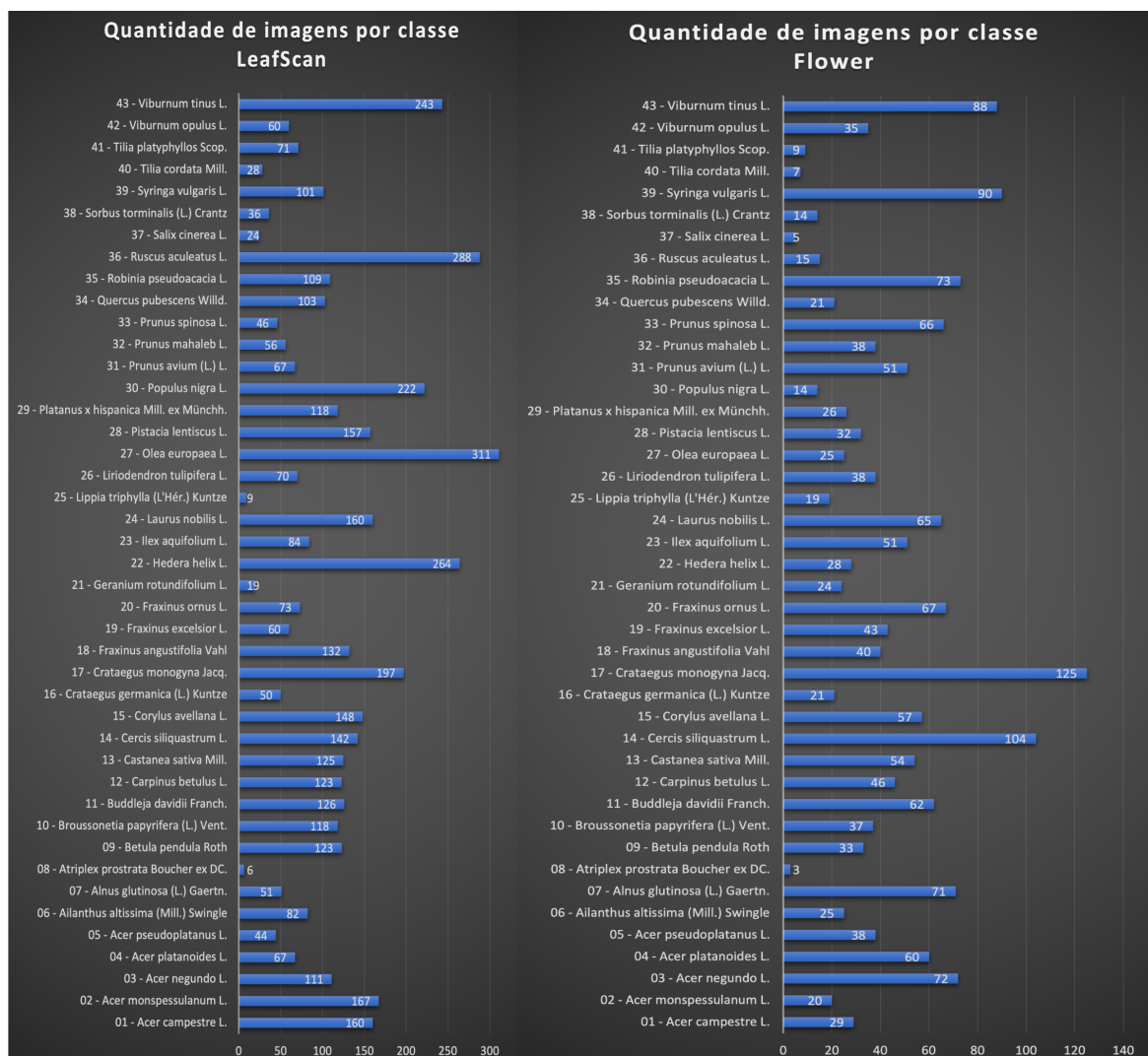


Fonte: O autor

4.2.1 Preparo das Imagens

A base de dados é desbalanceada, isso significa que as espécies não possuem as mesmas quantidades de representações, como exemplo podemos citar a espécie *Acer Campestre* L, que possui 160 imagens na sua base de treinamento, já a *Anemone Hepatica* L possui 20 imagens, por fim e respaldando o desbalanceamento citamos *Alnus Alnobetula* (Ehrh.) K.Koch com apenas 4 imagens. Na Figura 22 é possível visualizar a distribuição de imagens por classe, aferindo assim o desbalanceamento. Cada espécie foi referenciada por um número iniciando em 01 e indo até 43, esses números servirão para visualizar a matriz de confusão.

Figura 22. Imagens por classe - *LeafScan* e *Flower*

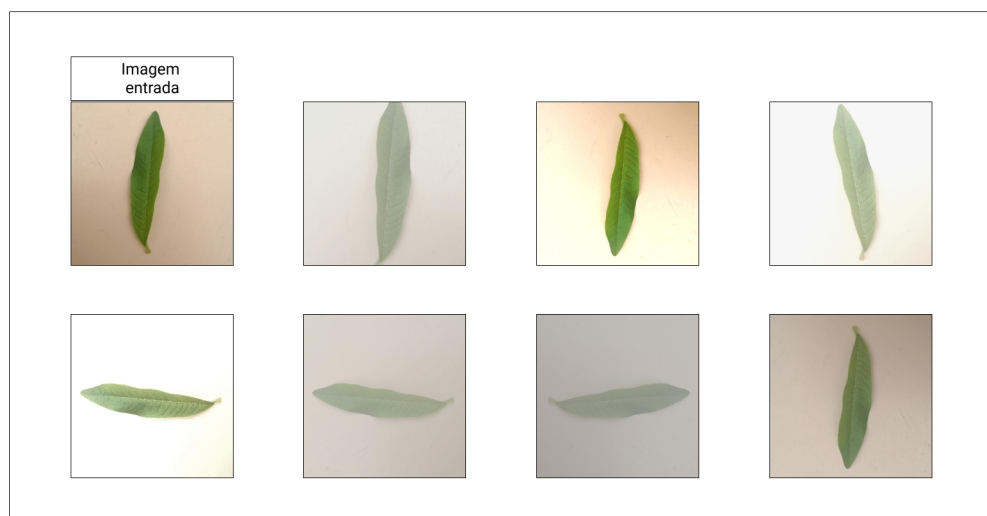


Fonte: O autor

4.2.2 Aumento dos Dados

Para que possamos balancear a quantidade de amostras por espécie, foi aplicado um aumento de dados do inglês *data augmentation*. De acordo com Ghazi et al. (2017) redes neurais profundas necessitam de uma quantidade grande de dados para que auxiliem a reduzir o *overfitting* que é uma situação onde o modelo tem um resultado muito satisfatório com a base de treinamento, no entanto, quando recebe o conjunto da base de teste o desempenho não é muito bom, fazendo com que o modelo não tenha uma boa capacidade de generalização. Será utilizado o pacote *Augmentor* (BLOICE et al. 2017) que é um pacote em Python para expandir e gerar dados de maneira artificial que possam ser utilizados nas tarefas de aprendizado de máquina. Para que possamos exponencializar o ganho que o aumento de dados traz para o processamento de uma rede neural, serão estabelecidas as seguintes operações: *Rotate* com pequenas variações de graus, *rotate* com 90, 180 e 270°, o *flip_top_bottom* e *flip_left_right*. *Perspective Skewing* é perspectiva enviesada que envolve a transformação de modo que pareça estar observando a imagem de um ângulo diferente, operando na inclinação dos cantos e também da direita para a esquerda. Ainda foram utilizadas variações randômicas de brilho, contraste, cor, distorção e ruído, deixando cada classe com 200 imagens. A Figura 23 mostra as imagens após algumas operações de aumento de dados.

