

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação



Dissertação

**Classificação de Distúrbios em Folhas de Macieiras Utilizando Redes Neurais
Convolucionais**

Lucas Garcia Nachtigall

Pelotas, 2016

Lucas Garcia Nachtigall

**Classificação de Distúrbios em Folhas de Macieiras Utilizando Redes Neurais
Convolucionais**

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação

Orientador: Prof. Dr. Ricardo Matsumura Araujo
Coorientador: Dr. Gilmar Ribeiro Nachtigall

Pelotas, 2016

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

N124c Nachtigall, Lucas Garcia

Classificação de distúrbios em folhas de macieiras utilizando redes neurais convolucionais / Lucas Garcia Nachtigall ; Ricardo Matsumura Araujo, orientador ; Gilmar Ribeiro Nachtigall, coorientador. — Pelotas, 2016.

84 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2016.

1. Aprendizado de máquina. 2. Inteligência artificial. 3. Redes neurais convolucionais. 4. Classificação de sintomas em plantas. I. Araujo, Ricardo Matsumura, orient. II. Nachtigall, Gilmar Ribeiro, coorient. III. Título.

CDD : 005



UNIVERSIDADE FEDERAL DE PELOTAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE PÓS-GRADUAÇÃO



5

DEFESA DE DISSERTAÇÃO

NOME DO ESTUDANTE	MATRICULA
LUCAS GARCIA NACHTIGALL	14101508
CURSO OU PROGRAMA	NÍVEL
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO	MESTRADO

MEMBROS DA BANCA EXAMINADORA	TÍTULO	ASSINATURA
Ricardo Matsumura Araujo (Orientador)	DOUTOR	
Cristian Cechinel (PPGC-UFPEL)	DOUTOR	
Paulo R. Ferreira Jr. (PPGC-UFPEL)	DOUTOR	
Sandro Camargo (Unipampa)	DOUTOR	

APRECIÇÃO SOBRE A DISSERTAÇÃO

NÃO SIGILOSA

Em 9 de Junho de 2016, os membros acima nomeados para a defesa da Dissertação do(a) estudante LUCAS GARCIA NACHTIGALL, matriculado no Programa de Pós-Graduação em Computação, consideraram a Dissertação aprovada, estabelecendo o título definitivo como sendo "Classificação de Distúrbios em Folhas de Macieiras Utilizando Redes Neurais Convolucionais" e estabelecem um prazo máximo de 30 dias para as correções e entrega da versão definitiva.

DADOS PESSOAIS DOS MEMBROS DA BANCA EXAMINADORA

NOME COMPLETO	CPF	ANO NASCIMENTO	TITULAÇÃO		
			Área	Local	Ano
Ricardo Matsumura Araujo (Orientador)	9488568028	1979	CCO-P	UFAC	2010
Cristian Cechinel (PPGC-UFPEL)	454628449-72	1977	ENG. INFORMÁTICA	UPH	2012
Paulo R. Ferreira Jr. (PPGC-UFPEL)	61699092034	1976	Computação	UFAC	2008
Sandro Camargo (Unipampa)	772242210-15	1975	CIENCIA COMP	UFRGS	2010

1ª Via – Coordenador do Curso; 2ª Via – Orientador; 3ª Via – PRPPG.
DISTRIBUIÇÃO A CARGO DA COORDENAÇÃO DO PROGRAMA.

Dedico esta dissertação à minha família e amigos.

AGRADECIMENTOS

Ao Professor Dr. Ricardo Matsumura Araujo, pelo apoio, confiança, orientação e, principalmente, amizade dispensados não só na realização deste trabalho, mas durante o curso.

À Universidade Federal de Pelotas, em especial ao Programa de Pós-Graduação em Computação, pela oportunidade concedida para a realização do curso de Mestrado.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), pelo apoio financeiro.

À Embrapa Uva e Vinho, em especial a Estação Experimental de Fruticultura de Clima Temperado, pelo apoio na coleta e rotulagem das imagens dos distúrbios avaliados neste trabalho.

Ao Grupo de Especialista das diferentes instituições de pesquisa na cultura da macieira, pela contribuição na identificação de distúrbios através dos formulários online.

Aos Professores do Programa de Pós-Graduação em Computação, pelos ensinamentos, colaborações e amizade.

Ao meu pai e Coorientador Gilmar Ribeiro Nachtigall, minha mãe Heliane Garcia Nachtigall, minha madrinha Tania Elisa Morales Garcia, ao meu irmão Matheus Garcia Nachtigall e a todos os familiares por me apoiarem e me guiarem durante toda a minha vida.

À minha namorada Amanda Carriconde Duquia, sempre dando apoio, incentivo e carinho, precisando ter quantidades extraordinárias e quase inacreditáveis de paciência e compreensão.

Aos colegas que me acompanham desde a graduação, Roger Machado, Renato Souza de Souza e Ricardo Almeida, pela companhia, apoio e ajuda nas disciplinas cursadas e estudos.

RESUMO

NACHTIGALL, Lucas Garcia. **Classificação de Distúrbios em Folhas de Macieiras Utilizando Redes Neurais Convolucionais**. 2016. 83 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

Na cultura de macieira, nos últimos anos, têm sido verificadas perdas de até 25% da produção, as quais estão relacionadas à diversos fatores, entre eles as mudanças climáticas, a falta de tecnologias inovadoras e competitivas para o setor da maçã, além do surgimento de novas pragas e/ou doenças e agravamento das perdas por aquelas já conhecidas. Boa parte das soluções tecnológicas para resolver/evitar estes problemas passa pelo rápido reconhecimento em campo das causas destas perdas, principalmente quando relacionadas a incidência de doenças, pragas e desequilíbrios nutricionais que atacam as plantas. Contudo, o método atualmente utilizado para identificar estes sintomas em macieiras baseia-se no conhecimento de especialistas em detectá-los nos pomares, utilizando como ferramenta publicações contendo imagens e descrições de possíveis sintomas a serem encontrados. Sendo assim, verifica-se a necessidade de implementar uma nova abordagem para tratar esse problema de uma forma menos onerosa e mais eficiente. Técnicas de aprendizado de máquina se tornaram populares pelo fato de poderem extrair informações e padrões a partir de dados e então prever novos resultados ou classificar novos dados apresentados. Algumas das técnicas que atualmente são o estado da arte na classificação de imagens são as Redes Neurais Convolucionais, junto com as Redes Neurais Artificiais e Máquinas de Vetores de Suporte. Por ser uma área pouco explorada na literatura, foi necessária a criação de um novo banco de imagens de modo a avaliar se estas técnicas de aprendizado de máquina mencionadas podem classificar corretamente os sintomas em folhas. Para isso, foi realizada a coleta de folhas de macieiras que apresentavam cinco diferentes sintomas. Deste modo, esta dissertação propõe a utilização e comparação destas técnicas de aprendizado de máquina, para a tarefa de classificação de dois sintomas de doenças, dois sintomas de deficiências nutricionais e um sintoma de danos por herbicida em folhas de macieiras. Após a criação de partições de imagens para treino, validação e teste, estes foram submetidos às técnicas de Redes Neurais Convolucionais, redes *Multilayer Perceptron* e Máquinas de Vetores de Suporte. Além disso, as partições de testes foram enviadas para especialistas da área agrônômica para a identificação dos sintomas, de modo a comparar a acurácia dos resultados obtidos. Os resultados mostraram que a abordagem proposta neste trabalho, utilizando Redes Neurais Convolucionais, obteve acurácias entre 93,3% e 97,3% de acerto, se

igualando ou obtendo resultados superiores aos obtidos por outras técnicas de aprendizagem de máquinas, bem como pelos especialistas selecionados. Desta forma, o uso de Redes Neurais Convolucionais poderá viabilizar o diagnóstico em folhas de macieiras de forma rápida, precisa e usual.

Palavras-chave: Aprendizado de Máquina, Inteligência Artificial, Redes Neurais Convolucionais, Classificação de Sintomas em Plantas.

ABSTRACT

NACHTIGALL, Lucas Garcia. **Disorders Classifications on Apple Tree Leaves with Convolutional Neural Networks**. 2016. 83 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

Over the past few years in Apple culture, losses of up to 25% of production have been verified, these losses are related to several factors, such as climate change, the lack of innovative and competitive technologies for the apple sector, in addition to the appearance of new pests and/or diseases and aggravation of losses for reasons already identified. Many of the technological solutions to solve or prevent these problems is the fast recognition in the field of the causes of these losses, especially when there is the incidence of diseases, pests and nutritional imbalances. Nevertheless, the existing method for identifying symptoms in apple rely on using the knowledge of experts to detect symptoms in orchards, using as aid publications containing pictures and descriptions of possible symptoms to be found. Thus, it shows a need for a new approach to addressing this problem on a less costly and more efficient manner. Machine learning techniques have become popular because they can learn information and patterns from data and then predict new results or classify new data submitted. Some of the techniques that are currently state of the art in image classification are the Convolutional Neural Networks, along with artificial neural networks and Support Vector Machines. As an poorly explored area in the literature, it showed the need to create a new image data set. In a way to analyse if the mentioned machine learning techniques can correctly classify the symptoms on leaves. Thus, this thesis proposes the use and comparison of these machine learning techniques to the classification task of symptoms of 2 types of diseases, two symptoms of nutritional deficiencies and one symptom of damage by herbicide in apple leaves. After the establishment of training, validation and test sets, these were used to train and test the techniques of Convolutional Neural Networks, Multilayer Perceptron Networks and Support Vector Machines. In addition, for better comparison of the results obtained, the test sets were sent to agronomy experts for identification of symptoms, in order to compare the accuracy on the results obtained. The results showed that the approach proposed in this paper, using Convolutional Neural Networks, obtained accuracies between 93.3 % and 97.3 %, matching or obtaining superior results to those obtained by other machine learning techniques, as well as by selected experts. Thus, the use of Convolutional Neural Networks may facilitate diagnosis in leaves of apple in quickly, accurately and usual manner.

Keywords: Machine Learning, Artificial Intelligence, Convolutional Neural Networks, Symptoms Classification on Plants.

LISTA DE FIGURAS

1	Neurônio Artificial. Fonte: Adaptado de MARTINS-FILHO; MOL; ROCHA (2005)	23
2	Função Sigmoide. Fonte: Adaptado de WIKIPEDIA (2016)	24
3	Função Tangente Hiperbólica. Fonte: Adaptado de ROPER (2014).	24
4	Função ReLU. Fonte: Adaptado de LI; KARPATY (2015)	25
5	Organização dos campos receptivos em uma CNN. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).	28
6	Compartilhamento dos parâmetros para criação de um mapa de características. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).	29
7	Rede neural convolucional com dois estágios. Fonte: Adaptado de LECUN et al. (2010).	30
8	Exemplo de convolução. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).	30
9	Exemplo de redução utilizando filtro $MAX\ 2 \times 2$ e deslocando 2×2 (sem sobreposição). Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).	31
10	Exemplo de filtros aprendidos na base MNIST (a) sem <i>Dropout</i> (b) utilizando <i>Dropout</i> . Fonte: Adaptado de HINTON et al. (2012).	32
11	Possível hiperplano de separação e hiperplano ótimo (em verde). Fonte: Adaptado de OLIVEIRA-JUNIOR (2010).	34
12	Transformação: problema não linearmente separável em um problema linearmente separável. Fonte: Adaptado de REBELO (2008).	34
13	Separação de classes através da aplicação da máquina RBF. Fonte: Adaptado de OLIVEIRA-JUNIOR (2010).	35
14	Folhas coletadas com sintomas de deficiência de Potássio. Fonte: Autor.	44
15	Folhas coletadas com sintomas de deficiência de Magnésio. Fonte: Autor.	44
16	Folhas coletadas com sintomas de Sarna. Fonte: Autor.	44
17	Folhas coletadas com sintomas de <i>Glomerella</i> . Fonte: Autor.	45
18	Folhas coletadas com sintomas de Danos por Herbicida. Fonte: Autor.	45
19	Resultados analíticos para as amostras de folhas de macieira normais e folhas com deficiência de potássio e de magnésio, este, em dois níveis de severidade (média e alta) (NACHTIGALL, 2015).	47
20	Identificação dos fungos <i>Glomerella cingulata</i> e <i>Venturia inaequalis</i> em microscópio (ALVES, 2015).	47

21	Esquema da metodologia de coleta e criação do banco de imagens. Fonte: Autor.	50
22	Arquitetura da rede CNN de Alex Krizhevsky. Adaptada de (KARNOWSKI, 2015)	52
23	Exemplo de questão do primeiro questionário enviado aos especialistas. Fonte: Autor.	54
24	Diagrama do sistema de votação desenvolvido. Fonte: Autor.	55
25	Diagrama do funcionamento do sistema de votação desenvolvido. Fonte: Autor.	56
26	Distribuição das acurácias obtidas para cada um dos cinco sintomas estudados em função do uso de diferentes números de imagens para treinamento visando a classificação pela técnica CNN no Grupo 2 de imagens. Fonte: Autor.	61
27	Relação entre as acurácias médias dos cinco sintomas em folhas de macieira e o número de imagens na partição de treinamento utilizada para a técnica CNN no Grupo 2 de imagens. Fonte: Autor.	62
28	Estimativa por intervalo de confiança ($IC_{1-\alpha}(p)$) com aproximação normal para cada proporção de acerto nas técnicas testadas, a 99% de confiança para o Grupo 1 de imagens. Fonte: Autor.	65
29	Estimativa por intervalo de confiança ($IC_{1-\alpha}(p)$) com aproximação normal para cada proporção de acerto nas técnicas testadas, a 99% de confiança para o Grupo 2 de imagens. Fonte: Autor.	67
30	Folha com deficiência de Magnésio, classificada pela técnica de CNN e por especialistas como sendo sintoma de sarna. Fonte: Autor.	69

LISTA DE TABELAS

1	Concentração de macro e micronutrientes em folhas de macieira normais e com três níveis de severidade de sintomas causados por toxidez do herbicida glifosato (NACHTIGALL, 2015).	47
2	Detalhes das Camadas da rede utilizada. Fonte: Autor.	51
3	Exemplo de Matriz de Confusão	57
4	Valores de <i>Kappa com sua Força de Concordância</i> . Fonte: Adaptado de LANDIS; KOCH (1977)	58
5	Número de folhas de macieira com sintomas que foram coletadas e rotuladas. Fonte: Autor.	59
6	Acurácias obtidas pelas classificação dos especialistas nas partições de testes dos Grupos 1 de imagens. Fonte: Autor.	63
7	Percentuais de acurácia na classificação de imagens de sintomas obtidos pelas abordagens propostas conforme o Grupo 1 de imagens. Fonte: Autor.	63
8	Matriz de confusão da CNN utilizando o Grupo 1 de imagens. Fonte: Autor.	64
9	Acurácias obtidas pelas classificação dos especialistas nas partições de testes dos Grupos 2 de imagens. Fonte: Autor.	65
10	Percentuais de acurácia na classificação de imagens de sintomas obtidos pelas abordagens propostas conforme o Grupo 2 de imagens. Fonte: Autor.	66
11	Matrizes de confusão para os resultados da classificação por CNN e resultados da classificação por especialistas com o sistema de votação, utilizando o Grupo 2 de imagens. Fonte: Autor.	68

LISTA DE ABREVIATURAS E SIGLAS

CNN	Convolutional Neural Networks
GLCM	Gray Level Co-occurrence Matrix
MLP	Multilayer Perceptron
MSE	Mean Square Error
ReLU	Rectified Linear Unit Layer
RNA	Redes Neurais Artificiais
RNP	Redes Neurais Profundas
SGD	Stochastic Gradient Descent
SGDM	Spatial gray-level dependence matrices
SVM	Support Vector Machine

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Objetivos	19
1.2	Estrutura da Dissertação	19
2	APRENDIZAGEM DE MÁQUINA	21
2.1	Redes Neurais Artificiais	21
2.1.1	Neurônio	22
2.1.2	Redes <i>Multilayer Perceptron</i>	24
2.2	Rede Neural Convolucional	26
2.3	Máquinas de Vetores de Suporte	32
2.3.1	Funções de Kernel	33
2.3.2	<i>Kernel</i> Gaussiano (RBF)	35
2.4	Considerações sobre o Capítulo	36
3	APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DE IMAGENS	37
3.1	Redes <i>Multilayer Perceptron</i>	37
3.2	Máquinas de Vetores de Suporte	38
3.3	Redes Neurais Convolucionais	38
3.4	Trabalhos Relacionados	40
3.5	Considerações sobre o Capítulo	42
4	METODOLOGIA	43
4.1	Banco de Imagens	43
4.1.1	Sintomas em Folhas	43
4.1.2	Veracidade dos Sintomas	46
4.1.3	Grupos de Imagens	48
4.2	Algoritmos	49
4.2.1	Redes Neurais Convolucionais	49
4.2.2	Redes <i>MultiLayer Perceptron</i>	52
4.2.3	Máquina de Suporte de Vetores	53
4.3	Classificação por Especialistas	53
4.4	Avaliação dos Resultados	57
5	RESULTADOS E DISCUSSÃO	59
5.1	Banco de Imagens	59
5.2	Desempenho por Número de Amostras por classe	61
5.3	Classificação de Imagens	62
5.3.1	Grupo 1	62

5.3.2	Grupo 2	64
5.4	Síntese	69
6	CONSIDERAÇÕES FINAIS	70
	REFERÊNCIAS	73
	APÊNDICE A IMAGENS UTILIZADAS PARA CRIAÇÃO DE CADA GRUPO .	81
A.1	Grupo 1	81
A.2	Grupo 2	81
	APÊNDICE B ESPECIFICAÇÃO CONFIGURAÇÕES UTILIZADAS PARA TREINO	82
B.1	Redes Neurais Convolucionais	82
B.2	Redes <i>Multilayer Perceptron</i>	82

1 INTRODUÇÃO

Estima-se que 80 a 90% dos danos causados por doenças e pragas que atacam a cultura da macieira ocorrem em folhas e frutos (VALDEBENITO-SANHUEZA et al., 2002; KOVALESKI, 2004). Entre as doenças, a sarna da macieira e a mancha de *Glomerella* são as mais importantes, enquanto que as podridões nos frutos são responsáveis por grande parte das perdas de maçãs durante o período de pós-colheita. No caso das pragas, cujas folhas e frutos servem como fonte alimentar ou hospedeiros, têm-se como principais problemas o ataque de mosca-das-frutas, grafolita e grandes lagartas. Já os distúrbios causados por excesso ou falta de nutrientes na macieira são visíveis principalmente nas folhas, durante a fase de desenvolvimento vegetativo, com exceção da deficiência de cálcio, cujos sintomas se manifestam nos frutos durante o armazenamento.

O rápido diagnóstico de sintomas causados por distúrbios nutricionais e fisiológicos, ataque de pragas e doenças em pomares de macieira é essencial para evitar prejuízos maiores com a disseminação da condição para áreas não afetadas, uma vez que aproximadamente 25% da produção de maçãs é perdida em função dessas causas (MEDEIROS, 2015; VALDEBENITO-SANHUEZA et al., 2006; KOVALESKI, 2004; NACHTIGALL; CARRARO; ALLEONI, 2007). Além disso, o diagnóstico correto é um processo indispensável para definir as estratégias de manejo/controle e, consequentemente, para o uso racional de fertilizantes e agrotóxicos. Nas condições de campo, normalmente são verificados dois problemas principais que podem impedir o diagnóstico de forma rápida: a) é inviável o monitoramento permanente de grande número de plantas por pessoas capazes de detectar os sintomas; b) conhecimentos insuficientes para identificar o problema corretamente da pessoa que detecta os sintomas.

Apesar de haver algumas soluções para estabelecer um diagnóstico dos principais sintomas de anomalias na cultura da macieira, como é o caso de publicações com fotos e explicações relativas à condição de ocorrência dos principais distúrbios nutricionais, pragas e doenças (VALDEBENITO-SANHUEZA et al., 2008), são necessárias tecnologias facilitadoras para um rápido diagnóstico. Os sistemas especialistas, como é o caso desenvolvido para a cultura da videira (FIALHO et al., 2012), são sistemas

que normalmente envolvem uma base de dados de referência que pode ser consultada pelo usuário, contudo apresentam grau de automação bastante lento e sua acurácia é relativamente baixa, especialmente quando se considera a importância do problema.

Deve-se considerar que os sistemas especialistas são ferramentas baseadas em um conjunto de regras que tentam representar, da maneira mais acurada possível, todo o universo de possibilidades do problema que se deseja resolver. No caso específico de um sistema desenvolvido para identificar distúrbios nutricionais, ataque de pragas e de doenças em plantas, como por exemplo, para uma determinada cultura agrícola, inicia-se com todos os distúrbios nutricionais e todas as pragas e doenças para as quais o sistema foi treinado e, através de uma série de perguntas, as possibilidades vão sendo sucessivamente refinadas, até que se chegue à resposta desejada.

Os sistemas especialistas, usados para diagnóstico de distúrbios nutricionais, de pragas e de doenças em plantas, normalmente necessitam considerar uma ampla gama de problemas, fazendo com que seja necessária a criação de um extenso conjunto de regras, as quais devem se relacionar de maneira coerente a fim de resultar em um bom diagnóstico. Um extenso conjunto de regras significa que o processo para se chegar a um diagnóstico acurado pode demandar que um grande número de perguntas seja respondido pelo usuário. Este fato aumenta consideravelmente a probabilidade de erros, uma vez que uma única pergunta respondida de maneira errônea pode fazer com que o sistema se desvie irremediavelmente da resposta correta.

No Brasil, o diagnóstico de sintomas em plantas utilizando Inteligência Artificial ainda não é habitualmente utilizado por produtores e técnicos no setor da agricultura. No caso específico da cultura da macieira, além de permitir a obtenção do diagnóstico de forma rápida, precisa e usual através da classificação digital de imagens, possibilitará construir uma interface para dispositivos móveis e web que servirá como ponte entre o usuário final e o núcleo da ferramenta de diagnóstico estabelecido. Além disso, teoricamente uma vez referendada a sua eficiência, o seu uso pode ser adaptado e estendido para outras culturas e sistemas que utilizem imagem como objeto de identificação. Deste modo, o uso de técnicas de aprendizado de máquinas nas tarefas de identificação e classificação de imagens, através de redes neurais, pode ter grande potencial para o diagnóstico de sintomas em plantas.

Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso central de animais, normalmente o cérebro humano, tendo a capacidade de realizar tarefas como reconhecimento e associação de padrões, aproximação de funções e filtragem. Estas redes são formadas por neurônios interconectados que computam valores de entrada.

O estudo de RNAs iniciou-se com MCCULLOCH; PITTS (1943) introduzindo a ideia de redes neurais como máquinas computacionais após realizarem uma analogia das células nervosas e o processo eletrônico; e com HEBB (1949) estudando o compor-

tamento de animais, que propôs um princípio de aprendizado em sistemas nervosos complexos, assim, postulando a primeira regra de aprendizagem auto-organizada. Os avanços continuaram com ROSENBLATT (1958) com a criação de uma rede com múltiplos neurônios para a resolução de problemas linearmente separáveis e a proposta do perceptron como o primeiro modelo de aprendizagem supervisionada.

Por dificuldades encontradas no tempo computacional para executar o treinamento das redes neurais, e a falta algoritmos de treinamento suficientemente rápidos para treinar redes profundas e amplas, diminui-se o foco nas pesquisas em Redes Neurais Profundas. Entretanto, em meados dos anos 2000, HINTON; OSINDERO; TEH (2006) propuseram uma nova técnica de treinamento, sendo capaz de realizar o treinamento de forma ágil, a qual aliada ao avanço tecnológico na capacidade de processamento, permitiram que as Redes Neurais Profundas voltassem a ser uma área promissora a ser estudada.

Apesar de não ser possível afirmar que arquiteturas profundas são superiores às rasas em todos os problemas (SALAKHUTDINOV; MURRAY, 2008), existem evidências de que quando a tarefa é complexa o suficiente e existe informação suficiente para capturar esta complexidade, estas redes profundas podem ter um melhor resultado (LAROCHELLE et al., 2007). Além disso, os modelos de arquitetura profunda podem aprender com menor envolvimento humano na construção do modelo antes do treinamento, menos padrões de exemplo e menor custo computacional, pois integram as etapas de extração de características e classificação em um mesmo modelo.

Atualmente é possível ver grandes organizações utilizando as Redes Neurais Profundas para a resolução de seus problemas, como por exemplo: o Facebook realizou contrato com Yann LeCun para coordenar um novo laboratório de inteligência artificial, com o intuito de desenvolver tarefas de aprendizado para possibilitar a marcação de pessoas em fotos automaticamente; a Google contratou Geoffrey Hinton e dois de seus estudantes para aprimorar as técnicas de aprendizado profundo e ajudar a lidar com o crescente número de dados que a Google possui, o Baidu junto com Andrew Ng na coordenação do novo laboratório no vale do silício; e outras organizações como, Yahoo que já mostraram interesse na área.

Sendo assim, esta dissertação propõe a utilização da técnica de Redes Neurais Convolucionais (*Convolutional Neural Network - CNN*) para a classificação de sintomas em folhas de macieiras, de modo a apresentar alternativas para tornar o diagnóstico de sintomas rápido, preciso e usual, bem como apresenta uma revisão bibliográfica em relação as RNAs e suas sucessoras, detalhando sua história, conceitos, modelos e aplicações, incluindo trabalhos relacionados as técnicas de aprendizado de máquina para a classificação de sintomas em plantas.

1.1 Objetivos

Como objetivo, esta dissertação buscou analisar a viabilidade da técnica de redes neurais convolucionais para a classificação de sintomas em folhas de macieira, causados por distúrbios nutricionais, danos causados por doenças e por toxidez de herbicida, visando reduzir as perdas decorrentes destes eventos em pomares de macieira. Além disso, apresenta uma revisão bibliográfica com os termos relevantes para este trabalho, de modo que esta possa ser utilizada como referência para futuros projetos.

Objetivos secundários:

- Criação de um banco de imagens contendo amostras dos sintomas de distúrbios em folhas de macieira.
- Comparação do desempenho das CNNs com o de outras técnicas de aprendizado de máquina e com especialistas na área de agronomia.
- Avaliação da influência do número de imagens nas partições de treinamento nos resultados finais de predição.

1.2 Estrutura da Dissertação

Este trabalho divide-se em sete Capítulos, incluindo este Capítulo inicial de Introdução onde é apresentado o problema de pesquisa, caracterizado pela ausência de ferramentas de reconhecimento automático de distúrbios em folhas de macieira para o diagnóstico fácil e precoce, na literatura analisada.

No Capítulo 2 são apresentados os conceitos e base histórica sobre Redes Neurais Artificiais, especificando sua formação, arquiteturas e processo de aprendizagem; a utilização e arquiteturas das Redes Neurais Convolucionais e os conceitos de Máquinas de Vetores de Suporte. Já no Capítulo 3 são apresentadas as técnicas utilizadas para a tarefa de reconhecimento e classificação de imagens, bem como trabalhos relacionados que as utilizam para a classificação de sintomas em plantas.

No Capítulo 4 são apresentados os protocolos metodológicos para a criação do banco de imagens, envolvendo sintomas causados por doenças, por deficiências nutricionais e por danos de herbicida em folhas de macieiras, bem como a descrição das abordagens escolhidas para a classificação das imagens coletadas, detalhando seus funcionamentos e arquiteturas, bem como as formas de avaliação destas técnicas de classificação.

No Capítulo 5 são apresentados e discutidos os resultados obtidos com a coleta e rotulagem das folhas de macieira com os sintomas que foram selecionados para a formação do banco de imagens, os resultados obtidos com a classificação das imagens, mencionadas na metodologia adotada, e das comparações entre as técnicas de aprendizado de máquina utilizadas e os resultados obtidos por especialistas.

Nos Capítulos 6 e 7 são apresentadas as conclusões em função dos resultados obtidos e as sugestões e considerações sobre possíveis melhorias e trabalhos futuros, respectivamente.

2 APRENDIZAGEM DE MÁQUINA

A evolução nos estudos de reconhecimento de padrões e teorias sobre inteligência artificial, inclui o aprendizado de máquina, que foi definido por SAMUEL (1959) como a área de estudo que provê a capacidade à computadores de aprender sem serem programados explicitamente (SIMON, 2013). Esta área de pesquisa explora a utilização de algoritmos e técnicas para aprender informações e padrões sobre dados existentes, e assim fazer previsões futuras ou reconhecimento de novos dados introduzidos.

Suas aplicações podem ser observadas nas áreas de filtragem de *spam* em mensagens, reconhecimento de caracteres, mecanismos de busca, mineração de dados e reconhecimento de imagens, vídeos e sons.

Sendo assim, este Capítulo apresenta uma breve motivação do uso de técnicas de aprendizagem de máquina, junto com algumas de suas técnicas e conceitos.

2.1 Redes Neurais Artificiais

A motivação ao estudo de Redes Neurais Artificiais (RNAs) advém da maneira que o cérebro humano processa informações, reconhecendo informações diferentemente de computadores convencionais.

De acordo com HAYKIN (1999), uma RNA pode ser definida como:

Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar o conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:

- *O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem;*
- *Forças de conexão entre neurônios, conhecidos como pesos sinápticos, são utilizados para armazenar o conhecimento adquirido.*

BARONE (1999) cita algumas das principais características das RNA's: sua capacidade de aprender através de exemplos e generalizar este aprendizado de modo a reconhecer novos exemplos similares; possuem um bom desempenho em tarefas pouco ou mal definidas, ou com pouca informação explícita de como as resolverem; robustez a ruídos, ou seja, dados e informações erradas ou falsas nos exemplos.

As definições encontradas neste Capítulo, são baseadas principalmente na obra de HAYKIN (1999), PAULA (2000), TAFNER; XEREZ; RODRIGUES FILHO (1995) e BISHOP et al. (1995).

2.1.1 Neurônio

Assim como o neurônio biológico, o neurônio artificial possui um ou mais sinais de entrada e apenas um sinal de saída. As informações podem ser recebidas através de sensores ou de outros neurônios artificiais que fazem parte da RNA. Estes sinais são processados e enviados para a saída. Os sinais de entrada (estímulos) devem chegar até o neurônio simultaneamente, isto é, todas as informações devem chegar ao núcleo do neurônio artificial ao mesmo tempo (TAFNER, 1996).

Um dos atributos de grande importância do neurônio artificial é o peso. Os pesos, também conhecidos por pesos sinápticos, são representados pela letra w (*weight*) e representam o grau de importância que determinada entrada possui em relação àquele determinado neurônio.

O valor do peso é alterado em função da intensidade do sinal de entrada, e dessa forma, o peso muda o seu valor representativo para a rede (processo de aprendizagem). Deduz-se que, quanto mais estimulada for uma entrada, mais estimulado será o peso correspondente, e quanto mais for estimulado um peso, mais significativo e influente o mesmo será para o resultado do sinal de saída do respectivo neurônio.

Matematicamente, os pesos são vistos como um vetor de valores $[w_1, w_2, \dots, w_n]$ para um neurônio, ou uma matriz de pesos, coleção de vetores, para um conjunto de neurônios.

O sinal de excitação do neurônio é resultante do somatório do produto dos sinais de entrada, representados por um vetor $[x_1, x_2, \dots, x_n]$, pelo vetor de pesos do neurônio ($\sum_{i=0}^n X_i W_i$ menos o valor correspondente a $X_0 W_0$ será explicado adiante e corresponde ao viés, representando um estímulo inicial a rede). Após esta operação, os sinais de entrada passam a ser chamados de entradas ponderadas.

A próxima tarefa a ser tomada pelo neurônio, é a de verificar se o valor resultante da soma entre o produto dos sinais de entrada pelos respectivos pesos atingiu ou não um valor predeterminado, chamado de limiar. Se o valor calculado atingiu o limiar, o mesmo é repassado adiante através da saída. Caso contrário, se o valor não atingiu o limiar, o sinal não será transferido. Esse processo de verificação é chamado de função de transferência, que também é conhecido como limiar lógico.

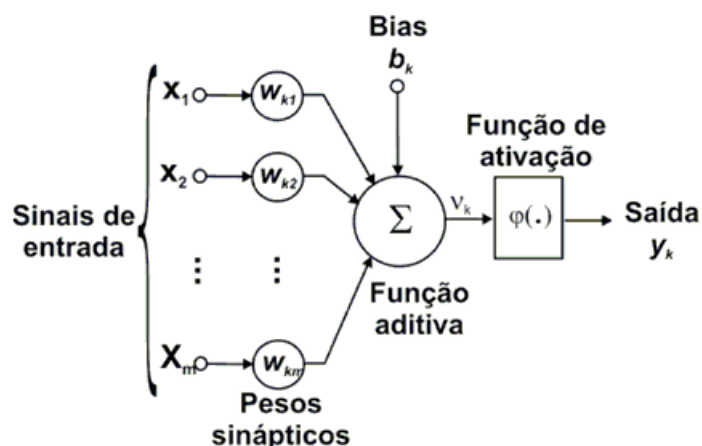


Figura 1: Neurônio Artificial. Fonte: Adaptado de MARTINS-FILHO; MOL; ROCHA (2005)

A resposta final da rede ou das camadas subjacentes está diretamente ligada com o resultado obtido pela função de transferência. Por isso, deve-se dar a devida atenção a este processo. A lógica neural pressupõe, que a intensidade dos sinais de entrada, dispara, ou não, o sinal do neurônio, fazendo com que este estimule o neurônio seguinte (TAFNER; XEREZ; RODRIGUES FILHO, 1995).

Além da função de transferência, há a função de ativação ou não-linear, a qual antecede a mesma e tem como função, suceder um nível de ativação dentro do próprio neurônio, ou seja, o neurônio, através desta função, decidirá o que fazer com o resultado da soma ponderada das entradas (ativar ou não). Essa decisão tem efeito somente ao respectivo neurônio artificial.

Função de Ativação Uma função de ativação representada por $\varphi(v)$ é o que define a saída de um neurônio, podendo ser dividida em três tipos:

- 1: Função Sigmoide: A função sigmoide é a forma mais comum de função de ativação utilizada na construção de RNAs. Ela é descrita como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não-linear. Enquanto a função linear assume o valor de 0 ou 1, uma função sigmoide assume um intervalo contínuo de valores entre 0 e 1. Apresentada na Figura 2.
- 2: Função Tangente Hiperbólica (Tanh): A função tangente hiperbólica é utilizada para comprimir um valor real para uma variação entre -1 e 1 e pode ser representada pela Equação 1 e apresentado na Figura 3.

$$\tanh(x) = 2\sigma(2x) - 1 \quad (1)$$

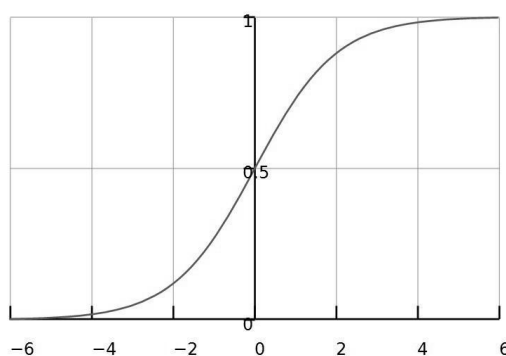


Figura 2: Função Sigmoidal. Fonte: Adaptado de WIKIPEDIA (2016)

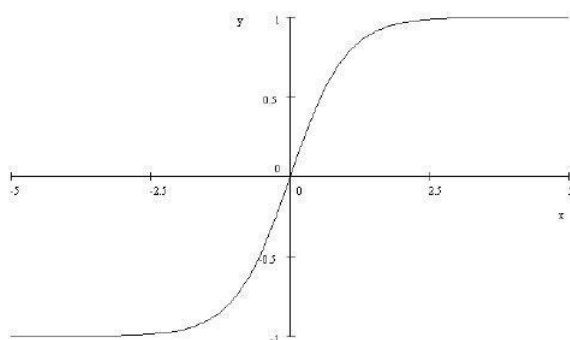


Figura 3: Função Tangente Hiperbólica. Fonte: Adaptado de ROPER (2014).

- 3: Função de Unidade Linear Retificada (*ReLU - Rectified Linear Unit*): A função ReLU é descrita com a Equação 2 e apresentado na Figura 4, ou seja, a ativação tem seu limiar em zero. A vantagem desta função de ativação é um aumento de até 6 vezes na aceleração da convergência do gradiente descendente estocástico em comparação com as funções sigmoide e Tanh (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Além disso, quando comparada com as funções sigmoide e Tanh que utilizam operações custosas para seu cálculo, a função ReLU pode ser facilmente implementada utilizando uma matriz com limiares de ativação em zero.

$$f(x) = \max(0, x) \quad (2)$$

2.1.2 Redes *Multilayer Perceptron*

Os problemas não linearmente separáveis podem ser resolvidos através das redes com uma ou mais camadas intermediárias, chamadas MLP (*Multilayer perceptron*). A alteração da arquitetura da rede, como a inserção de camadas ocultas e/ou o número de neurônios, a princípio, não parece ser problema, pois um dos principais agravantes passa a ser o algoritmo de treinamento para as redes multicamadas. Fator este que, devido à inexistência ou desconhecimento, causou uma atenuação nas pesquisas em

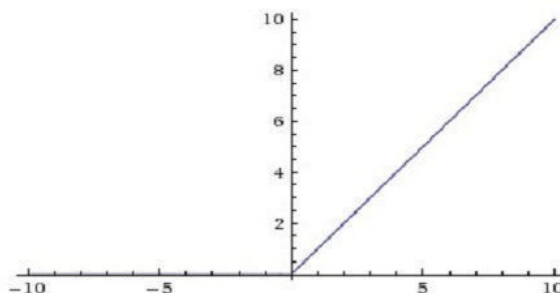


Figura 4: Função ReLU. Fonte: Adaptado de LI; KARPATY (2015)

redes neurais artificiais em meados da década de 70.

Uma das alternativas adotadas é dividir a rede em um conjunto de sub-redes, sendo uma sub-rede para cada camada, com um treinamento independente. Este método de subdivisão, muitas vezes, ou não é possível ou é muito complexo. Outra possibilidade seria realizar um treinamento completo, isto é, de uma só vez. O problema encontrado para este segundo método está em como realizar o treinamento dos nodos que pertencem à camada intermediária, visto que é extremamente complicado determinar que tipo de resposta desejada estes teriam, ou seja, como determinar o erro.

A aplicabilidade deste método está restrita a definição do erro nos nodos pertencentes às camadas intermediárias da rede. Se for utilizada uma função do tipo limiar, a avaliação do erro será complexa, visto que, os nodos das camadas intermediárias e de saída não terão como saber a margem de erro ou a diferença entre as respostas de seus nodos com relação às respostas desejadas. Uma das soluções para o problema apresentado seria a utilização de uma função de ativação não linear, a qual resolve o mesmo em parte, visto que a utilização deste tipo de função em redes multicamada resultaria na equivalência de uma rede de uma única camada (BRAGA; CARVALHO; LUDEMIR, 1998).

Adotou-se então, treinar as redes com mais de uma camada através de métodos baseados no gradiente descendente. Métodos baseados no gradiente descendente precisam ter a função de ativação contínua, diferenciável e não decrescente. A função adotada precisa informar os erros que a rede cometeu para as camadas anteriores, com uma boa precisão. Logo a função que mais se adapta a estas características é a função do tipo sigmoide (BRAGA; CARVALHO; LUDEMIR, 1998).

O processamento atribuído a cada neurônio pertencente à rede é resultante da combinação do processamento realizado pelos neurônios da camada anterior, que por sua vez estão atribuídos a este nodo da próxima camada. A medida com que cada camada intermediária da rede se aproxima da camada de saída há uma delimitação do espaço de decisão dos dados que está recebendo.

Para uma rede com duas camadas intermediárias, teríamos a primeira camada

oculta, delimitando o espaço de padrões de treinamento através das retas traçadas pelos neurônios. A segunda camada forma regiões convexas, onde o número de lados que compõe tal região é determinado pela quantidade de unidades conectadas a este neurônio, que por sua vez combina as retas que surgiram da camada anterior. Cada neurônio da camada de saída forma regiões, provenientes das combinações das regiões convexas (BRAGA; CARVALHO; LUDEMIR, 1998). Conclui-se que cada neurônio que compõe uma rede *Multilayer Perceptron* contribui para a detecção de características dos dados apresentados.

A determinação do número de camadas a ser utilizada influi de forma crucial no aprendizado da rede. O uso de um grande número de camadas intermediárias não é recomendado, visto que o erro ocorrido em uma camada é propagado a camadas anteriores da rede. A determinação do número de neurônios que pertence a camadas intermediárias é definida de forma empírica, e normalmente depende da distribuição dos padrões de treinamento e validação da rede.

Um uso excessivo de neurônios levará a rede a decorar a partição de treinamento, ao invés de extrair as características gerais (generalizar). Ao processo de memorização da partição de treinamento, dá-se o nome de sobre ajuste (*over fitting*). Um número razoavelmente pequeno de neurônios levará a rede a aumentar o tempo de treinamento, dificultando a determinação da representação ótima do problema proposto. Neste caso, alguns neurônios poderão ficar sobrecarregados, pois estes precisam lidar com um número elevado de restrições a serem analisadas (PAULA, 2000).

Pela forma como as redes RNAs recebem seus dados como entrada, surge uma nova limitação à tarefa de reconhecimento e classificação de imagens, onde, no caso de uma imagem, cada pixel deve ser tratado como um atributo para a rede. Em casos reais, onde imagens com resoluções de 12 mega pixels são facilmente encontradas, isso originaria entradas na rede com milhões de atributos, dificultando seu aprendizado e aumentando drasticamente o tempo de treinamento.

É possível analisar este problema no caso do banco de imagens CIFAR-10 (KRIZHEVSKY; HINTON, 2009), onde cada imagem possui apenas 32x32 pixels de resolução, que quando apresentados a uma RNA, possuirá 3.072 atributos (32 pixels de altura \times 32 pixels de largura \times 3 cores RGB). Apesar deste caso ainda parecer tratável por uma RNA, utilizando o hardware existente atualmente, quando escalando o problema para imagens coloridas de 256 x 256 pixels o número de entradas aumenta para 196.608, causando sobre ajuste, menor aprendizado e maior tempo de treino da rede.

2.2 Rede Neural Convolucional

FUKUSHIMA (1980) introduziu as CNN por que são baseadas na descoberta da

mecânica visual dos seres vivos, onde várias células são responsáveis por detectar luz em regiões chamadas campos receptivos, quanto mais complexas estas células, maiores são os campos receptivos. Após, é utilizado um operador convolutivo para que todas estas células realizem a mesma ação e enviem as informações necessárias ao cérebro.

Em 2011, o estado da arte em Redes Neurais *feedforward* Profundas, para reconhecimento de imagens, se alternou para redes com camadas convolucionais e de sub amostragem ou redução (*max-pooling*) cobertas por diversas camadas completamente conectadas para a classificação. Seu treinamento é normalmente realizado sem qualquer pré-treinamento não-supervisionado. Desde 2011, implementações desta técnica que utilizam GPUs para auxiliar seu processamento e treinamento, ganharam vários concursos de reconhecimento de padrões.

Na proposta da CNN, LECUN et al. (1989) consideraram o fato de que a projeção de uma arquitetura específica de uma rede neural, baseada numa informação a priori do problema a ser resolvido, é capaz de levar o modelo a uma melhor capacidade de generalização. Com os avanços, LECUN et al. (1998) se basearam no fato de que os extratores de características ideais para resolver problemas que lidam com a variabilidade e a riqueza de dados naturais podem ser incorporados na rede neural automaticamente durante o processo de aprendizagem. Dessa forma, procura-se evitar que a acurácia de um dado classificador seja determinada pela capacidade do projetista do modelo na escolha das características que melhor representam o problema apresentado.

