



**UNIVERSIDADE FEDERAL DO VALE DO SÃO
FRANCISCO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE
COMPUTAÇÃO**

VICTOR GUSTAVO DA SILVA OLIVEIRA

**VALIDAÇÃO DE MÉTODO DE DETERMINAÇÃO DE
ATRIBUTOS DE QUALIDADE EM UVAS UTILIZANDO
TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

JUAZEIRO - BA

2019

VICTOR GUSTAVO DA SILVA OLIVEIRA

**VALIDAÇÃO DE MÉTODO DE DETERMINAÇÃO DE
ATRIBUTOS DE QUALIDADE EM UVAS UTILIZANDO
TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

Trabalho apresentado à Universidade Federal
do Vale do São Francisco - Univasf, Campus
Juazeiro, como requisito da obtenção do
título de Bacharel em Engenharia de
Computação.

Orientador: Prof. D. Sc. Rosalvo Ferreira de
Oliveira Neto
Coorientador: Prof. D. Sc. Daniel dos Santos
Costa

JUAZEIRO - BA

2019

Para Admilson e Edileuza, pelo amor incondicional, neste ou em qualquer plano.

AGRADECIMENTOS

"Deus mora nos detalhes"

Este antigo ditado alemão diz muito sobre este trabalho. Os detalhes, quanto menores, mais complexos de serem evidenciados.

Agradeço a Deus, ou à força que nos move, que nos forjou e moldou. A matemática é muito complexa para justificar a nossa existência neste plano. Com certeza, existe um engenheiro por trás da criação.

Agradeço, acima de tudo e todos, aos meus pais, Admilson e Edileuza. Eles que nunca falharam em me dar amor, carinho, atenção e suporte, e me deram o prazer e a sorte de estarem presentes em tudo e por tudo. Desde o início da minha vida, as dificuldades enfrentadas, a troca do curso de veterinária pela engenharia. À minha mãe, a minha força motriz, o pilar de nós todos, sempre comigo nessa caminhada, todo o meu amor, orgulhá-la é o meu maior objetivo de vida. E ainda que as escolhas da vida tenham levado meu pai deste lado, sei que ele está comigo, comemorando mais este passo. A eles eu devo tudo que sou, e não há dinheiro no mundo que pague isto.

Aos meus irmãos, Samuel e Felipe, pelo amor, apoio e companheirismo. Não conseguirei nunca expressar a sorte que tive de, além de irmãos, ter os melhores amigos que poderia ter sempre, desde a infância até a vida adulta. Esta vitória também é de vocês.

À minha família, meus avós, tios, primos, sobrinha. Aos que acreditaram em mim, neste caminho que venho trilhando. Muito obrigado.

À Larissa, minha namorada e companheira, por todo o amor e carinho, pelo companheirismo, atenção e afeto. Você é parte especial da minha vida. Obrigado sempre por tudo o que faz por mim.

Aos amigos, novos e antigos. Aos amigos de infância, dos laboratórios, do AP, da JKC. Em especial, aos amigos de caminhada, por sua amizade incondicional, Murilo (Piguiroso), Julliana Melo, Daladier (Magão), Marlon Rocha, Daniel (Bebê), Dailton, Nilo Dantas, Carolina Batatinha, Gildeon, Anderson (Bodão), Francisco (Kiko). Sem vocês eu o mundo não teria tantas cores.

Aos professores, mestres e amigos, desde a escola. Aos professores das engenharias, e do colegiado de Engenharia de Computação. Aos professores Rosalvo Neto, pelos conselhos, puxões de orelha e orientação. A Daniel Costa e Rodrigo Ramos, orientadores e parceiros nesta empreitada. Ao professor Fábio Nelson, por toda a ajuda que me prestou.

O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo.
Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas
admiráveis.

José de Alencar

RESUMO

A produção vinícola no Vale do Submédio São Francisco vem crescendo a um ritmo constante, representando 15% da produção brasileira. As características geográficas da região, juntamente com a irrigação a tornam perfeita para a produção de vinhos, sendo o único lugar no mundo com 2,5 safras anuais. A determinação de atributos de qualidade nas uvas é uma importante etapa para a definição da melhor época de colheita, e os vinhos produzidos com as uvas nestas condições apresentam um maior valor agregado. Neste cenário, o presente trabalho visa validar um método de estimação de atributos de qualidade através de processamento de imagens RGB, processo digital de representação de imagens através de intensidades de cores primárias vermelha, verde e azul, e modelos de regressão utilizando *random forest*. O trabalho foi desenvolvido com base no processamento de imagens através de extração das intensidades de iluminação dos *pixels*, de regressão com *random forest* e redes neurais artificiais do tipo *multilayer perceptron*, para criação de modelos e estimação dos valores dos atributos de qualidade, e a comparação dos dois métodos de regressão, com análise dos resultados individuais dos mesmos juntamente ao método tradicional. Verificou-se a paridade entre os métodos estudados, e o seu melhor desempenho em relação ao método tradicional, conseguindo valores de MSE e coeficientes de correlação melhores.

Palavras-chave: Processamento de imagens, *random forest*, redes neurais, métodos não-destrutivos, atributos de qualidade.

ABSTRACT

Winery production in Submédio São Francisco Valley grows in a stable rythm, being responsible by 15% of Brazil wine's production. Geographics features region combined with irrigation makes it the perfect place to wine production, being the only place in the world with 2.5 harvests per year. Quality attributes determination in grapes are an important step to define ideal harvest season and the wine produced with the grapes harvested in the ideal moment have a higher finacial value added. In this scenario, the current work searches to validate an estimation method of quality attributes through RGB images processing, a digital process of image representation by primary collors red, green and blue, and regression models using random forest. The work was developed based in image processing for pixels intensity values extraction, random forest and multilayer perceptron neural networks regressions, for model creation and quality atributes values estimation, and comparison of the two studied methods alongside with the traditional method. It was verified the parity between studied methods, and their better performance in relation to the traditional method, with better MSE and correlation coefficients values.

Key-words: *Image processing, Random Forest, Python, Non-Destructible Methods, Quality Atributes.*

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de funcionamento do <i>random forest</i>	26
Figura 2 – Diagrama de funcionamento do <i>multilayer perceptron</i>	27
Figura 3 – Sistema de Aquisição das imagens.	28
Figura 4 – Amostras de imagens obtidas no processo.	28
Figura 5 – Processo de Vetorização de imagem da uva.	29
Figura 6 – Extração do raio da uva.	30
Figura 7 – Comparação Íris-Uva.	31
Figura 8 – Processo de Vetorização de imagem da uva.	32
Figura 9 – Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para SST.	35
Figura 10 – Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para Antocianinas.	35
Figura 11 – Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para Flavonóides.	36
Figura 12 – Gráficos de dispersão para os métodos de determinação dos atributos de qualidade.	38

LISTA DE QUADROS

LISTA DE TABELAS

Tabela 1 – Resumo de Trabalhos.	23
Tabela 2 – Descritores estatísticos para <i>Cabernet Sauvignon</i>	34
Tabela 3 – Desempenho de MSE dos métodos para SST.	36
Tabela 4 – Desempenho de MSE dos métodos para Antocianinas.	36
Tabela 5 – Desempenho de MSE dos métodos para Flavonóides.	37
Tabela 6 – Desempenho de r_P dos métodos para SST.	37
Tabela 7 – Desempenho de r_P dos métodos para Antocianinas.	37
Tabela 8 – Desempenho de r_P dos métodos para Flavonóides.	37

LISTA DE ABREVIATURAS E SIGLAS

MLP	<i>Multilayer Perceptron</i> (Perceptron Multicamadas).
SST	Sólidos Solúveis Totais
MSE	<i>Mean Squared Error</i> (Erro Quadrático Médio)
RBG	<i>Red, Green and Blue</i> (Vermelho, Verde, e Azul)
PLS	<i>Partial Least Squares</i> (Mínimos Quadrados Parciais)
PCR	<i>Principal Components Regression</i> (Regressão de Componentes Principais)
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
PCA-BPNN	<i>Principal Components Analysis with Back Propagation Neural Networks</i> (Análise de Componentes Principais - Redes Neurais com Propagação Reversa)
NIRS	<i>Near Infrared Spectroscopy</i> (Espectroscopia Próxima do Infravermelho)
SECV	<i>Standard Error of Cross Validation</i> (Erro Padrão de Validação Cruzada)
SEP	<i>Standard Error Prediction</i> (Predição de Erro Padrão)
LS-SVM	<i>Least Squared Support Vector Machine</i> (Mínimos Quadrados com Máquinas de Vetores de Suporte)
RNA	Redes Neurais Artificiais
ANN	<i>Artificial Neural Networks</i> (Redes Neurais Artificiais)
MSC	<i>Multiplicative Scatter Correction</i> (Correção de Dispersão Multiplicativa)
CP-ANN	<i>Counter Propagation Artificial Neural Networks</i> (Redes Neurais Artificiais com Contra-Propagação)
HSI	<i>Hue, Saturation and Intensity</i> (Matiz, Cor e Intensidade)
TA	<i>Tritatate Acidity</i> (Acidez Tritratada)
ROI	<i>Region Of Interest</i> (Região de Interesse)
ANOVA	<i>Analysis Of Variation</i> (Análise de Variação)
VIRS	<i>Visible Infrared Spectroscopy</i> (Espectroscopia do Infravermelho Visível)

SUMÁRIO

1	INTRODUÇÃO	12
1.1	CARACTERIZAÇÃO DO PROBLEMA	13
1.2	OBJETIVOS GERAIS	13
1.3	OBJETIVOS ESPECÍFICOS	13
1.4	ORGANIZAÇÃO DO TRABALHO	13
2	REFERENCIAL BIBLIOGRÁFICO	15
2.1	O ESPECTRO NA DETERMINAÇÃO DE ATRIBUTOS EM ALIMENTOS COM O USO DA INTELIGÊNCIA ARTIFICIAL	15
2.2	O USO DE PROCESSAMENTO DIGITAL DE IMAGENS E INTELIGÊNCIA ARTIFICIAL NA OBTENÇÃO DE CARACTERÍSTICAS EM ALIMENTOS	19
3	MATERIAIS E MÉTODOS	24
3.1	LINGUAGENS DE PROGRAMAÇÃO	24
3.1.1	A linguagem <i>Python</i>	24
3.1.2	A linguagem <i>R</i>	25
3.2	O <i>RANDOM FOREST</i>	25
3.3	REDES NEURAIS E O <i>MULTILAYER PERCEPTRON</i>	26
3.4	CRIAÇÃO DA BASE DE DADOS	27
3.4.1	Coleta das amostras	27
3.4.2	Criação de <i>Folds</i> para experimento	28
3.4.3	Processamento digital das imagens	29
3.4.4	Criação dos vetores de <i>pixels</i>	31
3.5	ANÁLISE FÍSICO-QUÍMICA	32
3.6	CRIAÇÃO DOS EXPERIMENTOS PARA A REGRESSÃO	33
3.7	CRIAÇÃO DOS MODELOS E ESTIMAÇÃO DOS VALORES DOS ATRIBUTOS	33
4	RESULTADOS	34
4.1	AVALIAÇÃO DOS MODELOS	34
4.2	AVALIAÇÃO DE DESEMPENHO DOS MODELOS	34
5	CONCLUSÕES	39
5.1	CONSIDERAÇÕES FINAIS	39
5.2	LIMITAÇÕES	39
5.3	TRABALHOS FUTUROS	40
	REFERÊNCIAS	41

1 INTRODUÇÃO

Na região do Vale do Submédio São Francisco, a agricultura irrigada tem papel transformador, fortalecendo a economia e gerando emprego e renda. Desde a década de 1950, diversos esforços são feitos na região visando seu desenvolvimento. A partir de então, e aliado a incentivos para o desenvolvimento de novas culturas na região, formou-se o Polo Fruticultor do Vale do São Francisco (HORA; XAVIER; MATTOSINHO, 2016). Hoje, este polo exporta mais de 90% da sua produção para Europa, Estados Unidos e Japão, exercendo papel central na produção, anteriormente só encontrado nas capitais (ARAÚJO; SILVA, 2013).