Figura 23. Aumento de dados utilizando *Augmentor*.



Fonte: O autor

A base de treinamento utilizada foi dividida em treino e validação, sendo respectivamente 80% e 20%, sendo assim, as imagens da folha da planta foram divididas e ficaram com 6880 imagens para treinamento e 1720 imagens para a validação, as flores 6880 imagens para treinamento e 1720 imagens para a validação do modelo. Na Tabela 2 é possível visualizar melhor como a divisão foi feita. Como a intenção é avaliar a fusão dos modelos, a parcela que será utilizada é aquela onde as classes se encontram em ambos os componentes, sendo esta divisão, um protocolo padrão experimental.

Quadro 2. Divisão da base, treino, validação e teste

Elemento	Imagens Base	Espécies treino	Treino	Validação	Imagens teste	Espécies teste
LEAFSCAN	8600	43	6880	1720	89	43
FLOWER	8600	43	6880	1720	89	43

Fonte: O autor

4.3 DEFINIÇÃO DOS MODELOS PROFUNDOS - ETAPA B

Alguns métodos de modelos profundos foram avaliados, uma vertente possível é a construção do zero de um modelo de aprendizado profundo. Outra opção é através de CNNs já pré-treinadas no ImageNet.

ImageNet é um grande banco de dados que foi projetado para uso e pesquisa de software de reconhecimento de objetos visuais, essa base que possui mais de 14 milhões de imagens que foram adicionadas manualmente, contendo mais de 20000 categorias, o objetivo do ImageNet é conseguir em média 1000 imagens para cada categoria. O objetivo de usar uma rede pré-treinada é a possibilidade de uma validação de acurácia em um tempo menor, pelo fato da rede já iniciar com determinados pesos (IMAGENET 2021).

Para este trabalho em específico, utilizamos rede neural já pré-treinada para que tende a reduzir o custo computacional para treinamento dos modelos e aumento de acurácia. Utilizamos as seguintes arquiteturas, as quais foram amplamente detalhadas na seção anterior, sendo elas: VGG16, NasNetMobile, EfficientNetB0,

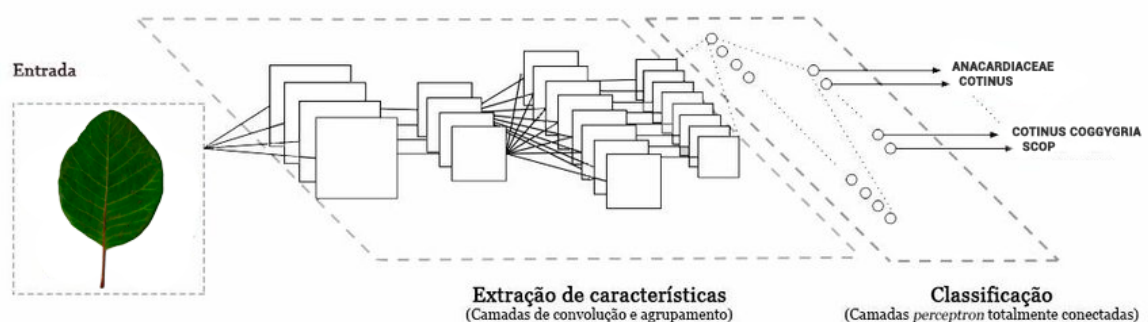
EfficienteNetB7, ResNet50, ResNet152, Xception, InceptionV3, InceptionResnetV2, todas as redes selecionadas alcançaram um ótimo desempenho no ImageNet e serão exploradas a seguir e sua arquitetura foi apresentada no estado da arte.

4.4 TREINAMENTO DOS MODELOS

Embora seja possível utilizar uma rede pré-treinada, ela deverá ser ajustada, o que é conhecido na literatura como afinação, do inglês *fine tuning*, para que seja capaz de se adaptar à realidade do trabalho proposto, às classes, e à atual base. É importante destacar que embora a rede seja pré-treinada ela não necessariamente possui um módulo para classificação de espécies de plantas, utilizando multicomponentes, como no presente caso que será a folha e a flor, então ela necessita de refinamentos.

Na Figura 24 mostra como é o funcionamento de uma CNN genérica, tendo a parte inicial como extração de características e a final como classificação, as camadas totalmente conectadas.

Figura 24. Fluxo de uma CNN.



Fonte: Adaptada Lima, et al. 2017

Durante o treinamento também foram utilizados alguns recursos aos quais foram expostos na revisão da literatura e foram utilizados no modelo apresentado. Para aprimoramento da rede foi utilizado como otimizador o SGD descida do gradiente estocástica que basicamente atualiza um conjunto de parâmetros com a finalidade de reduzir a função de erro (GHAZI et al. 2017), o SGD leva vantagem em relação ao GD descida do gradiente, pois nessa segunda o custo é muito mais alto, pois quando os

pesos são atualizados, precisam ser aplicados para todo conjunto de dados, logo quanto maior a base, maior será o tempo de processamento. O SGD atualiza o peso através de lotes, favorecendo quando aplicado a grandes quantidades de dados, o que normalmente é o caso quando falamos de machine learning. Quando utilizamos o SGD também podemos configurar alguns parâmetros que tendem a ajudar no treinamento do modelo e afetam o resultado.

A taxa de aprendizagem do inglês *Learning rate* seja talvez um dos hiperparâmetros mais importantes, ele controla a velocidade com o que o modelo se adapta ao problema, utilizamos em nosso modelo a taxa de $1e-3$. Outro hiperparâmetro que podemos configurar é o *Decay* que ajuda a reduzir o sobreajuste ou *overfitting* e também melhora o desempenho para novos dados. Momentum é outro parâmetro dentro da SGD que foi ajustado, ele influencia na quantidade que as mudanças realizadas no peso anterior afetam a atual. O Quadro 3 demonstra algumas variações de hiperparâmetros que foram realizadas, a fim de encontrar a melhor configuração para o modelo em questão, sendo testado na rede InceptionResnetV2 a qual apresentou o melhor desempenho, ao fim de 100 épocas.

Quadro 3. Configuração de Hiperparâmetros - Rede InceptionResnetV2

Learning Rate	Decay	Momentum	Batch_size	Acurácia (LeafScan)
1e-3	1e-4	0.9	20	0.69
1e-2	1e-4	0.5-0.8	30-60	0.59
1e-4	1e-4	0.5-0.8	30-60	0.56

Fonte: O autor

4.5 AVALIAÇÃO INDIVIDUAL DOS MODELOS - ETAPA D

Cada item é processado por uma CNN e então tem como retorno a classe a qual ela pertence, nesse instante é avaliado como cada componente individual está se comportando e qual foi o resultado, utilizando os seguintes componentes: folha e

flor. Então, esses dois componentes apresentam um desempenho, ou seja, uma acurácia individual após a primeira parte do processamento.

4.6 TÉCNICAS DE FUSÃO - ETAPA E

Nesta etapa foi investigado a fusão dos componentes. Essa é a segunda etapa do processamento, cada componente já foi avaliado individualmente, para que então seja capaz de combinar alguns elementos, aumentando o nível de acurácia para a classificação das plantas, dentre os métodos de fusão, utilizamos: produto e soma, outra abordagem utilizada é o stacking e o uso de algoritmos genéticos para fins de decidir quais classificadores estarão presentes na combinação. A soma consiste na combinação das saídas dos classificadores individuais de cada componente, realizando a soma entre as probabilidades, para o produto, aplica-se o mesmo princípio, no entanto, as probabilidades são multiplicadas. Para o stacking a saída dos classificadores é combinada em um único vetor de forma que sejam concatenadas e utiliza-se algoritmos rasos como meta-classificadores. Através dos algoritmos genéticos será fornecida qual a melhor combinação de classificadores possível para o problema apresentado, logo a função objetivo será maximizar a acurácia, trazendo o conjunto de classificadores como melhor solução.

