

**UNIVERSIDADE ESTADUAL DE PONTA GROSSA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

HENIKE GUILHERME JORDAN VOSS

**DESENVOLVIMENTO DE UM NARIZ ELETRÔNICO APLICADO À
DETERMINAÇÃO DO ESTÁGIO DE MATURAÇÃO EM
PESSEGUEIROS**

PONTA GROSSA

2019

HENIKE GUILHERME JORDAN VOSS

**DESENVOLVIMENTO DE UM NARIZ ELETRÔNICO APLICADO À
DETERMINAÇÃO DO ESTÁGIO DE MATURAÇÃO EM
PESSEGUEIROS**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada, curso de Mestrado em Computação Aplicada - Área de concentração Computação para Tecnologias em Agricultura - da Universidade Estadual de Ponta Grossa, como requisito parcial para obtenção do título de Mestre.

Orientador: Prof. Dr. Sergio Luiz Stevan Júnior

PONTA GROSSA

2019

V968 Voss, Henike Guilherme Jordan
Desenvolvimento de um Nariz Eletrônico aplicado à determinação do estágio de maturação em pessegueiros / Henike Guilherme Jordan Voss. .
Ponta Grossa, 2019.
93 f.

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

Orientador: Prof. Dr. Sergio Luiz Stevan Junior.

1. Nariz eletrônico (e-nose). 2. Compostos orgânicos voláteis (COVs).
3. Pêssego. 4. Odor da fruta. 5. Maturação. I. Stevan Junior, Sergio Luiz. II.
Universidade Estadual de Ponta Grossa. Computação para Tecnologias em
Agricultura. III.T.

CDD: 005.3

TERMO DE APROVAÇÃO

Henike Guilherme Jordan Voss

“DESENVOLVIMENTO DE UM NARIZ ELETRÔNICO APLICADO À DETERMINAÇÃO DE ESTÁGIO DE MATURAÇÃO EM PESSEGUEIROS”

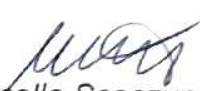
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:



Prof. Dr. Sérgio Luiz Stevan Junior
UTFPR



Prof. Dr. Rosane Falate
UEPG



Prof. Dr. Marcella Scoczynski Ribeiro Martins
UTFPR

Ponta Grossa, 15 de março de 2019.

V968 Voss, Henike Guilherme Jordan
Desenvolvimento de um Nariz Eletrônico aplicado à determinação do
estágio de maturação em pessegueiros / Henike Guilherme Jordan Voss. .
Ponta Grossa, 2019.
93 f.

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

Orientador: Prof. Dr. Sergio Luiz Stevan Junior.

1. Nariz eletrônico (e-nose). 2. Compostos orgânicos voláteis (COVs).
3. Pêssego. 4. Odor da fruta. 5. Maturação. I. Stevan Junior, Sergio Luiz. II.
Universidade Estadual de Ponta Grossa. Computação para Tecnologias em
Agricultura. III.T.

CDD: 005.3

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

AGRADECIMENTOS

Aos meus familiares, em especial aos meus pais Paulo e Sonia por sempre acreditarem em mim, me apoiarem em minhas decisões, além do incentivo ao estudo e valorização das minhas conquistas pessoais e profissionais.

A minha namorada, Aline, que sempre me incentivou e me apoiou, e mesmo nos momentos mais difíceis me deu forças para continuar em busca dos meus sonhos.

Ao meu primo Celso e meu irmão Daniel, amigos e parceiros, que sempre me fizeram pensar grande e acreditar que era possível alcançar meus objetivos.

Ao meu orientador Professor Doutor Sergio Stevan, pelos valiosos conselhos e orientações, além de seu empenho e paciência em todos os momentos do mestrado, acreditando no meu potencial e me incentivando nas pesquisas.

Aos professores que fazem parte do corpo docente do programa, pelos ensinamentos, conselhos e suporte, no qual contribuíram para a minha formação como pesquisador.

A Universidade Estadual de Ponta Grossa por ter me acolhido e contribuído para a minha formação acadêmica e profissional, dando condições de realizar as pesquisas.

Por último, à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela ajuda financeira, no qual contribuiu diretamente no desenvolvimento da pesquisa durante esses anos de mestrado.

