PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ – PUCPR
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA – PPGIa

VONCARLOS MARCELO DE ARAÚJO

RECONHECIMENTO DE ESPÉCIES DE PLANTAS A PARTIR DA IMAGEM DA FOLHA E DO USO DA APRENDIZAGEM PROFUNDA

CURITIBA – PR
2021

VONCARLOS MARCELO DE ARAÚJO

RECONHECIMENTO DE ESPÉCIES DE PLANTAS A PARTIR DA IMAGEM DA FOLHA E DO USO DA APRENDIZAGEM PROFUNDA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática (PPGIa) da Pontifícia Universidade Católica do Paraná (PUCPR), requisito parcial para a obtenção do título de Doutor em Informática.
Linha de pesquisa: Visão Computacional.

Orientadores:
Prof. Dr. Alceu de Souza Britto Júnior
Prof. Dr. Alessandro L. Koerich

CURITIBA – PR
2021

AGRADECIMENTOS

Ao meu orientador, Alceu de Souza Britto Jr, pelo apoio, não somente relativo à orientação deste estudo, mas por toda a ajuda prestada ao longo do período de formação acadêmica. Agradeço, também, ao Dr. Alessandro L. Koerich, por ter me recebido no período em que fiz pesquisa no Canadá, cuja sabedoria e conselhos enriqueceram esta pesquisa. Obrigado por acreditarem e investirem no meu potencial.

Aos amigos do PPGIa: André Brun, Bruna Delazeri, Cheila Cristina, Cleverton Vicentini, Diogo Olsen, Estefânia Fuzyi, Flávia Beuting, Flávio de Almeida e Silva, Jonathan de Matos, Francis Baranoski, Gustavo Bonacina, Irapuru Florido e Débora, Jhonatan Geremias, Kelly Wiggers, Luiz Giovanini, Marcelo Pereira, Marcos Monteiro Júnior, Ronan Assumpção Silva, Rodolfo Botto, Rodrigo Siega, Sandoval Ruppel, Teruo Maruyama e Vilmar Abreu.

Por fim, à minha família, por todo o apoio e incentivo.

RECONHECIMENTO DE ESPÉCIES DE PLANTAS A PARTIR DA IMAGEM DA FOLHA E DO USO DA APRENDIZAGEM PROFUNDA

Autor: Voncarlos Marcelo de Araújo
Orientador: Prof. Dr. Alceu de Souza Britto Junior
Coorientador: Prof. Dr. Alessandro L. Koerich

RESUMO

A identificação automática de plantas é considerada um desafio na área de reconhecimento de padrões devido à dificuldade na distinção de espécies semelhantes e sua ampla diversidade biológica. Modelos de aprendizagem profunda são recomendados pela literatura para tratar esse tipo de problema, contudo, dependem do uso de grandes conjuntos de dados. A escalabilidade é outro aspecto desejável em modelos de reconhecimento automático, isto é, novas espécies de plantas devem ser adicionadas de forma incremental. O presente estudo descreve dois métodos para o reconhecimento automático de plantas a partir da imagem digital do componente folha. Para tal, modelos de aprendizagem profunda, Redes Neurais Convolucionais (CNN, do inglês, *Convolutional Neural Network*) e Redes Neurais Siamesas (SNN, do inglês, *Siamese Neural Network)*, treinados sob dois pontos de vista da folha (geral e local), são utilizados em uma abordagem de classificação *Coarse-to-fine*, ou seja, gênero-espécie. Os métodos propostos também consideram estratégias para tratar desbalanceamento dos conjuntos de dados e garantir escalabilidade. Os resultados experimentais foram avaliados em dois conjuntos de dados de plantas (PlantCLEF 2015 e LeafSnap), e se mostraram promissores e favoráveis quando comparados com o estado da arte. O principal método proposto, baseado em SNN, obteve 0,87 e 0,96 de acurácia para os conjuntos de dados PlantCLEF 2015 e LeafSnap, respectivamente. Com isso, é possível afirmar que o método proposto consegue atingir alta performance sob dados desbalanceados, além de garantir a sua escalabilidade.

Palavras-chave: Reconhecimento de plantas, Rede Neural Siamesa, Rede Neural Convolucional.

RECOGNITION OF PLANT SPECIES FROM THE LEAF IMAGE AND THE USE OF DEEP LEARNING

Author: Voncarlos Marcelo de Araújo
Supervisor: Prof. Dr. Alceu de Souza Britto Junior
Co-Supervisor: Prof. Dr. Alessandro L. Koerich

ABSTRACT

Automatic plant identification is considered a challenge in pattern recognition due to difficulty in distinguishing between similar species and their wide biological diversity. Deep learning models are recommended in the literature to address this type of problem but generally rely on the use of large data sets. Scalability is another desirable aspect in automatic recognition models, that is, new plant species must be added incrementally. This work describes two methods for automatic plant recognition from the digital image of the leaf component. To this end, deep learning models, Convolutional Neural Networks (CNN) and Siamese Neural Networks (SNN), trained with two points of view of the leaf (global and local) are used in a *Coarse-to-fine* classification approach, that is, genus-species. The proposed methods still consider strategies to deal with unbalanced data and guarantee scalability. The experimental results are evaluated in two plant datasets (PlantCLEF 2015 and LeafSnap). The results obtained were promising and favorable when compared to the state of the art. The main proposed method, based on SNN, obtained 0.87 and 0.96 accuracy for the PlantCLEF 2015 and LeafSnap bases, respectively. With this, we demonstrate that the proposed method can achieve high performance under unbalanced data and guarantee scalability.

Keywords: Plant Recognition, Siamese Neural Network, Convolutional Neural Network.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DAS SIGLAS

ACO	*Ant Colony Optimization*
AHMTL	Aprendizagem Múltipla Hierárquica Profunda
BBF	*Best-Bin-First*
B-CNN	*Branch Convolutional Neural Network*
CNN	*Convolutional Neural Network*
DAG	*Directed Acyclic Graph*
DL	*Deep Learning*
DN	*Deconvolutional Network*
GIH	*Grayscale Intensity Histograms*
GLCM	*Gray Level Cooccurrence Matrix*
HD-CNN	Rede Neural Profunda Convolucional Hierárquica
HGO-CNN	*Hybrid Generic-Organ Convolutional Neural Network*
HOG	*Histogram of Gradients*
IDSC	*Inner-Distance Shape Context*
KNN	*K-Nearest Neighbors*
LBP	*Local Binary Pattern*
LDA	*Linear Discriminant Analysis*
LSH	*Locality Sensitive Hashing*
MARCH	*Multiscale-arch-height*
MLP	*Multilayer Perceptron*
NAG	*Nesterov Accelerated Gradient*
NB	*Naive Bayes*
NFC	*Neuro Fussy Classifier*
NN	*Neural Network*
PDA	*Penalized Discriminant Analysis*
PPGIa	Programa de Pós-Graduação em Informática
ReLU	Unidade Linear Retificada
RF	Árvores Aleatórias
RF	*Random Florest*
RNA	Rede Neural Artificial
SFS	*Search Forward Selection*
SGD	*Stochastic Gradient Descent*
SIFT	*Scale-invariant Feature Transform*
SMC	*System Man and Cybernetics*
SNN	*Siamese Neural Network*
SURF	*Speeded Up Robust Features*
SVM	Máquina de Vetor de Suporte
TAR	*Triangle Area Representation*
TL	*Transfer Learning*
TOA	*Triangle Represented by two Oriented Angles*
TSL	*Triangle Side Lengths Representation*
TSLA	*Triangle Represented by two Side Lengths and an Angle*
UHMT	*Hit or Miss Transform*
ZM	*Zernike Moments*

SUMÁRIO

1 INTRODUÇÃO

A diversidade biológica das plantas é indispensável ao ecossistema terrestre, visto que todos os seres vivos dependem, direta ou indiretamente, das inúmeras espécies de plantas que proporcionam diferentes formas de energia à natureza (CHAKI; PAREKH; BHATTACHARYA, 2015).

As plantas são consideradas os principais fornecedores de oxigênio na Terra, pois convertem gás carbônico em oxigênio (essencial para a maioria dos organismos vivos) por meio da fotossíntese. As diferentes espécies de plantas são utilizadas por ampla gama de aplicações industriais em diferentes setores, tais como: na nutrição, na produção de temperos, remédios medicinais e de biocombustíveis, assim como na geração de energia sustentável e renovável (ADAM *et al*., 2012; PROCHNOW *et al*., 2009).

Além disso, as plantas ajudam a regular o clima, servindo de *habitat* e comida para insetos e outros animais. Um bom conhecimento da flora é crucial para aumentar a produtividade agrícola e garantir a sustentabilidade do Planeta. Por isso, tornam-se importantes abordagens automáticas que permitam reconhecer espécies de plantas raras, ou até mesmo auxiliar na identificação de novas espécies.

Abordagens manuais de reconhecimento de plantas necessitam de profissionais com treinamento especializado, que possam identificar plantas por meio de informações-chave e comparações entre espécies foliares. Infelizmente, devido à grande diversidade de espécies presentes na natureza, profissionais como biólogos, muitas vezes, não têm informações suficientes para tratar da biodiversidade de maneira segura. Dessa forma, o reconhecimento automático de plantas por meio de imagens tem se mostrado uma área promissora nos últimos anos (ELHARIRI; EL-BENDARY; HASSANIEN, 2014; PRIYA; BALASARAVANAN; THANAMANI, 2012; WU *et al*., 2007).

Abordagens automáticas de reconhecimento de plantas costumam destacar a folha como um componente muito importante na tarefa de identificação devido às suas propriedades visuais, tais como: contornos, bordas, cor, textura e padrões de veias (ZHAO *et al*., 2015). Essas propriedades variam entre as diferentes espécies de plantas e são utilizadas como descritores (ou características) na criação de abordagens automáticas de reconhecimento por meio de imagens. As primeiras abordagens desenvolvidas para reconhecer plantas adotaram estratégias atualmente conhecidas como *handcrafted* (KEBAPCI; YANIKOGLU; UNAL, 2011; MOUINE; YAHIAOUI; VERROUST-BLONDET, 2012; LARESE *et al*., 2014;

YANIKOGLU; APTOULA; TIRKAZ, 2014; GHASAB *et al*., 2015). Tais estratégias dependem da habilidade de especialistas em visão computacional para extrair características discriminantes a serem utilizadas no treinamento de um ou mais classificadores.

Recentemente, a extração de características de alto nível baseada em aprendizagem profunda passou a ser utilizada em diferentes tarefas de visão computacional, tais como: classificação de objetos (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), reconhecimento de imagens (SIMONYAN; ZISSERMAN, 2014), segmentação semântica (LONG; SHELHAMER; DARRELL, 2015), dentre outras. Naturalmente, modelos profundos também são utilizados, com sucesso, no reconhecimento de plantas. O desafio mundial de reconhecimento automático de plantas por meio de imagens PlantCLEF 2015 (GOËAU; BONNET; JOLY, 2015) demonstra que abordagens que utilizam tais modelos, mais especificamente Redes Neurais Convolucionais, superam abordagens *handcrafted* com uma margem de 30% de acurácia. Desde então, diferentes abordagens de aprendizagem profunda têm sido criadas (LEE *et al*., 2017ab; GHAZI; YANIKOGLU; APTOULA, 2017; BARRÉ *et al*., 2017; BODHWANI; ACHARJYA; BODHWANI, 2019), entretanto, apesar do grande avanço na última década, a tarefa de reconhecimento automático de plantas continua sendo um desafio na área de reconhecimento de padrões.

1.1 MOTIVAÇÃO

A tarefa de reconhecer espécies, seja na flora ou na fauna, é essencial para o desenvolvimento sustentável do meio ambiente (ADAM *et al*., 2012). Com o abrangente conhecimento da diversidade biológica dos seres vivos, diversas informações sobre *habitat*, clima e ecossistemas passaram a ser interpretadas por profissionais da área. Aliado a isso, a evolução da tecnologia nos últimos anos permitiu que novos tipos de coleta de dados do meio ambiente sejam disponibilizados em diferentes domínios de aplicação. Nesse contexto, recursos como máquinas fotográficas e *smartphones* têm oferecido facilidades em larga escala na aquisição de dados ambientais, o que leva ao desafio de analisá-los e compreendê-los.

Construir um sistema automatizado para identificar espécies de plantas requer considerável experiência de domínio. A abordagem deve transformar os valores brutos dos *pixels* de uma imagem em uma representação que possa aprender e classificar um padrão específico entre várias espécies de plantas visualmente semelhantes.

Apesar da possibilidade de utilização dos avanços do reconhecimento visual, especialmente das Redes Neurais Convolucionais (CNN) (ZHANG *et al*., 2015; YANG *et al*.,

2015; PAWARA *et al*., 2017; LECUN; BENGIO; HINTON, 2015; BARRÉ *et al*., 2017; LEE *et al*., 2017a), treinadas sob um grande número de amostras (KRIZHEVSKY, 2009; DENG *et al*., 2009), ainda é comum, no contexto das plantas, haver uma distribuição desbalanceada de amostras de treinamento por classe. Conjuntos desbalanceados têm a chamada distribuição de cauda longa: algumas classes principais reivindicam a maioria das amostras, enquanto a maioria das outras classes menores são representadas, relativamente, por menos amostras (KRISHNA *et al*., 2016; GUO *et al*., 2016). Além disso, a dificuldade de captura de imagens de plantas raras ou em locais de difícil acesso pode, de fato, causar o desbalanceamento entre as categorias.

Segundo Joly *et al*. (2014), a aquisição de dados, particularmente na biodiversidade floral, ainda é limitada pela falta de conhecimento dos seres humanos em identificar a grande diversidade de espécies. Isso reforça a necessidade de abordagens automatizadas, capazes de discriminar elementos entre categorias visualmente semelhantes, em conjuntos de dados desbalanceados.

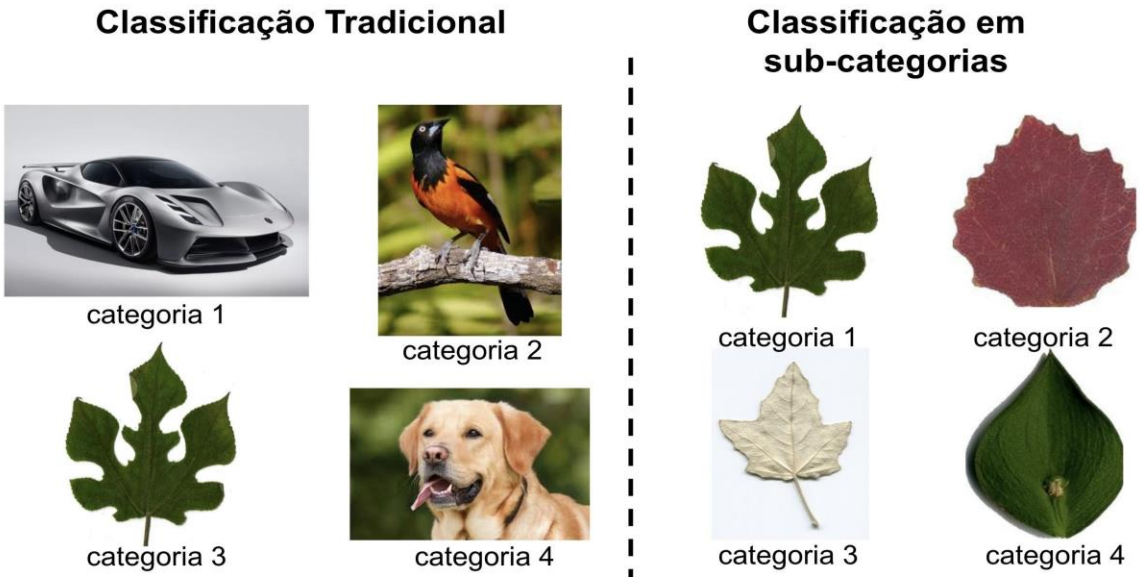
Atualmente, tem havido crescente interesse por pesquisas sobre classificação em subcategorias, também conhecida como *fine-grained*. O reconhecimento de subcategorias é um campo relativamente novo dentro do tema da classificação de objetos, que pode auxiliar na classificação de diferentes espécies de plantas (SFAR; BOUJEMAA; GEMAN, 2015; GE *et al*., 2016; ŠULC; MATAS, 2017; ZHANG *et al*., 2020).

1.2 DEFINIÇÃO DO PROBLEMA

Reconhecer plantas automaticamente não é apenas uma requisição de botânicos e ecologistas. Com a interdisciplinaridade e o avanço da tecnologia, essa tarefa é adotada em diferentes áreas, como na Agricultura, Medicina, Biologia, assim como pelo público em geral (conservacionistas, biólogos e visitantes de parques florestais). A criação de abordagens automatizadas de reconhecimento de plantas por meio de imagens do componente folha, no entanto, apresenta as seguintes dificuldades:

a) Classificação em subcategorias: Diferente da classificação tradicional de objetos, que visa encontrar uma categoria geral correta, tais como: um pássaro, um cachorro ou uma planta, o reconhecimento de plantas demanda a classificação de subcategorias (SFAR; BOUJEMAA; GEMAN, 2015; GE *et al*., 2016; ŠULC; MATAS, 2017; ARAÚJO *et al*., 2018; ZHANG *et al*., 2020). A Figura 1 ilustra o desafio da classificação em subcategorias.

Figura 1. Conceito de classificação tradicional de objetos versus classificação em subcategorias



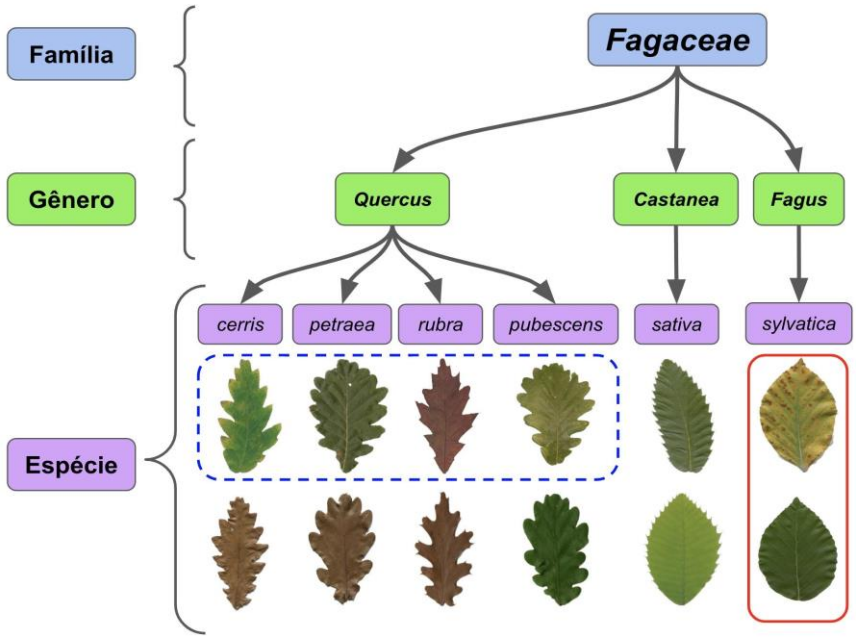
Fonte: elaboração própria (2021).

A classificação tradicional, geralmente, se refere à distinção de categorias de objetos muito diferentes, como uma categoria de carro ou de cachorro, por exemplo. Em uma classificação de subcategorias, todas elas pertencem à mesma categoria geral básica e, neste caso, as subcategorias são diferentes espécies de plantas.

Embora a classificação tradicional de imagens tenha progredido rapidamente nos últimos anos, ainda é tarefa desafiadora realizar um reconhecimento preciso e refinado em subcategorias de objetos. Existem dois aspectos que tornam o reconhecimento de subcategorias um problema desafiador à visão computacional, os quais são descritos a seguir e visualizados na Figura 2:

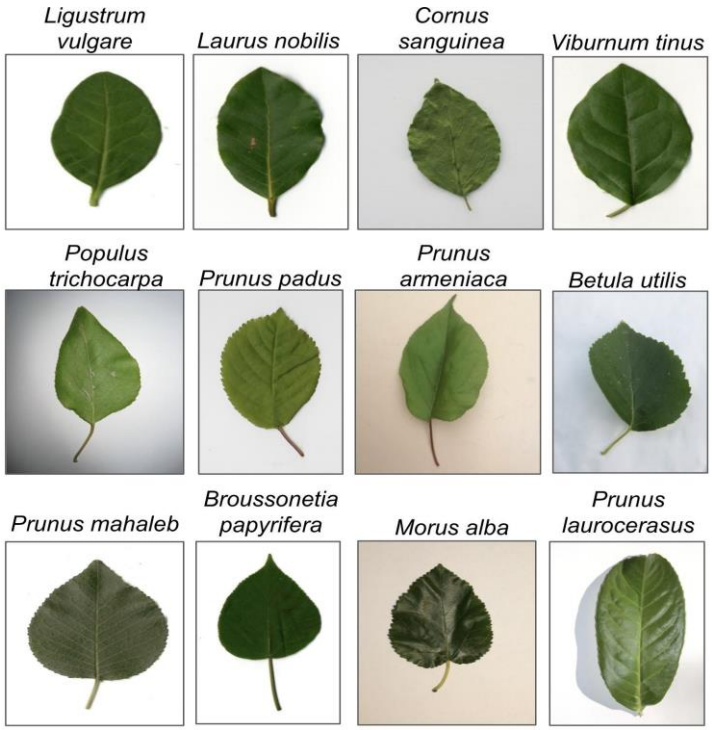
1. **Semelhança interespécies:** O retângulo pontilhado azul na Figura 2 ilustra um exemplo de diferenças sutis entre diferentes espécies de plantas. Algumas espécies têm formas idênticas e, às vezes, elas até compartilham semelhanças de cores, formas e texturas muito fortes. A Figura 3 apresenta uma diversidade de espécies que têm aspectos morfológicos similares.
2. **Alta variabilidade intraespécies:** O retângulo de linha vermelha na Figura 2 apresenta algumas variações que podem surgir em amostras presentes dentro da mesma classe. As imagens do componente folha geralmente são capturadas em diferentes cenários e suas características morfológicas podem variar dependendo da maturação da planta ou até mesmo variações de ângulos e iluminação. A Figura 4 ilustra algumas variações encontradas dentro de uma única espécie.

Figura 2. Classificação de subcategorias no contexto de plantas. Dificuldades intraespécies e interespécies



Fonte: elaboração própria (2021).
Legenda:
Retângulo pontilhado azul: interespécies (pequenas variações entre diferentes espécies).
Retângulo com linha vermelha: intraespécies (podem apresentar variações de fundo, oclusões, pose, cor, iluminação e estágios de maturação da planta dentro da mesma espécie).

Figura 3. Desafio interespécies: amostras de diferentes espécies com semelhanças sutis



Fonte: elaboração própria (2021).

Figura 4. Desafio intraespécies: variações encontradas dentro da espécie *Hedera helix*

Fonte: elaboração própria (2021).

b) Dados desbalanceados: A disponibilidade de uma distribuição uniforme de imagens rotuladas em conjuntos de dados de grande escala faz com que algumas abordagens tenham sucesso na tarefa de reconhecimento automático de objetos (KRAUSE *et al*., 2015). Alguns conjuntos de dados de reconhecimento visual comumente utilizados, por exemplo, CIFAR (KRIZHEVSKY, 2009) e ImageNet (DENG *et al*., 2009), exibem distribuições quase uniformes de classes rotuladas. Excepcionalmente, os conjuntos de dados do mundo natural têm distribuições desequilibradas (KENDALL; STUART; ORD, 1987), isto é, categorias dominantes reivindicam a maioria dos exemplos, enquanto grande parte das outras classes são representadas, relativamente, por poucos exemplos. Modelos treinados em tais dados têm um desempenho ruim para classes mal representadas (HORN; PERONA, 2017; JAPKOWICZ; STEPHEN, 2002; HE; GARCIA, 2009).

No reconhecimento de plantas algumas espécies são mais abundantes e fáceis de fotografar do que outras. Ao mesmo tempo, a falta de conhecimento especializado sobre plantas faz com que seja difícil aos humanos realizar anotações de rótulos reais e corretos em imagens, limitando ainda mais o número de amostras rotuladas disponíveis para cada espécie, visto que para coletar e construir um conjunto de dados robusto de plantas são necessárias redes sociais especializadas em imagens de plantas e sua validação por botânicos qualificados (JOLY *et al*., 2014). Consequentemente, algumas categorias sofrem do problema de dados desbalanceados, contendo substancialmente mais imagens de uma espécie em comparação à outra.

c) Quantidade de espécies: No contexto das plantas, a inserção de novas espécies na identificação se apresenta como dificuldade à visão computacional. Com o aumento do número

de espécies de plantas avaliadas, tais dificuldades – interespécies e intraespécies – se tornam ainda mais complexas. Além disso, há falta de dados rotulados em novas espécies devido à sua raridade e dificuldade de acesso a locais de captura. Uma solução escalável, portanto, que seja de fácil inclusão de novas espécies, deve ser explorada.

Não é trivial, contudo, criar modelos escaláveis em tarefas de reconhecimento de subcategorias, pois alguns modelos não generalizam pela carência de dados rotulados e ainda requerem retreinamento ou modificações para aprender, de forma consistente, novas categorias (TSAFTARIS; SCHARR, 2019). Normalmente, o retreinamento de um modelo de reconhecimento é utilizado para lidar com novos padrões adicionados, cujo processo é demorado e trabalhoso. A adição de novos padrões para reconhecimento também promove o aumento do número de pesos e parâmetros de modelos de reconhecimento. Tais circunstâncias tornam difícil treinar/ajustar, de forma eficiente, modelos de aprendizagem de maneira recorrente (treinamento após determinados intervalos de tempo ou com a adição de novas espécies de plantas). Até o presente momento, não se encontrou na literatura um sistema de reconhecimento escalável no desafio de reconhecimento de espécies de plantas.

Portanto, um diferencial em nosso trabalho em relação ao atual estado da arte, é a criação de um modelo de reconhecimento escalável, ou seja, o incremento de novas espécies está relacionado à passagem de um pequeno conjunto de referências, sem a necessidade de um retreinamento do sistema. Na prática, há forte necessidade de aprendizagem escalável no reconhecimento de subcategorias das plantas, já que especialistas estão constantemente explorando e aprendendo sobre novas espécies.

Para enfrentar esses desafios, busca-se, por meio desta tese, responder a algumas questões principais relacionadas ao reconhecimento de espécies de plantas:

- Sabendo que poucas amostras estão disponíveis em espécies de difícil acesso ou em novas espécies encontradas na natureza, seria possível inferir a transferência de aprendizagem por meio do uso de modelos profundos pré-treinados em outro domínio?
- Como aprender representações de características robustas que identifiquem diferenças sutis entre subcategorias?
- Uma estratégia hierárquica baseada na taxonomia disponível nas plantas poderia contribuir para mitigar os problemas relacionais à alta variabilidade intraclasse e a possível similaridade interclasse inerentes ao problema de classificação de espécies de plantas?
- O uso de diferentes representações da imagem da folha, considerando dois pontos de vista (geral e local), poderia contribuir no processo de classificação de espécies de plantas?

- É possível criar um modelo de reconhecimento de plantas escalável e eficiente que não necessite o retreinamento recorrente para cada nova espécie inserida?

1.3 OBJETIVOS

O objetivo geral deste estudo é propor dois métodos para o reconhecimento automático de plantas a partir da imagem digital do componente folha. Para tal, modelos de aprendizagem profunda, Redes Neurais Convolucionais (CNN, do inglês, *Convolutional Neural Network*) e Redes Neurais Siamesas (SNN, do inglês, *Siamese Neural Network*), treinados sob dois pontos de vista da folha (geral e local) são utilizados em duas abordagens de classificação *Coarse-to-fine*, ou seja, gênero-espécie.

As abordagens propostas neste estudo consideram, também, estratégias para tratar o desbalanceamento dos conjuntos de dados e garantir que novas espécies sejam reconhecidas sem a necessidade de um retreinamento do modelo de reconhecimento (escalabilidade). Para cumprir o objetivo geral, os seguintes objetivos específicos devem ser atingidos:

- Desenvolver uma classificação hierárquica por meio da taxonomia presente nas plantas. A estratégia *Coarse-to-fine* considera uma classificação por etapas, ou seja, gênero-espécie.
- Desenvolver a extração de características geral e local a partir de diferentes pontos de vista da imagem da folha da planta.
- Desenvolver um mecanismo de treinamento, utilizando poucas amostras por espécie para tratar desafios com dados desbalanceados.
- Garantir escalabilidade do método de tal forma que a inserção de novas espécies de plantas não demande o retreinamento dos modelos utilizados.

1.4 CONTRIBUIÇÕES

As contribuições científicas desta tese podem ser assim destacadas:

- Dois métodos de classificação de espécies de plantas sendo:
  1. O primeiro método baseado em Redes Neurais Convolucionais (CNN), no qual se utiliza a transferência de aprendizado de um conjunto de dados já existente juntamente com um ajuste fino.
  2. Um segundo método baseado em Redes Neurais Siamesas (SNN) que permite aprender padrões por meio da similaridade entre imagens.

- Esquema de representação do componente folha da planta baseado em dois pontos de vista (geral e local).
- Abordagem de classificação hierárquica *Coarse-to-fine* que considera a taxonomia hierárquica presente nas plantas (gênero-espécie).
- Estratégia para tratar desbalanceamento, considerando poucas amostras.
- Mecanismo para garantir escalabilidade em que novas espécies podem ser adicionadas facilmente, sem sobrecarregar o tamanho do modelo de aprendizagem (pesos e parâmetros), evitando o retreinamento.

Além das contribuições científicas supracitadas, destacam-se algumas colaborações sociais e ambientais deste estudo na criação de possíveis facilidades que podem surgir com base nos métodos propostos, sejam elas na Agricultura, na Educação ou no Ecossistema das plantas:

- Aplicações de reconhecimento automático: auxiliar especialistas e amantes da natureza na manipulação de nomes taxonômicos e no reconhecimento de plantas por meio de imagens.
- Reconhecimento em campo: educadores e admiradores da Ecologia podem usufruir de um possível aplicativo para ser utilizado como observação educacional ou entretenimento.
- Agências de controle biológico podem monitorar em tempo real o ecossistema das plantas. Com informações atualizadas é possível equilibrar o ecossistema e avaliar os impactos ambientais das atividades humanas a fim de planejar tomadas de decisões rápidas e efetivas.
- O compartilhamento e o armazenamento das informações das plantas podem facilitar o uso do conhecimento científico, beneficiando a sociedade na implantação de novas políticas ambientais.

1.5 PUBLICAÇÕES

As contribuições desta tese de doutorado foram publicadas nos seguintes artigos:

- ARAÚJO, V. M.; BRITTO, A. S.; BRUN, A. L.; KOERICH, A. L.; FALATE, R. *Multiple classifier system for plant leaf recognition*. IEEE International Conference on Systems, Man and Cybernetics (SMC). Banff, Canadá, 2017, pp. 1880-1885.

- ARAÚJO, V. M.; BRITTO, A. S.; BRUN, A. L.; KOERICH, A. L.; OLIVEIRA, L. E. S. *Fine- Grained Hierarchical Classification of Plant Leaf Images Using Fusion of Deep Models.* IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), Volos, 2018, pp. 1-5.

- ARAÚJO, V. M.; BRITTO, A. S.; KOERICH, A. L.; OLIVEIRA, L. E. S. *Two-View Fine-grained Classification of Plant Species*, 2021, pp. 1–18. O artigo está sob revisão da Revista Neurocomputing.

1.6 ORGANIZAÇÃO DA TESE

Esta tese está organizada em seis capítulos, conforme descrição a seguir: *Capítulo 1* – expõe as principais motivações, desafios e justificativas na tarefa de reconhecer espécies de plantas; *Capítulo 2* – aborda o referencial teórico, enfatizando as técnicas relacionadas ao tema central deste estudo; *Capítulo 3* – trata de uma revisão bibliográfica detalhada; *Capítulo 4* – expõe os dois métodos propostos para investigar as proposições iniciais, embasados na análise bibliográfica; *Capítulo 5* – apresenta os resultados obtidos, bem como comparações e análises realizadas; *Capítulo 6* – destaca as conclusões, as contribuições relevantes desta tese e as indicações de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados alguns conceitos sobre plantas que fundamentaram a construção deste estudo, especificamente quanto ao funcionamento do processo de organização taxonômica das plantas e sua categorização. Posteriormente, são apresentados extratores de características responsáveis por representar imagens, enfatizando o uso de modelos profundos, tais como: Redes Neurais Convolucionais (CNN) e Redes Neurais Siamesas (SNN). Ao final, apresenta-se as principais métricas de distância utilizadas no aprendizado de modelos profundos, cujo objetivo é estimar a similaridade entre duas imagens.

2.1 ORGANIZAÇÃO TAXONÔMICA DAS PLANTAS

O reconhecimento manual de plantas requer um especialista treinado para realizar o procedimento de forma complexa e demorada, com auxílio de livros como guias e ferramentas manuais de identificação (ELPEL, 2013). Normalmente, nesses guias encontram-se descrições das plantas para que os especialistas tenham em mãos informações importantes, tais como a taxonomia foliar.

A taxonomia é responsável por agrupar biologicamente organismos com base nas características comuns das plantas, formando diversos grupos com características morfológicas similares. Para cada grupo é dada uma nota, sendo que podem ser agregados para formar um supergrupo de maior pontuação, criando uma classificação hierárquica (JUDD *et al*., 2007). Os supergrupos criados por esse processo são referidos como taxonomia foliar ou *taxon*.

Geralmente, os livros ou guias utilizados pelos taxonomistas apresentam descrições organizadas por características de espécies semelhantes, de modo que outros biólogos possam utilizar o mesmo livro como suporte para reconhecer espécies desconhecidas. Na Botânica, essa tarefa implica em comparações entre características armazenadas e observadas pelo botânico que, em seguida, atribui um grupo taxonômico conhecido a uma planta específica, definindo, finalmente, a sua espécie.

Devido à grande variação de características fundamentais entre espécies de plantas (diversidade biológica), reconhecê-las pode ser uma tarefa onerosa e demorada. A rapidez no reconhecimento manual requer, muitas vezes, o conhecimento prévio da taxonomia foliar (à qual família ou gênero uma planta pertence). Ademais, algumas características morfológicas presentes em espécies desconhecidas podem ser similares entre diferentes famílias e gêneros,

causando confusões na sua identificação. A dificuldade, porém, ainda é maior pela escassez de taxonomistas qualificados (CARVALHO *et al*., 2007).

A categorização biológica das plantas é um passo crítico no processo taxonômico, pois informa hipoteticamente os componentes do *taxon*. Embora a disciplina de taxonomia em si não lide com as investigações quanto à forma como estão relacionadas umas às outras, ela serve para comunicar os resultados. Para isso, ela as classifica em ordens taxonômicas, segundo Linnaeus (1758), ou seja, em ordem do maior para o menor: Reino, Filo, Classe, Ordem, Família, Gênero e Espécie (Figura 5):

Figura 5. Taxonomia biológica das plantas, segundo Linnaeus (1758)

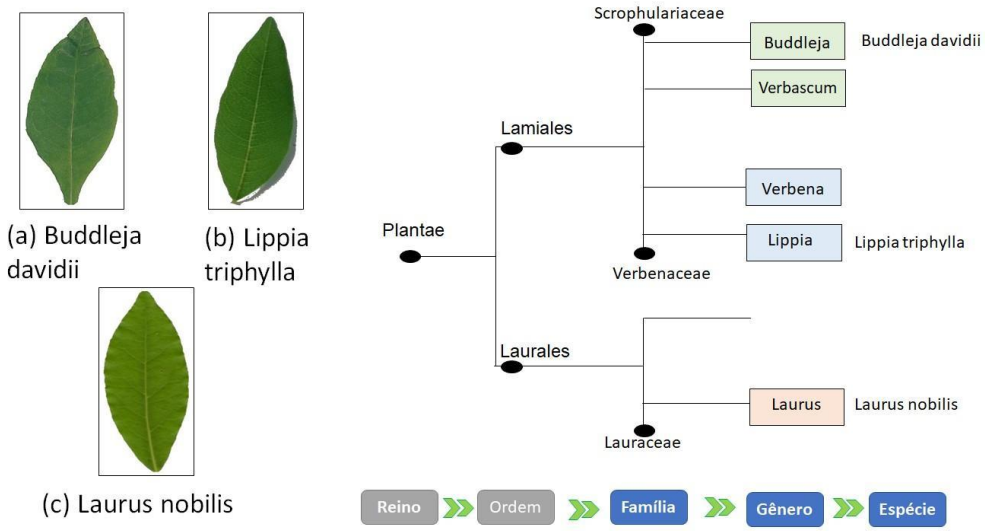
Fonte: elaboração própria com base em Linnaeus (1758).

Espécies botânicas são naturalmente organizadas em uma taxonomia hierárquica, em que distintos sistemas taxonômicos são usados para diferentes tipos de plantas. Por exemplo, o sistema APG III (AGPGROUP, 2009) é atualmente utilizado pela maioria dos botânicos a fim de classificar as plantas, sendo um sistema de taxonomia vegetal (Figura 6). A representação hierárquica pode permitir que especialistas da área reduzam seus erros de classificação de acordo com a "proximidade" das características foliares na hierarquia.

A representação hierárquica dificilmente é única, havendo muitas maneiras de decompor recursivamente os dados. Por exemplo, a hierarquia pode ser projetada manualmente usando grupos biológicos pré-definidos ou construídos automaticamente mediante o uso de características morfológicas presentes nas plantas. Na Figura 6 constata-se a existência de três espécies distintas: (a) *Buddleja davidii*; (b) *Lippia triphylla*; e (c) *Laurus nobilis*, as quais têm características morfológicas similares.