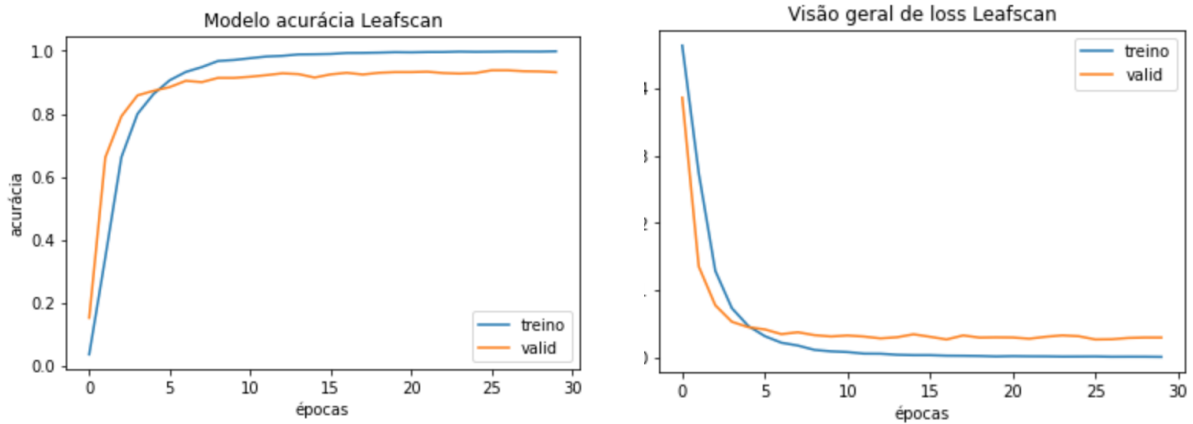
Para avaliação dos modelos será utilizado o top-1, top-3 e top-5, que significam, respectivamente: Top-1 é a precisão correta, onde o modelo previsto é exatamente o esperado. Top-3 considera a classificação correta se qualquer uma das três maiores previsões corresponder ao rótulo de destino, aplicando-se o mesmo princípio para o Top-5, no entanto, utilizando as cinco maiores previsões.

5 RESULTADOS E DISCUSSÕES

5.1 LEAFSCAN - FOLHA

Para resultados primários, ou seja, avaliação individual do modelo se tratando da Etapa D da metodologia e considerando apenas o conjunto de imagens de folha, também conhecidos na base como *leafscan*. A acurácia ao longo das épocas, pode ser observado na Figura 25, no primeiro gráfico a linha azul é referência aos resultados diante da base de treino e a linha laranja, através da base de validação, a base de treino foi dividida em treino e validação, ficando com 80% e 20%, respectivamente. Vale lembrar que para essa tarefa, a CNN utilizou 6880 imagens para treinamento da folha, e 1720 imagens para validação. Como as imagens possuem formatos variados, elas foram padronizadas antes da entrada da CNN, para o tamanho de 224 x 224 pixels. Foi utilizado como padrão a execução de 30 épocas. Também disponibilizamos uma visualização de perda (*loss*) que tem por princípio identificar o quão ruim foi a previsão do modelo para uma imagem, então quanto mais próximo do zero, melhor é a representação do modelo. A partir da Tabela 4 é possível visualizar a evolução do treinamento, 30 épocas foram utilizadas para o treinamento e validação do modelo, junto disso a partir da segunda coluna o tempo em segundos que o computador demorou para processar a CNN, seguida da perda na base de treino e acurácia na base de treino, então a perda na base de validação e a acurácia na base de validação. Importante destacar que a acurácia na base de treinamento do modelo já alcançou 99% para treinamento e 97% para validação. Embora tenhamos um total de nove redes pré-treinadas, o gráfico de evolução do treinamento e a tabela se referem ao melhor resultado individual para a folha, que neste caso compreende a rede neural InceptionResnet.

Figura 25. Gráfico de evolução de acurácia *leafscan* (treino e validação) e taxa de perda *loss* (InceptionResnetV2).



Fonte: O autor

Quadro 4. Treinamento Leafscan - Melhor resultado individual (InceptionResnetV2)

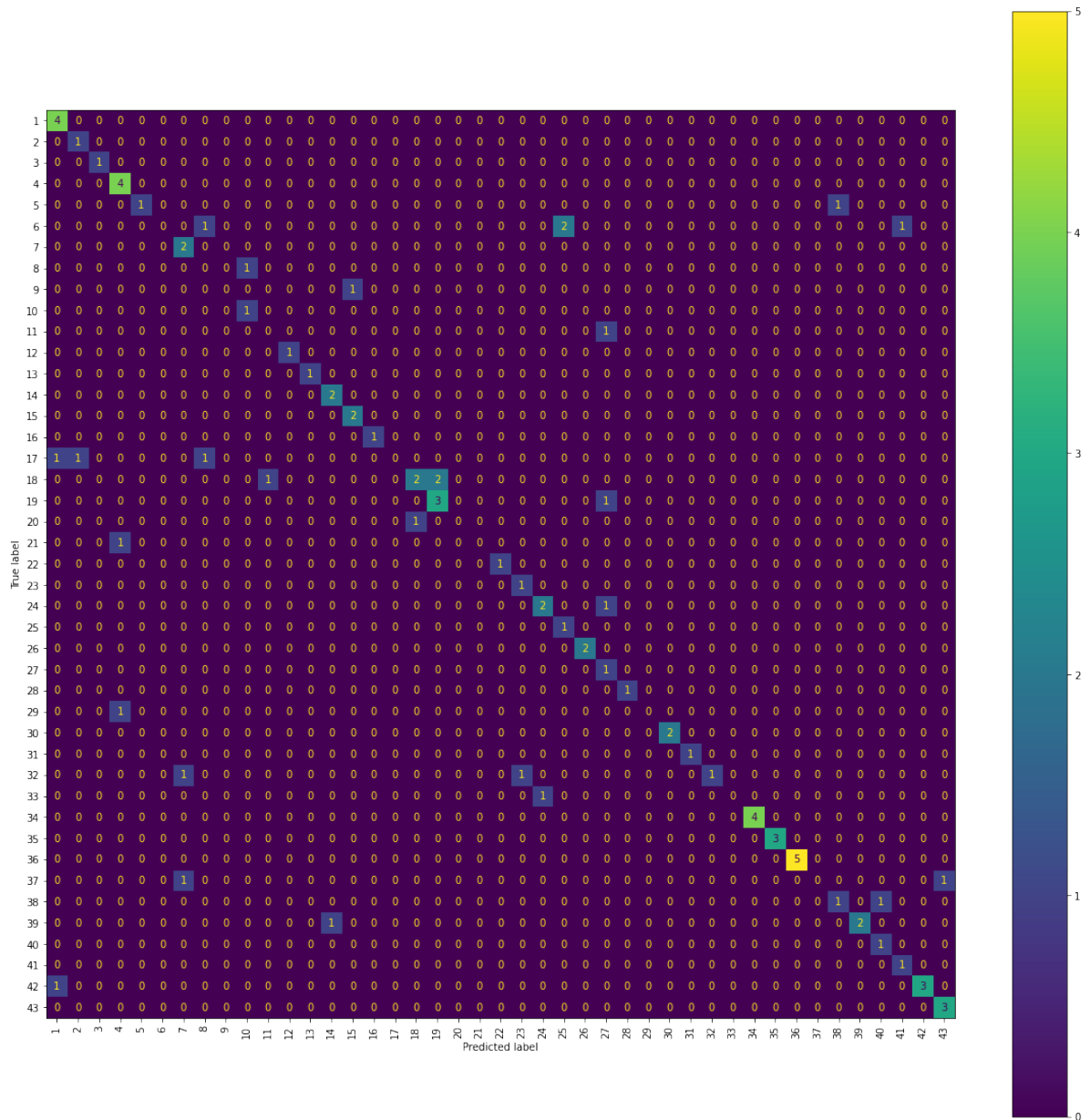
	Tempo (s)	loss	accuracy	val_loss	val_accuracy
Épocas 1/30	98	3.8730	0.0710	3.07210	0.3845
Épocas 30/30	48	0.0088	0.9993	0.0984	0.9771