Este cenário, juntamente com o clima semiárido da região, com poucas chuvas e sol abundante, tornou-a ideal para o cultivo de uvas, sendo o único lugar no mundo em que as videiras produzem, em média, duas safras e meia ao ano (ZANINI; ROCHA, 2010). Cultivada por milhares de anos pelas civilizações ao redor do mundo, a vinicultura acabou diversificando-se e dando origem às mais diversas variedades, tanto de mesa quanto para a produção de vinhos (PIRES, 2010).

Atualmente, o Polo Fruticultor do Vale do São Francisco corresponde a 15% do total de produção de vinhos brasileiros, e continua em crescimento (SOUSA et al., 2012). Ainda assim, de acordo com Bassoi et al. (2014), pouco se é investido em tecnologias agrícolas pelos produtores. Ao uso destas tecnologias na agricultura deu-se o nome de *Agricultura de Precisão*. Nesta, é empregada um conjunto de ferramentas e tecnologias com o objetivo de tomar melhores decisões com base em variações da área de produção, tanto geograficamente quanto de forma temporal, diminuindo custos, aumentando a produtividade e reduzindo o impacto ambiental na atividade agrícola (MOLIN; AMARAL; COLAÇO, 2015).

Segundo Binfare, Galvão e Castro (2016), a região do Vale do São Francisco apresenta em média 300 dias de sol por ano, e isto proporciona uma grande quantidade de açúcares nas uvas. A união de todos os fatores já descritos evidenciam o enorme potencial da região para este tipo de produção e o polo vitivinícola, que já vinha se consolidando, hoje é uma das forças motrizes da economia regional, e investimentos na área de vinicultura de precisão começam a ser feitos para a melhoria da qualidade dos bens produzidos na região.

Desta forma, a criação de tecnologia local é uma situação ideal para a produção, visto que proporciona um menor custo ao produtor e aumenta a sua produtividade, por um lado, e cria novas possibilidades de empreendedorismo e desenvolvimento da região e do seu entorno.

1.1 CARACTERIZAÇÃO DO PROBLEMA

A determinação do ponto de colheita é um dos fatores que melhoram a qualidade do vinho. Atualmente, para a definição do ponto de colheita das uvas, as vinícolas utilizam processos físico-químicos, com a determinação do ponto ótimo de maturação dado principalmente por três processos: Acidez Titulável Total (ATT), pH e teor de açúcar. A obtenção destes valores é feita em laboratório, e a relação entre os três determina características como sabor, coloração e a fixação de componentes necessários a seu sabor (SOUZA et al., 2013). Estes processos, além de exigirem profissionais qualificados (técnicos laboratoriais), demandam tempo e dinheiro, que são fatores importantes para os produtores, e tornam-se ainda mais importantes no caso do agricultor familiar ou do pequeno produtor.

Este trabalho é a validação de um método de determinação de atributos de qualidade (sólidos solúveis totais, antocianinas e flavonóides) em uvas (*Vitis vinifera L.*), da variedade *cabernet sauvignon*, utilizando *random forest* e redes neurais artificiais.

1.2 OBJETIVOS GERAIS

Utilizar o método de determinação de atributos de qualidade em uvas da casta *cabernet sauvignon* através de análise de imagens RGB, utilizando regressão através de técnicas de inteligência artificial, sendo elas *random forest* e redes neurais artificiais do tipo *multilayer perceptron* - MLP.

1.3 OBJETIVOS ESPECÍFICOS

- Criar nova base de dados a partir de imagens de bagas de uva da variedade *Cabernet Sauvignon*;
- Submeter a nova base ao processo de criação de modelos de regressão do *random forest* e redes neurais artificiais;
- Estimar os valores dos atributos de qualidade;
- Avaliar desempenho do método na casta analisada;

1.4 ORGANIZAÇÃO DO TRABALHO

A seguir, o trabalho foi dividido em referencial teórico e materiais e métodos. No referencial teórico, temos uma introdução de todos os conceitos abordados no trabalho, desde os conceitos de processamento digital de imagens para a agricultura, passando por suas aplicações, e finalmente, as tecnologias empregadas.

Na seção de materiais e métodos é mostrado todo o processo, desde a coleta das amostras,

como será feito o processamento das imagens para a criação da base de dados, o processo de treino e teste dos conjuntos de dados e, posteriormente, como será gerada a saída. Para a seção de Resultados e Discussão, são mostrados os desempenhos dos dois métodos de regressão para a base de dados, com a comparação de gráficos e tabelas obtidas. A seção de Conclusão aborda as considerações finais, limitações e trabalhos futuros do trabalho.

2 REFERENCIAL BIBLIOGRÁFICO

2.1 O ESPECTRO NA DETERMINAÇÃO DE ATRIBUTOS EM ALIMENTOS COM O USO DA INTELIGÊNCIA ARTIFICIAL

O uso de informações espectrais, sejam esses espectros visíveis (VIRS, *Visible Infrared Spectre*, espectro infravermelho) ou próximos do invisível (NIRS, *Near Infrared Spectre*, espectro próximo do infravermelho), são usados para o estudo de atributos de qualidade em várias situações envolvendo alimentos (Nicolai, Theron e Lammertyn (2007), Xia et al. (2007), Liu, Yang e Deng (2015)). A utilização destes recursos agregados à inteligência artificial é uma área de estudo bastante abrangente.

Liu et al. (2014) avaliaram a aplicação de informações multispectrais para a previsão de qualidades principais e estado de amadurecimento de morango, e compararam a performance dos modelos de previsão obtidos usando PLS (*Partial Least Squares - PLS*, em inglês, ou mínimos quadrados parciais, em português), SVM (*Support Vector Machines*, Máquinas de vetores de suporte, em inglês) e Redes Neurais com *Backpropagation*. No trabalho, as amostras foram separadas em três categorias, sendo que das duas primeiras categorias (verdes e maduras) foram obtidas as informações multispectrais imediatamente. Cada categoria tem 70 frutos. Para as amostras da terceira categoria, 70 frutos maduros foram deixados em temperatura ambiente por dois dias, para serem categorizadas como passadas. Para a criação dos modelos de predição de atributos de qualidade, foram utilizados PLS, SVM e Redes Neurais com *Backpropagation* (BPNN), com *leave-one-out cross-validation*. Para os modelos de análise espectral para determinação de estágio de maturação, foram utilizados SVM e PCA-BPNN. No PCA-BPNN, primeiro utiliza-se o PCA em todas as regiões do espectro, para determinar algumas poucas componentes principais como entrada do BPNN. A predição de firmeza e SST com os modelos BPNN apresentou melhor performance em relação ao SVM e PLS, embora para firmeza os resultados do PLS tenham se aproximado dos obtidos com BPNN. Para a classificação do estágio de maturação do fruto, já com o comportamento do PCA, foi possível visualizar a tendência dos dados mostrando que, para uma melhor classificação, a parte VIS do espectro apresenta uma melhor relação com os dados que a parte NIS, embora a utilização das duas melhore levemente o desempenho. Com PCA-BPNN, a precisão da classificação máxima alcançada foi 93,33%, e os melhores resultados foram obtidos com SVM, com precisão de classificação máxima de 100%.

O trabalho de Chauchard et al. (2004), objetivou o estudo das capacidades de SVM com Mínimos Quadrados de entregar modelos de calibração robustos e precisos para a predição de acidez total de uvas frescas de dados obtidos com sensores NIR portáteis,

e compará-los com Regressão de Mínimos Quadrados Parciais (*Partial Least Squares Regression - PLSR*, em inglês) e Regressão Multilinear (*Multiple Linear Regression - MLR*, em inglês). Para todos os métodos de regressão utilizados, a variação das ondas foram restringidos na janela de ondas curtas NIR, entre 680-1100 nm, incluindo o pico de clorofila (685 nm) e excluindo o pico de absorção de açúcar (830-920 nm), para evitar a influência concomitante com o constituinte químico mais influente nas uvas. Para os testes portáteis, apenas algumas ondas foram selecionadas. O conjunto de dados foi dividido em conjuntos de treino e teste usando o método *Venetian blinds*, de acordo com a acidez total. Os conjuntos têm a mesma distribuição de acidez total, com os conjuntos tendo 186 e 185 amostras, respectivamente. Para cada modelo, foi calculado o erro padrão de calibração (SEC), o erro padrão do *cross validation* com *leave-one-out* (SECV) e o coeficiente de determinação do *cross validation*, todos utilizando o conjunto de testes. Para o conjunto de treino, foi calculado o erro padrão da predição (SEP), o coeficiente de determinação, o viés e o erro padrão do viés corrigido (SEPC). Para definir a robustez de cada método, foi empregada uma técnica de geração de conjuntos de testes utilizando o conjunto de testes original e modificando algumas áreas de espectro, com ruídos simulados (Gaussiano, Multiplicativo, mudança de *baseline*, declive de *baseline*, mudança, alargamento e encolhimento de comprimento de onda, e perturbação da largura de onda. O método considerado com melhor desempenho foi a combinação de LS-SVM com pré-processamento utilizando transformação com variável padrão normal (*Standard normal variate - SNV*) e variáveis latentes obtidas por PLSR. A combinação de LS-SVM e SNV nas faixas de onda selecionadas para o estudo aparentam maior robustez se comparadas com o modelo MLR.

Ainda com o uso de informações espectrais e inteligência artificial, Chia, Rahim e Rahim (2011) estudaram a implementação de ANN (*Artificial Neural Network*, rede neural artificial) para predição de sólidos solúveis totais - SST de maçãs através de análise de espectroscopia VIS-SWNIR (*Visible and Shortwave Near Infrared*, ondas visíveis e curtas próximas ao infravermelho) e compara o resultado com o de um modelo de predição linear com Regressão de Componentes Parciais (*Partial Components Regression - PCR*). Para isto, 116 amostras de maçã Fuji adquiridas em supermercado foram armazenadas por no mínimo 24 horas antes de realizar a aquisição do espectro, para evitar qualquer alteração significativa de temperatura entre os indivíduos. Posteriormente, as maçãs foram lavadas, identificadas individualmente e foram separadas aleatoriamente entre o conjunto de treino (82 amostras) e o conjunto de testes (32 amostras). Foi realizada a aquisição de valores de espectro e, logo em seguida, a determinação de SST pelo método destrutivo convencional. Depois de pré-processados os dados, foi realizada uma redução de dimensionalidade com PCA (*Principal Component Analysis*, análise de componentes principais), de forma que poucas componentes principais pudessem representar o conjunto de espectros da amostra. Escolhida a melhor quantidade de componentes principais, as mesmas foram utilizadas como entradas para os modelos de MLR (*Multiple Linear Regression*, regressão linear múltipla)

e da rede neural. A arquitetura da rede neural foi feita da seguinte forma: foi utilizada na camada oculta uma função log-sigmóide, na camada de saída foi escolhida a transferência linear, o algoritmo de *backpropagation* da rede de treinamento foi o Levenberg-Marquardt, o gradiente descendente com pesos de *momentum* e função de aprendizagem com viés para a função de aprendizagem de peso/viés e a função de performance escolhida foi o MSE. O número de neurônios na camada oculta foi sendo aumentado caso a performance de predição não fosse satisfeita. O estudo mostrou que a rede neural com componentes principais (PCs-ANN) tem desempenho melhor que os modelos de predição linear para a predição de SST em maçãs, sendo que o método com dois neurônios alcançou melhores resultados que o método com somente um neurônio na camada oculta.