As redes neurais convolucionais fazem uso de campos receptivos locais, que são aqueles conectados a regiões específicas, compartilhamento de pesos e subamostragem espacial em sua arquitetura. Através do uso dos campos receptivos e do compartilhamento de pesos, é possível a detecção de características visuais elementares, como bordas, fins de linha ou quinas, em qualquer localização da entrada. Assim, é possível reduzir a quantidade de parâmetros livres na rede.

A CNN possui uma arquitetura de rede neural *feedforward* onde cada camada provê a entrada para a camada posterior. As camadas que realizam a convolução e subamostragem ou redução, englobando o processo de extração de características, localizam-se no início da rede. Ao final da CNN encontra-se uma rede completamente conectada que é responsável pela classificação do padrão de entrada. As características são extraídas de uma camada para outra da rede, onde um neurônio conecta-se a um campo receptivo local na camada anterior. Dessa forma, as características são combinadas sequencialmente para, por conseguinte, serem extraídas características de mais alto nível.

Considerando que um conjunto de características tenha sido extraído, sua localização exata perde a importância, podendo até mesmo atrapalhar o reconhecimento. Ao

invés disso, o foco é colocado na localização de cada característica com relação às demais. Assim, camadas sucessivas de convolução e subamostragem se alternam, de modo que a quantidade de mapeamentos de características cresça em detrimento da resolução espacial.

Seguindo as definições de LI; KARPATY (2015) e JURASZEK; SILVA; SILVA (2014), as CNN são arquiteturas multiestágios capazes de serem treinadas. Os campos receptivos são altamente correlacionados à localidade do estímulo na imagem capturada. As CNN utilizam este conceito forçando um padrão de conectividade entre as camadas de neurônios artificiais.

A Figura 5 mostra esta organização onde uma camada m está conectada a uma pequena sub região da camada $m - 1$. Neste exemplo a camada $m - 1$ corresponde a imagem de entrada. A camada superior m possui um campo receptivo de tamanho 3, onde cada neurônio recebe o estímulo de 3 neurônios da camada anterior. A camada $m + 1$ é semelhante à camada anterior, possuindo um campo receptivo de tamanho 3 com relação à camada anterior porém, com um campo receptivo de tamanho 5 com relação à imagem de entrada.

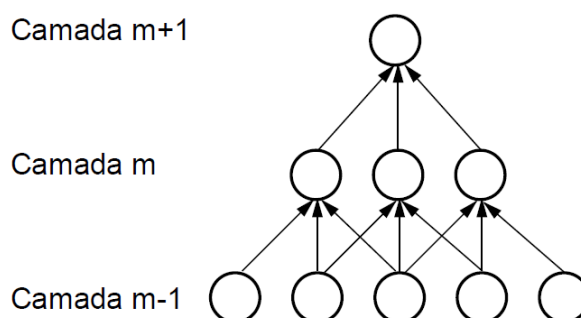


Figura 5: Organização dos campos receptivos em uma CNN. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).

Considerando essa analogia, cada campo receptivo é considerado um filtro não linear onde seus pesos devem ser aprendidos para que o neurônio seja ativado apenas quando um determinado estímulo estiver presente na área onde o filtro foi aplicado.

Cada filtro é aplicado à toda a imagem de entrada (ou camada anterior) de forma convolucional, o resultado da aplicação deste filtro é chamado de mapa de características (*feature map*). Cada mapa de características compartilha os mesmos parâmetros. A Figura 6 mostra o compartilhamento dos parâmetros. Esta estratégia garante que uma determinada característica será detectada pelo mapa de características independentemente da sua posição da imagem (ou mapa) de entrada.

As entradas de dados de cada estágio são um conjunto de mapas de características (*features maps*). Quando aplicada utilizando imagens coloridas, a entrada do primeiro estágio consiste nos três canais de cores da imagem. Cada vetor de duas

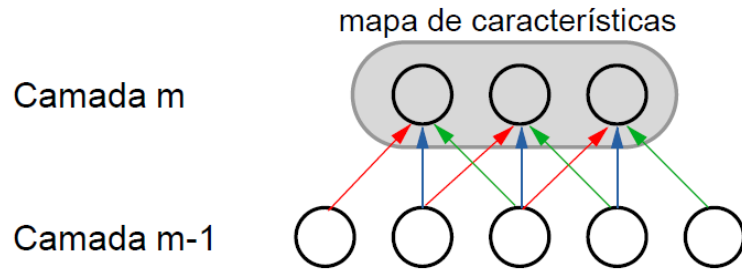


Figura 6: Compartilhamento dos parâmetros para criação de um mapa de características. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).

dimensões passa a funcionar como um mapa de características. Na saída de cada estágio, cada mapa corresponde a convolução do mapa de entrada por um filtro. A aplicação do filtro no mapa destaca algumas características. Cada filtro é responsável por destacar uma característica diferente. No primeiro estágio os filtros destacam linhas e gradientes em diferentes orientações.

Um mapa de característica é obtido efetuando a convolução de uma imagem de entrada por um filtro linear seguido da adição de um termo de *bias* e da aplicação de uma função não linear. Sendo a camada k , os filtros determinados por um conjunto de pesos W^k , um termo de *bias* b_k e o operador de convolução $*$, a Equação 3 (LECUN et al., 2010) mostra a obtenção do mapa de característica h^k para uma função não linear f .

$$h_{ij}^k = f((W^k * x)_{ij} + b_k) \quad (3)$$

Cada estágio é composto por três etapas, filtragem ou convolução (*filter bank layer* ou *convolutional layer*), etapa não linear (*non-linearity layer*) e etapa de redução ou subamostragem (*feature pooling layer*) que representa o campo receptivo. Uma CNN pode ser composta de um ou mais estágios onde cada um contém as três etapas. A Figura 7 mostra uma CNN com um único mapa de características de entrada (ex: uma imagem em tons de cinza) com dois estágios convolucionais C1+S1 e C2+S2. Detalhando as camadas presentes em uma arquitetura CNN temos:

Entrada (*input*): A camada de entrada contém os valores brutos dos píxeis da imagem, como apresentado anteriormente em uma imagem que possui valores de altura, largura e profundidade (número de canais de cores).

Camada de convolução (*Convolutional Layer*): A etapa de convolução ou filtragem realiza a convolução dos filtros W_i^k , correspondente ao i -ésimo filtro da camada k de tamanho $l_1 \times l_2$ na imagem. Cada filtro detecta uma característica particular em todas as localizações na imagem de entrada. Na primeira camada a imagem de entrada I_{mn} , sendo m a altura e n o comprimento, é representada através de um conjunto de mapas de características, um para cada canal de cor.

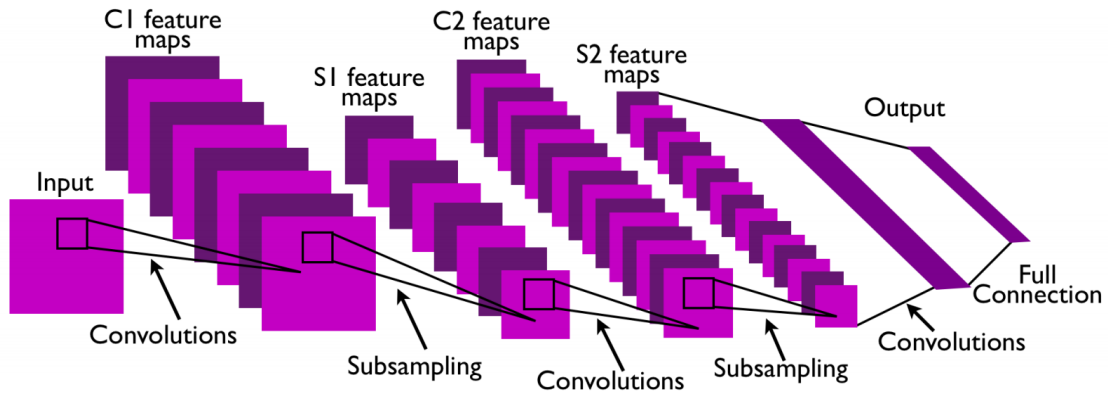


Figura 7: Rede neural convolucional com dois estágios. Fonte: Adaptado de LECUN et al. (2010).

Na Figura 8 é realizado um exemplo de convolução no qual a máscara é transladada em posições em que se encontra integralmente no domínio dos píxeis (e, portanto, reduzindo a imagem resultante). É possível observar que o filtro aplicado realça áreas onde existe uma grande diferença nos valores de forma vertical, como em um canto de um objeto por exemplo.

1	1	1	2
1	1	1	2
1	1	1	2

*

-1	-1	5
-1	-1	5
-1	-1	5

=

9	24
---	----

$$c1 = 1.(-1) + 1.(-1) + 1.5 + 1.(-1) + 1.(-1) + 1.5 + 1.(-1) + 1.(-1) + 1.5 = 9$$

$$c2 = 1.(-1) + 1.(-1) + 2.5 + 1.(-1) + 1.(-1) + 2.5 + 1.(-1) + 1.(-1) + 2.5 = 24$$

Figura 8: Exemplo de convolução. Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).

Camada de Unidade Linear Retificada (*ReLU - Rectified Linear Unit Layer*):

A etapa não linear é responsável por aplicar uma função não linear em cada um dos elementos dos mapas de características. HINTON et al. (2012) e ZEILER; FERGUS (2013) utilizam a função *ReLU* descrita na Equação 5, sendo z representado na Equação 4.

$$z = (W^T * x) + b \quad (4)$$

$$relu(z) = \max(0, z) \quad (5)$$

Camada de Sub-amostragem ou Redução (*Pool Layer*): A etapa *pooling* (também chamada *subsampling*) calcula a média ou máximo de uma vizinhança pré-determinada em cada um dos mapas de características. O resultado é um outro mapa com resolução menor que proporciona ausência de variação a pequenas translações. ZEILER; FERGUS (2013) demonstram um novo método para redução com base na distribuição das ativações em cada região. A Figura 9 mostra a aplicação da redução utilizando uma função MAX que seleciona o maior número em uma vizinhança 2×2 sem sobreposição.

$$f \left(\begin{array}{|c|c|c|c|} \hline 9 & 24 & 1 & 5 \\ \hline 36 & 10 & 1 & 6 \\ \hline 1 & 1 & 1 & 6 \\ \hline 1 & 1 & 1 & 6 \\ \hline \end{array} \right) = \begin{array}{|c|c|} \hline 36 & 6 \\ \hline 1 & 6 \\ \hline \end{array}$$

Figura 9: Exemplo de redução utilizando filtro *MAX* 2×2 e deslocando 2×2 (sem sobreposição). Fonte: Adaptado de JURASZEK; SILVA; SILVA (2014).

Camada de Normalização (*Normalization Layer*): Vários tipos de camadas de normalização foram propostas para seu uso em arquiteturas CNNs. Nesta dissertação e implementação utilizada, estas camadas possuem a intenção de implementar os regimes de inibição lateral observados em cérebros biológicos, onde esta inibição lateral é a capacidade de um neurônio reduzir a atividade de neurônios vizinhos, afetando o compartilhamento de parâmetros (WIKIPEDIA, 2015). Apesar desta camada ter sido utilizada, recentemente é possível notar uma diminuição no uso desse tipo de camada, por sua contribuição ser pequena (LI; KARPATHY, 2015).

Camada completamente conectada (*FC - Fully-Connected Layer*): Como visto nas redes RNAs, neurônios em uma camada completamente conectada tem conexões com todas as ativações realizadas na camada anterior. Ao término das etapas convolucionais, os mapas de características são transformados de 2 dimensões para 1 dimensão e utilizados no treinamento de um classificador totalmente conectado, com tamanho de saída correspondente à quantidade de categorias ou classes.

O resultado do classificador passa então pela função *softmax*, descrita na Equação 6, que garante que a soma de todas as probabilidades de saída da MLP resulta em 1, onde Y corresponde ao vetor de saída da MLP. A resposta final de qual categoria pode ser obtida através da seleção do item com maior probabilidade é descrita na Equação 7.

$$\text{softmax}(y_i) = \frac{e^{z_i}}{\sum_{j \in Y} e^{z_j}} \quad (6)$$

$$y^n = \text{argmax} P(Y = i | x^n, W, b) \quad (7)$$

Camada de *Dropout*: Para tentar minimizar o problema de sobre ajuste nas redes, é comum a utilização de técnicas de regularização. A regularização tem o objetivo de reduzir a quantidade de neurônios ativos quando uma determinada característica está presente na imagem. O objetivo é que apenas pequenas porções de neurônios sejam ativadas de acordo com as características observadas na imagem. A técnica proposta por (HINTON et al., 2012), chamada *Dropout* procura desativar aleatoriamente um conjunto de neurônios a cada iteração de treinamento. Com menos ativações o problema de sobre ajuste é reduzido, forçando cada camada da rede a se especializar em uma determinada característica de forma mais distinta. A Figura 10 mostra um exemplo dos filtros aprendidos com e sem a utilização do *Dropout*. É possível observar que, utilizando esta técnica, as características aprendidas pelos filtros serão muito mais distintas.

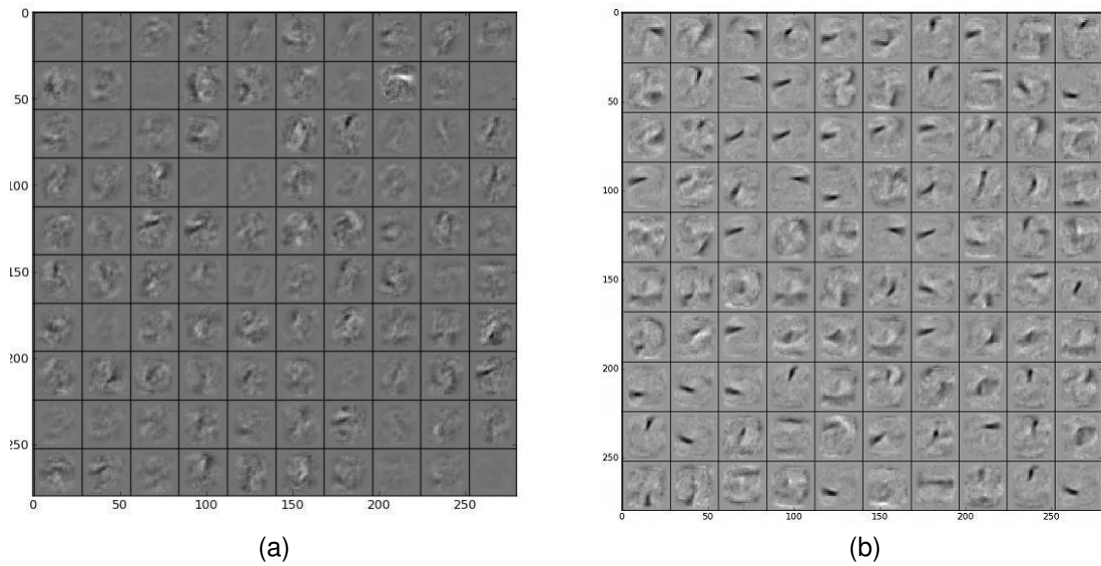


Figura 10: Exemplo de filtros aprendidos na base MNIST (a) sem *Dropout* (b) utilizando *Dropout*. Fonte: Adaptado de HINTON et al. (2012).

2.3 Máquinas de Vetores de Suporte

Seguindo as definições de OLIVEIRA-JUNIOR (2010) e LORENA; CARVALHO (2007), as Máquinas de Vetores de Suporte (*Support Vector Machines* - SVMs) têm a capacidade de resolver problemas de classificação e regressão, adquirindo com o aprendizado na etapa de treinamento, a capacidade de generalização. Considerando

um problema binário, o objetivo da SVM será de separar as instâncias das duas classes através de uma função que será obtida a partir dos exemplos conhecidos na fase de treinamento. O objetivo será produzir um classificador que funcione de forma adequada com exemplos não conhecidos, adquirindo assim a capacidade de prever as saídas de futuras novas entradas.

Algumas das principais características das SVMs que tornam seu uso atrativo são (SMOLA et al., 2000):

- Boa capacidade de generalização: os classificadores gerados por uma SVM, em geral, alcançam bons resultados de generalização. A capacidade de generalização de um classificador é medida por sua eficiência na classificação de dados que não pertençam à partição utilizada em seu treinamento. Na geração de preditores por SVMs, portanto, é evitado o *overfitting*, situação na qual o preditor se torna muito especializado na partição de treinamento, obtendo baixo desempenho quando confrontado com novos padrões.
- Robustez em grandes dimensões: as SVMs são robustas diante de objetos de grandes dimensões, como, por exemplo, imagens. Comumente há a ocorrência de *overfitting* nos classificadores gerados por outros métodos inteligentes sobre esses tipos de dados.
- Convexidade da função objetivo: a aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo global. Esta é uma vantagem sobre, por exemplo, as Redes Neurais Artificiais, em que há a presença de mínimos locais na função objetivo a ser minimizada.
- Teoria bem definida: as SVMs possuem uma base teórica bem estabelecida dentro da Matemática e Estatística.

Uma SVM constrói um classificador de acordo com um conjunto de padrões por ele identificados nos exemplos de treinamento, onde a classificação é conhecida.

Considerando o exemplo da Figura 11, nela existe um conjunto de classificadores lineares que separam duas classes, mas apenas um (em destaque) que maximiza a margem de separação (distância da instância mais próxima ao hiperplano de separação das duas classes em questão). O hiperplano com margem máxima é chamado de hiperplano ótimo, que será o objeto de busca do treinamento do classificador (GUNN et al., 1998).

2.3.1 Funções de Kernel

As funções de *kernel* têm a finalidade de projetar os vetores de características de entrada em um espaço de características de alta dimensão para classificação de

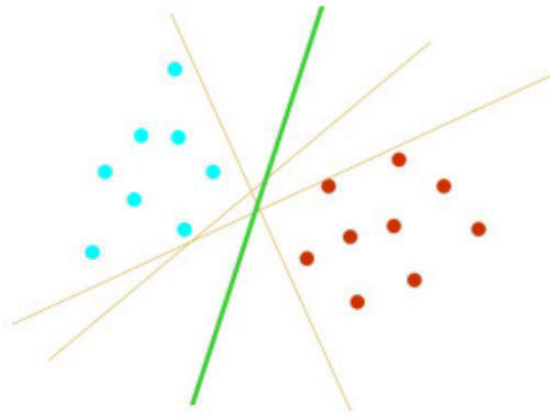


Figura 11: Possível hiperplano de separação e hiperplano ótimo (em verde). Fonte: Adaptado de OLIVEIRA-JUNIOR (2010).

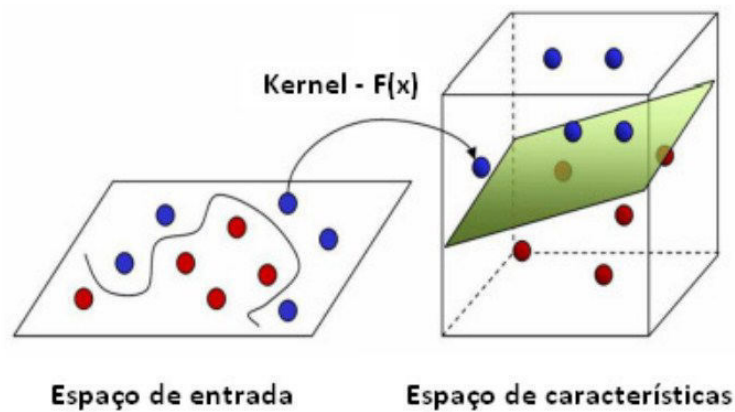


Figura 12: Transformação: problema não linearmente separável em um problema linearmente separável. Fonte: Adaptado de REBELO (2008).

problemas que se encontram em espaços não linearmente separáveis. Isso é feito, pois à medida que se aumenta o espaço da dimensão do problema, aumenta também a probabilidade desse problema se tornar linearmente separável em relação a um espaço de baixa dimensão. Entretanto, para obter uma boa distribuição para esse tipo de problema é necessário uma partição de treinamento com um elevado número de instâncias (GONÇALVES, 2015).

A Figura 12 mostra o processo de transformação de um domínio não linearmente separável, em um problema linearmente separável através do aumento da dimensão, onde é feito um mapeamento por uma função de *kernel* $F(x)$.

Uma função é definida como sendo uma função de kernel, se obedecer à seguinte teoria de Hilbert (RKHS - *Reproducing Kernel Hilbert Space*):

$$K(x_i, x_j) = (\phi(x_i) \times \phi(x_j)) \quad (8)$$

Onde ϕ deve pertencer a um domínio onde seja possível o cálculo do produto interno. Essas funções também satisfazem as condições do Teorema de Mercer (GONÇALVES, 2015).

Teorema de Mercer: uma determinada função é definida como sendo de *kernel*, se a matriz K é positivamente definida (auto valores maiores que zero), onde K é obtida por:

$$K = K_{ij} = K(x_i, x_j) \quad (9)$$

2.3.2 Kernel Gaussiano (RBF)

O *kernel* RBF é bastante utilizado para resolução de problemas de aprendizagem, inclusive é usado computacionalmente como padrão em muitas bibliotecas de linguagens de programação que utilizam o algoritmo SVM.

Na máquina RBF, diferentemente do *kernel* linear, é possível resolver problemas, originalmente, não linearmente separáveis, através do mapeamento para um espaço de maior dimensão. Existem dois parâmetros que podem ser variados em busca de um melhor resultado para o aprendizado do classificador, são eles: γ (gamma) e C (custo). No *kernel* RBF o número de funções radiais e os seus respectivos centros são definidos pelos vetores suporte obtidos. A Figura 13 mostra a aplicação de uma máquina RBF, onde é possível ver um aglomerado de pontos que representam as instâncias das classes, e a cor do fundo indica a qual classe pertence aquele *cluster* (ALMEIDA, 2007).