Fonte: O autor

Para a execução de 30 épocas, foi adicionado mais uma camada de *Dropout* a fim de evitar o *overfitting*, no entanto foi constatado que o erro na validação estabilizou, não sendo necessário tantas épocas para aprovar o modelo da maneira que se encontra. Após o treinamento e validação a base foi submetida à base de teste do LifeCLEF, nela todas as imagens são inéditas, não estiveram presentes nem no treinamento e nem na validação, a base de teste do *Leafscan* utilizada contém 89 imagens pertencentes a 43 espécies distintas, ou como estamos abordando aqui, 43 classes. Para tal execução no melhor modelo foi obtido 69% de acurácia para *leafscan*, sendo este o melhor resultado individual para a folha. Através da Figura 26 é possível visualizar a Matriz de Confusão e em seguida um gráfico de barras ascendente com o resultado de todas as redes que foram avaliadas. Todas as redes foram testadas tanto com as camadas congeladas para realizar a transferência de aprendizado quanto descongelando e atualizando os pesos das camadas. Os

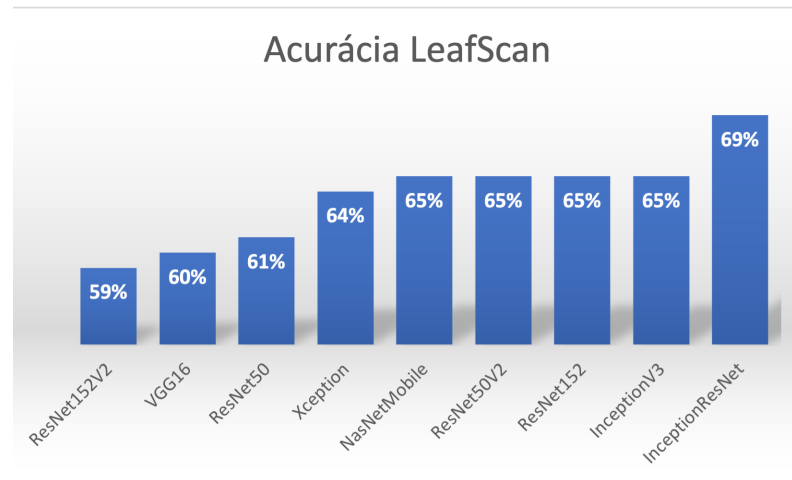
melhores desempenhos das redes para a folha foram obtidos utilizando o descongelamento das camadas e atualizando o peso, com exceção da VGG16 que teve suas camadas iniciais congeladas, e retreinando apenas os dois últimos blocos.

Figura 26. Matriz - LeafScan - Melhor resultado individual (InceptionResnetV2)



Fonte: O autor

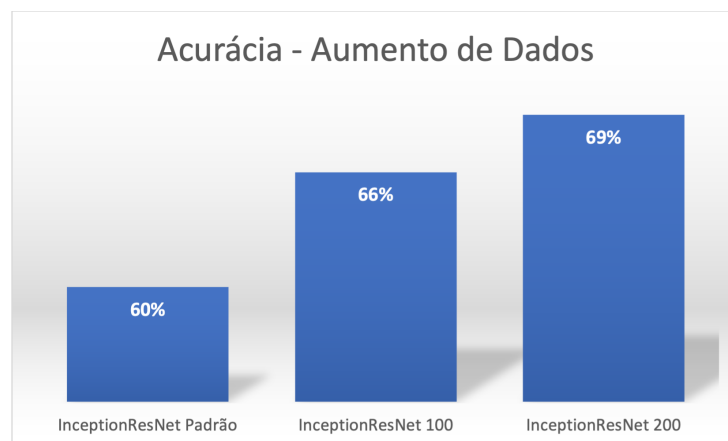
Figura 27. Acurácia LeafScan redes pré-treinadas - resultado individual



Fonte: O autor

Outro teste que foi realizado é para a comparação do aumento de dados, para isso foi aplicado o mesmo protocolo de execução para três diferentes variações. InceptionResnet Padrão é o resultado da rede sem o aumento de dados, InceptionResnet 100 significa um aumento de dados totalizando 100 imagens por classe utilizando recursos de rotação e flip_top_bottom e InceptionResnet 200, são 200 imagens por classe e variando mais os recursos, além das rotações já utilizadas, foram adicionadas brilho, contraste, cor, distorção, ruído e perspectiva enviesada. O aumento de dados fez com que a rede melhorasse o desempenho.

Figura 28. Acurácia LeafScan - Aumento de Dados

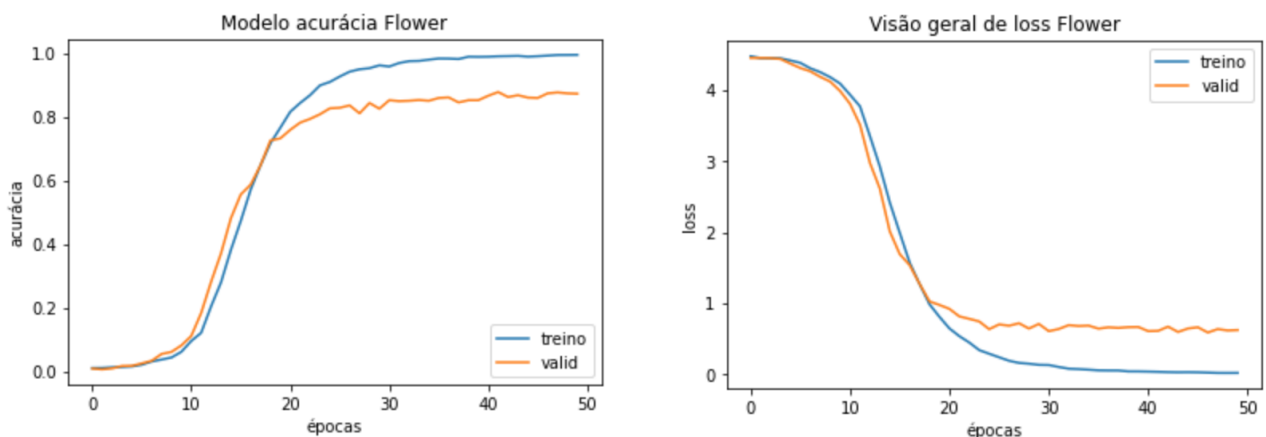


Fonte: O autor

5.2 FLOWER - FLOR

Para resultados primários no que se refere às flores. A acurácia ao longo das épocas, pode ser observado na Figura 29, onde no primeiro gráfico a linha azul é referência aos resultados diante da base de treino e a linha laranja, através da base de validação, a base de treino foi dividida em treino e validação, ficando com 80% e 20%, respectivamente. A CNN utilizou 6880 imagens para treinamento da folha, e 1720 imagens para validação. As imagens foram padronizadas antes da entrada da CNN, para o tamanho de 224 x 224 pixels. Foi utilizado como padrão a execução de 30 e 50 épocas. Também disponibilizamos uma visualização de perda (*loss*) que tem por princípio identificar o quão ruim foi a previsão do modelo para uma imagem, então quanto mais próximo do zero, melhor é a representação do modelo. A partir da Tabela 5 é possível visualizar a evolução do treinamento.

