



UNIVERSIDADE DE UBERABA  
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E EXTENSÃO.  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA  
MESTRADO PROFISSIONAL

LUCIANO RODRIGO FERREIRA

APLICAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NA DETECÇÃO DE  
PRESENÇA DE DOENÇAS NA FOLHA DO CAFEIRO

UBERABA-MG  
2020





LUCIANO RODRIGO FERREIRA

APLICAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NA DETECÇÃO DE  
PRESENÇA DE DOENÇAS NA FOLHA DO CAFEEIRO

Dissertação apresentada como parte dos requisitos para obtenção do Título de Mestre em Engenharia Química do Programa de Pós-Graduação em Engenharia Química – Mestrado Profissional da Universidade de Uberaba (PPGEQ-MP/UNIUBE).

Orientador:  
Prof. Dr. Mauro Luiz Begnini

UBERABA-MG  
2020

Catálogo elaborado pelo Setor de Referência da Biblioteca Central UNIUBE

F413a Ferreira, Luciano Rodrigo.  
Aplicação de técnicas de *machine learning* na detecção de presença de doenças na folha do cafeeiro / Luciano Rodrigo Ferreira. – Uberaba, 2020.  
47 f. : il. color.

Dissertação (mestrado) – Universidade de Uberaba. Programa de Mestrado em Engenharia Química, concentração: Desenvolvimento de Processos Químicos Agroindustriais.

Orientador: Prof. Dr. Mauro Luiz Begnini.

1. Café. 2. Café – Doenças e pragas. 3. Aprendizado de computador. I. Begnini, Mauro Luiz. II. Universidade de Uberaba. Programa de Mestrado em Engenharia Química. III. Título.

CDD 633.73

LUCIANO RODRIGO FERREIRA

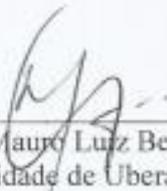
APLICAÇÃO DE TÉCNICAS DE *MACHINE LEARNING* NA DETECÇÃO DE  
PRESENÇA DE DOENÇAS NA FOLHA DO CAFEEIRO

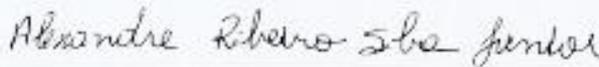
Dissertação apresentada como parte dos requisitos para obtenção do Título de Mestre em Engenharia Química do Programa de Pós-Graduação em Engenharia Química – Mestrado Profissional da Universidade de Uberaba (PPGEQ-MP/UNIUBE).

Área de Concentração: Desenvolvimento de Processos Químicos Agroindustriais

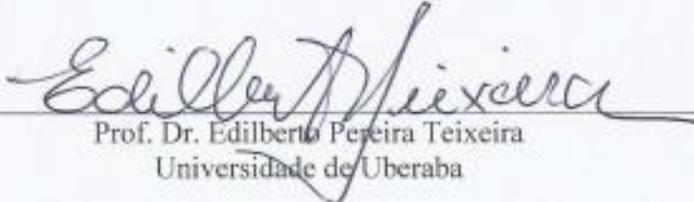
Aprovada em: 28 / 10 / 20.

BANCA EXAMINADORA:

  
\_\_\_\_\_  
Prof. Dr. Mauro Luiz Begnini  
Universidade de Uberaba

  
\_\_\_\_\_  
Prof. Dr. Alexandre Ribeiro Silva Júnior

Instituto Federal do Triângulo Mineiro

  
\_\_\_\_\_  
Prof. Dr. Edilberto Pereira Teixeira  
Universidade de Uberaba



À minha família, Nara, Sofia, Manuela e Luccas pelo companheirismo diário e apoio incondicional em todos os projetos da minha vida, sem vocês nada seria possível e não faria sentido.



"Agradeço aos meus amigos e professores que sempre me incentivaram a nunca desistir, em especial ao Prof. Dr. José Roberto Delalibera Finzer e o Prof. Dr. Mauro Luiz Begnini, pelo incentivo na conclusão deste mestrado".



“Não faz sentido olhar para trás e pensar: devia ter feito isso ou aquilo, devia ter estado lá. Isso não importa. Vamos inventar o amanhã, e parar de nos preocupar com o passado”.

*“Steve Jobs”*



## Resumo

Atualmente a tecnologia vem ganhando espaço em todas as atividades humanas, no trabalho, no lazer, na educação e de forma mais rápida que no passado. Como a tecnologia está presente em todos os setores, não seria diferente na agricultura e neste campo ela tem se mostrado uma ferramenta indispensável no apoio à tomada de decisão. O monitoramento agrícola tem proporcionado aos agrônomos, fazendeiros, técnicos agrícolas e demais profissionais, uma análise da lavoura de forma mais precisa, possibilitando a identificação antecipada de possíveis problemas, proporcionando iniciativas preventivas antes que um problema possa afetar uma área de cultivo. Não distante disso, a análise de imagens, através do estudo de padrões, pode ajudar a identificar possíveis doenças na lavoura que posteriormente poderão ser combatidas desde o início, evitando assim a propagação de uma praga. Com a intenção de colaborar na detecção de doenças nas folhas das árvores do café surgiu este trabalho, “Detecção de doenças na folha do café através de técnicas de *Machine Learning*”. Diante de um universo de possibilidades de linguagens de programação que se apresentam aptas para o desenvolvimento de soluções capazes de realizar estudos de padrões de imagens, foi escolhido o software MatLab. A escolha do MatLab se deve à sua constante aplicação em casos científicos que necessitam executar uma quantidade expressiva de cálculos e simulações. O MatLab é uma das tecnologias utilizadas para operações matemáticas que envolvam matrizes e álgebra linear. Ele também fornece suporte para projetos e testes de algoritmos mais rápidos e melhores. Isso facilita a análise de dados e o estudo das mudanças de comportamento. O MatLab trata imagens como uma matriz de pixels, o que facilita a análise de comportamento e as comparações entre as imagens. Como referência, foram armazenadas imagens de folhas de café saudáveis e também folhas de café que apresentam doenças, essas duas categorias já seriam suficientes para apresentar um resultado de semelhança durante a análise de outra folha qualquer, porém, pode ocorrer que a imagem a ser analisada não tenha nenhuma ligação com folhas de café, então, uma terceira categoria de imagens foi criada e denominada “outras”. As imagens são armazenadas em pastas distintas e para o teste foram analisadas 80 imagens em cada grupo, caso algum grupo tenha menos imagens que o outro o programa levará em conta a pasta com menos imagens e selecionará a mesma quantidade nas demais pastas para realizar o estudo. Uma nova imagem que não conste nos grupos já armazenados como banco de referências é apresentada ao algoritmo e então começa um processo que analisa todas as imagens saudáveis entre elas, todas as imagens com sinais de doença entre elas e todas as imagens diversas também entre elas. Diante das comparações entre cada imagem do grupo com as demais do próprio grupo é possível chegar à matriz de confusão, que apresenta o percentual de semelhança entre as imagens quando comparadas com cada grupo. Após a utilização da função bag-of-feature um histograma é gerado conforme o vocabulário de características analisadas pela função, neste caso, foram 1000 (mil) características por imagem. A imagem a ser comparada e que já foi formada no início do algoritmo é identificada como semelhante a um dos grupos do banco de imagens, sendo classificada através da função: `categoryClassifier=trainImageCategoryClassifier(trainingSets, bag)`; utilizando o treinamento realizado por outra função: `evaluate(categoryClassifier, trainingSets)`; Após esse processo é apresentada a classificação da imagem em questão, como doente, saudável ou outra. A análise permite identificar as semelhanças entre as imagens e, portanto, pode ser utilizada para distinguir diferentes doenças não só na folha do café, mas em outros tipos de cultivo. Mediante identificação será possível indicar qual a química, técnica agrícola ou outra solução profissional que seja ideal para o tratamento da planta.

Palavras-chaves: café, *Machine Learning*, MatLab, doenças, vocabulário, banco, imagens



## Abstract

Currently, technology has been gaining ground in all human activities, at work, in the laser, in education and more quickly than in the past. As technology is present in all sectors, it would be no different in agriculture and in this field it has proved to be an indispensable tool in supporting decision making. Agricultural monitoring has provided agronomists, farmers, agricultural technicians and other professionals with a more accurate analysis of the crop, enabling early identification of possible problems, providing preventive initiatives before a problem can affect a crop area. Not far from this, the analysis of images, through the study of patterns, can help to identify possible diseases in the crop that can later be combated from the beginning, thus preventing the spread of a plague. With the intention of collaborating in the detection of diseases in the leaves of the coffee trees, this work appeared, "Detection of diseases in the coffee leaf through *Machine Learning* techniques". Faced with a universe of possibilities for programming languages that are able to develop solutions capable of carrying out studies of image patterns, the MatLab software was chosen. The choice of MatLab is due to its constant application in scientific cases that need to perform a significant amount of calculations and simulations. MatLab is one of the technologies used for mathematical operations involving matrices and linear algebra. It also provides support for faster and better algorithm design and testing. This facilitates data analysis and the study of behavioral changes. MatLab treats images as a matrix of pixels, which facilitates behavior analysis and comparisons between images. As a reference, images of healthy coffee leaves and also coffee leaves that present diseases were stored, these two categories would be enough to present a similarity result during the analysis of any other leaf, however, it may happen that the image to be analyzed has no connection with coffee leaves, so a third category of images was created and called "other". The images are stored in different folders and for the test 80 images were analyzed in each group, if any group has fewer images than the other, the program will take into account the folder with fewer images and select the same amount in the other folders to carry out the study. A new image that is not included in the groups already stored as a reference bank is presented to the algorithm and then begins a process that analyzes all healthy images between them, all images with signs of disease between them and all the different images also between them. In view of the comparisons between each image of the group with the others of the group itself, it is possible to arrive at the confusion matrix, which shows the percentage of similarity between the images when compared with each group. After using the bag-of-feature function, a histogram is generated according to the vocabulary of characteristics analyzed by the function, in this case, there were 1000 (thousand) characteristics per image. The image to be compared and which was already informed at the beginning of the algorithm is identified as similar to one of the image bank groups, being classified through the function: `categoryClassifier = trainImageCategoryClassifier(trainingSets, bag)`; using training performed by another function: `evaluate(categoryClassifier, trainingSets)`; After this process, the classification of the image in question is presented, as sick, healthy or otherwise. The analysis allows to identify the similarities between the images and, therefore, can be used to distinguish different diseases not only in the coffee leaf, but in other types of cultivation. Through identification it will be possible to indicate which chemical, agricultural technique or other professional solution is ideal for treating the plant.

**Keywords:** coffee, machine learning, MatLab, diseases, vocabulary, bank, images.



## Lista de Figura

Figura 1: Consórcio Antiferrugem .....	7
Figura 2: Detector de Pragas .....	8
Figura 3: FMC Agrícola.....	8
Figura 4: FOXBayer.....	9
Figura 5: ADAMA .....	9
Figura 6: Folha com ferrugem.....	10
Figura 7: Cercosporiose na folha e fruto do cafeeiro .....	11
Figura 8: Mancha Manteigosa na folha do cafeeiro .....	11
Figura 9: Queima do Fio do cafeeiro .....	12
Figura 10: Roseliniose no cafeeiro.....	12
Figura 11: Nematóide das galhas do cafeeiro .....	13
Figura 12: Exemplo de Histograma .....	16
Figura 13: Divisões da Inteligência Artificial .....	18
Figura 14 : Modelo de um neurônio artificial .....	19
Figura 15 : Aprendizado Supervisionado.....	21
Figura 16 : Aprendizado Não Supervisionado .....	22
Figura 17 : Classificação através de Machine Learning.....	23
Figura 18: TOOLBOX com APPS.....	24
Figura 19: Tela inicial do APP Neural Pattern recognition.....	25
Figura 20: Galeria de imagens de folhas saudáveis .....	28
Figura 21 : Matriz de uma imagem no MatLab.....	29
Figura 22: Barra de Aplicativos MatLab - Software Image Batch Processor .....	30
Figura 23 : Descrição das pastas de imagens .....	34
Figura 24 : Quantidade de imagens por pasta .....	35
Figura 25 : Exibição de uma imagem do banco .....	35
Figura 26 : Execução da função bagOfFeatures.....	36
Figura 27 : Resultado da classe bagOfFeatures.....	36
Figura 28 : Classificação de categorias .....	38
Figura 29 : Criação da Matriz de confusão .....	38
Figura 30 : Imagem de uma folha doente avaliada pelo algoritmo .....	39
Figura 31 : Resultado final do Algoritmo .....	40



## **Lista de Gráficos**

Gráfico 1: Projeção do consumo de café no Brasil .....	03
Gráfico 2: Produção de café no Brasil.....	04
Gráfico 3: Cotações de mercado internacional.....	05



## **Lista de abreviações, siglas e símbolos.**

MatLab	Software de alta performance para cálculos com matrizes
ML	<i>Machine Learning</i> – aprendizado de máquina
Pixels	Menor elemento em um dispositivo de exibição
IBGE	Instituto Brasileiro de Geografia e Estatística
Conab	Companhia Nacional de Abastecimento
Cecafé	Conselho dos Exportadores de café
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
IA	Inteligência Artificial
iOS	Sistema Operacional da Apple
Android	Sistema Operacional cujo principal desenvolvedor é o Google
UFMS	Universidade Federal do Mato Grosso do Sul
RNA's	Redes Neurais Artificiais
ABIC	Associação Brasileira da Indústria de Café
BoW	Bag-Of-Word
BoF	Bag-Of-Feature



## SUMÁRIO

1. Introdução .....	1
1.1 O Mercado cafeeiro e suas tecnologias .....	1
1.2 O estudo do café.....	3
1.3 Café Arábica e Robusta (conilon) .....	4
2. Revisão de Literatura .....	6
2.1 Consórcio Antiferrugem.....	7
2.2 Detector de Pragas.....	7
2.3 FMC Agrícola .....	8
2.4 FOXBayer .....	8
2.5 ADAMA Alvo.....	9
2.2. Principais doenças na cafeicultura .....	10
2.3. Processamento de Imagens.....	14
2.4. Histograma .....	15
2.5. Reconhecimento de Padrões.....	16
2.6. Inteligência Artificial e Redes Neurais .....	17
2.7. Neurônios de uma Rede Neural .....	18
2.8. Algoritmos de Aprendizado .....	20
2.8.1. Aprendizado Supervisionado .....	20
2.8.2. Aprendizado Não Supervisionado.....	21
2.9. Relação entre redes neurais e reconhecimento de padrões.....	22
2.10. <i>Machine Learning</i> .....	23
2.11. O Software MatLab.....	24
3. Objetivos do Trabalho.....	25
3.1. Objetivos Gerais.....	25
3.2. Objetivos Específicos .....	25
4. Justificativa do Trabalho .....	26
5. Metodologia .....	26
5.1. Método Bag of Words.....	27
5.2. Método Bag of Features .....	27
5.3. Desenvolvimento do Código Fonte.....	30
6. Resultados e Discussões.....	31
7. Considerações Finais.....	40
8 - Propostas Futuras .....	41
9. Referências.....	42



## 1. Introdução

### 1.1 O Mercado cafeeiro e suas tecnologias

A revolução tecnológica reflete-se em todo o sistema produtivo. Diante de uma nova postura do mercado e a expressiva redução nos custos de informatização, o setor primário brasileiro mostra-se aberto à revolução da informação, da mesma maneira que há anos ocorreu com o setor industrial e de serviços (CEOLIN et al., 2008).

Os empresários agrícolas, mais especificamente os cafeicultores, devem intensificar suas ações para a adoção de novas práticas de gestão, incorporando em seus processos a utilização da tecnologia de informação (TI), pois ela possibilita, dentre outros benefícios, prover as informações necessárias à tomada de decisão. Sendo o café um produto historicamente importante no agronegócio brasileiro, é de interesse estudar a adoção de TI por parte dos empresários do setor cafeeiro (SANTOS e JESUS, 2005).

Atualmente, parte dos solos das áreas em produção brasileira pode ser classificada como “frágil”, devido à sua baixa aptidão agrícola ou de capacidade de uso e elevado potencial de degradação, em função, por exemplo, da elevada erosão ou de outros fenômenos decorrentes de sua instabilidade diante do uso e manejo. Portanto, explorar a discussão acerca de manejos agrícolas mais adequados e alternativas de uso desses solos, de forma a potencializar a sua capacidade de Prestação de Serviços Ambientais (PSA), torna-se estratégica para o país e para o atendimento do debate em torno de uma agricultura sustentável (TURETTA et al., 2017).