Janik et al. (2007), ainda na área de análise com espectro VIS, demonstrou as vantagens da ANN para determinar a concentração de antocianina em uvas vermelhas, e comparação do método de VIS-NIRS e os dados de referência, sendo estes obtidos com o método tradicional e o uso de regressão PLS (PLSR). Inicialmente, 3134 amostras de uva foram obtidas de 1999 até 2003, sendo estas o conjunto de treino e 250 uvas de 2004 para o conjunto de testes, compreendendo uvas de 9 variedades e 11 regiões diferentes. Depois, foi realizado o processo de homogeneização das amostras, com congelamento e descongelamento das amostras, extração de sólidos solúveis com uma mistura de álcool, e preparação para caracterização da espectroscopia. A obtenção dos dados de espectroscopia foram obtidos em comprimentos de onda de 400-2500 nm. No processo, as amostras homogeneizadas foram escaneadas com espectrômetro VIS-NIR com largura de onda variando entre 400-2500 nm com escala de 10 nm. Os valores obtidos foram então correlacionados com os valores de antocianina de referência utilizando o software *Unscrambler* e, para o treino com PLS, foi exportado para um formato intermediário. O espectro não foi pré-processado, com exceção de centralização de média e redução de granularidade por média de cinco pontos, embora um conjunto de treino com pré-processamento utilizando derivação de Savitzky-Golay tenha sido realizada. O conjunto de treino com PLS foi realizado com o espectro total das amostras do conjunto de testes de 2004 e *cross-validation* com *leave-one-out*. A regressão com RNA foi feita utilizando como conjunto de testes as amostras de 1999-2003 divididas aleatoriamente em dois subconjuntos, sendo o conjunto de treino com 75% das amostras e o conjunto de testes com os 25% restantes. A regressão com PLS de amostras de safras anteriores apresentou por vezes imprecisão devido à alta curvatura da regressão ou do viés. Utilizando os escores do PLS como entradas para um modelo com ANN, obteve-se melhores resultados na modelagem de relacionamentos não-lineares entre a intensidade espectral e a concentração de antocianina. Apesar disso, os modelos com ANN utilizando o espectro total não obtiveram desempenho satisfatório na previsão dos valores em amostras de novas safras, mesmo com os modelos robustos, dada uma curva em formato de "S" na regressão, indicando *overfitting*. Com seleção apropriada de larguras de ondas do espectro baseadas na correlação máxima ou na covariância, o resultado melhorou a um ponto de

similaridade com os obtidos com a entrada utilizando os escores de PLS. Da mesma forma que as entradas de PLS, os modelos de RNA foram desenvolvidos de escores com PCA. Contudo, se um número maior ou menor de escores for utilizado do PCA, pode-se ter *overfitting* ou *underfitting*, que diferenciam dos valores obtidos com as entradas de PLS. O método de RNA explorando a correlação entre os escores de PLS e valores analíticos gerou um modelo de regressão estável capaz de prever a concentração de antocianina total no espectro VIS-NIS de uvas de uma nova safra.

Ainda utilizando frutas, Munawar (2014) desenvolveu uma tese de doutorado onde os objetivos foram divididos em três partes principais: Determinar a viabilidade do método NIRS para estimar os atributos de qualidade tais como SST, acidez tritratada e ácido ascórbico de manga através de um modelo de calibração multivariado seguido de *cross validation*; Comparar métodos de pré-processamento espectral sobre os modelos de calibração e avaliar o impacto na precisão e robustez destes nos modelos de predição; Comparar os métodos de regressão não linear (SVMR - regressão com SVM, e ANN) na previsão de atributos de qualidade do fruto e comparar com os resultados de PLSR. Foram obtidas em supermercado local 58 mangas Kent de três países (Brasil, Espanha e Israel). Estas foram armazenadas em temperatura ambiente de 25°C e foram coletadas amostras de SST, TA (*Tritratate acidity*, acidez tritratada) e AA (ácido ascórbico) de cada amostra a cada dois dias, totalizando seis amostras para cada indivíduo. O dia zero foi o dia da compra das mangas e, para todos os dias de amostras, foram analisados 10 indivíduos, com exceção do último dia, em que só sobraram 8 indivíduos. As amostras de NIR foram obtidas antes das análises destrutivas e foram feitas no mesmo dia ou no máximo com um dia de diferença. Após coleta do espectro, foram cortadas individualmente na mesma localização de acordo com os valores de NIR, e foi retirada a polpa. Primeiro, foi analisada a concentração de ácido ascórbico, visto que ele pode ser oxidado pelo oxigênio presente na atmosfera. Cinco gramas da amostra da polpa foram maceradas e misturadas com 20 ml de ácido metafosfórico a 5% em uma proveta para prevenir a oxidação, e depois submetido ao processo de tritratção. O ácido ascórbico presente foi quantificado de acordo com sua reação com a solução de tritratamento, sendo o indicador do fim da reação a coloração vermelha (ou rósea). A quantificação de SST e TA foi feita simultaneamente, fazendo um suco com 20 gramas de polpa e no máximo 100 ml de água. O suco foi então submetido à centrífuga, e uma gota desse suco foi retirada e analisada através de um refratômetro analógico para determinação de SST em °BRIX. Para o TA, foi realizada tritratção automática com NaOH e determinada em mg.100g⁻¹. Os três atributos foram então medidos novamente e o valor utilizado foi a média das duas medições. No trabalho, foi estabelecido que a predição de atributos de qualidade principais da manga poderia ser obtida com qualidade através de NIRS. Pré-processamentos como a derivação, SNV e MSC (*Multiplicative Scatter Correction*, correção de dispersão multiplicativa) afetam a qualidade dos modelos de predição utilizando PCR e PLSR, sendo que o melhor modelo

de predição para os três atributos medidos foi alcançado usando o modelo PLSR com espectro não derivado. Pode-se assumir que, apesar do fato do coeficiente de determinação observado nos modelos de predição de SST e AA não terem sido muito altos, eles ainda podem ser usados na seleção e classificação da qualidade das mangas.

Em uma outra abordagem, dessa vez com derivados de carne, Prevornik et al. (2014) testa a classificação com ANN de presuntos curados a seco em diferentes classes de maturação baseado em informações espectrais obtidas dos mesmos. Em três fábricas de processamento de presunto comercial, foram coletados 128 presuntos curados a seco, obedecendo as especificações de produção dos mesmos. Para cada peça, foram cortadas fatias de 2 cm, que foram congeladas, processadas e moídas até virarem um pó fino. As amostras então foram submetidas ao processo de coleta de dados da espectroscopia, com comprimento de onda variando entre 400-2500 nm, com separação a cada 2 nm. Para o estudo, foram utilizados o método Konohen não-supervisionado de ANN e o método supervisionado ANN com *counter-propagation* (CP-ANN). Os valores de espectroscopia foram comprimidos em 262 pontos pela média sequencial de quatro médias distintas, e estes pontos foram usados como variáveis para o classificador Konohen de 4 x 4 neurônios com 100 épocas, sendo que cada teste foi feito com 32 amostras por vez. Foi gerado então, depois do treino, um mapa de Konohen, que é um grid com a quantidade de neurônios como suas dimensões, onde cada amostra foi espalhada. Assim, entre as amostras localizadas na região de um neurônio, aquelas que estivessem mais próximas eram selecionadas para o conjunto de testes e treino, onde cada amostra foi então separada de acordo com seu estágio de maturação (10, 12, 14 ou 16 meses), no final totalizando 64 amostras para o conjunto de teste e 64 para o de treino. A precisão da predição com o método desenvolvido pelos autores deste projeto foi de 79,7%, com 14,1% de incerteza e sem falsos positivos, utilizando uma rede de 12 x 12 neurônios com 200 épocas. Isto portanto indica uma relevância prática do uso do método para classificação de presunto do tipo *Kraški pršut*, sendo que, para outros tipos de presunto, seria necessária a validação do uso prático do mesmo.

2.2 O USO DE PROCESSAMENTO DIGITAL DE IMAGENS E INTELIGÊNCIA ARTIFICIAL NA OBTENÇÃO DE CARACTERÍSTICAS EM ALIMENTOS

O uso de imagens RGB para a extração e predição de características de alimentos é um ramo de pesquisa que atualmente tem crescido e apresentando bons resultados, como é possível verificar nos trabalhos de Nambi et al. (2016), Savakar e Anami (2009) e Pereira et al. (2018). Savakar e Anami (2009), em seu trabalho, estudaram o reconhecimento e a classificação de produtos agrícolas (grãos, frutas e flores) indianos a granel através de análise de amostra de imagens obtidas em produções de agricultura/horticultura. As amostras foram coletadas em diferentes localizações no distrito de Bijapur, do estado

de Karnataka, Índia, durante a safra de 2007 do Mercado Produtor de Agricultura e da Faculdade de Ciências Agrícolas de Bijapur. As imagens foram obtidas com auxílio de câmera digital, com iluminação em posição de 45° de inclinação em relação à mesa. Inicialmente, foi feita a extração dos valores de RGB das imagens coletadas e, a partir desses valores, foram calculados os valores de HSI (*Hue, Saturation, Intensity* - matiz, saturação e intensidade das cores) para cada amostra. Foi então calculada a média, variância e a extensão dos *pixels* das amostras. Depois, foi feito o cálculo com a escala de cinza para determinar os parâmetros de textura dos grãos, utilizando uma matriz de co-ocorrência. Desses valores são extraídos então média, variância, extensão, energia, probabilidade máxima, contraste, momento de diferença inversa e correlação. Foi utilizada uma rede neural com *back-propagation*, onde os valores obtidos foram usados como variáveis de entrada, chegando a um total de 44, e o número de neurônios da saída corresponderam à quantidade de classes existentes, com um total de 10 padrões de saída. As 400 amostras de cada conjunto foram divididas em duas metades, onde 200 imagens foram usadas para treino e outras 200 para testes, totalizando, para todos os conjuntos, 4000 (2000 para testes e 2000 para treino). Do conjunto de treino, foram retiradas 15% das amostras para validação do classificador. Foi proposto o método de classificação utilizando características de cor da imagem, de textura e a combinação dos dois para maior robustez, sendo este último o método proposto no trabalho. Utilizando as características de cor, a precisão foi de 87,5%, 78,4% e 75,7% para classificação de grãos, mangas e flores de jasmin, respectivamente. Com o uso de características de textura, a precisão aumentou, com valores de 90,8%, 80,2% e 85,8% para os mesmos produtos respectivamente. Quando combinados os dois conjuntos de características, obteve-se o desempenho máximo do sistema, com 94,1%, 84% e 90,1% de precisão. Acredita-se, com isso, que o uso do método proposto é válido para a classificação desses produtos, substituindo a classificação feita manualmente, com menor suscetibilidade a falhas.

Na área de produção animal, Chen et al. (2010) exemplificaram a viabilidade de uso da visão computacional e SVM para predição de gordura presente em carne, com foco em três pontos de interesse: separar a gordura subcutânea da ponta da costela de vaca usando análise digital de imagens; extrair as características de cor dessa gordura; desenvolver um classificador para determinar o escore da gordura da carne baseado nas características de cores propostas. Como objeto de pesquisa, cento e vinte e três peças de costelas de vaca foram compradas de um revendedor local, com representação de vários escores de gordura tipicamente encontrados nas fábricas da China, e envelhecidos por 72h a 4° C. Depois, cortes de músculo com espessura de 2,5 cm foram obtidas de cada peça, e foram estudadas suas características. As amostras foram divididas em um grupo de testes (38 amostras) e um grupo de treino (85 amostras). Seis características da cor da gordura foram extraídos das imagens (média e desvio padrão do espaço RGB), enquanto mais seis características foram calculadas (média e desvio padrão de *HUE*, saturação

e espaço de cores HSI), fazendo a conversão de RGB para HSI que foi necessária. As médias descreveram as propriedades médias da cor da gordura, enquanto os desvios padrão representam a desuniformidade da intensidade da cor em toda a gordura. O conjunto de testes serviu para testar a eficiência do classificador, onde os doze valores foram usados como entrada, e os resultados preditos foram comparados com os obtidos com métodos tradicionais de medição, e as taxas de acerto foram usadas para avaliar a performance do classificador. Então, concluiu-se que a segmentação da gordura subcutânea pode ser obtida utilizando técnicas de processamento de imagens, tais como a localização das bordas, limiarização e operações morfológicas. A melhor configuração do SVM, dado que os valores obtidos com os sensores tradicionais são de natureza não-linear com os valores obtidos pelas características extraídas do processamento das imagens, foi obtida procurando-se manualmente os melhores parâmetros, controlando a performance do classificador. A aplicação do classificador proposto resultou numa taxa de acerto de 97,4%. Assim, o escore de cor do corte de carne pode ser predito satisfatoriamente utilizando o método de visão computacional com processamento de imagens e técnicas de inteligência artificial utilizando SVM.