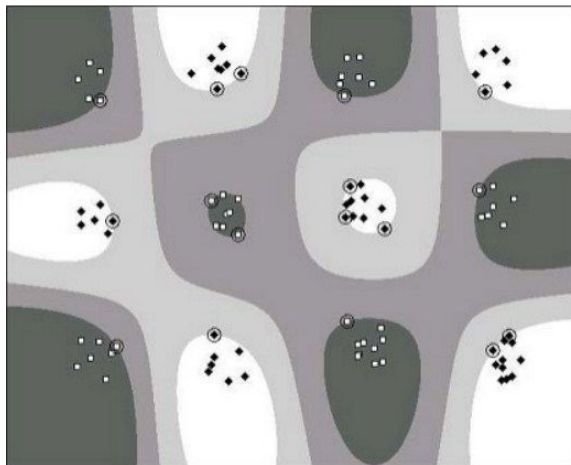


Figura 13: Separação de classes através da aplicação da máquina RBF. Fonte: Adaptado de OLIVEIRA-JUNIOR (2010).

Outro motivo para uso do *kernel* RBF é o baixo número de parâmetros (dois) que influenciam o resultado do aprendizado. No *kernel* polinomial são usados mais parâmetros que o RBF: C (custo), γ (gamma) e o grau (*degree*). Se C assumisse 11 valores, γ assumisse 10 no *kernel* RBF, seria uma combinação de 110 valores,

adicionando-se o grau e atribuindo-lhe 5 valores, a combinação iria para 550, aumentando consideravelmente o desempenho e complexidade do classificador.

2.4 Considerações sobre o Capítulo

Neste Capítulo foram apresentados conceitos de RNA's, as CNNs e SVMs, junto com seus desenvolvimentos, arquiteturas e funcionamentos.

Em relação as CNNs, a literatura consultada não especifica uma arquitetura padrão, na qual a quantidade de camadas necessárias está diretamente relacionada com o número de hierarquias de características necessário para representar e reconhecer a base de imagens a ser trabalhada. Uma base de imagens com categorias altamente distintas necessita de uma representação interna mais simples, enquanto bases onde diferentes categorias são muito semelhantes, necessitam de um nível de representações internas maior para ser capaz de distinguir uma categoria da outra.

O levantamento das principais ferramentas disponíveis e seus conceitos fundamentais foram importantes para a implementação das técnicas utilizadas neste trabalho, visando a resolução do problema apresentado.

3 APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DE IMAGENS

Até recentemente, muitas das técnicas de aprendizado de máquina e processamento de sinais exploravam arquiteturas com estruturas rasas. Essas arquiteturas tipicamente continham no máximo uma ou duas camadas de transformações de características não-lineares, como a MLP por exemplo. Estas arquiteturas rasas se mostraram efetivas para resolver muitos problemas simples ou bastante restritos, mas sua modelagem limitada e poder de representação podem causar dificuldades quando se lida com aplicações mais complexas, presentes no mundo real, como a fala humana, linguagens, imagens e cenas visuais. Desta forma, novas técnicas estão sendo desenvolvidas de modo a se obter capacidades de aprendizado ainda maiores na tarefa de reconhecimento e classificação de imagens.

Neste Capítulo serão apresentadas algumas das técnicas utilizadas para o reconhecimento e classificação de imagens, bem como trabalhos relacionados que as utilizam para a classificação de sintomas em plantas.

3.1 Redes *Multilayer Perceptron*

A técnica de *backpropagation*, que se tornou muito popular em 1970, fez muitos pesquisadores utilizarem este algoritmo para o treinamento de redes neurais artificiais, mas obtendo pouco sucesso quando se aumentava o número de camadas em problemas maiores. A tese de HOCHREITER (1991) identificou a razão desta falha por causa do *Vanishing gradient problem* (problema da diminuição do gradiente), onde este gradiente normalmente começa em pontos aleatórios e normalmente atinge um ótimo local não desejado quando os erros que deveriam se propagar diminuem a valores desprezíveis. Assim, conforme a profundidade da rede era aumentada, a severidade deste problema aumentava significativamente.

Para superar este problema, foram propostos diferentes métodos. Um deles foi a hierarquia de multi-níveis de redes, proposta por SCHMIDHUBER (1992), onde as redes foram treinadas um nível por vez de modo não-supervisionado, e ajustados uti-

lizando o *backpropagation*, fazendo cada nível aprender uma representação comprimida das observações realizadas, e esta representação foi alimentada para o próximo nível. Outro método são as redes de *long short term memory* (memórias de curto a longo prazo) proposto por HOCHREITER; SCHMIDHUBER (1997).

Assim, com os avanços das RNAs, foi possível sua aplicação em tarefas como identificação de homens e mulheres a partir de imagens com uma resolução muito baixa (TAMURA; KAWAI; MITSUMOTO, 1996). Já o trabalho de ROUSSEAU et al. (2000), faz uso de RNAs para classificar o comportamento de ratos, utilizando como entrada *frames* de vídeo.

Para a classificação do MNIST (LECUN et al., 1998), o trabalho de KUSSUL; BAIDYK (2001) explorou o aumento do número de neurônios para se obter resultados comparáveis ao estado da arte naquele momento. Também temos as RNAs sendo utilizadas para a classificação automática de imagens de alta resolução providas de satélites (FRATE et al., 2007)

3.2 Máquinas de Vetores de Suporte

Como mencionado anteriormente, as SVM constituem uma técnica embasada na Teoria de Aprendizado Estatístico (CORTES; VAPNIK, 1995), que receberam grande atenção no passado (HEARST et al., 1998; CRISTIANINI; SHAW-TAYLOR, 2000).

Os resultados da aplicação desta técnica são comparáveis aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs) (HAYKIN, 1999). Em algumas tarefas obtiveram resultados superiores, tal como na detecção de faces em imagens (HEARST et al., 1998), na categorização de textos (HEARST et al., 1998) e em aplicações em Bioinformática (ZIEN et al., 2000).

LIN et al. (2010) obtiveram o primeiro lugar no desafio de classificar o banco de imagens ImageNet, utilizando uma abordagem de extração de atributos com codificação de coordenadas locais e codificação de super vetores e, então, a aplicação de SVMs lineares.

Em 2015, apesar de não terem obtido a posição ganhadora no desafio de classificação, várias equipes de pesquisadores como "Tencent-Bestimage", "darkensemble", "DROPLET-CASIA" e "ESOGU MLCV" obtiveram resultados promissores em seu uso (RUSSAKOVSKY et al., 2015).

3.3 Redes Neurais Convolucionais

As redes CNNs se tornaram famosas por conseguirem realizar a detecção de dígitos escritos à mão, atualmente superando os resultados no banco de imagens MNIST (LECUN et al., 1998), alcançando um erro de 0.23% (CIRESAN; MEIER; SCHMIDHU-

BER, 2012) quando comparado a uma RNA cujo erro obtido foi 0,35% (CIRESAN et al., 2010). Entretanto, assim como as RNAs e RNP's, sem um poder computacional eficiente e técnicas para evitar o sobre ajuste na rede, se o número de camadas fosse aumentado, as redes deixavam de ter um bom desempenho em problemas complexos. Desde 2011, implementações desta técnica, que utilizam GPUs para auxiliar seu processamento e treinamento, ganharam vários concursos de reconhecimento de padrões. O estado da arte em Redes Neurais *feedforward* Profundas, para reconhecimento de imagens, se alternou para redes com camadas convolucionais e de *max-pooling* cobertos por diversas camadas puras de classificação. Seu treinamento vem sendo normalmente realizado sem qualquer pré-treinamento não-supervisionado.

Com a popularização de técnicas de inteligência artificial e aprendizado de máquina para a classificação de imagens, foram criados grandes bancos de imagens, utilizados para a comparação de resultados das novas técnicas desenvolvidas. Um destes banco de imagens é o ImageNet (DENG et al., 2009), que consiste em mais de 15.000.000 de imagens rotuladas em alta definição, em mais de 22.000 diferentes categorias.

Assim, várias técnicas foram pesquisadas de modo a se obter uma menor taxa de erro na classificação das imagens presentes neste banco de imagens. Uma das técnicas que se destacou para esta tarefa de classificação foi a Rede Neural Convolucional Profunda, se tornando o atual estado da arte.

GLOROT; BENGIO (2010) publicaram um estudo demonstrando a dificuldade do treinamento de RNP e o seu desempenho na base de dados *Small-ImageNet*, que é um conjunto de imagens com 37x37 píxeis em tons de cinza, divididas em 1000 classes. Como arquitetura foi utilizado cinco camadas ocultas com 1.000 unidades ocultas por camadas e 90.000 imagens para o treinamento. O resultado obtido foi de uma taxa de erro de aproximadamente 69%, utilizando as configurações apresentadas.

KRIZHEVSKY; SUTSKEVER; HINTON (2012) realizaram o treinamento de uma rede Neural Convolucional Profunda para a classificação do ImageNet e obtiveram o melhor resultado já reportado até então, utilizando os sub conjuntos presentes nas competições ILSVRC-2010 e ILSVRC-2012. A arquitetura utilizada possuía cinco camadas convolucionais e três camadas completamente conectadas, com um tempo de treinamento entre cinco e seis dias. Esta arquitetura obteve uma taxa de erro nos conjuntos top-1 e top-5 de 38,1% e 16,4%, respectivamente, utilizando um conjunto de cinco redes.

HE et al. (2014), também utilizando o banco de imagens ImageNet como base de dados, propuseram a utilização de *Spatial Pyramid Pooling* em CNNs profundas. Esta abordagem adicionou uma camada de pooleamento espacial piramidal para remover a restrição de tamanho das imagens de entrada. Segundo os autores, esta abordagem conseguiu obter aumento de desempenho satisfatório em questão de tempo de trei-

namento, e suas taxas de erros nos conjuntos top-1 e top-5 foram de 27,9% e 9,1%, respectivamente.

Baseados nos resultados obtidos por KRIZHEVSKY; SUTSKEVER; HINTON (2012), SIMONYAN; ZISSERMAN (2014) realizaram modificações na arquitetura proposta, utilizando filtros convolucionais pequenos de 3x3 de modo a se obter um desempenho melhor que os propostos anteriormente no estado da arte. Além disto, outro foco foi na profundidade da rede, a qual foi possível de ser aumentada pelo fato dos filtros serem de menor tamanho. Nesta arquitetura estavam presentes dezesseis camadas convolucionais e três camadas completamente conectadas. A taxa de erro nos conjuntos top-1 e top-5 foram de 23,7% e 6,8%, respectivamente, utilizando um conjunto de duas redes e técnicas de *multi-crop* e *dense evaluation*.

Além disso, as CNNs têm se mostrado boas alternativas nos mais diferentes problemas de visão computacional, como reconhecimento de caracteres (LECUN et al., 1998), reconhecimento de faces (LAWRENCE et al., 1997) e categorização de objetos (YU; XU; GONG, 2009).

3.4 Trabalhos Relacionados

Analisando a literatura, não foram encontrados trabalhos, utilizando as técnicas descritas, para a classificação de sintomas em folhas de macieiras, ou então, um banco de imagens com sintomas em folhas desta cultura que pudesse ser utilizado nesta dissertação. Mesmo assim, o conceito de se utilizar aprendizado de máquinas para detectar sintomas em plantas tem se mostrado uma alternativa promissora nos últimos anos. Assim, nesta sessão serão apresentados trabalhos realizados, onde diferentes abordagens foram utilizadas para a identificação ou classificação de sintomas em plantas, os quais podem contribuir para a comparação distinções com o trabalho aqui desenvolvido.

O trabalho de RUMPF et al. (2010) teve como objetivo: discriminar folhas de beterraba doentes e saudáveis, diferenciar entre os três tipos de doenças propostos no trabalho; e identificar doenças antes mesmo de sintomas específicos tornassem visíveis. Os autores utilizaram SVMs com uma função de base radial como *kernel* para realizar a identificação e classificação dos sintomas de folhas saudáveis ou doentes.

Para a captação de dados, foi utilizado um espectrorradiômetro portátil com uma sonda na planta e um suporte de clipe para a folha. O artigo não apresenta o número exato de exemplos utilizados para a abordagem proposta, mas com base na matriz de confusão apresentada, é possível notar que este número é igual ou maior que 3.570 exemplos.

Como entrada, os autores usaram nove índices espectrais de vegetação, relacionados com parâmetros fisiológicos como atributos para uma classificação automática e

a técnica de validação cruzada para treino e validação com todas as imagens presentes, resultando em uma classificação com acurácia de até 97% entre folhas saudáveis e folhas doentes, até 86% de acurácia na classificação entre os três sintomas de doenças escolhidos, e acurácias entre 65% e 90% para a detecção pré-sintomática de doenças nas plantas.

Já AL-HIARY et al. (2011) propõem uma metodologia para automaticamente detectar e classificar doenças em folhas de plantas a partir de imagens. O processo consiste em seis fases principais: aquisição de imagem, pré-processamento de imagem, segmentação de imagem, extração de atributos, análise estatística e classificação por uma rede MLP.

Os autores utilizaram um arquivo de treino e outro de teste, contendo 32 amostras para cada uma das seis classes de folhas, cada amostra foi formada de 10 atributos extraídos utilizando técnicas chamadas de Método de co-ocorrência de cor (*Color Co-occurrence Method*). A rede MLP utilizada foi composta de 10 camadas ocultas, cada uma contendo 10 neurônios seguido de uma camada de saída contendo 6 neurônios para a classificação. A rede utilizou a técnica de *feed forward back propagation* e a função de desempenho sendo a *Mean Square Error (MSE)*, o número de interações foi de 10.000 e o erro máximo permitido foi 10^{-5} . De acordo com os autores, esta abordagem obteve acurácia de 94%.

No trabalho de SANNAKKI et al. (2013) utilizaram uma rede MLP para classificar doenças foliares em uvas. Para esta abordagem, as imagens da folha passaram primeiramente por passos de remoção de fundo e pré processamento (*Anisotropic Diffusion*). O próximo passo foi a segmentação, que utilizou a técnica de *clustering* de dados, onde grandes quantidades de informação foram agrupados em *clusters* de segmentos menores com informações similares.

Assim, a imagem foi dividida em seis *clusters*, a média de cada *cluster* foi calculada e ordenada de forma ascendente, de acordo com as lesões apresentadas. Para o passo de extração de atributos das regiões com sintomas doentes, é calculada a matriz de co-ocorrência de níveis de cinza (*Gray Level Co-occurrence Matrix (GLCM)*) junto com a utilização de matrizes de dependências espaciais de níveis de cinza (*Spatial Gray-level Dependence Matrices (SGDM's)*), realizando a extração de 9 atributos.

A rede neural foi do tipo *feed forward Back Propagation Neural Network Classifier*, que consiste em três camadas, sendo estas, uma camada de entrada, uma camada oculta e uma camada de saída. Para o treino, 29 imagens foram usadas, em um total de duas classes. Foram utilizadas duas imagens validação e duas para teste. Os resultados deste trabalho, mostraram que a matriz de confusão apresentou 100% de acerto das 33 imagens testadas após o treino da rede.

REVATHI; HEMALATHA (2014a) trabalharam com doenças foliares de algodão, utilizando um banco de imagens com 270 imagens divididas em seis classes de doenças.

Para realizar a extração de atributos, a abordagem utilizou um método de extração de atributos baseados em características definidas manualmente, que consistia na borda da folha, cor e textura. Essa extração de atributos foi realizada por meio de otimização por enxame de partículas e classificadores SVM e *fuzzy*. Os autores obtiveram resultados de 91% e 94%, respectivamente.

Em uma segunda abordagem, REVATHI; HEMALATHA (2014b) utilizaram o mesmo banco de imagens, junto com a técnica de otimização por enxame de partículas para a extração de atributos, mas neste caso utilizando uma rede *Cross Information Gain Deep Forward Neural Network*, obtendo 95% de acerto.

3.5 Considerações sobre o Capítulo

Neste Capítulo foram relatadas as técnicas de aprendizado de máquina para a realização das tarefas de identificação e classificação de imagens, apresentando redes neurais, SVMs e CNNs. Além disso, foram apresentadas comparações de arquiteturas e modelos de redes utilizados para resolver o desafio da classificação do banco de imagens ImageNet.

Foram mostrados diferentes trabalhos onde são vistas abordagens para tentar solucionar o problema de identificação de sintomas em distintas plantas. É notável que todos os trabalhos apresentados obtiveram resultados satisfatórios com suas abordagens, de modo que diferentes técnicas e extração de atributos foram capazes de atingir acurácias de 91% a 100% de acerto em suas respectivas tarefas.

O referencial apresentado reforça uma das principais premissas na abordagem proposta nesta dissertação, de que não se utilize qualquer extração de atributos antes do aprendizado pela rede. Além disso, permite inferir sobre a forma de captação das imagens, sem o uso de ferramentas complementares, como as utilizadas no trabalho de RUMPF et al. (2010), que utilizaram um espectro radiômetro portátil para obter dados espectrais das folhas.

Sendo assim, com base na revisão bibliográfica realizada e nos trabalhos analisados, se tomou a decisão de se utilizar a técnica de redes CNN para analisar a viabilidade de detectar sintomas de doenças, distúrbios nutricionais e danos por herbicidas em folhas de macieira.

4 METODOLOGIA

Neste Capítulo são apresentados os protocolos metodológicos para a criação do banco de imagens, envolvendo sintomas causados por doenças, por deficiências nutricionais e por danos de herbicida em folhas de macieiras. Além disso, são apresentadas as abordagens escolhidas para a classificação das imagens coletadas, detalhando seus funcionamentos e arquiteturas, bem como as formas de avaliação destas técnicas de classificação.

4.1 Banco de Imagens

Após ser realizado um estudo na literatura sobre trabalhos que tratam da classificação de sintomas em folhas de macieiras, verificou-se a inexistência de banco de imagens de sintomas em folhas de macieira com características adequadas para a execução do trabalho. Desta forma, surgiu a necessidade da criação de um banco de imagens para verificar a viabilidade do uso das técnicas de Redes Neurais Convolucionais para a classificação das imagens, visando obter o diagnóstico de sintomas em folhas desta cultura.

Considerando o curto espaço de tempo para a coleta das imagens (apenas a safra 2014/15), concentrado nos meses de janeiro a abril, optou-se por selecionar cinco sintomas em folhas de macieiras. As folhas com sintomas foram coletadas nos pomares experimentais da Embrapa Uva e Vinho - Estação Experimental de Fruticultura de Clima Temperado, localizada em Vacaria/RS (28°30'49"S; 50°52'58"W).

4.1.1 Sintomas em Folhas

Os sintomas selecionados representam dois danos causados por desequilíbrios nutricionais (deficiência de potássio e de magnésio, apresentados respectivamente nos conjuntos de Figuras 14 e 15), dois danos por doenças (sarna da macieira - causada pelo fungo *Venturia inaequalis* e mancha de *Glomerella* - causada pelo fungo *Glomerella cingulata*, apresentados nos conjuntos de Figuras 16 e 17), respectivamente, e um dano causado por toxidez de herbicida (*glifosato*), apresentado no con-

junto de Figuras 18.



Figura 14: Folhas coletadas com sintomas de deficiência de Potássio. Fonte: Autor.



Figura 15: Folhas coletadas com sintomas de deficiência de Magnésio. Fonte: Autor.



Figura 16: Folhas coletadas com sintomas de Sarna. Fonte: Autor.

A identificação das folhas com sintomas nos pomares foi feita a partir da experiência dos profissionais da Embrapa Uva e Vinho, nas especialidades de nutrição de plantas e fitopatologia de plantas, bem como nas bases de informações técnicas sobre estes sintomas publicadas em livros especializados (NACHTIGALL; BASSO; FREIRE, 2004; VALDEBENITO-SANHUEZA et al., 2008). A descrição de cada um dos cinco sintomas é apresentada abaixo:

- Danos por deficiência de Potássio: A deficiência do potássio se manifesta na forma de clorose (amarelecimento) das folhas mais velhas, seguida de necrose



Figura 17: Folhas coletadas com sintomas de *Glomerella*. Fonte: Autor.



Figura 18: Folhas coletadas com sintomas de Danos por Herbicida. Fonte: Autor.

(escurecimento) nas margens das folhas, inicialmente nas mais velhas. A deficiência aguda se manifesta nas folhas adultas na forma de “queima das bordas” das folhas velhas, a partir das pontas das folhas.

- Danos por deficiência de Magnésio: Os sintomas se manifestam em folhas adultas e se caracterizam por amarelecimento das regiões entre as nervuras das folhas, cujas manchas, na forma de “V”, evoluem das margens da folha em direção à nervura central. Este sintoma evolui para necrosamento (escurecimento) dos tecidos atacados, culminando com a desfolha precoce da base do ramo.
- Danos por doença de Sarna: Os sintomas se caracterizam por pequenas manchas de cor verde-oliva nas folhas, as quais vão se tornando acinzentadas. As lesões são de formato circular, podendo estar isoladas ou coalescer e se espalhar por toda a superfície foliar. A infecção em folhas jovens provoca distorções na folha. Os sintomas podem aparecer em ambas às faces da folha.
- Danos por doença de Mancha de *Glomerella*: Os sintomas iniciais nas folhas caracterizam-se pela presença de manchas avermelhadas sem margens definidas, distribuídas ao acaso com tamanho variando de 1 a 4 mm. A lesão evolui e assume coloração amarelo-acinzentada podendo ou não apresentar margem marrom-avermelhada. No centro das lesões foliares mais antigas pode ocorrer a formação de pontos escuros correspondentes à frutificação do fungo.

- Danos por toxidez de Herbicida: Os sintomas iniciais caracterizam-se pela clorose entre as nervuras (amarelecimento) e pelas bordas das folhas necrosadas (escurecidas), que progridem, da borda para o centro da folha, para necrose generalizada, inclusive com perda de tecido foliar. Quando o herbicida chega às folhas através de vasos específicos (via xilema), este tende a se acumular na borda da folha e faz com que os sintomas sejam mais fortes e visualizados, primeiramente, nesses locais e nas folhas mais velhas.