A morfologia refere-se à formação estrutural das plantas. Nesse aspecto, é importante perceber que a hierarquia composta por essas três classes se encontra em espaços diferentes da taxonomia foliar. Por exemplo, a espécie (a) *Buddleja davidii* pertence ao Reino *Plantae*, Ordem *Lamiales*, Família *Scrophulariaceae* e Gênero *Buddleja*, enquanto outras espécies com aspectos morfológicos similares podem estar em outras camadas da hierarquia, como a espécie *Laurus nobilis*, a qual tem sua taxonomia pertencente à Ordem *Laurales*, Família *Lauraceae* e Gênero *Laurus* (Figura 6).

Figura 6. Folhas e seus respectivos posicionamentos na hierarquia taxonômica do sistema APG III



Fonte: elaboração própria com base em APG III (2009).
Legenda:
(a) Buddleja davidii
(b) Lippia triphylla
(c) Laurus nobilis

Observa-se que há semelhança visual entre as formas das folhas e, apesar da grande similaridade, existe expressiva distância entre essas espécies na hierarquia (o primeiro ancestral em comum entre as espécies é o nó raiz: Reino: *Plantae*).

2.2 APRENDIZADO PROFUNDO

Aprendizado profundo ou *Deep Learning* (DL) é um ramo da Aprendizagem de Máquina (*Machine Learning*) que visa modelar dados a partir do aprendizado da representação para classificação ou regressão. Historicamente, DL tem as suas raízes ligadas ao paradigma de aprendizagem conexionista, projetado para lidar com problemas de classificação linearmente separáveis, inspirados em neurônios biológicos (ROSENBLATT, 1958). Esforços realizados por pesquisadores da área buscaram alternativas para permitir a adaptação de vários hiperplanos na tentativa de classificar conjuntos de dados não linearmente separáveis.

Com base no estudo desenvolvido por Werbos (1988), passou-se a empregar vários hiperplanos a fim de criar classificadores não lineares e tratar tarefas mais complexas de aprendizado supervisionado. Esse feito foi realizado mediante a utilização de Redes Neurais Artificiais, especialmente após a implementação do algoritmo de treinamento de retro

propagação (*back-propagation*). Essa evolução acabou apoiando o design das redes conhecidas como *Multilayer Perceptron* (MLP) (CYBENKO, 1989).

Mais recentemente, o aprendizado profundo se apresentou como uma estratégia para modelar e classificar grandes quantidades de dados complexos, aprendendo representações de baixo e alto nível e usando uma combinação de funções não lineares, como as executadas pelo MLP (MALLAT, 2008). Os algoritmos de DL são projetados mediante o uso de uma arquitetura de rede com camadas convolucionais consecutivas. Uma etapa de treinamento é aplicada para atualizar os pesos da rede a fim de melhor representar os padrões de entrada de acordo com os rótulos esperados.

As Redes Neurais Convolucionais (CNN) estão entre as técnicas de DL mais populares para resolver problemas práticos de classificação. A CNN surgiu como uma alternativa para reduzir a densidade de unidades conectadas e logo se tornou referência no estado da arte, de acordo com vários pesquisadores, especialmente ao se lidar com tarefas de reconhecimento de objetos (STÉPHANE, 2009) ou de dígitos manuscritos (ALWZWAZY *et al*., 2016) e na detecção de face (YANG *et al*., 2015).

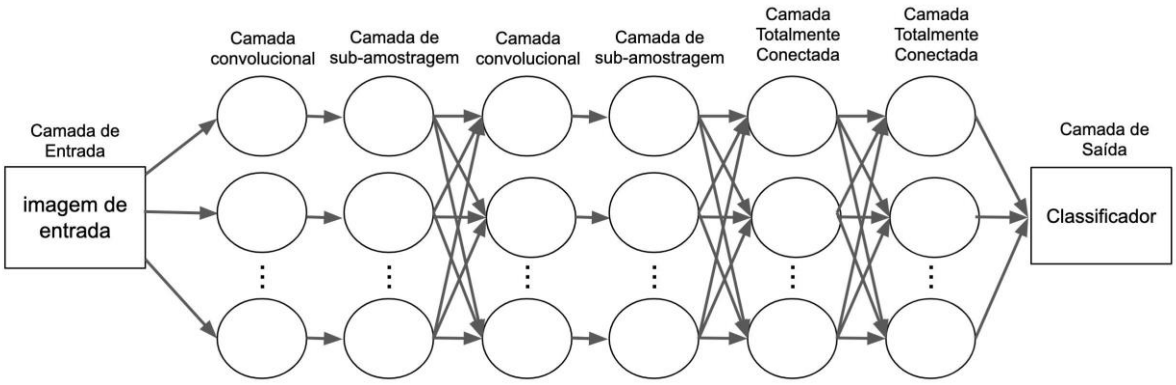
Nesse contexto é possível detalhar os processos de construção e treinamento das Redes CNNs. Primeiramente, apresenta-se o funcionamento das camadas que compõem a arquitetura de uma rede CNN (Seção 2.2.1), incluindo os algoritmos de treinamentos clássicos usados para otimizar o mapeamento de dados de entrada e saída (Seção 2.2.2). Além disso, discutem-se abordagens empregadas na literatura para evitar possíveis problemas no aprendizado da rede profunda (Seção 2.2.3). Finalmente, apresenta-se uma rede profunda de aprendizado métrico que mensura a similaridade entre amostras de imagens empregando duas CNNs equivalentes em sua arquitetura (Seção 2.2.4).

2.2.1 Arquitetura

Redes Neurais Convolucionais (CNNs) são organizadas em arquiteturas de multicamadas convolucionais (LECUN; BENGIO; HINTON, 2015). A primeira camada da rede é responsável por coletar os dados de entrada. As camadas posteriores, consideradas ocultas, realizam diferentes operações, tais como: convolução e redução de dimensionalidade dos dados de entrada. A última camada gera a saída da rede com rótulos preditos. É importante ressaltar que cada camada possui um conjunto de unidades associadas (neurônios) com o fim de calcular operações pré-definidas, como convolução, sub-amostragem e normalização (GOODFELLOW; BENGIO; COURVILLE, 2016).

Normalmente, as arquiteturas CNNs são organizadas da seguinte maneira: 1) camada de entrada; 2) camada de convolução com função de ativação; 3) camada de sub-amostragem; 4) camada totalmente conectada; e 5) camada de saída. A Figura 7 ilustra uma arquitetura típica empregada no domínio de classificação de imagens, sendo composta por uma camada de entrada, duas camadas convolucionais ocultas intercaladas com camadas de sub-amostragem, duas camadas totalmente conectadas e uma camada de classificação.

Figura 7. Ilustração de uma arquitetura tradicional CNN em tarefas de classificação de imagens



Fonte: elaboração própria (2021).
Obs.: Rede composta por uma camada de entrada, duas camadas escondidas convolucionais intercaladas com camadas de sub-amostragem, duas camadas totalmente conectadas e uma camada de classificação.

A estrutura CNN consiste em células simples e células complexas. Esta estrutura é inspirada nas estruturas do campo receptivo que são encontradas no córtex visual primário humano, descobertas pela primeira vez por Hubel e Wiesel (1962). A operação de convolução é utilizada graças à sua capacidade de extrair características de regiões locais da imagem (ABDEL-HAMID *et al.*, 2014), apoiando a identificação de padrões recorrentes a partir de dados de entrada. Nesse sentido, as recorrências estão associadas aos padrões mais semelhantes e prevalentes que ocorrem em uma entrada específica. A operação de convolução também produz uma representação robusta a variações, ou seja, as características produzidas ainda são representativas após determinados deslocamentos de entrada (LECUN; BENGIO, 1998; GOODFELLOW; BENGIO; COURVILLE, 2016).

A equação 2.1 estabelece uma operação de convolução a partir de uma imagem de entrada.

$$S_{i,j} = (I \otimes \Theta)_{i,j} = \sum_{x=1}^{m} \sum_{y=1}^{n} I_{i-x,j-y} \Theta_{x,y} \tag{2.1}$$

em que *I* é a imagem de entrada de tamanho $q \times p$; *i* e *j* correspondem aos índices de linhas e colunas de *I; x* e *y* são os índices de linhas e colunas da máscara de convolução; e Θ é a matriz que representa a máscara de convolução de tamanho $m \times n$. Normalmente, antes da operação de convolução é realizado um preenchimento com dígitos "zero" em torno da imagem, permitindo capturar informações relevantes presentes nas bordas das imagens de entrada.

A Figura 8 ilustra o processo realizado por uma unidade convolucional (ou neurônio), considerando uma imagem de entrada *I* com tamanho 4 × 4 com preenchimento de zeros. A figura também mostra uma máscara convolucional Θ com tamanho 3 × 3 que produz uma imagem de saída *S* de tamanho 4 × 4. Considerando este exemplo, a operação de convolução usada para obter $S_{1,1}$ é definida na Equação 2.2. Após a convolução, um viés multiplicado por algum peso pode ser adicionado da seguinte forma: $S_{1,1} + bw_0$, alterando o ponto de interceptação dessa função com o espaço de entrada das variáveis.

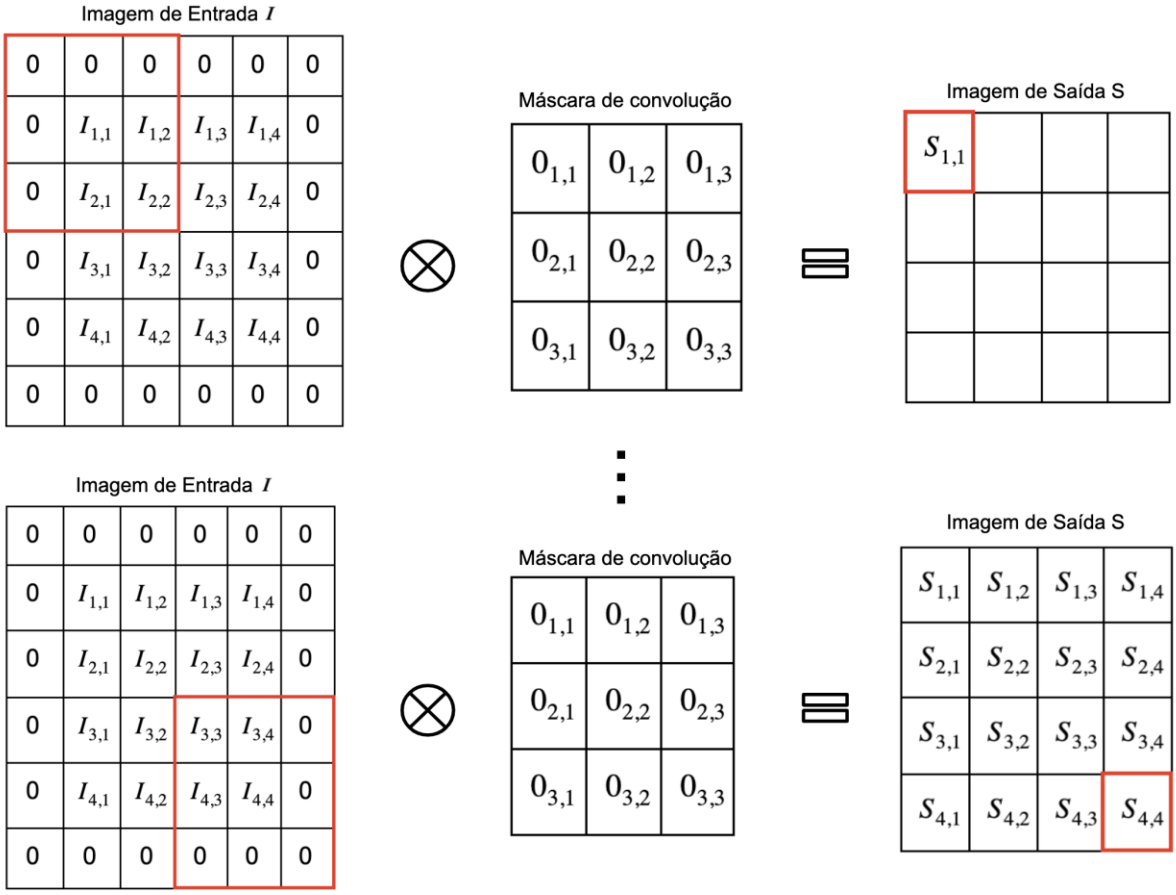
$$\begin{aligned} s_{1,1} = I_{0,0} \times \Theta_{3,3} + I_{0,1} \times \Theta_{3,2} + I_{0,2} \times \Theta_{3,1} \\ + I_{1,0} \times \Theta_{2,3} + I_{1,1} \times \Theta_{2,2} + I_{1,2} \times \Theta_{2,1} \\ + I_{2,0} \times \Theta_{1,3} + I_{2,1} \times \Theta_{1,2} + I_{2,2} \times \Theta_{1,1} \end{aligned} \tag{2.2}$$

Normalmente, uma função de ativação é aplicada posteriormente à operação de convolução a fim de normalizar os valores em um dado intervalo. Esse processo é útil em cenários supervisionados para definir a função de perda e destacar características relevantes na saída da operação convolucional. Muitas funções de ativação são empregadas pelas CNNs, apresentando a tangente hiperbólica (Equação 2.3) e a sigmoide (Equação 2.4) como as mais comuns (LECUN; BENGIO; HINTON, 2015; KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Estudos recentes afirmam que a função de Unidade Linear Retificada (ReLU) (Equação 2.5) melhora os resultados em determinados cenários (CIRESAN; MEIER; SCHMIDHUBER, 2012), evitando valores negativos e permitindo que os valores das saídas sejam maiores do que "um", diferentemente das outras duas funções que limitam o intervalo de suas saídas. Nesse sentido, a ReLU se assemelha melhor às magnitudes originais dos dados de entrada e tem sido amplamente utilizada em modelos CNNs.

$$y(x) = \frac{e^z - e^{-z}}{e^x + e^{-z}} \tag{2.3}$$

$$y(x) = \frac{1}{1 + e^{-x}} \tag{2.4}$$

Figura 8. Ilustração da operação de convolução



Fonte: elaboração própria (2021).
Obs.: Considerando uma imagem de entrada *I* de tamanho 4 × 4 com *zero-padding*, dado a máscara convolucional Θ de tamanho 3 × 3, para produzir uma saída *S* de 4 × 4. O resultado é multiplicação (*element-wise*) da região local da imagem *I* com a máscara de convolução Θ.

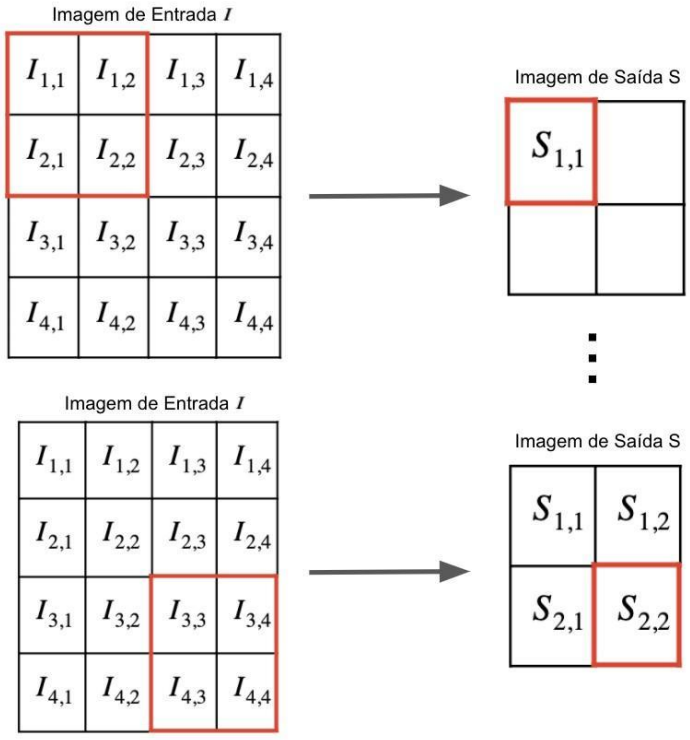
$$y(x) = max(0, x) \tag{2.5}$$

O passo seguinte para definir um modelo CNN se apresenta como uma função de redução, conhecida como sub-amostragem ou *pooling*. Essa função geralmente é aplicada na tentativa de manter as informações mais relevantes e reduzir as dimensões dos dados, resultando em características suficientemente robustas para distorções locais (LECUN *et al*., 1998). Duas operações principais de *pooling* são recomendadas pela literatura para lidar com tarefas de classificação de imagens, sendo *max-pooling* a mais comum, retornando o valor máximo de cada campo receptivo local à imagem processada (LECUN; BENGIO; HINTON, 2015). A outra operação de *pooling* é chamada de *average-pooling*, que apresenta a média de todos os campos receptivos locais (RANZATO *et al*., 2006). *Max-pooling* é amplamente utilizada em conjunto com arquiteturas CNNs, mantendo a magnitude das informações. Além disso, alguns

autores afirmam robustez à translação dos dados (LECUN; BENGIO; HINTON, 2015).

A Figura 9 ilustra um processo de *pooling* com tamanho 2 × 2 e passo 2, calculado por uma única unidade sob uma imagem de entrada *I* de tamanho 4 × 4. O passo ou *stride* se refere ao número de linhas e colunas percorridas pela máscara convolucional.

Figura 9. Ilustração da operação de *pooling*



Fonte: elaboração própria (2021).
Obs.: Considerando uma máscara convolucional de tamanho 2 × 2 com tamanho de passo (*stride*) de tamanho 2 × 2 aplicado em uma imagem de entrada de 4 × 4.

As últimas camadas da CNN são totalmente conectadas, em que as unidades calculam um produto interno semelhante às camadas ocultas do modelo *Multilayer Perceptron* (MLP). A Equação 2.6 define o produto interno, na qual *I* corresponde a imagem de entrada de tamanho *q* × *p*; *w* representa os pesos; e *b* é o viés ou o termo para alterar a interceptação da função de ativação utilizando as variáveis de entrada. Os pixels da imagem *I* e o viés *b* são elementos multiplicados por pesos e, posteriormente, todos os valores são somados (LECUN *et al.*, 1998). Vale ressaltar que uma função de ativação pode ser aplicada nas saídas dessa camada.

$$y(x) = \sum_{i=1}^{i=q} \sum_{j=1}^{j=p} (I_{i,j} * w_{i,j}) + b \times w_b \tag{2.6}$$

Nesse contexto, a última camada totalmente conectada (*softmax*) gera as probabilidades

de cada imagem de entrada ser associada às correspondentes classes de saída. Consequentemente, o número de unidades que compõem essas camadas é igual ao número de classes dos conjuntos de dados. Essa saída é usada para calcular uma função de perda e executar o treinamento, além de classificar exemplos desconhecidos. A função de perda mais utilizada é a entropia cruzada, que avalia a correlação entre probabilidades da classe correta e a obtida.

A Equação 2.7 define a função *softmax*, apresentando *C* como o número de classes e *x* como o vetor de saída (produzido pela camada totalmente conectada).

$$p_j = \frac{e_j^x}{\sum_{j=1}^{C} e_j^x} \tag{2.7}$$

Para uma tarefa de classificação, a função de perda é definida na Equação 2.8, na qual p_j é a probabilidade de a imagem de entrada pertencer à classe *j*. Com isso, y_j é igual a 1 quando *j* corresponde à classe correta, caso contrário, $y_j = 0$.

$$l = -\sum_{j=1}^{C} y_j \log p_j \tag{2.8}$$

Na literatura existem diversas arquiteturas CNN de sucesso. Por exemplo, o modelo LeNet (LECUN *et al*., 1998) conta com duas camadas de convolução seguidas de *pooling*, terminando com uma camada totalmente conectada. AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) contém cinco camadas de convolução seguidas de *pooling*, sendo duas camadas totalmente conectadas. Já a rede VGG-16 (SIMONYAN; ZISSERMAN, 2014) possui cinco blocos com camadas de convolução seguidas de *pooling* entre elas, e três camadas totalmente conectadas. Por fim, GoogLenet (SZEGEDY *et al*., 2014) abrange nove blocos, cada um com três camadas de convolução seguidas de *pooling*. Todas essas arquiteturas apresentaram resultados promissores quando aplicadas à área de reconhecimento de padrões, e são amplamente utilizadas em tarefas de reconhecimento e classificação de objetos.

2.2.2 Algoritmos de treinamento

O Gradiente Descendente Estocástico (*Stochastic Gradient Descent* – SGD) é um dos algoritmos mais populares para realizar o treinamento das Redes Neurais Convolucionais. Sua função é adaptar as máscaras convolucionais e os pesos das camadas totalmente conectadas, conforme a Equação 2.9.

$$w_{t+1} = w_t - \eta \frac{1}{N} \sum_{i=1}^{N} \nabla_{w_t} l(f(x_i, w_t), y_i) \tag{2.9}$$

onde x_i corresponde a um exemplo de treinamento rotulado por y_i; $f(x_i, w_t)$ representa a saída da CNN; η é a taxa de aprendizado; N é o número de exemplos; $l(\cdot)$ representa a função de perda; w_t é o peso atual; e w_{t+1} é o mesmo peso após sua adaptação. Segundo Bottou (2012), o gradiente descendente conduz a adaptação dos pesos da rede, considerando todos os exemplos de treinamento a cada iteração ou, então, um único exemplo de treinamento, conforme a Equação 2.10.

$$w_{t+1} = w_t - \eta \nabla_{w_t} l(f(x_i, w_t), y_i) \tag{2.10}$$

Considerando que o uso de um único exemplo por iteração consome muito tempo de processamento, o SGD é geralmente estimado a partir de um subconjunto de exemplos de treinamento, conhecido como *mini-batch*. A Equação 2.11 define o *mini-batch*, e segue os mesmos parâmetros da Equação 2.9, embora *n* represente o número de exemplos pertencentes a um subconjunto de treinamento (*mini-batch*).

$$w_{t+1} = w_t - \eta \frac{1}{n} \sum_{i=1}^{n} \nabla_{w_t} l(f(x_i, w_t), y_i) \tag{2.11}$$

A escolha da taxa de aprendizado do *mini-batch* deve ser levada em consideração. Uma taxa pequena pode levar à convergência lenta, enquanto uma taxa grande pode comprometer a convergência, causando flutuações em torno dos mínimos locais da função de perda. Nesse contexto, um termo conhecido por *momentum* é comumente adicionado na tentativa de evitar oscilações em funções de perda e melhorar a sua convergência (ROBBINS; MONRO, 1951). A Equação 2.12 define o SGD com um termo de *momentum* γ, no qual $\boldsymbol{v}$ representa o resultado do gradiente, $l(f(x_i, w_t), y_i)$ é a função de perda e $\boldsymbol{w}$ corresponde aos pesos.

$$\begin{aligned} v_t &= \gamma v_{t-1} - \eta \nabla_{w_t} l(f(x_i, w_t), y_i) \\ w_{t+1} &= w_t + v_t \end{aligned} \tag{2.12}$$

Ademais, um termo simples de regularização, conhecido como *weight decay*, é aplicado para melhorar a generalização e, também, evitar sobreajustes (*over-fitting*) (KROGH; HERTZ, 1992). O termo de regularização é atualizado com a função de perda para manter os valores em

torno de zero. O *weight decay* λ, geralmente é adicionado à função de perda, conforme definido na Equação 2.13, na qual $E(w_t)$ corresponde ao erro total e $\|w_t\|^2$ é a normalização L2 dos pesos.

$$E(w_t) = \frac{\lambda}{2}||w_t||^2 + \sum_{i=1}^{n} l(f(x_i, w_t), y_i) \tag{2.13}$$

Além do SGD existem ainda outros algoritmos que otimizam o treinamento de uma Rede CNN a fim de reduzir as perdas e fornecer resultados mais precisos. Nesse sentido, *Nesterov Accelerated Gradient* (NAG) (NESTEROV, 1983), Adadelta (ZEILER, 2012) e RMS-PROP (TIELEMAN; HINTON, 2012) são comumente utilizados na literatura.

2.2.3 Overfitting

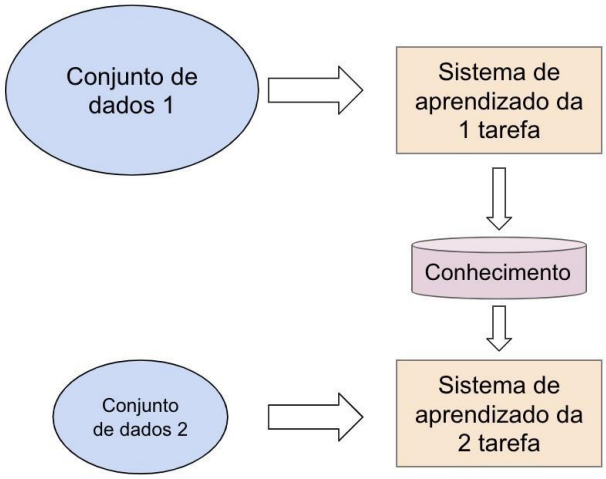
O sobre ajuste do modelo à base de treinamento é extremamente prejudicial à sua capacidade de generalização. Esse problema é conhecido na literatura como *overfitting*, uma vez que a rede tem bom desempenho no conjunto de treinamento, mas não funciona adequadamente em exemplos não vistos (amostras de teste). O problema surge quando o treinamento é realizado mediante o uso de uma arquitetura com um número excessivo de camadas e unidades operando em algum pequeno conjunto de treinamento (MELLO, 2019). Segundo Luxburg e Schölkopf (2011), o conjunto de treinamento deve ser representativo para fornecer garantias de aprendizado às arquiteturas complexas.

Geralmente, com base nesses aspectos, são aplicadas duas estratégias a redes profundas na tentativa de evitar o *overfitting*: i) o aumento do conjunto de treinamento; e ii) técnica de *dropout*.

O método de aumento de dados incrementa artificialmente o conjunto de treinamento, transformando exemplos de entrada e preservando os rótulos associados. Nesse cenário, translações e rotações estão entre as transformações mais comuns aplicadas em pequenas regiões de imagens na tentativa de produzir novas instâncias úteis (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). As técnicas de aumento de dados foram desenvolvidas para oferecer suporte a conjuntos insuficientes de dados.

Outra forma de fazer a convergência em tarefas com conjuntos insuficientes de dados é mediante a utilização da técnica de transferência de aprendizado (*Transfer Learning – TL*), amplamente utilizada nos casos em que a quantidade de dados de treinamento é pequena. A abordagem se concentra em transferir o conhecimento aprendido de uma base de dados maior e de mesmo contexto para melhorar o aprendizado de uma tarefa-alvo relacionada, como mostra a Figura 10.

Figura 10. Processo de transferência de aprendizado



Fonte: elaboração própria (2021).
Obs.: Na ilustração, o sistema de aprendizado 2 recebe um conhecimento adicional do sistema de aprendizado 1, sendo que ambos têm tarefas com contextos relacionados.

De forma complementar, o *dropout* se tornou uma técnica popular que visa reduzir a complexidade da rede e, consequentemente, evitar o *overfitting* (SRIVASTAVA; SALAKHUTDINOV; HINTON, 2013). Essa abordagem consiste em multiplicar a saída de cada unidade da camada oculta por um vetor de variáveis aleatórias independentes de uma distribuição de Bernoulli (HOGG; MCKEAN; CRAIG, 2019). Dessa forma, a arquitetura é treinada para cada dado de entrada e parte das unidades são inibidas (as saídas são multiplicadas por zero) para que não participem do processo de retropropagação. Durante a fase de teste, todas as unidades são consideradas. Os autores afirmam que o *dropout* forma uma combinação de diferentes arquiteturas, uma vez que o conjunto de unidades ativas cria configurações diferentes a cada iteração, e reduz a dependência entre as unidades que podem ser criadas pelo algoritmo de retropropagação.

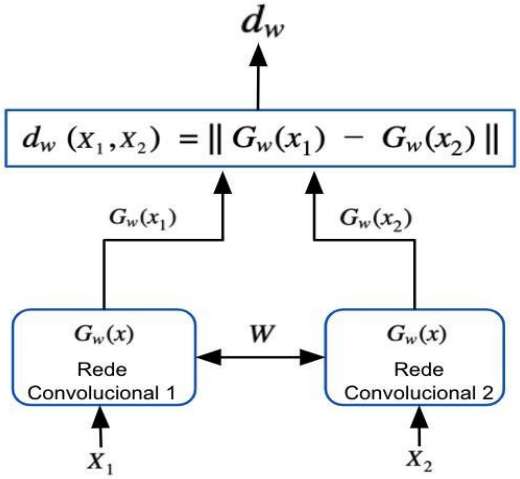
Em resumo, o aumento de dados e o *dropout* são empregados em arquiteturas de aprendizado profundo a fim de reduzir o *overfitting*, oferecendo garantias de convergência. Ambas as técnicas, no entanto, aumentam o tempo de processamento e requerem mais parâmetros para executar o treinamento. Nesse sentido, o aprendizado métrico profundo vem se destacando nos últimos anos por criar arquiteturas que podem ser independentes de grandes volumes de dados. A ideia é usar pares de imagens correspondentes, bem como métricas de distância para treinar uma arquitetura profunda, conhecida como Rede Neural Siamesa, a qual pode ser implementada com a utilização de CNNs.

2.2.4 Redes Neurais Siamesas

A Rede Neural Siamesa consiste em empregar redes gêmeas que aceitam entradas distintas, mas que são unidas por uma função de energia em seu topo (KOCH; ZEMEL; SALAKHUTDINOV, 2015). A prática pode ser compreendida como uma arquitetura de rede neural com duas sub-redes simétricas que utilizam parâmetros compartilhados. Ao final, utiliza-se uma função que calcula uma métrica de distância entre as representações de cada sub-rede. A arquitetura siamesa foi proposta pela primeira vez na utilização de um par de redes neurais homogêneas que mensurava a similaridade entre imagens de entrada para uma tarefa de verificação de assinatura (BROMLEY *et al*., 1993). Mais recentemente, a Rede Siamesa foi modificada para ser utilizada em outras aplicações, tais como: a classificação de texto (NECULOIU; VERSTE-EGH; ROTARU, 2016); o reconhecimento de digitais (ZHONG; YANG; DU, 2018), entre outras. Em Wang *et al*. (2015), a rede foi usada para biometria facial, medindo a similaridade entre faces.

Um exemplo de Rede Neural Siamesa é apresentado na Figura 11, onde X_1 e X_2 formam pares de imagens de entrada. A arquitetura retorna duas representações ($G_w(x)$), que são vetores de características extraídos das imagens de entrada. Juntamente com a representação, o rótulo y é levado em consideração no aprendizado supervisionado da rede, onde y é um rótulo binário, sendo que $y = 1$ determina que as imagens são semelhantes entre si, caso contrário $y = 0$, e W representa o compartilhamento de parâmetros. Finalmente, o sistema calcula a similaridade entre imagens por meio de uma métrica de distância d_w.

Figura 11. Visão geral de uma Rede Neural Siamesa



Fonte: elaboração própria com base em Koch, Zemel e Salakhutdinov (2015).

2.2.4.1 Métricas de distância

O aprendizado métrico é uma abordagem diretamente relacionada às métricas de distância que têm por objetivo estabelecer a similaridade ou dissimilaridade entre objetos (KAYA; BILGE, 2019). Atualmente, ampla variedade de métricas de distância são empregadas na literatura com o objetivo de reduzir a distância entre objetos similares, bem como aumentar a distância entre objetos dissimilares. Estudos atuais na área de métricas profundas estão diretamente relacionados a distâncias, tais como: Euclidiana e o Cosseno. Nesse sentido, antes de detalhar algumas das principais funções de distância de similaridade, vale ressaltar que elas têm algumas propriedades em comum (BELLET; HABRARD; SEBBAN, 2013).

Formalmente, uma função de distância é denotada por $D: \chi \times \chi \rightarrow \mathbb{R}$, que atribui um número com valor real a qualquer par de pontos do espaço de entrada $x_i, x_j \in \chi$, e que obedece às três propriedades a seguir:

- Identidade: $D(x_i, x_j) = 0$ se $x_i = x_j$;
- Simetria: $D(x_i, x_j) = D(x_j, x_i)$;
- Desigualdade Triangular: $D(x_i, x_j) + D(x_j, x_k) \geq D(x_i, x_k)$.

Geralmente, uma função de distância não obedece necessariamente a todas essas propriedades, por exemplo, ao se permitir $D(x_i, x_j) = 0$ para $x_i \neq x_j$, termina-se com uma pseudométrica. Ou seja, ao se omitir a simetria ou a desigualdade triangular, ainda assim se usa o termo geral de função de distância. Um conceito intimamente relacionado à função de distância é uma função de similaridade. A Equação 2.14 permite visualizar que uma função de similaridade está inversamente relacionada a uma função de distância.

$$D(x_i, x_j) = e^{-S(x_i, x_j)} \tag{2.14}$$

onde $D(x_i, x_j)$ é a função de distância e $S(x_i, x_j)$ é a função de similaridade. Ao se assumir que a função de similaridade é delimitada no intervalo de [0, 1], outra transformação amplamente usada é apresentada na Equação 2.15:

$$D(x_i, x_j) = 1 - S(x_i, x_j) \tag{2.15}$$

Embora se utilize o termo de função de distância ao longo desta tese, deve ficar claro que algoritmos que aprendem funções de distância também são usados em funções de

similaridade. Vários autores exploram a ideia de desenvolver algoritmos para aprender funções de distância a partir de dados de treinamento rotulados. Nesse sentido, passa-se a fornecer uma breve descrição de algumas funções de distância encontradas na literatura e utilizadas no aprendizado de métricas de similaridade.

- **Distância Euclidiana**: uma das funções de distância amplamente utilizada (que também é uma métrica) é definida na Equação 2.16:

$$D_{Euclidiana}(x_i, x_j) = \sqrt{(x_i - x_j)^2} = \sqrt{\sum_{k=1}^{d}(x_{ik} - x_{jk})^2} \tag{2.16}$$

- **Distância Mahalanobis**: a distância de Mahalanobis é uma generalização da distância Euclidiana que também leva em consideração as correlações do conjunto de dados e é invariável à escala. Na sua forma original, mede a distância de pares de vetores com base no pressuposto de que eles se originam da mesma distribuição subjacente. Formalmente, dada uma distribuição *p*, da qual se tem uma matriz de covariância S, Mahalanobis mensura a distância entre dois pontos x_i e x_j pela Equação 2.17:

$$D_{Mahalanobis}(x_i, x_j) = \sqrt{(x_i - x_j)^T S^{-1}(x_i - x_j)} = \sqrt{\sum_{k=1}^{d}\sum_{l=1}^{d} x_{ik} \sum_{kl}^{-1} x_{jl}} \tag{2.17}$$

  É possível observar que se a matriz de covariância S é uma matriz identidade, a distância de Mahalanobis se torna a distância Euclidiana. Se, porém, a matriz de covariância for restringida a ser diagonal, tem-se uma distância Euclidiana normalizada.

- **Distância Manhattan**: originalmente, essa distância foi proposta por Hermann Minkowsky (MINKOWSKY, 1910), sendo a sua definição apresentada na Equação 2.18:

$$D_{Manhattan}(x_i, x_j) = \sum_{k=1}^{d} |x_{ik} - x_{jk}| \tag{2.18}$$

  A métrica mede a distância, necessária para caminhar entre os dois pontos x_i e x_j. Essa função é análoga a “quarteirões de uma cidade”, onde a distância está disposta em quarteirões. Mais formalmente, é a soma dos comprimentos das projeções dos segmentos de linha entre os pontos nos eixos de um sistema de coordenadas.

- **Similaridade de cosseno**: é uma métrica de semelhança amplamente usada para agrupar dados direcionais (que lidam com vetores unitários), isto é, mede apenas a direção relativa

entre pares de vetores. Mais formalmente, ela pode ser denotada pela Equação 2.19:

$$Similaridade(x_i, x_j) = \frac{x_i, x_j}{||x_i|| \cdot ||x_j||} = \sum_{k=1}^{d} \frac{x_{ij} \cdot x_{jk}}{\sqrt{\sum_{k=1}^{d} x_{ik}^2} \cdot \sqrt{\sum_{k=1}^{d} x_{jk}^2}} \quad (2.19)$$

Nesse contexto, a distância entre dois pontos correlatos é apresentada na Equação 2.20:

$$D_{cosseno}(x_i, x_j) = 1 - cos^{-1}(Similaridade(x_i, x_j)) \quad (2.20)$$

Após breve síntese sobre métricas de distância que podem ser empregadas na Rede Neural Siamesa passa-se a detalhar alguns pontos complementares na utilização dessas métricas em tarefas de reconhecimento de objetos.

2.2.4.2 Treinamento e teste

O treinamento da rede neural siamesa é realizado utilizando pares de amostras positivos e negativos. Um par positivo significa que duas imagens pertencem à mesma categoria. Caso contrário, considera-se negativo.

A distância entre pares positivos e negativos é calculada por meio de uma função de distância. Para isso, denota X_1 e X_2 como par de entrada do conjunto de treinamento, cuja distância pode ser calculada por qualquer métrica apresentada nesta Seção, conforme Equação 2.21:

$$d_{\mathrm{w}}\ (X_1, X_2) = ||G_{\mathrm{w}}(x_1) - G_{\mathrm{w}}(x_2)|| \quad (2.21)$$

em que $G_w(x_1)$ e $G_w(x_2)$ são geradas como novas representações dos pares de imagens de entrada e d_w é usado para calcular a distância entre duas entradas. Finalmente, uma função de perda *Loss* (Equação 2.22) é responsável por garantir a aproximação dos pares de imagens da mesma classe e afastar aquelas diferentes (KOCH; ZEMEL; SALAKHUTDINOV, 2015).

$$Loss = -[y \log(d_{\mathrm{w}}) + (1 - y) \log(1 - d_{\mathrm{w}})] \quad (2.22)$$

onde y é a função de indicação binária que denota se os pares de imagens são similares ou não.