RESUMO

Em regiões com o predomínio da agricultura, a inspeção da qualidade e do índice de maturação dos frutos no pomar é geralmente analisado pela experiência do agricultor, o que pode estar sujeito a erros e gerar um custo maior de tempo e dinheiro. Dependendo do período de colheita dos frutos, a fruta pode ter diferentes qualidades em diferentes pontos da área de plantio. Os compostos orgânicos voláteis (COVs) variam de quantidade e tipo conforme as diferentes etapas de crescimento dos frutos. Assim, um equipamento para o monitoramento que gere uma resposta rápida e precisa do ciclo de crescimento dos pêssegos na cultura é desejável, aliado a um reduzido custo de comercialização. Para esse fim, os narizes eletrônicos mostram-se como uma alternativa, pois possibilita o monitoramento *online* dos COVs gerados pela cultura. Neste contexto, um protótipo foi desenvolvido para identificar a maturação do pêssego Eragil no ciclo de crescimento pré-colheita, que compreendeu o período de agosto a dezembro de 2018, no Brasil. Foram estudados modelos com os 13 sensores de gás feitos com semicondutor de óxido metálico (MOS) e a redução para 6 sensores com o auxílio do teste *Chi*-quadrado de Pearson, a fim de comparação. Amostras com 4 estágios de crescimento foram utilizadas para o treinamento e construção do modelo. A taxa de erro de 1,92% na etapa de teste das amostras, utilizando o método *Random Forest*, com análise discriminante linear, para o conjunto de dados reduzido para 6 sensores, mostra que o dispositivo é promissor para monitoramento de áreas com uma intensa emissão de COVs. Os resultados obtidos nessa pesquisa mostram-se satisfatórios, nas comparações quantitativas (erros) e qualitativas dadas pelo equipamento, com base no que foi encontrado na pesquisa bibliográfica, uma vez que as aplicações tendem a realizar a medida do fruto pós-colheita e não a avaliação do fruto durante o seu crescimento, no pomar.

Palavras-chaves: Nariz eletrônico (E-nose); Compostos Orgânicos Voláteis (COVs); pêssego; odor da fruta; maturação.

ABSTRACT

In regions with the predominance of agriculture, inspection of the quality and fruit maturity index in the orchard is usually analyzed by the farmer's experience, which can be subject to errors and generate a greater cost of time and money. Depending on the fruit harvest period, the fruit may have different qualities at different points in the planting area. Volatile organic compounds (VOCs) vary in quantity and type depending on the different stages of fruit growth. Thus, monitoring equipment that generates a rapid and accurate response to the growth cycle of the peaches in the crop is desirable, together with a low marketing cost. To this end, the electronic noses are an alternative, since it allows online monitoring of the VOCs generated by the crop. In this context, a prototype was developed to identify the maturation of the Eragil peach in the pre-harvest growth cycle, which comprised the period from August to December 2018, in Brazil. Models with the 13 gas sensors made with a metal oxide semiconductor (MOS) and the reduction to 6 sensors were studied with the aid of the Pearson's *Chi*-square test, for comparison. Samples with 4 growth stages were used for the training and construction of the model. The 1.92 % error rate in the sample test step using the Random Forest method with linear discriminant analysis for the reduced data set for 6 sensors shows that the device is promising for monitoring of areas with an intense emission of VOCs. The results obtained in this research are satisfactory, in the quantitative (errors) and qualitative comparisons given by the equipment, based on what was found in the bibliographical research, since the applications tend to carry out the measurement of the post-harvest fruit and not the evaluation of the fruit during its growth, in the orchard.

Keywords: Electronic nose (E-nose); Volatile Organic Compounds (VOCs); Peach; fruit flavor; maturation.

LISTA DE FIGURAS

Figura 1 – Pêssegos da cultivar Eragil, em diferentes estágios de maturação: (a) verde (b) amarelo (c) meio maduro (d) maduro	18
Figura 2 – Curva de crescimento sigmóide duplo do fruto de pêssego	20
Figura 3 – Semelhança entre o sistema biológico olfativo e um nariz eletrônico	22
Figura 4 – Comparativo entre as abordagens de seleção de subconjuntos de atributos . . .	27
Figura 5 – Fluxograma ilustrativo da metodologia adotada	44
Figura 6 – Protótipo do nariz eletrônico	47
Figura 7 – Diagrama de blocos do circuito do protótipo desenvolvido	47
Figura 8 – Equipamento montado no cultivar	51
Figura 9 – Fluxograma ilustrativo da etapa de pré-processamento dos dados	53
Figura 10 – A resposta típica dada pelos sensores com amostras de pêssego imaturo . . .	57
Figura 11 – A resposta típica dada pelos sensores com amostras de pêssego maduro . . .	58
Figura 12 – A resposta típica dada pelos sensores com amostras de pêssego super maduro	59
Figura 13 – Dados tridimensionais obtidos com 13 sensores pelo PCA	60
Figura 14 – Dados tridimensionais obtidos com 13 sensores pelo LDA	61
Figura 15 – Dados tridimensionais obtidos com 6 sensores pelo PCA	61
Figura 16 – Dados tridimensionais obtidos com 6 sensores pelo LDA	62