A cultura do café pode ser acometida por várias patologias desde o viveiro, ou seja, fase de incubação da muda de café, até o campo, onde o pé de café pode estar em fase de desenvolvimento e/ou produção. Tais doenças contribuem para a quebra na produção e na qualidade do fruto, que por consequência, produzirá um produto ruim e de difícil aceitação no mercado, gerando prejuízo ao produtor (CARVALHO, 2012).

A agricultura de precisão trata-se de um “gerenciamento de cada parâmetro em um local específico para reduzir o desperdício, aumentar os lucros e manter a qualidade do meio ambiente”. (ADAMCHUK, PERK E SCHERPERS, 2003).

A agricultura de precisão surge como um conjunto de tecnologias que são capazes de auxiliar o produtor rural a identificar as estratégias de manejo a serem adotadas para aumentar a eficiência no gerenciamento do processo de produção, podendo maximizar a rentabilidade das colheitas e reduzir os custos de aplicação de insumos, tornando a atividade mais competitiva (CARVALHO et al., 2006; OLIVEIRA et al., 2007; SILVA; SOUZA; ALVES, 2008)

Neste sentido, vem se destacando o desenvolvimento de aplicativos que colaboram com a tomada de decisão, muitos deles, fazendo uso de redes neurais computacionais, que utilizam técnicas de aprendizado e o estudo de comportamento e similaridades nas mais diversas situações, ou seja, algoritmos capazes de analisar dados e tomar decisões baseadas em experiências anteriores.

As aplicações do algoritmo de máxima verossimilhança (Maxver) são bastante conhecidas na literatura de sensoriamento remoto, estando este algoritmo presente em vários sistemas de softwares de processamento de imagens; e frequentemente, é usado como rotina de classificação padrão contra o qual outros algoritmos de classificação são comparados. Este

algoritmo identifica a forma, o tamanho e a orientação das amostras de treinamento (VIEIRA, 2000).

As tecnologias só fazem sentido se aplicadas com objetivo de proporcionar um benefício, independentemente de qualquer prerrogativa, os avanços tecnológicos devem contribuir para melhorar as condições de vida das pessoas e o meio em que elas vivem.

O último Censo Agropecuário do Instituto Brasileiro de Geografia e Estatística (IBGE) mostrou que 1,5 milhão de produtores rurais acessam dados por meio de dispositivos eletrônicos, número 1.900% superior ao de 10 anos atrás, o que revela boa adesão às soluções digitais. “A digitalização de processos, entretanto, demanda infraestrutura telecomunicações na área rural, ainda reduzida no Brasil. É nosso calcanhar de aquiles” (ZAPAROLLI, 2020).

No Brasil, é possível notar que o cultivo do café é uma tradição, em algumas regiões se destaca mais que em outras, mas de fato, existem muitas lavouras em potencial que com ajuda tecnológica podem impulsionar a produção aumentando a quantidade de grãos colhidos, dentre os diversos recursos tecnológicos possíveis o acompanhamento da detecção e prevenção de pragas e doenças pode ser fundamental.

Segundo o Conselho dos Exportadores de Café – CECAFÉ, no ano de 2018 as exportações de café no Brasil alcançaram a marca de 35.533.000 sacas, o que representa um crescimento de 14,90% em relação à 2017.

É preciso ressaltar que 89,5% foram de café verde o que corresponde à 31.815.000 sacas, das quais 29.335.000 são de café arábica e 2.480.000 da espécie conilo, já o restante, cerca de 11,5% (3.718.000 sacas) foram de café industrializado.

Em relação ao volume da produção brasileira, o ranking dos seis maiores estados produtores em 2018, em ordem decrescente, é o seguinte: Minas Gerais, em primeiro lugar, com 30,7 milhões de sacas, corresponde a 53% da produção; Espírito Santo, em segundo, com 12,81 milhões de sacas (22%); São Paulo, em seguida, com 6,07 milhões de sacas (10%); Bahia, 4,50 milhões de sacas (8%), Rondônia, 2,19 milhões de sacas (4%) e Paraná, sexto Estado produtor, com 2%, produz 1,05 milhão de sacas (EMBRAPA, 2018).

Os números expressivos em relação à produção brasileira de café são sem dúvida alguma a justificativa para investimentos que possam certificar a qualidade de toda essa produção, ou ainda, alavancar tal produção uma vez que já se tem consolidado o potencial agrícola do café na região.

Diante da expressividade do café como um dos produtos mais importantes na agricultura brasileira, torna-se óbvio que ele seja utilizado como objeto de estudo das principais tecnologias em desenvolvimento no país.

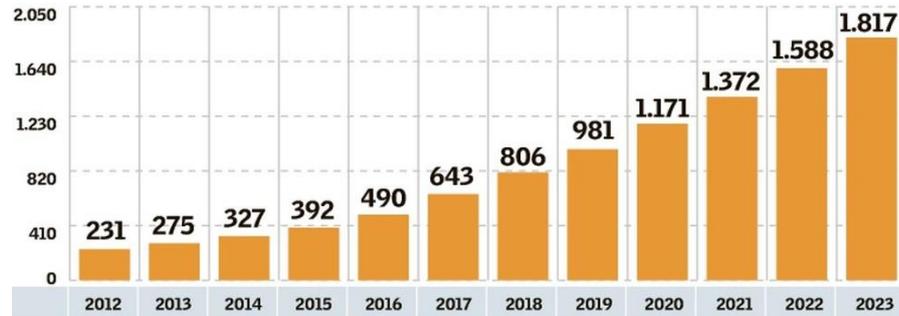
A expectativa do setor é que nos próximos anos o consumo do café aumente significativamente no mercado interno, pois, ao longo dos anos é o que tem acontecido, conforme demonstrado no Gráfico 1.

Gráfico 1 : Projeção do consumo de café no Brasil

## Cafés especiais no Brasil

Projeções para a evolução do consumo no país

■ Em milhares de sacas de 60 kg



Fonte: Rabobank

Fonte: <https://valor.globo.com/agronegocios/noticia/2020/01/10/consumo-de-cafe-vai-crescer-no-pais.ghtml>

Para abastecer e conseguir atender ao mercado e necessário evitar perdas na produção, já considerando o transporte, o processo de armazenamento e principalmente as pragas durante o cultivo.

### 1.2 O estudo do café

O cultivo de café no Brasil teve início no século XVIII, mas apenas no século seguinte a produção foi grande o bastante para torná-lo o principal produto da economia do país. O café foi o principal produto de exportação da economia brasileira durante o século XIX e o início do século XX, garantindo as divisas necessárias à sustentação do Império do Brasil e também da República Velha. (PINTO, 2020).

Atualmente, O Brasil é o maior exportador de café no mercado mundial e ocupa a segunda posição, entre os países consumidores da bebida. O Brasil responde por um terço da produção mundial de café, o que o coloca como maior produtor mundial, posto que detém há mais de 150 anos. Conforme dados do Ministério da Agricultura, Pecuária e Abastecimento (Mapa), a cafeicultura brasileira é uma das mais exigentes do mundo, em relação às questões sociais e ambientais, e há uma preocupação em se garantir a produção de um café sustentável. A atividade cafeeira é desenvolvida com base em rígidas legislações trabalhistas e ambientais. São leis que respeitam a biodiversidade e todas as pessoas envolvidas na cafeicultura, e que punem, rigorosamente, qualquer tipo de trabalho escravo e/ou infantil nas lavouras. As leis brasileiras estão entre as mais rigorosas entre os países produtores de café. (ABIC, 2020).

No Brasil a grande extensão territorial permitiu que a cultura do café se multiplicasse por diferentes regiões, relevos, climas e demais características peculiares de cada uma dessas regiões, por este motivo, o país produz variados tipos de grãos, propiciando atender diferenciados tipos de paladar e também de preços ao consumidor.

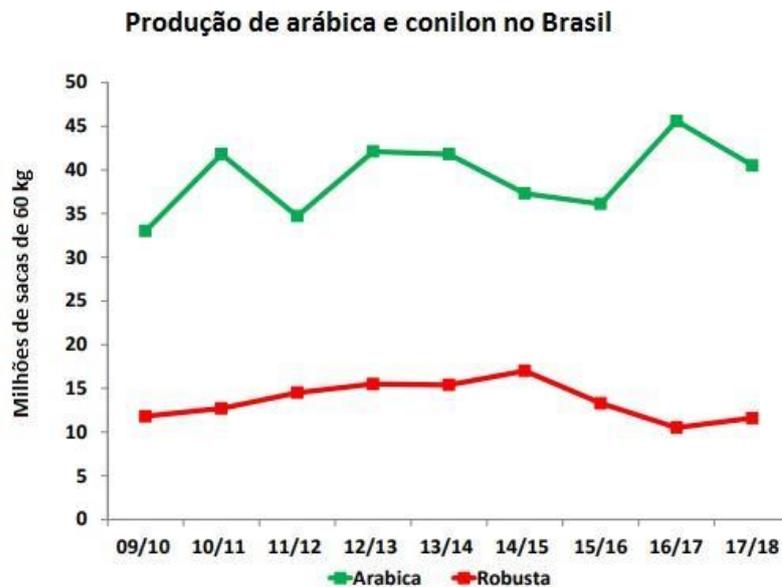
Tanta diversidade possibilitou o que chamamos de “Blends” (mistura de tipos), tendo como base o café de terreiro ou natural, o descascado, despulpado, os ácidos, encorpados e até os

aromáticos e especiais, sendo que, as duas principais espécies plantadas são, o arábica com aproximadamente 80% da área e o tipo Robusta (conilon) com o restante.

Atualmente, Minas Gerais concentra a maior produção de café do Brasil. Todo o estado corresponde a 50% da produção do café arábica e conta ainda com polos importantes de produção de cafés especiais. Ainda no Sudeste, o café Conilon ou Robusta como também é conhecido, tem forte produção no Espírito Santo e parte do Rio de Janeiro. Outra região de relevância na produção do Conilon, é Rondônia, na parte norte do país. Rondônia, inclusive, investiu em tecnologia e pesquisa e triplicou a produção de 2020, além de recuperar lavouras consideradas perdidas em 2016 (ALVES, 2020).

Estes dois tipos dominam basicamente o mercado e estão em expansão em relação a produtividade, o que tem atraído investidores para os negócios que envolvem o café, conforme demonstra o Gráfico 2.

Gráfico 2: Produção de café no Brasil



Fonte: Grão Gourmet

### 1.3 Café Arábica e Robusta (conilon)

O café do tipo Arábica é caracterizado pela produção de cafés mais finos e requintados. São plantados em altitudes superiores a 800 metros que são ideais para este tipo de produção de qualidade superior, possui aroma intenso, sabores variados, mais encorpados e com certa acidez.

O tipo Arábica possui frutos ovais e necessita de mais cuidados no manejo, enquanto a Robusta, possui frutos menores e valor inferior no mercado, além de possuir quase o dobro de cafeína em relação ao tipo Arábica (GRAOGOURMET, 2020).

Com um maior controle do aproveitamento na produção, através da utilização de ferramentas que possam colaborar com redução de perdas e prevenção de pragas, o ganho final pode ser ainda mais rentável para os produtores exportadores de café.

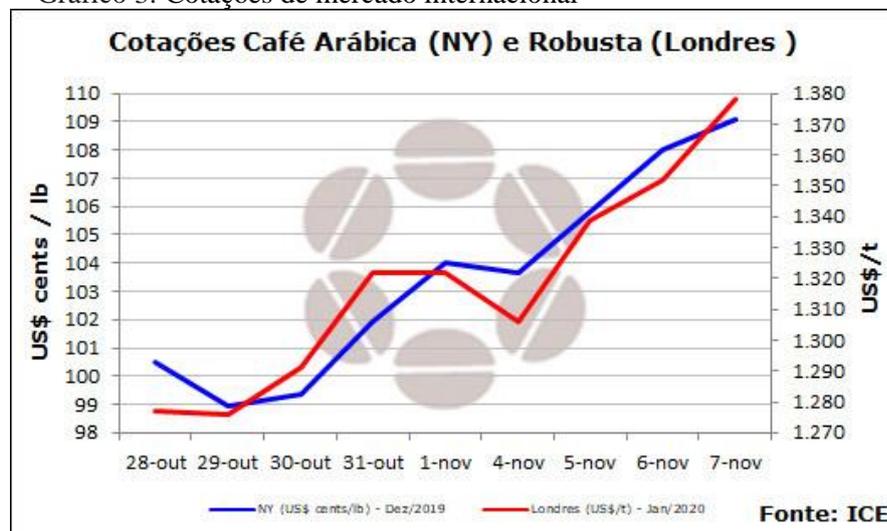
Mesmo o Robusta florescendo mais de uma vez ao ano, crescendo mais rapidamente e tendo uma resistência maior contra os parasitas, o Arábica continua correspondendo a 3/4 da produção mundial, devido a qualidade de aromas e a sua maior doçura. Dependendo da variedade de arábica de acordo com o *terroir* da fazenda onde é produzido, podemos conseguir notas muito variadas, como as notas floral e mel do Catiguá, as notas de castanhas e caramelo do Bourbon Amarelo, e o próprio *blend* Octavio com notas de caramelo e amêndoas (OCTAVIO, 2020).

A primeira estimativa do Instituto Brasileiro de Geografia e Estatística (IBGE) para a safra de 2020 é de 56,4 milhões de sacas de 60 kg, representando um aumento de 12,9% em relação à última safra. Os números de produção do arábica estão em 42,2 milhões de sacas de 60 kg, com crescimento de 22,1% também em relação ao mesmo período anterior (COOPERCAM, 2020).

Minas Gerais se destaca nacionalmente mais uma vez como maior produtor de café arábica, com 75% da produção nacional, seguido pelo estado do Espírito Santo que espera produzir 33,4% a mais que ano passado.

A exportação do café tem se mostrado em crescimento e o valor comercial está em expansão, tanto para o tipo Arábica, quanto para o Robusta, é o que indica as cotações de preço em Nova York e em Londres conforme pode-se observar no Gráfico 3.

Gráfico 3: Cotações de mercado internacional



Fonte: Safraes

Entre as doenças presentes na cafeicultura está a cercosporiose, causada pelo fungo "*Cercospora coffeicola*", uma das principais e mais antigas doenças conhecidas do cafeeiro. Pode ocorrer tanto no viveiro quanto no campo (CHAVES, SILVA, POZZA, 2014).

A nutrição mineral pode promover modificações morfológicas ou químicas que aumentam ou reduzem a resistência das plantas, tais modificações podem ocorrer nas paredes das células ou então na composição química dos compostos fenólicos.

Várias ferramentas tecnológicas propõem o acompanhamento do avanço destas doenças e principalmente a sua detecção, logo no início, o que possibilita uma tomada de decisão preventiva, dentre estas ferramentas e técnicas destaca-se o uso da Inteligência Artificial, mais especificamente a *Machine Learning*.

*Machine Learning* é uma tecnologia onde os computadores têm a capacidade de aprender de acordo com as respostas esperadas por meio associações de diferentes dados, os quais podem ser imagens, números e tudo que essa tecnologia possa identificar. *Machine Learning* é o termo em inglês para a tecnologia conhecida no Brasil como aprendizado de máquina (IBM, 2020).

A técnica *Machine Learning* é formada por um método de análise de dados que automatiza a construção de modelos analíticos, trata-se de um ramo da inteligência artificial, baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana.

A utilização de redes neurais já foi testada por Mizubuti et al. (1994) para descrever a epidemia de ferrugem do feijoeiro (*Uromyces appendiculatus*) e também por Pozza (1998) em identificação de epidemias de vassoura-de-bruxa (*Crinipelis pernicioso*) no cacauzeiro.

Assim, considerando os resultados alcançados em diversos patossistemas e a importância do diagnóstico de doenças na folha do café, o presente trabalho tem como objetivo avaliar as redes neurais através de técnicas de *Machine Learning* como método potencial para identificar doenças no cafeeiro.