Finalmente, para testar a Rede Neural Siamesa, um conjunto de imagens é organizado em vetores de características $C_i = (a_{1,i}, a_{2,i}..., a_{n,i})$ e $T_j = (a_{1,j}, a_{2,j}..., a_{n,j})$, onde C_i são vetores das imagens de referências de uma categoria i, T_j é o vetor de teste, e n é a quantidade de características extraídas de cada imagem. O teste deve ser comparado com todos os vetores das

referências do conjunto de dados para obter a classificação dos resultados. Os resultados, por sua vez, devem ser ordenados de acordo com as distâncias entre as amostras da métrica de similaridade escolhida.

2.3 CONSIDERAÇÕES FINAIS

Este capítulo apresentou as duas principais arquiteturas de rede utilizadas nesta tese – as Redes Neurais Convolucionais (CNN) e as Redes Neurais Siamesas (SNN). Nesse contexto, foram explorados conceitos como: parâmetros, tamanhos de máscaras convolucionais e número de unidades utilizadas na arquitetura da rede. Além disso, foram descritos os principais algoritmos de otimização da CNN, inclusive as suas técnicas para redução da chance de *overfitting*. As Redes Neurais Siamesas também foram detalhadas, uma vez que são empregadas para realizar o aprendizado por meio de uma métrica do cálculo de similaridade.

3 ESTADO DA ARTE

Neste capítulo são discutidos os principais estudos publicados pela literatura quanto aos aspectos relevantes e correlatos do reconhecimento de plantas por meio da imagem digital do componente folha. A grande maioria das pesquisas relacionadas está focada no processo de aquisição, pré-processamento, extração de características e aprendizado supervisionado de classificadores (COPE *et al*., 2012). Nesse contexto, as abordagens de extração de características de folhas utilizam, geralmente, informações sobre forma (ZHAO *et al*., 2015), textura (KADIR *et al*., 2014; CHAKI; PAREKH; BHATTACHARYA, 2015) e venação (LEE *et al*., 2017a) na construção de classificadores monolíticos ou *ensembles*, seguindo diferentes técnicas de indução.

O Congresso Mundial de Reconhecimento de Objetos ImageCLEF (Fórum de Avaliação de Linguagem Cruzada – *Cross Language Evaluation Forum*) tem demonstrado tendência ao uso de modelos profundos desde 2014, quando um dos participantes do Congresso explorou as Redes Neurais Convolucionais. O vencedor da tarefa de identificação de plantas PlantCLEF 2014 (CHEN *et al*., 2014) foi o único participante do congresso que utilizou um método baseado em aprendizagem profunda, treinando o modelo AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) para classificar 500 espécies de plantas, superando todos os participantes com grande margem. Após essa conquista bem-sucedida e os novos desenvolvimentos em aprendizado profundo, o congresso de 2015 foi dominada por sistemas baseados na mesma abordagem para fazer a classificação de 1.000 espécies de plantas (GOËAU; BONNET; JOLY, 2015). Desde então, vários pesquisadores têm utilizado soluções relacionadas a Redes Neurais Convolucionais.

O aprendizado profundo inovou na classificação de plantas, especialmente no que diz respeito à classificação dos problemas de granularidade fina, da qual as CNNs extraem características de alta discriminação entre subcategorias, o que tem sido importante para diferenciar espécies de plantas dentro dos diferentes gêneros existentes. Para corroborar, na competição PlantCLEF 2016, apenas oito dos 94 grupos de pesquisa obtiveram sucesso na tarefa de reconhecimento de plantas, sendo que todos empregaram arquiteturas CNNs (GOËAU; BONNET; JOLY, 2016).

Com a taxonomia biológica presente nas espécies de plantas, alguns autores têm explorado soluções de classificação hierárquica. A motivação está na redução do número de

classes em cada estágio da hierarquia, minimizando a complexidade do problema. Da mesma forma, métricas de similaridade estão sendo utilizadas no contexto de granularidade fina.

Neste capítulo faz-se o levantamento de estudos relacionados ao reconhecimento de espécies de plantas por meio da imagem da folha, com destaque às soluções de classificadores monolíticos (Seção 3.1), *ensembles* (Seção 3.2), modelos profundos (Seção 3.3), soluções encontradas no Congresso PlantCLEF (Seção 3.4) e soluções hierárquicas (Seção 3.5). Para finalizar o capítulo, apresenta-se uma tabela resumida dos principais estudos do estado da arte.

3.1 CLASSIFICADORES MONOLÍTICOS

No processo manual de reconhecimento, botânicos usam diferentes características para descrever as plantas a partir do componente folha, tais como: forma, textura, cor e estrutura das veias (PRASAD; KUDIRI; TRIPATHI, 2011). Os estudos apresentados como classificadores monolíticos utilizam um único e exclusivo modelo de classificação para resolver o problema de reconhecimento automático de folhas de plantas.

Normalmente, as características baseadas em forma consideram o contorno da planta e negligenciam as informações contidas no seu interior. O descritor de forma considera uma sequência de valores calculados em pontos obtidos ao redor do contorno de um objeto, iniciando em algum ponto e traçando o contorno no sentido horário ou anti-horário. Mouine, Yahiaoui e Verroust-Blondet (2013b) investigaram duas abordagens triangulares multiescala para a descrição da forma da folha: representação da área triangular (*Triangle Area Representation* – TAR) e representação do comprimento do lado do triângulo (*Triangle Side Lengths Representation* – TSL). O descritor TAR é calculado com base na área dos triângulos formados por pontos no contorno da forma, os quais fornecem informações sobre as propriedades da forma, como a convexidade ou a concavidade em cada ponto de contorno, o que permite alta capacidade de discriminação.

Além das duas abordagens triangulares em escala múltipla, Mouine, Yahiaoui e Verroust-Blondet (2013a), em um trabalho posterior, propuseram mais duas representações que denotam ângulos orientados a triângulo (*Triangle Represented by two Oriented Angles* – TOA), comprimentos de lados de triângulo e representação angular (*Triangle Represented by two Side Lengths and an Angle* – TSLA). O TOA usa apenas valores de ângulo para representar um triângulo, cuja orientação fornece informações sobre concavidades e convexidades locais, enquanto o TSLA é um descritor de contorno triangular em múltiplas escalas que descreve os triângulos por seus comprimentos e ângulos.

Mouine, Yahiaoui e Verroust-Blondet (2013a) avaliaram as suas abordagens em três bases de dados conhecidas na literatura – primeiro, a base de dados Swedish (SÖDERKVIST, 2001), que contém 1.125 imagens de folhas uniformemente distribuídas em 15 espécies; segundo, o conjunto de dados Flavia (WU *et al*., 2007), composto de 1.907 imagens de folhas pertencentes a 32 espécies; e, por fim, no Congresso ImageCLEF2011 (GOËAU *et al*., 2011), dois subconjuntos de dados, ou seja, o conjunto de dados "*scan*", que contém 4.870 imagens para treinamento e 1.760 imagens para teste, e o subconjunto "*scan-like*", que consiste em 1.819 imagens para treinamento e 907 imagens para teste. Os melhores resultados para as bases de dados foram atingidos com as abordagens TSLA e TSL para Swedish, com 96,53% e 95,73%, respectivamente. Já para Flavia, o TSLA obteve 69,93% de acurácia, enquanto na base de dados ImageCLEF2011, o TOA apresentou o melhor resultado, com 0.53 para a categoria "*scan*" e 0.63 para "*scan-like*", utilizando a métrica ImageCLEF2011 (GOËAU *et al*., 2011).

Algumas críticas sobre a possibilidade de utilizar apenas o contexto de forma na classificação de plantas é relatado no estudo de Zhao *et al.* (2015), em que os autores fazem duas observações sobre o descritor da forma utilizada por eles, denominado Contexto de Forma de Distância Interna (*Inner-Distance Shape Context* – IDSC). Primeiramente, verificaram que este descritor não consegue modelar, suficientemente, detalhes locais da folha, visto que o algoritmo é calculado com base em todos os pontos do contorno de uma maneira híbrida, na qual a informação global domina o cálculo. Como resultado, duas espécies distintas de plantas com formas globalmente semelhantes tendem a ser classificadas como da mesma espécie, mesmo apresentando detalhes locais diferentes.

Em segundo lugar, a estrutura de correspondência de pontos dos métodos de classificação de forma genérica não funciona bem para folhas que contêm formas ramificadas. Para resolver esse problema, Zhao *et al*. (2015) propuseram um recurso independente do IDSC, chamado I-IDSC que, ao invés de calcular informações globais e locais de maneira híbrida, as calcula de forma independente para que diferentes aspectos de uma forma de folha possam ser examinados individualmente. Os autores argumentam que, em comparação com IDSC, a vantagem do I-IDSC é tripla: (1) discrimina folhas com forma geral semelhante, mas com margens diferentes e vice-versa; (2) classifica com precisão as folhas com morfologias simples e ramificadas; e (3) mantém apenas as informações mais discriminativas, podendo ser computada de maneira mais eficiente. Como resultado, Zhao *et al.* (2015) confirmaram esses argumentos em cinco bases de dados: Swedish, ICL, Smithsonian, Plumbers Island e uma base de dados construída por eles. A base de dados ICL contém 6.000 imagens de folhas de 200 espécies, cada uma com 30 amostras; Smithsonian contém no total 343 imagens de folhas de

93 espécies; Plumbers Island consiste de 7.313 imagens de folhas de 249 espécies; e a base de dados construída por eles contém 279 imagens de folhas de 54 espécies.

Outra abordagem de forma foi utilizada por Wang *et al.* (2015), que desenvolveram um descritor chamado "Altura do Arco Multiescala" (*Multiscale-arch-height* – MARCH), construído com base nas medidas côncavas e convexas de arcos de vários níveis. Esse método extrai recursos de altura de arco hierárquico em diferentes extensões de acordes de cada ponto de contorno para fornecer um descritor de forma compacta e de várias escalas. Os autores afirmam que MARCH tem as seguintes propriedades: invariante à escala e rotação de imagens, compacidade e baixa complexidade computacional. O desempenho do método proposto foi avaliado e demonstrou ser superior às abordagens de IDSC (ZHAO *et al*., 2015) e TAR (MOUINE; YAHIAOUI; VERROUST-BLONDET, 2013b). MARCH foi avaliado nas bases de dados mencionadas anteriormente, e obteve resultado de 96,21%, 85,31%, 73,00% e 54,8% para Swedish, Flavia, ICL e ImageCLEF2012 (GOËAU et al., 2012), respectivamente, na categoria "*scan*".

A fim de explorar a estrutura de venação das plantas, Charters *et al*. (2014) propuseram um descritor chamado EAGLE para caracterizar os padrões de borda de folha dentro de um contexto espacial. O referido descritor também explora a estrutura vascular da folha, cujos padrões de borda entre as regiões vizinhas caracterizam a estrutura geral da venação e são representados em um histograma de relações angulares. Em combinação com o descritor *Speeded Up Robust Features* (SURF), os descritores estudados são capazes de fazer a fusão de seus gradientes e de caracterizar o gradiente local e os padrões de venação formados pelas bordas envolvidas. Resultados preliminares obtidos no banco de dados de imagens de folhas Swedish demonstraram que o descritor EAGLE é capaz de aumentar o desempenho do descritor local SURF efetivo em 6%.

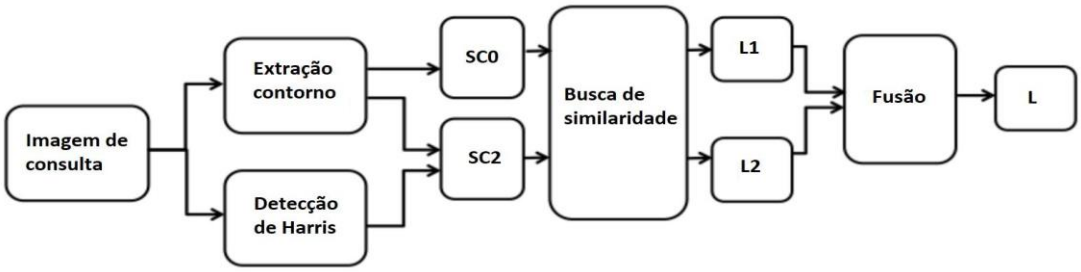
Com o objetivo de entender as características das veias das folhas, Larese *et al*. (2014) introduziram uma estrutura para identificar três espécies de leguminosas com base nas características das veias foliares. Os autores calcularam 52 características por folha (por exemplo, o número total de bordas, o número total de nós, o comprimento total da folha, o comprimento mediano, mínimo e máximo e a largura da veia). Os autores definiram e discutiram cada medida proposta e utilizaram quatro classificadores alternativos como Máquina de Vetores de Suporte (*Support Vector Machines* – SVM), de Kernel Linear e Gaussiano, Análise Discriminante Penalizada (*Penalized Discriminant Analysis* – PDA) e Árvores Aleatórias (*Random Florest* – RF). Os experimentos foram realizados a partir do uso de imagens que foram limpas com um processo químico (realçando o contraste das veias das folhas), o que

aumentou a sua precisão de 84,1% para 88,4% em comparação com imagens não processadas quimicamente. A base de dados avaliada consistiu de 433 espécies, sendo 211 de plantas de soja, 136 de feijão vermelho e 86 de feijão branco.

3.2 ENSEMBLES

As combinações de diferentes características extraídas de um objeto são frequentemente empregadas em sistemas de classificação com o objetivo de imitar a percepção humana. Por exemplo, no problema de reconhecimento de plantas, enquanto a forma da folha pode ser suficiente para distinguir algumas espécies, outras podem ter formas de folhas muito semelhantes às demais, mas ter padrões de textura distintos. A partir disso, Mouine, Yahiaoui e Verroust-Blondet (2013c) realizaram novos experimentos com o objetivo de fazer a combinação de dois descritores baseados em contexto de forma. A Figura 12 apresenta a arquitetura utilizada pelos autores.

Figura 12. Arquitetura proposta no estudo de Mouine, Yahiaoui e Verroust-Blondet (2013c)



Fonte: Mouine, Yahiaoui e Verroust-Blondet (2013c).

Na Figura 12, quando uma imagem de consulta é dada como entrada, é feita a extração de características a partir da extração de contorno e detector de Harris. O cenário SC2 proposto representa a extração do contorno juntamente com a aproximação dos pontos salientes da folha, computados pela detecção de Harris. O cenário SC0 captura as relações espaciais entre as pontas do contorno da folha. Para SC0 e SC2, uma técnica de busca de similaridade aproximada, baseada em um método de Hashing Sensível à Localidade (*Locality Sensitive Hashing* – LSH), é utilizada para verificar as distâncias das características extraídas, sendo L1 e L2, respectivamente, as listas de imagens mais semelhantes à imagem computada de entrada. Por fim, uma combinação de SC0 e SC2 retorna um *ranking* L de probabilidades. Os resultados mostram que uma combinação de margem e forma melhorou o desempenho da classificação avaliada nas bases de dados ImageCLEF nos anos de 2011 e 2012. Utilizando a métrica

ImageCLEF (GOËAU *et al*., 2011), em 2011 a combinação atingiu 0.78 para a categoria "*scan*" e 0.70 para "*scan-like*"; em 2012, 0.58 e 0.61 para "*scan*" e "*scan-like*", respectivamente.

Elhariri, El-Bendary e Hassanien (2014) estudaram propriedades estatísticas de primeira e segunda ordem de textura. As propriedades estatísticas de primeira ordem são: intensidade média, contraste médio, suavidade, assimetria do histograma de intensidade, uniformidade e entropia dos histogramas de intensidade em escala de cinza (*Grayscale Intensity Histograms* – GIH). As estatísticas de segunda ordem, também conhecidas como estatísticas de matriz de co-ocorrência em nível de cinza (*Gray Level Co-occurrence Matrix* – GLCM) são bem conhecidas pela análise de textura e definidas por uma imagem que define a distribuição de valores co-ocorrentes em determinado deslocamento. Os autores descobriram que o uso de propriedades estatísticas de primeira e segunda ordem de textura combinada melhorou a acurácia da classificação em comparação com o uso de propriedades individuais. Resultados preliminares mostraram que a combinação com o uso do classificador (*Linear Discriminant Analysis* – LDA) alcançou precisão de classificação de 92,65% em uma base de dados criada pelos autores com 340 imagens divididas em 30 espécies.

O estudo de Ghasab *et al*. (2015) empregou uma classificação chamada "Otimização em Colônia de Formigas" (*Ant Colony Optimization* – ACO) como um algoritmo de tomada de decisão, o qual foi empregado para investigar as melhores características discriminantes dentro de um espaço de busca. Os autores utilizaram características como forma, morfologia, textura e cor, e fizeram diversas combinações entre essas características para encontrar o melhor conjunto. As características selecionadas foram classificadas pela Máquina de Vetor de Suporte (*Support Vector Machine* – SVM). A eficiência do sistema foi testada em cerca de 2.050 imagens de folhas coletadas em duas bases de dados de plantas diferentes – Flavia e FCA – obtendo, respectivamente, 96,25% e 94,81% de acurácia nas bases de dados.

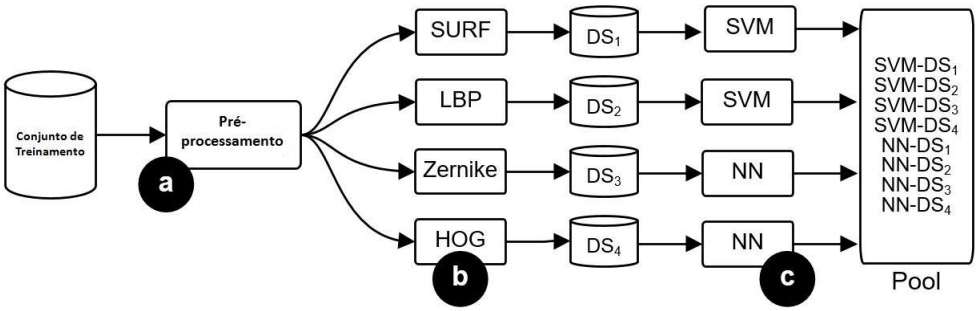
Além da combinação de classificadores em nível de características, há, também, a possibilidade de se fazer a combinação em nível de classificadores, a exemplo do estudo de Chaki, Parekh e Bhattacharya (2015). Os autores empregaram dois tipos de classificadores neurais: um *perceptron* de múltiplas camadas (*Multi Layer Perceptron* – MLP), usando retropropagação (*back-propagation*) e um classificador neuro fuzzy (*Neuro Fussy Classifier* – NFC). A textura da folha foi modelada usando o filtro Gabor e a matriz de co-ocorrência de níveis de cinza (GLCM), enquanto a forma da folha foi capturada com o uso de um conjunto de coeficientes de transformada de *Curvelet*, juntamente com momentos invariantes. As acurácias obtidas apenas por descritores baseados em textura são 81,6% com NFC e 87,1% com MLP. Por outro lado, utilizando apenas descritores baseados em forma, uma precisão

significativamente menor foi encontrada, ou seja, de 50,16% usando NFC e 41,6% usando MLP. Com este cenário, Chaki, Parekh e Bhattacharya (2015) combinaram os descritores de textura e forma, e a precisão da classificação subiu para 97,6% com NFC e caiu para 85,6% com MLP. Os resultados encontrados foram avaliados na base de dados Flavia.

Antes de concluir este subcapítulo relacionado à combinação de classificadores ou *ensembles*, compartilha-se um dos experimentos iniciais, publicado no Congresso *System Man and Cybernetics* (SMC), em 2017, relacionado a abordagens que empregam classificadores monolíticos e *ensembles* para o reconhecimento de folhas de plantas. Neste estudo, Araújo *et al*. (2017) construíram uma abordagem de reconhecimento de folhas de plantas, e avaliaram o impacto de classificadores *ensemble* sobre classificadores monolíticos. Utilizaram, para tanto, quatro diferentes características baseadas em textura e forma: *Local Binary Pattern* (LBP), *Histogram of Gradients* (HOG), *Speed of Robust Features* (SURF) e *Zernike Moments* (ZM). Todos esses tipos de características extraídos são usados para treinar um conjunto (*pool*) de diferentes classificadores, sendo que para obter um *pool* heterogêneo foram usados dois distintos algoritmos de aprendizado: *Support Vector Machine* (SVM) e *Neural Network* (NN).

Uma visão geral da metodologia proposta por Araújo *et al*. (2017) é apresentada na Figura 13, cujas imagens iniciais se referem ao estágio de pré-processamento (Figura 13a), que visa remover o fundo ruidoso da imagem e as estruturas indesejadas, como o caule da folha. O próximo passo é a extração de características (Figura 13b), realizada por quatro diferentes descritores, o que gera quatro representações de características para cada imagem de treinamento. Em seguida, cada classificador é treinado em uma das quatro representações de recurso (Figura 13c), resultando em um conjunto formado por oito classificadores.

Figura 13. Visão geral da metodologia proposta em Araújo *et al.* (2017)



Fonte: Araújo *et al.* (2017).
Notas:
(a) pré-processamento
(b) extração de características e geração do vetor de características
(c) treinamento do classificador e *pool* de classificadores

Uma vez que todos os classificadores são treinados, a estratégia de busca avançada (*Search Forward Selection* – SFS) (ROLI; GIACINTO; VERNAZZA, 2001) foi empregada para selecionar estaticamente um subconjunto de classificadores (*ensemble*) do *pool* original com o objetivo de melhorar a precisão de reconhecimento. Finalmente, os classificadores selecionados foram combinados durante o estágio de teste para designar uma espécie (rótulo) a uma instância de entrada.

Na Estratégia de Busca Avançada (SFS), a ideia é separar os classificadores com base na precisão e, em seguida, selecionar o melhor deles para participar do *ensemble*. Após isso, cada um dos sete classificadores restantes é combinado com o primeiro, usando uma regra de combinação apropriada (por exemplo, voto majoritário, máxima, média, etc.). Se for observado um aumento na acurácia em relação ao conjunto anterior, o classificador que fornece a maior precisão é inserido no conjunto, caso contrário o classificador é descartado. Ao final do processo de seleção, tem-se um conjunto de classificadores, cujo número depende daqueles que, ao serem combinados, alcancem o mais alto desempenho.

Como resultado, nas duas bases de dados – ImageCLEF 2011 e 2012 –, utilizadas no estudo de Araújo *et al*. (2017), verificou-se que a complementaridade de classificadores se destaca em relação aos demais utilizados individualmente no reconhecimento de plantas. Na Tabela 1 compara-se as matrizes de confusão do melhor classificador monolítico e o *ensemble* para a categoria de "*scan*" do conjunto de dados ImageCLEF 2011. Este cenário foi escolhido por permitir melhor interpretação visual, dado o seu pequeno número de classes.

Como se pode observar, a combinação de recursos de textura e forma (*ensemble*) aumenta o desempenho de todas as classes em relação ao melhor classificador individual. Por exemplo, percebe-se redução significativa nas confusões entre as classes G (*Ilex aquifolium*) e M (*Quercus ilex*). Para confirmar tal afirmação, a Figura 14 mostra a diferença de texturas entre as imagens das classes G (Figura 14b) e M (Figura 14a). Dessa maneira, descritores de forma, se utilizados de maneira individual, não seriam discriminantes o suficiente para distinguir as duas espécies.

Tabela 1. Matrizes de confusão para o melhor classificador monolítico (ZM + NN) e o melhor *ensemble* gerado para a categoria "*scan*" do conjunto de dados ImageCLEF 2011

Classificador Monolítico

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	3	0	1	0	0	5	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	1
C	0	0	10	0	1	0	0	0	0	0	0	0	0
D	0	0	0	22	3	1	0	0	0	0	0	0	3
E	0	0	0	0	3	0	0	0	2	0	0	0	2
F	1	0	0	0	0	18	0	0	0	0	0	0	0
G	0	0	0	1	1	0	16	0	0	0	0	0	8
H	0	0	0	0	0	0	0	21	0	0	1	0	0
I	0	0	0	0	0	0	0	0	2	0	0	0	14
J	0	1	0	0	0	0	0	0	1	1	0	0	5
K	0	0	0	0	0	0	0	0	0	0	1	0	0
L	0	0	0	0	0	0	0	0	0	0	0	2	0
M	0	0	0	0	1	0	4	0	0	0	0	0	24

Estratégia de SMC

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	7	0	1	0	0	1	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	1	0	0	0	0
C	0	0	11	0	0	0	0	0	0	0	0	0	0
D	0	0	4	25	0	0	0	0	0	0	0	0	0
E	0	0	0	0	4	0	0	0	0	0	0	0	3
F	1	0	0	0	0	18	0	0	0	0	0	0	0
G	0	0	0	0	0	0	22	0	1	0	0	1	2
H	0	0	0	0	0	0	0	22	0	0	0	0	0
I	0	0	0	0	0	0	0	0	7	0	0	0	9
J	0	0	0	0	0	0	0	0	0	6	0	0	2
K	0	0	0	0	0	0	0	0	0	0	1	0	0
L	0	0	0	0	0	0	0	0	0	0	0	2	0
M	0	0	0	0	2	0	0	0	2	0	0	0	25

Fonte: Araújo *et al*. (2017).

Figura 14. Exemplo de um cenário difícil enfrentado pelas estratégias

(a) (b)

Fonte: Araújo *et al*. (2017).
Legenda:
a) classe M, pertencente à espécie *Quercus ilex*
b) classe G, uma instância da espécie *Ilex aquifolium*

3.3 MODELOS PROFUNDOS

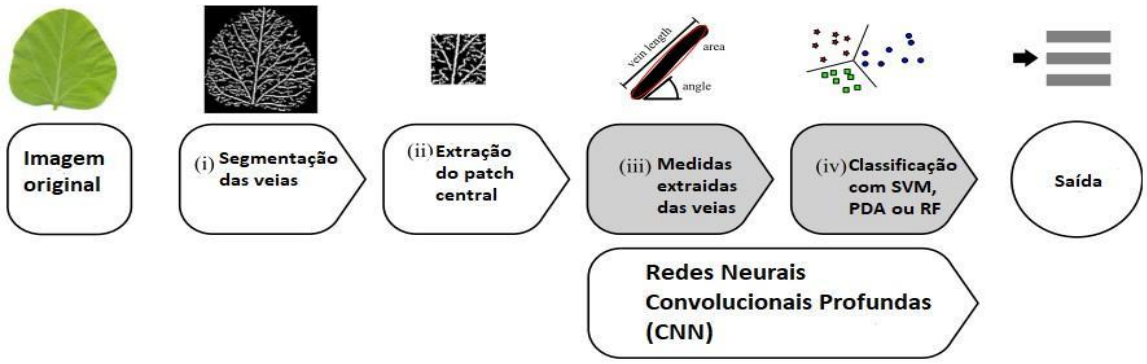
Zhu *et al*. (2018) realizaram extensos experimentos em vários conjuntos de dados de plantas, constatando desempenho notável, aprendendo representações e usando modelos profundos em comparação às abordagens monolíticas ou *ensembles* que, na literatura, são conhecidos como modelos *handcrafted features*. Por conta disso, diversas abordagens utilizando CNN foram avaliadas no contexto de reconhecimento de plantas. Lee *et al.* (2017a) utilizaram modelos CNN para a extração de características, e Redes De-Convolucionais (*Deconvolutional Network* – DN) para plotar características extraídas durante as camadas convolucionais da CNN. Dois tipos de amostragens foliares são avaliados: a) folha inteira, que expõe informações da forma da folha; e b) *patches* de folhas, que traz informações detalhadas, como, por exemplo, as suas veias. Os autores observaram que as características foliares mudam de acordo com a mudança da amostragem da folha.

De acordo com os experimentos preliminares e a visualização dos resultados do estudo

de Lee *et al.* (2017a), a CNN treinada usando folhas inteiras e *patches* foliares extraem diferentes níveis de informação contextual. Com isso, os autores apresentaram arquitetura que combinou esses dois tipos de amostragem, chamando-a de "Fusão Precoce", que realiza o treinamento da rede considerando as duas amostragens e ao final da rede as diferentes saídas são combinadas por uma regra de fusão. O método proposto foi avaliado na base de dados Flavia (WU *et al*., 2007), que consiste em 32 classes e obteve um resultado de 94% de performance.

Tentativa do uso de CNNs na identificação de plantas utilizando padrões morfológicos das veias foi realizado por Grinblat *et al*. (2016). Primeiramente, os autores extraíram os padrões das veias usando o descritor de Transformação de Perda ou Acerto (*Hit or Miss Transform* – UHMT), aplicado para extrair padrões das veias, cuja saída é uma imagem binária. Em seguida, extraíram um *patch* central da imagem original com o objetivo de eliminar possíveis influências da forma da folha. Logo após mensuraram as características das veias, tais como: total de veias encontradas na imagem, média do tamanho das veias, número de nós, entre outras. Na classificação, utilizaram três diferentes algoritmos de aprendizagem de máquina: Máquina de Vetor de Suporte (SVM), Análise Discriminante Penalizada (PDA) e Árvores Aleatórias (RF). A Figura 15 apresenta toda a etapa descrita anteriormente, em que os autores substituíram as etapas (iii) e (iv) por uma classificação que utilizou CNN, avaliando a sua abordagem proposta. O modelo CNN proposto por Grinblat *et al*. (2016) consiste de seis camadas, sendo cinco convolucionais e uma *softmax*. Para o treinamento os autores utilizaram SGD com os seguintes hiperparâmetros: *batch-size* com 20 amostras, *dropout* = 0.5, e taxa de aprendizagem iniciada em 0.01. A base de dados utilizada consistiu em três espécies diferentes de leguminosas: feijão branco, feijão vermelho e soja. O número total de amostras contou com 866 imagens, divididos da seguinte forma: 422 imagens correspondem a folhas de soja; 272 imagens a folhas de feijão vermelho; e 172 a folhas de feijão branco. A classificação do modelo CNN foi de 90,2%, 98,3% e 98,8% para feijão branco, feijão vermelho e soja, respectivamente, performance superior ao protocolo de classificação que utilizou classificadores de bases como SVM, PDA ou RF.

Figura 15. Trabalho de Grinblat *et al.* (2016)



Fonte: Grinblat et al. (2016).
Notas:
Estágio (i) faz o pré-processamento na imagem
Estágio (ii) o elemento de forma é ocultado por meio de um recorte central na folha original
Estágios (iii) e (iv) extração e classificação do método, respectivamente
Obs.: os estágios acinzentados foram substituídos por uma rede convolucional profunda

A combinação de CNNs também tem sido utilizada para aumentar o desempenho de classificação de modelos profundos. No estudo de Ghazi, Yanikoglu e Aptoula (2017), os autores exploraram três arquiteturas CNNs populares: GoogleNet, AlexNet e VGGNet, usando transferência de conhecimento (*transfer learning*) e técnicas de aumento de dados (*data augmentation*). Como resultado, Ghazi, Yanikoglu e Aptoula (2017) demonstraram o impacto do uso da transferência de conhecimento, bem como alguns efeitos relacionados às mudanças de hiperparâmetros das Redes CNNs (iterações, tamanho de lote (*batch-size*), aumento de dados, etc.) no desempenho da classificação. Aplicaram, também, a técnica de aumento de dados nas fases de treinamento e teste, baseando-se na extração de *patches* randômicos da imagem original das folhas e, também, numa rotação na imagem, resultando em um cenário com 80 imagens adicionais. O tamanho das imagens foi redimensionado para 224 x 224 para GoogleNet e 227 x 227 para AlexNet e VGGNet. Uma fusão de soma baseado no *score* de cada imagem adicional foi aplicado com vistas à combinação das saídas de cada instância de uma imagem, bem como dos diferentes modelos de CNN utilizados. Ghazi, Yanikoglu e Aptoula (2017) ainda avaliaram diversos hiperparâmetros da rede CNN, variando o número de iterações da rede de 100 a 500 mil e incrementando o tamanho do *batch* em 20, 40 e 60. A avaliação do método proposto pelos autores foi realizada na base de dados PlantCLEF 2015. Os resultados demonstraram que a elevação do número de iterações de 100 para 500 mil aumentou significantemente a performance de todos os modelos avaliados, mostrando que as redes são resistentes ao *overfitting* devido à arquitetura dos modelos e à aplicação da técnica de aumento de dados. Outra conclusão foi a melhora da precisão com o aumento do tamanho do *batch*, o

que, consequentemente, também aumenta a duração do treinamento. O aumento de dados com 80 amostras adicionais para cada instância elevou a acurácia da validação em relação ao teste realizado com 10 amostras adicionais. Por fim, os autores concluíram que aumentar o número de iterações, o número de amostras e o tamanho do *batch* melhora o desempenho de classificação, embora também haja aumento no custo computacional do modelo.

Outro método baseado na arquitetura CNN foi proposto por Tan *et al*. (2020), chamado *D-Leaf*, cuja arquitetura é composta por seis camadas, das quais três são de convolução (estágio de convolução, ReLU e *pooling*) e três são totalmente conectadas. Para validar a sua abordagem, todas as imagens foram levadas a um método morfométrico convencional que calculou as medidas morfológicas baseadas nas veias das folhas por meio do filtro de Sobel. Após o pré-processamento das folhas, as características foram extraídas a partir de três modelos diferentes de CNN: AlexNet pré-treinada, AlexNet com fine-*tunning* e D-*Leaf*.

Cinco modelos de classificação foram empregados: *Support Vector Machine* (SVM), Rede Neural Artificial (RNA), k-vizinho mais próximo (KNN), Naive *Bayes* (NB) e CNN. O desempenho do método *D-Leaf*, juntamente com os demais modelos foram validados em três conjuntos de dados publicamente disponíveis, que são MalayaKew, Flavia e Swedish, e alcançou uma precisão de teste de 94,88% em comparação aos modelos AlexNet (93,26%) e AlexNet com *fine-tunning* (95,54%). Além disso, os modelos de CNN tiveram melhor desempenho do que as medidas morfométricas tradicionais, portanto, os resultados *D-Leaf* demonstraram eficácia na tarefa de identificação de espécies de plantas.

Mais recentemente, modelos baseados em aprendizado métrico profundo passaram a ser empregados no reconhecimento de objetos, os quais têm por objetivo compreender representações de imagens via métrica de similaridade. Foram encontrados três estudos na literatura, os quais utilizaram métricas de similaridade em seus modelos, empregando Redes Neurais Convolucionais Siamesas (SNN), especificamente na tarefa de reconhecimento de plantas: o estudo de Wang e Wang (2019) utilizou abordagem de aprendizado com poucas amostras de imagens (*few-shot-learning*) baseadas na SNN para reconhecer folhas de plantas; a distância Euclidiana foi usada para mensurar a distância entre representações; e a SNN foi construída com a arquitetura GoogLeNet. O método proposto foi avaliado em três conjuntos de dados de plantas: Flavia, Swedish e LeafSnap com pequeno número de amostras de aprendizado em cada base de dados. Os resultados experimentais mostraram que utilizando apenas 20 amostras de treinamento por classe, a precisão de classificação da abordagem foi de 95,32%, 91,37% e 91,75%, respectivamente, para os conjuntos de dados Flavia, Swedish e Leafsnap.

Em outro trabalho similar, Zhi-Yong *et al*. (2018) propõem nova abordagem chamada

“Rede Siamesa de Estrutura Espacial”, a qual foi usada para aprender a representação de pares semelhantes e dissimilares de imagens. Pares similares foram formados com a mesma categoria de plantas, e pares dissimilares usaram diferentes espécies de plantas. Os autores avaliaram o seu desempenho no Congresso PlantCLEF 2015 utilizando redes neurais recorrentes, e o resultado foi de 0.84, superando todos os métodos concorrentes a partir da métrica *S* disponibilizada pelo referido Congresso.

Recentemente, Figueroa-Mata e Mata-Montero (2020) utilizaram a métrica de similaridade para discriminar espécies de plantas em base de dados desbalanceadas. Os autores avaliaram a possibilidade de os modelos SNN serem melhores do que os CNN em relação ao desempenho e custo computacional. Além disso, novas espécies (20 espécies do conjunto de dados da Costa Rica) nunca vistas pelo modelo SNN foram avaliadas sem treinamento do modelo proposto. Em seu primeiro experimento, os autores concluíram que, para conjuntos de dados com menos de 20 imagens por espécie, o SNN teve melhor desempenho que o CNN, além de melhorar o desempenho em classes desbalanceadas e ter menor custo computacional. O segundo experimento mostrou que o SNN pode ser genérico quando novas espécies de plantas são enviadas para o reconhecimento sem a necessidade de retreinamento do modelo proposto.

3.4 CONGRESSO *PLANTCLEF*

Nesta seção são explorados os estudos que participaram do Congresso mundialmente conhecido na tarefa de reconhecimento de plantas, cujo objetivo foi disponibilizar uma base de dados robusta para que qualquer pesquisador pudesse ter acesso à aplicação de suas metodologias. O Congresso iniciou em 2011 com uma pequena base de dados, contendo poucas amostras e espécies. Recentemente, em seu último evento, em 2019, a base de dados disponibilizada contou com mais de 10.000 espécies de plantas e milhares de imagens.

Nos primeiros eventos do Congresso, realizados de 2011 a 2013, as metodologias consistiam em características básicas (*handcrafted features*), com técnicas de pré-processamento de imagem, extração de características e uso de classificadores individuais (monolíticos) e combinados (*ensemble*). Nesse período, os resultados não alcançaram o crescimento esperado dada as metodologias utilizadas na época, entretanto, em 2014 houve uma mudança neste aspecto. Um grupo de pesquisadores, Chen *et al*. (2014) utilizou Redes Neurais Convolucionais para resolver o problema de reconhecimento de plantas e obtiveram resultado

considerável em relação aos demais pesquisadores que até então utilizavam abordagens tradicionais.