No trabalho de Nambi et al. (2016), foi feita a comparação de diferentes atributos de imagens extraídas de imagens RGB para melhorar a predição de estágio de maturação de manga usando algoritmos simples, e tornar o processo mais adequado para aplicações de visão de máquina de baixo custo. Na preparação do seu experimento, aproximadamente 25 kg de mangas foram coletadas em duas regiões da cidade de Tamil Nadu, Índia. Então, foram aplicados processos para amadurecimento das amostras, e estas foram coletadas de três em três aleatoriamente durante o período de amadurecimento para a análise das características. No total, 114 amostras foram obtidas no final do processo para a criação dos conjuntos de dados. A determinação dos atributos físico-químicos (SST e ácido tritratado) foram feitas com as polpas homogeneizadas das mangas, sendo o SST determinado com uso de refratrômetro digital, e TA através da tritratção com 0,1 N de NaOH. A aquisição da coloração foi feita através de medidor digital, e as características de textura através de um analisador de texturas. As imagens obtidas em câmara escura foram então processadas para separar os canais RGB, depois outro tratamento foi realizado para converter o espaço de cores para HSV. Cada canal de cor foi separado e as intensidades dos *pixels* foi medida na escala de 0 a 255. A seguir, foi utilizado o método de Otsu para encontrar a região de interesse, separando-se o fundo da ROI (*Region Of Interest*, região de interesse) através do método de limiarização. Os valores dos *pixels* dos canais RGB foram normalizados, e no final foram obtidas 16 características como entradas (Média intensidade de R, G e B com e sem normalização na ROI; Quantidade de *pixels* de R, G e B com e sem normalização; Diferença dos índices de vermelho para verde, vermelho para azul e verde para azul; Média dos valores de H, S e V). A determinação do período de classificação foi separada em cinco estágios, determinados como a qualidade de amadurecimento do fruto. Como resultado,

foi evidenciado que as características das regiões de interesse (ROI) obtidas em RGB apresentaram maior eficiência quando da predição do estágio de maturação dos frutos, sendo viável com estes resultados construir um algoritmo de visão de máquina para a previsão do estágio ou do nível de maturação da manga, com o uso de imagens RGB simples e a aplicação do algoritmo desenvolvido no próprio trabalho.

Recentemente, a classificação de estágio de amadurecimento de mamão utilizando *random forest* e de técnicas de processamento de imagens foi proposta por Pereira et al. (2018). Para esta tarefa, de cada amostra, foram obtidas duas fotos (uma para cada lado) e realizadas análises físico-químicas (firmeza da polpa, pH, sólidos solúveis, carotenóides totais e ácido ascórbico). As amostras foram classificadas em três estágios de maturação de forma visual. Depois, seguindo os escores de firmeza de polpa, foi determinado o escore para cada categoria. Foi utilizado o classificador Random Forest através da biblioteca *randomForest* em R, com 500 árvores de decisão e hiper-parâmetros padrões. Análise de Variância (ANOVA) foi realizada em *MATLAB* para a identificação das diferenças das propriedades físico-químicas entre os estágios de maturação. Para avaliação do modelo gerado pelo *random forest*, foram criados os conjuntos de treino (utilizando *cross-validation*) e teste. O conjunto de treino foi usado para fazer ajustes finos nos hiper-parâmetros do classificador, enquanto o de teste serviu para testar o desempenho. Seguindo este método então, como resultados obtidos, o primeiro e o segundo estágio de maturação foram classificados corretamente, enquanto o terceiro estágio apresentou uma dificuldade maior na classificação. Apesar disso, em todos os estágios, algumas frutas foram incorretamente preditas em pelo menos uma iteração. Para alguns frutos, não se pode validar a relação de cor de superfície e propriedades internas.

Para que se tenha um sumário geral, a Tabela 1 a seguir mostra um resumo dos trabalhos citados separados por autor, técnica utilizada no trabalho e as variáveis que foram utilizadas para o desenvolvimento do projeto.

Tabela 1 – Resumo de Trabalhos.

Autor	Técnica utilizada	Variáveis Alvo Utilizadas
Liu et al. (2014)	Regressão utilizando PLS, SVM e BPNN	Espectro entre 405-970 nm
Chauchard et al. (2004)		Espectro entre 300-1600 nm, com intervalo de 3.3 nm
Chia, Rahim e Rahim (2011)	Regressão utilizando ANN	Espectro entre 400-1000 nm com intervalo de 6 nm
Janik et al. (2007)	Regressão utilizando ANN	Espectro entre 400-2500 nm com intervalo de 10 nm
Munawar (2014)	Regressão utilizando SVM e ANN	Espectro entre 1000-2500 nm com intervalo de 2 nm
Prevolnik et al. (2014)	Classificação utilizando ANN	Espectro entre 400-2500 nm com intervalo de 2 nm
Savakar e Anami (2009)	Classificação utilizando BPNN	Média, variância e <i>range</i> de RGB, HSI, HUE, saturação e intensidade; média, variância, <i>range</i> , energia, probabilidade máxima, contraste, momento de diferença inversa e correlação das matrizes de co-ocorrência de RGB
Chen et al. (2010)	Regressão utilizando SVM	Média e desvio padrão de RGB e HSI
Nambi et al. (2016)	Regressão utilizando PCA e análise de <i>clusters</i> hierárquicos	Média de RGB, RGB normalizado e HSV; Área ou quantidade de <i>pixels</i> de RGB e RGB normalizado; Índice do diferencial normalizado de vermelho até verde, vermelho até azul e verde até azul.
Pereira et al. (2018)	Classificação utilizando <i>Random Forest</i>	Firmeza de polpa, pH, SST, carotenóides totais e ácido ascórbico

Fonte: O autor

3 MATERIAIS E MÉTODOS

3.1 LINGUAGENS DE PROGRAMAÇÃO

As linguagens de programação são a forma de comunicação entre pessoas e computadores. É por intermédio delas que o computador recebe ordens e entende o que precisa ser feito (TUCKER; NOONAN, 2009). As linguagens de programação modernas surgiram da necessidade de facilitar esta comunicação, visto que a linguagem de máquina, que é essencialmente a que o computador entende, era muito complexa para se trabalhar e, para situações mais complexas, tornava o trabalho quase impossível. Assim, as linguagens de programação modernas possuem ferramentas e facilidades, que tornam o desenvolvimento mais simples e prazeroso (GUDWIN, 1997).

Este trabalho utiliza duas linguagens de programação: *Python* e *R*.

3.1.1 A linguagem *Python*

Python é uma linguagem de programação que começou a ser desenvolvida em 1990 por Guido van Rossum. É uma linguagem interpretada, interativa e orientada a objetos, totalmente modularizada. Sua construção é simples, podendo ser expandida por uma infinidade de bibliotecas, que podem ser adicionadas facilmente ao sistema (SANNER et al., 1999).

Para a manipulação de imagens utilizando esta linguagem, foi desenvolvida a biblioteca *OpenCV* (*Open Source Computer Vision*, Visão Computacional em Código Aberto). A *OpenCV* é uma biblioteca de código aberto e *cross-platform*, isto é, pode ser utilizada em várias plataformas de desenvolvimento diferentes, proporcionando ferramentas para aplicações e experimentos de visão computacional, através do uso de abstrações de detalhes de implementação de *hardware* e *software* (HOWSE, 2013).

Visão Computacional abrange, além da manipulação de imagens, técnicas que procuram emular a visão humana, tendo como entrada uma imagem, e sua saída uma interpretação da mesma como um todo ou parcialmente (MARENGONI; STRINGHINI, 2009). Nos processos de visão computacional, segundo Gonzales e Wintz (1987), temos três níveis principais: baixo-nível, médio-nível e alto-nível. O baixo-nível compreende as mais básicas manipulações, como ajuste de contraste ou cores da imagem. O médio-nível abrange as técnicas de segmentação de imagens, como recortes ou separação de camadas, e também o reconhecimento de formas e objetos. O alto-nível são os processos relacionados às tarefas humanas, extraíndo as características da imagem para a obtenção dos dados desejados.

Pela facilidade de trabalho, uma curva de aprendizado suave, documentação abrangente e facilidade de trabalho com diferentes conjuntos de dados, a linguagem foi utilizada neste trabalho.

3.1.2 A linguagem *R*

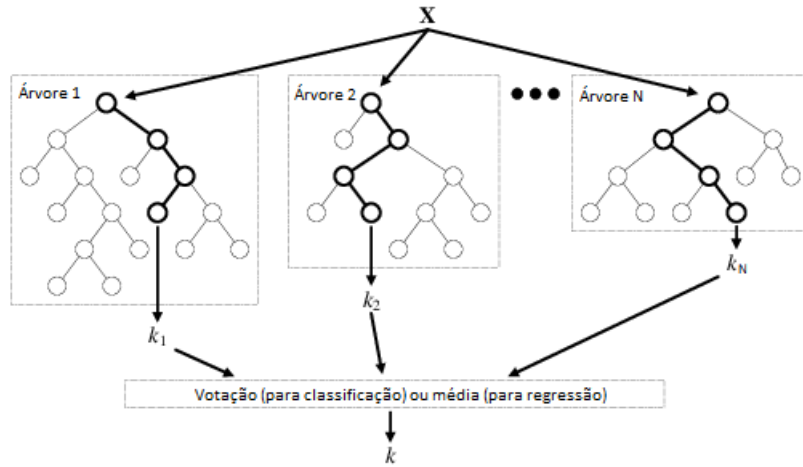
A linguagem *R*, por sua vez, foi desenvolvida como um subsistema de computação para estatística e geração de gráficos. De acordo com Torgo (2016), foi inicialmente desenvolvida em 1996 por Ihaka e Gentleman, da Universidade de Auckland, Nova Zelândia, baseada na linguagem S desenvolvida nos laboratórios da AT&T Bell, e desde o seu surgimento, é utilizada amplamente no meio acadêmico para análises estatísticas. Devido ao seu alto poder de customização, pode-se trabalhar uma infinidade de possibilidades, desde o uso de ferramentas e pacotes já desenvolvidos, ou mesmo o desenvolvimento de novas técnicas pelos usuários (TEAM, 2000). A inserção da inteligência artificial, mais especificamente da mineração de dados na linguagem *R*, foi um passo natural visto que, por sua definição, a mineração de dados visa extrair informação e conhecimentos novos e úteis através de conjuntos de dados, o que é compartilhado por diversas outras áreas, como a estatística. Assim, o ambiente *R* torna-se um poderoso aliado nesta categoria (TORGO, 2016).

3.2 O *RANDOM FOREST*

O método de classificação e regressão denominado *random forest* foi proposto por Breiman (2001), e é denominado *ensemble* de classificadores, ou seja, uma conjunção dos mesmos. O *random forest* é baseado na técnica de *ensemble* chamada *bagging*. Nesta, várias árvores de classificação são criadas, e a partir destas são gerados pesos na decisão tomada por cada uma a partir do conjunto de dados, de forma independente, e no final é dada a predição a partir do voto da maioria das árvores (LIAW; WIENER et al., 2002).

Para o *random forest*, é adicionado um componente de aleatoriedade onde, além de construir árvores usando partes do conjunto de dados, muda-se a forma como as árvores de decisão ou regressão são montadas. Enquanto nas árvores de decisão tradicionais os nós são construídos baseados no conjunto de melhores atributos entre todas as variáveis, no *random forest* é feita a escolha dos atributos em um conjunto aleatório com o mesmo tamanho para cada árvore de decisão da floresta (LIAW; WIENER et al., 2002). Apesar de contra-intuitivo e de apresentar resultados pouco satisfatórios quando tomamos uma árvore em separado, as árvores em conjunto apresentam robustez, e uma qualidade de resultados considerado satisfatório (BREIMAN, 2001). O diagrama da Figura 1 apresenta o funcionamento do *random forest*, onde k_1 , k_2 , k_N são os melhores valores estimados por cada árvore, e k é o valor final obtido.