LISTA DE TABELAS

Tabela 1 – Exemplos de aplicações utilizando narizes eletrônicos	14
Tabela 2 – Produção de pêssego no município de Ponta Grossa - Paraná.	17
Tabela 3 – COVs identificados por GC / MS no processo de crescimento dos pêssegos.	19
Tabela 4 – Faixa de detecção de partículas e preços dos sensores MQ e TGS utilizados no estudo	45
Tabela 5 – Principais características do sensor de temperatura e pressão atmosférica BMP180	48
Tabela 6 – Principais características do sensor analógico de umidade do ar HHH-4030	48
Tabela 7 – Principais características do microcontrolador Atmel ATmega2560	49
Tabela 8 – Estágios de maturação do pêssego e seus respectivos dias e classes de gases	50
Tabela 9 – Valores das constantes utilizadas nas equações de ajustes dos sensores	52
Tabela 10 – Coeficiente V de Cramer para cada um dos sensores utilizados nos experimento	60
Tabela 11 – Resultados obtidos com os classificadores para o conjunto de dados com 13 sensores	63
Tabela 12 – Resultados obtidos com os classificadores para o conjunto de dados com 6 sensores	63
Tabela 13 – Matriz de confusão para todos os classificadores sem metodo de redução como entrada utilizando 13 sensores	64
Tabela 14 – Matriz de confusão para todos os classificadores com o PCA como entrada utilizando 13 sensores	65
Tabela 15 – Matriz de confusão para todos os classificadores com o LDA como entrada utilizando 13 sensores	65
Tabela 16 – Matriz de confusão para todos os classificadores sem metodo de redução como entrada utilizando 6 sensores	66
Tabela 17 – Matriz de confusão para todos os classificadores com o PCA como entrada utilizando 6 sensores	66
Tabela 18 – Matriz de confusão para todos os classificadores com o LDA como entrada utilizando 6 sensores	67
Tabela 19 – Resumo dos resultados dos trabalhos relacionados de maturação do pêssego	68
Tabela 20 – Características dos sensores da família MQ e TGS utilizados neste trabalho	82
Tabela 21 – Características do sensor de temperatura e pressão atmosférica BMP180	84
Tabela 22 – Características do sensor analógico de umidade do ar HHH-4030	84
Tabela 23 – Características do microcontrolador Atmel ATmega2560	85

LISTA DE ABREVIATURAS E SIGLAS

AIA	Ácido indolacético
CART	<i>Classification And Regression Tree</i> - Árvore de Classificação e Regressão
CCD	<i>Charge Coupled Device</i> - Dispositivo de Carga Acoplada
COV	Compostos Orgânicos Voláteis
DFA	<i>Discriminant Function Analysis</i> - Análise de Funções Discriminantes
EEPROM	<i>Electrically-Erasable Programmable Read-Only Memory</i> - Memória de apenas leitura programável e apagável eletricamente
ELM	<i>Extreme Learning Machine</i> - Máquina de Aprendizagem Extrema
GC-MS	<i>Gas Chromatography and Mass Spectrometry</i> - Cromatografia gasosa e espectrometria de massa
GPS	<i>Global Positioning System</i> - Sistema de Posicionamento Global
I ² C	<i>Inter-Integrated Circuit</i> - Circuito Inter-Integrado
KNN	<i>K-Nearest Neighbours</i> - K-vizinhos mais próximos
LDA	<i>Linear Discriminant Analysis</i> - Análise Discriminante Linear
LiPo	Polímero de Lítio
MDA	<i>Multivariate Data Analysis</i> - Análise de Dados Multivariados
MMH	<i>Maximum Marginal Hyperplane</i> - Hiperplano Marginal Máximo
MOS	<i>Metal-Oxide-Semiconductor</i> - Semicondutor de óxido metálico
MSPC	<i>Multivariate Statistical Process Control</i> - Controle de Processo Estatístico Multivariado
NPK	Nitrogênio, Fósforo e Potássio
PCA	<i>Principal Component Analysis</i> - Análise de Componentes Principais
PCR	<i>Principal Component Regression</i> - Regressão de Componentes Principais
pH	Potencial Hidrogeniônico

PLS	<i>Partial Least Squares Interpolation</i> - Interpolação Parcial de Mínimos Quadrados
ppm	Partículas por milhão
PWM	<i>Pulse Width Modulation</i> - Modulação por largura de pulso
R	Coefficiente de correlação
R ²	Coefficiente de determinação
RF	<i>Random Forest</i> - Floresta aleatória
RNA	<i>Artificial Neural Network</i> - Rede Neural Artificial
SEP	<i>Standard Error of Prediction</i> - Erro padrão de previsão
SIMCA	<i>Independent Modelling of Class Analogy</i> - Modelagem Suave Independente de Analogias de Classe
SR	<i>Stochastic Resonance</i> - Ressonância estocástica
SVM	<i>Support Vector Machine</i> - Máquina de Vetores de Suporte
UART	<i>Universal Asynchronous Receiver-Transmitter</i> - Receptor-transmissor Universal Assíncrono
USB	<i>Universal Serial Bus</i> - Barramento Serial Universal