## 2. Revisão de Literatura

A tecnologia está implícita em todo produto, serviço e procedimento produzidos ou utilizados, porém, a mudança na tecnologia que afeta o processo e/ou o produto e serviços, quando adotada, caracteriza-se como inovação (MATTOS; GUIMARÃES, 2005)

A agropecuária não deve mais ser considerada atrasada e pobre tecnologicamente, mas sim uma atividade moderna e altamente dinâmica. De um lado, as propriedades vinculadas às modernas cadeias produtivas presentes no Brasil podem ser caracterizadas pela sua gestão eficaz e no uso de tecnologias; por outro lado, há um grande conjunto de propriedades que apresenta ineficiência na gestão e carência de tecnologias (GARCIA; VIEIRA FILHO, 2014).

No caso da cafeicultura a necessidade de monitoramento é constante, muitas técnicas e projetos já apresentam soluções que tem colaborado com o agricultor na tomada de decisão, e a proposta de um software de monitoramento vem ganhando espaço através de tecnologias como o estudo de imagens por satélite (georreferenciamento) e estudos químicos relacionados ao solo.

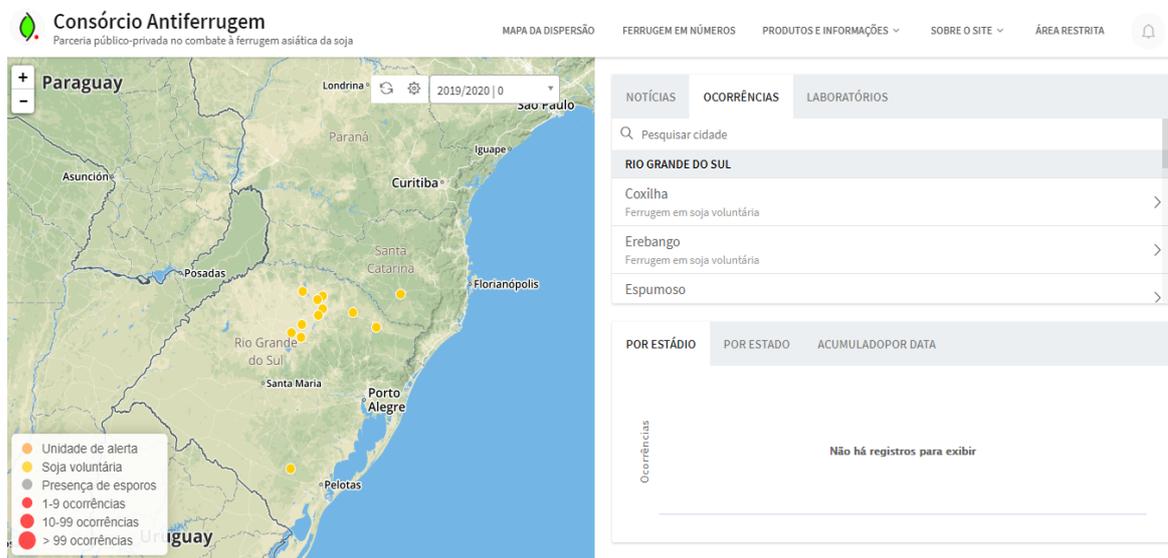
O comércio desenvolvedor de sistemas de softwares é munido de tecnologias voltadas ao agronegócio. Há sistemas para a agricultura de precisão, nas quais permitem realizar simulações espaciais em todas as áreas de gerenciamento predefinidas, calibrar os modelos para simular histórico de variabilidade espacial (THORP et al., 2008).

Dentre as diversas propostas de uso de softwares que colaboram com o diagnóstico de problemas enfrentados pelas plantações alguns estão se tornando conhecidos, mas é necessário avaliar as peculiaridades de cada um deles e se eles são capazes de identificar os reais problemas conforme cada tipo de cultura. Seguem alguns exemplos:

## 2.1 Consórcio Antiferrugem

O website Consórcio Antiferrugem permite o acesso facilitado as informações e ao monitoramento da dispersão da ferrugem-asiática da soja no Brasil. Através da plataforma web e aplicativos móveis disponíveis para iOS e Android, produtores e técnicos de todo país podem acompanhar o avanço desta importante doença no decorrer das safras através de ferramentas simples e intuitivas, conforme a Figura 1. (SOJA, 2020)

Figura 1: Consórcio Antiferrugem



Fonte: [www.consorcioantiferrugem.net](http://www.consorcioantiferrugem.net)

## 2.2 Detector de Pragas

O Software de Detecção de Pragas é um projeto inovador para pesquisa e aplicação em agronegócio, coordenado pelo Prof. Robson Soares Silva da Fábrica de Software da UFMS de Ponta Porã, e desenvolvido pelo aluno de Ciência da Computação Willians Magalhães Primo, tem por objetivo indicar a possibilidade da existência de determinada praga e as principais características da mesma. Assim o produtor deve procurar um especialista para ver se a suspeita é verdadeira ou não, pois podem existir situações em que determinadas imagens ainda não consigam ser detectadas totalmente pelo aplicativo (APKPURE, 2016), conforme observado na Figura 2.

Figura 2: Detector de Pragas



Fonte: <https://apkpure.com>

### 2.3 FMC Agrícola

Criado pela FMC Agricultural Products, o aplicativo compara imagens de acervo para identificação de plantas daninhas, pragas e insetos. Também oferece informações agronômicas. Disponível para iOS e Android (FMC, 2019), conforme a Figura 3.

Figura 3: FMC Agrícola



Fonte: <https://www.fmcagricola.com.br>

### 2.4 FOXBayer

Foi criado para identificar pragas nas lavouras de soja, trigo, feijão e algodão, a partir de um banco de fotos com imagens em alta resolução. O aplicativo desenvolvido pela multinacional alemã Bayer, fornece ainda recomendações para o tratamento. Disponível para iOS e Android, (AGROBAYER, 2019) conforme a Figura 4.

Figura 4: FOXBayer

Serviços agronômicos

SOFTWARES PARA O AGRONEGÓCIO

SOFTWARES AGRO

Laboratórios de Análise de Fibras de Algodão

AD - Monitoramento através de Agrodigitalização

globalfarm  
AGRIBUSINESS INTELLIGENCE  
www.globalfarm.com.br

Agribusiness Intelligence - GlobalFarm

Assessoria no gerenciamento e no levantamento de custos de produção, planejamento de safra, planejamento de

AQILA

A INTEGRAÇÃO ENTRE A GESTÃO E A OPERAÇÃO DA SUA LAVOURA

Com o plataforma AQILA, o agricultor pode monitorar as práticas labores que afetam o resultado de sua região.

AGRO1

Aqila – Aplicativo de Monitoramento de Lavouras

Implantação do Aplicativo Aqila para controle e organização das operações que envolvem a condução das lavouras e o

Estação Meteorológica "Cultivo Inteligente" - AgroSmart

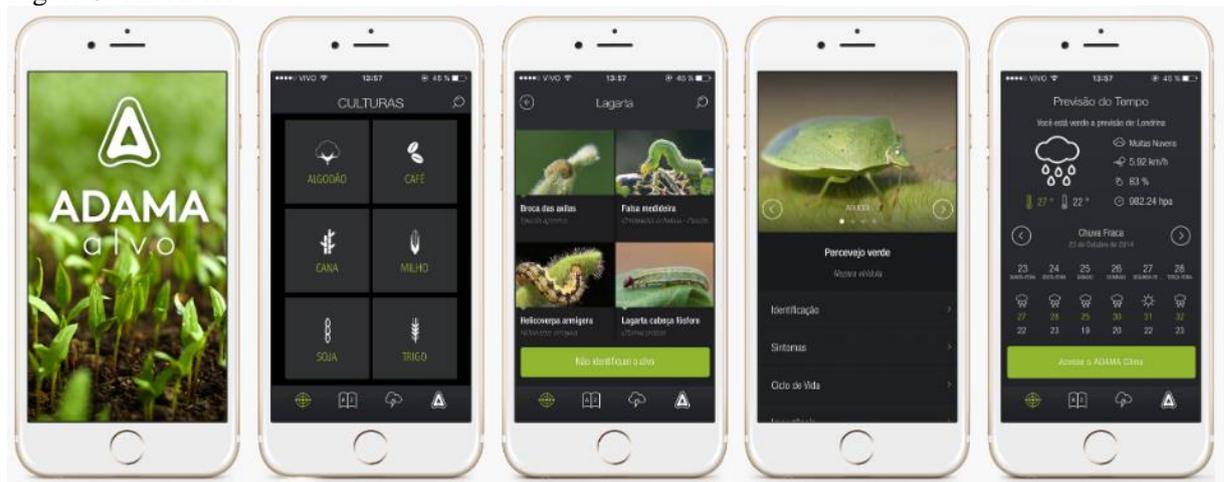
Estação Meteorológica com conectividade satélite e licença de uso "Cultivo Inteligente".

Fonte: <https://www.redeagroservices.com.br>

## 2.5 ADAMA Alvo

Ferramenta desenvolvida pela empresa Adama é capaz de identificar pragas a partir da comparação com imagens de um banco de dados, que ajudam no tratamento de doenças na soja, trigo, café, milho, algodão e cana-de-açúcar, (ADAMA, 2020) conforme a Figura 5.

Figura 5: ADAMA



Fonte: <https://www.adama.com/>

Devido às inúmeras possibilidades que o agronegócio dispõe em relação à produtos tecnológicos que possam contribuir de forma abrangente, principalmente considerando as plantações e lavouras de soja, milho, algodão, e vários outros, torna-se necessário focar em um produto para que se possa iniciar o processo de desenvolvimento de um software, analisar as principais doenças e começar classificando um banco de dados de imagens.

## 2.2. Principais doenças na cafeicultura

**2.2.1 Ferrugem:** Uma doença muito comum que acomete os cafezais é a ferrugem, causada pelo fungo *Hemileia Vastatrix*. A ferrugem figura como uma das mais severas doenças que atacam os cafeeiros (ZAMBOLIM, 1985). Ela provoca queda das folhas e seca dos ramos de forma precoce, o que prejudica diretamente a produção da planta no ano seguinte. As perdas ocasionadas pela presença da ferrugem são estimadas em valores de 35%, no mínimo, a 50% nos piores casos quando há ocorrência de estiagem prolongada, na face inferior das folhas observam-se pústulas de cor amarelo alaranjada. Ocorre principalmente em condições de campo, conforme pode ser notada na Figura 6.

Figura 6: Folha com ferrugem



Fonte: <http://www.sapc.embrapa.br>

**2.2.2. Cercosporiose:** A Cercosporiose é causada por um fungo que comumente ataca as folhas e frutos em desenvolvimento dos cafeeiros. Esta doença é responsável por causar lesões necróticas arredondadas com coloração castanho claro, ou escuro, e com centro acinzentado envolto por um halo de cor amarela. As folhas atacadas caem rapidamente e ocasionam a secagem de ramos inteiros, o que implica na redução da produção na planta (SOUZA, et al., 2013).

As lesões são circulares, concêntricas, centro escuro com bordas de cores, variando do pardo-claro ao marrom claro até o marrom-escuro. Halo clorótico ao redor das lesões, pode ocorrer tanto no campo como em viveiro tanto em folha quanto em frutos, conforme Figura 7.

Figura 7: Cercosporiose na folha e fruto do cafeeiro



Fonte: <http://www.sapc.embrapa.br>

**2.2.3. Mancha manteigosa nas folhas do cafeeiro:** Nos últimos anos, a mancha manteigosa tem revelado um agravante na sintomatologia, no campo, observa-se um forte definhamento com grande número de ramos mamortos, provocando declínio vegetativo e produtivo (FERREIRA et al., 2005).

Nas folhas e frutos surgem manchas de cor clara de aspecto oleoso que depois se transformam em lesões necróticas, conforme Figura 8.

Figura 8: Mancha Manteigosa na folha do cafeeiro



Fonte: <http://www.sapc.embrapa.br>

**2.2.4. Queima do fio do cafeeiro:** Em estágios mais avançados da doença, as folhas secas desprendem-se do ramo, ficando penduradas ou suspensas por filamento branco do micélio, formado por uma agregação de hifas (Matiello, 1991; Godoy et al., 1997). Observe a Figura 9.

Figura 9: Queima do Fio do cafeeiro



Fonte: <http://www.sapc.embrapa.br>

**2.2.5. Roseliniose no cafeeiro:** Comumente observada em cafezais instalados em áreas previamente ocupadas por florestas, as quais não foram corretamente destocadas, a doença é causada pelo fungo *Rosellinia bunodes* (Berk. & Br.)

Trata-se de uma doença do sistema radicular da planta, causando escurecimento das raízes e fácil desprendimento da casca (CARVALHO; CHAULFON, 2000; MARCOLAN et al., 2009; ZAMBOLIM et al., 2009).

Os sintomas iniciais são amarelecimento e murchamento das plantas, queda das folhas e morte dos ramos. As raízes ficam escuras e a casca se desprende, ocorre normalmente após os 4 anos de idade da planta, conforme Figura 10.

Figura 10: Roseliniose no cafeeiro



Fonte: <http://www.sapc.embrapa.br>

**2.2.6. Nematóide das galhas do cafeeiro:** Nematoides do gênero *Meloidogyne* são tidos como os mais importantes da agricultura por serem patógenos do tipo cosmopolitas, possuírem ampla gama de hospedeiros e causarem prejuízos da ordem de milhões de dólares anualmente no mundo (FREITAS et al., 1999). Relatos de diferentes partes do mundo têm mostrado a importância do nematoide-das-galhas na agricultura. São conhecidas mais de 500 espécies de plantas em mais de 50 famílias botânicas as quais o nematoide das galhas é capaz de causar danos (AGRIOS, 2005).

A planta fica nanica, desfolhada e clorótica. Nas raízes, veem-se em profusão, galhas de cor marrom claro a marrom escuro, conforme Figura 11.

Figura 11: Nematóide das galhas do cafeeiro



Fonte: <http://www.sapc.embrapa.br>

A Tabela 1 a seguir demonstra as principais doenças do cafeeiro e a medida de controle a ser adotada.

Tabela 1 : Principais doenças do cafeeiro e seu controle

<b>Doença</b>	<b>Medida de Controle</b>
Ferrugem	Controle químico (cúpricos e triazóis); Variedades resistentes
Mancha mantelgosa	Controle químico (cúpricos; clorotalonil, triazóis) Uso de quebra-ventos; limpeza de restos culturais
Queima do fio	Eliminação das partes atacadas; Controle químico (fungicidas cúpricos)
Roseliniose	Eliminação de tocos e restos de árvores Arranquio e queima de planta atacadas
	<b>No viveiro:</b>
	Uso de sementes sadias e/ou certificadas; Uso de substrato Tratado Uso de nematicidas na sacola
	<b>No campo:</b>
Nematóide das galhas	Eliminação de plantas doentes; Uso de nematicidas na cova de plantio; Uso de variedades resistentes (ex: "Apostã")
	<b>No viveiro</b>
	Controle de irrigação Adubação equilibrada; Uso de fungicidas cúpricos; Cuidados na aclimação; evitar excesso de sol;
Cercosporiose	<b>No campo</b>
	Pulverização preventiva com fungicidas cúpricos Cuidado na adubação nitrogenada

Fonte: <http://www.sapc.embrapa.br>

### 2.3. Processamento de Imagens

Atualmente, diversos campos da atividade humana têm recorrido a imagens para armazenar, exibir e fornecer informações sobre o mundo que nos rodeia. O desafio que fica para os pesquisadores, engenheiros e empresários, é extrair rapidamente informação valiosa a partir de dados brutos de imagens. Este é o objetivo primário do processamento de imagens, ou seja, converter imagens em informação (RSI, 2005).

O processamento de imagens consiste na aplicação de técnicas que possam representar uma determinada imagem com o objetivo de extrair mais facilmente informações nela presente, esse processo busca representar a imagem de modo que o resultado final seja adequado para uma aplicação concreta e específica.

De forma geral, funções de processamento que incluam reconhecimento e interpretação estão associadas com aplicações de análise de imagens nas quais o objetivo é a extração automática ou mesmo parcialmente automática de informação a partir de uma imagem (GONZALES e WOODS, 1987).

Dentre os diversos recursos existentes o mais comum é a construção de um histograma, que é uma representação de características de uma imagem apresentado através de gráficos de frequência com o objetivo de ilustrar como uma determinada amostra de dados está distribuída.