4.1.2 Veracidade dos Sintomas

Mesmo que a coleta das folhas com sintomas tenha sido realizada por especialista na área agrônômica e com experiência na cultura da macieira, era necessário comprovar cientificamente a veracidade dos sintomas nas folhas coletadas, de modo a garantir a eficiência dos resultados. Para isso foram utilizadas três estratégias diferentes conforme o tipo de sintoma:

1. Para os sintomas causados por desequilíbrios nutricionais, selecionaram-se folhas com sintomas de deficiência de potássio e de magnésio, após a obtenção das imagens, e amostras de folhas normais, cada uma composta de 100 folhas. As amostras foram secas a 60 graus centígrados em estufa de circulação forçada de ar até peso constante e encaminhadas para o laboratório para a realização de análises químicas, visando à quantificação das concentrações totais dos nutrientes de interesse (potássio e magnésio). Os resultados obtidos comprovam que as amostras com sintomas representam efetivamente as deficiências de potássio e magnésio. Na Figura 19 são apresentados os resultados analíticos para as amostras de folhas normais e com deficiência de potássio, onde se verifica que as mostras de folhas com sintomas visuais de deficiência de potássio (que originaram as imagens para este sintoma) continham apenas 38% da concentração de potássio em relação às folhas normais. Também na Figura 19 são apresentados os resultados analíticos para as amostras de folhas normais e com deficiência de magnésio (em dois níveis de severidade), onde se verifica que as amostras de folhas com sintomas visuais de deficiência de magnésio (que originaram as imagens para este sintoma) continham 63% e 40% da concentração de magnésio em relação às folhas normais.
2. Para os sintomas causados por danos por doenças (sarna da macieira - causada pelo fungo *Venturia inaequalis*, e mancha de glomerella - causada pelo fungo *Glomerella Cingulata*), após a obtenção das imagens, selecionaram-se amostras de 100 folhas com sintomas previamente identificados para as duas doenças. Estas amostras foram levadas ao laboratório de fitopatologia para a incubação e multiplicação do agente causal (fungos). Após foi feito o isolamento

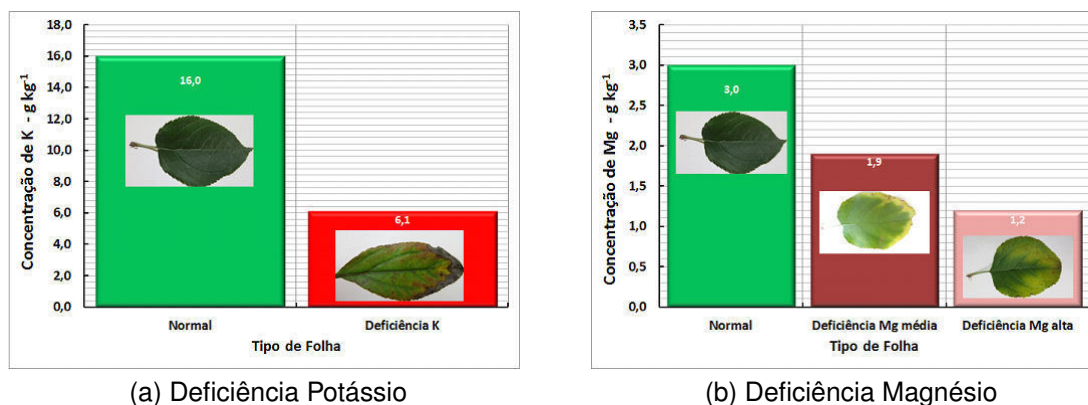


Figura 19: Resultados analíticos para as amostras de folhas de macieira normais e folhas com deficiência de potássio e de magnésio, este, em dois níveis de severidade (média e alta) (NACHTIGALL, 2015).

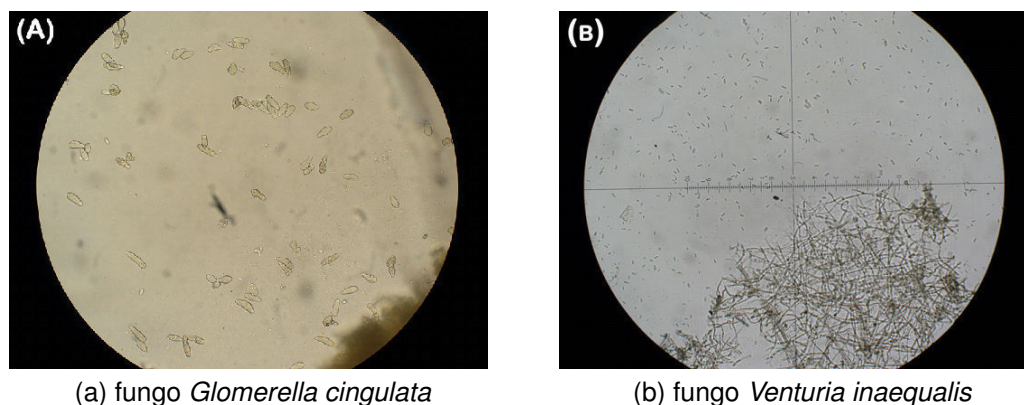


Figura 20: Identificação dos fungos *Glomerella cingulata* e *Venturia inaequalis* em microscópio (ALVES, 2015).

Tipo de Folha	N	P	K	Ca	Mg	S	Cu	Zn	Fe	Mn	B
	g kg ⁻¹						mg kg ⁻¹				
Folha Normal	18,0	1,5	17,0	19,0	4,6	1,9	6,0	127	78	631	36
Folha Sintoma Leve	20,0	1,6	18,2	19,0	4,8	1,8	7,0	127	86	607	47
Folha Sintoma Médio	19,0	1,8	18,0	21,0	5,1	2,2	8,0	173	105	798	41
Folha Sintoma Forte	18,0	1,7	18,7	17,0	4,4	2,1	8,0	162	107	685	40

Tabela 1: Concentração de macro e micronutrientes em folhas de macieira normais e com três níveis de severidade de sintomas causados por toxidez do herbicida glifosato (NACHTIGALL, 2015).

dos fungos e a sua identificação em microscópio, permitindo a comprovação dos agentes causais e seus respectivos danos nas folhas de macieira. Na Figura 20 é representada a identificação dos fungos *Glomerella cingulata* e *Venturia inaequalis*, causadores das doenças mancha de *Glomerella* e sarna da macieira, respectivamente, realizada a partir das amostras com sintomas coletadas. Os resultados obtidos comprovam que as amostras com sintomas representam efetivamente as doenças selecionadas.

3. Para os sintomas causados por danos por toxidez de herbicida (glifosato), optou-se por realizar a análise química das folhas com sintomas, visando à quantificação das concentrações totais dos nutrientes que poderiam causar confundimento dos sintomas observados, casos estivesse em concentrações abaixo do normal, uma vez que a análise do princípio ativo do herbicida é de difícil caracterização, dado que este é rapidamente degradado na planta, após a sua absorção e origem do sintoma de toxidez. Deste modo, selecionaram-se amostras de folhas com sintomas de toxidez por herbicida, após a obtenção das imagens, e amostras de folhas normais, cada uma composta de 100 folhas. As amostras foram secas a 60 graus centígrados em estufa de circulação forçada de ar até peso constante e encaminhadas para o laboratório para a realização de análises químicas. Os resultados obtidos comprovam que as amostras com sintomas não representam efetivamente distúrbios nutricionais, uma vez que as amostras de folhas com sintomas causados pelo herbicida glifosato, (em três níveis de severidade) não apresentaram concentrações de nutrientes diferentes das folhas normais (Tabela 1). Todos os nutrientes analisados nas amostras com sintomas de toxidez por herbicida estão dentro da faixa considerada como normais para a cultura da macieira (NACHTIGALL; BASSO; FREIRE, 2004).

4.1.3 Grupos de Imagens

Após a coleta das folhas, as imagens destas foram captadas individualmente, utilizando câmera fotográfica de 12 MP e fundo branco. Uma nova análise das imagens foi realizada por um agrônomo, para verificar se era possível realizar a classificação dos sintomas com base no que era apresentado nas imagens. No caso das imagens ou folhas apresentassem defeitos ou estivessem fora do padrão de captura utilizado, estas eram descartadas do banco de dados

Uma vez feita a captura e análise de todas as imagens, houve a criação de um grupo de imagens (denominado Grupo 1), com a inclusão de 290 imagens de cada sintoma selecionadas aleatoriamente, visando padronizar uma mesma quantidade de imagens, de modo a evitar que o uso de quantidades diferentes de imagens por sintoma causasse um treinamento incorreto na rede. Estas 290 imagens foram redimensionadas para a resolução de 256 x 256 píxeis, de modo a adequá-las a ferramenta a ser utilizada e ao equipamento disponível.

As imagens do Grupo 1 foram divididas em três partições para cada sintoma: treino, validação e teste. Inicialmente, foram separadas para a partição de treino 15 imagens de cada tipo de doença, totalizando 75 imagens. As 275 imagens restantes de cada tipo de doença foram separadas em 70% para treinamento (192 imagens) e 30% para validação (83 imagens). O organograma da metodologia de coleta, rotulagem e criação do banco de dados é apresentado na Figura 21.

Após as etapas de treino, validação, escolha da melhor arquitetura das técnicas selecionadas e fase de testes das imagens do Grupo 1, foi realizada uma análise preliminar dos resultados. Como na análise dos resultados obtidos pelos especialistas em agronomia (metodologia descrita na Seção 4.3), verificou-se que apenas armazenar informações das acurácias era insuficiente para a utilização de diversas métricas e desenvolver uma matriz de confusão destes resultados, sendo necessária uma nova análise. Além disso, verificou-se a necessidade de avaliar o crescimento da acurácia em relação ao número de amostras utilizadas no treinamento proposto para as diferentes técnicas.

Assim, foi criado um novo grupo de imagens (denominado Grupo 2) escolhidas aleatoriamente, dividindo 15 imagens para a partição teste e 275 para a partição de treinamento. As imagens do Grupo 2 também foram aplicadas nas arquiteturas das técnicas de aprendizado de máquina desenvolvidas, utilizando as partições de treino e validação do Grupo 1. A partição de testes, constituída de imagens inéditas, novamente foi submetida a análise de acurácia de classificação de sintomas pelas diferentes técnicas, bem como foi submetidas aos especialistas escolhidos da primeira etapa (Grupo 1), com o armazenando dos dados necessários para as métricas escolhidas.

Para a análise do crescimento da acurácia em relação ao número de amostras de cada sintoma, a partir das imagens do Grupo 2, foram criadas novas partições de treinamento contendo 5, 10, 20, 50, 100, 150, 200 e 250 imagens de cada um dos cinco sintomas, as quais foram submetidas à análise de acurácia, utilizando a técnica que apresentou maior eficiência para a classificação de sintomas nas imagens do Grupo 1.

4.2 Algoritmos

Nesta Seção serão apresentadas as configurações utilizadas para o treinamento, teste e abordagens selecionadas.

4.2.1 Redes Neurais Convolucionais

Com os processos de coleta e divisão das imagens realizados, utilizou-se a ferramenta DIGITS® para realizar o treinamento de diferentes tipos de arquiteturas de redes CNN, de modo a se obter a melhor configuração para este sistema. Para esta tarefa foi utilizado o Grupo 1 de imagens, apresentado na Figura 21, empregando as partições de treinamento e validação.

Para este procedimento foram analisados casos na literatura que tinham diferentes números de imagens e configurações de CNN. As redes testadas variaram de redes rasas, contendo uma ou duas camadas convolucionais seguidas pelas camadas de subamostragem e completamente conectadas, até redes com múltiplas camadas

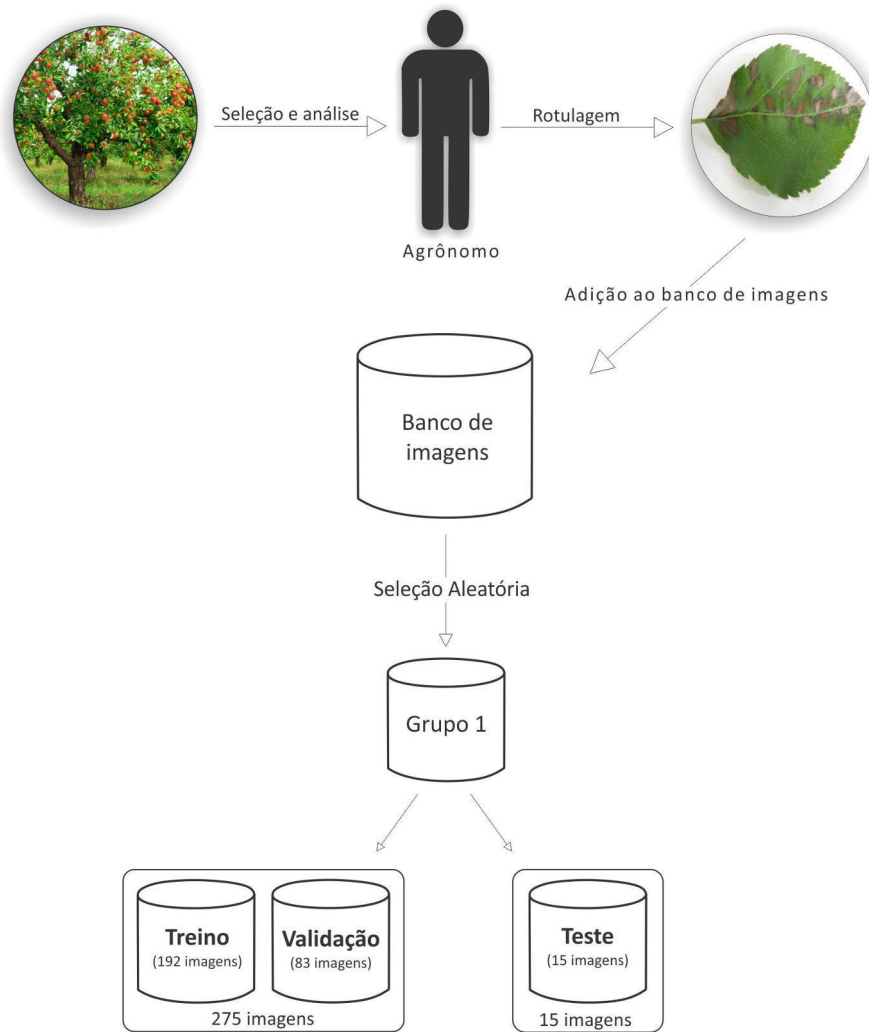


Figura 21: Esquema da metodologia de coleta e criação do banco de imagens. Fonte: Autor.

como utilizaram SIMONYAN; ZISSERMAN (2014). Realizados os diferentes testes, notou-se que uma arquitetura baseada no modelo apresentado por Alex Krizhevsky em 2012 chamado de AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) foi capaz de obter uma taxa de acerto satisfatória e realizar o treinamento da rede em até cinco horas no hardware disponível.

Esta arquitetura utilizada foi constituída das camadas descritas abaixo e apresentadas na Figura 22.

- Camada de entrada.
- 5 camadas convolucionais (*conv*), de modo que: a primeira e a segunda, foram seguidas de camadas *ReLU*, camadas de normalização e de *pooling*; a terceira e quarta foram seguidas por uma camada *ReLU*; a quinta camada seguida por uma camada *ReLU* e uma de *pooling*.
- 2 camadas completamente conectadas (*full* ou *fc* na imagem 22), seguindo a quinta camada convolucional, contendo 4.096 neurônios e uma camada completamente conectada de saída contendo cinco neurônios.

Detalhes específicos de cada camada estão apresentados na Tabela 2, onde:

C : Camada Convolucional.

M : Camada de *Max Pooling*.

N : Camada de Normalização.

F : Camada Totalmente Conectada (*Full connected layer*).

Camada	1	2	3	4	5	6	7	8
Tipo	C+M+N	C+M+N	C	C	C+M	F	F	F
Canais	96	256	384	384	256	4096	4096	5
Filtro	11*11	5*5	3*3	3*3	3*3	-	-	-
Passo da conv.	4*4	1*1	1*1	1*1	1*1	-	-	-
Pooling	3*3	3*3	-	-	3*3	-	-	-
Passo Pooling	2*2	2*2	-	-	2*2	-	-	-
Padding	2*2	1*1	1*1	1*1	1*1	-	-	-

Tabela 2: Detalhes das Camadas da rede utilizada. Fonte: Autor.

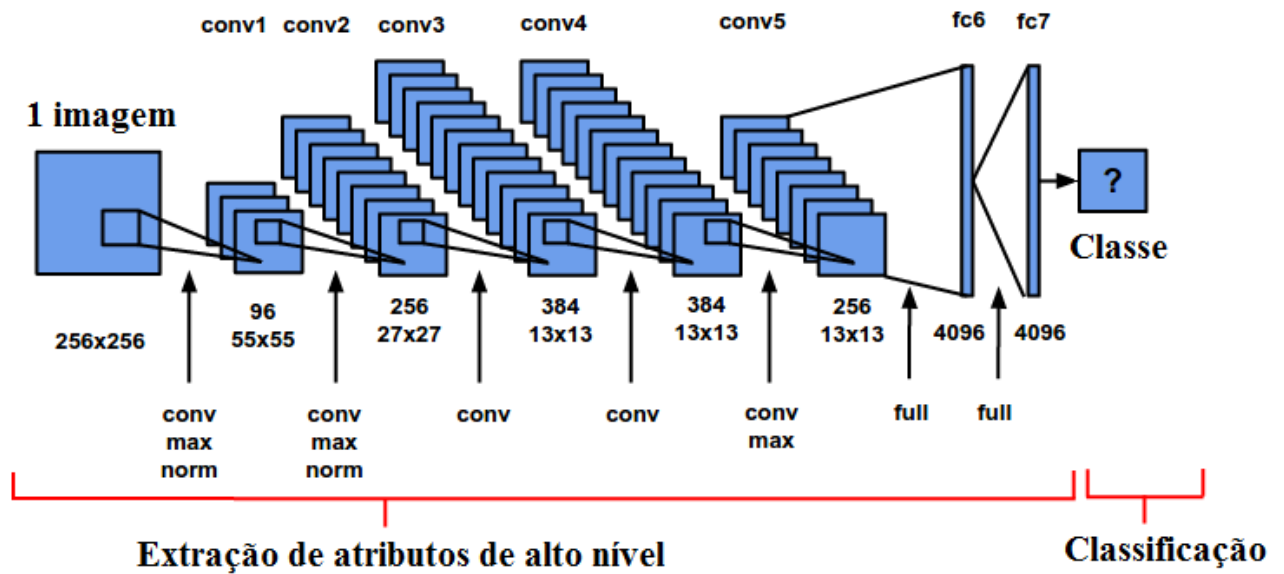


Figura 22: Arquitetura da rede CNN de Alex Krizhevsky. Adaptada de (KARNOWSKI, 2015)

Para o treinamento, a rede foi treinada por 300 épocas, onde notou-se uma convergência na taxa de aprendizado apresentada pela ferramenta utilizada. O valor da taxa de aprendizagem (*Learning rate*) foi de 0.0001. Mais informações sobre o treinamento são detalhas no Apêndice A.1.

4.2.2 Redes *MultiLayer Perceptron*

Para as configurações das redes MLP, houve dificuldades em se encontrar trabalhos na literatura, onde estas foram usadas para a classificação de imagens com altas resoluções sem a utilização de técnicas de extração de atributos. Sendo assim, a abordagem para este caso foi a realização de diversos testes com diferentes configurações de rede, também utilizando a ferramenta DIGITS®, onde foram alterados o número de neurônios por camadas e o número de camadas, variando de uma até três camadas.

Nestes testes das redes MLP também foram aplicados os mesmos protocolos apresentados anteriormente, utilizando o Grupo 1 de imagens, e as partições de treinamento e validação. Sendo assim, a melhor configuração encontrada foi obtida utilizando uma rede de seis camadas: onde a primeira camada foi a entrada da rede; as segunda e terceira camadas possuindo 200 e 500 neurônios respectivamente, sendo estas seguidas por camadas ReLU de função de ativação; e a última camada possuindo cinco neurônios de saída, de acordo com o número de classes do problema.

Para o treinamento, esta rede também foi treinada por 300 épocas. E utilizou-se a mesma taxa de aprendizagem à aquela das CNNs. Mais informações sobre o treinamento são detalhas no Apêndice A.1.

Para o treinamento, a rede teve 300 épocas, utilizando uma taxa de aprendizagem (*Learning rate*) de 0.0001. Mais informações sobre o treinamento são detalhas no Apêndice A.2.

4.2.3 Máquina de Suporte de Vetores

Para as SVMs, foram utilizadas as ferramentas LibSVM (CHANG; LIN, 2011), PowerMeanSVM (WU, 2012) e LibLinear (RONG et al., 2008). Seguindo os protocolos anteriores, também foi utilizado o Grupo 1 de imagens, com as partições de treino e validação de modo a obter as melhores configurações desta técnica. Como estas ferramentas não têm suporte à entradas no formato de imagem, primeiro foi necessária a conversão das imagens para o formato aceito. Esta conversão foi feita transformando primeiro as imagens em arquivos de texto, onde cada pixel da imagem foi transformado em três valores RGB, que variam de 0 a 255, e então estes valores foram inseridos em um arquivo de texto, seguindo os padrões indicados pela documentação apresentada em CHANG; LIN (2011).

Ainda seguindo as recomendações presentes na documentação das ferramentas, o arquivo de entrada teve seus valores escalados para valores entre 0 e 1. Para a ferramenta LibSVM, utilizou-se o *kernel* Gaussiano linear (KEERTHI; LIN, 2003), de modo a possibilitar a verificação da diferença desta abordagem com a da ferramenta LibLinear que utiliza um kernel linear. Já a ferramenta PMSVM utilizou como padrão o chi-square *kernel*.

4.3 Classificação por Especialistas

Como estratégia para contornar a situação de ausência de trabalhos similares ao desenvolvidos nesta dissertação, os quais permitiriam a comparação de eficiência da abordagem proposta para classificar os diferentes sintomas avaliados com os obtidos em outras situações, optou-se por comparar a acurácia obtida pelas técnicas de aprendizado de máquinas utilizadas com a acurácia obtidas por especialistas na cultura da macieira. Para isso, se obteve o apoio da Embrapa Uva e Vinho - Estação Experimental de Fruticultura de Clima Temperado, para a seleção destes especialistas em macieira. Este conjunto de especialistas foi formado por pesquisadores com títulos de mestrado e doutorado, atuando nas áreas de fitopatologia, pós-colheita de frutos, nutrição de plantas, fitotecnia e entomologia.