Figura 29. Gráfico de evolução de acurácia *flower* (treino e validação) e taxa de perda



Fonte: O autor

Quadro 5. Evolução do treinamento Flower - Melhor resultado individual (Resnet50)

	Tempo (s)	loss	accuracy	val_loss	val_accuracy
Épocas 1/30	77	4.2930	0.0127	2.1539	0.4056
Épocas 50/50	57	0.0077	0.9995	0.0743	0.9545

Fonte: O autor

Para a execução de 50 épocas, foram adicionadas mais duas camadas de *Dropout* a fim de evitar o *overfitting*. Os pesos foram carregados do ImageNet e então as camadas foram congeladas para a transferência de aprendizado. Após o treinamento e validação a base foi submetida à base de teste, nela todas as imagens são inéditas, não estiveram presentes nem no treinamento e nem na validação, a base de teste da Flor utilizada contém 89 imagens pertencentes a 43 espécies distintas, ou como estamos abordando aqui, 43 classes. Para tal execução foi obtida 67% de acurácia para *flower*. Sendo possível a visualização através da Matriz de Confusão. Especificando mais a transferência de aprendizado, para a flor os melhores desempenhos foram alcançados congelando as camadas para as redes ResNet50, ResNet152, NasNetMobile e Xception. Já para as redes InceptionResNetV2 e InceptionV3 os melhores resultados foram descongelando e atualizando o peso das camadas e VGG16 que teve suas camadas iniciais congeladas, e retreinando apenas os dois últimos blocos.

Figura 30. Matriz de Confusão - Flower - Melhor resultado individual (Resnet50)

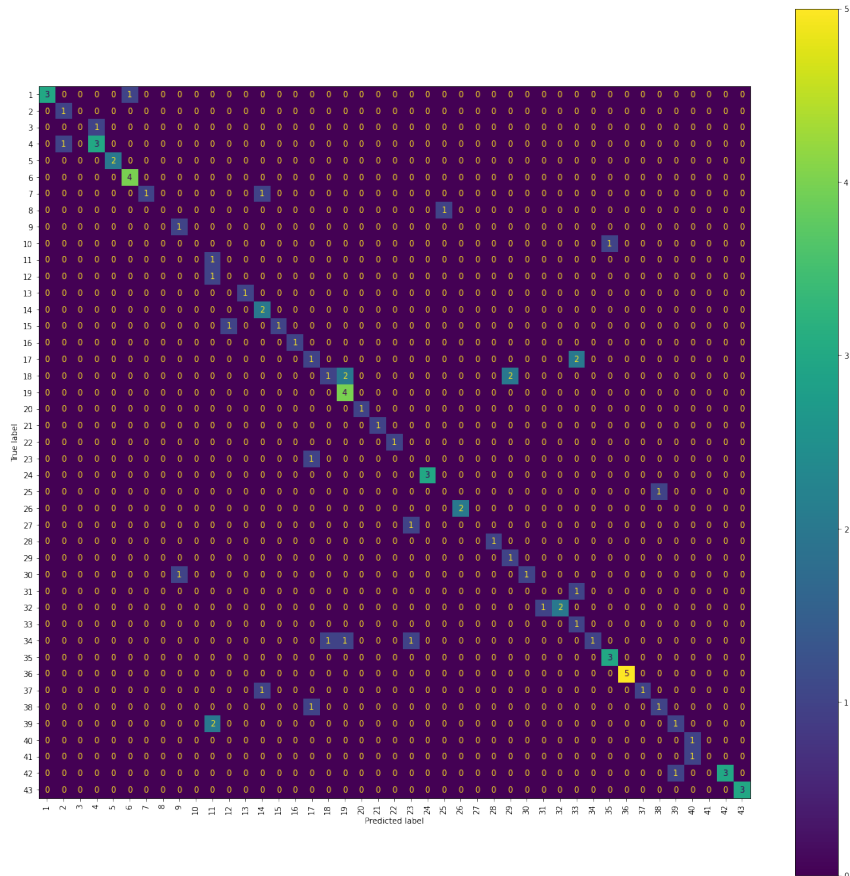
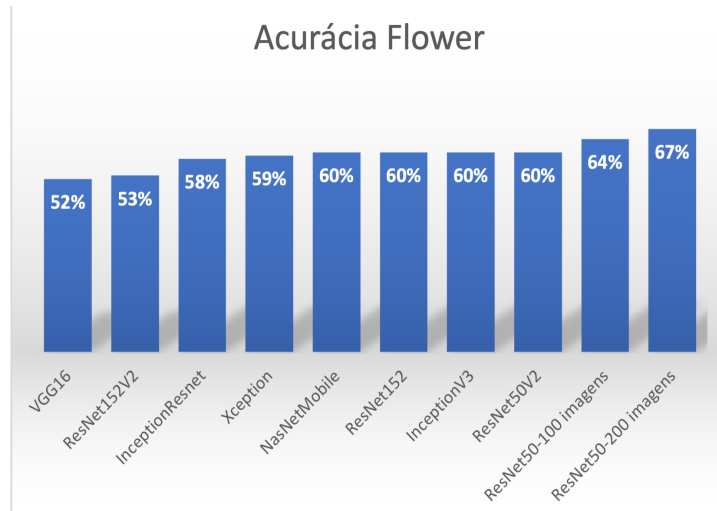


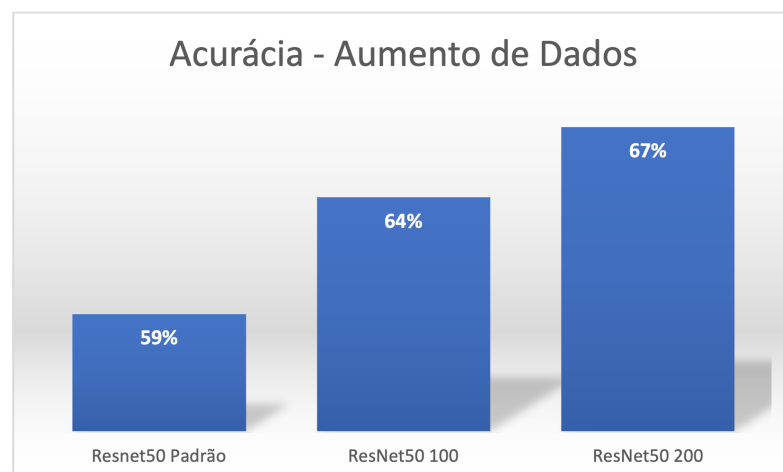
Figura 31. Acurácia Flower todas as redes pré-treinadas - resultado individual



Fonte: O autor

De semelhante modo ao que foi realizado para o Leafscan, foi realizado o teste para Flower com objetivo de mensurar o impacto do aumento de dados para os resultados, logo, Resnet 50 Padrão se refere ao resultado sem aumento de dados, Resnet50 100 totalizando 100 imagens por classe utilizando recursos de rotação e flip_top_bottom e Resnet50 200 com um aumento deixando 200 imagens por classe e variando mais os recursos, além das rotações já utilizadas, foram adicionadas brilho, contraste, cor, distorção, ruído e perspectiva enviesada.