Os autores também testaram classificadores tradicionais utilizando Transformação de Características Invariantes à Escala (*Scale-invariant Feature Transform* – SIFT) para extração de características onde são modeladas com misturas gaussianas e Fisher Vector e, por fim, utilizaram SVM para fazer a classificação. O resultado, entretanto, não foi satisfatório e, por isso, passaram a explorar Redes Convolucionais com o treinamento da rede e utilização de uma rede pré-treinada sobre o conjunto de dados ImageNET (DENG *et al*., 2009).

A CNN gerada tem cerca de 60 milhões de parâmetros e consiste de cinco camadas convolucionais, algumas das quais são seguidas por *max-pooling* e três camadas totalmente conectadas (*fully-connected*) com uma camada final de *softmax*. Essa rede é encontrada na literatura como AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Com o uso dos classificadores tradicionais e CNN, uma fusão das características extraídas em cada método trouxe o melhor resultado do Congresso, obtendo 0.456 de performance em uma base de dados com 500 espécies. Todos os resultados de performance apresentados pelos autores nesta Seção foram obtidos utilizando a métrica *S* do congresso PlantCLEF.

No ano seguinte, em 2015, praticamente todos os métodos submetidos ao Congresso PlantCLEF utilizaram CNNs. Sungbin (2015) utilizou uma nova CNN chamada GoogleNet (SZEGEDY *et al*., 2014), similarmente pré-treinada sob banco de dados ImageNet, com ajuste fino (*fine*) ao conjunto de treinamento de plantas. Sungbin (2015) usou uma estratégia de treinamento que consistiu em dividir aleatoriamente o conjunto de treinamento PlantCLEF em cinco subconjuntos, de modo a obter cinco classificadores CNN complementares, sendo a combinação supostamente mais estável. As saídas para cada CNN foram obtidas combinando os resultados da classificação das imagens com o método de fusão de borda. Na base de 2015, o autor obteve um *score* de 0.667 na classificação das plantas.

No ano de 2016, o Congresso disponibilizou o mesmo conjunto de treinamento usado em 2015, mas aumentou o desafio para o reconhecimento dos dados de teste. Com isso, Hang, Tatsuma e Masaki (2016) utilizaram um sistema baseado no modelo CNN VGGNet (SIMONYAN; ZISSERMAN, 2014), substituindo a última camada de *pooling* por camada de *pooling* de Pirâmide Espacial. Os autores também utilizaram uma técnica de aumento de dados, obtendo um resultado de 0.742 utilizando a métrica *S*.

Há pouco tempo, no Congresso de 2017, a base de dados já continha mais de 1.7 milhão de imagens distribuídas em de 10.000 espécies de plantas. Lasseck (2017) obteve o melhor resultado, atingindo 0.885 de acurácia na tarefa de classificação. O autor usou conjuntos de

CNNs pré-treinados no ImageNet baseados em três arquiteturas (GoogLeNet, ResNet-152 e ResNeXT), cada uma treinada com técnicas de *bagging*. Uma estratégia de aumento de dados multiplicou por cinco o número de imagens de treinamento, cuja técnica consistiu em cortes aleatórios na imagem, inversão horizontal, variações de saturação, brilho e rotação. As imagens de teste também foram aumentadas e as previsões resultantes foram calculadas pela média.

Com o grande avanço nos sistemas automáticos de reconhecimento de plantas, o objetivo do Congresso de 2018 consistiu em fazer uma comparação entre especialistas humanos e sistemas automáticos de reconhecimento. Para os dados de treinamento foram fornecidos todos os conjuntos de dados dos anos anteriores do Congresso PlantCLEF. O conjunto de testes foi obtido após processo colaborativo dos especialistas da área. Ao final, foi selecionado um subconjunto de 75 espécies de plantas, ilustrado por 216 imagens relacionadas a 33 famílias e 58 gêneros.

O melhor resultado do ano de 2018 foi obtido pelo estudo de Šulc, Picek e Matas (2018) que utilizaram seis Redes Neurais Convolucionais (CNNs) com base em duas arquiteturas de ponta (Inception-ResNet-v2 e Inception-v4). As CNNs foram inicializadas com pesos pré-treinados no ImageNet e, em seguida, ajustadas com diferentes hiperparâmetros e com o uso de aumento de dados (rotação horizontal aleatória, distorção de cores e cortes aleatórias para alguns modelos). Um diferencial do trabalho desses autores é que supondo forte distribuição desbalanceada das classes entre o teste e os conjuntos de treinamento, os resultados das CNNs foram ajustados de acordo com estimativa das probabilidades com base em um algoritmo de maximização da expectativa. Um resultado de 0.884 de performance no top-1 durante o Congresso foi obtido por essa equipe com o melhor *ensemble*.

Mais recentemente, em 2019, a principal novidade foi avaliar o reconhecimento automatizado das plantas em regiões com deficiência de dados, focado em um conjunto de 10.000 espécies localizadas, principalmente, na região do Planalto das Guianas e na Floresta Tropical Amazônica, conhecidas por terem uma das maiores diversidades de plantas e animais do mundo. A maioria das equipes considerou que o conjunto de dados de treinamento era muito ruidoso e desbalanceado. Mesmo assim, o melhor resultado obtido foi o trabalho de Chulif *et al.* (2019), que utilizou um conjunto de CNNs com arquiteturas Inception-v4 e Inception-ResNet-v2, e alcançou uma performance top-1 de 0.246 para o conjunto de testes.

Os resultados da edição 2019 do Congresso PlantCLEF foram significativamente inferiores aos das edições anteriores, confirmando a suposição de que a flora tropical é inerentemente mais difícil de reconhecer do que uma flora mais generalista. O Congresso concluiu que o desempenho das Redes Neurais Convolucionais diminuiu devido ao baixo

número de imagens de treinamento da maioria das espécies e ao maior grau de ruído que ocorre nesses dados.

Em resumo, todos os estudos expostos nesta seção atingiram os melhores resultados no Congresso Mundial de Plantas PlantCLEF nos anos de 2014 a 2019. Percebe-se, com isso, um avanço na arquitetura das redes utilizadas em cada estudo, em que o número de parâmetros e camadas aumenta de acordo com a arquitetura utilizada. É possível constatar, também, que a combinação dos modelos CNNs em alguma metodologia é utilizada para potencializar os resultados, entretanto, quanto mais camadas tiver a estrutura de uma rede, maior será o seu tempo de treinamento. O uso de técnicas de aumento de dados também faz com que problemas de classificação em classes desbalanceadas sejam contornados, embora o processo seja moroso.

3.5 ABORDAGENS HIERÁRQUICAS

Abordagens recentes de reconhecimento de plantas estão sendo baseadas na hierarquia taxonômica. Zhu *et al.* (2019) dividem as plantas em grupos hierárquicos, tais como Famílias, Gêneros e Espécies, os quais podem ser consultados recursivamente de acordo com as características discriminativas da planta. A grande maioria dos estudos relacionados e apresentados até o momento utiliza um conjunto solitário ou combinado de características da folha (forma, textura, veias, etc.), além de classificadores tradicionais para discriminar classes de folhas de um conjunto de dados de plantas. Tais abordagens funcionam quando as espécies de plantas no conjunto de dados são, na maior parte, homogêneas e podem ser separadas por um único arranjo de características. Na natureza, porém, as folhas podem ter tipos ilimitados de variedades em arranjos geométricos, formas, matizes e texturas. Adicionalmente, podem apresentar estruturas morfológicas básicas, ramificadas, torcidas ou até mesmo fragmentadas (falta de partes da folha).

Para levar em conta tais variedades heterogêneas, Chaki, Parekh e Bhattacharya (2018) propuseram uma abordagem de classificação arquitetural hierárquica, onde cada nó da estrutura hierárquica utiliza uma característica visual específica da planta, estando conectada a um arranjo de classificadores personalizados. Os resultados de cada nó são combinados para favorecer uma compreensão mais completa das características da folha. Cada característica extraída pelos autores tem o objetivo de fazer o agrupamento de elementos similares e compor o sistema de classificação hierárquica. A extração de textura, por sua vez, tem o objetivo de identificar folhas com padrões de textura proeminentes ou estruturas da veia na superfície da folha. Por outro lado, as características de contorno discriminam folhas simples de outras

ramificadas.

Na morfologia foliar das plantas, as folhas simples são caracterizadas por uma única lâmina de folhas, enquanto as folhas ramificadas têm vários folhetos dentro de uma única unidade. Quando uma imagem de teste é avaliada, o classificador é responsável por verificar se é uma folha simples ou ramificada, formando um classificador hierárquico baseado nas características morfológicas das plantas (CHAKI; PAREKH; BHATTACHARYA, 2018).

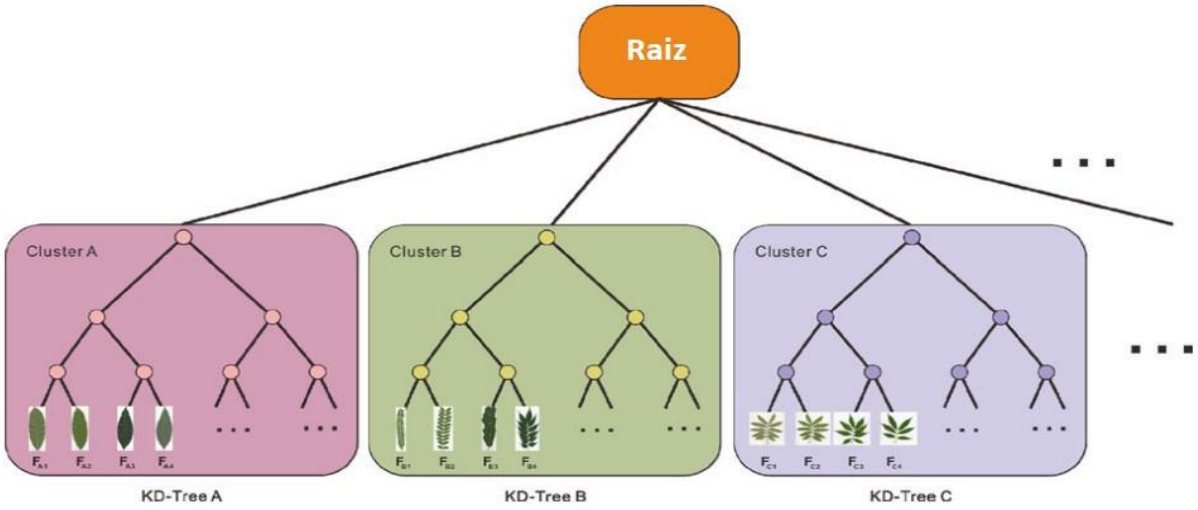
A ideia da classificação hierárquica é decompor o problema original em subproblemas, compartilhando propriedades mais homogêneas. Muitos problemas de classificação do mundo real referem-se à classificação hierárquica, em que as classes previstas são organizadas em estruturas de árvore ou em Gráfico Acíclico Direto (*Directed Acyclic Graph* – DAG). Alguns utilizam a hierarquia de classes, incluindo o compartilhamento de exemplos de treinamento em categorias semelhantes (FERGUS *et al*., 2010) ou combinando informações de diferentes níveis da hierarquia (ZWEIG; WEINSHALL, 2007).

Wu *et al*. (2015) propuseram um modelo de classificação hierárquica baseado na similaridade das características encontradas na base de dados. Os autores dividiram a base de dados em subconjuntos baseados em *clustering* das características extraídas, em que cada característica da folha é indexada paralelamente com árvores hierárquicas de dimensão K (*KD-trees*) para obter uma recuperação eficiente da imagem da folha. A base de dados avaliada pelos autores é do Jardim Botânico do Sul da China, que contém 23.025 espécies. Cerca de 80 imagens de cada espécie foram coletadas sob diferentes condições de posição, rotação, orientação, ângulo de visão e iluminação, portanto, cerca de 1,5 milhão de imagens foliares formaram o grande banco de dados para reconhecimento automático de folhas. Para fazer a extração das características os autores utilizaram Histogramas de Gradientes.

Para a geração da árvore hierárquica, Wu *et al*. (2015) classificaram as folhas que tinham características similares, como forma e textura, por meio do *clustering.* A Figura 16 mostra que os *clusters* foram construídos por *KD-trees*, agrupando as características extraídas $F_1, F_2, ..., F_n$ e obtendo *m clusters* $C_1, C_2, ..., C_m$. Juntamente com as *KD-trees* foi utilizado um algoritmo *Best-Bin-First* (BBF), comumente usado para encontrar uma solução aproximada para o problema de busca do vizinho mais próximo em espaços de alta dimensão. BBF é um algoritmo aproximado que retrocede de acordo com uma fila de prioridades baseada na proximidade.

Os melhores resultados foram atingidos com o classificador K-vizinhos mais próximo (*K-Nearest Neighbors* – KNN) com K = 5, atingindo 90% de acertos na base de dados avaliada.

Figura 16. Hierarquia foliar apresentada por meio da metodologia *KD-tress*

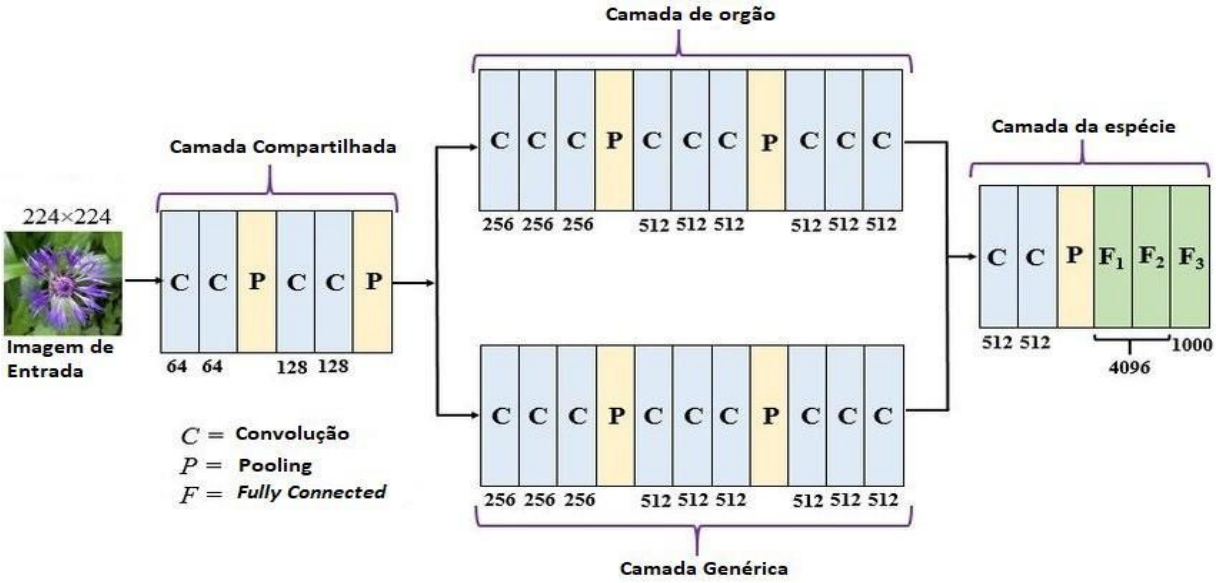


Fonte: Wu *et al*. (2015).

Sfar, Boujemaa e Geman (2015) apresentaram uma estratégia hierárquica para classificar plantas, similar ao estudo de Wu *et al*. (2015). Os autores construíram uma hierarquia estruturada em árvore recursiva construída de baixo para cima (*bottom-up*), fundindo grupos semelhantes, e tratando cada espécie como um *cluster* único no início e depois juntos (ou aglomerados), sucessivamente, em pares de agrupamentos, até que todos estivessem mesclados em um único agrupamento. Para tanto, foi utilizado critério de norma euclidiana, ou seja, em cada passo foi definida a dissimilaridade entre dois clusters. Para este estudo, os autores utilizaram descritores de forma e textura tais como: Hough, EOH, HSV e Fourier, utilizando o classificador SVM. Os experimentos foram avaliados sobre quatro bases de dados: Swedish, Flavia, Smithsonian e ImageCLEF2011.

Os estudos citados permitem construir uma hierarquia das plantas com base nas características similares das folhas. Há, também, estudos que constroem a sua hierarquia fundamentada na informação contida na base de dados. Lee *et al*. (2017a), utilizando esse princípio, propuseram uma arquitetura chamada Rede Neural Convolucional Híbrida de Órgãos Genéricos (*Hybrid Generic-Organ Convolutional Neural Network* – HGO-CNN) que integra Redes Neurais Convolucionais para extrair características das folhas e classificá-las conforme a arquitetura apresentada na Figura 17.

Figura 17. Arquitetura hierárquica proposta por Lee *et al.* (2017a): HGO-CNN



Fonte: Lee *et al.* (2017a).

O HGO-CNN usa uma CNN de dois caminhos com o propósito de extrair características baseadas em órgãos das plantas. Na camada de órgãos são treinados sete tipos de rótulos de órgãos pré-definidos da planta (ramo, planta inteira, flor, fruta, folha, caule ou folha sobre escâner (*leafscan*)). Após treinar a camada de órgão e fazer a sua classificação, passa-se a treinar a camada de espécies com base nos seus rótulos retornadas pelo órgão predito na etapa anterior. A base de dados avaliada por Lee *et al*. (2017a) foi PlantCLEF 2015, que obteve um resultado de 0.80 na categoria *leafscan*. Posteriormente, os autores estenderam o seu estudo, propondo um novo *framework* baseado no HGO-CNN com pequenas alterações como, por exemplo, utilizando redes neurais recorrentes. O nome dado a esse novo *framework* foi Plant-StructNet (LEE; CHAN; REMAGNINO, 2018), que tem como objetivo modelar dependências contextuais de alto nível entre visões das plantas, compreendendo órgãos variados ou diferentes pontos de vista de um órgão similar.

Recentemente, Zhu *et al.* (2019) introduziram um modelo hierárquico CNN de dois estágios, em que o primeiro consiste em reconhecer a família associada à planta, e o segundo utiliza uma estratégia de extração de mapas de calor da imagem para encontrar partes discriminativas entre espécies de plantas. Destarte, os autores utilizaram a taxonomia foliar das plantas para classificar hierarquicamente famílias e espécies de plantas em dois estágios. Ademais, conduziram experimentos em dois conjuntos de dados: Malayakew e ICL, utilizando um modelo CNN Xception, atingindo uma precisão de 99% em ambos os conjuntos. Zhu *et al*. (2019) utilizaram, para tanto, a técnica de aumento de dados e dividiram a sua base em 90% para treinamento e 10% para teste. Concluíram que com o conhecimento prévio dos rótulos da

família, a CNN pode identificar melhor as características biológicas das plantas, além das características visuais mais intuitivas, levando a um melhor desempenho nas tarefas de classificação.

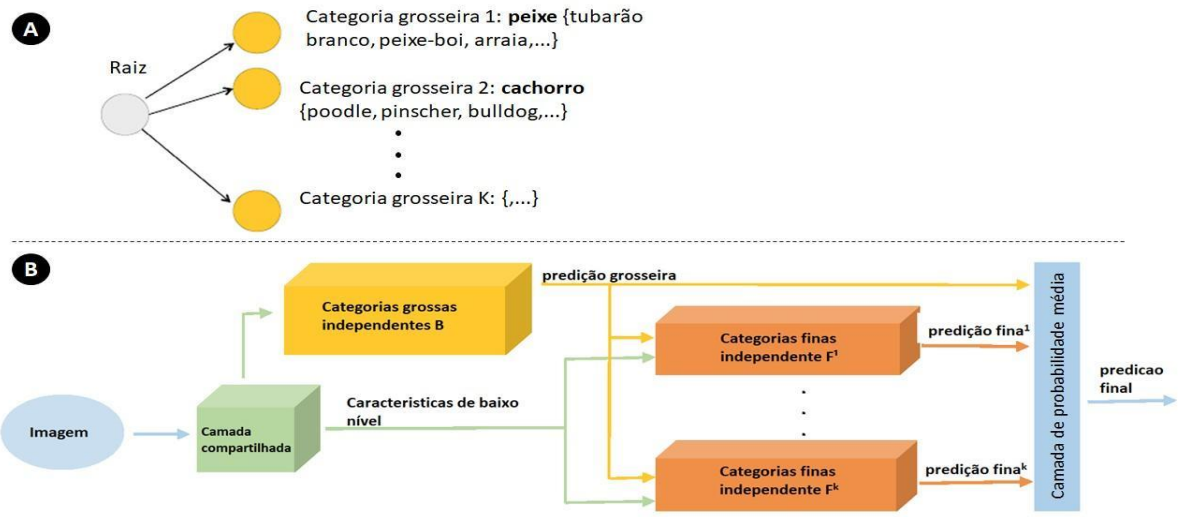
O constante crescimento dos conjuntos de dados de plantas favorece o aumento dos desafios de classificação de subcategorias, isto é, são necessárias características mais específicas na discriminação entre espécies. Pensando nisso, Zhang *et al*. (2020) propuseram um modelo de Aprendizagem Múltipla Hierárquica Profunda (AHMTL) a partir da utilização de um mecanismo de atenção. A base de dados utilizada neste estudo foi composta por apenas um grupo de famílias de plantas, chamadas *Orchid*, dividida em 158 gêneros e 2.608 espécies. Com a utilização da taxonomia foliar da planta, a classificação foi realizada hierarquicamente e separada por grupos com diversas amostras não sobrepostas. Foram gerados seis grupos com, no máximo, 1.000 espécies, sem sobreposição, sendo que para cada grupo foi aplicada uma classificação empregando o modelo CNN AlexNet para fornecer saídas de classificação para cada grupo gerado. Finalmente, Zhang *et al*. (2020) combinaram todas as saídas das seis CNNs para obter a representação final das características e fornecer a previsão final para cada espécie de planta. O objetivo do mecanismo de atenção foi colocar um rótulo nas características extraídas de uma imagem de consulta chamada "Não está no grupo", quando um grupo não tem a espécie predita que está sendo avaliada. Os autores concluíram que o mecanismo de atenção usado no algoritmo AHMTL pode remover os componentes de características inúteis e aprender informações profundas mais discriminativas e específicas da planta. Além disso, a classificação hierárquica, guiada pela taxonomia da planta, substitui a classificação tradicional *softmax*, de modo que as relações interespécies entre plantas em larga escala sejam minimizadas.

Além de métodos hierárquicos voltados ao reconhecimento de plantas, modelos hierárquicos também são utilizados com sucesso em diversas tarefas de classificação de imagens em larga escala. Yan *et al*. (2015), por exemplo, propuseram uma Rede Neural Profunda Convolucional Hierárquica (HD-CNN) que incorpora um classificador de Rede Neural Convolucional na hierarquia de categorias com mais de 1.000 classes, o qual decompõe uma tarefa de classificação de imagem em duas etapas, conhecida como *Coarse-to-fine*. Primeiro, classes fáceis são separadas e chamadas "categorias grosseiras" (*Coarse*) e um classificador CNN baseado no modelo VGG-16 é aplicado; segundo, classes mais desafiadoras são conduzidas a outro classificador CNN, chamadas "categorias finas" (*Fine*). Um HD-CNN segue o paradigma de classificação de grosseiro para fino (*Coarse-to-fine*) e probabilisticamente integra previsões de classificadores de uma categoria grossa para fina. O desempenho da tarefa de classificação de imagens foi avaliado no conjunto de dados CIFAR

100 e ImageNet. Yan *et al*. (2015) mostraram que o HD-CNN pode obter um erro menor do que um modelo CNN não hierárquico.

Yan *et al.* (2015) mencionam que sua arquitetura é composta por uma hierarquia categórica em que as categorias finas são agrupadas em categorias grossas como apresentado na Figura 18a. Categorias grossas são consideradas classes que estão no topo da hierarquia, como, por exemplo, a categoria 1 “peixe”. Já as categorias finas são consideradas os nós subsequentes da hierarquia do nó pai, por exemplo, “tubarão branco”, “peixe-boi”, “arraia”, etc.

Figura 18. Arquitetura apresentada por Yan *et al.* (2015) para classificação hierárquica



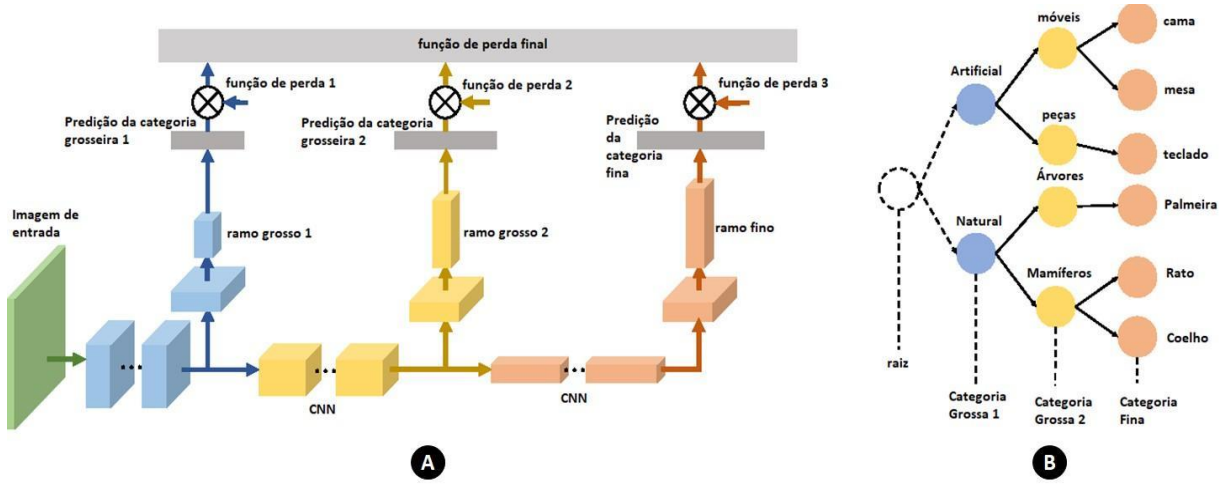
Fonte: Yan *et al*. (2015).

A arquitetura HD-CNN é ilustrada na Figura 18b, a qual é dividida em quatro partes: camadas compartilhadas, categorias grossas independentes B, múltiplas categorias finas e uma camada de probabilidade média. Nas camadas compartilhadas (bloco verde) são recebidas imagens como entrada e extraídas características de baixo nível. No bloco amarelo são representadas categorias grosseiras independentes, que produzem uma predição a ser utilizada no estágio de categorias finas. Yan *et al.* (2015) afirmam que a categoria grossa serve para duas propostas: os pesos são utilizados para combinar as predições feitas na camada fina, permitindo uma condicional execução da categoria fina, onde se verifica se as categorias grosseiras contêm subclasses das categorias finas. Os blocos alaranjados são conjuntos de camadas independentes finas, em que cada categoria gera suas predições. Por fim, no bloco azul, são combinadas as predições das categorias grossas e finas. Segundo os autores, a razão da combinação é que as camadas precedentes em redes profundas são responsáveis por características de baixos níveis, que podem ser cantos ou contornos, enquanto camadas finais extraem características mais

específicas da classe.

No estudo de Zhu e Bain (2017), os autores propuseram uma variante do tradicional modelo CNN VGG-16, chamado de *Branch Convolutional Neural Network* (B-CNN). Esse modelo faz a predição em múltiplas saídas ordenadas por categorias grossas e finas ao longo das camadas convolucionais concatenadas, correspondendo à estrutura hierárquica das classes-alvo, nas quais são conduzidas em forma de um conhecimento *a priori* de suas saídas. A performance do modelo foi avaliada nas bases de dados MNIST, CIFAR-10 e CIFAR-100. A arquitetura do modelo proposto pelos autores está ilustrada na Figura 19a e sua respectiva árvore de rótulos hierárquicos na Figura 19b.

Figura 19. Arquitetura B-CNN apresentada por Zhu e Bain (2017) para classificação hierárquica



Fonte: Zhu e Bain (2017).

A Rede Convolucional apresentada na Figura 19a possui múltiplas camadas, e no meio da estrutura existem os ramos de cada categoria grosseira, os quais produzem uma predição do nível da hierarquia correspondente. No topo da arquitetura de cada ramo, *softmax* são utilizados para produzir a saída final com as probabilidades de cada ramo da rede combinadas por alguma regra de fusão, sendo que cada ramo da arquitetura corresponde a uma Rede Convolucional com camadas completamente conectadas. Segundo Zhu e Bain (2017), na parte de classificação, o modelo B-CNN retorna como saída várias predições correspondendo a cada nível da árvore de rótulos (Figura 19b). Para fazer a avaliação nas bases de dados, os autores construíram manualmente a árvore de rótulos a fim de utilizar a arquitetura hierárquica e, com isso, eles obtiveram 99,40%, 88,22% e 64,42%, respectivamente, de acurácia para as bases de dados de MNIST, CIFAR-10 e CIFAR-100.

Pode-se verificar que os trabalhos de Yan *et al*. (2015) e de Zhu e Bain (2017) têm arquiteturas parecidas e seguem o mesmo princípio de que categorias grosseiras e finas são separadas por algum critério de classificação, gerando uma estrutura hierárquica de classes. Consequentemente, obtiveram resultados superiores se comparado com modelos tradicionais ou planos que classificam todas as classes de uma única vez. A forma como são conduzidos e tratados cada nível da estrutura hierárquica, entretanto, faz com que uma rede tenha influências diferentes na classificação final de imagens.

3.6 CONSIDERAÇÕES FINAIS

Em resumo, diferentes características distinguem a grande biodiversidade de espécies de plantas. Por exemplo, enquanto a forma da folha pode ser suficiente para distinguir espécies de uma base de dados, outras podem ter formas de folhas muito semelhantes entre si. Nenhum descritor de característica, único ou de determinado tipo de característica, pode ser suficiente para separar a diversidade das categorias existentes, tornando a seleção de características um problema desafiador em cenários de subcategorias.

Vários estudos primários mostraram os benefícios da fusão de características das folhas tais como: textura, bordas, contornos, veias, etc. possibilitam um aumento na performance de classificação (KEBAPCI; YANIKOGLU; UNAL, 2011; CHAKI; PAREKH; BHATTACHARYA, 2015; MOUINE; YAHIAOUI; VERROUST-BLONDET, 2013c; CAGLAYAN; GUCLU; CAN, 2013). A textura, muitas vezes, pode ser ofuscada pela forma como uma característica dominante atua na classificação de folhas e flores, no entanto, ela fornece e captura informações complementares sobre a folhagem, permitindo descrever finas nuances ou microtexturas da superfície da folha (YANIKOGLU; APTOULA; TIRKAZ, 2014).

Na análise foliar, não se espera que a cor seja tão discriminativa quanto a forma ou a textura, uma vez que a maioria das espécies de plantas contém algumas folhas coloridas, e um tom de verde que também varia muito sob diferentes formas de iluminação (YANIKOGLU; APTOULA; TIRKAZ, 2014). Apesar da baixa variabilidade interclasses em termos de cor, há alta oscilação intraclasse, ou seja, até mesmo as cores das folhas pertencentes à mesma espécie podem apresentar ampla gama de cores, dependendo da estação e do estado geral da planta (por exemplo, nutrientes e água). Vale destacar, também, que muitas folhas secas ficam com tonalidade marrom, o que comprova que a cor não costuma ser um recurso útil para a análise de folhas.

Independentemente das complicações anteriormente mencionadas, a cor ainda pode

contribuir para a identificação da planta, considerando folhas que exibem tonalidade singular (YANIKOGLU; APTOULA; TIRKAZ, 2014). Uma investigação mais aprofundada sobre a cor das folhas, no entanto, é necessária, já que para a análise de flores ela desempenha importante papel. Como característica, a cor também é conhecida por sua baixa dimensionalidade e complexidade computacional, sendo conveniente para aplicações em tempo real.

No estudo de Lee *et al*. (2017a) consta que combinar dados foliares tais como: planta inteira e estrutura das veias é uma alternativa para melhorar o desempenho da classificação, pois são as principais características aprendidas pelo classificador CNN em cada camada de convolução da rede. Os resultados e discussões do estudo apontam para uma questão que ficou aberta: quantas camadas convolucionais são necessárias na CNN para obter melhor capacidade de otimização na modelagem de dados da planta? Com base em inúmeras publicações sobre a classificação de objetos que utilizam CNN, os autores observaram um aumento dramático na profundidade das camadas dos modelos CNN para alcançar os resultados promissores.

Com o passar dos anos houve constante crescimento dos modelos convolucionais, ou seja: cinco camadas convolucionais em AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), 16 camadas em VGG16 (SIMONYAN; ZISSERMAN, 2014), 21 camadas em GoogleNet (SZEGEDY *et al*., 2014) e 164 camadas em ResNet (ZHANG *et al*., 2018). Esse crescimento foi exposto no Congresso PlantCLEF, e mostrou que durante os anos 2014 a 2019 diferentes modelos CNNs foram utilizados, iniciando por AlexNet, GoogleNet, VGGNet, ResNet, entre outros. Entende-se, portanto, que quanto maior a profundidade da rede, maior o tempo de treinamento do modelo de classificação. As redes profundas de CNN exigem grandes quantidades de dados de treinamento, limitando à base de dados com grandes quantidades de amostras para boa generalização.

O uso dos modelos CNNs em problemas de classificação, dependendo do tamanho da base de dados, com quantidades pequenas de amostras em classes, causa o problema de *over-fitting*, quando o treinamento fica viciado pelo fato de o modelo não conseguir aprender padrões para determinar uma classe. Para resolver o problema é necessário o uso de uma técnica de aumento de dados. Pensando nisso, Zhang *et al.* (2015) propôs uma CNN usando esta técnica, em que primeiramente foram implementadas transformações multiformas (por exemplo, rotação e translação, etc.) para ampliar o conjunto de dados sem alterar os seus rótulos. Essa técnica contribui enormemente para o desempenho do CNNs, pois é capaz de reduzir o grau de ajuste excessivo e melhorar a capacidade de generalização das CNNs. Em seguida, os autores treinaram uma CNN para classificar os dados de folhas aumentadas com três grupos de conjuntos de testes e, finalmente, concluíram que o método de aumento de dados é bastante

viável e eficaz. A exatidão alcançada por seu algoritmo supera outros métodos de aprendizado supervisionado no popular conjunto de dados Flávia. Como esse método gera mais dados para treinamento e testes, ele pode tornar o algoritmo CNNs mais robusto à variação das imagens, provando ser eficaz para reduzir o problema de *over-fitting*.

Embora o método de aumentar dados melhore a performance do classificador CNN, evidencia-se a dependência de grande quantidade de dados. Na identificação de plantas, novas espécies são encontradas e catalogadas diariamente por botânicos e, muitas vezes, não existe quantidade expressiva de amostras de imagens, deixando, assim, o processo de geração de modelo ineficiente para novos dados. Da mesma forma, a aplicação da técnica de aumento de dados para essas amostras não resolve completamente o problema, visto que o treinamento de um grande conjunto de amostras se torna computacionalmente caro. Finalmente, uma metodologia ideal deve ser escalar, isto é, independente de retreinar esses modelos onerosos cada vez que uma nova espécie desconhecida é inserida no modelo de reconhecimento.

Nos estudos apresentados anteriormente, a classificação hierárquica se mostra uma estratégia inteligente a ser seguida, uma vez que o problema a ser tratado tem uma estrutura hierárquica padronizada, e o fato de dividir um problema de classificação em subproblemas traz a possibilidade de discriminar categorias consideradas interclasse e intraclasse, com um nível de percepção mais acurado. Torna-se necessário, porém, encontrar alguma forma de extrair características robustas que sejam discriminantes entre objetos similares dentro de uma subcategoria. Classificadores monolíticos, *ensembles* e CNNs têm características individuais que podem fazer essa separabilidade, contudo, é necessário encontrar uma metodologia capaz de envolver essas características, sem a necessidade de um retreinamento do sistema, tampouco exigir grande quantidade de dados para o treinamento do classificador.

Partindo desse princípio, as abordagens de aprendizado métrico profundo demonstraram em seus estudos (ZHI-YONG *et al*., 2018; WANG; WANG, 2019; FIGUEROA-MATA; MATA-MONTERO, 2020) a possibilidade de criar um sistema de reconhecimento a partir da utilização de poucas amostras de treinamento. Outra questão destacada pelas redes métricas profundas é a escalabilidade dos sistemas a ser avaliada na implementação de uma Rede Neural Convolucional Siamesa (SNN), a qual identifica espécies de plantas por meio de uma métrica de similaridade entre amostras de espécies e que pode ser independente de um modelo robusto de classificação.

Para finalizar este capítulo, nas Tabelas 2, 3 e 4 é apresentada uma visão geral dos estudos encontrados na literatura. Na Tabela 2 sintetizam-se alguns estudos que empregaram metodologias tradicionais, tais como classificadores monolíticos ou *ensembles*. Já na Tabela 3

constam alguns estudos que empregaram modelos profundos (CNNs) e métricos (SNN).

Para finalizar, na Tabela 4 sintetizaram-se metodologias hierárquicas para resolver o problema de reconhecimento de plantas. É possível verificar que todos os estudos que utilizaram modelos profundos optaram por utilizar a técnica de aumento de dados para melhorar a acurácia de seus classificadores.

Alguns estudos, todavia, empregaram metodologia em que a classificação de uma amostra de teste é dada como certa quando o modelo de classificação gerado contém grande quantidade de imagens de treinamento, ou seja, classes que têm poucas amostras de imagens ou são consideradas desbalanceadas são confundidas com outras e classificadas erroneamente pelo classificador. Este é outro aspecto que poucas metodologias apresentadas conseguiram tratar.

Percebeu-se, também, que são poucos os estudos que utilizam a estrutura hierárquica para resolver o problema de classificação de plantas, uma vez que, *a priori*, é mais natural descobrir a família ou o gênero de determinada planta pelos seus fenômenos taxonômicos do que diretamente a sua espécie.

Finalmente, até onde se sabe, apenas três estudos (ZHI-YONG *et al*., 2018; WANG; WANG, 2019; FIGUEROA-MATA; MATA-MONTERO, 2020) exploraram aprendizado métrico profundo na tarefa de reconhecimento de plantas.