SUMÁRIO

1	Introdução	13
1.1	Organização do trabalho	15
1.2	Objetivo geral	16
1.3	Objetivos específicos	16
2	Revisão da Literatura	17
2.1	Produção do pêssego	17
2.2	Características do fruto	18
2.3	Ciclo de crescimento dos pêssegos	19
2.4	Narizes Eletrônicos	21
2.4.1	Tipos e características de sensores de gás em Narizes Eletrônicos	23
2.5	Pré-processamento dos dados	23
2.5.1	Análise de redundância e correlação	25
2.5.2	Seleção de variáveis	25
2.5.3	Análise de Dados Multivariados - MDA	27
2.5.3.1	Análise de Componentes Principais - PCA	28
2.5.3.2	Análise Discriminante Linear - LDA	29
2.6	Métodos de Reconhecimento de Padrões	29
2.6.1	K-vizinhos mais próximos - KNN	29
2.6.2	Máquinas de vetor de suporte - SVM	30
2.6.2.1	Conjunto de dados linearmente separáveis	31
2.6.3	Floresta aleatória - RF	32
2.6.4	Redes Neurais Artificiais - RNA	33
2.6.4.1	<i>Extreme Learning Machine</i> - ELM	34
2.7	Avaliação de classificadores	35
2.7.1	Validação Cruzada <i>k-fold</i>	35
3	Estado da Arte	36
3.1	Industrial/Urbano	36
3.2	Agricultura	38
3.2.1	Solo	38
3.2.2	Fruticultura	40
3.2.2.1	Bananas	41
3.2.2.2	Pêssegos	41
4	Materiais e Métodos	44

4.1	Construção/Parametrização do Equipamento	44
4.2	Protocolo de Experimentação	49
4.3	Procedimentos de Pré-Processamento	51
4.4	Procedimentos de Processamento dos dados	53
4.5	Procedimentos de Avaliação dos dados	54
5	Resultados e Discussões	56
5.1	Respostas geradas pelos sensores	56
5.2	Remoção dos atributos equivalentes	56
5.3	Redução da dimensionalidade dos dados	58
5.4	Resultados das classificações e discussões	60
5.5	Comparação com outros trabalhos	68
6	Conclusões	71
7	Trabalhos futuros	73
	 Referências	 74
	 Apêndices	 81
	APÊNDICE A – Características dos sensores utilizados neste trabalho	82
	APÊNDICE B – Diagrama esquemático do circuito do nariz eletrônico	86
	APÊNDICE C – Layout da placa de circuito impresso do nariz eletrônico	87
	APÊNDICE D – Resumo dos dados obtidos nos experimentos	88

1 INTRODUÇÃO

O pêssego (*Prunus pérsica*) é um fruto que se originou da China, que também é a maior produtora mundial do fruto, representando um volume de 38% da produção em todo o planeta, além de ser o país que mais consome a fruta (CAMILLO, 2009). No mundo todo são produzidos cerca de 17,5 milhões de toneladas de pêssegos por ano.

Mesmo o pêssego ter sido implantado no Brasil há muito tempo, o país ainda não é capaz de produzir conforme sua potencialidade. Dados mostram que o Brasil possui uma área cultivada de pêssegos superior a 24000 hectares, com produção de cerca de 238,5 mil toneladas por ano (BALKE, 2016). O Brasil ocupa a 12^a posição do ranking mundial de produção em volume. Dentre os estados brasileiros com maior produção, está o Rio Grande do Sul seguido de Santa Catarina, São Paulo e Paraná (CAMILLO, 2009).

O pêssego Eragil, que é amplamente cultivado no Paraná, especialmente na região dos Campos Gerais, pode ser tanto produzido com o propósito de utilização na indústria, quanto também ao consumo *in natura*. Apresenta polpa de coloração amarelada e que não adere ao caroço, e este, por sua vez, possui tamanho grande, formato oblongo, com cerca de 50% da coloração externa avermelhada (BRACKMANN et al., 2009). É uma fruta de alta sazonalidade, sendo os meses de outubro a dezembro aqueles que concentram cerca de 79% da oferta do produto (SANTOS, 2007).

As frutas, e especialmente os pêssegos, geram diferentes COVs (Compostos Orgânicos Voláteis) e sua qualidade, medida pelo aroma, sabor e cor, altera constantemente durante toda a fase de crescimento e maturação, que ocorre no período de pré-colheita (BAIETTO; WILSON, 2015). Ao medir essas propriedades, alguns métodos instrumentais usados são manuais e destrutivos¹. A medição de forma não destrutiva da qualidade interna dos frutos está se tornando importante para a indústria e para os consumidores (RAJKUMAR et al., 2012).

Uma indicação comum de maturação das frutas é o mecanismo biológico. Nos lugares onde a agricultura predomina, a inspeção da qualidade e da maturação dos frutos é dada pela experiência do agricultor, o que pode estar sujeito a erros e descuidos (CHEN et al., 2018). Além disso, caso seja considerado um período específico para a colheita dos frutos, podem haver ocasiões em que os frutos estejam com qualidades distintas em diferentes pontos da área plantada, o que se enquadra na agricultura de precisão. Visto isso, é desejável um equipamento que monitore os diferentes graus de maturação nos pontos de cultivo, e dado que isto pode ser identificado pelos COVs, os narizes eletrônicos mostram-se uma alternativa robusta para este fim, além de possibilitar o monitoramento contínuo da cultura.