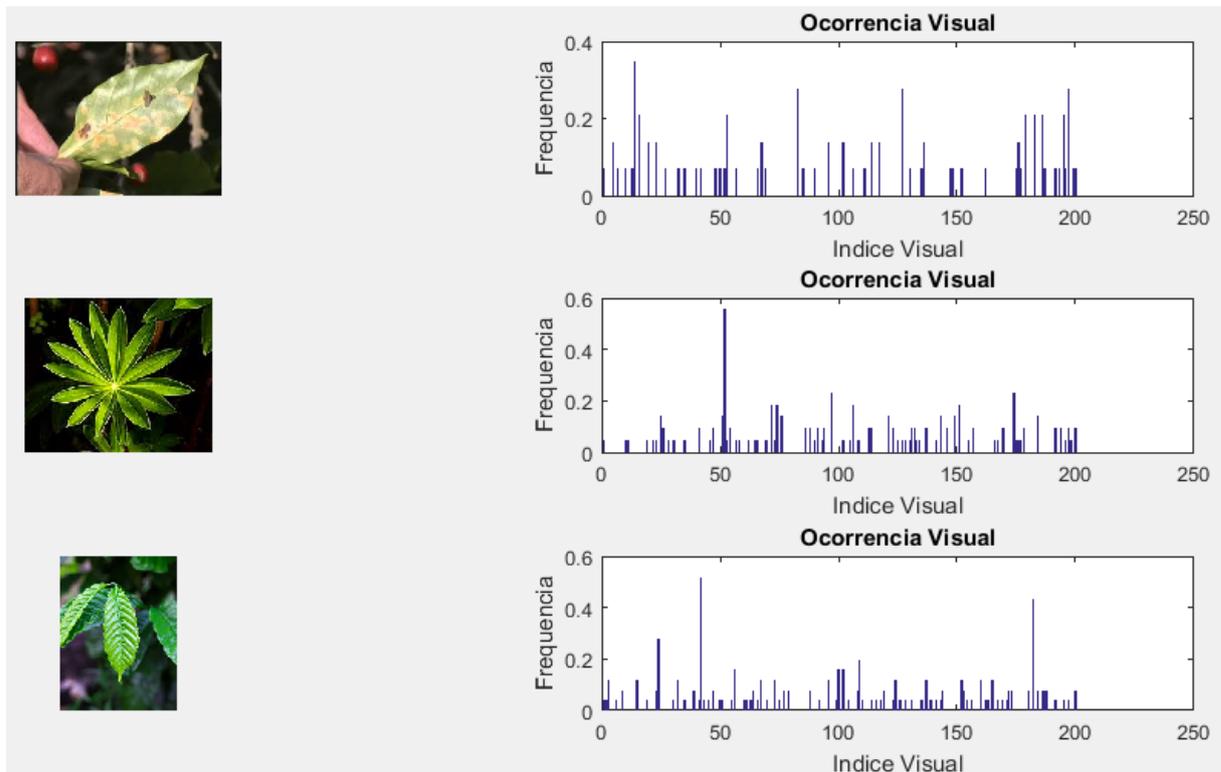
O Histograma é um gráfico de colunas utilizado na estatística. É composto de vários retângulos adjacentes, representando a tabela de frequências com perda de informações (valores agrupados por classes) de um conjunto de valores. Na escala horizontal, marcam-se os intervalos de classes, e cada intervalo é a base de cada retângulo ou barra; na escala vertical, marcam-se as alturas dos retângulos ou barras, que são as respectivas frequências absolutas das classes (LOPES, 1999).

#### **2.4. Histograma**

Histograma é uma ferramenta utilizada para a análise de estatísticas, tanto qualitativa como quantitativa, representada em um gráfico que mostra a frequência de acontecimentos dentro de um espectro pré-determinado e que pode ser classificado. Os dados evidenciados e classificados dentro de um histograma são chamados de amostras e elas expressam um comportamento (MAPPA, 2020)

O histograma busca representar através de um gráfico, características relacionadas a uma imagem, quanto mais características analisadas, mais detalhado será o histograma, o processo acontece após uma conversão para intensidades de cor cinza, sendo que, para cada nível de intensidade (cinza), o número de pixels com aquele nível é identificado, também é analisado se a imagem se encaixa dentro dos possíveis níveis e se a mesma não está utilizando todos os níveis disponíveis. Conforme Figura 12.

Figura 12: Exemplo de Histograma



Fonte: acervo do autor

## 2.5. Reconhecimento de Padrões

Segundo PAO (1989), o conhecimento de reconhecimento de padrões é importante devido às ocorrências na vida humana tomarem forma de padrões. A formação da linguagem, o modo de falar, o desenho das figuras, o entendimento das imagens, tudo envolve padrões.

O que pode parecer simples para os seres humanos, na verdade é um grande desafio para a tecnologia, o reconhecimento de padrões pode acontecer de forma praticamente instantânea quando observado por um humano ou até por animais, mas tecnicamente não é tão simples quando envolve sistemas computacionais.

O termo “Reconhecimento de Padrões” surgiu nos anos 60, e a princípio com a finalidade de manipular formas simples, mas foi em 1973 que a definição dada por Duda e Hart foi considerada a mais precisa sobre o termo.

“Campo interessado no reconhecimento de regularidades significativas em ambientes ruidosos ou complexo, feito por máquinas, ou a procura por uma estrutura de dados” (DUDA e HART, 1973).

As motivações de estudos de reconhecimento de padrões estão geralmente ligadas às necessidades de uma interface de comunicação entre pessoas e máquinas, através de linguagens naturais e a construção de máquinas inteligentes que possam realizar pequenas tarefas como se fossem humanos.

Deve-se levar em conta que as pessoas aplicam seus próprios métodos de reconhecimento de padrões em praticamente todas as áreas da atividade humana. Na prática computacional, pode-se enumerar algumas áreas onde vem sendo usado Reconhecimento de Padrões (KASABOV, 1996; SMITH, 1998; TAFNER, 1998):

- Análise, segmentação e pré-processamento de imagens;
- Reconhecimento de faces;
- Identificação de impressões digitais;
- Reconhecimento de caracteres;
- Análise de manuscritos;
- Computer Vision;
- Entendimento e reconhecimento de voz;
- Diagnóstico médico.

Em muitos casos em que o reconhecimento de padrão é utilizado o objetivo em si pode ser considerado apenas uma ocorrência exata de um padrão de comportamento, já em outras situações, basta que a ocorrência seja aproximada.

No desenvolvimento deste projeto a simples aproximação do reconhecimento de padrão já pode ser considerada suficiente para tomada de decisão, porém, a porcentagem relacionada a essa aproximação deve estar o mais próximo possível dos 100% para aceitação.

Obviamente não podemos considerar uma porcentagem abaixo de 50% pois assim a semelhança de uma imagem com um grupo seria menor que o restante de opções para comparação, mas, atingir os 100% também seria para imagem exata, assim, é preciso considerar um parâmetro acima de 80% para creditar a semelhança de uma imagem com um grupo de outras imagens.

## 2.6. Inteligência Artificial e Redes Neurais

Em 1955, John McCarthy, professor do “*Dartmouth College*” criou o termo inteligência artificial e passou a organizar eventos, conferências, seminários, sobre o tema, sendo pioneiro nessa área.

A inteligência artificial é uma das ciências mais recentes, teve início após a Segunda Guerra Mundial e, atualmente, abrange uma enorme variedade de subcampos, desde áreas de uso geral, como aprendizado e percepção, até tarefas específicas como jogos de xadrez, demonstração de teoremas matemáticos, criação de poesia e diagnóstico de doenças. A inteligência artificial sistematiza e automatiza tarefas intelectuais e, portanto, é potencialmente relevante para qualquer esfera da atividade intelectual humana. Nesse sentido, ela é um campo universal (RUSSELL; NORVIG, 2004).

Rosenblatt, em 1958, já descreveu um modelo computacional probabilístico para descrever a organização e o estoque de informação no cérebro humano, o qual batizou de “*Perceptron*”, que, na verdade, era um modelo de rede neural artificial.

Seguindo nesse caminho as inovações tecnológicas que envolviam inteligência artificial acabaram por ser responsáveis em grandes momentos de desenvolvimento da humanidade, principalmente na revolução industrial, permitindo avanços em diversos produtos, como a máquina a vapor, a eletricidade, o motor de combustão, a computação e a robótica.

Atualmente a inteligência artificial tem se mostrado com um dos recursos tecnológicos mais importantes do momento, principalmente por estar aliada a técnicas de “*Machine Learning*”, que permite um sistema computacional ser melhorado em relação ao seu desempenho toda vez que o mesmo é executado, ou seja, “aprender” sem a interferência humana.

A Inteligência artificial é a área da ciência da computação focada na construção de máquinas e computadores capazes de simular comportamento inteligente, tornando-os capazes de fazer tarefas associadas à inteligência humana, como reconhecimento de voz, tomada de decisão, percepção visual e tradução de idiomas (SIEGEL et al., 2018).

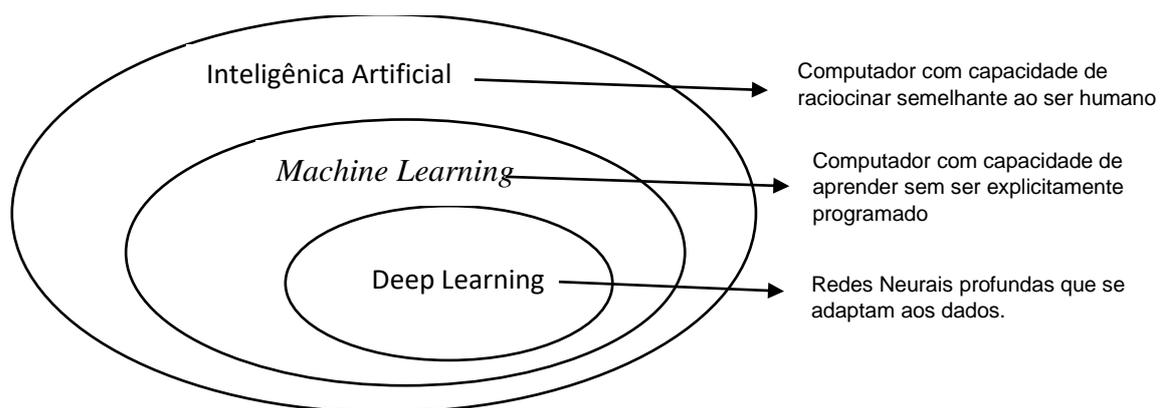
Em 1950, Alan Turing criou procedimentos para detectar se uma máquina conseguiria se passar por um humano em uma conversa por escrito, esse procedimento ficou conhecido como teste de Turing.

Marvin Minsky em 1951, criou uma calculadora de operações matemáticas que era capaz de simular ligações entre neurônios, o que motivou o governo americano e entidades privadas dos Estados Unidos a investirem na área de inteligência artificial.

A Inteligência Artificial ainda é tabu dependendo do assunto abordado, ainda não se sabe se o homem vai ser capaz de criar a real inteligência artificial, ou ao menos desvendar os princípios do cérebro humano que é a base sua criação. Hoje, o que se sabe é que seus conceitos desenvolvidos ao longo de anos têm trazido grandes benefícios para humanidade e que de um modo geral ela sempre vai inovar e evoluir gradativamente (STAIRS; REYNOLDS, 2006).

As divisões correspondentes às áreas da Inteligência Artificial, podem ser observadas abaixo na Figura 13.

Figura 13: Divisões da Inteligência Artificial



Fonte: acervo do autor

## 2.7. Neurônios de uma Rede Neural

As redes neurais artificiais se diferenciam pela sua arquitetura e pela forma como os pesos associados às conexões são ajustados durante o processo de aprendizado. A arquitetura de uma

rede neural restringe o tipo de problema no qual a rede poderá ser utilizada, e é definida pelo número de camadas (*camada única* ou *múltiplas camadas*), pelo número de nós em cada camada, pelo tipo de conexão entre os nós (*feedforward* ou *feedback*) e por sua topologia (HAYKIN, 2001).

As Redes Neurais Artificiais (RNA's), são assim definidas, por se assemelharem com as redes neurais do sistema nervoso biológico, em termos de estrutura complexidade e interligações, que são realizadas pelos neurônios.

A habilidade do ser humano em realizar funções complexas e principalmente a sua capacidade de aprender advém do processamento paralelo e distribuído da rede de neurônios do cérebro. Os neurônios do córtex, a camada externa do cérebro, são responsáveis pelo processamento cognitivo. Um novo conhecimento ou uma nova experiência pode levar a alterações estruturais no cérebro. Tais alterações são efetivadas por meio de um rearranjo das redes de neurônios, reforçando ou inibindo algumas sinapses (HAYKIN, 2001).

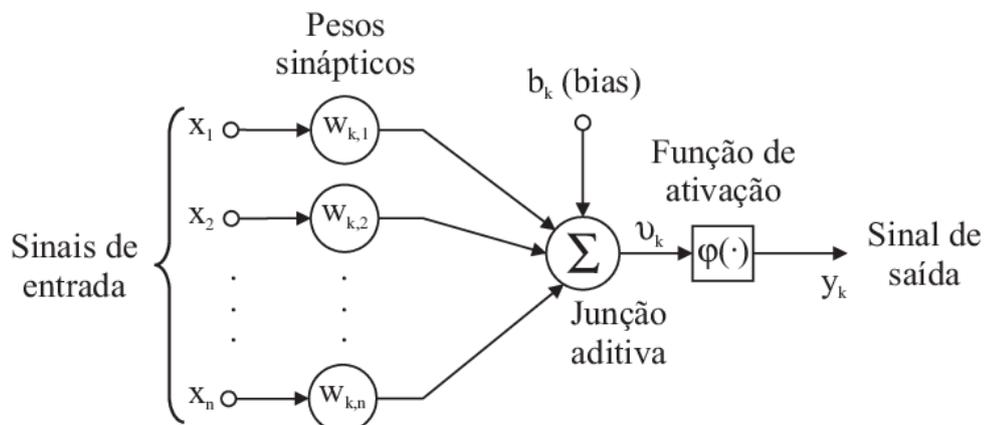
No caso das Redes Neurais Artificiais os neurônios são elementos de processamento simples, que possuem a capacidade de realizar operações como cálculos em paralelo, para processar dados e representar o conhecimento.

Redes neurais são sistemas computacionais, de hardware ou software, que imitam as habilidades computacionais dos sistemas biológicos (sistema nervoso) usando um grande número de simples neurônios artificiais interconectados. (MAREN et al,1990).

Desde que as redes neurais foram consolidadas como técnica de pesquisa para tomada de decisão, várias topologias surgiram com a finalidade de resolver problemas diferentes, como por exemplo, o processamento de linguagem natural, reconhecimento de fala e de imagens e a previsão de valores.

Um neurônio segue uma estrutura onde, dadas as entradas binárias combinadas por uma soma ponderada através da atribuição de pesos, gera-se uma entrada mais efetiva. Observe a estrutura de um neurônio artificial na Figura 14.

Figura 14 : Modelo de um neurônio artificial



Fonte: acervo do autor

## **2.8. Algoritmos de Aprendizado**

O conjunto de instruções definidos para adaptar parâmetros de uma rede neural de forma que a mesma possa aprender uma determinada função é conhecido como “Algoritmo de aprendizado”, sendo que, existe uma grande diversidade de algoritmos capazes de realizar aprendizado e treinamento, se distinguindo basicamente em dois tipos: supervisionado e não supervisionado.

Diferentes algoritmos de aprendizado de máquina podem ser aplicados em uma mesma amostra de dados e alguns desses algoritmos podem apresentar melhor desempenho que outros. Entretanto, para um dado domínio, não existem garantias que um determinado algoritmo seja necessariamente melhor que outro, ou seja, não existe uma análise matemática capaz de determinar qual algoritmo terá um melhor desempenho em relação aos outros (KOHAVI, SOMMERFIELD & DOUGHERTY 1997).

O Aprendizado de Máquina visto de uma maneira prática pode ser entendido como um conjunto de métodos, técnicas e ferramentas próprias para a aquisição automatizada de conhecimento a partir de conjuntos de dados (ROCHA et al., 1997)

Redes neurais artificiais são capazes de melhorar uma experiência ou resultado a medida em que a mesma se ajusta a novos valores de entrada, ou seja, possuem a capacidade de aprender a medida que novos exemplos são executados.