Figura 32. Acurácia Flower - Aumento de Dados

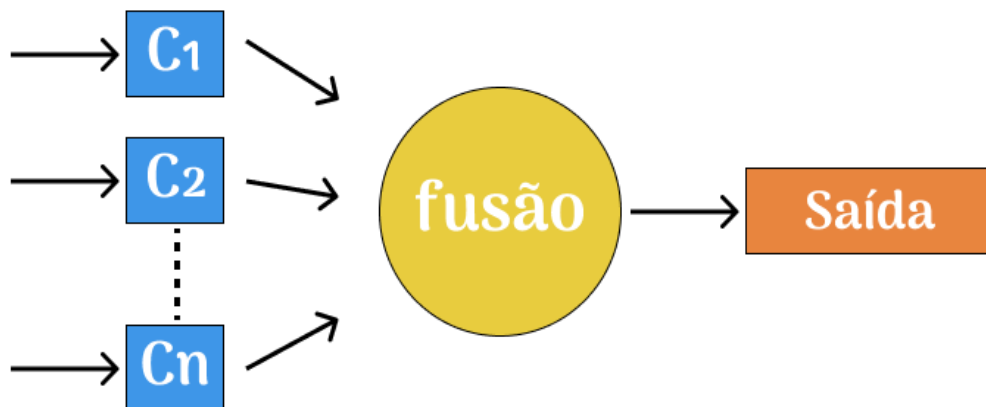


Fonte: O autor

5.3 FUSÃO DO MODELOS

Após as avaliações dos modelos individualmente, foi realizado o experimento de combinar as saídas dos classificadores para cada classe. Proposta por Kittler et al. (1998), a soma apresenta resultados que a *posteriori* não irão se desviar muito dos resultados a *priori* e pode ser visualizada através da Figura 33.

Figura 33. Fluxo da fusão



Fonte: O autor

Para a fusão, foram utilizadas a técnica de soma, que consiste em somar as probabilidades das saídas de cada classe, também utilizado o produto, onde o princípio é o mesmo que o da soma, no entanto, é realizada a multiplicação das probabilidades de cada classe. A criação de um meta-classificador, realizando o stacking entre vários algoritmos rasos e por fim o uso de algoritmos genéticos para informar qual a melhor combinação de classificadores de modo que a acurácia seja maximizada. Para isso, foram utilizadas as 43 classes de cada espécie, somando e multiplicando o vetor de probabilidades de cada imagem.

5.3.1 Seleção Estática de Classificadores usando Algoritmo Genético

O algoritmo genético irá otimizar a escolha dos classificadores para a melhor solução possível. Utilizando como entrada do algoritmo genético as probabilidades para cada classe resultantes dos 18 modelos treinados, sendo eles nove para folha e

nove para flor e definindo uma função objetivo de modo que venha maximizar a acurácia do modelo, o algoritmo genético buscou a melhor combinação de classificadores a ser utilizada entre os 18 disponíveis. No Quadro 6 é possível visualizar os parâmetros utilizados para a melhor solução encontrada.

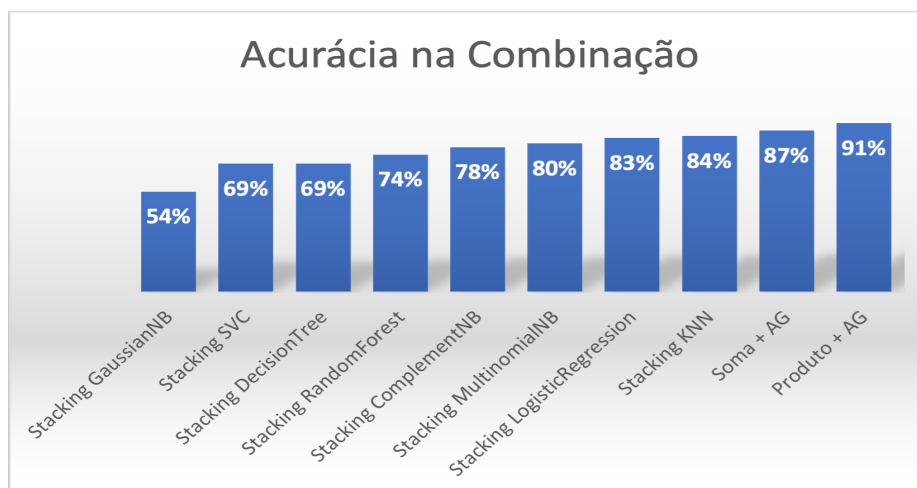
Quadro 6. Parâmetros Algoritmos Genéticos

max_num_ iteration	population_ size	mutation_ probability	elit_ratio	crossover_ probability	Best
200	300	0.2	0.1	0.8	X
200	30	0.2-0.8	0.2-0.8	0.1-0.7	-
2000	3000	0.2-0.8	0.2-0.8	0.1-0.7	-

Fonte: O autor

Para tal tarefa, aumentamos para 91% a acurácia do modelo usando algoritmos genéticos para encontrar a melhor combinação dos modelos, maximizando a acurácia média geral em 23 pontos percentuais. A partir da Figura 34 é possível visualizar o desempenho da fusão.

Figura 34. Acurácia na Combinação dos modelos

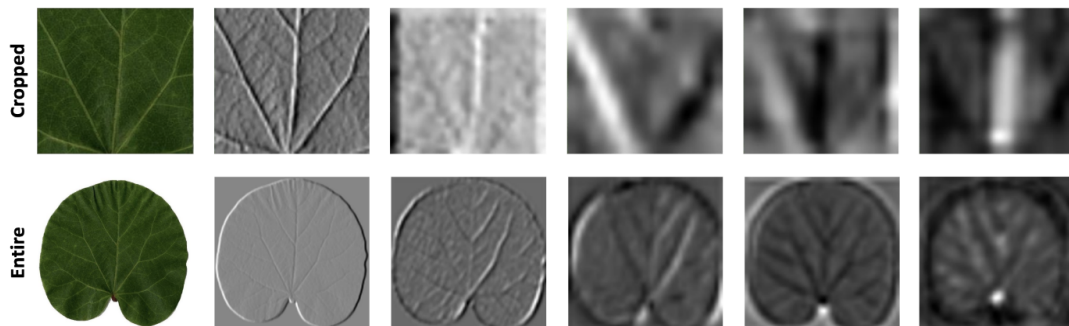


Fonte: O autor

5.4 COMPARAÇÃO DOS RESULTADOS OBTIDOS

Considerando a base de dados LifeCLEF 2015, podemos citar (ARAÚJO et al. 2020), que realizou uma classificação refinada a partir de duas visões da planta. No primeiro estágio da classificação era revelado o gênero da planta e em seguida sua espécie. O processo utilizou SCNN que calcula a similaridade entre uma imagem de folha e outra de referência. No primeiro processamento é considerada uma imagem inteira da folha da planta de modo que o formato e a cor sejam utilizadas para a classificação. Na segunda etapa do processamento é utilizada uma referência da imagem mais local, ou seja, o centro da imagem é recortado para que possa considerar a textura e os padrões de veias da folha, sendo mostrado na Figura 36. As duas entradas que eram realizadas, a planta inteira e recortada.

Figura 36. Imagens de entrada na SCNN



Fonte: Araújo et al. 2020

A sua acurácia considerando apenas a espécie foi de 75%, no entanto, o objetivo de seu trabalho era combinar para sua classificação o gênero e a espécie de cada planta, para tal alcançou uma acurácia de 87%, em contraste, tivemos uma acurácia de 91% combinando a folha e a flor da planta. A partir da Figura 37, seguem exemplos de imagens onde as predições foram 100% corretas.