Figura 1 – Diagrama de funcionamento do *random forest*.



Fonte: Adaptado de Internet¹

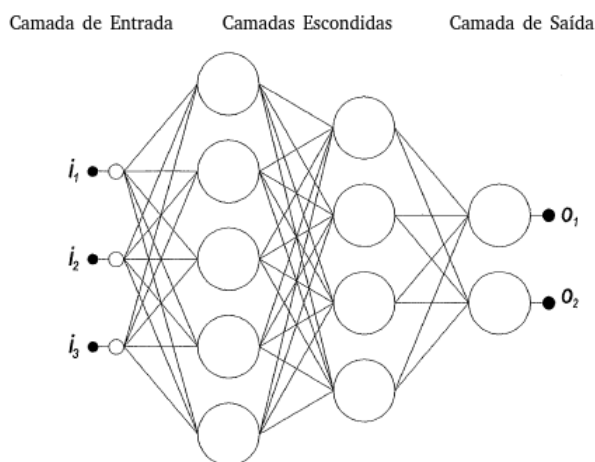
3.3 REDES NEURAIS E O *MULTILAYER PERCEPTRON*

A inteligência artificial tem por premissa básica simular o comportamento humano e sua capacidade de decisão e escolha. Uma rede neural artificial é um sistema desenvolvido para emular o cérebro humano, através de dispositivos físicos ou técnicas de programação, utilizando-se de neurônios para a criação de uma rede complexa e capaz de tomar decisões. Para este estudo, foi utilizado um determinado tipo de rede neural artificial, chamada de *multilayer perceptron* (HAYKIN, 2007).

O *multilayer perceptron* é uma rede neural baseada em retropropagação utilizando neurônios do tipo *perceptron*. O primeiro neurônio do tipo *perceptron* foi desenvolvido em 1958 por Rosenblatt. O seu funcionamento se dá com um conjunto de entradas com valores alimentando o sistema composto por camadas compostas de neurônios interligados entre si, e através de um sistema de pesos onde, a partir destes, se forma uma combinação linear de onde é possível obter uma camada de saída única, através de uma função não-linear de ativação dos neurônios. A Figura 2 mostra uma rede do tipo *multilayer perceptron*, onde i_1 , i_2 e i_3 são vetores de entrada e k_1 , k_2 , vetores de saída (GARDNER; DORLING, 1998; PANCHAL et al., 2011).

¹ Disponível em: <<https://goo.gl/YeQHhw>>. Acesso em out. 2018.

Figura 2 – Diagrama de funcionamento do *multilayer perceptron*.



Fonte: Adaptado de Gardner e Dorling (1998)

3.4 CRIAÇÃO DA BASE DE DADOS

A construção da base de dados dá-se pela coleta das uvas, aquisição das imagens, pré-tratamento e obtenção dos vetores de *pixels*. Estes foram então utilizados como entrada para a criação dos modelos de regressão.

3.4.1 Coleta das amostras

As amostras de uvas da casta *Cabernet Sauvignon* (*Vitis vinifera* L.) foram obtidas em um vinhedo produtor da região do Vale Submédio São Francisco, no município de Lagoa Grande, BA (9.05363; -40.19868). No total, foram obtidas 576 bagas que foram submetidas ao processo de análise.

Foram realizadas oito coletas direto no parreiral, onde foram escolhidas em cada coleta doze bagas de seis cachos diferentes, e de cada parreira eram escolhidos dois cachos. Cada coleta foi realizada com separação temporal de uma semana, com início em 06/07/2017, e fim em 21/08/2017.

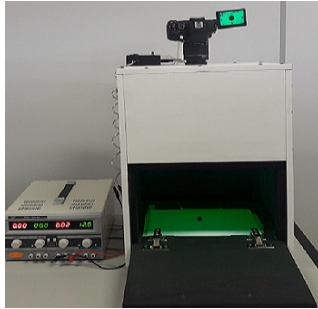
A obtenção das imagens das uvas, de forma individual, foi feita em uma caixa especialmente projetada para este fim, no Laboratório de Energia na Agricultura (LENA) da Universidade Federal do Vale do São Francisco (UNIVASF). Esta é, no seu interior, pintada de preto fosco, para que o ruído gerado pela reflexão das luzes na uva fosse mínimo. Acima da caixa ficava posicionada a câmera de marca e modelo Canon EOS REBEL T5i, com velocidade ISO-100, escala f/5,6, distância focal de 48 mm e tempo de exposição de 0,8s. As uvas, dentro da caixa, eram iluminadas por trios de LEDs *solderless* (sem solda) de marca e modelo CREE (*CREE LED Lighting Solutions Inc.*) XPE2, com temperatura de cor variando entre 3200 K a 3500 K, com 3 W de potência, nas cores vermelho, verde,

azul, branco quente e branco frio. Os leds foram dispostos com distância angular de 120° entre si.

A Figura 3 mostra a caixa, montada (Figura 3a), com a fonte de alimentação dos LEDs ligada e o sistema de iluminação totalmente ligado (Figura 3b).

Figura 3 – Sistema de Aquisição das imagens.

(a) Caixa completa.



(b) Sistema de iluminação LED.



Fonte: O Autor.

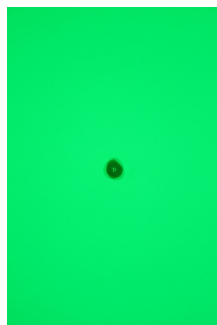
A seguir, tem-se um exemplo das imagens obtidas pelo processo com este sistema (figura 4). Estas, em alta resolução ($3456 \times 5184 \text{ pixels}$), com separação por cores (Azul, Verde, Vermelho, Branco Quente, Branco Frio). Neste trabalho, foram utilizadas as imagens Branco Quente somente.

Figura 4 – Amostras de imagens obtidas no processo.

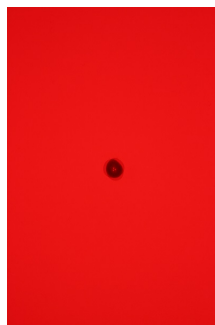
(a) Azul



(b) Verde



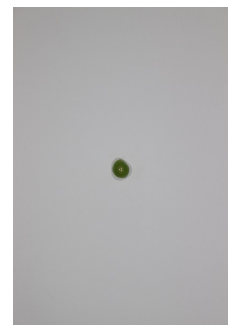
(c) Vermelho



(d) B. Quente



(e) B. Frio



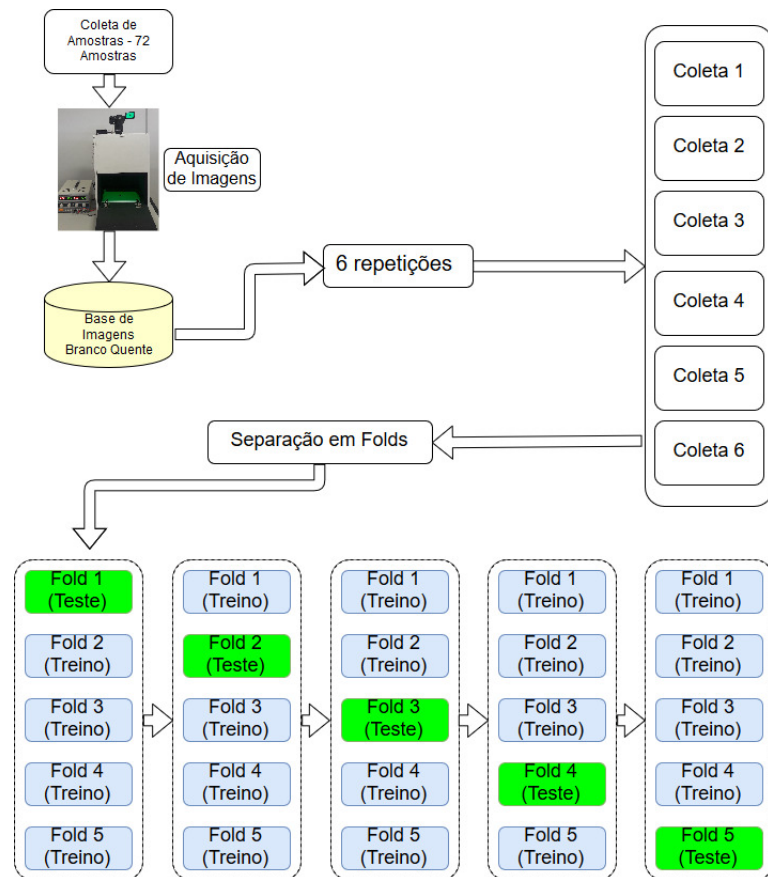
Fonte: O autor.

3.4.2 Criação de *Folds* para experimento

Após obtidas as imagens, foi feito o processo de separação das amostras, montando os experimentos para a avaliação baseado no modelo de *K-Fold Cross Validation* (Validação cruzada com *K-Fold*). Este método consiste em dividir o conjunto de dados em subconjuntos menores, independentes entre si, com indivíduos aleatórios sem repetição do conjunto principal de dados, e realizando para cada subconjunto, um experimento onde, um por

vez, cada subconjunto funcionou como conjunto de validação (testes), enquanto os outros, agregados, foram o conjunto de treino para a regressão. Assim, foram feitos K experimentos, com K sendo a quantidade de *folds* criada, gerando então K resultados diferentes para o mesmo conjunto de dados, que foram então avaliados (ARLOT; CELISSE et al., 2010). Esta abordagem foi escolhida em detrimento ao método adotado tradicionalmente, como em Pereira et al. (2018) pois, segundo Larson (1931), se realizados teste e treino com a mesma base de dados, o resultado é considerado extremamente otimista, o que pode invalidar o experimento realizado, ou gerar uma incerteza sobre seu resultado. Neste trabalho, portanto, foi escolhido o método *K-fold Cross Validation*, com K=5. Assim, os subconjuntos tem proporcionalmente o mesmo tamanho, gerando robustez estatística no processo. A figura 5 mostra o processo.

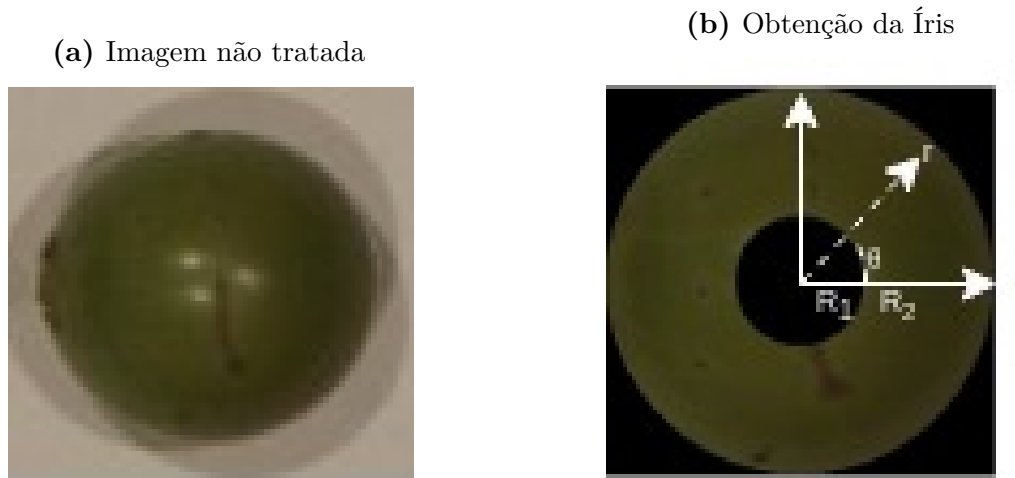
Figura 5 – Processo de Vetorização de imagem da uva.



Fonte: O autor.