Tabela 2. Trabalhos relacionados: abordagens tradicionais (*handcrafted*)

Abordagem	Pré-processamento	Características	Metodologia	Base de Dados	Estudos
Tradicional	Otsu segmentação	Forma	Detecção de Harris	ImageCLEF 2011 e 2012	MOUINE; YAHIAOUI; VERROUST-BLONDET (2013b)
	N/A	Forma	I-IDSC	Swedish, ICL, Smithsonian, Plumbers Island e base de dados criada	ZHAO *et al*., 2015
	N/A	Forma	SURF + Bag of Visual Words	Swedish	CHARTERS *et al*., 2014
	Segmentação das veias (UHMT)	Veias	SVM (linear e Gaussian kernels), PDA e RF	INTA	LARESE *et al*., 2014
	N/A	Forma	Hough, Fourier e EoH	ImageCLEF 2011	MOUINE; YAHIAOUI; VERROUST-BLONDET (2012)
	N/A	Forma	Múltiplas representações triangulares	Swedish, Flavia, ImageCLEF 2011 e 2012	MOUINE; YAHIAOUI; VERROUST-BLONDET (2013a)
	N/A	Forma	Distância de centróide triangular multiescala	ImageCLEF 2012, Swedish, Smithsonian, Flavia	YANG; WEI; YU (2016)
	N/A	Contorno	Momentos de HU e histograma de curvatura	50 espécies europeias	CERUTTI *et al*. (2013)
	Binarização da imagem	Contorno	Altura do arco multiescala	Swedish, Flavia, ICL e categoria" scan" ImageCLEF 2012	WANG *et al*. (2015)
	Binarização da imagem	Contorno	Ângulo-R	Flavia, ImageCLEF 2012	CAO; WANG; BROWN (2016)
	N/A	Forma; textura	SURF, EoH, Fourrier, Hough	Swedish, ImageCLEF 2011 e 2012	JOLY *et al*. (2014)
	Otsu segmentação	Forma; textura e cor	Fourier, Gabor, RGB e LSH histogram	ImageCLEF 2012	YANIKOGLU; APTOULA; TIRKAZ (2014)
	Otsu segmentação e remoção do pecíolo (caule)	Forma; veias	Fourier, EoH, Hough, DFH com KNN	ImageCLEF 2011	MZOUGHI *et al.* (2016)
	Otsu segmentação	Margem e forma	Correlação espacial e detecção de Harris	Flavia, ImageCLEF 2011 e 2012	MOUINE; YAHIAOUI; VERROUST-BLONDET (2013c)
	Erosão morfológica e binarização da imagem	Forma, textura, cor e veias	LDA, RF, GLCM e HSV	UCI	ELHARIRI; EL-BENDARY; HASSANIEN (2014)
	Binarização da imagem	Forma, morfologia, textura e cor	ACO, SVM	FCA e Flavia	GHASAB *et al*. (2015)
	Correção de translação, rotação e escala	Textura, contorno e momentos invariantes	*Neuro-fuzzy controller* (NFC) e MLP	Flavia	CHAKI; PAREKH; BHATTACHARYA (2015)

Fonte: elaboração própria (2021).

Tabela 3. Trabalhos relacionados: abordagens profundas e métricas

Abordagem	Pré-processamento	Características	Metodologia	Base de Dados	Estudos
Profunda	Segmentação por região de interesse (ROI), *data augmentation*	CNN pré-treinada	AlexNet	PlantCLEF 2014	CHEN *et al.* (2014)
	Data augmentation	Combinação Bordafuse das CNNs geradas	*Fine-tuning* com GoogLeNet	PlantCLEF 2015	SUNGBIN (2015)
	Data augmentation	Alterou a última camada de *polling* por *Spatial Pyramid Pooling layer*	VGG-16	PlantCLEF 2016	HANG; TATSUMA; MASAKI (2016)
	Data augmentation	Combinou os modelos CNNs	GoogLeNet, ResNet e ResNeXT	PlantCLEF 2017	LASSECK (2017)
	Data augmentation	CNN pré-treinada	3 camadas CNN (2 conv; 1 *softmax*)	Flavia	ZHANG *et al*. (2015)
	Segmentação das veias (UHMT) e *data augmentation*	Padrões das veias combinadas com CNN	6 camadas CNN (5 conv; 1 *softmax*)	INTA	GRINBLAT *et al.* (2016)
	Transfer learning e *data augmentation*	Combinação de imagens patches e inteiras	AlexNet	PlantCLEF 2015	LEE *et al*. (2017a)
	Transfer learning e *data augmentation*	Combinação CNNs	GoogLeNet, AlexNet, e VGGNet	PlantCLEF 2015	GHAZI; YANIKOGLU; APTOULA, (2017)
	Transfer learning e *data augmentation*	CNN com classificação SVM	VGG-19 20 camadas (16 conv; 3 FC; 1 *softmax*)	PlantCLEF 2015	ZHU *et al.* (2018)
	Transfer learning e algoritmo de Sobel	Combinação de classificadores	AlexNet pré-treinada, AlexNet com *fine-tunning* e *D-Leaf*	PlantCLEF 2015	TAN *et al*. (2020)
Métrica	N/A	Redes Neurais Recorrentes na arquitetura SNN	Rede Neural Siamesa (SNN)	PlantCLEF 2015	ZHI-YONG *et al*. (2018)
	Few-shot-learning (5 até 20 amostras por classe)	GoogleNet na arquitetura SNN	Rede Neural Siamesa (SNN)	LeafSnap, Swedish e Flavia	WANG; WANG (2019)
	Few-shot-learning (5 até 30 amostras por classe)	Arquitetura CNN criada por eles (similar ao utilizado no trabalho de KOCH; ZEMEL; SALAKHUTDINOV, (2015)	Rede Neural Siamesa (SNN)	Flavia	FIGUEROA-MATA; MATA-MONTERO (2020)

Fonte: elaboração própria (2021).

Tabela 4. Trabalhos relacionados: abordagens hierárquicas

Abordagem	Pré-processamento	Características	Metodologia	Base de Dados	Estudos
Hierárquica	N/A	Combinação de CNNs na hierarquia	VGG-16:2 CNNs, 1 para gênero e outra para espécie	PlantCLEF 2015	LEE *et al.* (2017)
	Segmentação da Imagem	Hierarquia composta pela estrutura morfológica das plantas	Seleção de forma em diferentes partes da folha	Flavia	CHAKI; PAREKH; BHATTACHARYA (2018)
	Segmentação	Hierarquia por similaridade	*KD-trees*	Base de dados própria	WU *et al.* (2015)
	Segmentação	Hierarquia por similaridade	Conjunto confidente (CS)	Swedish, Flavia, Smithsonian: e Image-CLEF 2011	SFAR; BOUJEMAA; GEMAN (2015)
	N/A	Combinação de CNNs na hierarquia	Redes neurais recorrentes	PlantCLEF 2015	LEE; CHAN; REMAGNINO (2018)
	Data augmentation e recorte na região central da imagem	Classificação de dois níveis (Famílias e espécies)	Mapa de calor, utilizando CNN Xception	Malayakew e ICL	ZHU *et al*. (2019)
	N/A	Aprendizagem múltipla hierárquica profunda (AHMTL)	AlexNet CNN	Orchid	ZHANG *et al.* (2020)

Fonte: elaboração própria (2021).

4 MÉTODO PROPOSTO

Este capítulo apresenta duas abordagens para facilitar o reconhecimento de plantas por meio de imagens do componente folha. Primeiramente, passa-se a explorar uma classificação hierárquica mediante a utilização da taxonomia presente nas plantas (Seção 4.1), a utilização de diferentes pontos de vista da folha em cada estágio da hierarquia (Seção 4.2) e o pré-processamento da imagem utilizada em cada ponto de vista (Seção 4.3). Posteriormente, na Seção 4.4, apresenta-se o primeiro método proposto, utilizando as Redes Neurais Convolucionais (CNN). Na Seção 4.5, descreve-se o segundo método, empregando as Redes Neurais Convolucionais Siamesas (SNN). Finalmente, na Seção 4.6 apresentam-se as métricas de avaliação e as ferramentas utilizadas na implementação dos métodos (Seção 4.7).

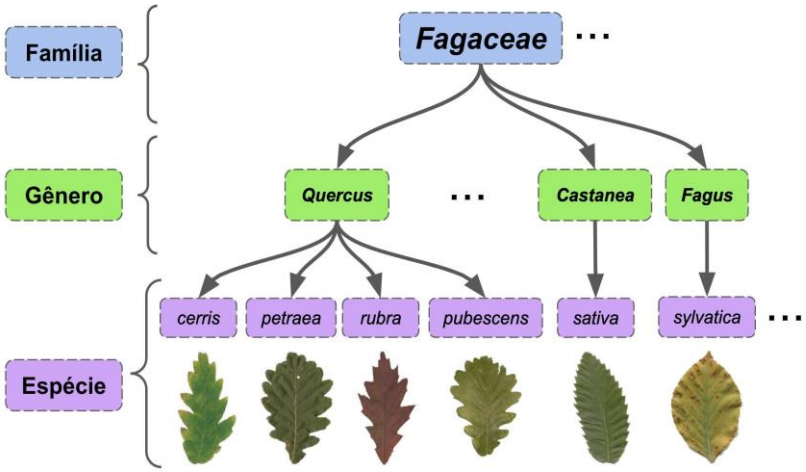
4.1 DEFINIÇÃO DA HIERARQUIA TAXONÔMICA DAS PLANTAS

A taxonomia das plantas está dividida biologicamente numa hierarquia ordenada por Reino, Filo, Classe, Ordem, Família, Gênero e Espécie. A Figura 20 apresenta a estrutura hierárquica taxonômica disponível nos conjuntos de dados das plantas utilizadas neste estudo. Observa-se a existência de Famílias, Gêneros e Espécies organizados de acordo com as suas semelhanças morfológicas.

Na visão dos botânicos, níveis mais altos da hierarquia contêm características mais discriminativas (SCHUH; BROWER, 2009). Por exemplo, na identificação de uma planta desconhecida, inicialmente é identificada a família ou o gênero à qual a planta pertence para, posteriormente, reconhecer a sua espécie. Essa estratégia de dividir e conquistar diminui a complexidade de reconhecer as subcategorias. Além disso, as espécies de plantas podem ser classificadas hierarquicamente por meio das ramificações da árvore, de uma maneira conhecida pela comunidade científica como Grossa à Fina ou *Coarse-to-fine*. Essa abordagem (*Coarse-to-fine*) reconhece por etapas, primeiramente, as classes de níveis mais altos da árvore (classes grosseiras) para, posteriormente, discriminar as classes subordinadas (classes finas).

Segundo He *et al*. (2020), *Coarse-to-fine* reduz a complexidade interespécies, pois por ser um problema de reconhecimento de subcategorias, algumas espécies de plantas podem ter fortes semelhanças visuais interespécies, o que leva a aprender classificadores interrelacionados de maneira independente. Há, também, menor complexidade discriminativa na classificação hierárquica, o que reduz o número de categorias passadas aos demais níveis hierárquicos, fazendo com que a classificação *Coarse-to-fine* seja implementada nos métodos propostos.

Figura 20. Divisão hierárquica da taxonomia das plantas empregada neste estudo (Famílias, Gêneros e Espécies)



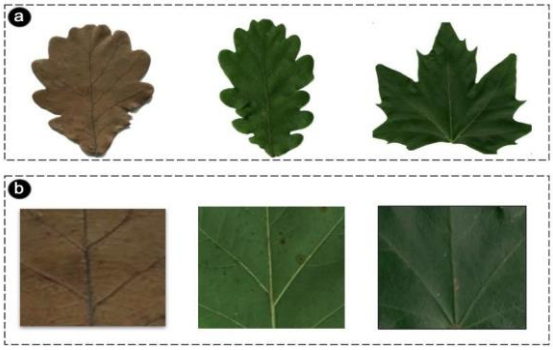
Fonte: elaboração própria (2021).

4.2 REPRESENTAÇÃO DA PLANTA SOB DOIS PONTOS DE VISTA

Diversas abordagens utilizam a imagem da folha como objeto principal na extração de características para o reconhecimento de plantas, embora haja dificuldade em discriminar espécies com diferenças sutis. Por esse motivo, a imagem da planta é representada a partir de dois pontos de vista da folha (geral e local). A perspectiva inicial extrai características gerais (cores, formas, bordas e contornos), enquanto a segunda foca na extração de características mais detalhadas da folha da planta (texturas e padrões de veias).

Extraíram-se, assim, características gerais utilizando a imagem da folha inteira e, de outro modo, as características locais foram extraídas a partir de um recorte no centro das imagens das folhas. Nas Figuras 21a e 21b são apresentados exemplos das perspectivas gerais e locais para três amostras de folha.

Figura 21. Exemplos dos dois pontos de vista da folha

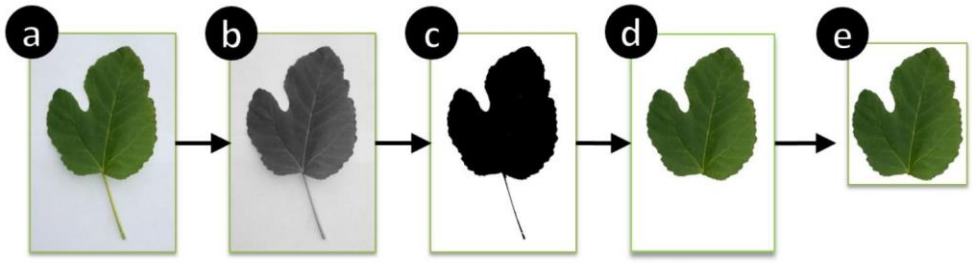


Fonte: elaboração própria (2021).
Legenda: (a) ponto de vista geral; (b) ponto de vista local

4.3 PRÉ-PROCESSAMENTO

O pré-processamento das imagens das folhas ocorreu em duas etapas: primeiramente removeram-se as estruturas indesejáveis da folha (Figura 22) e, posteriormente, realizou-se o corte central na imagem filtrada (Figura 21b).

Figura 22. Pré-processamento da imagem da folha



Fonte: elaboração própria (2021).
Legenda:
a) imagem original
b) imagem em escala de cinza
c) operação de *threshold*
d) operação de *top-hat*
e) imagem final após aplicação de *bounding box*.

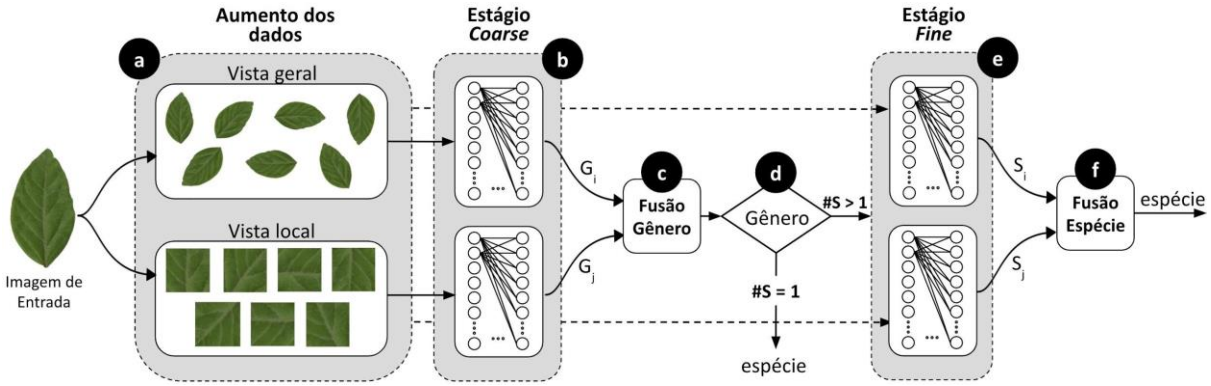
Na Figura 22 observa-se que a primeira etapa de pré-processamento converte a imagem colorida em níveis de cinza (Figura 22b), removendo o fundo da imagem por meio do método *threshold* de Otsu (Figura 22c). Assim, todos os *pixels* com valor abaixo do limiar estimado pelo método de Otsu são classificados como primeiro plano, e os restantes como plano de fundo. Com a folha segmentada de seu fundo, a próxima etapa é remover o caule da folha usando o algoritmo de *top-hat*. A transformação de *top-hat* determina um conjunto de estruturas finas que se destacam na imagem. Esse método funciona aplicando à imagem uma estrutura circular definida, denominada "elemento estrutural", com um diâmetro superior à largura dos elementos a serem excluídos (o caule da folha). Após a retirada do caule pelo método de *top-hat* (Figura 22d), a imagem resultante é submetida à etapa de definição do *bounding box,* método utilizado para focar em informações relevantes da folha. A imagem final do pré-processamento (Figura 22e) é utilizada como ponto de vista geral. A resolução final da imagem é de 224×224 *pixels*.

Na segunda etapa de pré-processamento, a imagem filtrada, contendo apenas a folha, foi recortada em seu centro por meio do algoritmo proposto por Bloice, Stocker e Holzinger (2017) e redimensionou-se a resolução da imagem cortada para 32×32, 64×64 e 128×128. A ideia foi avaliar o desempenho do método proposto considerando imagens com diferentes resoluções de entrada.

4.4 PRIMEIRA ABORDAGEM: CNN

A visão geral do método proposto com a CNN é apresentada na Figura 23. O método utiliza a estratégia de classificação hierárquica *Coarse-to-fine*, avaliando categorias de gêneros no primeiro estágio da hierarquia (*Coarse* – Figura 23b) e, posteriormente, categorias de espécies das plantas no segundo estágio (*Fine* – Figura 23e).

Figura 23. Visão geral do método proposto com a CNN



Fonte: elaboração própria (2021).
Legenda:
a) Etapa de aumento de dados
b) Duas CNNs pré-treinadas, uma para cada ponto de vista (geral e local) sob categorias de Gêneros
c) Fusão das probabilidades de cada vista de Gênero
d) Verificação do número de espécies encontradas no gênero predito
e) Duas CNNs pré-treinadas, uma para cada vista sob categorias de Espécies
f) Predição da espécie da planta utilizando a fusão das probabilidades de cada ponto de vista.

Inicialmente, as imagens das folhas das plantas são submetidas à etapa de aumento de dados (Figura 23a) conforme a Seção 4.4.1. Em seguida, no estágio *Coarse* empregou-se duas CNNs pré-treinadas utilizando o conjunto de dados ImageNet (DENG *et al*., 2009) (Figura 23b), uma CNN para cada ponto de vista (geral e local) utilizando categorias de gêneros.

Na próxima etapa (Figura 23c), após o treinamento dos modelos, realizou-se a combinação dos resultados de classificação, quando recebida uma imagem de entrada (teste). A combinação foi realizada por meio de um esquema de fusão apresentado na Seção 4.4.3. Na etapa seguinte (Figura 23d) tem-se como saída da classificação *Coarse* o primeiro gênero predito, que é utilizado como entrada do próximo estágio hierárquico (*Fine*). Para esse fim, apenas as categorias de espécies do único gênero predito são consideradas. Similarmente, na classificação *Fine* (Figura 23e), as características das espécies são extraídas de cada ponto de vista, sendo o treinamento realizado sob subconjuntos com categorias de espécies.

Finalmente, a fusão dos resultados ocorre apenas entre as espécies contidas no gênero

predito (Figura 23f) para a mesma imagem de entrada. É importante notar que o segundo estágio da hierarquia é acionado apenas quando o gênero predito no primeiro estágio é composto por mais de uma espécie (S > 1). Caso contrário, é possível identificar a espécie da planta diretamente (S=1) (Figura 23d). Uma descrição detalhada de cada etapa deste método é apresentada nas próximas seções.

4.4.1 Aumento dos dados

O principal objetivo de se aumentar o conjunto de dados está relacionado ao desbalanceamento de algumas espécies, sendo criadas novas imagens com variações de rotações e ângulos para categorias mal representadas. Dada uma imagem inteira (ponto de vista geral), seis novas imagens são obtidas girando a imagem original em 45º, 90º, 180º, 225º, 270º e 330º no sentido horário. Da mesma forma, foram realizadas as mesmas rotações na imagem recortada (ponto de vista local). Finalmente, com o aumento dos dados, obteve-se um total de sete imagens inteiras e sete imagens recortadas com diferentes rotações para cada amostra de entrada. Como ilustrado na Figura 23a, considerando a imagem de entrada original, foram geradas 14 novas imagens.

4.4.2 Redes Neurais Convolucionais (CNN)

Os modelos de reconhecimento são gerados com o treinamento em diferentes subconjuntos, conforme a Tabela 5. O subconjunto “A” utiliza no treinamento todas as categorias de gênero da base de dados sob o ponto de vista geral das folhas (imagens inteiras). O subconjunto “B” é treinado com gêneros sob o ponto de vista local (imagens recortadas da folha). Subconjuntos “A” e “B” são responsáveis por gerar modelos CNNs do primeiro estágio da hierarquia (Figura 23b) e, posteriormente, combinar a predição dos resultados de uma imagem de teste (Figura 23c). Os subconjuntos “C” e “D” também são utilizados para obter diferentes pontos de vista (geral e local), mas treinados sob categorias de espécies, enquanto os modelos gerados são utilizados no segundo estágio da classificação hierárquica (Figura 23e). A técnica de aumento de dados foi aplicada em todos os subconjuntos da Tabela 5 utilizando a base de dados PlantCLEF 2015, bem como dividiu-se a base de dados em 70% dos dados para treinamento e 30% para validação. Portanto, em cada subconjunto (A, B, C ou D) temos 123.578 imagens de treinamento e 52.962 imagens de validação divididas em suas respectivas classes.

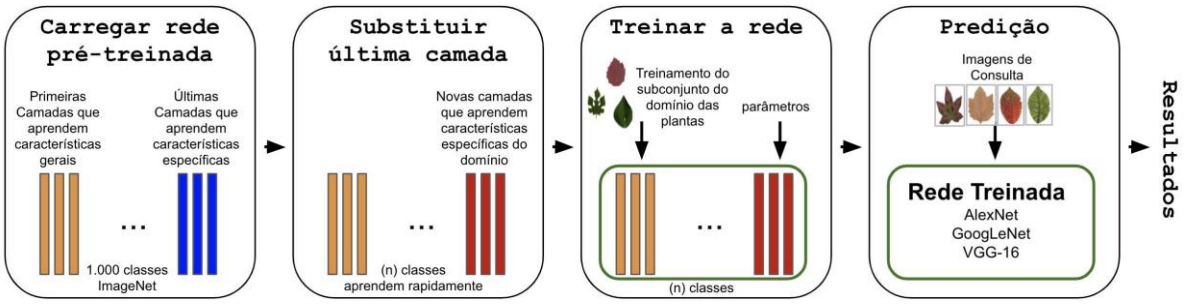
Tabela 5. Treinamento dos modelos de reconhecimento com diferentes subconjuntos

Classes	Treinamento	Estágio Hierarquia	Categoria	Ponto de vista
260	Subconjunto A	*Coarse*	Gênero	Geral
260	Subconjunto B	*Coarse*	Gênero	Local
351	Subconjunto C	*Fine*	Espécie	Geral
351	Subconjunto D	*Fine*	Espécie	Local

Fonte: dados da pesquisa (2021).

O processo de treinamento de cada modelo usou a técnica de transferência de aprendizado a partir de uma rede pré-treinada sobre o conjunto de dados ImageNet (DENG *et al*., 2009). Além disso, foi realizado o ajuste fino dos dados, isto é, as primeiras camadas da rede pré-treinada são congeladas e apenas as camadas totalmente conectadas são responsáveis pela extração de característica mais específicas dos dados de entrada durante o treinamento (a rede funciona como um extrator de característica de imagem). Cada arquitetura de rede utilizada neste estudo (AlexNet, GoogLeNet e VGG-16) passa a ser detalhada, seguindo as etapas de transferência de aprendizado e ajuste fino, conforme apresentado na Figura 24.

Figura 24. Transferência de aprendizado e ajuste fino aplicadas às redes AlexNet, GoogLeNet e VGG-16 com base no método proposto



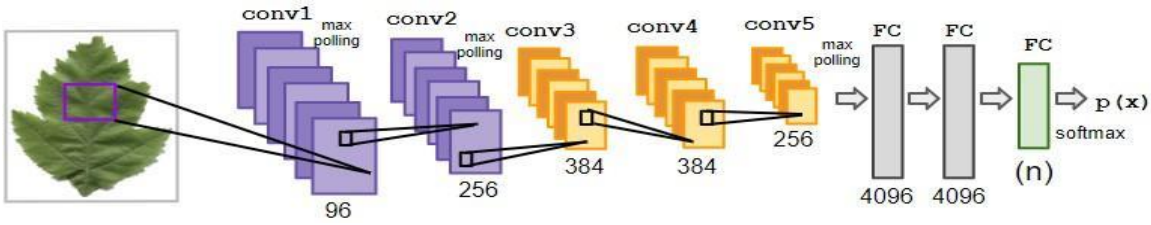
Fonte: elaboração própria (2021).
Obs.: (n) refere-se à quantidade de categorias encontradas na base de dados de plantas.

A arquitetura AlexNet foi desenvolvida por Krizhevsky, Sutskevsky e Hinton (2012). A rede é composta por uma arquitetura de três camadas de convolução e duas camadas totalmente conectadas (Figura 25). A primeira camada convolucional filtra uma imagem de entrada de tamanho 224 × 224 × 3, com 96 *kernels* de tamanho 11 × 11 × 3, e uma distância (*stride*) de quatro *pixels* (distância do deslizamento do campo receptivo na imagem). A segunda camada convolucional toma como entrada a saída da primeira camada convolucional e a filtra com 256 *kernels* de tamanho 5 × 5 × 48. Nas duas primeiras camadas são utilizadas as funções de *max-pooling* e LRN (normalização dos valores dos neurônios), o que provoca a redução do número de parâmetros pela ação do *pooling* e a normalização dos valores. A terceira camada

convolucional tem 384 *kernels* de tamanho 3×3×256 conectados às saídas (normalizada e agrupada) da segunda camada convolucional. A quarta camada convolucional tem 384 *kernels* de tamanho 3 × 3 × 192, e a quinta camada convolucional tem 256 *kernels* de tamanho 3 × 3 × 192.

Após a passagem por essas cinco camadas de extração de características, os dados são classificados em duas camadas totalmente conectadas. *softmax* é utilizada para calcular as probabilidades da saída correta dada uma imagem de entrada (ALOM *et al*., 2018). As duas camadas totalmente conectadas têm 4.096 neurônios, sendo que a saída desta última camada é fornecida a um *softmax* de 1.000 saídas que produz uma distribuição sobre 1.000 rótulos de classe. Dessa forma, nesta arquitetura aplicou-se o ajuste fino, removendo a última camada da arquitetura original de AlexNet, a qual foi substituída por uma camada totalmente conectada *softmax* p(x) com (n) saídas. O tamanho de (n) é definido igual à contagem de categorias dos conjuntos de dados utilizados neste estudo.

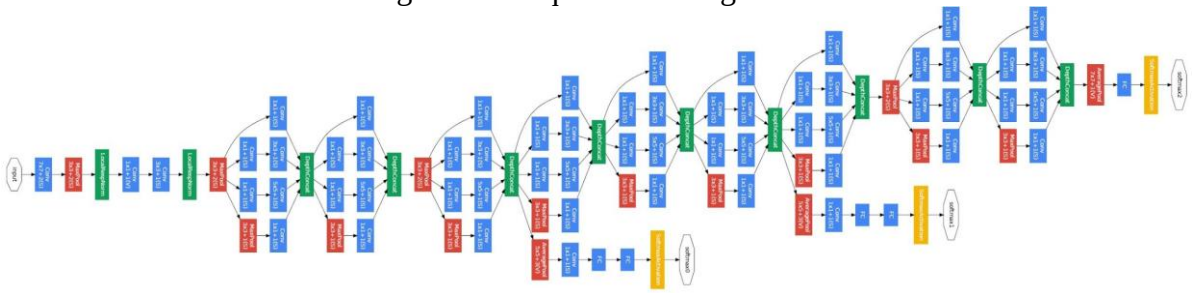
Figura 25. Arquitetura AlexNet



Fonte: elaboração própria com base em Krizhevsky, Sutskever e Hinton (2012).

Proposta por Szegedy *et al*. (2014), a arquitetura GoogLeNET foi desenvolvida com o objetivo de reduzir a complexidade computacional em relação às demais arquiteturas. O diferencial dessa arquitetura é a inclusão da camada de *Inception*, representada por camadas compostas de filtros de *kernels* com diferentes tamanhos (ALOM *et al*., 2018). A Figura 26 ilustra a arquitetura da rede GoogLeNET.

Figura 26. Arquitetura GoogLeNet



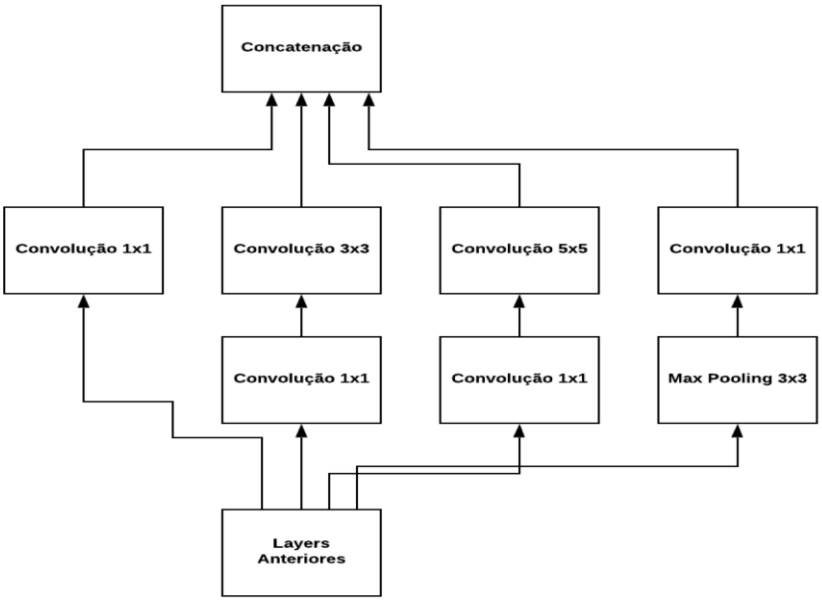
Fonte: Szegedy *et al*. (2014).

As camadas em azul são associadas à operação de convolução; as vermelhas ao *pooling*;

amarelo ao *softmax* e, finalmente, em verde as camadas referentes à normalização e concatenação dos valores. O comportamento conjunto das camadas de convolução e *pooling* permite tratar cada camada como um módulo *Inception* na rede. No total, GoogLeNet contém 22 camadas, os filtros da rede são de tamanho 1 × 1, 3 × 3 e 5 × 5, utilizando *max-pooling* e tamanho do *stride* = 2 (SZEGEDY *et al*., 2014).

Na passagem de uma camada a outra, um procedimento interno efetua quatro convoluções (uma 1 × 1, uma 3 × 3, uma 5 × 5 e um *max-pooling*). Ao final de um *Inception* é aplicado um filtro de 1 × 1, e a normalização dos valores (ALOM *et al.*, 2018). Um exemplo da camada *Inception* é apresentado na Figura 27.

Figura 27. Arquitetura de uma camada *Inception*

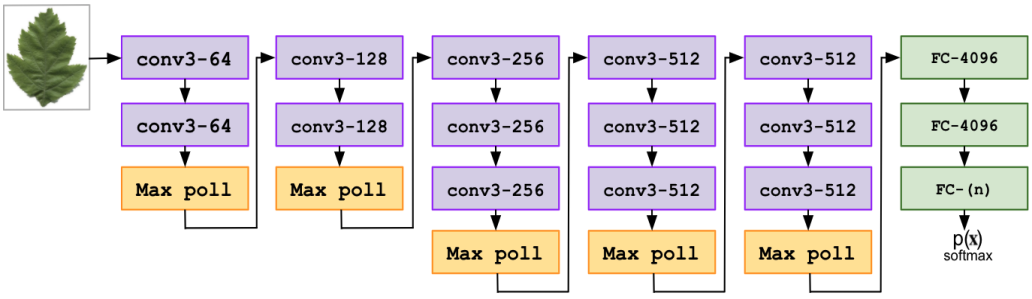


Fonte: Szegedy *et al*. (2014).

A arquitetura VGG-16 (SIMONYAN; ZISSERMAN, 2014a) é apresentada na Figura 28. Diferente das outras arquiteturas, esta aumenta a profundidade da rede, inserindo mais camadas convolucionais e diminuindo o tamanho dos *kernels*, fixando em 3 × 3 cada camada. A rede VGG-16 completa é composta por cinco blocos com camadas de convolução seguidas de *max-pooling* e três camadas totalmente conectadas. O primeiro bloco da rede tem duas camadas convolucionais de 64 filtros; o segundo bloco possui duas camadas de 128 filtros cada um; o terceiro bloco tem três camadas de 256 filtros; e os blocos 4 e 5 têm três camadas de 512 filtros cada um. A função de ativação ReLU é utilizada em cada camada da rede e *max-pooling* de tamanho 2 × 2 com *stride* = 2 são utilizados após cada camada de convolução a fim de reduzir a dimensionalidade. As últimas três camadas são totalmente conectadas, sendo as duas primeiras de tamanho 4.096, e a última com o tamanho de (n) igual à quantidade de categorias

da base de dados.

Figura 28. Arquitetura VGG-16



Fonte: elaboração própria com base em Simonyan e Zisserman (2014).

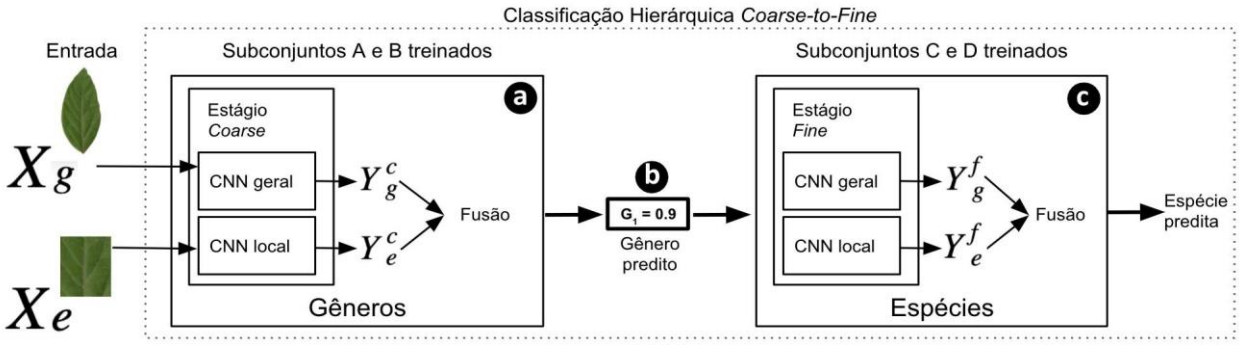
4.4.3 Reconhecimento e esquema de fusão

Utilizando os modelos treinados na Seção 4.4.2 foi possível realizar o reconhecimento de uma imagem de entrada com a fusão das saídas preditas em cada estágio da classificação hierárquica *Coarse-to-fine*. Foram realizadas duas etapas de fusão utilizando a regra de soma. Assumiu-se um conjunto de *Y* classificadores, sendo Y_g^c, Y_e^c (treinados sobre subconjuntos A e B) e Y_g^f, Y_e^f (treinados sobre subconjuntos C e D), onde cada classificador *Y* produz na saída $[P_{Yi}(W_j|X)]$ que representa o suporte para a hipótese de que o vetor *X* seja da classe W_j. O rótulo predito de cada etapa de fusão é encontrado pela regra de soma definida na equação 4.1. *k* denota os dois termos da fusão (*g*, *e*).

$$U(X) = argmax \sum_{i=1}^{k=2} P_{Yi}(W_j|X) \tag{4.1}$$

Dada uma imagem de entrada *X*, o índex máximo é encontrado pela fusão das probabilidades de Y_i em *U*(*X*), o que significa que a primeira fusão é gerada pelas probabilidades dos classificadores Y_g^c, Y_e^c (Figura 29a), onde *c* representa predição para o estágio *Coarse* utilizando o ponto de vista geral *g* ou local *e*. Após a combinação da primeira etapa da classificação hierárquica, considerou-se apenas o primeiro gênero predito (Figura 29b) para ser levada à segunda etapa *Fine*. Na segunda etapa a mesma regra de fusão é implementada (Figura 29c), porém a fusão das probabilidades dos classificadores Y_g^f, Y_e^f são apenas entre as espécies encontradas no gênero predito, em que *f* representa predições do estágio *Fine*, utilizando o ponto de vista geral *g* ou local *e*. Finalmente, após a fusão das etapas, retornou-se à espécie predita final.

Figura 29. Reconhecimento e esquema de fusão do primeiro método proposto: CNN



Fonte: elaboração própria (2021).