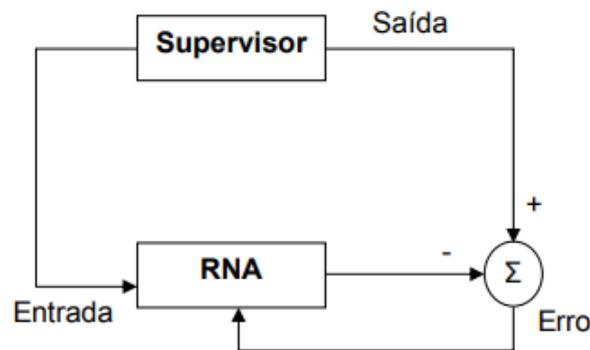
Conforme LORENA e CARVALHO (2007), as técnicas de aprendizado de máquinas empregam um princípio de inferência denominado indução, no qual é possível obter conclusões genéricas a partir de um conjunto particular de exemplos. Estas técnicas de aprendizados indutivos podem ser divididas em dois principais tipos: os supervisionados e os não supervisionados.

### **2.8.1. Aprendizado Supervisionado**

Um algoritmo é executado sobre um conjunto expressivo de amostras e um modelo de classificação é definido de tal maneira que seja possível classificar dados de entrada desconhecidos a partir desse modelo (BROWNLEE, 2016).

Neste modelo de aprendizado a entrada e saída desejada para a rede são fornecidas por um supervisor externo, buscando sempre o ajuste dos parâmetros da rede, até encontrar uma ligação entre os pares de entrada e saída fornecidos anteriormente, conforme a Figura 15.

Figura 15 : Aprendizado Supervisionado



Fonte: acervo do autor

A função do supervisor é indicar o comportamento bom ou ruim para a rede, objetivando o direcionamento do processo de treinamento, em seguida a rede tem sua saída calculada e comparada com o resultado desejado, caso ainda esteja com erros, os pesos das conexões são reajustados a fim de minimizar a diferença na próxima execução até atingir a solução desejada ou se aproximar o máximo possível dela.

O aprendizado de máquina é composto de algoritmos orientados por dados, ou seja, são os dados de entrada que informam qual é a resposta ideal.

De acordo com MITCHELL (1997), o erro de um aprendizado supervisionado pode ser calculado como a diferença entre a saída desejada e a saída gerada.

Neste projeto o algoritmo irá classificar em um grupo as imagens consideradas “doentes”, em outro grupo as imagens consideradas “saudáveis” e por fim as imagens chamadas de “outras”, uma vez que o treinamento acontece de forma recursiva buscando semelhanças entre as imagens do próprio grupo, e ao apresentar uma imagem de entrada ao sistema o mesmo consegue identificar suas características como mais semelhantes a um dos grupos já treinados.

### 2.8.2. Aprendizado Não Supervisionado

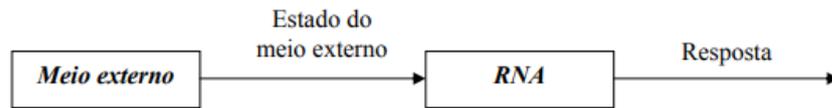
Diferentemente do aprendizado supervisionado, no não supervisionado não se utiliza referências, ou seja, não ocorre um treinamento com o conhecimento do ambiente. LORENA e CARVALHO (2007).

Sem a presença de um supervisor para acompanhar o desempenho do aprendizado, o que realmente acontece nessa técnica é que somente os padrões de entrada são informados para rede e a partir do momento em que a rede estabelece uma harmonia com as regularidades da entrada de dados, a mesma desenvolve uma habilidade de formar representações para codificar características dos dados de entrada, identificando novas classe e grupos.

De acordo com SOUTO et al. (2003), as técnicas de aprendizado não supervisionado são mais utilizadas quando o entendimento dos dados é feito através de padrões ou tendências.

Este tipo de aprendizado só é possível quando existe redundância nos dados de entrada, caso contrário não seria possível encontrar padrões ou características dos dados de entrada, conforme demonstrado na Figura 16

Figura 16 : Aprendizado Não Supervisionado



Fonte: acervo do autor

A situação ideal para utilização do aprendizado não supervisionado é quando você não tem todos os dados sobre os resultados esperados, comparando com o projeto em questão, seria o mesmo que não classificar as imagens anteriormente em grupos e sim deixar que o algoritmo decida e forme os grupos analisando cada uma das imagens de todo banco.

No método não supervisionado, nenhuma informação é passada ao algoritmo anteriormente, obviamente que o mesmo não pode inventar, mas sim, analisar e separar em grupos, ou seja, para cada imagem do banco o mesmo iria realizar o processo de análise e depois agrupá-la com uma outra similar e assim por diante com as próximas imagens que vão sendo analisadas.

## 2.9. Relação entre redes neurais e reconhecimento de padrões

O reconhecimento de padrões é por natureza uma técnica não exata, pois alguns padrões podem ser classificados conforme sua estrutura, já outros, não seguem uma estrutura pré-definida, estando refém da natureza dos dados disponíveis e sua aplicação concreta.

Segundo FUKUSHIMA (1988), não há soluções gerais para o reconhecimento, a escolha é influenciada pela natureza dos dados disponíveis, da aplicação e da tecnologia.

A possibilidade da generalização em uma rede neural é uma das vantagens quando associada ao reconhecimento de padrões, no caso em questão, como as folhas de uma árvore é possível imaginar a grande variedade de formas que os diversos tipos de imagens de folhas podem conter, a inclinação, o tamanho, a luminosidade e a cor, são algumas das variáveis que podem afetar o reconhecimento do objeto.

Uma rede neural é capaz de aprender alguns tipos de folhas, dos mais variados tamanhos, das mais variadas cores e então distinguir sempre que for testada, se uma imagem é mesmo de uma folha.

De certa forma, as redes neurais podem ser consideradas um método de modelagem altamente recomendável para trabalhar com sistemas complexos, que não podem ser descritos adequadamente por um conjunto de regras ou equações exatas, mas que são capazes de identificar padrões quando os mesmos existem.

## 2.10. Machine Learning

Segundo KOHAVI e PROVOST (1998), o *Machine Learning* explora a construção e o estudo de algoritmos que podem aprender e fazer previsão de dados. O ML veio para contribuir de maneira significativa na análise de dados e no aperfeiçoamento dos resultados, afinal, quando a organização dá atenção à inteligência organizacional esperam-se ótimos resultados focando decisões assertivas para alavancar o negócio.

Para WITTEN & FRANK (2005, p. 29), "*machine learning* tem preocupação em formular modelos que podem ser generalizados, com intuito de responder questões a partir de possíveis hipóteses."

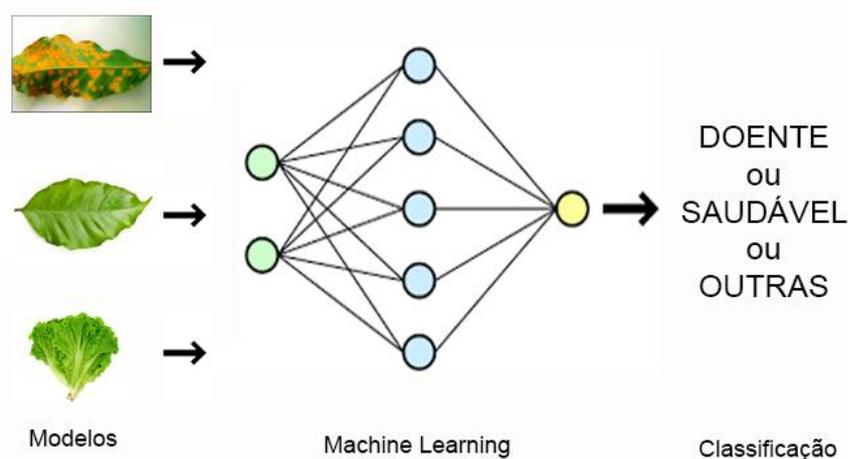
Na prática a "*Machine Learning*" é uma área da computação derivada da Inteligência Artificial que utiliza dados para produzir um modelo que pode ser aplicado em uma determinada tarefa, seja na indústria, segurança, agropecuária, medicina ou robótica, tais aplicações estão avançando diariamente principalmente aliando soluções que fazem uso de processamento de imagens.

O *Machine Learning* é um sistema capaz de adquirir e acumular conhecimento e, assim, melhorar a performance em tarefas específicas, é um dos tipos de algoritmos usados na IA, que desenvolve programas que aprendem a fazer previsões com base em dados, sem necessitar do auxílio de um programador. O ML é utilizado em aplicações como processamento de imagens, filtro de spam, recomendações de música, e detecção de fraudes entre outras (BISHOP, 2006).

A principal diferença entre Inteligência Artificial e *Machine Learning* se dá devido ao fato que a Inteligência Artificial é um conceito amplo que envolve mecanismos computacionais que se baseiam no comportamento humano, enquanto a *Machine Learning*, é parte da Inteligência Artificial, que trata diretamente o treinamento dos algoritmos modificando seu comportamento de forma autônoma e tendo como base a sua própria experiência.

A mudança comportamental acontece através de regras que procuram melhorar o desempenho de uma tarefa ou aplicação e até tomar uma decisão mais apropriada conforme o contexto, tais regras, são geradas com base no reconhecimento de padrões dentro dos dados analisados, conforme representado na Figura 17.

Figura 17 : Classificação através de *Machine Learning*



Fonte: acervo do autor

## 2.11. O Software MatLab

Na área científica é muito comum o uso do software MatLab para situações que envolvam estatísticas, simulações, cálculos complexos com grande quantidade de dados e até inteligência artificial.

MatLab é um poderoso sistema desenvolvido para trabalhar com análise numérica envolvido em resolução de problemas científicos, de engenharia, biomedicina e outros. O nome MatLab deriva de MATrix LABoratory, uma vez que o sistema foi concebido para tornar a manipulação de matrizes particularmente fácil (HAHN; VALENTINE, 2007).

O MatLab é um software de propriedade da MathWorks, desenvolvido para ter uma linguagem script própria que assemelha-se com a descrição matemática, pois facilita a manipulação de matrizes, análise numérica, processamento de sinais e construção de gráficos (MATHWORKS, 2007).

Atualmente o software dispõe de uma biblioteca abrangente em relação às funções matemáticas, proporcionando geração de gráficos e manipulação de dados que auxiliam o trabalho do programador.

Sua sintaxe é um pouco diferenciada, sua flexibilidade e sua capacidade de expansão, utilizando as ToolBoxes, permite ao MatLab se tornar especialista em diversas áreas de conhecimento, desde análise de elementos finitos, inteligência artificial e análise de sinais até a depuração de processamento em tempo real. (SALVADOR, 2016).

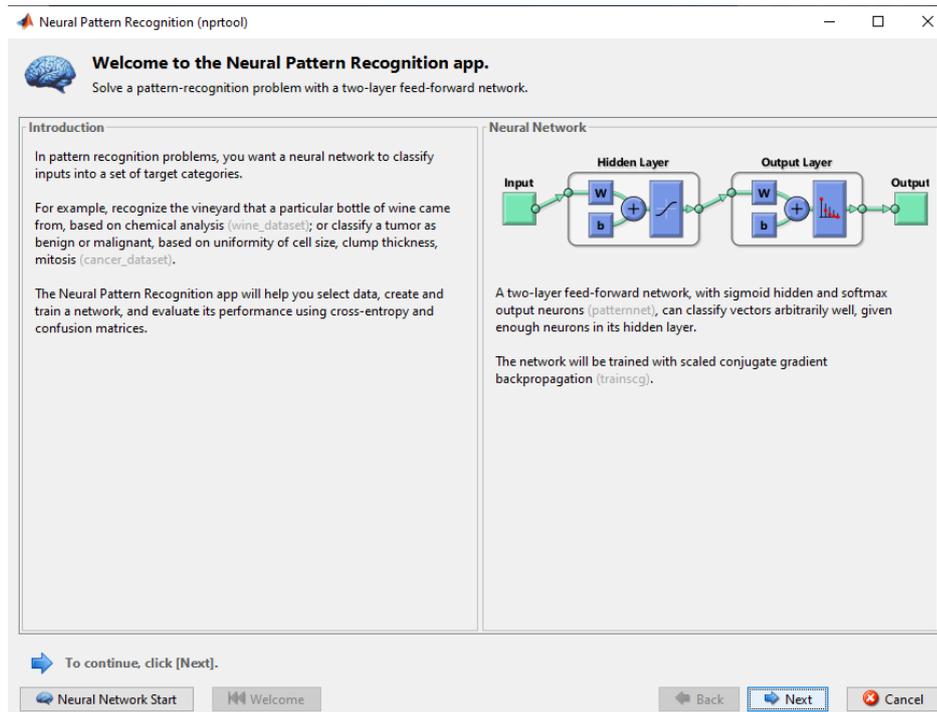
O APP “Neural Net Pattern Recognition”, foi utilizado neste projeto como referência para o desenvolvimento do código fonte, conforme Figura 18

Figura 18: TOOLBOX com APPS



Fonte: acervo do autor

Figura 19: Tela inicial do APP Neural Pattern recognition



Fonte: acervo do autor

A linguagem ou ambiente de programação MatLab permite que os usuários escrevam suas próprias bibliotecas. Assim, o usuário pode enriquecer a linguagem, incorporando a ela novas funções.

### 3. Objetivos do Trabalho

#### 3.1. Objetivos Gerais

O objetivo deste trabalho é a construção de um algoritmo que utilizando técnicas de redes neurais mais especificamente de *Machine Learning*, possa realizar o reconhecimento de padrões entre diversos tipos de folhas de café, identificando se a imagem de uma folha faz parte de um grupo doente ou saudável da respectiva planta.

#### 3.2. Objetivos Específicos

- Coletar de imagens de folhas saudáveis e doentes do cafeeiro para serem utilizadas como banco de imagens.
- Identificar as ferramentas disponíveis no MatLab para aplicação de *Machine Learning*
- Estudar as funções e recursos que realizam testes envolvendo redes neurais e mais especificamente as técnicas de *Machine Learning*.
- Criar o código fonte para identificação das imagens conforme as técnicas de *Machine Learning*.
- Testar imagens, gerando histogramas, matriz e realizando comparações.
- Ajustar o código fonte.

#### 4. Justificativa do Trabalho

A opção por um trabalho como este decorre da importância da tecnologia aplicada em soluções para agricultura, proporcionando maior produtividade, maior lucro e ações antecipadas no combate a pragas e doenças que acometem as plantações.

O seguimento agrário é responsável por uma grande parcela da economia do país, e torna-se cada vez mais necessário o investimento em soluções tecnológicas que possam ser aplicadas diretamente no campo.

Neste projeto o objetivo é que possa colaborar inicialmente comprovando que a identificação de doenças é possível através das técnicas de Inteligência Artificial, mais especificamente com uso do *Machine Learning*.

A princípio este trabalho busca identificar plantas doentes ou não, porém, é válido afirmar que uma vez comprovada a capacidade de uma aplicação *Machine Learning* identificar a planta doente, também se torna possível aprimorar a aplicação para identificar futuramente qual tipo de doença foi encontrada nas imagens analisadas, e então, propor tratamento químico.

#### 5. Metodologia

A fase de análise de requisitos é fundamental para o sucesso do processo de desenvolvimento do software. Nesta fase, o projetista (engenheiro de requisitos) especifica as funções e desempenho do software, indica a interface do software com outros sistemas, estabelece as restrições de projeto do software (PRESSMAN, 1994).

O processo de análise do projeto teve início baseando-se nos requisitos, envolvendo o estudo de softwares mais propícios para o desenvolvimento do algoritmo de análise dos dados, tais recursos necessitavam contemplar soluções eficientes no uso da inteligência artificial e *Machine Learning*.

De acordo com (KOTONYA, 1998; SOMMERVILLE, 1997), Engenharia de Requisitos é um termo que engloba todas as atividades envolvidas na descoberta, documentação e manutenção de um conjunto de requisitos para um sistema computacional. O uso do termo “engenharia” implica que técnicas sistemáticas e repetíveis devem ser utilizadas para garantir que os requisitos do sistema sejam completos, consistentes e relevantes.