Para analisar o percentual de acertos destes especialistas, foi desenvolvido e encaminhando individualmente um questionário online sem limite de tempo, para classificação dos mesmos sintomas da partição de testes do Grupo 1 de imagens. Em sua página inicial, este questionário continha uma explicação de seu funcionamento, incluindo uma imagem exemplo de como seriam as questões (Figura 23). Ainda na

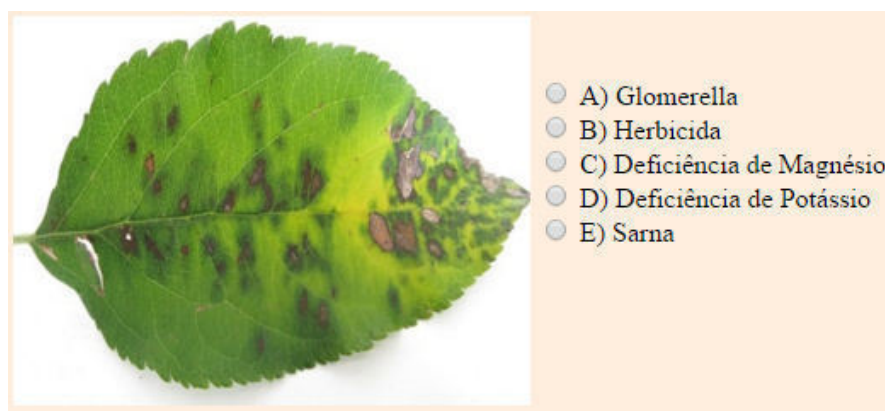


Figura 23: Exemplo de questão do primeiro questionário enviado aos especialistas.
Fonte: Autor.

página inicial era solicitada a identificação do especialista, através do seu e-mail, bem como um questionário para a identificação de quais sintomas o especialista já possuía conhecimento e experiência. Caso o especialista não possuísse conhecimento prévio em um dos sintomas, no início do questionário de classificação dos sintomas era aberta uma nova janela apresentando uma breve explicação do sintoma e uma imagem de exemplo.

Neste questionário as imagens foram apresentadas aos pesquisadores de forma aleatória e estes deveriam selecionar um dos cinco sintomas existentes, caso a questão fosse deixada sem marcação ela era considerada errada. Foram selecionados 12 especialistas para este questionário, sendo que o resultado de cada participante só era revelado após a conclusão de todos os especialistas, de modo a evitar que este realizasse diversas vezes o teste para melhorar sua pontuação. No final, foi informado a cada participante somente o seu próprio resultado, sendo que estes resultados e nomes dos participantes não foram ou serão divulgados publicamente.

Terminado o questionário pelo participante, o resultado era enviado ao autor deste trabalho com as informações do número de questões corretas de cada um dos cinco sintomas, junto da acurácia final de cada especialista.

Analisando os dados recebidos, observamos que em um caso real, um produtor pode pedir a opinião profissional de diversos especialistas para a classificação dos sintomas apresentados nas folhas, onde cada especialista faz uso do seu conhecimento específico em relação ao sintoma apresentado. Sendo assim, para uma melhor comparação com o trabalho realizado, um novo formulário foi desenvolvido e enviado a 10 especialistas que obtiveram as maiores acurácias entre os 12 participantes no questionário anterior, visando a criação de um sistema de votação com melhor capacidade de classificação.

Como apenas seis destes especialistas responderem ao questionário enviado, optou-se por incluir o agrônomo que realizou a coleta e rotulagem das imagens para

responder o novo formulário. Esta decisão foi tomada pelo fato de que cada especialista é perito em uma área de pesquisa, e em um caso real, um produtor pode se utilizar de mais de um especialista, ou então de alguém capaz de identificar diferentes sintomas em plantas. Sendo assim, um especialista que teve contato constante com todos os sintomas descritos nessa dissertação seria capaz de fornecer uma métrica melhor de comparação para a abordagem proposta.

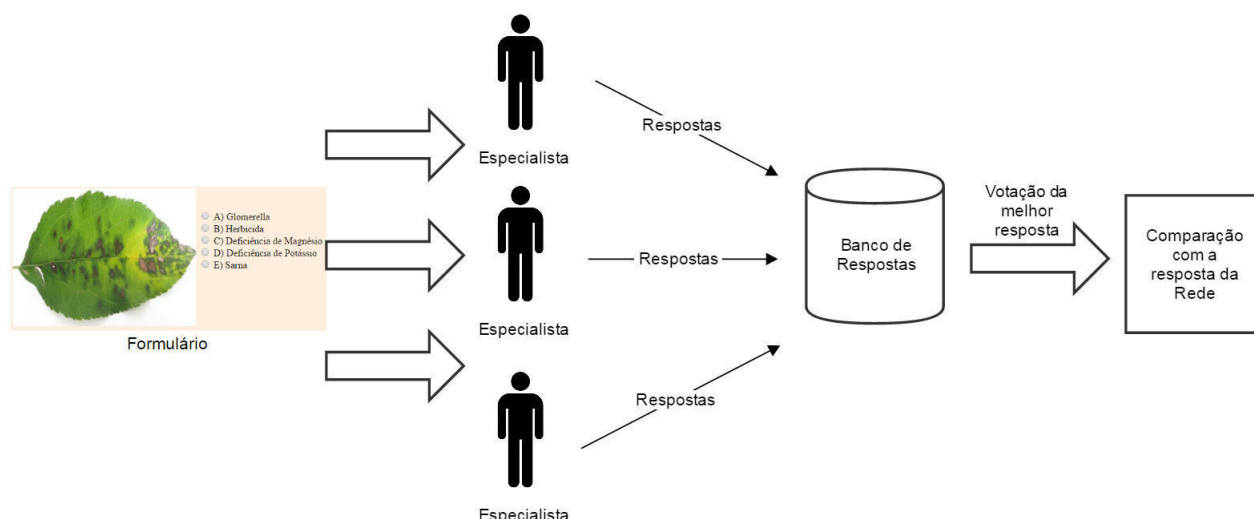


Figura 24: Diagrama do sistema de votação desenvolvido. Fonte: Autor.

A premissa deste novo questionário foi de armazenar todas as respostas dos especialistas, e assim criar um sistema de votação, apresentado na Figura 24. Por esta metodologia, foi selecionada a resposta mais votada entre todos os participantes do questionário, a qual foi considerada como final.

Caso existisse um número igual de votos entre diferentes imagens, a resposta selecionada pelo sistema foi baseada no cálculo da média de acurácia entre os especialistas, selecionando a opção votada, cuja a média de acertos foi maior.

Esta proposta fica clara com a simulação apresentada na Figura 25, onde temos o especialista 1 e 2 selecionando a opção A como resposta, e os especialistas 3 e 4 selecionando a opção B como resposta. Como a média de acertos dos especialistas 1 e 2 foi de 91,3% e a média de acertos dos especialistas 3 e 4 foi de 76,6%, o sistema escolhe a resposta A como sua resposta final.

Para este segundo questionário foi utilizado a partição de testes do Grupo 2 de imagens. Além disso, neste segundo questionário todos os especialistas possuíam acesso a janela de ajuda contendo todas as imagens e descrições dos sintomas.

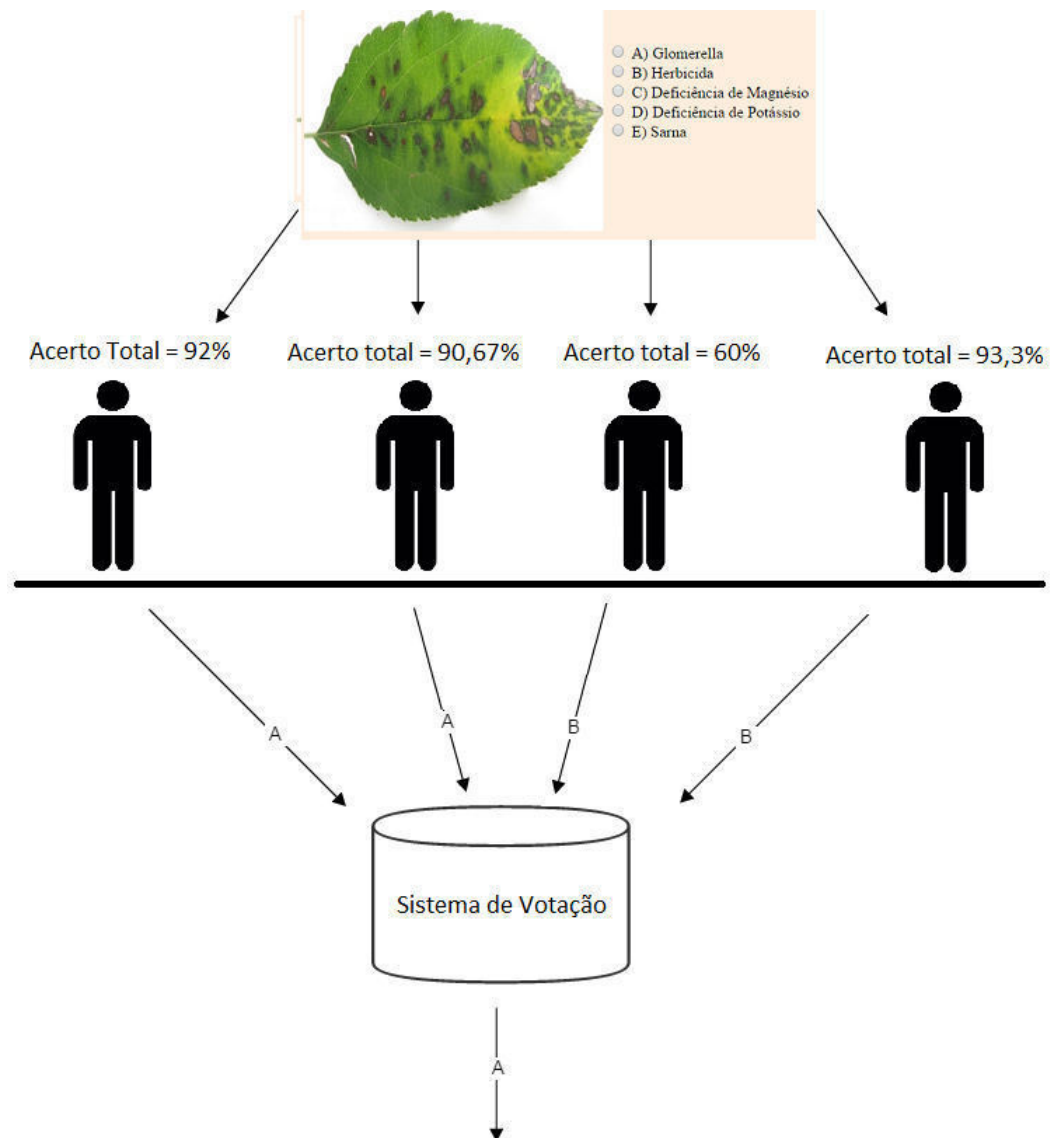


Figura 25: Diagrama do funcionamento do sistema de votação desenvolvido. Fonte: Autor.

4.4 Avaliação dos Resultados

A análise dos dados considerou os resultados obtidos pelas abordagens de aprendizado de máquina selecionadas, utilizando a partição de teste de cada um dos três Grupos de imagens (treinadas através das partições de treino e validação de seu respectivo Grupo), e os resultados obtidos dos dois questionários submetidos aos especialistas.

De modo a conceituar as métricas usadas, a Tabela 3 apresenta um exemplo de Matriz de Confusão com duas classes: A e B, onde podem ser classificadas como: VP (Verdadeiros Positivos), VN (Verdadeiros Negativos), FP (Falsos Positivos) e FN (Falsos Negativos).

	A	B
A	VP	FN
B	FP	VN

Tabela 3: Exemplo de Matriz de Confusão

Para a análise e comparação dos resultados obtidos, foi selecionada a acurácia como métrica de comparação. Para este trabalho, a acurácia representa a porcentagem de imagens classificadas corretamente, formalizada pela Equação 10:

$$\text{Acurácia} = \frac{\text{número de imagens corretamente classificadas}}{\text{número total de imagens na partição}} \quad (10)$$

Além disso, para se obter uma análise por classe das redes, também se optou por utilizar o cálculo do *Recall*, de modo a verificar se: dado um exemplo, o classificador é capaz de classificá-lo corretamente. O *Recall* é dado pela Equação 11, onde: **Recall** é a proporção de casos corretamente classificados dentro de uma classe; **VP** é o número de predições na classe correta; **FN** é o número de predições em classes erradas.

$$\text{Recall} = \frac{VP}{VP + FN} \quad (11)$$

Outra métrica utilizada foi a Precisão, que trata da capacidade do classificador em verificar se obtida uma predição sobre uma classe, quão correta essa classificação pode ser considerada. O cálculo da Precisão é dado pela Equação 12, onde: **Precisão** é a chance de casos corretamente classificados dentro de uma classe; **VP** é o número de predições na classe correta; **FP** é o número de predições erradas na classe indicada.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (12)$$

Por fim, de modo a se ter uma métrica adicional de comparação entre classificadores, foi utilizada a métrica *Kappa* (LANDIS; KOCH, 1977) cuja descrição é apresentada na Tabela 4. Desta forma, foi obtida uma relação do valor de *Kappa* com a força de concordância (*strenght of agreement*) do classificador. Na Equação 14, a acurácia total é dada pela Equação 10 e a acurácia aleatória é dado pela Equação 13.

$$\text{Acurácia Aleatória} = \frac{1}{\text{Número de Classes}} \quad (13)$$

$$\text{Kappa} = \frac{\text{Acurácia Total} - \text{Acurácia Aleatória}}{1 - \text{Acurácia Aleatória}} \quad (14)$$

Valor de <i>Kappa</i>	Força de Concordância
<0,00	Medíocre
0,21-0,20	Fraco
0,21-0,40	Razoável
0,41-0,60	Moderado
0,61-0,80	Substancial
0,81-1,00	Próximo ao Perfeito

Tabela 4: Valores de *Kappa* com sua Força de Concordância. Fonte: Adaptado de LANDIS; KOCH (1977)

A importância de utilizar ambas as métricas de *Recall* e *Precisão* se dá pelo fato de que, mesmo em casos onde a *Precisão* é de 100%, ou seja, sempre que o classificador realizar a predição de um exemplo como uma classe definida, ela estará correta, é também necessária a informação de quantos exemplos foram classificados de forma incorreta, apresentada na métrica *Recall*.

Cabe destacar que as métricas de *Recall*, *Precisão* e *Kappa*, foram apenas utilizadas na técnica de CNN (por ser o foco principal dessa dissertação) e com a classificação dos especialistas (por simular um caso mais próximo ao real), de modo a verificar a viabilidade de utilização da abordagem proposta para a classificação de imagens com sintomas em folhas de macieira.

Para avaliar a confiabilidade de acurácia para cada proporção de acerto nas técnicas testadas, foram obtidas as estimativas por intervalo a 99% de confiança das proporções de acertos em cada grupo para cada método de classificação utilizado. Os intervalos foram comparados com os obtidos para a proporção sob a hipótese de nulidade de que as respostas foram dadas de maneira casual, ou seja, a taxa de acertos seria de 20% ($p_0 = 0,20$).

5 RESULTADOS E DISCUSSÃO

Nesse Capítulo serão apresentados e discutidos os resultados obtidos desde a coleta e rotulagem das folhas de macieira com os sintomas que foram selecionados para a formação do banco de imagens, até os resultados obtidos com a classificação das partições de teste dos três Grupos de imagens, quando aplicados nas abordagens mencionadas na metodologia adotada. Também será apresentada comparação entre os resultados obtidos pelas técnicas de aprendizado de máquina utilizadas e os resultados obtidos pelos especialistas.

5.1 Banco de Imagens

A coleta das folhas com sintomas para esta dissertação foi limitada entre os meses de janeiro e abril de 2015, considerando a tomada de decisão em realizar o trabalho (início) e a proximidade do outono (fim), já que após esta fase ocorre o amarelecimento e queda das folhas, comprometendo as amostras e, conseqüentemente, prejudicando a identificação dos sintomas. A coleta e rotulagem dos sintomas buscou obter a maior variabilidade possível de condições de ocorrência do mesmo sintoma nas folhas, visando obter maiores chances de aprendizagem e classificação dos sintomas pelos métodos selecionados.

Sintoma	Amostras
Mancha da Glomerella	581
Dano por Herbicida	325
Deficiência de Magnésio	355
Deficiência de Potássio	341
Dano por Sarna	393

Tabela 5: Número de folhas de macieira com sintomas que foram coletadas e rotuladas. Fonte: Autor.

Mesmo com um período limitado de tempo para a obtenção das imagens, ainda assim foi possível a coleta e rotulagem de um total de 1.995 imagens dos cinco sintomas selecionados (Tabela 5), após o descarte de situações de imagens com baixo foco ou com dúvidas sobre a condição do sintoma apresentado. Não foi possível a criação de uma base de dados homogênea, contendo um número padrão de imagens para cada classe, o qual variou entre 581 e 325 imagens. Porém, para a execução e conclusão deste trabalho, esta base de dados permitiu, de forma eficiente, a utilização das abordagens selecionadas, nos diferentes Grupos de imagens criados. As imagens utilizadas para a realização dos treinos e testes dos três Grupos de imagens selecionados são apresentadas no Apêndice A deste trabalho

A análise quanto à robustez do banco de imagens utilizado neste trabalho, quando comparada a de outros trabalhos que utilizaram metodologias semelhantes, mostra grande variabilidade de situações. No trabalho de SANNAKKI et al. (2013), que teve como objetivo, o diagnóstico e classificação de doenças da folha da uva utilizando redes neurais, foram utilizadas 33 imagens coletadas com o auxílio de especialistas para a identificação e rotulagem para a divisão em duas classes. Essas imagens então foram divididas em três partições: 29 na partição de treinamento, duas na partição de validação e duas na partição de testes. Além disso, o trabalho relata o uso de imagens adquiridas pela internet para complementar o banco de imagens proposto.

O trabalho de AL-HIARY et al. (2011), que teve como objetivo classificar imagens de doenças em plantas utilizando extração de atributos e redes MLP, utilizou 192 imagens para treino e 192 para teste, divididas igualmente entre seis classes.

No trabalho de REVATHI; HEMALATHA (2014a), que utilizou a técnica de SVM para diagnosticar doenças em folhas de algodão, a metodologia apresentada detalha a coleta de 270 imagens divididas em seis classes, a coleta teve auxílio e rotulagem de fazendeiros da região. As imagens foram redimensionadas para uma resolução de 150 x 150 píxeis e divididas em 270 imagens para treinamento e 120 para teste.

Já o trabalho de RUMPF et al. (2010), teve uma abordagem de inocular as doenças escolhidas em folhas, e então realizar a captura de 3.570 imagens, divididas em quatro classes conforme os sintomas se desenvolviam. Uma distinção entre a metodologia proposta por RUMPF et al. (2010) e a desta dissertação, está no fato de que nas imagens obtidas por estes autores, a captura foi realizada com foco específico nos sintomas e não em uma folha completa da planta.

Embora o banco de imagens utilizado nesta dissertação, composto por 1.995 amostras de cinco distintos sintomas em folhas, possa representar um número pequeno de imagens para o treinamento de redes CNN, ainda assim, apresenta número de imagens superior à maioria dos trabalhos citados que obtiveram êxito nas situações de classificação de imagens, o que indica que a estratégia adotada foi adequada para alcançar os objetivos propostos.

Cabe destacar que, os bancos de imagens apresentados nos trabalhos relacionados citados, não estavam compartilhados abertamente para o uso. O banco de imagens utilizado nesta dissertação, junto com as partições utilizadas para treino e teste de cada grupo de imagens, estão disponíveis para acesso em: <https://www.dropbox.com/s/b81z064ohynhlgn/DataSet-AppleLeaves.zip> e no Apêndice A.

5.2 Desempenho por Número de Amostras por classe

Na Figura 26 são apresentados os resultados da Acurácia obtidos para cada um dos cinco sintomas estudados, quando utilizados diferentes números de imagens para treinamento visando à classificação utilizando a técnica CNN. Verifica-se que para os sintomas de *Glomerella* e *Sarna* os índices de acurácia foram elevados mesmo com número pequeno de imagens, em comparação com os demais sintomas.

A Figura 27 mostra que existe um incremento logarítmico na precisão em função do número de amostras utilizadas, onde estes acréscimos na precisão passam a ser menores a partir de 150 amostras para treinamento. Este resultado é particularmente importante, pois indica que o número de amostras utilizadas neste trabalho (275) para o treinamento das diferentes técnicas está situado na faixa de obtenção de maior acurácia de classificação das imagens utilizadas.

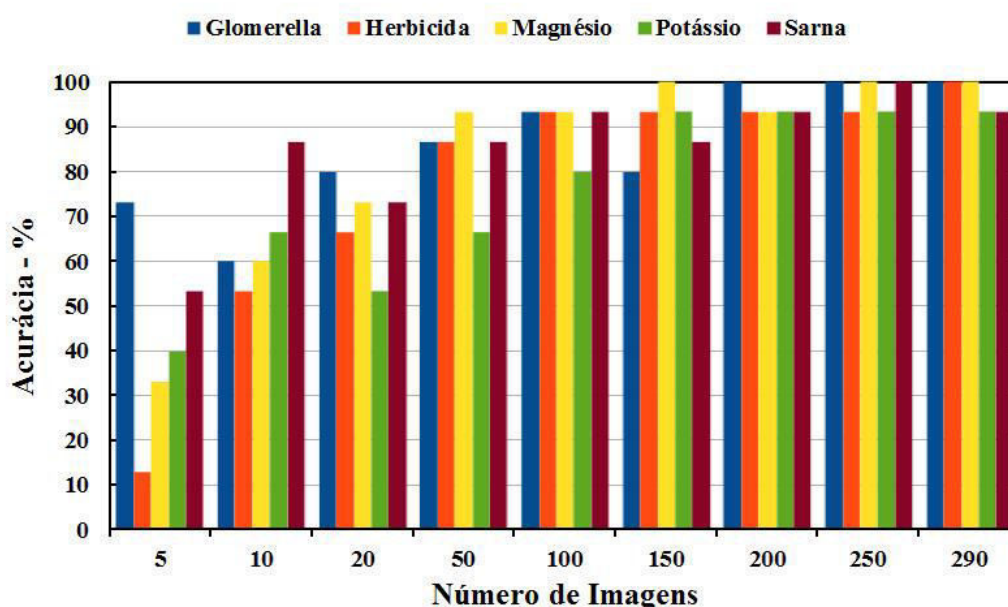


Figura 26: Distribuição das acurácias obtidas para cada um dos cinco sintomas estudados em função do uso de diferentes números de imagens para treinamento visando à classificação pela técnica CNN no Grupo 2 de imagens. Fonte: Autor.