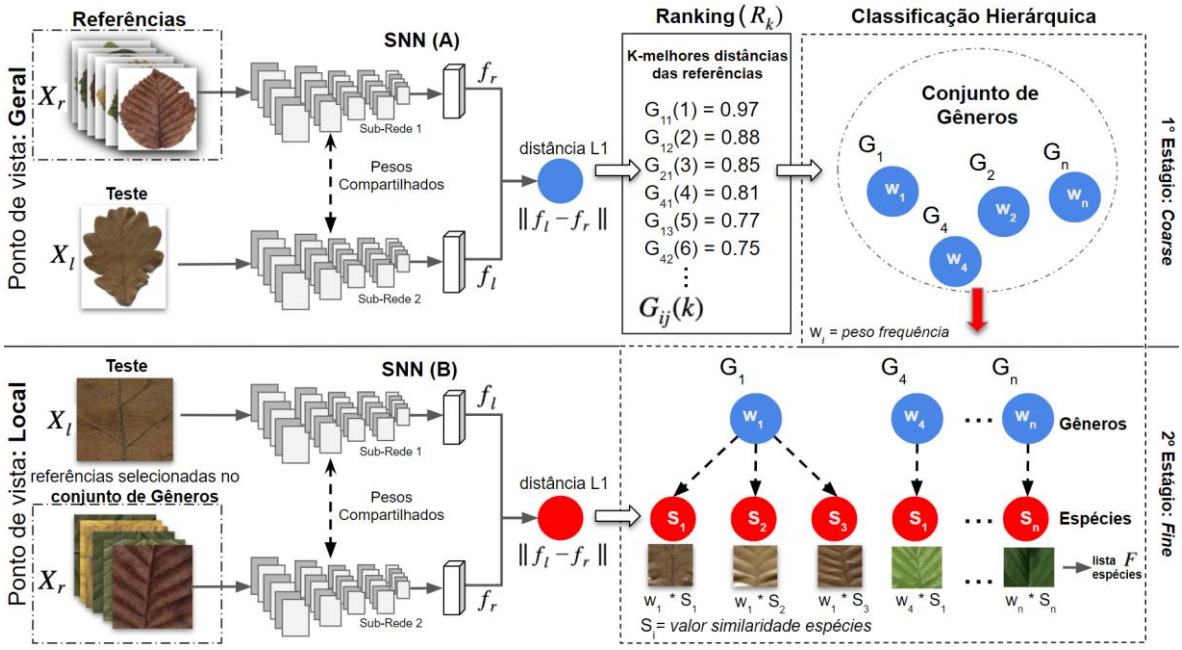
4.5 SEGUNDA ABORDAGEM: SNN

Neste método, o modelo estima a semelhança entre duas imagens. Para este propósito, uma Rede Neural Convolucional é usada para extrair características de duas imagens e mensurar a similaridade, calculando a distância entre as representações obtidas. Na Seção 4.5.1 apresenta-se uma visão geral do método proposto com a utilização do SNN, detalhando-se os principais componentes do aprendizado métrico, tais como: codificador CNN para representar imagens, função de perda e estratégia de treinamento. Na Seção 4.5.2 explana-se a estratégia de classificação hierárquica *Coarse-to-fine*. Finalmente, na Seção 4.5.3 apresenta-se o esquema de fusão utilizado no reconhecimento de plantas.

4.5.1 Redes Neurais Siamesas (SNN)

Apresenta-se, a seguir, uma visão geral do método proposto, utilizando SNN juntamente com a implementação da classificação hierárquica *Coarse-to-fine* e com diferentes pontos de vista da folha. A estratégia de classificação *Coarse-to-fine* reconhece o gênero da planta no primeiro estágio utilizando SNN (A). Posteriormente, no segundo estágio as espécies são reconhecidas com SNN (B), como ilustrado na Figura 30. SNN (A) mensura a similaridade entre uma imagem de teste X_l e imagens de referências X_r previamente escolhidas para representar cada gênero.

Figura 30. Visão geral do método proposto utilizando SNN



Fonte: elaboração própria (2021).

A saída do primeiro estágio é um conjunto de gêneros preditos, ordenado pelas K-melhores distâncias (maior similaridade) encontradas na lista (R_k). Determinou-se o conjunto de gêneros no primeiro estágio, mensurando a distância entre a imagem de teste e cada referência de gênero. Em seguida, calculou-se w_i como um peso de frequência para cada gênero do conjunto predito. O cálculo do peso w para cada gênero é dado pela quantidade de vezes em que referências *j* de um mesmo gênero *i* aparece na lista R_k. Por exemplo, na Figura 30, se considerarmos apenas $R_k = 6$, o peso de frequência w do gênero G1 será = 3, G4 = 2 e G2 = 1. Utilizou-se a informação de frequência de gênero para ponderar o processo de fusão no segundo estágio.

No segundo estágio, as mesmas imagens de testes e referências são inseridas, mas são alteradas para o ponto de vista local (imagens cortadas) a fim de que a similaridade entre pares de imagens seja mensurada por SNN (B), considerando apenas as espécies que pertencem ao conjunto de gêneros encontrados no primeiro estágio. Finalmente, a saída do segundo estágio é uma lista *F* das espécies de plantas ponderadas por um esquema de fusão entre as saídas do primeiro e segundo estágio.

Nas próximas seções apresenta-se o método SNN proposto em detalhes.

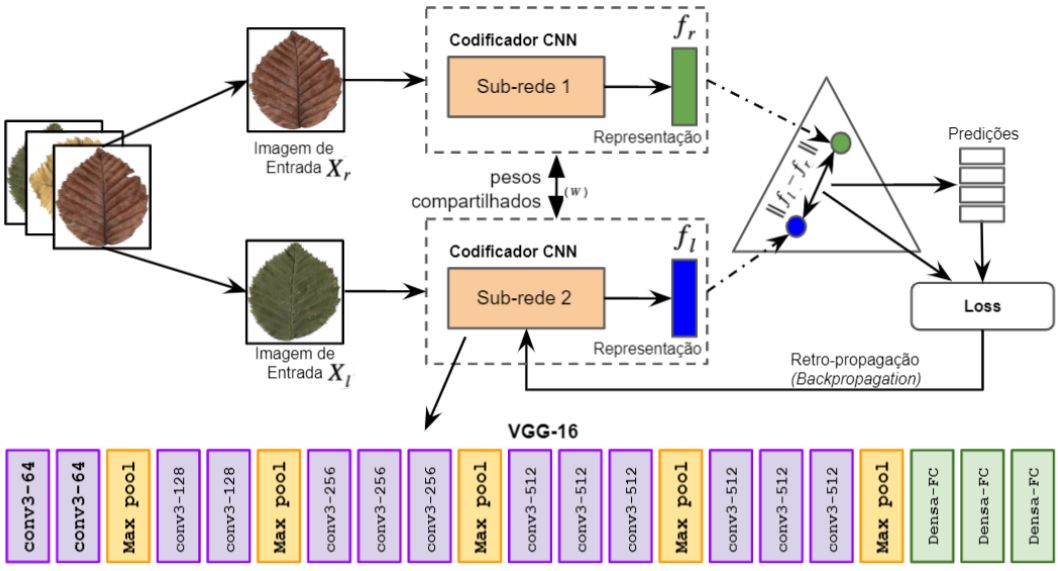
4.5.1.1 Aprendizado Métrico Profundo

Cabe, inicialmente, destacar que o Aprendizado Métrico Profundo é realizado em duas

SNNs distintas: SNN (A) e SNN (B). A SNN (A) é treinada sob categorias de gêneros da planta com ponto de vista geral (imagens inteiras) da folha, enquanto a SNN (B) é treinada sob categorias de espécies da planta com ponto de vista local (imagens recortadas) da folha. A Figura 31 apresenta a forma como é realizado o aprendizado métrico da SNN (A), treinada sob o ponto de vista geral, embora o mesmo procedimento seja realizado para SNN (B), utilizando ponto de vista local (imagens recortadas) da folha.

Na Figura 31, a Rede Neural Siamesa (SNN) proposta utiliza duas redes idênticas em paralelo (Sub-Rede 1 e Sub-Rede 2), ambas compartilhando a mesma arquitetura e pesos. Cada rede, porém, aceita uma imagem de entrada diferente, sendo as respectivas saídas combinadas de forma a realizar uma predição final. Mais especificamente, o objetivo é ter duas redes neurais idênticas que recebem a entrada de pares de imagens e que aprendem uma função para produzir uma métrica de similaridade como saída.

Figura 31. Arquitetura da Rede Neural Siamesa profunda para o reconhecimento de plantas



Fonte: elaboração própria (2021).

A grande preocupação é com a forma de treinar tais redes combinadas a fim de que possam aprender uma função de similaridade. Idealmente, uma Rede Neural Convolucional poderia ser utilizada sem qualquer restrição. Seria desejável, contudo, utilizar um modelo CNN personalizado que conseguisse atingir o mesmo nível de desempenho na representação de características de modelos pré-treinados. Tal modelo de CNN, no entanto, requer grandes volumes de dados de treinamento para produzir representações robustas. Como há conjuntos de dados limitados para plantas, aproveitou-se o poder dos modelos CNN pré-treinados sob grandes conjuntos de dados (ImageNET) (DENG *et al*., 2009) que nos últimos anos mostraram

resultados promissores na resolução de problemas de visão computacional. Dessa forma, utilizou-se uma VGG-16 pré-treinada aplicando um ajuste fino (*fine-tuning*), cujo objetivo é aprender a estimar a similaridade entre pares de imagens. Ajustamos a VGG-16 descongelando os dois últimos blocos (Bloco 4 e Bloco 5) para que seus pesos sejam atualizados em cada época à medida que treinamos nosso modelo.

Considerando que há duas imagens de entrada X_r e X_l, passou-se a primeira imagem X_r para o codificador CNN da sub-rede 1. Recebeu-se, então, uma representação de características de X_r denotado como f_r = VGG-16(X_r) onde f_r é um vetor de característica gerado da imagem X_r. Similarmente, a segunda imagem X_l é levada ao codificador CNN da sub-rede 2, compartilhando os mesmos pesos, *w*, para obter uma representação de características de X_l denotada por f_l. Com as duas representações geradas (f_r, f_l), finalmente, a camada de saída original *softmax* da rede pré-treinada VGG-16 foi substituída por uma camada que calcula uma métrica de distância *L1* (Equação 4.2) entre as últimas camadas totalmente conectadas de cada codificador CNN gêmeo.

$$L1(f_l, f_r) = \sum_{k=0}^{M-1} ||f_{l,i} - f_{r,i}|| \tag{4.2}$$

O valor de *L1* será menor se as imagens de entrada (X_r e X_l) forem similares e vice-versa. O valor de distância é incorporado à função de perda (descrita na Seção 4.5.1.2) para ajustar o codificador CNN por meio da retropropagação (*back-propagation*) e aprimorar a representação das características. Uma camada linear totalmente conectada (densa) e uma função de ativação sigmoide são utilizadas para converter a distância a uma probabilidade *p*, que indica se uma imagem de entrada pertence ou não à mesma classe-alvo. Na validação do modelo siamês, dado como entrada um conjunto de imagens de referência N_r e uma imagem de teste *t*, a classe-alvo predita é computada na equação a seguir, onde $o(t)$ denota a classe verdadeira da imagem de teste e $\hat{o}(.)$ é a classe predita.

$$\hat{o}_{N_r}(t) = o(\operatorname{argmax} p(t, X_i)),\ X_i \in N_r \tag{4.3}$$

4.5.1.2 Função de perda

A partir do uso de uma função de perda apropriada, o codificador CNN pode aprender parâmetros a fim de obter uma representação melhorada da imagem de entrada. A ideia é que o codificador CNN seja penalizado pela função de perda de acordo com as classes das imagens de entrada. Tal função faz com que o modelo crie representações de características mais

semelhantes se as classes-alvo forem iguais, e representações de características dissimilares se as classes forem diferentes. A função de perda é formulada da seguinte maneira:

$$Loss = -[y \log(d_{\mathrm{w}}) + (1 - y) \log(1 - d_{\mathrm{w}})] \tag{4.4}$$

em que o valor de *y* é o rótulo verdadeiro, que será 1 quando duas imagens de entrada forem semelhantes; 0 se forem diferentes; d_{w} é a métrica de distância entre as representações de características das imagens de entrada.

4.5.1.3 Treinamento

Detalhou-se a estratégia de treinamento deste modelo siamês no Algoritmo 1. Como se pode observar, para cada época de treinamento *numEpocas*, lotes (*Batch*) contendo pares de imagens com seus respectivos rótulos (1-classes iguais, 0-classes diferentes) são criados usando imagens do conjunto de treinamento e seu respectivo rótulo (X_T, y_T) (linha 2 do Algoritmo 1). Dessa forma, o SNN extrai vetores de características f_l e f_r dos pares de imagens de entrada $X_{esquerda}$ e $X_{direita}$, respectivamente. A linha 7, por sua vez, mensura a distância *L1* entre os vetores de características extraídos (f_l, f_r) como denotado na Equação 4.2, onde *M* é o tamanho do mapa de características.

De acordo com Aggarwal, Hinneburg e Keim (2001), a distância *L1* é consistentemente mais preferível do que outras métricas de distância (por exemplo, distância euclidiana e de cosseno) para vetores com dimensões elevadas. Por exemplo, ao usar o modelo VGG-16 para compor a SNN, tem-se vetores de características de 4.096 entradas. Além disso, o estudo que deu visibilidade mundial às Redes Siamesas utiliza a distância *L1* (KOCH; ZEMEL; SALAKHUTDINOV, 2015).

A distância computada (d_{w}, na linha 7) é a entrada da última camada da SNN, onde *Loss* é computado (linha 8) utilizando a Equação 4.4. Finalmente, na linha 9 realizou-se a atualização dos pesos e parâmetros da Rede SNN.

Tabela 6. Algoritmo 1 de treinamento

Algoritmo 1: Algoritmo de treinamento

Entrada: *numEpocas*, conjunto treinamento (X_T , y_T), *tamanho_batch*

para i = 1 **ate** *numEpocas* **faca**
  Batch = *obtenha_pares_batch* (X_T, y_T, tamanho_batch)
  para j = 1 **ate** *tamanho_batch* **faca**
    $X_{esquerda}$, $X_{direita}$, y = *Batch*(j)
    f_l = *extração_características* ($X_{esquerda}$)
    f_r = *extração_características* ($X_{direita}$)
    d_w = *distância_L1* (f_l, f_r) utilizando Equação 4.2
    δ = *computa_loss*(d_w, y) utilizando Equação 4.4
    atualiza_pesos_rede(δ)
  fim-para
fim-para

Fonte: adaptado de Snell, Swersky e Zemel (2017).

4.5.1.4 Parâmetros

As representações (vetores de características) das duas imagens de entrada foram geradas com a utilização de uma camada densa de 4.096 neurônios. Em seguida, adicionou-se uma camada personalizada ao modelo para encontrar a distância *L1*, calculando a diferença absoluta entre as representações e, finalmente, uma camada densa com uma unidade sigmoide para gerar pontuações de similaridade. O algoritmo Estocástico Gradiente Descendente (SGD), com *momentum* de 0.9, foi usado para realizar o ajuste fino do modelo SNN. No ajuste fino, apenas os pesos dos blocos 4 e 5 da VGG-16 são atualizados juntamente com as camadas densas. Além disso, uma taxa de aprendizado de 0.001 foi definida com um decaimento de 0.5 a cada 512 iterações, num total de 2.048 iterações. Hiperparâmetros como tamanho do lote (*batch-size*) e número de iterações foram definidos empiricamente como 32 e 2.048, respectivamente.

4.5.2 Estratégia de classificação hierárquica *Coarse-to-fine*

A estratégia *Coarse-to-fine* apresentada na Figura 30 considera a taxonomia botânica hierárquica das plantas. Nesse aspecto, no primeiro estágio da classificação hierárquica, classificou-se uma imagem de teste com SNN (A) utilizando gêneros, para que as suas espécies fossem definidas com a utilização da SNN (B) no segundo estágio. Para cada espécie, selecionaram-se aleatoriamente alguns exemplos rotulados para compor os subconjuntos de

imagens de referências. O número de referências por espécie foi experimentalmente definido (de uma para seis amostras), e as referências para cada gênero foram as mesmas selecionadas para cada espécie. Uma lista classificada (R_k) de referências de gênero ordenadas pela similaridade foi a saída do primeiro estágio hierárquico.

O estágio *Coarse*, baseado na representação global da folha, forma um conjunto de gêneros candidatos. Estes foram utilizados para selecionar as espécies a serem avaliadas na segunda etapa da classificação hierárquica, considerando a representação local da folha. Finalmente, o gênero (classificação *Coarse*) e as espécies (classificação *Fine*) foram combinados para produzir uma lista de classificação final *F* das melhores hipóteses de espécies de plantas.

Vale ressaltar que as referências de cada gênero foram selecionadas aleatoriamente, contemplando cada uma de suas espécies. Para tanto, o algoritmo de seleção tentou adicionar pelo menos uma referência para representar cada espécie dentro do gênero. Por exemplo, o gênero *Prunus* possui três espécies, então o algoritmo selecionou aleatoriamente duas amostras diferentes de cada espécie para formar o conjunto de seis referências. É importante dizer que os conjuntos de dados usados têm um máximo de seis espécies por gênero, no entanto, mesmo que um gênero tenha mais espécies do que o número de referências consideradas no sistema, observaram-se resultados promissores (ver os experimentos com menos de seis referências nas Tabelas 13 e 14). Na maioria dos casos, contudo, as características visuais para famílias e gêneros foram relativamente semelhantes no contexto de espécies de plantas.

4.5.3 Reconhecimento e esquema de fusão ponderada

O reconhecimento das espécies de plantas foi realizado pela fusão dos resultados do primeiro e segundo estágios da Figura 30. Considerando uma amostra de teste X_l, a saída do primeiro estágio é uma lista R_k de referências de gêneros ordenadas pela similaridade fornecida por SNN (A). É importante lembrar que é possível ter até seis referências para cada gênero, a depender do número de imagens de referência definidas. Assim, R_k é uma lista com k referências de gênero, onde cada G_{ij} é a j_{th} referência do gênero i. Pode-se, portanto, computar a frequência em que referências de gênero aparecem em R_k, o qual é usado como peso (w) para calcular a saída do segundo estágio.

No segundo estágio, o valor de similaridade S_i entre uma imagem de teste (X_l) e imagens de referência (X_r) das espécies encontradas dentro de cada gênero G_i presente no ranking R_k é mensurado por SNN (B), considerando o ponto de vista local da imagem da folha. Finalmente,

a saída do segundo estágio é uma lista de espécies *F* ordenada pelo *score* ζ, calculado conforme descrito na Equação 4.5:

$$\zeta = \frac{\mathrm{w}_i \cdot S_i}{\sum \mathrm{w}_i} \tag{4.5}$$

em que w_i é o peso do gênero i_{th} computado no primeiro estágio. Com esse esquema de fusão, foram combinados pontos de vista gerais e locais da imagem da folha, considerando uma estratégia hierárquica que reduz o número de espécies no segundo estágio e diminui a complexidade de classificação.

A saída de SNN (A) são pontuações de similaridade da vista geral, enquanto a saída do SNN (B) são pontuações de similaridade da vista local. Avaliaram-se diferentes regras (soma, produto e votação por maioria) para combinar as pontuações gerais e locais produzidas no primeiro e segundo estágios do sistema. Os melhores resultados foram observados ao usar o esquema baseado na Equação 4.5. Conforme mencionado, a frequência de gênero w é usada como peso nas suas respectivas espécies. A lógica por detrás disso é que a espécie correta, geralmente, pertence ao gênero que mais aparece na lista (R_k) de gêneros candidatos. Assim, utilizou-se tal frequência como ponderação na segunda etapa, sendo uma informação importante na decisão final da espécie.

4.6 MÉTRICAS DE AVALIAÇÃO

Foram utilizadas duas métricas diferentes, sendo uma para cada base de dados. A métrica de avaliação utilizada para o conjunto de dados PlantCLEF 2015 (GOËAU; BONNET; JOLY, 2015) foi disponibilizada pelo próprio Congresso, e é denominada classificação de *score* médio (*S*), definida na Equação 4.6:

$$S = \frac{1}{U} \sum_{u=1}^{U} \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} S_{u,p,n} \tag{4.6}$$

onde *U* é o número de usuários que capturam imagens de plantas (que têm pelo menos uma amostra de imagem nos dados de teste); (P_u) é o número de plantas individuais observadas pelo *u*-th usuário; ($N_{u,p}$) é a quantidade de fotos tiradas da *p*-th planta observada pelo *u*-th usuário; e ($S_{u,p,n}$) é a pontuação entre 0 e 1 para as *n*-th imagens tiradas da *p*-th planta observada pelo *u*-th usuário.

Para a base de dados LeafSnap, a métrica de avaliação é calculada de acordo com a Equação 4.7:

$$\text{Acc} = \frac{\text{número de exemplos classificados corretamente}}{\text{total de número de exemplos}} \tag{4.7}$$

4.7 IMPLEMENTAÇÃO

As implementações foram realizadas na linguagem *Python*, por meio do uso das bibliotecas Keras e TensorFlow. Os experimentos foram realizados em uma máquina disponibilizada pelo Programa de Pós-Graduação em Informática (PPGIa) com processador i7-3770, de 3.40 GHz (64 bits) e 16 GB de memória RAM, equipada com uma placa gráfica NVIDIA Titan X, com 3.584 núcleos e 12 GB de memória.

4.8 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados dois métodos para a tarefa de reconhecimento de espécies de plantas. O primeiro método usa Redes Neurais Convolucionais (CNN). A transferência de aprendizado, ajuste fino e o aumento de dados são utilizados para tratar categorias desbalanceadas e obter melhores representações das imagens. No segundo método utiliza-se uma rede de aprendizado métrico conhecida como Rede Neural Siamesa (SNN), que emprega duas CNNs simétricas para um par de imagens de entrada e, ao final, mensura a similaridade entre estas imagens.

Em ambos os métodos emprega-se uma classificação hierárquica que reduz a complexidade no problema de reconhecimento de plantas. Além disso, dois pontos de vista (geral e local) das imagens das plantas são considerados, os quais permitem uma representação complementar da imagem da folha.

5 RESULTADOS EXPERIMENTAIS

Neste capítulo descrevem-se os resultados experimentais para ambas as abordagens propostas (CNN e SNN) na tarefa de reconhecimento de plantas. Na Seção 5.1 é apresentado o protocolo experimental realizado sobre os conjuntos de dados PlantCLEF 2015 e LeafSnap. Posteriormente, nas Seções 5.2 e 5.3 são descritos os experimentos realizados na primeira e segunda abordagem, respectivamente. Finalmente, na Seção 5.4, comparam-se os métodos propostos com o estado da arte.

5.1 PROTOCOLO EXPERIMENTAL

Duas bases de dados foram selecionadas para formar o protocolo experimental: Plant-CLEF 2015 e LeafSnap, as quais são descritas em detalhes nas próximas seções.

5.1.1 PlantCLEF 2015

O conjunto de dados PlantCLEF 2015 (GOËAU; BONNET; JOLY, 2015) é composto por imagens de várias estruturas de plantas, tais como: frutas, flores, folhas, caules e plantas inteiras. O escopo deste trabalho está relacionado apenas à estrutura da folha da planta. Este grupo contém um total de 351 espécies distintas divididas em 12.610 imagens rotuladas em seu conjunto de treinamento. O conjunto de teste contém 221 imagens com seus rótulos verdadeiros, distribuídas em 60 espécies. As informações taxonômicas das plantas são pré-estabelecidas em XMLs, vinculados a cada imagem da base de dados. O uso de metadados permite viabilizar a separabilidade das imagens em grupos taxonômicos: Famílias, Gêneros e Espécies. A Tabela 7 mostra uma síntese do conjunto de dados original PlantCLEF 2015.

5.1.2 LeafSnap

O conjunto de dados LeafSnap abrange 184 espécies de árvores do Nordeste dos Estados Unidos, cuja base possui 7.719 imagens capturadas por dispositivos móveis. A aquisição das imagens foi realizada em ambientes externos, com isso, vários graus de ruído foram encontrados, como desfoque, variações de iluminação, sombras, etc. A divisão hierárquica dessa base de dados está organizada em 35 Famílias, 73 Gêneros e 184 Espécies.

Para avaliar o conjunto de dados LeafSnap, foram selecionadas 15 imagens aleatórias por cada classe (espécies) para compor o conjunto de teste, totalizando 2.760 imagens. Os detalhes desta base de dados são encontrados na Tabela 7.

Tabela 7. Número original de classes e imagens nos conjuntos de dados PlantCLEF 2015 e LeafSnap separados por grupos taxonômicos

Base de dados	Grupo Taxonômico	Classes Treinamento	Classes Teste	Imagens de treinamento	Imagens de teste
PlantCLEF 2015	Famílias	205	29	12.610	221
	Gêneros	260	43		
	Espécies	351	60		
LeafSnap	Famílias	35	35	7.719	2.760
	Gêneros	73	73		
	Espécies	184	184		

Fonte: dados da pesquisa (2021).

5.1.3 Organização dos dados

Definidos os dois métodos, passa-se a descrever a forma de organização dos dados de cada um deles. As abordagens diferem na forma como são conduzidos os dados de entrada na rede, para tanto, utiliza-se a técnica de aumento de dados no primeiro método (CNN) e a formação de pares de amostras no segundo (SNN).

Com relação ao aumento de dados, a Tabela 7 apresenta a organização das amostras dos conjuntos de dados antes de aplicar a estratégia de aumento dos dados. Em função do desbalanceamento entre classes, entretanto, aplicou-se a técnica de aumento de dados apenas no conjunto de dados PlantCLEF 2015, conforme mencionado na Seção 4.4.1. O conjunto de dados de treinamento original foi dividido em 70% para treinamento e 30% para validação. Após a etapa de aumento de dados foram gerados subconjuntos de treinamento e validação com 123.578 e 52.962 imagens, respectivamente, conforme detalhado na Tabela 8. Vale ressaltar que a mesma quantidade de imagens dos subconjuntos de treinamento e validação são utilizados para os diferentes pontos de vista da planta (gerais e locais).

Tabela 8. Conjuntos de treinamento e validação após aplicação da técnica de aumento de dados na base de dados PlantCLEF 2015.

Grupo Taxonômico	Classes Treinamento	Classes Teste	Imagens de treinamento	Imagens de validação
Famílias	205	29	123.578	52.962
Gêneros	260	43		
Espécies	351	60		

Fonte: dados da pesquisa (2021).

Na formação de pares de amostras, diferentemente da abordagem CNN, um dos objetivos do método SNN é reconhecer espécies de plantas com número reduzido de amostras disponíveis, ignorando o uso da técnica de aumento de dados. Dessa forma, definiu-se o número de amostras de treinamento para cada espécie entre um e seis (número mínimo de amostras de treinamento encontrada em uma classe) ao criar conjuntos de amostras rotuladas necessárias para treinar a SNN. Além disso, para avaliar a escalabilidade da SNN considerando poucas classes, reduzimos o número de classes de 351 para 60 no treinamento do modelo, embora, temos 351 saídas para uma imagem de teste. A Tabela 9 mostra o número total de amostras de treinamento por grupo taxonômico (famílias, gêneros e espécies) ao considerar até seis amostras por espécie no conjunto de dados PlantCLEF 2015 e LeafSnap.

Tabela 9. Total de imagens de treinamento por famílias, gêneros e espécies quando utilizadas seis amostras por categoria

Base de dados	Grupo Taxonômico	Número de classes	Número total de imagens
PlantCLEF 2015	Famílias	29	174
	Gêneros	43	258
	Espécies	60	360
LeafSnap	Família	35	210
	Gêneros	73	438
	Espécies	184	1.104

Fonte: dados da pesquisa (2021).

Com os conjuntos de amostras da Tabela 9, foram gerados subconjuntos de treinamento supervisionado contendo pares de imagens positivos e negativos. Um par positivo significa que duas imagens pertencem à mesma categoria sendo, então, rotulado como “1”. Caso contrário, se pertencerem a categorias diferentes, considera-se negativo e este é rotulado como “0”.

Conforme recomendado por Melekhov, Kannala e Rahtu (2016), geraram-se subconjuntos de treinamento contendo mais pares negativos do que positivos. A Tabela 10 mostra o número de amostras positivas e negativas dos subconjuntos utilizados para treinamento da SNN. O treinamento da SNN é realizado conforme descrito na Seção 4.5.1.3.

Tabela 10. Número de amostras positivas e negativas dos subconjuntos de treinamento usados no método SNN

Base de dados	Grupo Taxonômico	Exemplos positivos	Exemplos negativos
PlantCLEF 2015	Famílias	200	300
	Gêneros	400	600
	Espécies	800	1.200
LeafSnap	Família	300	450
	Gêneros	600	900
	Espécies	1.000	1.500

Fonte: dados da pesquisa (2021).

5.2 EXPERIMENTOS DA PRIMEIRA ABORDAGEM: CNN

Inicialmente, esta Seção apresenta algumas análises necessárias à definição do método que utiliza CNN, embora essas definições também sejam fundamentais à implementação do segundo método proposto (SNN). A Seção 5.2.1 descreve diferentes modelos de redes pré-treinadas (AlexNet, GoogLeNet e VGG16Net), usadas para experimentos comparativos, a fim de encontrar aquela com os melhores resultados experimentais sob as mesmas condições. A Seção 5.2.2 define a classificação hierárquica grossa-a-fina (*Coarse-to-fine*); a Seção 5.2.3 estabelece a configuração da classificação hierárquica utilizando duas visões; e a Seção 5.2.4 apresenta uma quarta análise que mostra a importância da representação proposta a partir de dois pontos de vista da imagem da folha. Finalmente, na Seção 5.2.5, valida-se o método CNN proposto.

5.2.1 Modelos pré-treinados

As performances de reconhecimento de AlexNet, GoogLeNet e VGG16 são mostradas na Tabela 11. Esses experimentos foram realizados sob o conjunto de validação da base PlantCLEF 2015, apresentado na Tabela 8, realizando um experimento preliminar que considerou apenas o ponto de vista geral das folhas. Observa-se, na Tabela 11, que a VGG16 forneceu melhores resultados nas mesmas condições experimentais, apresentando uma acurácia de 75,09%.

Tabela 11. Performance de diferentes arquiteturas CNN, considerando imagens de folhas inteiras (ponto de vista geral)

Grupo taxonômico	Modelo	Acurácia (%)
Espécies	AlexNet	71,98
	GoogLeNet	73,30
	VGG16	75,09

Fonte: dados da pesquisa (2021).

Alguns parâmetros foram utilizados em comum para comparar essas arquiteturas. Com isso, o algoritmo Estocástico Gradiente Descendente (SGD) foi implementado com *momentum* = 0.9 e o número de épocas foi definido empiricamente como 30. A taxa de aprendizagem foi iniciada com 0.001, diminuindo o seu valor por um fator de 0.5 a cada sete épocas. O tamanho do lote (*batch-size*) foi definido como 32, utilizando *dropout* = 0.5.

5.2.2 Classificação de famílias, gêneros e espécies

Nesta seção avalia-se a forma de definição dos níveis grosso-a-fino (*Coarse-to-fine)* da estratégia hierárquica proposta. Para tanto, realizou-se o reconhecimento das imagens foliares utilizando o modelo pré-treinado VGG16, considerando individualmente cada grupo taxonômico (família, gênero e espécie) sob diferentes pontos de vista: geral e local. Os resultados observados no conjunto de dados de validação PlantCLEF 2015 são mostrados na Tabela 12.

Tabela 12. Performance individual considerando cada grupo taxonômico: Famílias, Gêneros e Espécies sob diferentes pontos de vista da folha: geral e local, e com múltiplas resoluções: 32×32, 64×64 e 128×128

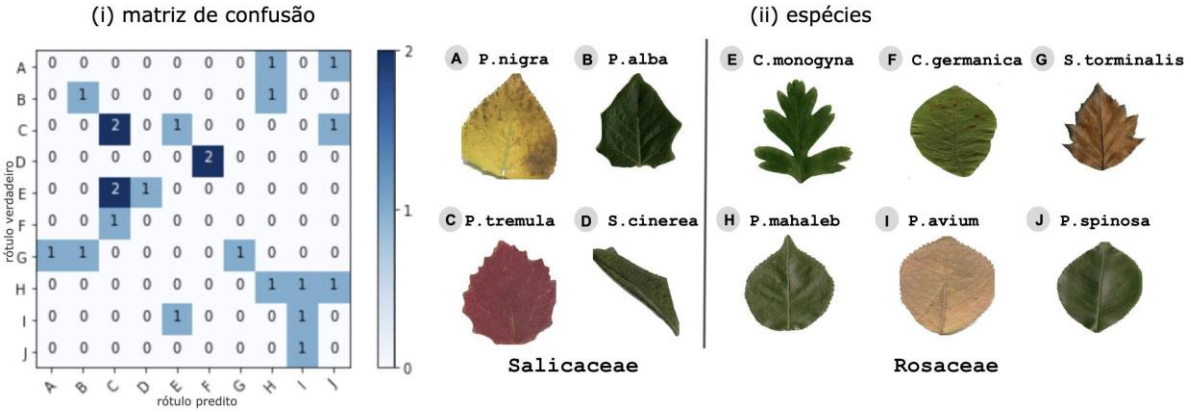
Grupo Taxonômico	**Performance (%)**			
	Visão geral	**Visão local**		
		32 x 32	**64 x 64**	**128 x 128**
Família	69,51	63,01	60,47	59,16
Gênero	85,89	67,33	64,41	62,35
Espécie	75,09	74,00	72,35	71,17

Fonte: dados da pesquisa (2021).

Individualmente, o uso do ponto de vista local (imagens cortadas) tem desempenho inferior se comparado ao uso da vista geral (imagens inteiras) em todos os grupos taxonômicos, indicando que se deve evitar o uso da vista local no primeiro estágio (*Coarse*) da hierarquia. Além disso, os resultados da Tabela 12 demonstram que o modelo VGG16 tem melhor performance ao utilizar gêneros. Da mesma forma, é possível observar que para a representação da visão local é melhor o uso de imagens recortadas de resolução 32×32 em comparação com resoluções de 64×64 e 128×128.

Para melhor compreender os resultados do reconhecimento de espécies de plantas por cada grupo taxonômico, matrizes de confusão são apresentadas nas Figuras 33 (i) e 34 (i). A Figura 33 (i) mostra uma matriz de confusão usando as famílias *Salicaceae* e *Rosaceae* avaliada sob o conjunto de teste PlantCLEF 2015. Os piores acertos são observados nas espécies *S.cinerea*, *C.monogyna* e *S.torminalis*. Percebe-se que o modelo treinado sob grupo de famílias não é capaz de classificar adequadamente as espécies de folhas que possuem diferentes características morfológicas dentro da mesma família. A espécie *S.cinerea* (Figura 33d) gerou confusão com a espécie *C.germanica* (Figura 33f), provavelmente causada pela cor verde e formato fino da folha, que é bem diferente das demais espécies da família *Salicaceae*.

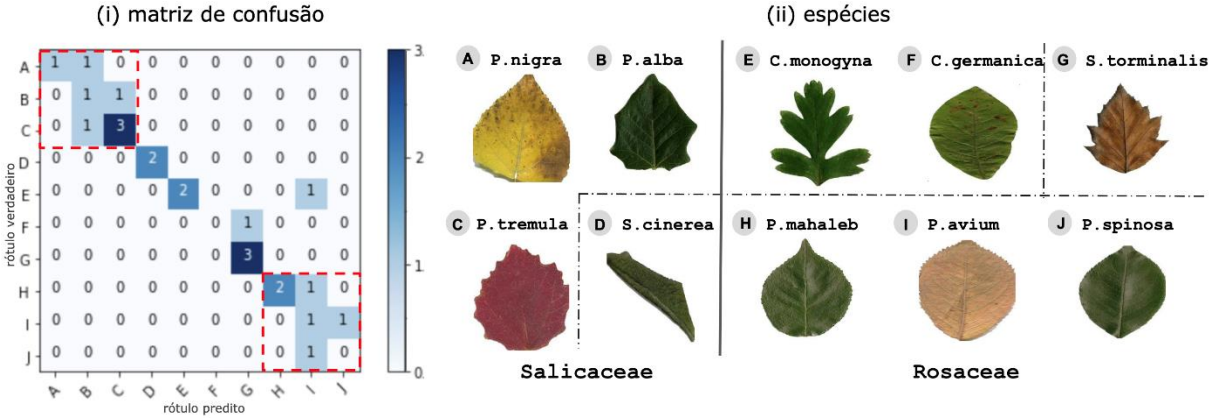
Figura 33. Performance da VGG16 treinada sob grupos de Famílias



Fonte: elaboração própria (2021).
Legenda:
(i) matriz de confusão
(ii) espécies separadas por famílias: *Salicaceae e Rosaceae*

Por outro lado, *C.monogyna* (Figura 33e) e *S.torminalis* (Figura 33g) têm características morfológicas diferentes das folhas presentes na família *Rosaceae*, causando erros de classificação. Destaca-se, todavia, que as confusões dentro dos grupos de família ocorrem entre espécies que contêm atributos distintos em termos de cor, forma e características morfológicas (desafio intraespécies).

Figura 34. Performance da VGG16 treinada sob grupos de Gêneros



Fonte: elaboração própria (2021).
Legenda:
(i) matriz de confusão
(ii) espécies separadas por famílias: *Salicaceae e Rosaceae* e gêneros *Populus, Salix, Crataegus, Sorbus e Prunus*

A Figura 34 (i) mostra uma matriz de confusão para as mesmas duas famílias *Salicaceae* e *Rosaceae*, usando agora uma VGG16 treinada sob grupos de gêneros, ou seja, separou-se as espécies em grupos de gêneros que são representados pelas linhas pontilhadas na Figura 34 (ii).

Ao separá-las, são formados dentro da família *Salicaceae*, dois gêneros (*Populus e Salix)* e três gêneros (*Crataegus, Sorbus e Prunus*) na família Rosaceae.

Observa-se, na Figura 34 (i), que o modelo VGG16 treinado com gênero melhora o desempenho de reconhecimento devido à separabilidade das características morfológicas da folha. As maiores confusões, no entanto, ocorrem com espécies semelhantes dentro do mesmo grupo de gênero (destacado por linhas pontilhadas em vermelho na Figura 34 (i). Normalmente, as espécies do mesmo gênero possuem características visuais semelhantes, tais como a forma. Por exemplo, as três espécies mais semelhantes são *P.nigra* (Figura 34a), *P.alba* (Figura 34b) e *P.tremula* (Figura 34c), pertencentes ao gênero *Populus*. Essas semelhanças refletem confusões locais, uma vez que as três espécies têm curvaturas e esboços similares.