Uma análise de requisitos compreende identificar cada etapa do processo de desenvolvimento do projeto e as suas especificidades, dentre as etapas foram definidos os seguintes passos:

- Análise de Requisitos
- Escolha do software (MatLab)
- Seleção de imagens para compor o banco de imagens
- Desenvolvimento do algoritmo
- Ajustes no algoritmo e nas imagens do banco
- Resultados
- Conclusão

Desde modo, foi escolhido o MatLab que tem se mostrado muito superior, inclusive já consolidado como ferramenta matemática capaz de auxiliar no processo de simulação envolvendo uma grande quantidade de dados.

Após a escolha do software o próximo passo foi escolher as imagens para compor o banco de dados e em seguida foram implementadas as funções nativas do MatLab que são capazes de realizar o treinamento e a classificação das imagens do banco.

### 5.1. Método Bag of Words

No método Bag of Words a classificação é feita a partir de uma representação alternativa do texto na forma de um histograma. Para gerar essa representação alternativa é necessária uma etapa prévia de agrupamento (*clustering*) [CARUANA and NICULESCU-MIZIL, 2006]

O bag of words (saco de palavras), faz uso da linguagem natural para representar de forma simplificada a recuperação de uma informação, neste modelo ou técnica, não existe nenhuma consideração com a gramática ou ordem em que as palavras acontecem, simplesmente identifica a quantidade de vezes que uma palavra ocorre em um determinado texto.

Com essa modelagem traçamos um vetor onde as palavras (ou termos) são as colunas e as linhas são compostas pelo número de vezes que a palavra aparece no texto, após a realização de cálculos baseados em número de vezes que a palavra apareceu no texto e a quantidade de textos analisados, alguns valores são atribuídos a cada palavra.

Se uma palavra aparece com muita frequência em um documento, ela não deve ser importante então será atribuída uma pontuação baixa a ela.

Por outro lado, se uma palavra aparece com menos frequência nesse documento, então será atribuída uma pontuação mais alta a essa palavra. Tendo assim o efeito de destacar palavras distintas.

### 5.2. Método Bag of Features

O Bag of Features é um método em visão computacional com raízes em uma técnica clássica para classificação de texto conhecida como ~Bag of Words (BoW) [SEBASTIANI 2002].

No contexto de visão computacional, o método possui forma similar ao Bag of Word, porém o agrupamento é feito a partir de “features” retiradas de imagens, ao invés de texto. Há diversas técnicas para determinar regiões de imagem relevantes para a extração de “Features” [RUI et al. 1999], como algoritmos de detecção de cantos [AWRANGJEB et al. 2012],

Para todo algoritmo ou script criado no “MatLab” é recomendado inicialmente que as variáveis que por ventura tenham sido utilizadas e que ocupam espaço na memória sejam definitivamente reiniciadas, zeradas ou esvaziadas, através do comando “clear all”.

*Sintaxe:*

```
clear all
close all
```

Em seguida a imagem que será avaliada é carregada para uma variável denominada “imgbusca”, enquanto as pastas que contém o banco de imagens são carregadas para uma outra variável “imgSets”, preservando a estrutura das pastas e de forma recursiva, além disso, a referida função permite armazenar os dados de uma imagem em um objeto no MatLab sem precisar carregá-las na memória, ou seja, serão utilizadas no ‘Workspace’, conforme demonstra a Figura 20.

*Sintaxe:*

```
imgbusca = 'caso3.jpg';
imgSets = imageSet('C:\Projeto', 'recursive');
```

Figura 20: Galeria de imagens de folhas saudáveis

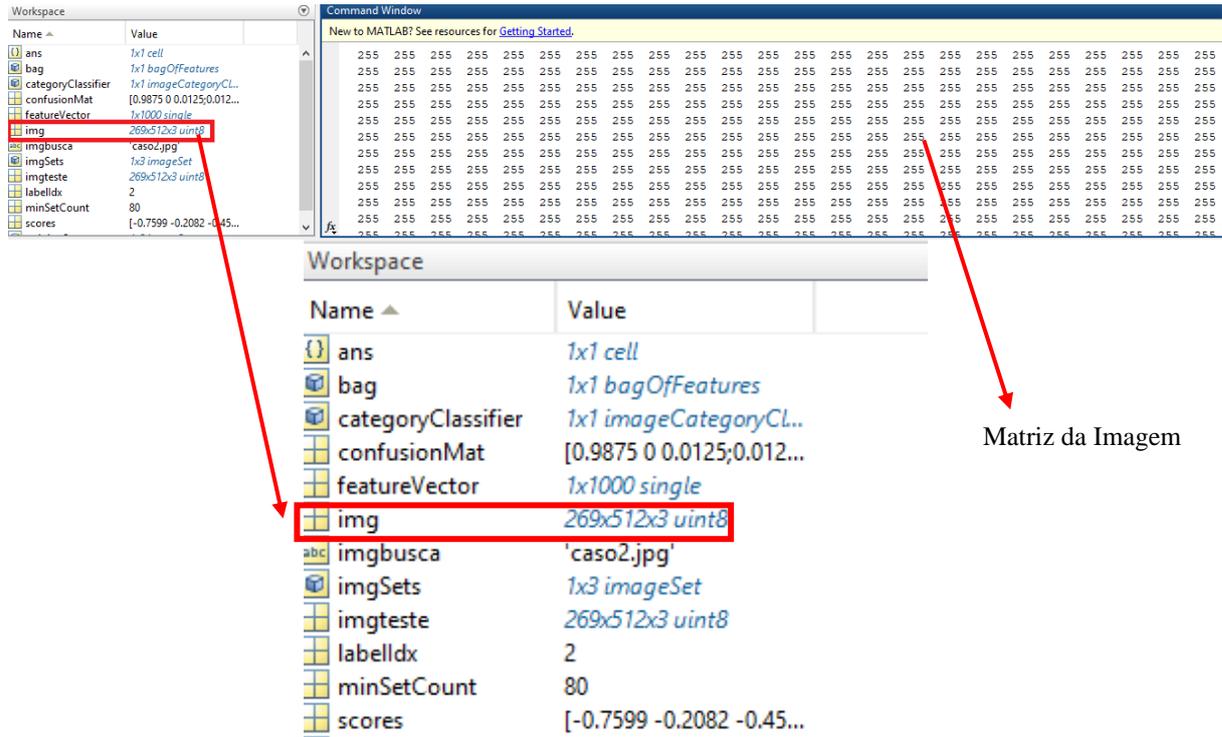


Fonte: acervo do autor

Quando uma imagem é carregada para o ambiente do MatLab a mesma é associada a uma variável. No exemplo abaixo a variável “img” foi preenchida por uma imagem do banco de teste e o resultado foi uma matriz, conforme Figura 21.

*Sintaxe:* `img = read(caso2.jpg)`

Figura 21 : Matriz de uma imagem no MatLab



Fonte: acervo do autor

Assim que as imagens foram carregadas, o treinamento de cada categoria foi realizado através da função “trainingSets” e as imagens válidas de cada pasta foram contadas, pois, o programa utiliza sempre como referência a pasta com menor número de elementos e então seleciona o mesmo número nas demais, tal procedimento é realizado para que o número de imagens em cada categoria seja exatamente igual

Através da função “Bag-Of-Features” as imagens são identificadas em seus grupos e utilizando um vocabulário com 1000 pontos avaliados em cada imagem, a função é capaz de detectar as semelhanças entre elas.

Foi gerado o histograma que permite avaliar os diferentes pontos em cada imagem analisadas pelo software conforme a quantidade estipulada no vocabulário da função de padrões.

Foi possível observar que a quantidade de pontos analisados nas imagens reflete diretamente no histograma gerado pelo MatLab, quando mais detalhes analisados em uma imagem, o que chamamos de vocabulário da imagem, mais detalhes são apresentados no histograma, a ponto que se torna necessário a sua exibição em maior escala para visualização.

Em seguida, foi gerada a matriz de confusão, onde os dados são comparados entre as imagens do mesmo grupo e também com os demais grupos, um índice de semelhança é estabelecido para cada grupo.

Através da matriz de confusão foi possível estipular uma porcentagem de precisão que o método comparativo pode alcançar, neste caso foi possível chegar a 98% de precisão.

Finalmente após a classificação, chegamos ao resultado, onde o algoritmo apontou qual grupo de imagens mais se assemelhava à imagem em análise naquele momento.

### 5.3. Desenvolvimento do Código Fonte

O desenvolvimento do código fonte foi elaborado conforme estudos realizados através dos aplicativos: “Neural Pattern Recognition (nprtool) ” e “Image Batch Processor”, que são encontrados na barra de aplicativos do MatLab, conforme Figura 22.

Figura 22: Barra de Aplicativos MatLab - Software Image Batch Processor



Fonte: acervo do autor

Os estudos e o desenvolvimento do código envolveram principalmente as técnicas dos métodos de classificação “BagOfWord” e “Bag-Of-Features”.

Código fonte:

```
clear all;
close all;

imgbusca = 'caso3.jpg';

imgSets = imageSet('C:\Projeto','recursive');
minSetCount = min([imgSets.Count]);
trainingSets = partition(imgSets,minSetCount,'randomize');
[trainingSets.Count]

bag =
bagOfFeatures(trainingSets,'VocabularySize',1000,'PointSelection','Detector
')

img = read(imgSets(1), randi(imgSets(1).Count));
featureVector = encode(bag,img);
figure
subplot(3,2,1);
imshow(img);
subplot(3,2,2);
bar(featureVector);
title('Ocorrencia Visual');
xlabel('Indice Visual');
ylabel('Frequencia');

img = read(imgSets(2), randi(imgSets(2).Count));
featureVector = encode(bag,img);
```

```

subplot(3,2,3);
imshow(img);
subplot(3,2,4);
bar(featureVector);
title('Ocorrência Visual');
xlabel('Índice Visual');
ylabel('Frequência');

img = read(imgSets(3), randi(imgSets(3).Count));
featureVector = encode(bag, img);
subplot(3,2,5);
imshow(img);
subplot(3,2,6);
bar(featureVector);
title('Ocorrência Visual');
xlabel('Índice Visual');
ylabel('Frequência');

img = imread(fullfile('C:\analisar', imgbusca));
featureVector = encode(bag, img);
figure
subplot(3,2,1);
imshow(img);
subplot(3,2,2);
bar(featureVector);
title('Ocorrência Visual');
xlabel('Índice Visual');
ylabel('Frequência');

categoryClassifier = trainImageCategoryClassifier(trainingSets, bag);
confusionMat = evaluate(categoryClassifier, trainingSets);

imgteste = imread(fullfile('C:\analisar', imgbusca));
figure
imshow(imgteste)
[labelIdx, scores] = predict(categoryClassifier, imgteste);
categoryClassifier.Labels(labelIdx)

```

## 6. Resultados e Discussões

O resultado esperado foi alcançado, uma vez que, o algoritmo consegue realizar a tarefa de classificação por padrões através do uso de “*Machine Learning*” com aprendizado supervisionado.

A aplicação do algoritmo pode ser modificada conforme o banco de imagens a ele apresentado, o que torna possível a identificação de outros objetos ou produtos como a cana de açúcar, a soja, o algodão e não só a folha do café.

O primeiro passo que buscava a identificação entre folha saudável e folha doente era imprescindível para comprovar que outros grupos de imagens também possam utilizar o algoritmo para treinamento, classificando por exemplo grupos de tipos de doenças, neste caso, ao invés do algoritmo retornar simplesmente se a planta está ou não doente, o mesmo é capaz

de identificar qual grupo de doenças e então recomendar conforme estudos laboratoriais, quais os produtos químicos devem ser utilizados no combate à doença.

Todos estes aspectos são melhorias e aprimoramentos que o algoritmo permite realizar em estudos futuros, assim como, as aplicações de novas tecnologias, como por exemplo transformá-lo em um código para web ou que possa se tornar um aplicativo para smartphone, onde o usuário poderia através de uma câmera capturar uma imagem de uma folha qualquer e o sistema identificar a doença logo em seguida.

Foi possível notar que as funções de inteligência artificial utilizadas no MatLab são sensíveis às configurações aplicadas, no caso da função “*Bag-Of-Features*” por exemplo, o vocabulário utilizado na função diz respeito à quantidade de características a serem analisadas em cada imagem do banco.

Outro ponto observado é que a medida que o banco de imagens vai ficando maior, com uma quantidade maior de fotos, a quantidade de pontos analisados deve ser também aumentada, para que a análise seja perfeita, caso contrário serão muitas imagens e poucos pontos analisados em cada uma delas, o que pode ocasionar erros por semelhança de detalhes.

Em testes do início do projeto o número de imagens era pequeno, apenas 30 por grupo, então, não havia necessidade de avaliar uma grande quantidade de características por imagem, e com apenas 200 pontos analisados já era possível identificar o grupo correto, mas, à medida em que o banco de imagens foi alimentado com mais figuras, passando para 80 em cada grupo, foi necessário aumentar para 1000 o número de características analisadas por foto, para que, o algoritmo não resultasse em erros ao analisar imagens muito semelhantes.

Também é possível mudanças no algoritmo de forma que, ao identificar uma nova imagem como semelhante a um dos grupos de imagens do banco, o mesmo possa copiá-la para a respectiva pasta, alimentando ainda mais o banco de imagens, ressaltando que, no início do algoritmo já existe uma função que antes do processo de treinamento dar início, seleciona a mesma quantidade de imagens em cada pasta ou grupo para não causar pesos diferentes na análise, ou seja, é importante que o número de imagens seja sempre igual em todas as categorias do banco.

Tal aprimoramento faria com que o sistema fosse capaz de “aprender” cada vez mais, não só com as imagens já apresentadas, mas a cada teste realizado com figuras novas classificadas e posteriormente incorporadas ao banco de imagens.

Resultados gerados após a execução do código fonte:

```
ans =
```

```
80 80 80
```

```
Creating Bag-Of-Features from 3 image sets.
```

```
-----
```

```
* Image set 1: doentes.
```

```
* Image set 2: outros.
```

```
* Image set 3: saudavel.
```

- \* *Selecting feature point locations using the Detector method.*
- \* *Extracting SURF features from the selected feature point locations.*
- \*\* *detectSURFFeatures is used to detect key points for feature extraction.*
  
- \* *Extracting features from 80 images in image set 1...done. Extracted 12566 features.*
- \* *Extracting features from 80 images in image set 2...done. Extracted 14743 features.*
- \* *Extracting features from 80 images in image set 3...done. Extracted 17631 features.*
  
- \* *Keeping 80 percent of the strongest features from each image set.*
  
- \* *Balancing the number of features across all image sets to improve clustering.*
- \*\* *Image set 1 has the least number of strongest features: 10053.*
- \*\* *Using the strongest 10053 features from each of the other image sets.*
  
- \* *Using K-Means clustering to create a 1000 word visual vocabulary.*
- \* *Number of features : 30159*
- \* *Number of clusters (K) : 1000*
  
- \* *Initializing cluster centers...100.00%.*
- \* *Clustering...completed 33/100 iterations (~0.42 seconds/iteration)...converged in 33 iterations.*
  
- \* *Finished creating Bag-Of-Features*

*bag =*

*bagOfFeatures with properties:*

*VocabularySize: 1000*  
*StrongestFeatures: 0.8000*  
*PointSelection: 'Detector'*  
*Upright: 1*

*Warning: MatLab has disabled some advanced graphics rendering features by switching to software OpenGL. For more information, click here.*

*Training an image category classifier for 3 categories.*

- 
- \* *Category 1: doentes*
  - \* *Category 2: outros*
  - \* *Category 3: saudavel*
  
  - \* *Encoding features for category 1...done.*
  - \* *Encoding features for category 2...done.*
  - \* *Encoding features for category 3...done.*
  
  - \* *Finished training the category classifier. Use evaluate to test the classifier on a test set.*

*Evaluating image category classifier for 3 categories.*

-----

\* *Category 1: doentes*

\* *Category 2: outros*

\* *Category 3: saudavel*

\* *Evaluating 80 images from category 1...done.*

\* *Evaluating 80 images from category 2...done.*

\* *Evaluating 80 images from category 3...done.*

\* *Finished evaluating all the test sets.*

\* *The confusion matrix for this test set is:*

<i>KNOWN</i>	<i>PREDICTED</i>		
	<i>/ doentes</i>	<i>outros</i>	<i>saudavel</i>
<i>doentes</i>	<i>/ 0.99</i>	<i>0.00</i>	<i>0.01</i>
<i>outros</i>	<i>/ 0.01</i>	<i>0.99</i>	<i>0.00</i>
<i>saudavel</i>	<i>/ 0.00</i>	<i>0.04</i>	<i>0.96</i>

\* *Average Accuracy is 0.98.*

*ans =*

*'doentes'*

*>>*

## 6.1. Resultados Comentados:

Importando as imagens e verificando a descrição das pastas importadas, conforme demonstra a Figura 23:

Figura 23 : Descrição das pastas de imagens

```

Command Window
//
>> {treinamento.Description}

ans =

1x3 cell array

    {'doentes'}    {'saudavel'}    {'testes'}

```

Fonte: acervo do autor



A função “*bagOfFeatures*” seleciona pontos a serem analisados, ela permite delimitar a quantidade de pontos através da função “*VocabularySize*”, que neste exemplo foi programada para 1000 pontos de análise, a seleção é randomizada e o treinamento da rede é iniciado, conforme é demonstrado na Figura 26.