O desenvolvimento da tecnologia de sensores permitiu aos narizes eletrônicos se apre-

¹ As amostras medidas no processo não poderão ser mais aproveitadas, ou seja, deverão ser descartadas.

sentar como dispositivos simples e com alta precisão de detecção, sendo estes dispositivos cada vez mais usados como alternativa aos métodos tradicionais (WU et al., 2017; ZHU et al., 2017; LIN; ZHANG, 2016; JIANG; WANG, 2016; MEN et al., 2018b). Tabela 1 resume algumas das várias aplicações possíveis para os narizes eletrônicos.

Tabela 1 – Exemplos de aplicações utilizando narizes eletrônicos

Aplicação	Referências
Área médica	Voss et al. (2012), Lorwongtragool et al. (2014), Li et al. (2017), Huang et al. (2018)
Emissão do solo	Sudarmaji e Kitagawa (2016), Dorji, Pobkrut e Kerdcharoen (2017), Pineda e Pérez (2017), Bieganowski et al. (2018)
Qualidade de alimentos	Aleixandre et al. (2015), Wojnowski et al. (2017b), Chen et al. (2018)
Qualidade de bebidas	Ragazzo-Sanchez et al. (2008), Mutalib, Jaswir e Akmeliawati (2013), Santos e Lozano (2015), Nurul et al. (2017), Wei et al. (2017), Aleixandre et al. (2018)
Qualidade do ar	Bagula et al. (2012), Abraham e Pandian (2013), Deshmukh et al. (2015), Peterson et al. (2017), Blanco-Novoa et al. (2018), Kim e Hwangbo (2018), Laref et al. (2018)

As aplicações mais comuns de uso de sensores na agricultura são para prover informação precisa de colheita, solo, temperatura e condições meteorológicas. A gestão agrícola depende de um espectro de sensores que vai desde a detecção aérea remota, estações meteorológicas de campo, sensores ambientais de efeito estufa, sensores eletroquímicos, narizes eletrônicos, biocombustíveis e sofisticadas redes de sensores sem fio (PINEDA; PÉREZ, 2017).

O termo nariz eletrônico (*Electronic Nose*) é geralmente confundido com um dispositivo que é capaz de detectar odores como um nariz humano, sendo, de fato, esta definição equivocada. Os narizes eletrônicos normalmente são constituídos de uma matriz de sensores químicos baseados em materiais de óxido metálico. Mas, como os narizes humanos, esses dispositivos não são capazes de identificar as substâncias separadamente em uma determinada amostra. Pode-se fazer uma analogia dos dados obtidos por esses sensores com a impressão digital, uma vez que é muito difícil de encontrar duas substâncias que gerem a mesma resposta para todo o conjunto de sensores, sendo possível classificar substâncias de acordo com suas características (PINEDA; PÉREZ, 2017). Os Narizes Eletrônicos são constituídos em quatro partes: matriz de sensores, unidade de processamento de sinais, banco de dados e reconhecimento de padrões. Essas quatro partes simulam, respectivamente, a aquisição de informações por neurônios sensoriais do receptor olfatório humano, a codificação do nervo olfatório (bulbo), a memória cerebral e o processamento de informações pelo sistema olfativo humano (HUANG et al., 2018).

Os narizes eletrônicos podem ser usados para monitorar e medir os compostos odoríferos e gasosos, emitidos de várias indústrias para o meio ambiente. Os lançamentos de gases poluentes no ar por parte das indústrias são um dos principais problemas, uma vez que essas emissões são

altamente dinâmicas e complexas. Além disso, dependendo da composição lançada, as questões ambientais são afetadas, não só a nível local, mas também a nível global (DESHMUKH et al., 2015).

Os narizes eletrônicos também podem ser aplicados na agroindústria, porque permitem a aquisição de dados *online* sobre a natureza e qualidade química e física das plantas, produtos vegetais e animais e gases liberados de produtos agrícolas ao longo de todo o ciclo de produção de alimentos. A capacidade de monitoramento contínuo desses dispositivos fornece um meio de garantir que os métodos de produção e os resultados atinjam as especificações de qualidade (padrões) e as exigências das agências reguladoras e do consumidor para obter uma excelente salabilidade nos mercados comerciais (WILSON, 2013).

As indústrias agrícolas tornaram-se altamente dependentes dos narizes eletrônicos devido à capacidade destes equipamentos reconhecer a presença de misturas específicas de gases que são produzidas ou liberadas durante ou como subproduto de vários processos de fabricação. As características aromáticas dos produtos agrícolas, particularmente na indústria alimentar, contribuem imensamente para o valor do produto. Por essas razões, o controle de qualidade das características de aroma dos produtos manufaturados é de suma importância porque a consistência do produto é essencial para manter o reconhecimento e a satisfação da marca para o consumidor (WILSON; BAIETTO, 2009).

A literatura mostra que as aplicações realizam a medida do fruto pós-colheita e não o monitoramento dos cultivares durante o seu crescimento no pomar, com base nos COVs exalados pelos pêssegos. Com isso, este trabalho relata-se o estudo e desenvolvimento de um protótipo de um nariz eletrônico a fim de monitorar o ciclo de crescimento do pêssego em áreas do cultivar, desde a pós-floração até a pós-colheita, com o monitoramento a campo. Além disso, é necessário o estudo das diferentes ferramentas para processamento e classificação dos dados gerados *online* pelo dispositivo, a fim de se obter uma classificação precisa entre os estágios de crescimento do fruto e que não seja suscetível a possíveis erros na aplicabilidade do mesmo.