Apesar da similaridade entre as espécies, os acertos na matriz de confusão fornecida pela VGG16, treinada sob categorias de Gênero (Figura 34 (i)), são melhores se forem comparadas ao modelo VGG16 treinado sob categorias de Família (Figura 33 (i)). Esses experimentos preliminares indicaram que a estratégia hierárquica *Coarse-to-fine* deve considerar a classificação do Gênero no primeiro estágio ao invés da Família ou Espécie. Além disso, os resultados são melhores quando empregado o ponto de vista geral das imagens da folha no primeiro estágio do que uma vista local, como destacado na Tabela 12.

Ainda existem, no entanto, erros relacionados a espécies muito semelhantes em aparência (como aquelas apresentadas nas linhas pontilhadas em vermelho na Figura 34 (i)), os quais requerem esforços adicionais para serem solucionados. Para lidar com esse problema, propõe-se combinar as características dos diferentes pontos de vista da imagem da folha por meio do emprego da classificação hierárquica *Coarse-to-fine*.

5.2.3 Classificação hierárquica sob dois pontos de vista

A Tabela 13 mostra o uso dos grupos taxonômicos por meio da utilização de combinações hierárquicas diversificadas, bem como o ponto de vista utilizado em cada estágio da hierarquia. Por exemplo, em ID #1 avaliou-se a classificação da hierarquia usando Família + Espécie, o que significa que Família é utilizada para treinar e gerar o modelo VGG16 no primeiro estágio da hierarquia, e Espécie é treinado posteriormente. A solução, portanto, é testada na forma de classificação da hierarquia *Coarse-to-fine*. Para avaliar diversificadas combinações, na quarta coluna da Tabela 13, escolheu-se o tipo de ponto de vista (geral ou local) a ser utilizado em cada estágio da hierarquia. Se duas marcações são selecionadas, significa que o ponto de vista geral é usado no estágio *Coarse* da hierarquia, e o ponto de vista

local no estágio *Fine*. Caso contrário, apenas o ponto de vista marcado é usado para ambos os níveis. Os resultados obtidos na Tabela 13 referem-se ao conjunto de dados PlantCLEF 2015.

Tabela 13. Combinações diversificadas na classificação hierárquica Coarse-*to-fine*, de acordo com diferentes grupos taxonômicos e pontos de vista das plantas

Modelo	ID #	Combinação Hierárquica	Ponto de Vista		Acurácia (%)
			Geral	Local	
VGG16	1	Família + Espécies		✓	60,21
	2	Família + Espécies	✓		62,76
	3	Família + Espécies	✓	✓	67,55
	4	Gênero + Espécies		✓	64,33
	5	Gênero + Espécies	✓		78,98
	6	Gênero + Espécies	✓	✓	83,11

Fonte: dados da pesquisa (2021).

Os piores desempenhos são obtidos quando as combinações são realizadas com grupos taxonômicos de Família definidos no primeiro estágio da classificação hierárquica (ID #1, ID #2 e ID #3), mesmo utilizando diferentes pontos de vista. O desempenho individualmente ruim do grupo Família obtido na Tabela 12 faz com que todas as combinações que usam esse grupo tenham precisões inferiores devido à propriedade da classificação hierárquica (classificação por etapas). Isso faz com que os erros persistam nos seguintes níveis de classificação, independentemente da taxonomia ou do tipo de ponto de vista usado: ID #4, ID #5 e ID #6 iniciam usando Gênero no estágio *Coarse* da classificação hierárquica e Espécies no estágio *Fine*. A performance, contudo, tem avanço suave com ID #5 e ID #6, enquanto ID #4 permanece com baixo desempenho devido ao uso do ponto de vista local em ambos os níveis da hierarquia, ignorando características discriminativas como forma e curvatura da folha. Já ID #5 usa características gerais para ambos os níveis, em que erros ocorrem no segundo estágio, pois dentro do mesmo gênero existem espécies com características semelhantes, difíceis de discriminar usando imagens inteiras (desafio intraespécies). Finalmente, ID #6 traz o melhor resultado ao usar dois pontos de vista em uma estratégia de classificação hierárquica (gênero-espécies).

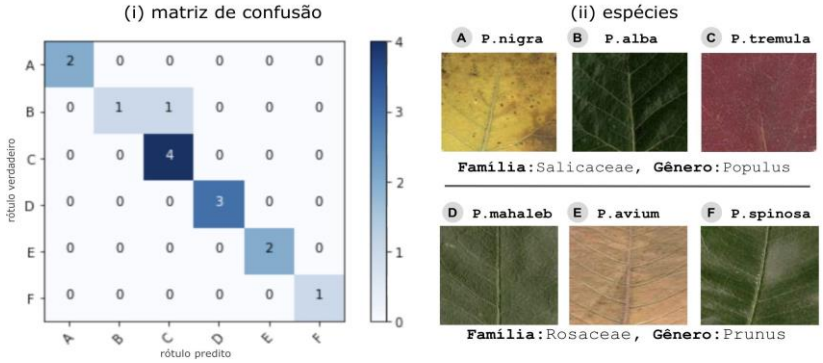
Os resultados alcançados sem a utilização da classificação hierárquica da Tabela 12 foram de 75,09%, e tiveram como saída final o uso do reconhecimento tradicional focado em grupos de espécies. Ao usar a classificação hierárquica (Tabela 13), o mesmo caso aumentou a precisão em 3,89% pontos percentuais com ID #5 (usando apenas imagens com ponto de vista geral), confirmando a suposição de que a estratégia de hierarquia pode melhorar o desempenho da classificação. Além disso, usando ID #6 a precisão aumentou em 8,02% pontos percentuais,

melhorando ainda mais a classificação com o uso do ponto de vista local da planta numa solução de classificação hierárquica *Coarse-to-fine*.

Na Figura 35, finalmente, percebe-se o impacto causado pelo uso do ponto de vista local no segundo estágio da hierarquia. Para mostrar a sua eficácia, construiu-se a matriz de confusão para casos em que os erros persistem quando utilizado apenas o ponto de vista geral. Esses casos já foram apresentados na matriz de confusão da Figura 34 (i), destacados por linhas vermelhas pontilhadas que são plotadas em nova matriz de confusão na Figura 35 (i), utilizando um novo ponto de vista (local) (Figura 35 (ii)). Constatou-se aumento no desempenho de reconhecimento em quase todas as espécies quando o ponto de vista local é empregado, bem como observou-se que as espécies do mesmo gênero (linhas vermelhas pontilhadas na Figura 34 (i)) assumem características gerais semelhantes. Quando, porém, essas espécies são plotadas no ponto de vista local (Figura 35 (ii)), elas apresentam diferenças mínimas e discriminantes entre si. A extração de características locais (padrões de veias e textura da folha da planta) em imagens recortadas possibilita distinguir espécies semelhantes. Comparados com o ID #5 na Tabela 13, que usa apenas o ponto de vista geral em ambos os níveis, os resultados melhoram em 4,13% ao adotar a vista local no segundo estágio da hierarquia (ID #6).

Após essas análises definiu-se o método proposto neste estudo, utilizando classificação hierárquica *Coarse-to-fine*, Gênero na classificação *Coarse* e Espécie como classificação *fine*. Empregou-se, ainda, representações diferentes da imagem da folha da planta, e pontos de vista gerais e locais para o primeiro e segundo níveis de hierarquia, respectivamente. Ambos os métodos propostos, foram avaliados mediante o emprego da classificação *Coarse-to-fine* e a utilização de dois pontos de vista.

Figura 35. Desempenho de VGG16 treinado com ponto de vista local em grupos de Gênero

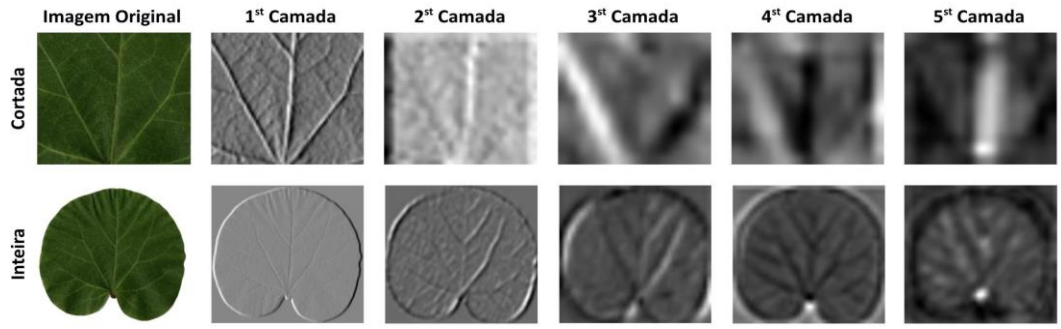


Fonte: elaboração própria (2021).
Legenda:
(i) matriz de confusão
(ii) Espécies separadas por famílias: *Salicaceae* e *Rosaceae* e Gêneros: *Populus e Prunus.*

5.2.4 Importância da representação de dois pontos de vista

Como é possível observar na Tabela 13, a representação com dois pontos de vista (ID #6) é uma alternativa promissora para aumentar o desempenho da classificação hierárquica proposta. Com os experimentos apresentados nas duas últimas Seções (5.2.2 e 5.2.3), foram obtidas duas intuições essenciais sobre as características das folhas. Em primeiro lugar, a forma da folha por si só não é a escolha certa para identificar plantas devido à ocorrência comum de folhas com contornos semelhantes, especialmente em espécies intimamente relacionadas. Nessas situações, o padrão de venação é uma poderosa característica discriminativa. Em segundo lugar, a estratégia de combinar características gerais e locais apresentadas nas imagens das folhas é realmente promissora. O que a CNN aprende, no entanto, em cada ponto de vista (imagem inteira e imagem cortada)? Para responder a essa pergunta foram exploradas as camadas CNN de cada representação (Figura 36).

Figura 36. Representação da planta utilizando dois pontos de vista da folha – mapa de características das camadas da CNN para imagens inteiras e cortadas



Fonte: elaboração própria (2021).

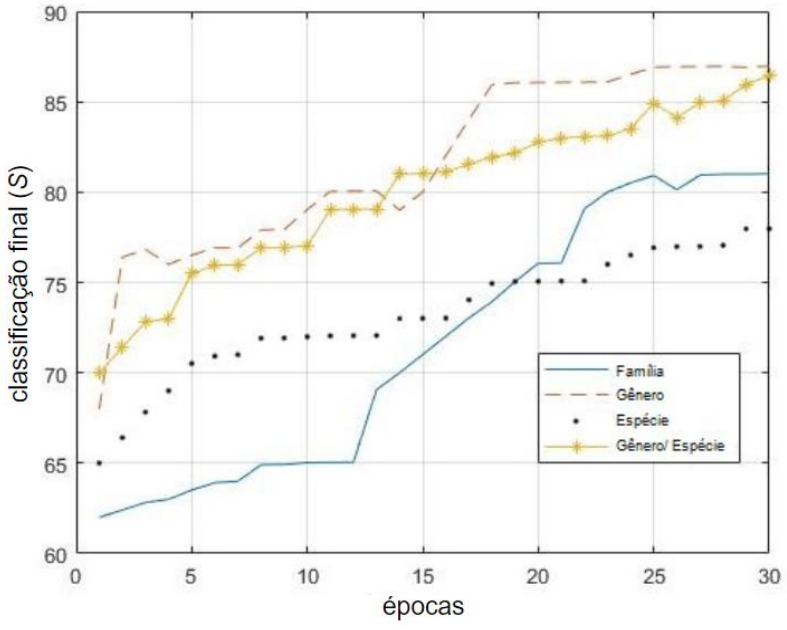
Observou-se que em cada camada da CNN, sob diferentes perspectivas, informações semelhantes são extraídas, principalmente no sentido de que ambos os modelos fornecem recursos de baixo nível nas primeiras camadas, enquanto nas camadas mais profundas é possível observar recursos mais específicos da imagem. As perspectivas diferem já que os modelos têm como entrada diferentes visões da folha da planta. Por exemplo, a primeira camada convolucional tende a extrair características de baixo nível, como bordas. É evidente, no entanto, a complementaridade entre as duas redes. A visão global fornece informações de bordas, contornos e formas da folha inteira, enquanto a visão local fornece características específicas das veias das folhas. A complementaridade pode ser observada visualmente entre cada camada correspondente. Em geral, a imagem inteira da folha (visão geral) fornece características relacionadas à forma e textura. Por outro lado, as características extraídas da

imagem da folha recortada (visão local) tendem a capturar padrões locais inerentes de venação. Essa representação de folha geral para específica permite combinar informações da forma, textura e venação, como normalmente é feito na classificação manual por taxonomistas de plantas.

5.2.5 Performance Geral – CNN

Na Figura 37 são apresentados os resultados da classificação final (utilizando a métrica *S*) de cada grupo taxonômico individualmente (Família, Gênero e Espécie), apenas empregando a combinação dos pontos de vista, realizando a fusão das probabilidades por meio da regra de soma. Os resultados obtidos sob o conjunto de teste PlantCLEF 2015 foram de 81,01%, 86,97% e 77,96% para Família, Gênero e Espécie, respectivamente, sem considerar a classificação hierárquica *Coarse-to-fine*. Como mencionado, o grupo Família foi descartado nessa abordagem de classificação hierárquica *Coarse-to-fine*, sendo a classificação iniciada pelo gênero no estágio *Coarse*, seguido pelas Espécies no estágio *Fine*. Na Figura 37, a classificação hierárquica *Coarse-to-fine* é representada pela linha amarela (Gênero/Espécie), que atingiu a classificação final (*S*) de 86,44%. Isso mostra que o método proposto foi capaz de lidar melhor com os desafios de reconhecimento de subcategorias na tarefa de reconhecimento de planta.

Figura 37. Comparação da performance de classificação dos grupos individuais (Família, Gênero e Espécie) e o método proposto hierárquico *Coarse-to-fine* (Gênero/Espécie) com dois pontos de vista sob o conjunto de dados PlantCLEF 2015



Fonte: elaboração própria (2021).

5.3 EXPERIMENTOS DA SEGUNDA ABORDAGEM: SNN

Similarmente, este método utiliza dois pontos de vista de imagens das folhas, ou seja, emprega uma classificação hierárquica *Coarse-to-fine*, considerando o Gênero no primeiro estágio da hierarquia (*Coarse*), e Espécie no segundo estágio (*Fine*). Algumas importantes modificações, entretanto, tornaram esta proposta a principal abordagem deste estudo. A primeira modificação é apresentada na Seção 5.3.1, na qual se destaca a performance da SNN, reduzindo o número de amostras de treinamento em até seis exemplos por espécie. Além disso, utilizou-se a classificação hierárquica para passar um conjunto de categorias candidatas de um estágio para outro.

Posteriormente, na Seção 5.3.2, apresenta-se o desempenho do método SNN proposto para o reconhecimento de espécies de plantas. Finalmente, avaliou-se a competência do uso de Redes Neurais Siamesas em dados desbalanceados (Seção 5.3.3), e mensurou-se a sua escalabilidade (Seção 5.3.4) e estabilidade (Seção 5.3.5).

5.3.1 Performance Geral – SNN

Nesta Seção, avalia-se a classificação hierárquica SNN, proposta com dois pontos de vista baseada em métricas de similaridade. As Tabelas 14 e 15 apresentam os resultados dos conjuntos de dados PlantCLEF 2015 e LeafSnap, respectivamente. Os melhores resultados de ambos os conjuntos de dados foram alcançados com a utilização de seis amostras de referências por cada classe (N_r = 6) e 30 referências candidatas de gênero retornadas no *ranking* de distâncias de similaridade R_k do primeiro estágio da hierarquia. No segundo estágio, a predição da imagem de teste considera o primeiro resultado retornado na lista F de espécies (top-k = 1). Além disso, alcançou-se 1.0 de performance com top-k = 5 para ambos os conjuntos de dados em suas respectivas métricas de performance, conforme relatado na Tabela 16.

Tabela 14. Performance de classificação (S) do método proposto SNN para o conjunto de dados PlantCLEF 2015

Estágio Hierárquico	**(R_k)**	**Número de Referências (N_r)**		
		1	**3**	**6**
1st (Gênero)	5	0.72	0.77	0.77
1st (Gênero)	15	0.86	0.84	0.81
1st (Gênero)	**30**	**0.98**	**0.96**	**0.95**
1st (Gênero)	50	0.99	0.98	0.98
2nd (Espécies)	1	0.81	0.86	0.87

Fonte: dados da pesquisa (2021).
Legenda:
N_r: número de amostras de referência por categoria na fase de classificação.

Tabela 15. Performance de classificação (*acc*) do método proposto SNN para o conjunto de dados LeafSnap

Estágio Hierárquico	(R_k)	Número de Referências (N_r)		
		1	**3**	**6**
1st (Gênero)	5	0.91	0.92	0.95
1st (Gênero)	15	0.96	0.96	0.98
1st (Gênero)	**30**	**0.99**	**0.98**	**0.98**
1st (Gênero)	50	0.99	0.98	0.97
2nd (Espécies)	1	0.91	0.95	0.96

Fonte: dados da pesquisa (2021).
Legenda:
N_r: número de amostras de referência por categoria na fase de classificação.

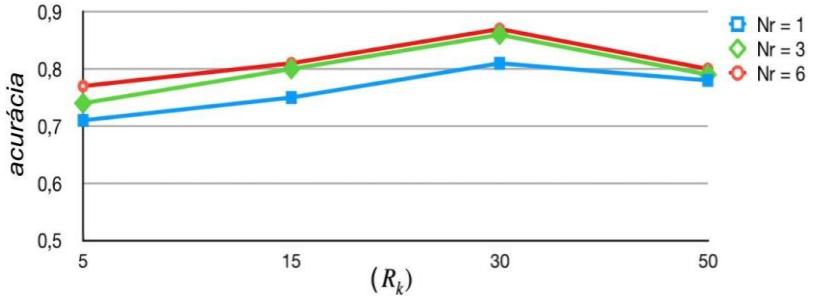
Tabela 16. Performance final do método SNN proposto considerando N_r = 6, R_k = 30 no primeiro estágio (*Coarse*) e top-k = 1, 3 e 5 no segundo estágio (*Fine*)

Base de dados	Método	Top-1	Top-3	Top-5
PlantCLEF 2015	SNN	0.87	0.94	1,0
	CNN	0.78	0.81	0.85
LeafSnap	SNN	0.96	0.99	1.0
	CNN	0.88	0.90	0.93

Fonte: dados da pesquisa (2021).
Obs.: o método CNN é comparado para cada conjunto de dados utilizando os top-5 resultados.

Importante notar que o número de referências N_r tem impacto nos resultados. Avaliou-se, portanto, na Figura 38, diferentes valores de N_r em distintas quantidades de referências de Gênero, candidatas retornadas na lista R_k. Espera-se que à medida que a lista R_k cresça, o desempenho também aumente. Nota-se, porém, a diminuição no desempenho ao usar R_k = 50, o que está diretamente relacionado à classificação hierárquica *Coarse-to-fine*, em cujo estágio *Coarse* foram definidas as classes de Gêneros a serem levadas ao estágio *Fine*, usando as referências de Gênero candidatas que constam na lista R_k.

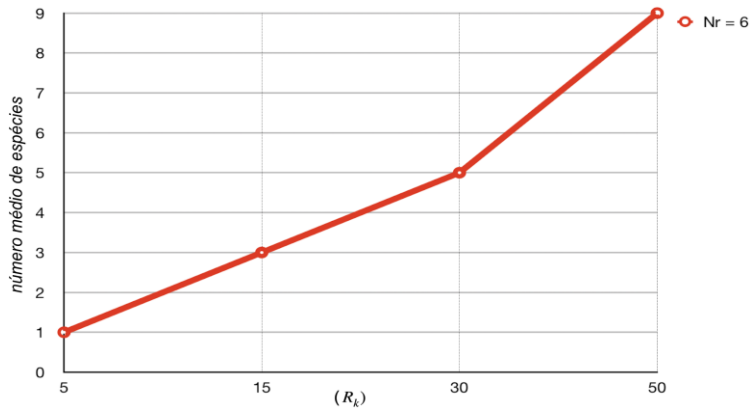
Figura 38. Precisão da classificação considerando diferentes números de referências N_r e quantidade de referências de Gêneros retornadas na lista R_k, usando o conjunto de dados PlantCLEF 2015



Fonte: dados da pesquisa (2021).

A Figura 39 visa mostrar o número de espécies (em média) avaliadas na segunda etapa do sistema ao considerar diferentes tamanhos (valores k) da lista de classificação R_k. As espécies selecionadas são aquelas que pertencem ao conjunto de gêneros formados pela lista R_k. O número de espécies pode variar de acordo com o tamanho da lista, por exemplo, com N_r = 6 (número de referências), o número médio de espécies levadas ao segundo estágio para R_k = 50, 30 e 15 é 9, 5 e 3, respectivamente. Tal análise corrobora a suposição deste estudo de que uma classificação *Coarse-to-fine* pode reduzir o número de classes que serão avaliadas na segunda etapa de classificação, minimizando a complexidade da tarefa de reconhecimento entre subcategorias.

Figura 39. Número médio de espécies fornecido pelo estágio *Coarse* ao usar (N_r = 6) em relação a diferentes quantidades (5, 15, 30, 50) de referências candidatas de Gênero retornadas na lista R_k



Fonte: dados da pesquisa (2021).

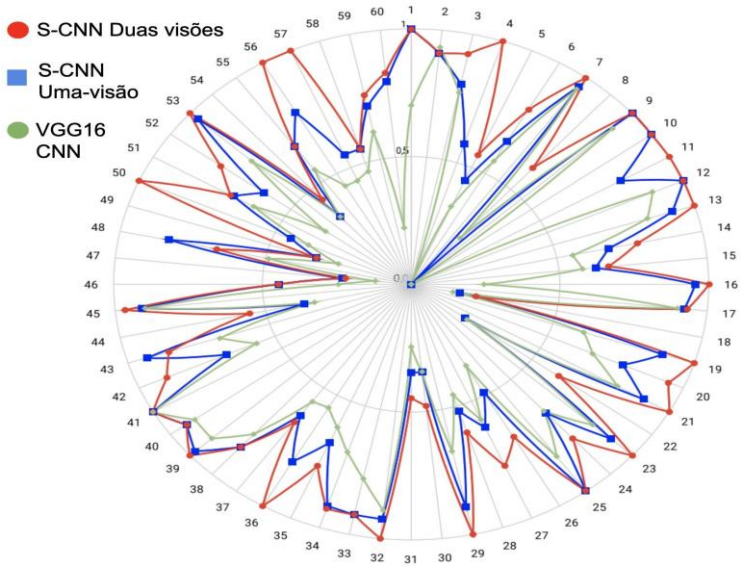
5.3.2 Análise detalhada do método proposto – SNN

A Figura 40 apresenta um gráfico de teia de aranha a fim de comparar a performance de reconhecimento para cada espécie de planta. Considerou-se o método proposto SNN usando apenas um ponto de vista (representação geral em ambos os níveis da hierarquia), com dois pontos de vista (representações gerais e locais) e o modelo VGG16 pré-treinado. Os resultados são apresentados na métrica *S* sob o conjunto de dados de teste PlantCLEF 2015.

Neste caso, o método proposto SNN, utilizando dois pontos de vista (linha vermelha) melhora as taxas de reconhecimento de várias espécies em comparação com a proposta que utiliza apenas um ponto de vista (linha azul). O uso da vista local que explora imagens de folhas recortadas foi utilizado para reduzir o conflito entre espécies semelhantes, corroborando com o experimento preliminar tratado na Seção 5.2.3. Por exemplo, a oitava espécie, denominada *quercus cerris*, tem 0.0 de precisão ao usar apenas a representação de um ponto de vista (geral). Por outro lado, a performance de reconhecimento aumentou para 0.55 quando usado o método

proposto com duas vistas. Dessa forma, quando se considera apenas uma vista (geral), a espécie *quercus cerris* se confunde com *quercus petraea*, *quercus rubra* e *quercus pubescens*.

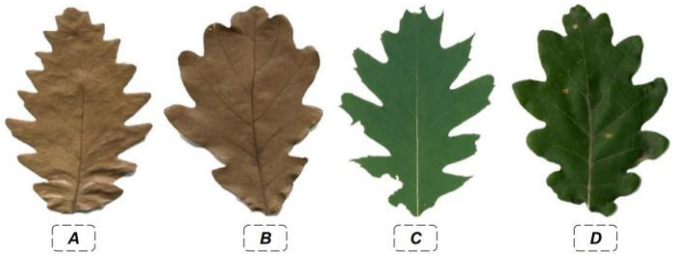
Figura 40. Comparação entre a proporção de espécies corretamente reconhecidas na classificação *Coarse-to-fine* utilizando um ponto de vista (SNN uma visão), o método proposto usando duas vistas (SNN duas visões) e VGG16 pré-treinado



Fonte: elaboração própria (2021).
Obs.: 60 espécies são avaliadas a partir do conjunto de dados PlantCLEF 2015.

A Figura 41 mostra quatro espécies de folhas, ou seja, *quercus cerris*, *quercus petraea*, *quercus rubra* e *quercus pubescens*, respectivamente. Visivelmente, essas espécies têm características morfológicas semelhantes que explicam a confusão quando apenas uma vista é adotada.

Figura 41. Amostras de espécies confusas (representação do ponto de vista geral)

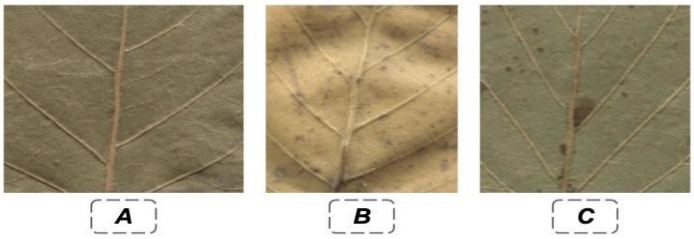


Fonte: elaboração própria (2021).
Legenda:
a) *quercus cerris*
b) *quercus petraea*
c) *quercus rubra*
d) *quercus pubescens*

É importante observar, no entanto, que a taxa de precisão de algumas espécies diminuiu com o uso do método SNN proposto e com a utilização de dois pontos de vista. Este é o caso

da quadragésima oitava espécie da Figura 40, que é a *betula pendula*. Observou-se, ainda, que há confusão entre *betula pendula*, *betula betulus* e *betula avellana*, pois elas têm texturas e padrões de veias muito semelhantes, conforme mostra a Figura 42.

Figura 42. Amostras de espécies confusas (representação do ponto de vista local)



Fonte: elaboração própria (2021).
Legenda:
a) *betula pendula*
b) *betula betulus*
c) *betula avellana*

Por fim, foram comparadas as abordagens SNN com o modelo VGG16 pré-treinado, apresentado na Seção 5.2.5. A Tabela 17 mostra as precisões finais de VGG16, SNN (uma vista) e SNN (duas vistas) sob o conjunto de testes para ambos os conjuntos de dados: PlantCLEF 2015 e LeafSnap.

Tabela 17. Precisões finais para VGG16 e SNN (uma vista geral) e o método proposto (dois pontos de vista para cada conjunto de dados) – PlantCLEF 2015 e LeafSnap

Abordagem	**PlantCLEF 2015 (*S*)**	**LeafSnap (*acc*)**
VGG16	0.78	0.88
SNN um ponto de vista	0.81	0.92
SNN dois pontos de vista	**0.87**	**0.96**

Fonte: dados da pesquisa (2021).

Os resultados finais proporcionaram uma melhora nas taxas de reconhecimento com o uso do método SNN proposto, o qual garantiu a eficácia da abordagem proposta.

5.3.3 Impacto em dados desbalanceados

A Tabela 18 mostra o impacto da abordagem proposta em dados desbalanceados. Para tanto, descreve-se a precisão de cada espécie da base de dados PlantCLEF 2015 ao usar o método proposto SNN e o modelo VGG16 pré-treinado. A segunda e a terceira coluna da Tabela 18 apresentam o número de imagens de treinamento utilizados por cada espécie para ambos os modelos. É importante mencionar que o modelo VGG16 é treinado usando o número original

de amostras de treinamento, mostrado na Tabela 6 para o conjunto de dados PlantCLEF 2015, enquanto o modelo SNN é treinado usando apenas seis amostras por classe.

Observou-se que para quase todas as espécies, o desempenho do modelo SNN é igual ou melhor do que o modelo VGG16. A maior parte dos erros do modelo CNN ocorre nas espécies com poucas amostras para treinamento. A espécie *L. triphylla* (linha em negrito na Tabela 18) possui nove amostras para treinamento e sete de teste. Com isso, é mais difícil para VGG16 gerar um modelo apropriado para discriminar *L. triphylla* entre todas as espécies com tão poucas amostras disponíveis para treinamento.

Tabela 18. Espécies e número de imagens de treinamento e teste do conjunto de dados PlantCLEF 2015

Espécie	Treinamento		Teste	Acertos		Espécie	Treinamento		Teste	Acertos	
	VGG	SNN		VGG	SNN		VGG	SNN		VGG	SNN
V. opulus	60	6	9	8	8	M. papyrifera	115	6	1	1	0
V. tinus	243	6	3	3	3	M. carica	116	6	1	1	1
L. styraciflua	61	6	3	3	3	M. rubra	88	6	1	1	1
A. prostrata	6	6	1	0	1	Fra. excelsior	60	6	9	4	5
A. cotinus	165	6	4	3	3	Fra. ornus	73	6	1	1	1
A. pistacia	157	6	1	1	1	Fra. vahl	132	6	7	4	4
I. aquifolium	84	6	1	1	1	O. europaea	311	6	1	1	1
H. helix	264	6	9	5	6	S. bulgaris	101	6	3	2	3
R. aculeatus	290	6	25	22	23	P. hispanica	119	6	5	4	4
A. trichomanes	6	6	1	0	1	C. monogyna	197	6	3	2	2
A. vulgaris	6	6	1	0	1	C. germanica	50	6	1	1	1
B. glutinosa	51	6	2	2	2	P. avium	67	6	2	2	2
B. pendula	122	6	6	2	4	P. mahaleb	56	6	3	2	3
B. betulus	124	6	4	3	3	P. spinosa	46	6	1	1	1
B. avellana	148	6	2	1	1	U. minor	382	6	2	2	2
C. australis	190	6	4	1	1	Po. alba	197	6	2	1	2
F. cercis	142	6	2	2	2	Po. nigra	222	6	2	1	2
F. robinia	109	6	3	3	3	Po. tremula	97	6	4	3	4
Q. cerris	125	6	9	2	5	S. cinerea	24	6	2	2	2
Q. sylvatica	96	6	2	1	2	A. pseudo.	44	6	3	2	3
Q. pubescens	104	6	6	3	2	A. sacchar.	42	6	5	2	4
Q. sativa	77	6	2	2	1	A. negundo	111	6	1	1	1
Q. petraea	76	6	1	1	1	A. platanoide	67	6	5	3	5
Q. rubra	48	6	2	1	2	A. campestre	160	6	13	9	12
G. genarium	19	6	1	1	1	A. monspess	167	6	1	1	1
Gi. Biloba	127	6	9	8	9	**L. triphylla**	**9**	**6**	**7**	**2**	**6**
L. nobilis	151	6	3	2	3	A. altissima	82	6	4	1	3
L. tulipifera	70	6	2	1	2	T. baccata	10	6	2	1	2
T. tilia	71	6	4	2	3	S. torminalis	36	6	3	3	3
T. cordata	28	6	3	1	2	B. davidii	126	6	1	1	1

Fonte: dados da pesquisa (2021).
Obs.: Os acertos são a quantidade de predição correta por espécie realizada pela abordagem proposta SNN ou VGG16.

Por outro lado, para o mesmo caso (espécie *L.triphylla*), o modelo SNN treinado com apenas seis amostras foi capaz de reconhecer corretamente seis das sete amostras de teste. A razão é que SNN não é treinado para aprender um classificador tradicional de espécies de plantas, mas uma métrica de distância para fornecer a semelhança entre duas imagens (imagem de referência e teste). O modelo SNN, portanto, é capaz de lidar com dados desbalanceados, consequentemente, usa um total de 360 imagens para treinar o modelo, e empregando apenas seis amostras por classe atingiu um total de 182 acertos, enquanto VGG16 acertou 147 amostras de teste utilizando 12.610 imagens de treinamento.

Da mesma forma, foram realizados experimentos sob dados desbalanceados no conjunto de dados LeafSnap. A Tabela 19 mostra o uso de diferentes tamanhos de subconjuntos de treinamento, iniciando em 1, 3, 6, 10, 15, 25 e entre 30 a 300 amostras por cada espécie. Para testar o conjunto de dados LeafSnap, selecionaram-se aleatoriamente 15 amostras de cada classe para compor o conjunto de teste.

Tabela 19. Precisão alcançada para diferentes quantidades de imagens de treinamento por classe (subconjuntos) utilizando o método proposto SNN e o modelo VGG16 para o conjunto de dados LeafSnap

Experimento	**Subconjunto**		**SNN**	**VGG16**
	Treinamento	**Teste**	**(*acc*)**	**(**acc**)**
#1	1	15	0.51	0.53
#2	3	15	0.85	0.52
#3	6	15	0.96	0.53
#4	10	15	0.94	0.65
#5	15	15	0.92	0.78
#6	25	15	0.91	0.84
#7	[30-300]	15	0.88	0.88

Fonte: dados da pesquisa (2021).

Em relação ao primeiro experimento (#1) da Tabela 19 destaca-se a dificuldade de obter um bom classificador para os métodos propostos, uma vez que SNN precisa de mais de um par de amostras positivas e negativas para treinar e atualizar a função de perda para convergir o modelo siamês, enquanto o modelo VGG16 requer um número substancial de amostras de treinamento para fornecer resultados sólidos. A principal vantagem do SNN é observada nos experimentos #2 e #3, os quais superam amplamente o modelo VGG16, empregando poucas amostras de treinamento. O desempenho do SNN começa a cair nos experimentos #4, #5, #6 e #7, devido a grande quantidade de pares de imagens (positivas e negativas) que são gerados para treinamento, causando *over-fitting* na Rede Siamesa. Diferentemente, o desempenho do

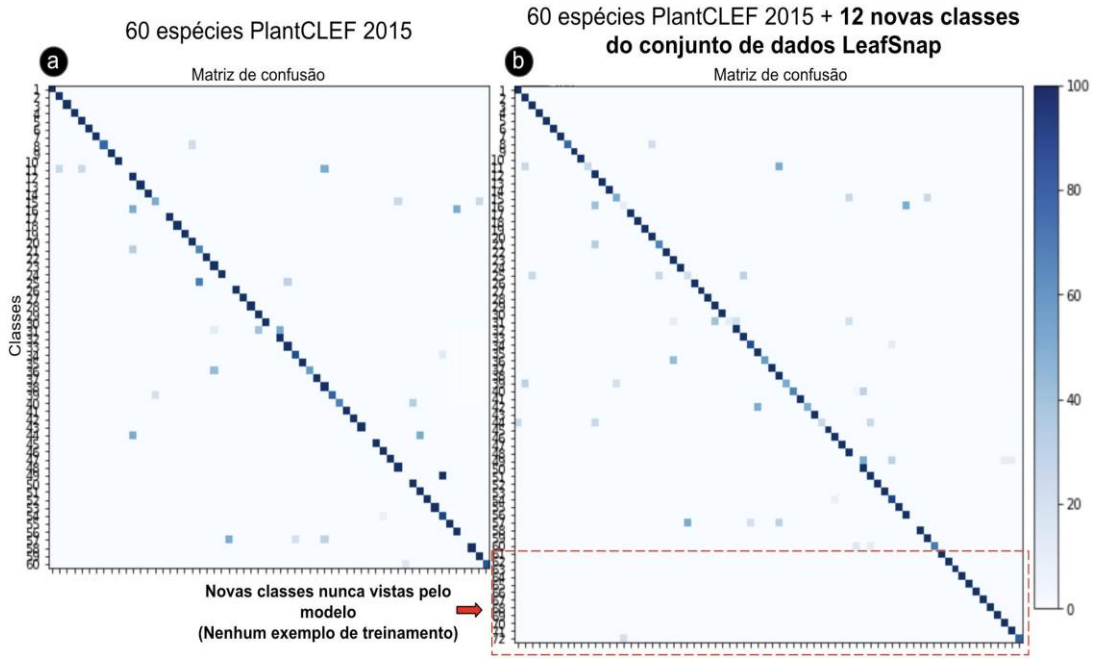
VGG16 começa a aumentar no experimento #6, alcançando 0.84 de precisão (*acc*) ao usar 25 amostras de treinamento por espécie.

No experimento #7 tem-se o melhor resultado para VGG16 com 0.88 de precisão quando utilizadas de 30 a 300 amostras de treinamento. SNN, porém, supera VGG16 no experimento #3, atingindo 0.96 de precisão usando apenas seis amostras de treinamento por classe. Esses experimentos corroboram a competência de SNN em lidar com dados desbalanceados.

5.3.4 Escalabilidade

A escalabilidade do método proposto pode ser avaliada considerando espécies de plantas não vistas durante a etapa de treinamento. Para alcançar esse objetivo com o emprego da metodologia da Rede Siamesa, basta adicionar imagens de referências das novas espécies a serem consideradas. Uma das principais vantagens da abordagem proposta é que não requer o retreinamento do modelo proposto, evitando excessivo consumo de tempo.

Figura 43. Matriz de confusão com destaque à escalabilidade de novas espécies de plantas desconhecidas pelo modelo



Fonte: elaboração própria (2021).

Para realizar esse experimento, reuniram-se as duas bases de dados apresentadas (PlantCLEF 2015 e LeafSnap), e procedeu-se ao treinamento do modelo SNN apenas com as imagens da base de dados PlantCLEF 2015. Após, utilizaram-se as espécies da base de dados LeafSnap, que foram inseridas na classificação de forma independente, sem fazer o treinamento

dessas classes. Ao final, obteve-se um conjunto de larga-escala com 244 espécies para serem classificadas, sendo 60 espécies da base de dados PlantCLEF 2015 e 184 espécies do conjunto LeafSnap.