Figura 26 : Execução da função bagOfFeatures

```

Command Window
Creating Bag-Of-Features from 3 image sets.
-----
* Image set 1: doentes.
* Image set 2: outros.
* Image set 3: saudavel.

* Selecting feature point locations using the Detector method.
* Extracting SURF features from the selected feature point locations.
** detectSURFFeatures is used to detect key points for feature extraction.

* Extracting features from 80 images in image set 1...done. Extracted 12566 features.
* Extracting features from 80 images in image set 2...done. Extracted 14743 features.
* Extracting features from 80 images in image set 3...done. Extracted 17631 features.

* Keeping 80 percent of the strongest features from each image set.

* Balancing the number of features across all image sets to improve clustering.
** Image set 1 has the least number of strongest features: 10053.
** Using the strongest 10053 features from each of the other image sets.

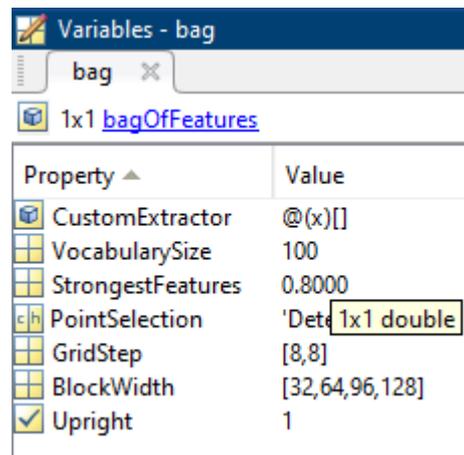
* Using K-Means clustering to create a 1000 word visual vocabulary.
* Number of features      : 30159
* Number of clusters (K)  : 1000

* Initializing cluster centers...100.00%.
* Clustering...completed 24/100 iterations (~0.98 seconds/iteration)...converged in 24 iterations.

```

Fonte: acervo do autor

Figura 27 : Resultado da classe bagOfFeatures



Property	Value
CustomExtractor	@(x)[]
VocabularySize	100
StrongestFeatures	0.8000
PointSelection	'Detector' <b>1x1 double</b>
GridStep	[8,8]
BlockWidth	[32,64,96,128]
Upright	1

Fonte: acervo do autor

### Trecho para geração da classe:

```

bag=bagOfFeatures(trainingSets,'VocabularySize',1000,'PointSelection','Detector')
img = read(imgSets(1), randi(imgSets(1).Count));
featureVector = encode(bag,img);

```

Dando sequência, um objeto de cada classe é selecionado para geração de histogramas, a função de cada um é simplesmente demonstrar que os histogramas são diferentes para cada categoria de imagens, os histogramas aqui não possuem qualquer interferência no processo de seleção ou treinamento do algoritmo, são simplesmente ilustrativos para demonstrar as diferenças das características dos elementos de cada conjunto do banco de dados.

```

figure
subplot(3,2,1);
imshow(img);
subplot(3,2,2);
bar(featureVector);
title('Ocorrencia Visual');
xlabel('Indice Visual');
ylabel('Frequencia');

img = read(imgSets(2), randi(imgSets(2).Count));
featureVector = encode(bag,img);
subplot(3,2,3);
imshow(img);
subplot(3,2,4);
bar(featureVector);
title('Ocorrencia Visual');
xlabel('Indice Visual');
ylabel('Frequencia');

img = read(imgSets(3), randi(imgSets(3).Count));
featureVector = encode(bag,img);
subplot(3,2,5);
imshow(img);
subplot(3,2,6);
bar(featureVector);
title('Ocorrencia Visual');
xlabel('Indice Visual');
ylabel('Frequencia');

img = imread(fullfile('C:\analisar',imgbusca));
featureVector = encode(bag,img);
figure
subplot(3,2,1);
imshow(img);
subplot(3,2,2);
bar(featureVector);
title('Ocorrencia Visual');
xlabel('Indice Visual');
ylabel('Frequencia');

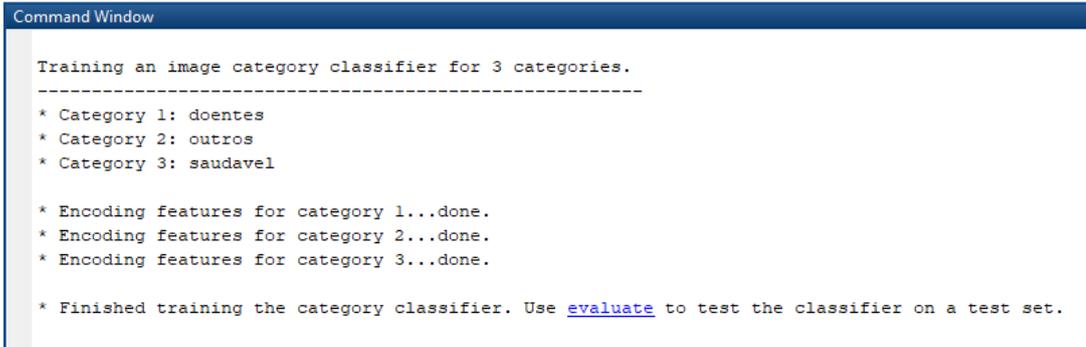
```

Através da função “trainImageCategoryClassifier” as imagens são classificadas como mesmo grupo, conforme demonstrado na Figura 28.

*Sintaxe:*

```
categoryClassifier = trainImageCategoryClassifier(trainingSets, bag);
```

Figura 28 : Classificação de categorias



```
Command Window
Training an image category classifier for 3 categories.
-----
* Category 1: doentes
* Category 2: outros
* Category 3: saudavel

* Encoding features for category 1...done.
* Encoding features for category 2...done.
* Encoding features for category 3...done.

* Finished training the category classifier. Use evaluate to test the classifier on a test set.
```

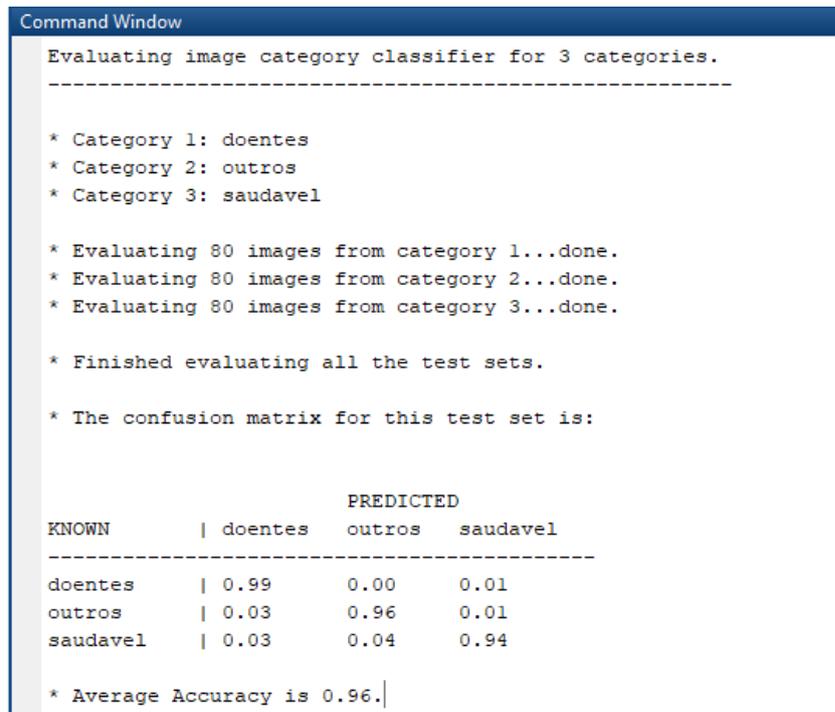
Fonte: acervo do autor

Após o treinamento, a função “evaluate” realiza a geração da matriz de confusão, onde é possível notar a porcentagem de semelhança das imagens quando comparadas de forma recursiva com as demais imagens do mesmo grupo e também com os outros grupos, observe a Figura 29:

*Sintaxe:*

```
confusionMat = evaluate(categoryClassifier, trainingSets);
```

Figura 29 : Criação da Matriz de confusão



```
Command Window
Evaluating image category classifier for 3 categories.
-----
* Category 1: doentes
* Category 2: outros
* Category 3: saudavel

* Evaluating 80 images from category 1...done.
* Evaluating 80 images from category 2...done.
* Evaluating 80 images from category 3...done.

* Finished evaluating all the test sets.

* The confusion matrix for this test set is:

          PREDICTED
KNOWN    | doentes  outros  saudavel
-----
doentes  | 0.99    0.00    0.01
outros   | 0.03    0.96    0.01
saudavel | 0.03    0.04    0.94

* Average Accuracy is 0.96.
```

Fonte: acervo do autor

Quando comparadas as fotos do grupo “doentes” com as demais fotos do mesmo grupo é possível notar uma semelhança de 99%, mas quando o grupo “doentes” é comparado com o grupo “outros”, a semelhança é 0% e com o grupo saudável 1% apenas.

Comparando então o grupo “outros” com o grupo doentes, tem-se apenas 3% de semelhança, mas quando a comparação acontece com elementos do próprio grupo esse percentual é de 96% e com o grupo saudável apenas 1%.

A comparação partindo do grupo “saudável” com o grupo “doentes” apresenta apenas 3% de semelhança, 4% quando comparadas com o grupo “outros”, já com as fotos do mesmo grupo a semelhança agora é de 94%, algumas variações de porcentagem podem não abranger os 100% nas comparações, isso acontece devido ajustes de casas decimais e arredondamentos do software.

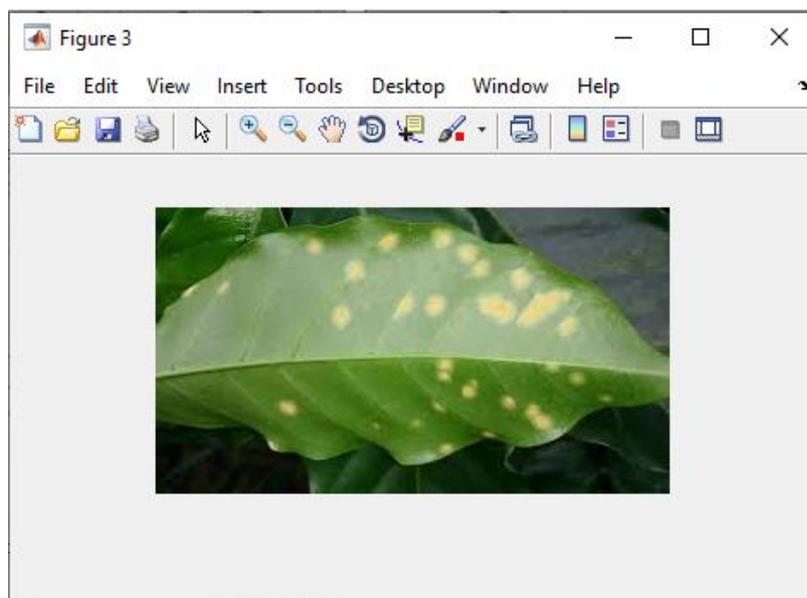
Sendo assim, a matriz de confusão nos apresenta uma eficácia de 96% na análise das imagens que serão submetidas ao teste de semelhança com os grupos, porcentagem essa considerável e totalmente aceitável para comparação.

A imagem em análise é carregada para uma variável utilizando a função `imread` e da função `imshow` apresenta a mesma na tela, conforme Figura 30.

*Sintaxe:*

```
imgteste = imread(fullfile('C:\analisar',imgbusca));
figure
imshow(imgteste)
```

Figura 30 : Imagem de uma folha doente avaliada pelo algoritmo



Fonte: acervo do autor

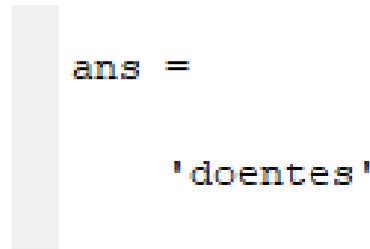
Apesar da função “evaluate” ter realizado a construção da matriz de confusão é a função “predict” que exibe a mensagem final classificando a imagem analisada como semelhante a uma das categorias do banco de imagens.

*Sintaxe:*

```
[labelIdx, scores] = predict(categoryClassifier, imgteste);
categoryClassifier.Labels(labelIdx)
```

A resposta para a imagem analisada é então apresentada ao final da execução do algoritmo, conforme demonstrado na Figura 31.

Figura 31 : Resultado final do Algoritmo



```
ans =
      'doentes'
```

Fonte: acervo do autor

## 7. Considerações Finais

O objetivo do trabalho foi alcançado uma vez que se propôs a construção de um algoritmo de identificação de imagens e comparação entre um grupo considerado doente e outro saudável através das técnicas de Redes Neurais, mais especificamente utilizando “*Machine Learning*” para o reconhecimento de padrões.

Já comprovada a eficácia do software MatLab em soluções que envolvam inteligência artificial e a importância da aplicação das *Machine Learning* em todos os seguimentos de desenvolvimento de softwares que possam colaborar com a tomada técnicas de decisão, pode-se notar que as técnicas são perfeitamente aplicáveis ao tema proposto.

A precisão da análise das características das funções de *Machine Learning* se mostrou surpreendente após os ajustes necessários para criação dos histogramas, uma vez que, mesmo diante de diferentes cores, luminosidade, perspectivas ou outras influências que uma figura pode conter, o resultado foi assertivo.

Foi possível confirmar a eficiência do método bag-of-feature que se mostrou eficaz na classificação das fotos e foi capaz de identificar uma nova imagem como semelhante à um dos grupos.

Obviamente que as imagens para o banco foram previamente selecionadas e algumas delas foram descartadas por estarem em má qualidade ou até mesmo em distâncias de foco muito diferentes, seria muito mais complexo identificar a doença em uma foto panorâmica por exemplo, por isso, é necessário a pré-seleção de fotos realmente válidas para os grupos em questão.

Na possibilidade de alimentar o banco de imagens com cada imagem testada e classificada pelo algoritmo, deve-se considerar um acompanhamento das imagens testadas com intuito de preservar o bom funcionamento do algoritmo e garantir que somente imagens corretas sejam carregadas para as devidas pastas com imagens semelhantes.

## **8 - Propostas Futuras**

- Identificar doenças específicas relacionadas ao café
- Readequação do código fonte com possibilidade de transformação em aplicativo para Website
- Readequação para utilizar o software em outros seguimentos da agricultura
- Recomendações para tratamento químico preventivos e com pesticidas conforme doença identificada

O atual trabalho possibilita o desenvolvimento futuro de aplicativos para Smartphones ou Websites através da tradução do algoritmo (código-fonte) para uma linguagem de programação web.

Será possível identificar qual doença mais provável dentro de grupos de doenças pré-estabelecidas e não somente como folha doente ou não, a partir de então, recomendar tratamentos químicos preventivos ou pesticidas conforme indicações químicas para cada tipo de doença.

Para o pequeno agricultor uma ferramenta como essa se torna fundamental para identificar não só no cafeeiro, mas também em outros tipos de cultivo, caso a ferramenta permita, pois, o mesmo conseguirá agir no combate das doenças de forma mais rápida podendo evitar gastos com profissionais e testes de identificação das doenças.