1.1 ORGANIZAÇÃO DO TRABALHO

Este capítulo apresentou uma introdução ao estudo abordado e tratará dos objetivos desta dissertação. Os capítulos subsequentes tratam dos seguintes tópicos: o capítulo 2 apresenta a revisão da literatura referente ao assunto abordado, o capítulo 3 descreve o estado da arte de aplicações de narizes eletrônicos na agricultura, o capítulo 4 demonstra os materiais e métodos adotados a fim de alcançar os objetivos propostos, o capítulo 5 apresenta os resultados obtidos, além de realizar uma discussão, o capítulo 6 expõe as conclusões obtidas com o desenvolvimento do trabalho e o capítulo 7 mostra as sugestões de trabalhos futuros. No final do documento, na seção de apêndices, encontram-se informações relevantes para o desenvolvimento da dissertação, que inclui a caracterização dos sensores utilizados (apêndice A), o diagrama esquemático do

circuito do protótipo (apêndice B), o *layout* da placa de circuito impresso (apêndice C), além dos dados gerados pelos sensores, com métricas estatísticas (apêndice D).

1.2 OBJETIVO GERAL

Identificar o amadurecimento dos pêssegos Eragil, *in loco*, de forma não-invasiva, através da presença dos gases exalados por estes, durante o ciclo de crescimento do fruto, desde o período de plena floração até o período pós-colheita, para que seja possível discriminar as diferentes fases do crescimento. Para isso, o desenvolvimento de um dispositivo de nariz eletrônico e de um sistema de processamento que permita com que as informações obtidas sejam disponibilizadas para o usuário se mostra como uma alternativa adequada.

1.3 OBJETIVOS ESPECÍFICOS

Os objetivos específicos do trabalho estão descritos a seguir:

- Construir uma placa de circuito impresso baseada em uma matriz de sensores a ser utilizada em dispositivos microcontroladores de prototipagem rápida, a fim de facilitar o uso deste dispositivo em propriedades rurais;
- Desenvolver os *firmwares* que serão responsáveis pela aquisição e captura dos dados;
- Coleta e formação da base de dados, com os dados adquiridos durante o monitoramento dos ciclos de maturação do pêssego;
- Identificar e implementar os métodos adequados para a redução da dimensionalidade dos dados;
- Identificar e implementar os algoritmos adequados para a classificação dos dados adquiridos;
- Demonstrar a capacidade do equipamento em classificar os estágios de crescimento do cultivar;
- Comparar os resultados obtidos do equipamento desenvolvido com aqueles encontrados na literatura.

2 REVISÃO DA LITERATURA

Neste capítulo serão apresentados as seções referentes a produção (seção 2.1), características (seção 2.2) e ciclo de crescimento (seção 2.3) dos pêssegos. Então, será dada uma fundamentação teórica sobre narizes eletrônicos (seção 2.4). Após isso, serão apresentados os procedimentos utilizados neste estudo para extração de características das respostas dos sensores, que envolve: pré-processamento (seção 2.5), aplicação de métodos de reconhecimento de padrões (seção 2.6) e avaliação dos classificadores (seção 2.7).

2.1 PRODUÇÃO DO PÊSSEGO

A introdução do pêssego no Brasil ocorreu no ano de 1532, em São Paulo, e, a partir disto foi distribuído em vários estados, sendo que os melhores resultados em relação à produção e qualidade foram obtidos na região sul do país, por possuírem as melhores condições climáticas para a sua produção (BALKE, 2016).

Uma vez que o pêssego apresenta grande diversidade e é importante no fator econômico, este tem sido amplamente cultivado no mundo, desde zonas frias temperadas a zonas subtropicais, sendo que são produzidos cerca de 17,5 milhões de toneladas de pêssego no mundo por ano. A América do Norte e a Europa são responsáveis por 4,38 milhões de toneladas de pêssegos produzidos, com destaque para a Itália, Espanha e Estados Unidos. A produção na Ásia chega perto de 9,87 milhões de toneladas, onde só a China, que é a maior produtora mundial da fruta, contempla 8,2 milhões de toneladas (BALKE, 2016).

O município de Ponta Grossa, no estado do Paraná, caracteriza-se por possuir alguns cultivares do fruto, que auxilia na demanda na região dos Campos Gerais, e segundo IPARDES (2019) já é responsável por uma parcela de representação na economia da região, conforme é mostrado na Tabela 2.

Tabela 2 – Produção de pêssego no município de Ponta Grossa - Paraná.