3.4.3 Processamento digital das imagens

Depois deste processo, as imagens passaram por um processo de tratamento para a extração das informações. Num primeiro momento, é localizado o centro de imagem da uva, e dele traçados dois raios, onde o primeiro visou retirar a área de incidência do LED nas uvas, de forma a reduzir o ruído quando da obtenção da média dos *pixels*, sendo o

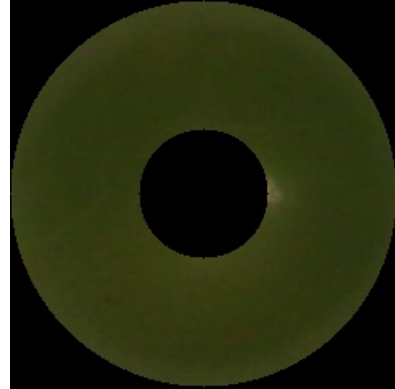
Figura 6 – Extração do raio da uva.

Fonte: O Autor.

??

processo de obtenção das médias explicados posteriormente neste trabalho, e o segundo raio define a borda exterior da uva, de modo a ter uma padronização do tamanho das imagens a serem analisadas. O primeiro raio, denominado r_1 , possui tamanho fixo, definido pela maior distância de incidência da iluminação dos LEDs na uva. A parte interna desse raio é removida da imagem, eliminando então o problema do ruído. O segundo raio, r_2 , corresponde ao limite externo da uva, de modo que é obtida para a análise a porção intermediária destes dois raios. O tamanho de r_2 também é fixo, sendo estabelecido como o tamanho da área da menor uva disponível para a análise a Figura ?? exemplifica este processo.

Depois que é feita esta separação, é feito o processo de conversão de coordenadas cartesianas para polares, utilizando o método *Rubber Sheet* de Daugman (DAUGMAN, 2004). Este método é originado para tratar da segmentação de íris, excluindo-se a área da pupila. Como é possível observar na Figura 7, a área de análise da uva se assemelha muito à íris sem a pupila, de modo que as devidas analogias e os nomes adotados seguirão, a partir deste ponto, como *pupila* e *íris* da uva.

Figura 7 – Comparação Íris-Uva.**(a)** Íris Humana.**(b)** Íris de uva.

Fonte: (a): Internet². (b): O autor.

Com a relação estabelecida, é fácil observar a aplicabilidade do método no trabalho. O que o *Rubber Sheet* faz é uma conversão de coordenadas cartesianas para polares, de forma que as coordenadas cartesianas de cada ponto da íris são transpostas para um par de coordenadas polares (r, θ) , onde r é a diferença entre o raio máximo da pupila e o raio máximo da íris, e θ é a variação angular de 0 até 2π (DAUGMAN, 2004). Portanto, considerando (x, y) coordenadas cartesianas e (r, θ) coordenadas polares, a conversão é dada pela equação:

$$I(x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \quad (3.1)$$

Onde:

- $x(r, \theta) \rightarrow (1 - r)x_p(\theta) + rx_i(\theta)$;
- $y(r, \theta) \rightarrow (1 - r)y_p(\theta) + ry_i(\theta)$;
- x e y representam as coordenadas cartesianas na região da íris;
- x_p e y_p representam os limites das coordenadas da pupila junto a θ ;
- x_i e y_i representam os limites das coordenadas da íris junto a θ ;

3.4.4 Criação dos vetores de *pixels*

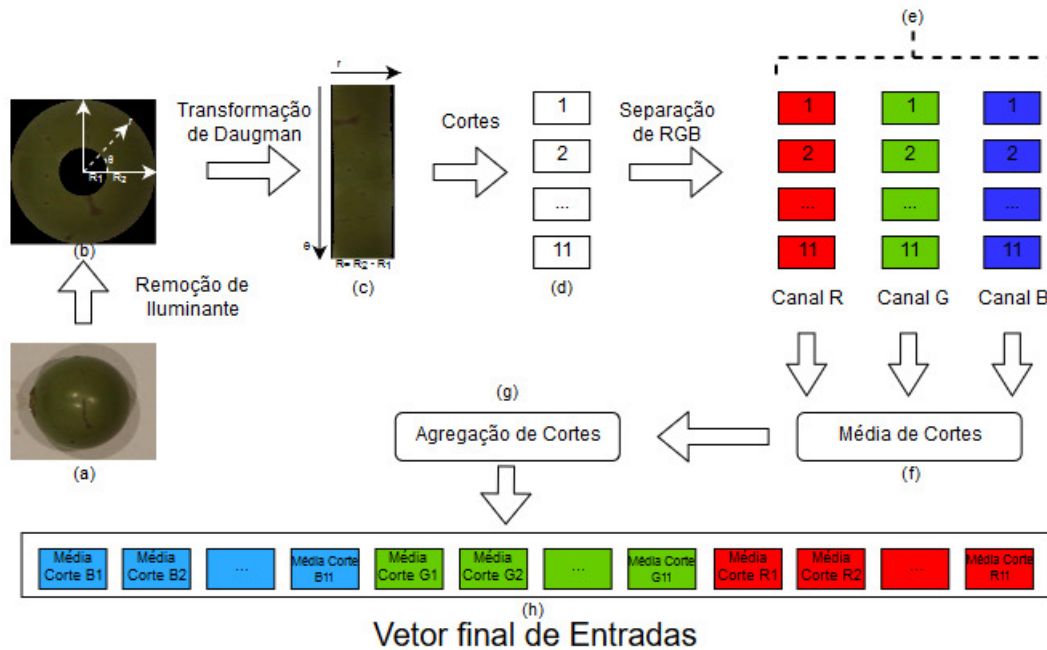
Depois de convertidas as imagens utilizando o método de Daugman, as novas imagens obtidas passaram por um processo de segmentação, onde cada imagem é subdividida em N outras imagens menores, com o objetivo de ampliar o espaço de amostragem característica das uvas, reduzindo a quantidade de valores de entrada para o modelo, se comparado com o método tradicional de análise. Para este trabalho, devido ao tamanho

² Disponível em: <<https://goo.gl/vwcKXq>>. Acesso em out. 2018.

das imagens resultantes, foi adotado o valor de N igual a 6, para que assim se obtenham segmentos com quantidades de *pixels* iguais, e regiões com tamanhos significativos para a determinação.

A partir deste ponto, já temos as imagens da forma ideal para a obtenção dos vetores de dados. Ainda utilizando a biblioteca *OpenCV*, com a linguagem *Python*, foi possível extrair a intensidade de cada cor primária (Vermelha, Verde, Azul) de cada *pixel* separadamente, e obter três vetores de intensidade dessas cores. Obtidos estes valores em cada segmento, foi calculada então a média para cada intensidade, ou seja, cada um dos segmentos vai ter então três valores de médias, para a intensidade de cor vermelha, verde e azul em cada *pixel* totalizando, para cada uva, 18 valores que serão utilizados como variáveis de entrada para a determinação dos atributos. A Figura 8 mostra este processo.

Figura 8 – Processo de Vetorização de imagem da uva.



(a) Imagem original; (b) Remoção de iluminante e determinação de área de interesse; (c) Processamento pelo Método de Daugman; (d) Corte da uva em subregiões; (e) Separação dos espaços de cores; (f) Média dos espaços de cores dos cortes; (g) Junção dos 18 valores de médias; (h) Vetor final de entradas para a determinação dos atributos. Fonte: O autor.

3.5 ANÁLISE FÍSICO-QUÍMICA

Neste estágio, as uvas foram submetidas individualmente ao processo de obtenção dos atributos de qualidade físico-químicos. Neste passo, os frutos foram destruídos e o sumo foi analisado em um refratômetro digital HI 96804, da Hana Instruments, com *range* de medida variando entre 0% e 85%, precisão de 0,2%, e resultado expresso em % ($^{\circ}$ BRIX) (HANNA INSTRUMENTS BRASIL, 2012).

Os valores de antocianinas totais foram obtidos baseados no método proposto por Francis (1982), adaptado para este experimento. Foram coletadas 0,5 g de casca e polpa dos frutos, que foram macerados por um minuto, misturados em uma solução de etanol a 95% e HCl, numa proporção de 85:15. Após a maceração, o produto foi armazenado em recipientes e mantidos em geladeira, protegidos da luz, por 24h. Depois deste tempo, o conteúdo foi então coletado para análise em espectrômetro com faixa de 535 nm. Os resultados são expressos em $cgkg^{-1}$, obtidos através da equação:

$$X = (abs * \text{fator de diluição}) / E1\% \quad (3.2)$$

onde X é a saída desejada (antocianinas ou flavonóides), o *fator de diluição* é definido como 5000, e *E1%* é o coeficiente de extinção de 98,2% para antocianinas, e 76,6% para flavonóides.

3.6 CRIAÇÃO DOS EXPERIMENTOS PARA A REGRESSÃO

Os vetores de cada uva foram então agrupados em arquivos *.csv*, onde cada linha do arquivo corresponde a uma uva, cada um com os 18 valores das uvas, e no final de cada linha ficaram os valores de SST, antocianinas e flavonóides totais para cada uva, de forma que foram três arquivos separados: um com valores de SST, outro com valores de antocianinas e um último com valores de flavonóides. Este processo foi utilizado para teste e treino de cada experimento. Portanto, houveram cinco pastas, representando os experimentos, cada uma com um conjunto de teste e um de treino para cada um dos atributos, totalizando então seis arquivos por pasta.

A partir daí, teremos então todo o experimento preparado para ser submetido à regressão, utilizando o *random forest*.

3.7 CRIAÇÃO DOS MODELOS E ESTIMAÇÃO DOS VALORES DOS ATRIBUTOS

Depois, com as entradas de teste e treino prontas, será o momento de criar o modelo utilizando o *random forest* e o *multilayer perceptron*. Foram utilizadas as funções *randomForest* e *RSNNS*, disponíveis na linguagem *R* para a criação dos modelos e estimação dos valores. A partir daí, teremos arquivos *.csv* de retorno contendo os valores de referência e os valores preditos, para que sejam feitas as análises necessárias.

4 RESULTADOS

Neste capítulo, são apresentados os resultados obtidos a partir dos experimentos, em forma de gráficos comparativos das saídas, tendo como resultados o MSE, o coeficiente de correlação r_P e gráficos de dispersão. O projeto seguiu toda a metodologia apresentada anteriormente, não havendo modificações na proposta.

4.1 AVALIAÇÃO DOS MODELOS

Os modelos criados foram avaliados utilizando-se o r_P e o MSE. As equações dos mesmos são descritas em 4.1 e 4.2 (procurar citações).

$$r_P = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}) \times (y_i - \bar{y})}{(n-1) \times \sigma_r \times \sigma_p} \quad (4.1)$$

$$MSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4.2)$$

com \hat{y}_i o valor predito pelo modelo, y_i a referência, $\bar{\hat{y}}$ a média dos valores de referência, $\bar{\hat{y}}$ a média dos valores preditos, n é a quantidade de amostras, σ_r o desvio padrão dos valores de referência e σ_p dos valores preditos.

4.2 AVALIAÇÃO DE DESEMPENHO DOS MODELOS

Na Tabela 2 mostra um sumário dos descritores estatísticos das análises físico-químicas para as uvas *Cabernet Sauvignon*, sendo eles antocianinas totais, flavonóides e SST.

Tabela 2 – Descritores estatísticos para *Cabernet Sauvignon*.