Figura 37. Imagens *Acer Campestre* (flor e folha)



Fonte: O autor

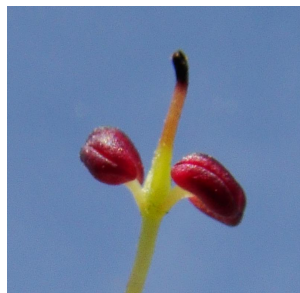
Figura 38. Imagens *Viburnum Opulus L* (flor e folha)



Fonte: O autor

A classe *Acer Campestre* e *Viburnum Opulus L* possui 4 imagens em sua base de teste, todas elas foram preditas corretamente. Na Figura 39 relatamos o pior caso, onde a *Fraxinus Angustifolia Vahl* possuindo 5 imagens em sua base teste, previu corretamente apenas uma.

Figura 39. Imagens *Fraxinus Angustifolia Vahl* (flor e folha)



Fonte: O autor

A partir da Figura 40 é possível visualizar, quais foram as classes que a classe *Fraxinus Angustifolia* Vahl previu incorretamente e qual a classe a imagem pertence realmente.

Figura 40. Imagens *Buddleja Davidii* Franch predita pela *Fraxinus Angustifolia* Vahl



Fonte: O autor

Figura 41. Imagens *Fraxinus excelsior* L. predita pela *Fraxinus Angustifolia* Vahl



Fonte: O autor

Figura 42. Imagens *Fraxinus ornus* L. predita pela *Fraxinus Angustifolia* Vahl



Fonte: O autor

Visualmente as folhas possuem uma similaridade muito grande, fato esse que prejudicou a predição correta das imagens. A Tabela 7 compara os resultados alcançados com alguns trabalhos que utilizaram a mesma base de imagens LifeCLEF 2015.

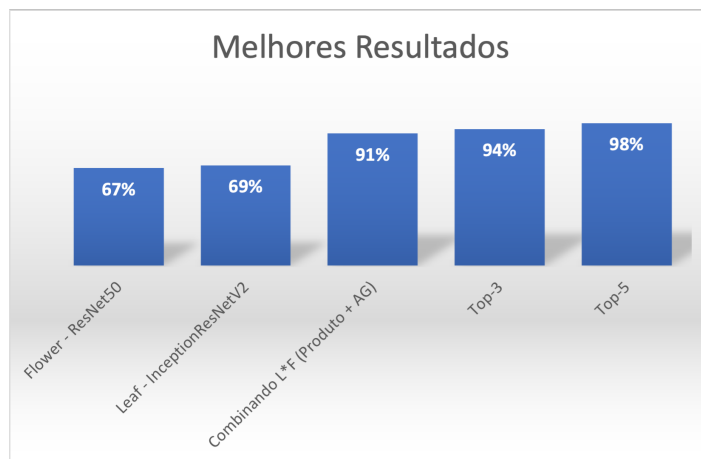
Quadro 7. Comparação com trabalhos que realizaram combinações

Referência	Metodologia empregada	Base utilizada	Acurácia
Lee (2018)	Combinação de modelos	LifeCLEF 2015	0,73
Ghazi (2017)	Combinação de modelos	LifeCLEF 2015	0,80
Araújo (2020)	Combinação de gênero e espécie da folha	LifeCLEF 2015	0,87
Método proposto	Combinação da saída dos classificadores otimizados com AG	LifeCLEF 2015	0,91

Fonte: O autor

Considerando o Top-3 onde o algoritmo calcula como uma classificação correta quando a imagem certa está entre as três maiores probabilidades a acurácia atingiu 94% e no Top-5 alcançou 98% de acurácia.

Figura 43. Melhores resultados (Individuais, combinação, top-3 e top-5)



Fonte: O autor

6 CONCLUSÕES

Neste trabalho foi implementado técnicas de aprendizagem profunda, através de redes neurais pré-treinadas no imagenet. Os componentes da planta utilizados no trabalho foram folha e flor, de modo que inicialmente as imagens foram pré-processadas, realizando aumento de dados para balancear as classes e aumentar a representatividade de algumas espécies que possuíam poucos exemplares. Em seguida, foram submetidos à tarefa de classificação, através de dezenove diferentes redes neurais convolucionais com ajuste fino, obtendo um resultado e uma acurácia individual. Após isso, foram realizadas algumas combinações de técnicas de fusão de modelos, utilizando as probabilidades de classe, sendo soma, produto e a criação de um meta-classificador os métodos consumidos na combinação, após isso a escolha dos classificadores para o resultado e acurácia final, foi feita através da otimização usando algoritmos genéticos.

Os resultados experimentais para a tarefa de classificação individual, alcançaram 69% de acurácia para a folha, também chamada de Leafscan, sob a rede InceptionResnet, para a flor, o resultado individual alcançou 67% de acurácia quando processada através da Resnet50. Após as técnicas de fusão, os melhores resultados foram 84% de acurácia no uso do stacking através de um meta-classificador KNN. Quando somada as probabilidades de cada classe atingiu 87% de acurácia, sendo a escolha dos classificadores através do algoritmo genético. O ápice e melhor resultado experimental do trabalho foi 91% de acurácia ao recorrer ao produto na combinação, sendo também este otimizado na escolha com o uso de algoritmos genéticos e 94% de acurácia no top-3 e 98% no top-5. Foi observado um aumento de 22 pontos percentuais quando comparado o melhor modelo individual com a combinação final de folha e flor.

Em relação aos trabalhos futuros, é interessante utilizar mais componentes da planta para fortalecer a classificação, uma outra abordagem importante é realizar um pré-processamento nas imagens de modo que consiga determinar melhor o que será utilizado na rede neural, segmentar uma imagem removendo itens desnecessários da imagem, como fundo e outros atributos que porventura apareçam, também utilizar uma combinação variada de fusão dos classificadores para investigar qual a melhor para o caso específico. Outro aspecto que pode ser vislumbrado é atuar nas imagens

onde a rede neural não conseguiu classificar corretamente, para que possa mapear quais as dificuldades da rede para possivelmente atenuar e reduzir o erro.

REFERÊNCIAS

ABADI, Martín et. al. TensorFlow: A System for Large-Scale Machine Learning. **USENIX Association**, p. 265-283, 2016.

ARAÚJO, Voncarlos. **Identificação de Espécies de Plantas Utilizando Combinação de Classificadores**. 2016. Tese (Mestrado em Computação Aplicada) – Universidade Estadual de Ponta Grossa, 2016.

AHN, Jin Mo et al. Google Inception v3 architecture: A schematic view of the Inception v3 model. **PLOS ONE**, 2018.

ARAÚJO, V. et al. Multiple Classifier System for Plant Leaf Recognition. **IEEE International Conference on Systems, Man, and Cybernetics (SMC)**, p. 1880-1885, out 2017.

ARAÚJO, H. D. Flávio et al. Redes Neurais Convolucionais com Tensorflow: Teoria e Prática. Sociedade Brasileira de Computação. III Escola Regional de Informática do Piauí. **Livro Anais-Artigos e Minicursos**, 2017 v.1, p. 382-406.

ARAÚJO, V. et al. Fine-grained Hierarchical Classification of Plant Leaf Images using Fusion of Deep Models, **IEEE International Conference on Tools with Artificial Intelligence**, p. 1-5, nov 2018.

ARAÚJO, V. et al. Two-View Fine-grained Classification of Plant Species. **Elsevier Neurocomputing**, v. 467, p. 427-441, 7 jan 2022

BLOICE, Marcus D. et al. Augmentor: An Image Augmentation Library for Machine Learning. **The Journal of Open Source Software**, 11 ago 2017.