## 9. Referências

- ABIC. Associação Brasileira da Indústria do café. Disponível em: <<https://www.abic.com.br/o-cafe/historia/o-cafe-brasileiro-na-Atualidade-2/>>. Acesso em 15 Ago. 2020
- ADAMA BRASIL. Adama Alvo. Disponível em <https://www.adama.com/brasil/pt/adama-inovacao/adama-alvo>. Acesso em: 30 Mai. 2019.
- ADAMCHUK, Viacheslav I.; PERK, Richard L.; SCHEPERS, James S. Precision Agriculture: Applications of Remote Sensing in Site-Specific Management. Universidade de Nebraska, 2003.
- AGRIOS, G. N. Plant Pathology, 5. ed. Oxford, UK: Academic Press, 2005.
- ALVES, Virgínia. Dia Mundial do café: Os números que comprovam o Brasil no topo de produção e exportação. Cafeicultura. 2020. Disponível em: <<https://revistacafeicultura.com.br/?mat=68988#:~:text=Atualmente%20Minas%20Gerais%20concentra%20a,de%20produ%C3%A7%C3%A3o%20de%20caf%C3%A9s%20especiais.&text=Outra%20regi%C3%A3o%20de%20relev%C3%A2ncia%20na,na%20parte%20norte%20do%20pa%C3%ADs.>>> Acesso em: 14 abril 2020.
- AWRANGJEB, M., Lu, G., and FRASER, C. S. (2012). Performance comparisons of contourbased corner detectors. IEEE Transactions on Image Processing.
- BISHOP, C. M. Patternrecognitionandmachinelearning. [S.l.]: springer, 2006.
- BROWNLEE, J. Master *Machine Learning* Algorithms: Discover how They Work and Implement Them from Scratch. [s.n.], 2016.
- CARUANA, R. and NICULESCU-MIZIL, A. (2006). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning, pages 161–168. ACM.
- CARVALHO, Vicente L.;DA CUNHA, Rodrigo L.;SILVA, Nathan R.N. Alternativas de Controle de Doenças do Cafeeiro. Coffee Science, 2012.
- CARVALHO, G. R. et al. Avaliação de produtividade de progênies de cafeeiro em dois sistemas de plantio. Ciência e Agrotecnologia, Lavras, v. 30, n. 5, p. 838-843, set./out. 2006.
- CARVALHO, V. L.; CHALFOUN, S. M. Doenças do Cafeeiro: diagnose e controle. Belo Horizonte: Epamig, 2000, 44 p. (EPAMIG: Boletim Técnico, 58).
- CASTRO. Eduardo José Ribeiro de. Rede de Requisitos. Jul. 2017. Disponível em: <<http://rederequisitos.com.br/o-que-sao-requisitos-e-requisitos-de-software/>>. Acesso em: 10 Set. 2020.

CEOLIN, A. C.; ABICHT, A. D. M.; CORRÊA, A. D. F.; PEREIRA, P. R. R. X.; SILVA, T. D. Sistemas de informação sob a perspectiva de custos na gestão da pecuária de corte gaúcha. *Custos e @gronegocio on-line*, Recife, PE, v. 4, Edição Especial, maio 2008.

CHAVES, Eugênio. SILVA, Marília Goulart. POZZA, Edson Ampélio. POZZA, Adélia Aziz Alexandre. Manejo da Cercosporiose, doença de extrema agressividade na cultura do café. *Revista Cultivar*. UFLA. n. 184. set. 2014.

CONAB – Companhia Nacional de Abastecimento. Histórico Mensal do Café – Janeiro de 2019. Disponível em: <https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuário-e-extrativista/analises-do-mercado/historico-mensal-de-cafe/>. Acesso em: 30 Jul. 2019.

CONSORCIO ANTIFERRUGEM. Mapa da Dispersão, 2019. Disponível em: <http://www.consorcioantiferrugem.net/#/main>. Acesso em: 20 Ago. 2019.

COOPERCAM. Cooperativa dos Cafeicultores de Campos Gerais e Campo do Meio Ltda. Disponível em: < [http://coopercam.com.br/noticias/selecionada/ibge\\_previsao\\_cafe\\_2020](http://coopercam.com.br/noticias/selecionada/ibge_previsao_cafe_2020) > Acesso em 20 Set. 2020.

DETECTOR DE PRAGAS. 2014. Disponível em: <https://apkpure.com/br/detector-de-pragas/com.ufms.cppp.detectordepragas>. Acesso em 15 Mar 2019.

DEVMIDIA. Introdução a Requisitos de Softwares. Disponível em: <<https://www.devmedia.com.br/introducao-a-requisitos-de-software/29580#:~:text=Requisitos%20s%C3%A3o%2C%20al%C3%A9m%20de%20fun%C3%A7%C3%B5es,necess%C3%A1ria%20para%20satisfazer%20um%20objetivo.>> Acesso em: 12 Set. 2020

DUDA, R. O. and HART, P. E., *Patternclassificationandsceneanalysis*. Johnhiley Sons, New York, 1973.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Doenças do Cafeeiro. Disponível em: [http://www.sapc.embrapa.br/arquivos/consorcio/publicacoes\\_tecnicas/folder\\_doencascafeeiro.pdf](http://www.sapc.embrapa.br/arquivos/consorcio/publicacoes_tecnicas/folder_doencascafeeiro.pdf). Acesso em 20 Jun 2019.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Produção dos Cafés do Brasilequivale a 36% da produção mundial em 2018. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/34724227/producao-dos-cafes-do-brasil-equivale-a-36-da-producao-mundial-em-2018>. Acesso: 15 Jun. 2019.

FAPESP. Revista Pesquisa. Disponível em: <<https://revistapesquisa.fapesp.br/agricultura-4-0>>. Acesso: 20 Mar. 2020.

FERREIRA, J. B.; SILVA, E. H.; FERNANDES, K. D.; PEREIRA, R. B.; ABREU, M. S.; PEREIRA, I. S. Efeito de fungicidas no controle da seca de ramos do cafeeiro (*C. arabica* L.) com mancha manteigosa (*Colletotrichum* spp.). **Fitopatologia Brasileira**, Brasília, v. 30, p. 111, 2005. Resumo. Suplemento.

FREITAS, L. G.; OLIVEIRA, R. D. L.; FERRAZ, S. Introdução à nematologia: cadernos didáticos. Viçosa, MG: UFV, 1999.

FRITZENLAB. Eletrônica & Eletrônicos. Disponível em: <https://fritzenlab.com.br/2016/10/termometro-com-led-rgb-e-arduino-attiny85/tabela-codigo-cores-rgb/>. Acesso em: 20 Ago 2019.

FUKUSHIMA, KUNIIHIKO, “Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition”, *Neural Networks*, Vol 1.

GARCIA, J. R.; VIEIRA FILHO, J. E. R. Política agrícola brasileira: produtividade, inclusão e sustentabilidade. *Política Agrícola*, Brasília, jan/mar. 2014.

GLOBORURAL. Disponível em: <https://revistagloborural.globo.com/Tecnologia-no-Campo/noticia/2016/02/aplicativo-agiliza-controle-de-pragas-e-doencas-no-campo.html>. Acesso em: 10 Dez 2018.

GONZALES, R. C.; WOODS, R. E. Digital image processing. New York: Addison-Wesley Publishing Company, Inc, 1987.

GRAO GOURMET. Produção de Café no Brasil. Disponível em: <https://www.graogourmet.com/blog/previsao-de-producao-de-cafe-20172018-e-estimada-em-159-milhoes-de-sacas/> Acesso em : 12 Ago. 2020.

HAHN, Brian D; VALENTINE, Daniel T., *Essential MatLab for Engineers and Scientists*, 3a Edição, Butterworth-Heinemann. 2007.

KASABOV, K. N. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. USA, Massachusetts Institute of Technology Press, 1996.

KOHAVI, R., SOMMERFIELD, D. & DOUGHERTY, J. (1997). *Data mining using MLC++: A machine learning library in C++*. *International Journal on Artificial Intelligence Tools*.

KOTONYA, G.; SOMMERVILLE, I. “Requirements Engineering: process and techniques”. New York: John Wiley & Sons Ltd, 1998, 282p.

KOHAVI, Ron e PROVOST, Foster. (1998): “**Glossary of terms**” *Machine Learning* **30**. Disponível em: <http://bit.ly/2CIYKNC>

LOPES, Paulo Afonso. *Probabilidades & Estatística*. 1. ed. Rio de Janeiro: R&A, 1999.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, vol.14, no2, 2007.

Disponível em:

[http://seer.ufrgs.br/index.php/rita/article/viewPDFInterstitial/rita\\_v14\\_n2\\_p43-67/3543](http://seer.ufrgs.br/index.php/rita/article/viewPDFInterstitial/rita_v14_n2_p43-67/3543).

Acesso em: 16 out. 2010.

MAPPA. Histograma: o que é e como funciona a nova ferramenta da Mappa. Disponível em: <https://mappa.ag/o-que-e-histograma-mappa/> Acesso em: 02 mai 2020.

MARCOLAN, A. L.; RAMALHO, A. R.; MENDES, A. M.; TEIXEIRA, C. A. D.; FERNANDES, C. de F.; COSTA, J. N. M.; VIEIRA JÚNIOR, J. R.; OLIVEIRA, S. J. de M.; FERNANDES, S. R.; VENEZIANO, W. Cultivo dos cafeeiros Conilon e Robusta para

Rondônia. 3. ed. rev. atual. Porto Velho: Embrapa Rondônia: Emater-RO, 2009. (Embrapa Rondônia. Sistema de produção, 33).

MAREN, A.J., HARSTON, C.T. e PAP, R.M. Hand book of Neural Computing Applications. San Diego: Academic Press, 1990.

MATIELLO, J.B. O café: do cultivo ao consumo. São Paulo: Globo, 1991.

MATTOS, J. R. L.; GUIMARÃES, L. S. Gestão da tecnologia e inovação: uma abordagem prática. São Paulo: Saraiva, 2005.

MITCHEL, Tom. (1997): *Machine Learning*. McGraw-Hill Science/Engineering/Math. Disponível em: <http://www.cs.ubbcluj.ro/~gabis/ml/ml-books/McGrawHill%20-%20Machine%20Learning%20-Tom%20Mitchell.pdf>.

MIZUBUTI, E.S.G., MAFFIA, L.A. & SIVA, C.A.B. Potencial do uso de redes neuronais em simulação de epidemias. Fitopatologia Brasileira, Brasília, 1994.

NOTÍCIAS AGRÍCOLAS. Disponível em:  
<<https://www.noticiasagricolas.com.br/noticias/cafe/256885-dia-mundial-do-cafe-os-numericos-que-comprovam-o-brasil-no-topo-de-producao-e-exportacao.html#.X27J5chKjcc>>  
Acesso em: 20 Mai 2020

OCTAVIO, Café. Café Arábica x Café Robusta. 30 Mar. 2020. Disponível em: <<https://octaviocafe.com.br/cafe-arabica-x-cafe-robusta/>> Acesso em: 30 Set. 2020

OLIVEIRA, E. et al. Eliminação de linhas em cafeeiros adensados por meio semi mecanizado. Ciência e Agrotecnologia, Lavras, v. 31, n. 6, p. 1826-1830, nov./dez. 2007.

PAO, Y., 1989, Adaptive Pattern recognition and neural networks. Addison-Wesley

PINTO, Tales dos Santos. "Raízes do café no Brasil"; *Brasil Escola*. Disponível em: <https://brasilestola.uol.com.br/historia/o-cafe-no-brasil-suas-origens.htm>. Acesso em 14 de mai de 2020.

POZZA, E.A. Desenvolvimento de sistemas especialistas e redes neuronais e suas aplicações em fitopatologia. Viçosa: UFV, 1998.

PRESSMAN, R.S. Software engineering: a practitioner's approach. European Edition, 1994, 801p.

PROJECTOR CENTRAL. The Deep Dive on Bit Depth. Disponível em: <https://www.projectorcentral.com/All-About-Bit-Depth.htm>. Acesso em 30 Jul. 2019.

QUEIROZ, José Eustáquio Rangel de. GOMES, Herman Martins. Introdução ao Processamento Digital de Imagens> Disponível em:  
<<http://www.dsc.ufcg.edu.br/~hmg/disciplinas/graduacao/vc-2014.1/Rita-Tutorial-PDI.pdf>>. Acesso em: 01 Dez.2018.

REKOGNITION. Disponível: <<https://console.aws.amazon.com/rekognition/home>>. Acesso em: 20 Nov. 2018

REDE AGRO SA. Softwares para Agronegócio, 2019. Disponível em:<<https://www.redeagroservices.com.br/Products/15624/software-agro>>. Acesso: 15 Mai. 2019.

ROCHA, C. A. J., PADILHA, T.P.P., HORST, P. S., & REZENDE, S. O. (1997). Processo de desenvolvimento de sistemas baseados em conhecimentos considerando aspectos de qualidade. In Anais da SEPAI – Semana Paraense de Informática – SUCESU, Belém – PA.

RSI. Image Processing in IDL. Ed. 2005

RUI, Y., HUANG, T. S., and CHANG, S.-F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, 10(1):39–62.

RURAL PECUÁRIA. Minas Gerais segue no topo como maior estado produtor de café do Brasil, 2019. Disponível em: <http://ruralpecuaria.com.br/noticia/minas-gerais-segue-no-topo-como-maior-estado-produtor-de-cafe-do-brasil.html>. Acesso em: 22 Jul. 2019.

RUSSEL, Stuart; NORVIG, Peter. Inteligência Artificial. 2. Ed. Rio de Janeiro: Campos, 2004.

SAFRAES. Café sobe nos mercados interno e externo na semana. Disponível em: <<https://www.safraes.com.br/mercado/cafe-sobe-nos-mercados-interno-externo-na-semana>> Acesso em: 10 Mai. 2020.

SALVADOR.Valter. O que é o MatLab. Ago. 2016. Disponível em: <<https://www.portalgsti.com.br/2016/08/o-que-e-o-MatLab.html> >. Acesso em 20 Mai. 2020.

SANTOS, Marcos Eduardo dos. JESUS, José Carlos dos Santos. Fatores que comprometem a adoção de tecnologia de informação em empresas cafeeiras. Disponível em: <[https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-65132005000300014&lang=pt](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132005000300014&lang=pt)> Acesso em 17 ago. 2020

SEBASTIANI, F. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 2002.

SILVA, F. M. da et al. Efeitos da colheita manual na bienalidade do cafeeiro em Ijaci, Minas Gerais. *Ciência e Agrotecnologia*, Lavras, maio/jun. 2010.

SMITH, John. Integred Spatial and Feature Image Systems: Retrieval, Analysis and Compression, Doctor thesis, 1997.  
<ftp://ftp.ctr.columbia.edu/CTRResearch/advent/public/public/jrsmith/thesis/>, (27 de Maio de 1998, 02:45].

SOJA, EMBRAPA. Consórcio Antiferrugem. Disponível em: <<http://www.consorcioantiferrugem.net/#/main>> Acesso em 22 ago. 2020

SOMMERVILLE, Ian; SAWYER, Peter. “Requirements Engineering – a good practice guide”. New York: John Wiley & Sons Ltd, 1997, 391p.

STAIR, Ralph M.; REYNOLDS, George W. Princípios de Sistemas de Informação. São Paulo: Thomson, 2006.

TAFNER, Malcon Anderson. Estagiamento Automático do Sono usando uma Rede Neural Artificial com Alta Representação Cortical. Florianópolis, 1998. Exame de Qualificação - Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina.

THORP, K. R. et al. Methodology for the use of DSSAT models for precision agriculture decision support. *Computers and Electronics in Agriculture*, Washington, v. 64, n. 2, p. 276-285, 2008.

TURETTA, D, P. A.; CASTRO, S. S.; POLIDORO, C. J. Solos, sustentabilidade e provisão de serviços ecossistêmicos. *Boletim informativo da SBCS*, 2017.

WITTEN, Ian H. e FRANK, Eibe. (2005): **Data mining** : practical *machine learning* tools and techniques. Elsevier.

ZAMBOLIM, L.; SOBREIRA, D. G.; SOUZA, A. F.; COSTA, H. Manejo integrado de doenças do conilon (*Coffea canephora*). In: Zambolim, L. (Ed.). *Tecnologias para produção do café conilon*. Viçosa, MG: UFV, 2009.

ZAPAROLLI, Domingos. Agricultura 4.0. *Revista Pesquisa*. <Disponível em: <https://revistapesquisa.fapesp.br/agricultura-4-0/> >. Acesso em 17 set. 2020