Produção de pêssego no município de Ponta Grossa		
Parâmetro	Valor	Unidade
Área colhida	40	Hectare (ha)
Produção	650	Tonelada (t)
Rendimento médio	16250	Quilograma/hectare (kg/ha)
Valor	1430	Mil reais (R\$1.000,00)

Fonte: (IPARDES, 2019)

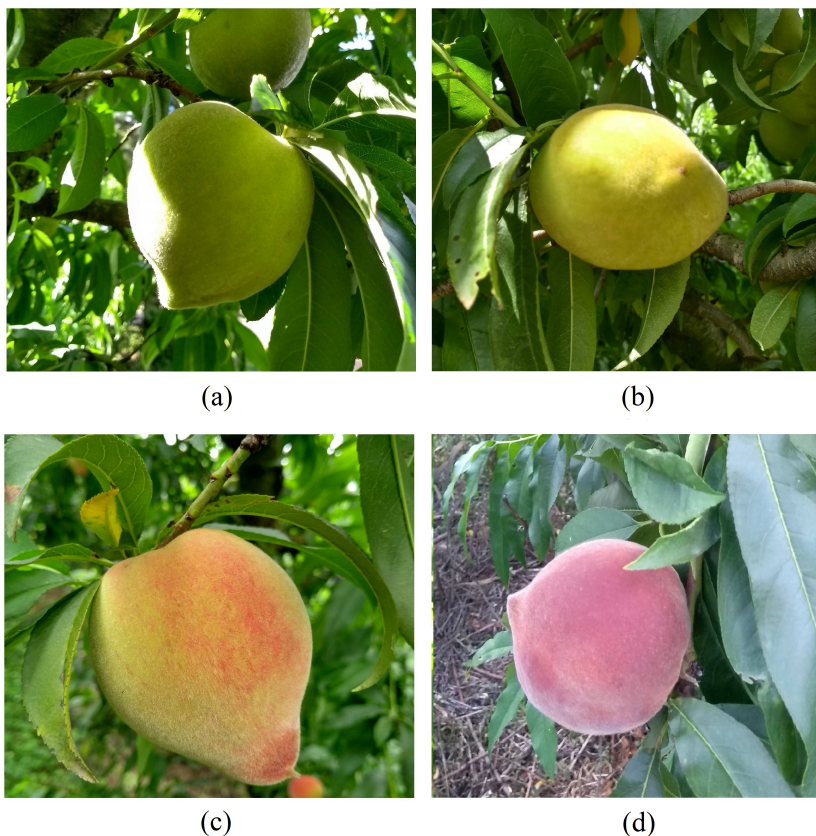
O Brasil ocupa a décima segunda posição no ranking mundial de produção bruta do fruto, com 238,5 mil toneladas em uma área total de 24,2 mil hectares, sendo que a região sul é responsável pela produção de mais de 95% do total produzido no Brasil (BALKE, 2016).

2.2 CARACTERÍSTICAS DO FRUTO

O pessegueiro *Prunus persica* (L.) Batsch é uma espécie de frutífera, cultivada em clima temperado, que mais tem sido pesquisada e adaptada às condições de clima subtropical ou temperado quente (CAMILLO, 2009). O fruto é muito apreciado em grande parte do planeta, por sua aparência e sabor. Segundo Camillo (2009), o consumo per capita, no Brasil, é de cerca 700 a 800g ao ano, com perspectivas para consumo de cerca de 100 mil toneladas de pêssegos por ano.

Aqueles que consomem o fruto, a qualidade deste pode ser dada por uma combinação de características e atributos como: aparência, firmeza, sabor e valores nutricionais. Durante o processo de amadurecimento do pêssego, é evidente a alteração da coloração da polpa e da casca, e amolecimento da polpa. A mudança na coloração ocorre com perda da cor verde e desenvolvimento do tons de cores como o amarelo, o vermelho, além de outras que são características do cultivar (SANTOS, 2007). A Figura 1 ilustra os frutos do cultivar Eragil em diferentes estágios evolutivos de maturação no pessegueiro, em que (a) o fruto está verde, (b) o fruto possui uma coloração amarelada, (c) meio maduro e (d) maduro, tendendo a super maturação.

Figura 1 – Pêssegos da cultivar Eragil, em diferentes estágios de maturação: (a) verde (b) amarelo (c) meio maduro (d) maduro



Fonte: O Autor

Balke (2016) mostra que o período de oferta do pêssgo brasileiro se concentra nos meses de setembro até o mês de janeiro. Para o cultivar Eragil, a plena floração geralmente ocorre no fim de agosto e a maturação inicia-se ao final do mês de outubro.

A cultivar Eragil vem sendo difundida para a produção de pêssegos de mesa. Uma vez que essas cultivares de polpa amarela predominam no cultivo industrial, as cultivares de polpa branca, que correspondem a maior parte da oferta para mesa, normalmente atendem mercados regionais, pois em geral apresentam problemas de transporte a longas distâncias (CAMILLO, 2009). A cultivar Eragil possui como características ter frutos de tamanho médio a grande, de polpa firme com coloração amarela e mais avermelhada próxima ao caroço. O caroço é solto. Há poucas cultivares de polpa amarela destinadas para o consumo *in natura* (CAMILLO, 2009). O fruto do Eragil apresenta tamanho médio, com peso superior a 120g, sendo considerado um cultivar de duplo propósito, ou seja, pode ser destinado ao consumo *in natura* ou industrial (BALKE, 2016). A colheita do pêssgo Eragil, na região dos Campos Gerais, é normalmente realizada no final do mês de novembro.