CHRISTOPHER, Mark, et al. Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. **Scientific Reports 8**, 12 nov 2018.

DE FARIA, Elisangela Lopes. **Redes neurais convolucionais e máquinas de aprendizado extremo aplicadas ao mercado financeiro brasileiro**. 2018. Tese (Doutorado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, set 2018.

FARIA, Gedson; ROMERO, Roseli A. Francelin. Navegação de robôs móveis utilizando aprendizado por reforço e lógica fuzzy. **SBA Controle & Automação**, set 2002.

FERGUSON, Max, et al. Automatic Localization of Casting Defects with Convolutional Neural Networks. **IEEE International Conference on Big Data**, p. 1726-1735, 2017.

GHAZI, Mostafa Mehdipour, et al. Plant identification using deep neural networks via optimization of transfer learning parameters. **Elsevier Neurocomputing**, v. 235, p. 228-235, abr 2017.

GLOROT, Xavier; BORDES, Antoine; BENGIO, Y. Deep Sparse Rectifier Neural Networks. **Journal of Machine Learning Research**. 15, 01 jan 2010.

GOEAU, H. et al. Deep learning for plant identification: how the web can compete with human experts. **Biodiversity Information Science and Standards** 2, 22 mai 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning** (Vol.1). Cambridge: MIT Press, 2016.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens**. São Paulo: Pearson Prentice Hall. 2010.

GUYER, D. E. et al. Machine Vision and Image Processing for Plant Identification. **Transactions of the ASAE - American Society of Agricultural Engineers (USA)** 1986.

HE, Kaiming, et al. Deep Residual Learning for Image Recognition. **Computer Vision and Pattern Recognition**, 2015.

IMAGECLEF. **Tarefa da planta LifeCLEF**. 2015. Disponível em: <https://www.imageclef.org/lifeclef/2015/plant> Acesso em: 23 jun 2021.

IMAGECLEF. **Repositório de imagens**. 2015. Disponível em: <http://otmedia.lirmm.fr/LifeCLEF/PlantCLEF2015/Packages/>. Acesso em: 23 jun 2021.

IMAGENET. **Imagenet**. Disponível em: <https://www.image-net.org>. Acesso em: 23 jun 2021.

KAEHLING, L. P.; LITTMAN, M. L.; MOORE, A. W. Reinforcement Learning: A Survey. **Journal of Artificial Intelligence Research** 4. p. 237-285, 1986.

KAVYA, R. Feature Extraction Technique for Robust and Fast Visual Tracking: A Typical Review. **International Journal of Emerging Engineering Research and Technology**, vol. 3, p. 98-104, 2015.

KERAS. **Keras**. Disponível em: <https://keras.io/> Acesso em: jun 2021.

KITTLER, Josef, et al. On Combining Classifiers. **IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE**, v. 20, p. 226-239, 1998.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in Neural Information Processing Systems** **25**, 2012.

JARRET, Kevin, et al. What is the Best Multi-Stage Architecture for Object Recognition? **International Conference on Computer Vision**, 2009.

LEE, H. S.; CHANG, S. C.; REMAGNINO, P. Multi-Organ Plant Classification based on Convolutional and Recurrent Neural networks, **IEEE Transactions on Image Processing**, v. 27, p. 4287-4301, set 2018.

LEONARDO, Matheus Macedo, et al. Deep Feature-Based Classifiers for Fruit Fly Identification (Diptera: Tephritidae). **Conference on Graphics, Patterns and Images**, p. 41-47, 2018.

LIMA, A. et al. Redes neurais convolucionais aplicadas ao processo de classificação de cultivares de guaranazeiros. Encontro nacional de inteligência artificial e computacional. **Anais** Uberlândia: Sociedade Brasileira de Computação, out. 2017, p. 855-864.

LIU, T. et al. Comparing Fully Convolutional Networks, Random Forest, Support Vector Machine, and Patch-based Deep Convolutional Neural Networks for Object-based Wetland Mapping using Images from small Unmanned Aircraft System. **GIScience & Remote Sensing**, jan 2018.

LORENA, Ana Carolina; DE CARVALHO, André C. P. L. F. Uma introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, p. 43-67, 2007.

MARQUES FILHO, O.; NETO, H. V. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999.

MASSAGO, S. **Introdução ao Algoritmo Genético**. São Carlos: UFSCAR, 2013.

MEHTA, Parmita, et al. Multilabel multiclass classification of OCT images augmented with age, gender and visual acuity data. **Biorxiv**, 2018.

MITTAL, Parul, et al. Combined Classifier for Plant Classification and Identification from Leaf Image based on Visual Attributes. **International Conference on Intelligent Circuits and Systems**, p. 184-187, 2018.

MITCHEL, T. **Machine Learning**. New York: McGraw-Hill, 1997.

MJOLSNESS, E. Machine Learning for Science: State of the Art and Future Prospects. **Science**, 2001.

MURPHY, Grace E. P.; ROMANUK, Tamara N. A meta-analysis of declines in local species richness from human disturbances. **Ecology and Evolution**, v. 4, p. 91-103, 2014.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions, **Elsevier Pattern Recognition** v. 29, p. 51-59, jan 1996.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. São Paulo: Thomson Learning, 2008.

PEDROSA, Glauco V. **Caracterização e recuperação de imagens usando dicionários visuais semanticamente enriquecidos**. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Universidade de São Paulo, 2015.

PELLUCCI, Paulo R. S. et al. Use of machine learning techniques in recognition of portuguese named entities. **Exacts**, v. 4, pg. 73-81, 2011.

PEREIRA, F.; BURGESS, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. **Advances in Neural Information Processing Systems 25**. NY: Curran Associates, 2012.

REYES, A.K.; CAICEDO, J.C.; CAMARGO, J.E. Fine-tuning deep convolutional networks for plant recognition. **CLEF**, 2015.

GOTARDE, Felipe R.; DALLILO, Felipe D. Algoritmos Genéticos: Uma revisão bibliográfica sobre o conceito. **Revista Científica Semana Acadêmica**, 2021

RZANNY, M. et al. Flowers, leaves or both? How to obtain suitable images for automated plant identification. **Plant Methods**, jul 2019.

SAXEN, Frerk et al. Face Attribute Detection with MobileNetV2 and NasNet-Mobile. **International Symposium on Image and Signal Processing and Analysis**, p. 176-180, 2019.

SEELAND, M. et al. Image-based classification of plant genus and family for trained and untrained plant species, **BMC Bioinformatics**, jan 2019.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-scale Image Recognition. **Computer Vision and Pattern Recognition**, 2015.

SHUSTROV, D. et al. Fine-Grained Wood Species Identification Using Convolutional Neural Networks, **Conference on Image Analysis**, p. 67-77, mai 2019.

SZEGEDY, Christian, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. **Computer Vision and Pattern Recognition**, 2016.

TAN, Mingxing; QUOC V. L. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. **Machine Learning**, 2020.

WATTANAPORNPRON, Warin, et al. Ensemble of Multiple Classifiers for Multilabel Classification of Plant Protein Subcellular Localization. **Life**, 2021.

WEISS, Karl, et al. A survey of transfer learning. **Journal Big Data** 3, 2016.

WU, Stephen Gang. et al. **A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network**. IEEE International Symposium on Signal Processing and Information Technology, 2007.

YANIKOGLU, B., et al. **Automatic plant identification from photographs**. Machine Vision and application 25, p. 1369-1383, 2014.

ZOPH, Barret, et al. **Learning Transferable Architectures for Scalable Image Recognition**. Computer Vision and Pattern Recognition, 2018.