Parâmetro	Amostras	Média	Min	Max	Intervalo	σ	Variância
SST	432	11,4	3,9	23,5	19,6	5,9	34,9
Ant.	432	86,2	0,8	197,5	196,7	82,6	6830,4
Flav.	432	39,3	9,7	153,3	143,6	26,8	720,7

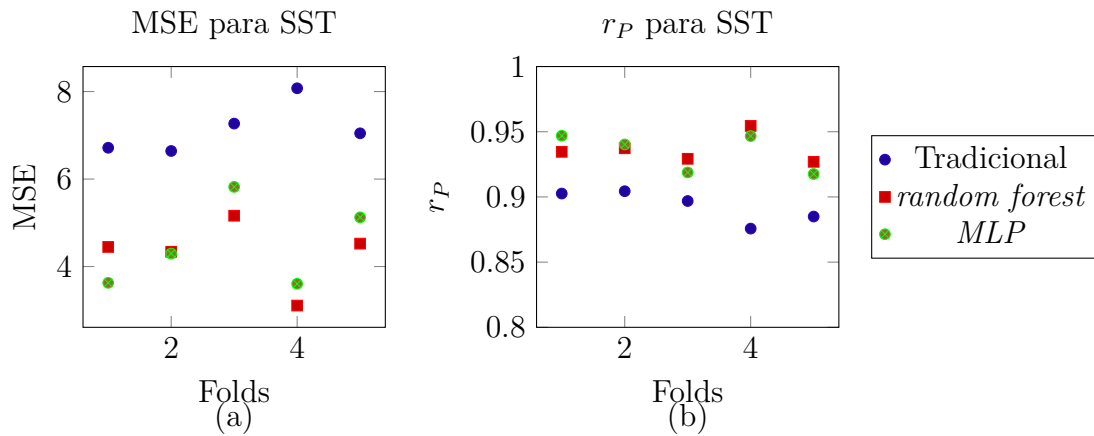
Fonte: O Autor.

É possível verificar uma grande variância entre as amostras. Isto se deve ao fato de que existe uma variação expressiva entre as amostras, evidenciando a robustez de modelos obtidos deste conjunto.

Nas Figuras 9, 11 e 10 a seguir, temos os valores de MSE (a) e r_P (b) de SST, antocianinas e flavonóides para as amostras. São mostrados os valores de referência, e dos

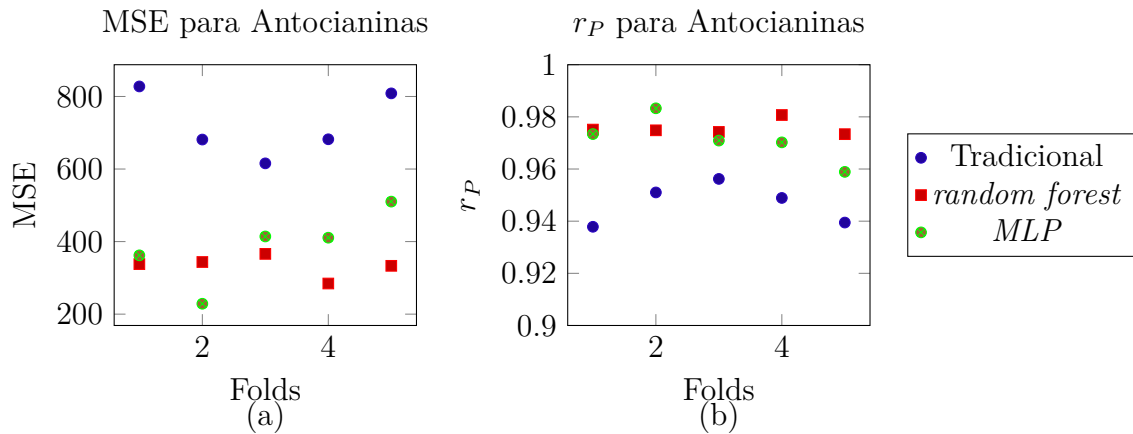
métodos *random forest* e *multilayer perceptron*, de acordo com a descrição da seção de Materiais e Métodos. Os valores de referência foram encontrados obtendo-se os valores de intensidade de todos os *pixels* da uva, obtendo a média para cada canal de intensidade de cor, e obtendo os valores através de regressão multivariada, com os valores dos canais de intensidade como variáveis independentes, e os valores de SST, antocianinas e flavonóides como variáveis dependentes.

Figura 9 – Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para SST.

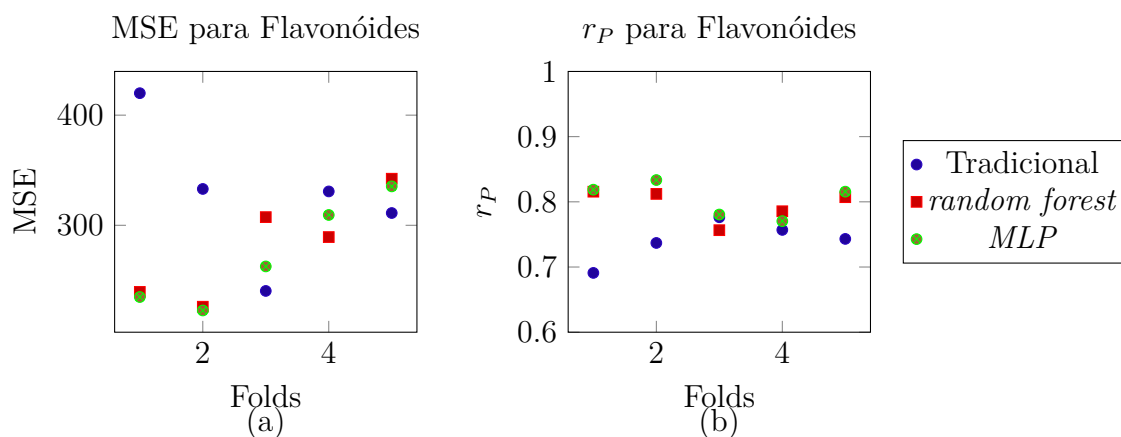


Fonte: O Autor.

Figura 10 – Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para Antocianinas.



Fonte: O Autor.

Figura 11 — Gráficos de desempenho dos métodos para MSE (a) e r_P (b) para Flavonóides.

Fonte: O Autor.

Como é possível observar nos gráficos, os métodos utilizando *random forest* e *MLP* apresentam MSE menor que o método tradicional, apresentando os resultados médios mostrados nas Tabelas 3, 4 e 5. Como para o MSE quanto menor o valor, melhor o desempenho, observa-se que os métodos de inteligência artificial possuem resultados mais adequados.

Tabela 3 — Desempenho de MSE dos métodos para SST.

Método	Mín	Máx	Média
Tradicional	6,642	8,075	7,149
<i>random forest</i>	3,108	5,162	4,316
<i>MLP</i>	3,606	5,821	4,495

Fonte: O autor.

Tabela 4 — Desempenho de MSE dos métodos para Antocianinas.

Método	Mín	Máx	Média
Tradicional	615,625	827,659	723,029
<i>random forest</i>	284,444	366,082	333,077
<i>MLP</i>	228,647	510,179	385.154

Fonte: O autor.

Tabela 5 – Desempenho de MSE dos métodos para Flavonóides.

Método	Mín	Máx	Média
Tradicional	240,392	419,942	327,040
<i>random forest</i>	226,058	342,280	280,913
<i>MLP</i>	222,642	335,304	272,961

Fonte: O autor.

Como podem ser verificados, os valores mínimos, máximos e médios para o MSE dos métodos para os três atributos de qualidade estudados são menores para os métodos de inteligência artificial, em detrimento ao método tradicional. Fica também evidente que os dois métodos, *random forest* e *MLP*, apresentam desempenho semelhante, onde o *random forest* consegue uma vantagem mínima, nesses quesitos, em relação aos atributos SST e Antocianinas. Para flavonóides, observa-se que o *MLP* tem um desempenho também levemente melhor.

Para o coeficiente de correlação, r_P , temos os resultados sintetizados nas Tabelas 6, 7 e 8 a seguir.

Tabela 6 – Desempenho de r_P dos métodos para SST.

Método	Mín	Máx	Média
Tradicional	0,876	0,904	0,893
<i>random forest</i>	0,927	0,954	0,937
<i>MLP</i>	0,918	0,947	0,934

Fonte: O autor.

Tabela 7 – Desempenho de r_P dos métodos para Antocianinas.

Método	Mín	Máx	Média
Tradicional	0,938	0,956	0,947
<i>random forest</i>	0,973	0,981	0,976
<i>MLP</i>	0,959	0,983	0,971

Fonte: O autor.

Tabela 8 – Desempenho de r_P dos métodos para Flavonóides.

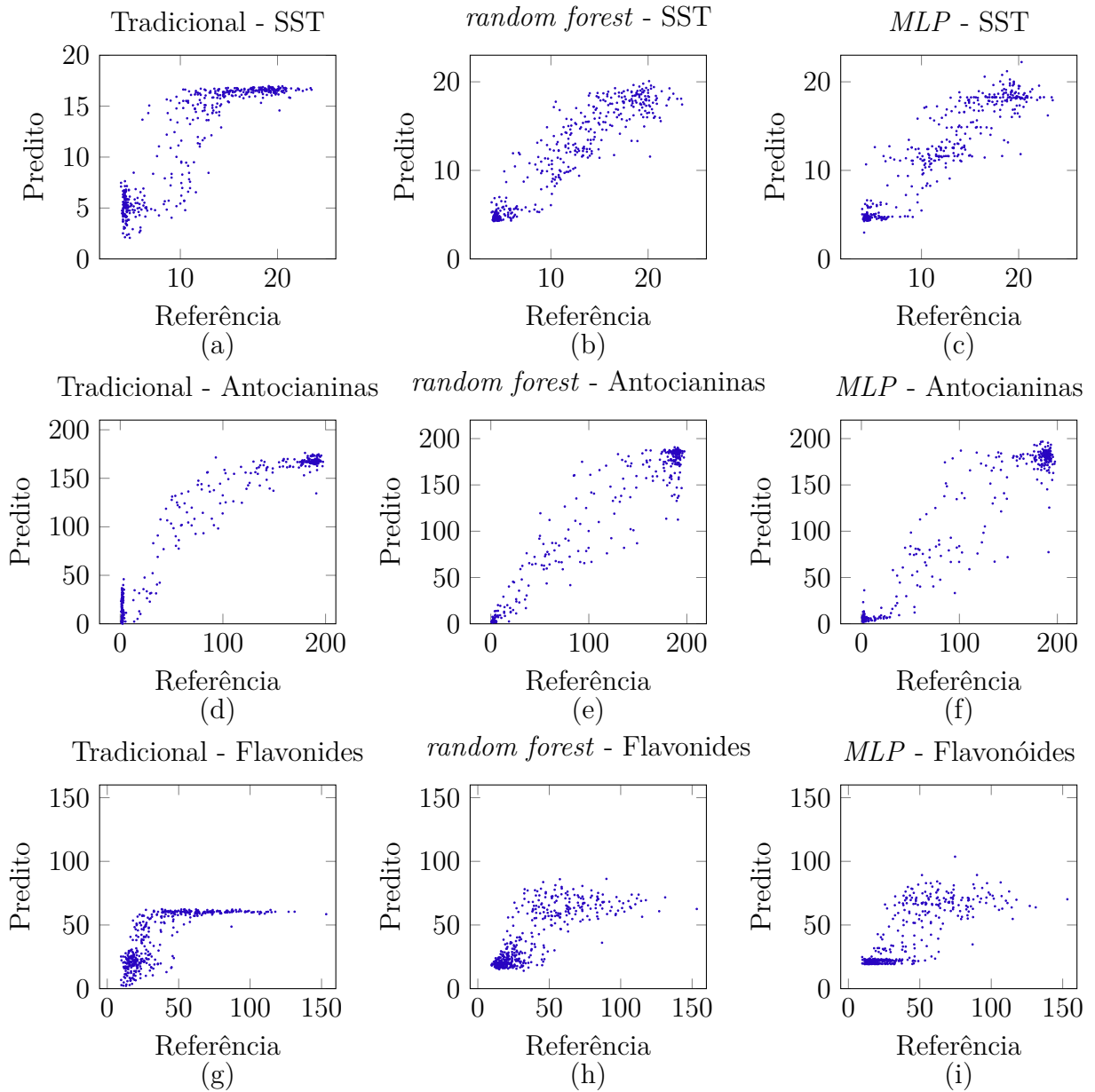
Método	Mín	Máx	Média
Tradicional	0,691	0,776	0,741
<i>random forest</i>	0,757	0,816	0,795
<i>MLP</i>	0,770	0,833	0,804

Fonte: O autor.