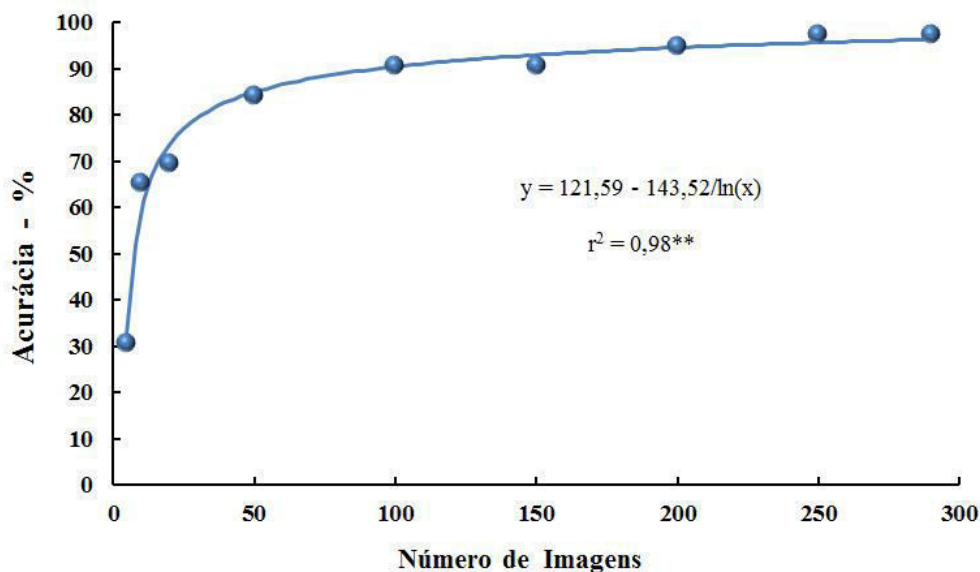


Figura 27: Relação entre as acurácias médias dos cinco sintomas em folhas de macieira e o número de imagens na partição de treinamento utilizada para a técnica CNN no Grupo 2 de imagens. Fonte: Autor.

5.3 Classificação de Imagens

5.3.1 Grupo 1

Os resultados apresentados na Tabela 6, mostram as Acurácias obtidas pelos pesquisadores no Grupo 1 de imagens. Foram selecionados 12 especialistas, que responderam ao questionário enviado. No cálculo da média da Acurácia dos resultados obtidos, estes especialistas obtiveram uma acurácia de 51,2%.

Na Tabela 7, estão apresentadas as acurácias obtidas pelas abordagens testadas, utilizando as partições de testes do Grupo 1 de imagens. Verifica-se que os menores percentuais de acerto na classificação de imagens de sintomas, utilizando técnicas de aprendizado de máquina, foram obtidos quando utilizou-se as técnicas LibLinear e PMSVM.

Quando comparadas as acurácias obtidas pela técnica de CNN, utilizando o Grupo 1 de imagens, obteve-se um aumento de acurácia de 9,3% quando comparados a redes MLP, e 60,0% quando comparada a técnica de LibSVM, que obteve a maior acurácia entre as ferramentas de SVM utilizadas.

Na análise dos resultados de classificação de sintomas obtidos pelos especialistas, deve-se considerar que para o Grupo 1 de imagens foi utilizado o melhor resultado obtido entre todos os especialistas que responderam ao primeiro questionário enviado, o que representou um percentual de acerto de 93,3%.

Quando a matriz de confusão da rede CNN foi analisada utilizando a partição de testes do Grupo 1 (Tabela 8) e o *Recall* como métrica, foi possível observar um erro maior nas folhas com sintomas de sarna. Realizando uma análise da precisão, pode-

Especialista	Área	Acurácia
1	Fitopatologia	93,3%
2	Pós-colheita	72,0%
3	Fitopatologia	69,3%
4	Nutrição	66,7%
5	Entomologia	57,3%
6	Nutrição	54,7%
7	Melhoramento	49,3%
8	Fitotecnia	44,0%
9	Fitotecnia	40,0%
10	Fitotecnia	38,7%
11	G. Ambiental	24,0%
12	Nutrição	5,3%
Média	-	51,2%

Tabela 6: Acurácias obtidas pela classificação dos especialistas nas partições de testes dos Grupos 1 de imagens. Fonte: Autor.

Técnica	Grupo 1
CNN	93,3%
MLP	84,0%
LibLinear	28,0%
PMSVM	28,0%
LibSVM	33,3%
Especialistas	93,3%

Tabela 7: Percentuais de acurácia na classificação de imagens de sintomas obtidos pelas abordagens propostas conforme o Grupo 1 de imagens. Fonte: Autor.

se notar uma tendência da rede classificar erroneamente imagens que apresentam sintomas de deficiência de Magnésio. Além disso, mesmo se obtendo uma acurácia menor com o Grupo 1 de imagens, com base na métrica *Kappa*, essa classificação também é considerada como quase perfeita.

Imagem Rotulada	Classificação							
	Sintoma	Glomerella	Herbicida	Def. Magnésio	Def. Potássio	Sarna	Recall	Precisão
	Glomerella	15	0	0	0	0	100,0%	93,7%
	Herbicida	0	14	0	1	0	93,3%	93,3%
	Def. Magnésio	0	0	15	0	0	100,0%	88,2%
	Def. Potássio	0	0	0	14	1	93,3%	93,3%
	Sarna	1	0	2	0	12	80,0%	100,0%
Acurácia								93,3%
Kappa								0,92

Tabela 8: Matriz de confusão da CNN utilizando o Grupo 1 de imagens. Fonte: Autor.

Na estimativa por intervalo de confiança ($IC_{1-\alpha}(p)$) com aproximação normal para cada proporção de acerto nas técnicas testadas, a 99% de confiança, verificou-se que alguns dos intervalos coincidiram em parte com o intervalo para $p_0 = 0,20$ (20%), que seria a proporção esperada caso fosse casual a resposta a cada imagem (Figura 28). Estas situações ocorreram para as técnicas Liblinear, PMSVM e LibSVM, no Grupo 1 de imagens.

5.3.2 Grupo 2

Para a Tabela 9, foram selecionados apenas os seis especialistas que responderam ao segundo questionário, juntamente com o agrônomo que realizou a coleta e rotulagem das imagens. Realizando o cálculo da média das acurácias apresentadas na Tabela 9, se obteve uma média de 71,9% de acurácia entre os especialistas.

A Acurácia média de 51,2% (Tabela 6), obtida por especialistas no Grupo 1 de imagens, e a Acurácia média de 71,9% (Tabela 9) pelo Grupo 2 de imagens, representam a dificuldade da classificação de imagens de sintomas para determinadas áreas de especialidade, mesmo considerando um grupo de pessoas especialista na cultura da macieira. Provavelmente o percentual de acerto na classificação de sintomas em folhas de macieira por indivíduos não especialistas seja inferior a estes resultados obtidos. Considerando que na prática a pessoa que detecta os sintomas possui, provavelmente, conhecimentos insuficientes para identificar o problema corretamente, a

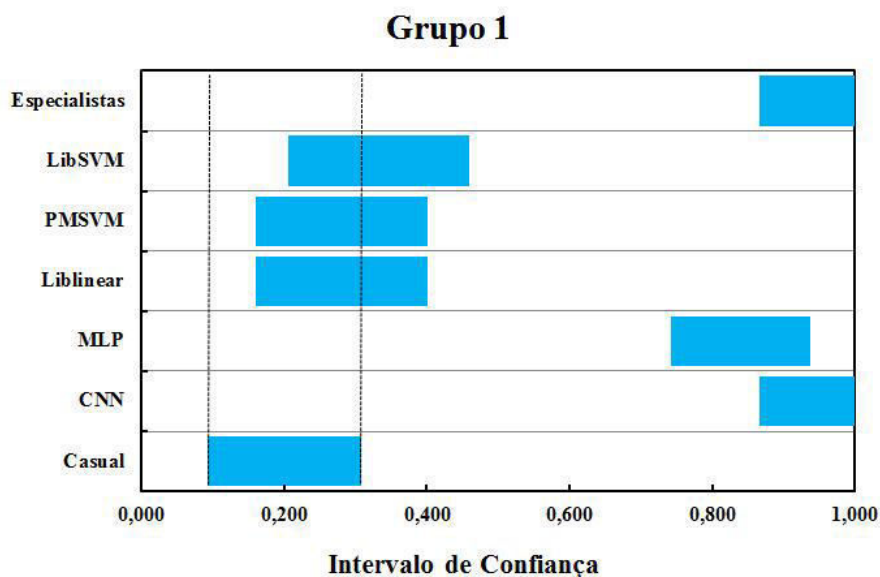


Figura 28: Estimativa por intervalo de confiança ($IC_{1-\alpha}(p)$) com aproximação normal para cada proporção de acerto nas técnicas testadas, a 99% de confiança para o Grupo 1 de imagens. Fonte: Autor.

eficiência do diagnóstico poderá ser insuficiente para minimizar perdas de produção do pomar de macieira.

Especialista	Área	Acurácia
1	Solos	93,3%
2	Fitopatologia	92,0%
3	Pós-colheita	90,6%
4	Fitopatologia	70,6%
5	Nutrição	60,0%
6	Fitotecnia	60,0%
7	G. Ambiental	37,3%
Média	-	71,9%

Tabela 9: Acurácias obtidas pelas classificação dos especialistas nas partições de testes dos Grupos 2 de imagens. Fonte: Autor.

Para o Grupo 2, os menores percentuais de acerto foram obtidos com as técnicas LibSVM e PMSVM, contudo, com valores superiores que aos menores valores do Grupo 1, apresentado na Tabela 10

Para o Grupo 2 de imagens, o resultado dos especialistas foi computado pelas

Técnica	Grupo 2
CNN	97,3%
MLP	77,3%
LibLinear	61,3%
PMSVM	50,6%
LibSVM	48,0%
Especialistas	96,0%

Tabela 10: Percentuais de acurácia na classificação de imagens de sintomas obtidos pelas abordagens propostas conforme o Grupo 2 de imagens. Fonte: Autor.

respostas obtidas pelo sistema de votação, representando 96% de acurácia. Além disso, a técnica de CNN obteve uma acurácia 20% maior que a acurácia obtida pela rede MLP e 36% maior que a abordagem utilizando LibLinear.

A baixa acurácia obtida pela técnica de SVM neste trabalho, como visto nos resultados das ferramentas LibLinear, PMSVM e LibSVM, difere dos resultados obtidos no trabalho de REVATHI; HEMALATHA (2014a), onde a técnica de SVM foi utilizada para alcançar acurácias de até 91%, utilizando extrações de atributos. Esta condição favorece a hipótese de que a baixa acurácia obtida pela técnica de SVM, obtida nesta dissertação, esteja relacionada ao não emprego das técnicas de extração de atributos na abordagem apresentada, ocasionando um número extremamente alto de atributos para estes algoritmos, dificultado seu treinamento e classificação.

No Grupo 2 não houve coincidência entre os intervalos de confiança das técnicas utilizadas e o intervalo de confiança para a resposta casual (20%) Figura 29. Isso possivelmente ocorreu por diferentes imagens terem sido utilizadas no treinamento e teste de cada Grupo, onde no Grupo 1 pode ter sido testado com imagens de mais difícil classificação, ou um treinamento mais específico dos sintomas, não generalizando suficientemente para uma correta classificação de novas imagens.

Para os dois Grupos avaliados, o intervalo de confiança da técnica CNN e o resultado obtido pelos especialistas apresentaram valores bem distintos do intervalo de confiança caso a resposta esperada para cada imagem fosse casual.

Cabe destacar que a acurácia obtida neste trabalho para a classificação de imagens de sintomas em folhas de macieira, através da técnica CNNs (valores superiores a 97%), foi similar ou superior as acurácias obtidas por RUMPF et al. (2010) para classificar folhas saudáveis e folhas doentes de beterraba, por AL-HIARY et al. (2011), para classificar doenças em folhas de plantas, e por REVATHI; HEMALATHA (2014a) para classificar folhas doentes de algodão.

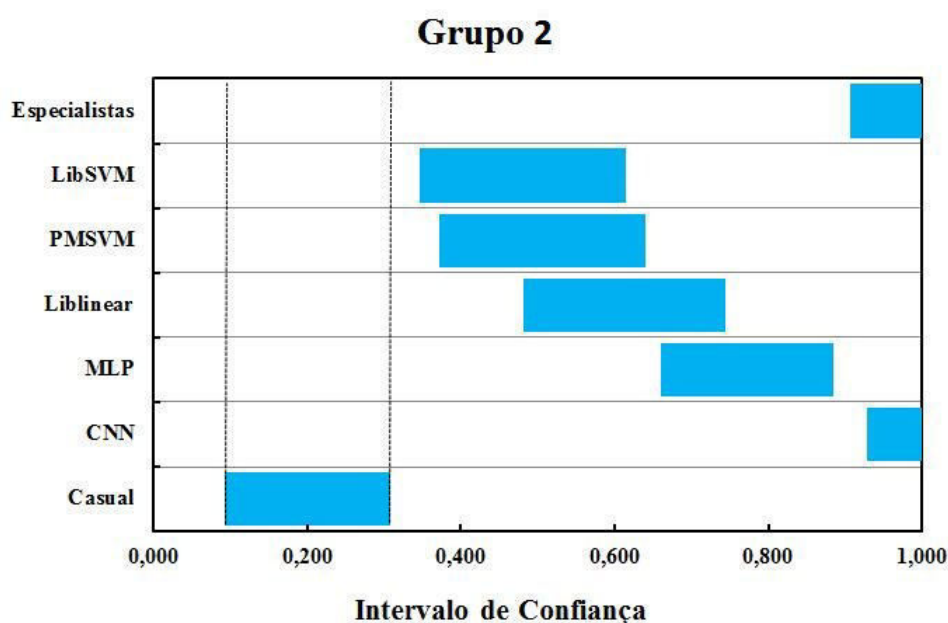


Figura 29: Estimativa por intervalo de confiança ($IC_{1-\alpha}(p)$) com aproximação normal para cada proporção de acerto nas técnicas testadas, a 99% de confiança para o Grupo 2 de imagens. Fonte: Autor.

Já o trabalho de SANNAKKI et al. (2013) para classificar doenças foliares em videiras, mostra uma acurácia de 100% na classificação de duas classes. No entanto, na metodologia apresentada, apenas duas imagens foram utilizadas para teste da rede neural apresentada. Além disso, nos resultados apresentados, os autores falam de um arquivo de entrada da rede contendo 33 imagens, e a seguir, foi apresentada uma matriz de confusão mostrando uma classificação 100% correta de 33 imagens.

As matrizes de confusões da CNN e da Classificação por Especialistas aplicadas no Grupo 2, utilizando o sistema de votação, são apresentados na Tabela 11. Quando comparadas a métrica *Kappa*, é possível visualizar um valor de 0,97 por parte das CNNs e um valor de 0,95 pelo sistema de votação. Nota-se uma pequena melhoria na técnica de CNN, mas ambas são classificadas pela métrica como próximas a uma classificação perfeita.

Analisando as Precisões obtidas, pode-se notar que a classificação utilizando as CNNs, não possui uma classe ao qual os exemplos tendem a ter uma maior predição. Contudo, na classificação utilizando o sistema de votação, nota-se que a Precisão para a classificação da sarna foi de 88,2%, ou seja, existe uma maior tendência aos especialistas a classificar um exemplo como sintoma de sarna.

No entanto, acredita-se que pelo fato da partição de testes ter sido pequena, não foi possível obter conclusões definitivas de quais sintomas têm uma classificação mais simples ou recebem mais predições, tanto para os pesquisadores quanto para as redes, pelo motivo de que ambas as abordagens obtiveram resultados satisfatórios e as

classificações errôneas não ocorreram em um único sintoma.

Quando comparadas as técnicas utilizando a métrica do *Recall*, não é possível destacar em ambas as abordagens, uma classe em que ocorram um maior número de exemplos classificados de forma incorreta.

Estes resultados destacam a importância de se utilizar as métricas *Recall*, *Precisão* e *Kappa* na análise dos resultados. Pois mesmo que a *Precisão* seja de 100%, ou seja, sempre que o classificador realizar a predição de um exemplo como uma classe definida como correta, é também necessária a informação de quantos exemplos foram classificados de forma incorreta, apresentada pela métrica *Recall*. Utilizando as definições de Força de Concordância apresentados na Tabela 4, observa-se que os valores de *Kappa* (Tabela 11) estão entre 0,81 e 1,00, podendo ser classificados como próximos ao perfeito.

		Classificação													
		CNN						Especialistas							
Imagem Rotulada	Sintoma	Glomerella	Herbicida	Def. Magnésio	Def. Potássio	Sarna	<i>Recall</i>	<i>Precisão</i>	Glomerella	Herbicida	Def. Magnésio	Def. Potássio	Sarna	<i>Recall</i>	<i>Precisão</i>
	Glomerella	15	0	0	0	0	100,0%	93,7%	15	0	0	0	0	100,0%	100,0%
	Herbicida	0	15	0	0	0	100,0%	100,0%	0	14	1	0	0	93,3%	100,0%
	Def. Magnésio	0	0	15	0	0	100,0%	100,0%	0	0	14	0	1	93,3%	93,3%
	Def. Potássio	0	0	0	14	1	93,3%	100,0%	0	0	0	14	1	93,3%	100,0%
	Sarna	1	0	0	0	14	93,3%	93,3%	0	0	0	0	15	100,0%	88,2%
Acurácia								97,3%							96,0%
Kappa								0,97							0,95

Tabela 11: Matrizes de confusão para os resultados da classificação por CNN e resultados da classificação por especialistas com o sistema de votação, utilizando o Grupo 2 de imagens. Fonte: Autor.

Porém, um ponto merece destaque, já que as duas abordagens (CNN e classificação por especialistas com o sistema de votação) classificaram erroneamente a mesma imagem, apresentada na Figura 30, que deveria ter sido classificada como deficiência de potássio e, em ambos os casos, foi classificada como sintomas de sarna. Nesta imagem, é possível notar um brilho mais intenso na folha quando houve a captura. Como se acredita que em uma grande escala de imagens é comum a ocorrência de ruídos nas imagens, optou-se por não retirar essa amostra, ou imagens semelhantes do banco de imagens. Contudo, a qualidade das imagens do banco de imagens é fundamental para a eficiência das técnicas empregadas.



Figura 30: Folha com deficiência de Magnésio, classificada pela técnica de CNN e por especialistas como sendo sintoma de sarna. Fonte: Autor.

5.4 Síntese

Nesse Capítulo foram apresentados os resultados obtidos pelas abordagens propostas pelo processo metodológico. Com base nas análises dos resultados foi possível observar que a técnica de CNN, utilizando o banco de imagens proposto, obteve resultados superiores, em relação às outras técnicas utilizadas na tarefa de classificação de imagens.

Os melhores resultados para a classificação de imagens de sintomas em folhas de macieira foram obtidos pela técnica CNNs, que obtiveram percentual de acerto igual ou superior às outras abordagens utilizadas. No Grupo 1 de imagens, esta técnica obteve uma acurácia de 93,3%, valor igual à melhor acurácia obtida pelos membros do grupo especialistas no Grupo 1. Para o Grupo 2 de imagens, esta abordagem obteve uma acurácia de 97,3%, classificando corretamente três imagens a mais do que o especialista com o melhor resultado e uma imagem a mais do que o sistema de votação.

Com base nos experimentos realizados e resultados apresentados, a técnica de CNN obteve acurácia igual ou superior à alcançada por pesquisadores e especialistas na classificação de sintomas causados por doenças, por deficiências nutricionais e por danos de herbicida em folhas de macieira. Mesmo dispondo de um número limitado de imagens, estes resultados mostram a viabilidade da técnica proposta para resolver o problema indicado neste trabalho.

Considerando a eficiência da técnica de CNN na classificação de sintomas causados por doenças, por deficiências nutricionais e por danos de herbicida em folhas de macieira, e possível vislumbrar a extrapolação desta metodologia para outros sintomas em folhas e frutos de macieira, bem como para outras culturas, desde que considerada a necessidade de adequação do banco de imagens. Desta forma, será possível o desenvolvimento de sistema que faça uso dos classificadores treinados para uso em campo.

6 CONSIDERAÇÕES FINAIS

Esta dissertação contextualizou o problema na identificação e classificação de sintomas causados por doenças e distúrbios nutricionais e fisiológicos em folhas de macieira, bem como as dificuldades encontradas atualmente para solucionar este problema, o que justificou a proposição de um sistema para identificação e classificação automática de sintomas utilizando técnicas de aprendizado de máquina.

Para fundamentar a técnica proposta, foi apresentado a história e desenvolvimento das redes neurais desde a criação de seus conceitos iniciais em 1943 até as suas aplicações atuais utilizando CNNs. Foram apresentados os conceitos básicos como unidades de formação das redes, tipos de arquiteturas, tipos de aprendizagem, métodos de treinamento, tarefas e problemas que podem ser solucionados pelas RNAs e CNNs.