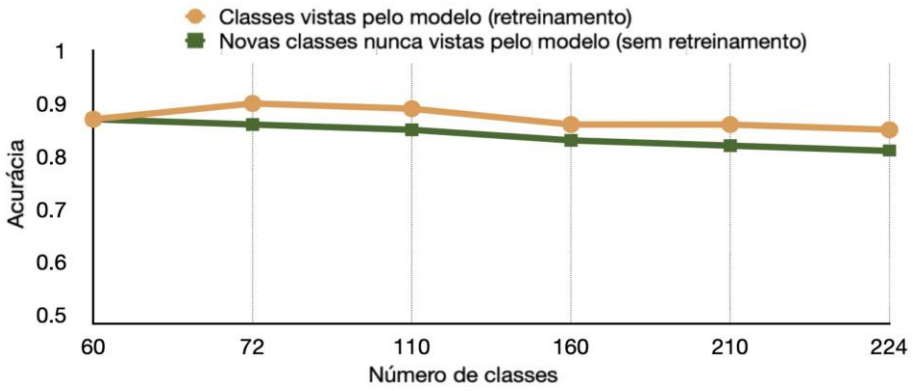
A Figura 43a apresenta a matriz de confusão do conjunto de dados PlantCLEF 2015 com 60 classes. Inicialmente, foram adicionadas 12 novas classes que nunca foram vistas pelo modelo e plotou-se novamente a matriz de confusão com 72 classes (Figura 43b). As 12 novas classes incluídas no problema foram obtidas do conjunto de dados LeafSnap. Comparando as duas matrizes de confusão, percebeu-se que a precisão antes de adicionar as novas classes era de 0.87 e depois da adição ela diminuiu para 0.86. A caixa vermelha na Figura 43b identifica as 12 novas classes do conjunto LeafSnap. A classificação para as novas classes obteve três erros em apenas uma das 12 espécies.

Após essa primeira análise percebeu-se que o método SNN pode ser escalável, entretanto, para confirmar essa hipótese, avaliou-se a escalabilidade utilizando toda a base de dados LeafSnap. Adicionaram-se, assim, 184 novas espécies nunca vistas pelo modelo, mensurando a sua adaptação na classificação. Além disso, realizou-se uma comparação entre espécies vistas (com retreinamento) e não vistas (sem retreinar) pelo modelo SNN.

A Figura 44 mostra o impacto na precisão do método proposto ao adicionar todas as novas classes não vistas pelo modelo (linha verde com quadrado). A precisão antes de adicionar as novas classes foi de 0.87 para as 60 classes do conjunto de dados PlantCLEF 2015. Após adicionar 12, 50, 100, 150 e 184 novas classes que pertencem ao conjunto de dados LeafSnap, a precisão caiu para 0.86, 0.85, 0.83, 0.82 e 0.81, respectivamente. É importante notar que, mesmo adicionando 184 novas espécies distintas de outro conjunto de dados, o método proposto manteve precisão próxima à alcançada para apenas 60 espécies, as quais nunca foram vistas pelo modelo e tampouco utilizadas em seu treinamento.

Além disso, para cada novo subconjunto adicionado (12, 50, 100, 150 e 184 novas classes) calculou-se o desempenho do sistema originalmente treinado no conjunto de dados PlantCLEF 2015 (60 espécies) com e sem processo de retreinamento. Como esperado, sempre houve melhor desempenho com o retreinamento do sistema (linha laranja com círculo). Notou-se, porém, pequena perda de desempenho (até 0.4), conforme Figura 44, que compara classes vistas (com retreinamento) e classes não vistas (sem retreinamento). Assim, este experimento mostrou que o método proposto escala relativamente bem. Ademais, mostrou que o sistema teve bom desempenho no caso de uma avaliação cruzada dos conjuntos de dados.

Figura 44. Escalabilidade do método proposto SNN, considerando espécies de folhas vistas e não vistas



Fonte: elaboração própria (2021).

A Tabela 20 avalia o tempo necessário para processar uma única imagem de teste e demonstrar se o desempenho computacional pode sofrer consumo excessivo de tempo quando aumenta a proporção da escalabilidade das classes. Para calcular o tempo computacional, a fase de testes foi avaliada e, por meio dos experimentos, observou-se que o tempo computacional cresce mais lentamente do que uma função linear à medida que aumenta o número de espécies.

Tabela 20. Tempo computacional para classificar uma folha de planta, considerando a escalabilidade de classes

Número de classes	Tempo (segundos)
60 (PlantCLEF 2015)	0.2010
72 (PlantCLEF 2015 + LeafSnap)	0.3965
110 (PlantCLEF 2015 + LeafSnap)	0.9276
160 (PlantCLEF 2015 + LeafSnap)	1.3830
210 (PlantCLEF 2015 + LeafSnap)	1.6789
244 (PlantCLEF 2015 + LeafSnap)	1.8458

Fonte: dados da pesquisa (2021).

5.3.5 Estabilidade

Como o resultado final pode ser afetado pela seleção aleatória de amostras de referência usadas para representar as espécies de plantas, avaliou-se a estabilidade do método em tal situação. Para tanto, realizaram-se cinco execuções do método proposto com diferentes conjuntos de referências de plantas (sem repetições), com aumento de dados para espécies com poucas amostras de treinamento, usando rotações na imagem da folha original, conforme apresentado na Seção 4.4.1. A Tabela 21 mostra os resultados das cinco execuções dos conjuntos de teste PlantCLEF 2015 e LeafSnap. Como se pode ver, o desempenho do método é estável mesmo usando diferentes conjuntos de referências.

Tabela 21. Execuções com conjuntos distintos de imagens de referência

Base de dados	Execução					Média
	1	2	3	4	5	
PlantCLEF 215	0.87	0.86	0.84	0.85	0.87	**0.86**
LeafSnap	0.96	0.94	0.93	0.95	0.93	**0.94**

Fonte: dados da pesquisa (2021).
Obs.: seis referências são selecionadas aleatoriamente para compor os conjuntos sem repetição de amostras de referência para cada conjunto de dados, PlantCLEF 2015 e LeafSnap

5.4 COMPARAÇÃO COM O ESTADO DA ARTE

Os resultados dos métodos propostos (SNN e CNN) foram comparados com os estudos da literatura que utilizaram os conjuntos de dados PlantCLEF 2015 e LeafSnap, conforme consta na Tabela 22. Cinco dos sete estudos de reconhecimento de plantas usam modelos CNN (SUNGBIN, 2015; LEE *et al*., 2017b; GHAZI; YANIKOGLU; APTOULA, 2017; BARRÉ *et al*., 2017; BODHWANI; ACHARJYA; BODHWANI, 2019), enquanto os outros dois estudos usam modelos métricos SNN (ZHI-YONG *et al*., 2018; WANG; WANG, 2019).

Tabela 22. Comparação com o estado da arte na tarefa de reconhecimento de plantas para conjuntos de dados PlantCLEF 2015 e LeafSnap

Trabalho	Abordagem	PlantCLEF 2015 (*S*)	LeafSnap (*acc*)
(SUNGBIN, 2015)	CNN	0.76	-
(LEE *et al*., 2017ab)	CNN	0.80	-
Método Proposto (CNN)	**CNN**	**0.86**	-
(GHAZI; YANIKOGLU; APTOULA, 2017)	CNN	0.84	-
(ZHI-YONG *et al*., 2018)	SNN	0.84	-
(BARRÉ *et al*., 2017)	CNN	-	0.86
(BODHWANI; ACHARJYA; BODHWANI, 2019)	CNN	-	0.93
(WANG; WANG, 2019)	SNN	-	0.91
Método Proposto (SNN)	**SNN**	**0.87**	**0.96**

Fonte: dados da pesquisa (2021).

Constata-se que ambos os métodos propostos relatam quase os mesmos resultados e não há diferenças significativas nos resultados para a base PlantCLEF 2015 (0.86 para CNN e 0.87 para SNN). A abordagem principal SNN, no entanto, funciona melhor em cenários com dados desbalanceados, pois o uso da arquitetura SNN permite treinar o modelo com poucas amostras, evitando grande custo computacional devido à estratégia de aumento de dados usada na abordagem CNN. Além disso, SNN torna o método escalonável, em que novas espécies de plantas podem ser facilmente integradas ao sistema sem serem retreinadas.

Há pontos positivos em relação aos métodos SNN existentes na literatura (ZHI-YONG *et al*., 2018; WANG; WANG, 2019). Para contornar o problema de reconhecimento de plantas, Zhi-Yong *et al*. (2018) utilizaram informações adicionais, considerando múltiplas visualizações dos órgãos das plantas, como a folha, planta inteira e flores. Diferentemente, a proposta desta tese utilizou apenas as características disponíveis nas folhas, das quais foram extraídas duas diferentes representações (características gerais e locais). Provou-se, assim, a capacidade de minimizar a dificuldade de reconhecimento de plantas com características semelhantes.

Wang e Wang (2019), por sua vez, mostraram o poder do SNN em representar categorias desbalanceadas usando um pequeno subconjunto de amostras por classe. Os experimentos avaliados pelos autores utilizaram apenas 10 classes de plantas com diferentes subconjuntos de treinamento (5, 10, 15 e 20 amostras de treinamento), nos quais os melhores resultados são realizados com 20 amostras por classes.

Contrariamente, o método SNN aplicado nesta tese usou apenas seis amostras de treinamento por cada classe, sendo avaliados conjuntos de dados bem definidos com uma grande quantidade de classes. O uso da Rede Siamesa, contudo, empregada numa solução hierárquica com dois pontos de vista, ajuda a aumentar o desempenho, utilizando um subconjunto menor de amostras de treinamento daquele apresentado por Wang e Wang (2019). Diferentemente de todos esses métodos, utilizou-se a abordagem *Coarse-to-fine* com uma representação de duas visualizações da imagem da folha. Como se pode ver, o método proposto SNN supera os estudos relacionados, além de produzir solução escalável.

5.5 CONSIDERAÇÕES FINAIS

Neste capítulo objetivou-se avaliar e discutir os métodos propostos CNN e SNN sob duas bases de dados de reconhecimento de espécies de plantas: PlantCLEF 2015 e LeafSnap. Para atingir este objetivo investigou-se a possibilidade do uso de uma classificação hierárquica mostrar resultados superiores a uma classificação tradicional. Além disso, consideraram-se dois pontos de vista diferentes das plantas: uma vista geral, que apresenta características gerais da planta (contornos, bordas e formas), e uma vista local, que foca em características específicas, tais como texturas e padrões das veias da folha. Baseados nesses resultados experimentais, a proposta de classificação hierárquica, juntamente com a representação de duas vistas, mostrou-se alternativa interessante para tratar os desafios relacionados ao reconhecimento de espécies de plantas. Além dos resultados promissores, algumas lacunas na primeira abordagem proposta (CNN) ficaram em aberto: (1) uso de um grande volume de dados de treinamento; (2) fusões

entre os estágios da classificação hierárquica; (3) escalabilidade do sistema.

Na busca por uma nova solução, a hipótese criada por esta tese foi desenvolver um sistema escalável. Para isso, propôs-se a segunda abordagem que utilizou Redes Siamesas (SNN), a qual demonstrou por meio dos experimentos realizados, a possibilidade de fornecer as contribuições esperadas, permitindo avaliar o impacto em dados desbalanceados que envolvem um pequeno número de amostras para representar cada categoria da base dados. Além disso, evitou-se o uso de uma técnica de aumento de dados para criar um sistema robusto de dados desbalanceados.

O uso de uma fusão ponderada demonstra que a passagem entre os estágios hierárquicos, criando um conjunto de classes de confiança, pode aumentar a performance dos resultados e diminuir a complexidade do problema entre classes. Por fim, avaliou-se a habilidade da SNN em criar um sistema escalável, inserindo novas espécies de folhas de planta sem a necessidade de um retreinamento do sistema.

Na comparação com estudos correlatos, apenas três empregaram SNN e obtiveram resultados promissores. A abordagem desta tese difere no quesito de apresentar a escalabilidade do modelo proposto, uma vez que foram utilizadas até seis amostras de treinamento por categoria. Os resultados da primeira abordagem CNN, portanto, foram importantes para destacar algumas vertentes, tornando o método proposto SNN a principal abordagem para a tarefa de reconhecimento de plantas.

6 CONCLUSÃO

Neste estudo foram apresentadas duas abordagens para a tarefa de reconhecimento de plantas por meio da imagem da folha. Na primeira destacou-se o uso de um modelo profundo CNN tradicional, que teve por objetivo empregar uma classificação hierárquica baseada na taxonomia das plantas para minimizar a complexidade da classificação. Além disso, dois pontos de vista da planta foram empregados para extrair características discriminantes da folha, os quais contornam os problemas relacionados ao desafio de reconhecimento de subcategorias.

Resultados experimentais da primeira abordagem indicaram que ao usar a taxonomia das plantas para realizar uma classificação hierárquica foram minimizados os problemas interespécies e intraespécies encontrados na tarefa de reconhecimento. Com isso, concluiu-se que com a extração de dois pontos de vista da imagem da folha, como características gerais (formas, contornos e cores) e locais (textura e padrões de veias), é possível prover a complementaridade à representação do problema, melhorando os resultados de classificação.

Apesar dos resultados promissores, as limitações da abordagem baseada em CNN estão relacionadas a necessidade de grandes volumes de dados de treinamento não disponíveis no caso de reconhecimento de plantas, assim como a sua baixa escalabilidade. A solução baseada em SNN foi apresentada como alternativa e demonstrou nos experimentos realizados ser capaz fornecer resultados competitivos utilizando poucas imagens de treinamento e garantindo escalabilidade.

A SNN foi aplicada com sucesso utilizando pequenos subconjuntos de dados de imagens de folhas de plantas que possuem uma a seis imagens de referência por espécie. O aprendizado métrico profundo da SNN é responsável por reconhecer espécies de plantas com pequeno número de exemplos rotulados, as quais apresentaram resultados melhores em relação a abordagens que utilizam a técnica de aumento de dados, como visto nos resultados experimentais deste trabalho. Os resultados também demonstraram que a SNN contorna problemas de base de dados desbalanceados, uma vez que não é treinada para aprender um classificador tradicional regular de espécies de plantas, mas uma métrica de distância para fornecer semelhança ou dissemelhança entre pares de imagens.

Além disso, o SNN torna o método proposto escalável, e novas espécies de plantas podem ser facilmente adicionadas ao sistema sem realizar o retreinamento do modelo. Para validar essa suposição, usou-se a SNN treinada sobre o conjunto de dados PlantCLEF 2015 para 60 espécies, e 184 novas espécies do conjunto de dados LeafSnap foram adicionadas para

classificação, sem realizar o retreinamento do modelo. A precisão antes e após a adição de novas espécies foi de 0.87 e 0.81, respectivamente. O método SNN, portanto, consegue reconhecer espécies não vistas durante a fase de treinamento e atingir a escalabilidade do sistema.

Planeja-se, num trabalho futuro, lidar com codificadores automáticos (*autoencoders*) para aprender representações dentro de uma arquitetura SNN com propriedade hierárquica e, assim, reconhecer plantas a partir da utilização do componente folha. A avaliação de outros elementos da planta tais como: caule, fruto, flor e planta inteira também se torna necessária, a combinação de outros elementos podem ajudar na precisão do reconhecimento das plantas. Além disso viabilizar o uso de outras arquiteturas recentes de redes neurais convolucionais (ResNet, ResNeXt e DenseNet) na metodologia proposta. Implementar um mecanismo de rejeição por limiar de similaridade. Com isso, caso haja a insuficiência de dados de referências para uma nova espécie, o sistema pode rejeitar e avisar que a nova espécie inserida não se encontra no modelo treinado. Finalmente, avaliar o método proposto em outros conjuntos de dados, afim de constatar sua eficácia em demais variados contextos.

REFERÊNCIAS

ABDEL-HAMID, O. *et al.* Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, v. 22, n. 10, 2014, pp. 1533-1545.

ADAM, Z. *et al.* Ficus deltoidea: a potential alternative medicine for Diabetes Mellitus. *Evidence-Based Complementary and Alternative Medicine.* Hindawi Publishing Corporation, v. 2012, 2012.

AGGARWAL, C. C.; HINNEBURG, A.; KEIM, D. A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. *In*: BUSSCHE, J.; VIANU, V. (Eds.). *Database Theory — ICDT 2001*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 420-434. ISBN 978-3-540-44503-6.

AGPGROUP, T. A. P. G. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: Apg iii. *Botanical Journal of the Linnean Society*, v. 161, n. 2, 2009, pp. 105-121.

ALOM, M. Z. *et al*. The history began from alexnet: a comprehensive survey on deep learning approaches. *CoRR*, abs/1803.01164, 2018.

ALWZWAZY, H. A. *et al*. Handwritten digit recognition using convolutional neural networks. *International Journal of Innovative Research in Computer and Communication Engineering*, v. 4, 2016, pp. 101-106.

ARAÚJO, V. M. *et al*. Multiple classifier system for plant leaf recognition. *In*: *2017 IEEE Int'l Conference on Systems, Man and Cybernetics (SMC)*. [S.l.: s.n.], 2017, pp. 1880-1885.

ARAÚJO, V. M. *et al*. Fine-grained hierarchical classification of plant leaf images using fusion of deep models. *In*: *IEEE 30th International Conference on Tools with Artificial Intelligence, ICTAI 2018*, 5-7 November 2018, Volos, Greece, [s.n.], 2018, pp. 1-5.

BARRÉ, P. *et al*. LeafNet: a computer vision system for automatic plant species identification. *Ecological Informatics*, v. 40, May 2017, pp. 50-56. ISSN 15749541.

BELLET, A.; HABRARD, A.; SEBBAN, M. *A survey on metric learning for feature vectors and structured data*, 2013.

BLOICE, M. D.; STOCKER, C.; HOLZINGER, A. *Augmentor:* an image augmentation library for machine learning, 2017.

BODHWANI, V.; ACHARJYA, D. P.; BODHWANI, U. Deep residual networks for plant identification. *Procedia Computer Science.* Elsevier B.V., v. 152, 2019, pp. 186-194. ISSN 18770509.

BOTTOU, L. Stochastic Gradient Descent Tricks. *In*: BOTTOU, L. *Neural Networks:* tricks of the trade. Second Edition. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 421-436. ISBN 978-3-642-35289-8.

BROMLEY, J. *et al*. Signature verification using a “siamese” time delay neural network. *In*: *Proceedings of the 6th International Conference on Neural Information Processing Systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. (NIPS’93), pp. 737-744.

CAGLAYAN, A.; GUCLU, O.; CAN, A. B. A plant recognition approach using shape and color features in leaf images. *In*: PETROSINO, A. (Ed.). *Image Analysis and Processing – ICIAP 2013*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 161-170. ISBN 978-3-642-41184-7.

CAO, J.; WANG, B.; BROWN, D. Similarity based leaf image retrieval using multiscale R-angle description. *Information Sciences.* Elsevier Inc., v. 374, 2016, pp. 51-64. ISSN 0020-0255.

CARVALHO, M. *et al.* Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. 2007.

CERUTTI, G. *et al*. Understanding leaves in natural images – A model-based approach for tree species identification. *Computer Vision and Image Understanding.* Elsevier Inc., v. 117, n. 10, 2013, pp. 1482-1501. ISSN 10773142.

CHAKI, J.; PAREKH, R.; BHATTACHARYA, S. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, v. 58, 2015, pp. 61-68. ISSN 0167-8655.

CHAKI, J.; PAREKH, R.; BHATTACHARYA, S. Plant leaf classification using multiple descriptors: a hierarchical approach. *Journal of King Saud University – Computer and Information Sciences*, 2018. ISSN 1319-1578.

CHARTERS, J. *et al*. Eagle: a novel descriptor for identifying plant species using leaf lamina vascular features. *In*: *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. [S.l.: s.n.], 2014, pp. 1-6. ISSN 1945-7871.

CHEN, Q. *et al*. IBM research Australia at lifeclef2014: Plant identification task. *In*: *CLEF*. [S.l.: s.n.], 2014.

CHULIF, S. *et al.* Plant identication on amazonian and guiana shield flora: neuon submission to lifeclef 2019 plant. *In*: *CLEF*, [S.l.: s.n.], 2019.

CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. *Multi-column Deep Neural Networks for Image Classification*, 2012.

COPE, J. S. *et al*. Plant species identification using digital morphometrics: a review. *Expert Systems with Applications*, v. 39, n. 8, 2012, pp. 7562-7573. ISSN 0957-4174.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, v. 2, n. 4, 1989, pp. 303-314. ISSN 1435-568X.

DENG, J. *et al*. ImageNet: a Large-Scale Hierarchical Image Database. *In*: *CVPR09*. 2009.

ELHARIRI, E.; EL-BENDARY, N.; HASSANIEN, A. E. Plant classification system based on leaf features. *In*: *2014 9th International Conference on Computer Engineering Systems (ICCES)*. [S.l.: s.n.], 2014, pp. 271-276.

ELPEL, T. J. *Botany in a Day:* the patterns method of plant identification. 6th, 2013.

FERGUS, R. *et al*. Semantic label sharing for learning with many categories. *In*: DANIILIDIS, K.; MARAGOS, P.; PARAGIOS, N. (Ed.). *Computer Vision – ECCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 762-775. ISBN 978-3-642-15549-9.

FIGUEROA-MATA, G.; MATA-MONTERO, E. Using a convolutional siamese network for image-based plant species identification with small datasets. *Biomimetics*, v. 5, n. 1, 2020.

GE, Z. *et al*. Fine-grained classification via mixture of deep convolutional neural networks. *In*: *2016 IEEE Winter Conference on Applications of Computer Vision.* IEEE, 2016, pp. 1-6.

GHASAB, M. A. J. *et al*. Feature decision-making ant colony optimization system for an automated recognition of plant species. *Expert Systems with Applications*, Elsevier Ltd, v. 42, n. 5, 2015, pp. 2361-2370. ISSN 09574174.

GHAZI, M. M.; YANIKOGLU, B.; APTOULA, E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*, v. 235, 2017, pp. 228-235. ISSN 0925-2312.

GOËAU, H.; BONNET, P.; JOLY. Plant identication in an open-world (lifeclef 2016). *In*: *CLEF'2016:* Conference and Labs of the Evaluation Forum. [S.l.: s.n.], 2016, pp. 28-439.

GOËAU, H.; BONNET, P.; JOLY, A. LifeCLEF Plant Identification Task 2015. *In*: *CLEF:* Conference and Labs of the Evaluation Forum. Toulouse, France: [s.n.], 2015.

GOËAU, H. *et al*. The ImageCLEF 2011 plant images classification task. *In*: *CLEF'2011: Conference and Labs of the Evaluation Forum*. [S.l.: s.n.], 2011, pp. 1-23.

GOËAU, H. *et al*. The ImageCLEF 2012 plant identification task. *In*: *CLEF'2012: Conference and Labs of the Evaluation Forum*. 2012. ISSN 16130073.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*, 2016.

GRINBLAT, G. L. *et al.* Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, v. 127, 2016, pp. 418-424. ISSN 0168-1699.

GUO, Y. *et al*. Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. *CoRR*, abs/1607.08221, 2016.

HANG, S. T.; TATSUMA, A.; MASAKI, A. Bluefield (kde tut) at lifeclef 2016 plant identification task. *CLEF (Working Notes)*, 2016.

HE, G. et al. Feature Selection-Based Hierarchical Deep Network for Image Classification. IEEE Access, v. 8, p. 15436–15447, 2020. ISSN 21693536.

HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, 2009, pp. 1263-1284.

HERMANN MINKOWSKY. Geometrie der Zahlen, volume 1, 1910.

HOGG, R. V.; MCKEAN, J. W.; CRAIG, A. T. *Introduction to mathematical statistics*. [S.l.: s.n.], 2019. ISBN 9780134686998 0134686993.

HORN, G. V.; PERONA, P. The devil is in the tails: fine-grained classification in the wild. *ArXiv*, abs/1709.01450, 2017.

HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, v. 160, n. 1, jan 1962, pp. 106-154. ISSN 0022-3751.

JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem: a systematic study. *Intell. Data Anal.* IOS Press, NLD, v. 6, n. 5, out. 2002, pp. 429-449. ISSN 1088-467X.

JOLY, A. *et al*. Interactive plant identification based on social image data. *Ecological Informatics*, v. 23, 2014, pp. 22-34. ISSN 15749541.

JUDD, W. *et al*. *Taxonomy*. In Plant Systematics – a phylogenetic approach. 3th ed. *In*: JUDD, W. *et al*. [S.l.: s.n.], 2007.

KADIR, A. *et al.* Leaf classification using shape, color, and texture features. *CoRR*, 2014.

KAYA, M.; BILGE, H. Deep metric learning: A survey. *Symmetry*, v. 11, n. 9, 2019.

KEBAPCI, H.; YANIKOGLU, B.; UNAL, G. Plant image retrieval using color, shape and texture features. *The Computer Journal*, v. 54, n. 9, Sept. 2011, pp. 1475-1490. ISSN 0010-4620.

KENDALL, M. G.; STUART, A.; ORD, J. K. *Kendall's Advanced Theory of Statistics*. USA: Oxford University Press Inc., 1987. ISBN 0195205618.

KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R. Siamese neural networks for one-shot image recognition. 2015.

KRAUSE, J. *et al*. The unreasonable effectiveness of noisy data for fine-grained recognition. *CoRR*, abs/1511.06789, 2015.

KRISHNA, R. *et al*. Visual genome: connecting language and vision using crowdsourced dense image annotations. *CoRR*, abs/1602.07332, 2016.

KRIZHEVSKY, A. *Learning multiple layers of features from tiny images*. [S.l.], 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Curran Associates, Inc*., 2012, pp. 1097-1105.

KROGH, A.; HERTZ, J. A. A simple weight decay can improve generalization. *In*: *Advances in Neural Information Processing Systems 4*. [S.l.]: Morgan Kaufmann, 1992, pp. 950-957.

LARESE, M. G. *et al.* Automatic classification of legumes using leaf vein image features. *Pattern Recognition*, v. 47, n. 1, 2014, pp. 158-168. ISSN 0031-3203.

LASSECK, M. Image-based plant species identification with deep convolutional neural networks. *CLEF (Working Notes)*, 2017.

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech and time series. *In*: *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998, pp. 255-258. ISBN 0262511029.

LECUN, Y.; BENGIO, Y.; HINTON, G. *Deep learning*, 2015.

LECUN, Y. *et al*. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, 1998, pp. 2278-2324.

LEE, S. H. *et al.* How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, v. 71, 2017a, pp. 1-13. ISSN 0031-3203.

LEE, S. H. *et al*. HGO-CNN: hybrid generic-organ convolutional neural network for multi-organ plant classification. *In*: *2017 IEEE Int'l Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017b, pp. 4462-4466.

LEE, S. H.; CHAN, C. S.; REMAGNINO, P. Multi-organ plant classification based on convolutional and recurrent neural networks. *IEEE Transactions on Image Processing*, v. 27, n. 9, Sept. 2018, pp. 4287-4301. ISSN 1057-7149.

LINNAEUS, C. V. Systema naturae, sive regna tria naturae systematice proposita per classes, ordines, genera, species. *In*: LINNAEUS, C. V. [S.l.: s.n.], 1758.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. *In*: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015, pp. 3431-3440. ISSN 1063-6919.

LUXBURG, U.; SCHÖLKOPF, B. Statistical learning theory: models, concepts and results. *In*: GABBAY, D. M.; HARTMANN, S.; WOODS, J. (Eds.). *Inductive Logic*. North-Holland, 2011 (Handbook of the History of Logic, v. 10), pp. 651-706.

MALLAT, S. *A Wavelet tour of signal processing.* The sparse way. 3th ed. USA: Academic Press, Inc., 2008. ISBN 0123743702.

MELEKHOV, I.; KANNALA, J.; RAHTU, E. Siamese network features for image matching. *In*: *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2016, pp. 378-383.

MELLO, R. F. On the shattering coefficient of supervised learning algorithms. *ArXiv*, abs/1911.05461, 2019.

MOUINE, S.; YAHIAOUI, I.; VERROUST-BLONDET, A. Advanced shape context for plant species identification using leaf image retrieval. *In*: *2nd ACM Int'l Conf. on Multimedia Retrieval*. [S.l.: s.n.], 2012, pp. 1-8. ISBN 978-1-4503-1329-2.

MOUINE, S.; YAHIAOUI, I.; VERROUST-BLONDET, A. A shape-based approach for leaf classification using multiscale triangular representation. *In*: *3rd ACM Int'l Conf. on Multimedia Retrieval*. [S.l.: s.n.], 2013a. p. 127–134.

MOUINE, S.; YAHIAOUI, I.; VERROUST-BLONDET, A. Combining leaf salient points and leaf contour descriptions for plant species recognition. *In*: KAMEL, M.; CAMPILHO, A. (Ed.). *Image Analysis and Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013b, pp. 205-214. ISBN 978-3-642-39094-4.

MOUINE, S.; YAHIAOUI, I.; VERROUST-BLONDET, A. Plant species recognition using spatial correlation between the leaf margin and the leaf salient points. In: *2013 IEEE International Conference on Image Processing*. [S.l.: s.n.], 2013c, pp. 1466-1470. ISSN 1522-4880.

MZOUGHI, O. et al. Semantic-based automatic structuring of leaf images for advanced plant species identification. *Multimedia Tools and Applications*, v. 75, n. 3, 2016, pp. 1615-1646.

NECULOIU, P.; VERSTEEGH, M.; ROTARU, M. Learning text similarity with Siamese recurrent networks. *In*: *Proceedings of the 1st Workshop on Representation Learning for NLP*. Berlin, Germany: Association for Computational Linguistics, 2016, pp. 148-157.

NESTEROV, Y. A method for solving the convex programming problem with convergence rate o($1/k^2$). *In*: NESTEROV, Y. [*S.l.* : *s.n.*], 1983.

PAWARA, P. *et al*. Data augmentation for plant classification. *In*: BLANC-TALON, J. *et al.* (Eds.). *Advanced Concepts for Intelligent Vision Systems*. Cham: Springer International Publishing, 2017, pp. 615-626. ISBN 978-3-319-70353-4.

PRASAD, S.; KUDIRI, K. M.; TRIPATHI, R. C. Relative sub-image based features for leaf recognition using support vector machine. *In*: *Proceedings of the 2011 International Conference on Communication, Computing & Security*. New York, NY, USA: ACM, 2011 (ICCCS '11), pp. 343-346. ISBN 978-1-4503-0464-1.

PRIYA, C. A.; BALASARAVANAN, T.; THANAMANI, A. S. An efficient leaf recognition algorithm for plant classification using support vector machine. *International Conference on Pattern Recognition, Informatics and Medical Engineering*, 2012, pp. 428-432.

PROCHNOW, A. *et al*. Bioenergy from permanent grassland – a review: 1. biogas. *Bioresource Technology*, v. 100, n. 21, 2009, pp. 4931-4944. ISSN 0960-8524.

RANZATO, M. *et al*. Efficient learning of sparse representations with an energy-based model. *In*: *Proceedings of the 19th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2006 (NIPS'06), pp. 1137-1144.

ROBBINS, H.; MONRO, S. A stochastic approximation method. *Annals of Mathematical Statistics*, v. 22, 1951, pp. 400-407.
ROLI, F.; GIACINTO, G.; VERNAZZA, G. Methods for designing multiple classifier systems. *In*: *2nd Int'l Workshop on Multiple Classifier Systems*. [S.l.: s.n.], 2001, pp. 78-87.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958, pp. 65-386.

SCHUH, R. T.; BROWER, A. V. Z. *Biological Systematics:* principles and applications. [S.l.]: Cornell University Press, 2009.

SFAR, A. R.; BOUJEMAA, N.; GEMAN, D. Confidence sets for fine-grained categorization and plant species identification. *International Journal of Computer Vision*, v. 111, n. 3, Feb. 2015, pp. 255-275. ISSN 1573-1405.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

SNELL, J.; SWERSKY, K.; ZEMEL.S.R. *Prototypical Networks for Few-shot Learning*: Advances in Neural Information Processing Systems (NIPS), 2017.

SÖDERKVIST, O. *Computer vision classification of leaves from swedish trees*, 2001, 74 p.

SRIVASTAVA, N.; SALAKHUTDINOV, R.; HINTON, G. E. Modeling documents with deep boltzmann machines. *ArXiv*, abs/1309.6865, 2013.

STÉPHANE, M. Chapter 1 - Sparse Representations. *In*: STÉPHANE, M. (Ed.). *A Wavelet Tour of Signal Processing*. 3th ed. Boston: Academic Press, 2009, pp. 1-31. ISBN 978-0-12-374370-1.

ŠULC, M.; MATAS, J. Fine-grained recognition of plants from images. *Plant Methods, BioMed Central*, v. 13, n. 1, 2017, pp. 1-14. ISSN 17464811.

ŠULC, M.; PICEK, L.; MATAS, J. Plant recognition by inception networks with test-time class prior estimation. *In*: *CLEF*. [S.l.: s.n.], 2018.

SUNGBIN, C. Plant identification with deep convolutional neural network snumedinfo at lifeclef plant identification task 2015. *CLEF (Working Notes)*, 2015.

SZEGEDY, C. *et al*. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

TAN, J. W. *et al*. Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, v. 17, n. 1, 2020, pp. 82-90. ISSN 15579964.

TIELEMAN, T.; HINTON, G. *Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude*. Coursera: Neural Networks for Machine Learning, 2012.

TSAFTARIS, S. A.; SCHARR, H. Sharing the right data right: a symbiosis with Machine learning. *Trends in plant science*, v. 24, n. 2, 2019, pp. 99-102. ISSN 1360-1385.

WANG, B. *et al*. MARCH: Multiscale-arch-height description for mobile retrieval of leaf images. *Information Sciences*, v. 302, 2015, pp. 132-148. ISSN 00200255.

WANG, B. I. N.; WANG, D. Plant leaves classification: a few-shot learning method based on Siamese Network. *IEEE Access*, IEEE, v. 7, 2019, pp. 151754-151763.

WANG, W. *et al*. Face recognition based on deep learning. *In*: ZU, Q. *et al.* (Ed.). *Human Centered Computing*. Cham: Springer International Publishing, 2015, pp. 812-820. ISBN 978-3-319-15554-8.

WERBOS, P. J. Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, v. 1, n. 4, 1988, pp. 339-356. ISSN 0893-6080.

WU, H. *et al*. Automatic leaf recognition from a big hierarchical image database. *Int. J. Intell. Syst.*, John Wiley & Sons, Inc., New York, NY, USA, v. 30, n. 8, ago. 2015, pp. 871-886. ISSN 0884-8173.

WU, S. G. *et al.* A leaf recognition algorithm for plant classification using probabilistic neural network. *ISSPIT 2007 - 2007 IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 11-16. ISSN 2162-7843.

YAN, Z. *et al.* HD-CNN: hierarchical deep convolutional neural network for large scale visual recognition. *In*: *ICCV'15:* Proc. IEEE 15th International Conf. on Computer Vision. [S.l.: s.n.], 2015.

YANG, C.; WEI, H.; YU, Q. Multiscale triangular centroid distance for shape-based plant leaf recognition. *In*: *European Conf. on Artificial Intelligence*. [S.l.: s.n.], 2016, pp. 269–276. ISBN 9781614996729.

YANG, S. *et al*. From facial parts responses to face detection: a deep learning approach. *In*: *The IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015.

YANIKOGLU, B.; APTOULA, E.; TIRKAZ, C. Automatic plant identification from photographs. *Machine Vision and Applications*, v. 25, n. 6, 2014, pp. 1369-1383. ISSN 14321769.

ZEILER, M. D. Adadelta: an adaptive learning rate method. *ArXiv*, abs/1212.5701, 2012.

ZHANG, C. *et al*. A convolutional neural network for leaves recognition using data augmentation. *In*: *2015 IEEE International Conference on Computer and Information Technology*, 2015, pp. 2143-2150.

ZHANG, H. *et al*. Aggregating diverse deep attention networks for large-scale plant species identification. *Neurocomputing*, Elsevier B.V., v. 378, 2020, pp. 283-294. ISSN 18728286.

ZHANG, K. *et al*. Residual networks of residual networks: Multilevel residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 28, n. 6, Jun. 2018, pp. 1303-1314. ISSN 1051-8215.

ZHAO, C. *et al*. Plant identification using leaf shapes – a pattern counting approach. *Pattern Recognition*, v. 48, n. 10, 2015, pp. 3203-3215. ISSN 0031-3203. Discriminative Feature Learning from Big Data for Visual Recognition.
ZHI-YONG, G. *et al*. Spatial-structure siamese network for plant identification. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 32, n. 11, 2018, p. 1850035.

ZHONG, D.; YANG, Y.; DU, X. Palmprint Recognition Using Siamese Network. *In*: ZHOU, J. *et al*. (Eds.). *Biometric Recognition*. Cham: Springer International Publishing, 2018, pp. 48-55. ISBN 978-3-319-97909-0.

ZHU, H. *et al*. Plant identification based on very deep convolutional neural networks. *Multimedia Tools and Applications*, Feb 2018. ISSN 1573-7721.

ZHU, X.; BAIN, M. B-CNN: branch convolutional neural network for hierarchical classification. *CoRR*, abs/1709.09890, 2017.

ZHU, Y. *et al.* TA-CNN: two-way attention models in deep convolutional neural network for plant recognition. *Neurocomputing*, Elsevier B.V., 2019. ISSN 0925-2312.

ZWEIG, A.; WEINSHALL, D. Exploiting object hierarchy: combining models from different category levels. *In*: *2007 IEEE 11th International Conference on Computer Vision*. [S.l.: s.n.], 2007, pp. 1-8. ISSN 1550-5499.