Fica evidente que os métodos propostos neste estudo apresentam melhores resultados que o tradicional. Além disso, corroboram os resultados analisados a partir do MSE.

Os gráficos de dispersão da Figura 12 mostram os valores preditos comparados com a referência. Fica evidente a maior linearização dos pontos nos gráficos dos métodos propostos, em detrimento ao método tradicional. Este fato corrobora então com os valores dos coeficientes de correlação descritos anteriormente.

Figura 12 – Gráficos de dispersão para os métodos de determinação dos atributos de qualidade.



Fonte: O Autor.

5 CONCLUSÕES

5.1 CONSIDERAÇÕES FINAIS

Os resultados obtidos com este trabalho demonstram o poder da inteligência artificial, com a utilização dos métodos de regressão *random forest* e *MLP*, para a determinação dos atributos de qualidade em uvas, superando o método tradicional em todos os atributos analisados.

A utilização de técnicas de processamento de imagens RGB, como a transformação de Daugman e a correlação das imagens obtidas com a íris humana, além de proporcionarem uma integração de conhecimentos grandiosa, evidenciam o estado da arte no que tange à inteligência artificial, e proporcionam um breve vislumbre de uma área que é atual, com crescimento elevado, que hoje figura como uma das pontas de lança da Ciência da Computação.

Para além disso, a criação de um método de baixo custo para a determinação dos atributos de qualidade é um ponto importante, em especial para o Vale do Submédio São Francisco, visto que a diminuição de custos é vantajosa para o desenvolvimento da produção vinícola, principalmente para os pequenos produtores, oferecendo uma alternativa confiável e que tem uma boa relação de custo-benefício.

5.2 LIMITAÇÕES

Por se tratar de uma abordagem nova para o problema proposto, este trabalho apresentou desafios interessantes ao seu desenvolvimento. Sendo um trabalho derivado de um projeto voluntário de pesquisa na UNIVASF, houve mais tempo para o desenvolvimento do mesmo, e ainda assim todo este tempo foi bem aproveitado para o desenvolvimento das técnicas e metodologias abordadas neste trabalho. Ao longo de aproximadamente dois anos de trabalho, podemos citar alguns destes como sendo:

- A criação da base de dados, que demandou muito trabalho mecânico e muito tempo, para que a mesma tivesse uma robustez satisfatória;
- As novas técnicas e tecnologias de processamento de imagem, que apesar de serem facilitadas com as bibliotecas e linguagens utilizadas, ainda assim demandaram tempo para que fossem corretamente aplicadas;
- A escolha do melhor método de tratamento dos dados, sendo que a segmentação foi escolhida por seu melhor desempenho. No entanto, outras técnicas foram abordadas

antes, como o uso de toda a imagem como um único vetor de entradas, ou o uso dos valores de intensidade como entradas individuais;

- A aplicação dos dois métodos de regressão, exigindo um estudo mais aprofundado nos métodos matemáticos em que se baseiam, e na aplicação correta para a confiabilidade deste trabalho.

Apesar disto, este trabalho se mostrou um aprendizado grandioso e extremamente relevante, com um nível técnico ideal para uma boa curva de aprendizado.

5.3 TRABALHOS FUTUROS

Um trabalho que é sugerido para o futuro dessa pesquisa é o desenvolvimento de uma aplicação para dispositivos móveis, que se utilize da câmera do próprio dispositivo para obter a imagem da uva e apresentar os valores estimados pelo classificador, para que o produtor possa ter a comodidade e a facilidade de conseguir estes resultados em campo, diminuindo o desperdício e aumentando a eficiência da produção.

REFERÊNCIAS

- ARAÚJO, G. J. F.; SILVA, M. M. Crescimento econômico no semiárido brasileiro: o caso do polo frutícola petrolina/juazeiro. **Caminhos de Geografia**, v. 14, n. 46, 2013. Citado na página 12.
- ARLOT, S.; CELISSE, A. et al. A survey of cross-validation procedures for model selection. **Statistics surveys**, The author, under a Creative Commons Attribution License, v. 4, p. 40–79, 2010. Citado na página 29.
- BASSOI, L. H. et al. Agricultura de precisão em fruticultura. **Embrapa Clima Temperado-Capítulo em livro científico (ALICE)**, In: BERNARDI, AC de C.; NAIME, J. de M.; RESENDE, AV de; BASSOI, LH; INAMASU ..., 2014. Citado na página 12.
- BINFARE, P. W.; GALVÃO, P. L. d. A.; CASTRO, C. A. T. Enoturismo: possibilidades e desafios para o desenvolvimento regional do turismo na região vinícola do vale do são francisco-nordeste brasileiro. Universidad de La Laguna. Servicio de Publicaciones, 2016. Citado na página 12.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 25.
- CHAUCHARD, F. et al. Application of ls-svm to non-linear phenomena in nir spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 71, n. 2, p. 141–150, 2004. Citado 2 vezes nas páginas 15 e 23.
- CHEN, K. et al. Color grading of beef fat by using computer vision and support vector machine. **Computers and Electronics in Agriculture**, Elsevier, v. 70, n. 1, p. 27–32, 2010. Citado 2 vezes nas páginas 20 e 23.
- CHIA, K. S.; RAHIM, H. A.; RAHIM, R. A. A comparison of principal component regression and artificial neural network in fruits quality prediction. In: IEEE. **2011 IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA)**. [S.l.], 2011. p. 261–265. Citado 2 vezes nas páginas 16 e 23.
- DAUGMAN, J. How iris recognition works. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 14, n. 1, p. 21–30, 2004. Citado 2 vezes nas páginas 30 e 31.
- FRANCIS, F. Analysis of anthocyanins. **Anthocyanins as food colors**, Academic press New York, v. 1, p. 280, 1982. Citado na página 33.
- GARDNER, M. W.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998. Citado 2 vezes nas páginas 26 e 27.

GONZALES, R. C.; WINTZ, P. Digital image processing. Addison-Wesley, Reading, Massachusetts, p. 503 p.: ill.; 24 cm, 1987. Disponível em: <<http://infoscience.epfl.ch/record/25541>>. Citado na página 24.

GUDWIN, R. R. Linguagens de programação. **Campinas: DCA/FEEC/UNICAMP**, p. 24, 1997. Citado na página 24.

HANNA INSTRUMENTS BRASIL. **Instruction Manual - HI 96804 Refractometer for Invert Sugar Measurements**. 2012. Disponível em: <https://hannainst.com.br/descargas/1106/manuais-em-ingles/13999/man_ing_hi96804.pdf>. Acesso em: 14 jan. 2019. Citado na página 32.

HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007. Citado na página 26.

HORA, G. B. d.; XAVIER, M. G. P.; MATTOSINHO, C. M. d. S. Inovação na indústria vitinícola do vale do são francisco. IFS, 2016. Citado na página 12.

HOWSE, J. **OpenCV computer vision with python**. [S.l.]: Packt Publishing Ltd, 2013. Citado na página 24.

JANIK, L. et al. The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. **Analytica chimica acta**, Elsevier, v. 594, n. 1, p. 107–118, 2007. Citado 2 vezes nas páginas 17 e 23.

LARSON, S. C. The shrinkage of the coefficient of multiple correlation. **Journal of Educational Psychology**, Warwick & York, v. 22, n. 1, p. 45, 1931. Citado na página 29.

LIAW, A.; WIENER, M. et al. Classification and regression by randomforest. **R news**, v. 2, n. 3, p. 18–22, 2002. Citado na página 25.

LIU, C. et al. Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. **PloS one**, Public Library of Science, v. 9, n. 2, p. e87818, 2014. Citado 2 vezes nas páginas 15 e 23.

LIU, C.; YANG, S. X.; DENG, L. Determination of internal qualities of newhall navel oranges based on nir spectroscopy using machine learning. **Journal of Food Engineering**, Elsevier, v. 161, p. 16–23, 2015. Citado na página 15.

MARENGONI, M.; STRINGHINI, S. Tutorial: Introdução à visão computacional usando opencv. **Revista de Informática Teórica e Aplicada**, v. 16, n. 1, p. 125–160, 2009. Citado na página 24.

MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. **Agricultura de precisão**. [S.l.]: Oficina de Textos, 2015. Citado na página 12.

MUNAWAR, A. A. **Multivariate analysis and artificial neural network approaches of near infrared spectroscopic data for non-destructive quality attributes prediction of Mango (Mangifera indica L.)**. Tese (Doutorado) — Niedersächsische Staats-und Universitätsbibliothek Göttingen, 2014. Citado 2 vezes nas páginas 18 e 23.

NAMBI, V. E. et al. Comparison of various rgb image features for nondestructive prediction of ripening quality of “alphonso” mangoes for easy adoptability in machine vision applications: A multivariate approach. **Journal of Food Quality**, v. 6, n. 39, p. 816–825, 2016. Citado 3 vezes nas páginas 19, 21 e 23.

NICOLAI, B. M.; THERON, K. I.; LAMMERTYN, J. Kernel pls regression on wavelet transformed nir spectra for prediction of sugar content of apple. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 85, n. 2, p. 243–252, 2007. Citado na página 15.

PANCHAL, G. et al. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. **International Journal of Computer Theory and Engineering**, v. 3, n. 2, p. 332–337, 2011. Citado na página 26.

PEREIRA, L. F. S. et al. Predicting the ripening of papaya fruit with digital imaging and random forests. **Computers and Electronics in Agriculture**, Elsevier, v. 145, p. 76–82, 2018. Citado 4 vezes nas páginas 19, 22, 23 e 29.

PIRES, A. P. M. **Composição química e actividade antioxidante de folhas de diferentes castas de videira**. Tese (Doutorado) — Instituto Politécnica de Bragança, Escola Superior Agrária, 2010. Citado na página 12.

PREVOLNIK, M. et al. Classification of dry-cured hams according to the maturation time using near infrared spectra and artificial neural networks. **Meat science**, Elsevier, v. 96, n. 1, p. 14–20, 2014. Citado 2 vezes nas páginas 19 e 23.

SANNER, M. F. et al. Python: a programming language for software integration and development. **J Mol Graph Model**, v. 17, n. 1, p. 57–61, 1999. Citado na página 24.

SAVAKAR, D. G.; ANAMI, B. S. Recognition and classification of food grains, fruits and flowers using machine vision. **International Journal of Food Engineering**, De Gruyter, v. 5, n. 4, 2009. Citado 2 vezes nas páginas 19 e 23.

SOUSA, A. et al. Global competitive dynamics and innovation in the brazilian wine sector: An analysis of vale do são francisco pole. In: 35TH WORLD CONGRESS OF VINE AND WINE (OIV 2012)-PROCEEDINGS-IZMIR (TURKEY). [S.l.], 2012. Citado na página 12.

SOUZA, J. d. L. et al. Pré-colheita e colheita de uvas na serra gaúcha (cooperativa vinícola aurora de bento gonçalves/rs) e legislação de bebidas derivadas da uva e do vinho no mapa-sc. 2013. Citado na página 13.

TEAM, R. C. R language definition. **Vienna, Austria: R foundation for statistical computing**, 2000. Citado na página 25.

TORGO, L. **Data mining with R: learning with case studies**. [S.l.]: Chapman and Hall/CRC, 2016. Citado na página 25.

TUCKER, A.; NOONAN, R. **Linguagens de Programação-: Princípios e Paradigmas**. [S.l.]: AMGH Editora, 2009. Citado na página 24.

XIA, J.-F. et al. Application of wavelet transform in the prediction of navel orange vitamin c content by near-infrared spectroscopy. **Agricultural Sciences in China**, Elsevier, v. 6, n. 9, p. 1067–1073, 2007. Citado na página 15.

ZANINI, T. V.; ROCHA, J. M. da. O enoturismo no brasil: um estudo comparativo entre as regiões vinícolas do vale dos vinhedos (rs) e do vale do são francisco (ba/pe). **Revista Turismo em Análise**, v. 21, n. 1, p. 68–88, 2010. Citado na página 12.