Para a realização de testes das abordagens selecionadas para o problema existente, foi necessária a criação de um banco de imagens contendo sintomas presentes nas folhas de macieiras. Estes sintomas foram: dois danos causados por desequilíbrios nutricionais (deficiência de potássio e de magnésio), dois danos por doenças (sarna da macieira - causada pelo fungo *Venturia inaequalis* e mancha de glomerella - causada pelo fungo *Glomerella cingulata*, e um dano causado por toxidez de herbicida (*glifosato*). Este banco de imagens possui um total de 1.995 imagens e está disponível abertamente para ser utilizado em futuras pesquisas e projetos.

Realizados os testes, a abordagem proposta, utilizando CNNs para a resolução do problema de classificação de sintomas em folhas de macieira, obteve resultado satisfatório, com acurácia entre 93,3% e 97,3%. Estas acurácias foram superiores as das outras técnicas de aprendizado de máquina testadas (MLP e SVM) e iguais ou superiores aos resultados obtidos pelos especialistas.

A avaliação da influência do número de imagens nas partições de treinamento nos resultados finais de predição, mostrou que a relação apresenta crescimento logarítmico, aumentando conforme é estendido o número de imagens de cada classe na partição de treinamento. O número de imagens utilizadas neste trabalho (275) para o treinamento das diferentes técnicas está situado na faixa de obtenção de maior acu-

rácia de classificação das imagens utilizadas.

Estes resultados apresentados, comprovam a viabilidade da técnica CNN para resolver o problema de classificação de distúrbios em folhas de macieiras, atingindo assim, o objetivo desta dissertação.

A eficiência dos resultados obtidos nesta dissertação, sugerem que trabalhos futuros expandam consideravelmente o banco de imagens desenvolvido, aprimorando e padronizando ainda mais a metodologia de coleta, detalhando modelos de câmera recomendados, tamanho de folha, luminosidade do ambiente e distância entre a câmera e a folha. Além disso, sugere-se a inclusão de novos sintomas presentes em folhas, bem como a inclusão de sintomas em frutos de macieira, além da adição de folhas e frutos saudáveis no banco, de modo que seja possível realizar a distinção entre saudáveis e com sintomas. Além disso, esta metodologia poderá ser extrapolada para outras culturas que enfrentam problemas similares na identificação de sintomas nas plantas.

De modo a aprimorar os resultados obtidos, sugere-se o desenvolvimento de metodologia para viabilizar o treinamento da CNN utilizando imagens contendo exclusivamente os sintomas, como sugerido por RUMPF et al. (2010) e AL-HIARY et al. (2011). Ou seja, utilizar a imagem somente do ponto contendo o sintoma, ao contrário de apresentar uma imagem contendo a totalidade da folha com sintoma, para o aprendizado da rede. Esta abordagem não foi utilizada neste trabalho porque necessitaria uma maior interação humana para definir as áreas doentes da folha.

Em relação a eficiência das outras técnicas de aprendizado de máquina testadas, com base na análise de trabalhos relacionados, acredita-se que seja possível alcançar acurácias satisfatórias, podendo até ser iguais ou superiores as CNN, desde que ocorra a extração de atributos de forma específica ao problema proposto.

Considerando as expectativa de aprimorar as limitações de hardware atuais, sugere-se testar técnicas de classificação de imagens mais recentes e profundas, de modo que seja possível tratar bancos de imagens maiores e com diferentes classes.

Cabe destacar que para o trabalho desenvolvido nesta dissertação, devido as limitações de hardware, não foi possível realizar testes utilizando a ferramenta GoogLeNet (SZEGEDY et al., 2014), visando melhorias dos resultados. Estas limitações de hardware também se refletiram nos treinamentos das redes, onde o maior *batch* aceito foi de 2. Esta limitação se torna evidente, ao comparar os resultados de MOHANTY; HUGHES; SALATHE (2016), divulgado de forma preliminar em abril de 2016. No trabalho destes autores foi realizada a comparação da técnica de AlexNet (utilizada nessa dissertação), com a GoogLeNet, utilizando banco de imagens PlantVillage com 54.306 imagens de folhas de 14 espécies de plantas e 17 sintomas, divididos em 38 classes (HUGHES; SALATHE, 2016). Nos treinamentos das redes, os autores utilizaram tamanho de *batch* de 24 para o GoogLeNet e 100 para o AlexNet. Para o trei-

namento e teste, possivelmente, utilizaram o *cluster* “Deneb” da EPFL (LAUSANNE, 2016), com 16 nodos de GPU, onde cada nodo possuía 4 placas K40 NVIDIA, obtendo acurácia de 99.35%. Desta forma, os resultados destes autores validam os resultados desta dissertação, em termos de motivações, metodologia e eficiência das técnicas utilizadas.

Por fim, sugere-se desenvolver um sistema compatível com dispositivos *web* e *mobile*, onde indivíduos possam facilmente acessar e enviar imagens para classificação, ou realizar o envio de novas imagens de sintomas e plantas, que serão agregadas ao sistema após a realização de sua análise e rotulagem seguindo os protocolos já definidos.

REFERÊNCIAS

- AL-HIARY, H.; BANI-AHMAD, S.; REYALAT, M.; BRAIK, M.; ALRAHAMNEH, Z. Fast and accurate detection and classification of plant diseases. **Machine learning**, [S.l.], v.14, p.5, 2011.
- ALMEIDA, F. Support Vector Machine. **Universidade Federal de Campina Grande**, [S.l.], 2007.
- ALVES, S. M. **Incubação, multiplicação e identificação de fungos em folhas de macieira**. Embrapa Uva e Vinho: [s.n.], 2015. Dados não publicados.
- BARONE, D. A. C. Projeto Revox. **Versão eletrônica em <http://www.ucs.tche.br/revox>**, [S.l.], 1999.
- BISHOP, C. M. et al. Neural networks for pattern recognition. **Oxford University Press**, [S.l.], 1995.
- BRAGA, A. d. P.; CARVALHO, A.; LUDEMIR, T. B. Fundamentos de redes neurais artificiais. **XI Escola brasileira de computação**, [S.l.], 1998.
- CHANG, C. C.; LIN, C. J. LIBSVM: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, [S.l.], v.2, p.27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- CIRESAN, D. C.; MEIER, U.; GAMBARDELLA, L. M.; SCHMIDHUBER, J. Deep big simple neural nets excel on handwritten digit recognition. **Neural Computation**, [S.l.], p.14, 2010.
- CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2012 IEEE CONFERENCE ON, 2012. **Anais...** [S.l.: s.n.], 2012. p.3642–3649.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, [S.l.], v.20, n.3, p.273–297, 1995.

CRISTIANINI, N.; SHAW-TAYLOR, J. **An introduction to support vector machines and other kernel-based learning methods**. [S.l.]: Cambridge university press, 2000.

DENG, J.; DONG, W.; SOCHER, R.; LI, L. J.; LI, K.; FEI FEI, L. Imagenet: A large-scale hierarchical image database. In: COMPUTER VISION AND PATTERN RECOGNITION, 2009. CVPR 2009. IEEE CONFERENCE ON, 2009. **Anais...** [S.l.: s.n.], 2009. p.248–255.

FIALHO, F. B.; GARRIDO, L. d. R.; BOTTON, M.; MELO, G. W. B. d.; FAJARDO, T. V. M.; NAVES, R. d. L. **Diagnóstico de doenças e pragas na cultura da videira usando o sistema especialista Uzum**. Bento Gonçalves/RS: Embrapa Uva e Vinho, 2012. Accessed: 2016-02-02, <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/73877/1/cot128.pdf>.

FRATE, F. D.; PACIFICI, F.; SCHIAVON, G.; SOLIMINI, C. Use of neural networks for automatic classification from high-resolution images. **Geoscience and Remote Sensing, IEEE Transactions on**, [S.l.], v.45, n.4, p.800–809, 2007.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological cybernetics**, [S.l.], v.36, n.4, p.193–202, 1980.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 2010. **Anais...** [S.l.: s.n.], 2010. p.249–256.

GONÇALVES, A. R. Máquina de Vetores Suporte. **Acesso em**, [S.l.], v.21, 2015.

GUNN, S. R. et al. Support vector machines for classification and regression. **ISIS technical report**, [S.l.], v.14, 1998.

HAYKIN, S. Neural Networks: A Comprehensive Foundation. Printice-Hall. **Inc., New Jersey**, [S.l.], 1999.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. **arXiv preprint arXiv:1406.4729**, [S.l.], 2014.

HEARST, M. A.; DUMAIS, S. T.; OSMAN, E.; PLATT, J.; SCHOLKOPF, B. Support vector machines. **Intelligent Systems and their Applications, IEEE**, [S.l.], v.13, n.4, p.18–28, 1998.

HEBB, D. O. **The organization of behavior**: A neuropsychological approach. [S.l.]: John Wiley & Sons, 1949.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, [S.l.], 2012.

HINTON, G.; OSINDERO, S.; TEH, Y. W. A fast learning algorithm for deep belief nets. **Neural computation**, [S.l.], v.18, n.7, p.1527–1554, 2006.

HOCHREITER, S. Untersuchungen zu dynamischen neuronalen Netzen. **Master's thesis, Institut für Informatik, Technische Universität, München**, [S.l.], 1991.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, [S.l.], v.9, n.8, p.1735–1780, 1997.

HUGHES, D. P.; SALATHE, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. **arXiv preprint arXiv:1511.08060**, [S.l.], 2016.

JURASZEK, G. D.; SILVA, A. G.; SILVA, A. T. **Reconhecimento de Produtos por Imagem Utilizando Palavras Visuais e Redes Neurais Convolucionais**. 2014. Dissertação de Mestrado — UDESC, Florianópolis, SC.

KARNOWSKI, J. **AlexNet + SVM**. Online; accessed 22/03/2016.

KEERTHI, S. S.; LIN, C. J. Asymptotic behaviors of support vector machines with Gaussian kernel. **Neural computation**, [S.l.], v.15, n.7, p.1667–1689, 2003.

KOVALESKI, A. Pragas. In: **NACHTIGALL, G.R. (Ed.) Maçã**.. Produção. Bento Gonçalves: Embrapa Uva e Vinho; Brasília: Embrapa Informação Tecnológica, 2004. p.103–115.

KRIZHEVSKY, A.; HINTON, G. **Learning multiple layers of features from tiny images**. [S.l.]: Citeseer, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS**, 2012. **Anais...** [S.l.: s.n.], 2012. p.1097–1105.

KUSSUL, E.; BAIDYK, T. Rosenblatt perceptrons for handwritten digit recognition. In: **NEURAL NETWORKS, 2001. PROCEEDINGS. IJCNN'01. INTERNATIONAL JOINT CONFERENCE ON**, 2001. **Anais...** [S.l.: s.n.], 2001. v.2, p.1516–1520.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, [S.l.], p.159–174, 1977.

LAROCHELLE, H.; ERHAN, D.; COURVILLE, A.; BERGSTRA, J.; BENGIO, Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: MACHINE LEARNING, 24., 2007. **Proceedings...** [S.l.: s.n.], 2007. p.473–480.

LAUSANNE, É. P. F. **Deneb and its extension, Eltanin**. Online; accessed 27-April-2016, <http://scitas.epfl.ch/hardware/deneb-and-eltanin>.

LAWRENCE, S.; GILES, C. L.; TSOI, A. C.; BACK, A. D. Face recognition: A convolutional neural-network approach. **Neural Networks, IEEE Transactions on**, [S.l.], v.8, n.1, p.98–113, 1997.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. **Neural computation**, [S.l.], v.1, n.4, p.541–551, 1989.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v.86, n.11, p.2278–2324, 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. et al. Convolutional networks and applications in vision. In: ISCAS, 2010. **Anais...** [S.l.: s.n.], 2010. p.253–256.

LI, F.-F.; KARPATY, A. **CS231n: Convolutional Neural Networks for Visual Recognition**. Online; accessed 28-April-2016.

LIN, Y.; LV, F.; ZHU, S.; YANG, M.; COUR, T.; YU, K.; CAO, L.; LI, Z.; TSAI, M.; ZHOU, X. et al. Imagenet classification: fast descriptor coding and large-scale svm training. **Large scale visual recognition challenge**, [S.l.], 2010.

LORENA, A. C.; CARVALHO, A. C. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, [S.l.], v.14, n.2, p.43–67, 2007.

MARTINS-FILHO, L. S.; MOL, A. A.; ROCHA, R. Desenvolvimento de ferramenta computacional para auxílio ao projeto de gemas lapidadas. **Rem: Revista Escola de Minas**, [S.l.], v.58, n.4, p.367–373, 2005.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, [S.l.], v.5, n.4, p.115–133, 1943.

MEDEIROS, H. A. d. Diagnóstico e caracterização de perdas por podridões pós-colheita em maçãs. In: 2015, ENCONTRO NACIONAL SOBRE FRUTICULTURA DE CLIMA TEMPERADO, Fraiburgo. **Anais...** Caçador: Epagri., 2015. v.14, n.11-12, p.48–52.

MOHANTY, S. P.; HUGHES, D.; SALATHE, M. Inference of Plant Diseases from Leaf Images through Deep Learning. **arXiv preprint arXiv:1604.03169**, [S.l.], 2016.

NACHTIGALL, G. R. **Resultados Analíticos de folhas de macieira**. Embrapa Uva e Vinho: [s.n.], 2015. Dados não publicados.

NACHTIGALL, G. R.; BASSO, C.; FREIRE, C. J. S. Nutrição e adubação de pomares. In: **NACHTIGALL, G.R. (Ed.) Maçã**:. Produção. Bento Gonçalves: Embrapa Uva e Vinho; Brasília: Embrapa Informação Tecnológica, 2004. p.66–77.

NACHTIGALL, G. R.; CARRARO, H. R.; ALLEONI, L. R. F. Potassium, Calcium, and Magnesium Distribution in an Oxisol under Long-Term Potassium-Fertilized Apple Orchard. **Communications in soil science and plant analysis**, New York, v.38, n.11-12, p.1439–1449, 2007.

OLIVEIRA-JUNIOR, G. Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado. **Ciência da Computação, Universidade Federal de Pernambuco**, [S.l.], 2010.

PAULA, M. B. Reconhecimento de palavras faladas utilizando redes neurais artificiais. **Monografia de final de curso. UFPEL**, [S.l.], 2000.

REBELO, L. D. T. **Avaliação automática do resultado estético do tratamento conservador do cancro da mama**. 2008. Dissertação de Mestrado — FEUP, Porto, Portugal.

REVATHI, P.; HEMALATHA, M. Cotton Leaf Spot Diseases Detection Utilizing Feature Selection with Skew Divergence Method. **International Journal of Scientific Engineering and Technology (ISSN: 2277-1581)**, [S.l.], v.3, p.22–30, 2014.

REVATHI, P.; HEMALATHA, M. Identification of Cotton Diseases Based on Cross Information Gain_Deep Forward Neural Network Classifier with PSO Feature Selection. **International Journal of Engineering and Technology (IJET) ISSN**, [S.l.], 2014.

RONG, E. F.; CHANG, K. W.; HSIEH, C. J.; WANG, X. R.; LIN, C. J. LIBLINEAR: A Library for Large Linear Classification. **Journal of Machine Learning Research**, [S.l.], v.9, p.1871–1874, 2008.

ROPER, L. D. **The Hyperbolic Tangent World**. Accessed: 2016-04-18, <http://www.roperld.com/science/Mathematics/HyperbolicTangentWorld.htm#references>.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, [S.l.], v.65, n.6, p.386, 1958.

ROUSSEAU, J.; VAN LOCHEM, P.; GISPEN, W.; SPRUIJT, B. Classification of rat behavior with an image-processing method and a neural network. **Behavior Research Methods, Instruments, & Computers**, [S.l.], v.32, n.1, p.63–71, 2000.

RUMPF, T.; MAHLEIN, A. K.; STEINER, U.; OERKE, E. C.; DEHNE, H. W.; PLÜMER, L. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. **Computers and Electronics in Agriculture**, [S.l.], v.74, n.1, p.91–99, 2010.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. Image-Net Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, [S.l.], v.115, n.3, p.211–252, 2015.

SALAKHUTDINOV, R.; MURRAY, I. On the quantitative analysis of deep belief networks. In: MACHINE LEARNING, 25., 2008. **Proceedings...** [S.l.: s.n.], 2008. p.872–879.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, [S.l.], v.3, n.3, p.210–229, 1959.

SANNAKKI, S. S.; RAJPUROHIT, V. S.; NARGUND, V.; KULKARNI, P. Diagnosis and classification of grape leaf diseases using neural networks. In: COMPUTING, COMMUNICATIONS AND NETWORKING TECHNOLOGIES (ICCCNT), 2013 FOURTH INTERNATIONAL CONFERENCE ON, 2013. **Anais...** [S.l.: s.n.], 2013. p.1–5.

SCHMIDHUBER, J. Learning to control fast-weight memories: An alternative to dynamic recurrent networks. **Neural Computation**, [S.l.], v.4, n.1, p.131–139, 1992.

SIMON, P. **Too Big to Ignore: The Business Case for Big Data**. [S.l.]: John Wiley & Sons, 2013. v.72.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, [S.l.], 2014.

SMOLA, A.; BARTLETT, P.; SCHOLKOPF, B.; SCHUURMANS, D. Introduction to large margin classifiers. In: **Advances in Large Margin Classifiers**. [S.l.]: MIT Press, 2000. p.1–28.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. **arXiv preprint arXiv:1409.4842**, [S.l.], 2014.

TAFNER, M. A. **Reconhecimento de palavras isoladas usando redes neurais artificiais**. 1996. Tese (Doutorado em Ciência da Computação) — Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis.

TAFNER, M. A.; XEREZ, M.; RODRIGUES FILHO, I. W. **Redes neurais artificiais: introdução e princípios de neurocomputação**. [S.l.]: Eko, 1995.

TAMURA, S.; KAWAI, H.; MITSUMOTO, H. Male/female identification from 8×6 very low resolution face images by neural network. **Pattern Recognition**, [S.l.], v.29, n.2, p.331–335, 1996.

VALDEBENITO-SANHUEZA, R. M.; MAFFIOLETTI, M.; COMPARIM, C.; KRASNIAK, J.; BOGO, A.; ARCARI, R. **Características e controle da podridão "olho de boi" nas maçãs do Sul do Brasil**. Bento Gonçalves: Embrapa Uva e Vinho, 2006. 13p. (Embrapa Uva e Vinho. Circular Técnica, 66).

VALDEBENITO-SANHUEZA, R. M.; V; BECKER, W.; BONETI, J.; KATSURAYAMA, Y.; CZERMAINSKI, A. B. **Manejo das doenças de verão na produção integrada de maçã**. Bento Gonçalves: Embrapa Uva e Vinho, 2002. 11p. (Embrapa Uva e Vinho. Circular Técnica, 36).

VALDEBENITO-SANHUEZA, R.; NACHTIGALL, G.; KOVALESKI, A.; SANTOS, R. d.; SPOLTI, P. **Manual de identificação e controle de doenças, pragas e desequilíbrios nutricionais da macieira**. Bento Gonçalves: Embrapa Uva e Vinho, 2008. 58p.

WIKIPEDIA. **Lateral inhibition** — **Wikipedia, The Free Encyclopedia**. Online; accessed 28-April-2016.

WIKIPEDIA. **Sigmoid function** — **Wikipedia, The Free Encyclopedia**. Online; accessed 18-April-2016.

WU, J. Power mean SVM for large scale visual classification. In: COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2012 IEEE CONFERENCE ON, 2012. **Anais...** [S.l.: s.n.], 2012. p.2344–2351.

YU, K.; XU, W.; GONG, Y. Deep learning with kernel regularization for visual recognition. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2009. **Anais...** [S.l.: s.n.], 2009. p.1889–1896.

ZEILER, M. D.; FERGUS, R. Stochastic pooling for regularization of deep convolutional neural networks. **arXiv preprint arXiv:1301.3557**, [S.l.], 2013.

ZIEN, A.; RÄTSCH, G.; MIKA, S.; SCHÖLKOPF, B.; LENGAUER, T.; MÜLLER, K.-R.
Engineering support vector machine kernels that recognize translation initiation sites.
Bioinformatics, [S.l.], v.16, n.9, p.799–807, 2000.

[Anexo]

APÊNDICE A IMAGENS UTILIZADAS PARA CRIAÇÃO DE CADA GRUPO

Neste Apêndice estão os *links* para os arquivos contendo quais imagens foram utilizadas para a criação de cada grupo de imagens.

A.1 Grupo 1

Treino: https://www.dropbox.com/s/os1qe2oqvesjo8k/train_group_1.txt.

Teste: https://www.dropbox.com/s/z3xezs4o1l0dp9f/test_group_1.txt.

A.2 Grupo 2

Treino: https://www.dropbox.com/s/2xrme0fy4ven432/train_group_3.txt.

Teste: https://www.dropbox.com/s/strl4wlvmlizjm/test_group_3.txt.

APÊNDICE B ESPECIFICAÇÃO CONFIGURAÇÕES UTILIZADAS PARA TREINO

Neste Apêndice estão os arquivos de configurações de treinamento utilizados pela ferramenta DIGITS®.

B.1 Redes Neurais Convolucionais

```
base_lr: 0.0001
display: 86
max_iter: 137600
lr_policy: "step"
gamma: 0.2
momentum: 0.9
weight_decay: 0.0005
stepsize: 27520
solver_mode: GPU
net: "train_val.prototxt"
solver_type: SGD
```

O arquivo de arquitetura de rede pode ser encontrado em : <https://www.dropbox.com/s/qkyf6brid543go9/CNN.prototxt>.

B.2 Redes *Multilayer Perceptron*

```
base_lr: 0.0001
display: 171
max_iter: 412500
lr_policy: "step"
gamma: 0.2
momentum: 0.9
weight_decay: 0.0005
```

```
stepsize: 82500  
solver_mode: GPU  
net: "train_val.prototxt"  
solver_type: SGD
```

O arquivo de arquitetura de rede pode ser encontrado em : <https://www.dropbox.com/s/0phf7yxwqp3n8sy/MLP.prototxt>.