2.3 CICLO DE CRESCIMENTO DOS PÊSSEGOS

Utilizando GC-MS (*Gas Chromatography and Mass Spectrometry* - cromatografia gasosa e espectrometria de massa) (HORVAT et al., 1990; BARALDI et al., 1999) identificaram os principais COVs exalados pelo pêssgo e as árvores frutíferas durante e após a fase de plena floração. A Tabela 3 mostra os principais gases gerados como subprodutos nesse processo.

Tabela 3 – COVs identificados por GC / MS no processo de crescimento dos pêssegos.

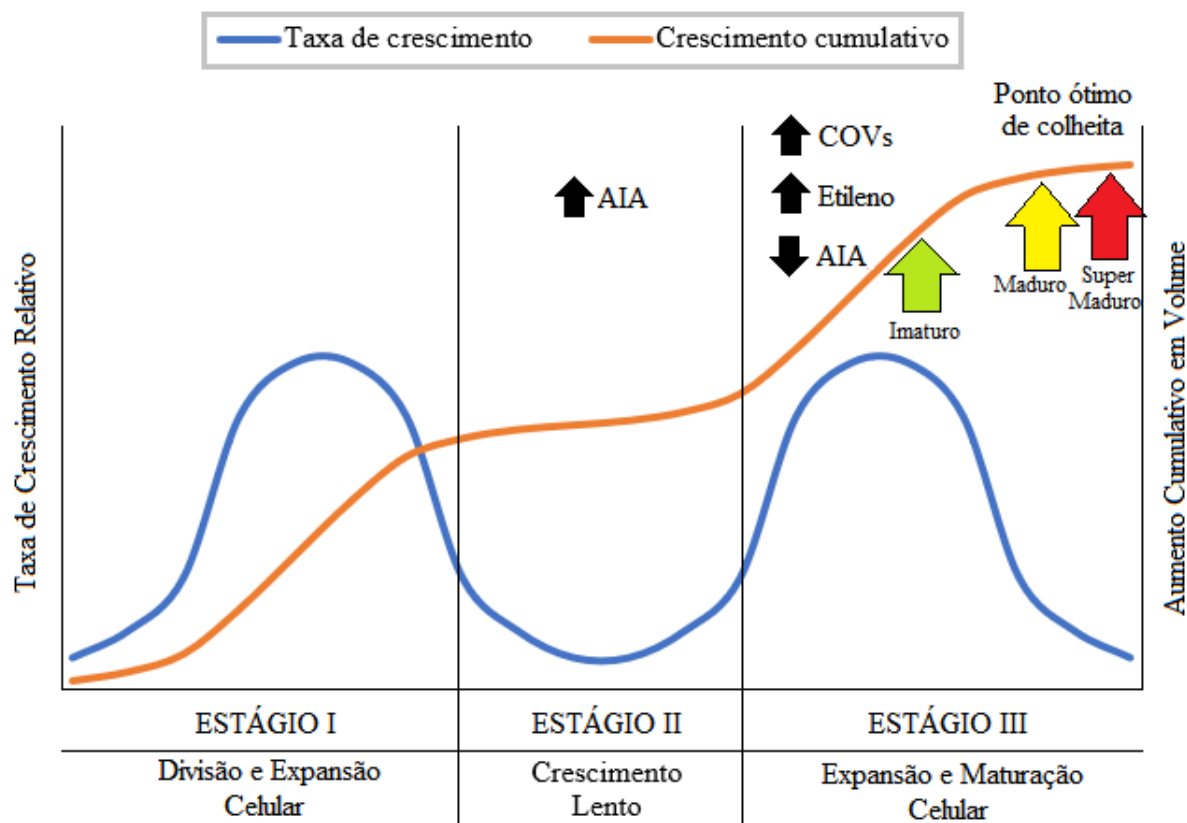
Classe	Composto	Classe	Composto
Aldeídos	Hexanal	Lactonas	Terpinoleno
	(E)-2-hexenal		Linalol
	Benzaldeído		γ -hexalactona
Álcoois	1-Hexanol		γ -heptalactona
	(E)-2-hexenol		γ -octalactona
	(Z)-3-hexenol		γ -nonalactona
	Acetato de hexila		γ -decalactona
Esters	Acetato de (E)-2-hexenilo		γ -dodecalactona
Cetonas	Acetoína		δ -decalactona
Terpenoides	α -terpineno		δ -dodecalactona
	γ -terpineno		

Fonte: (HORVAT et al., 1990; BARALDI et al., 1999)

A Figura 2 ilustra a curva sigmoide dupla, típica da taxa de crescimento do fruto do pêssgo, além da curva de crescimento cumulativo (integral ou somatória da taxa de crescimento). Os picos elevados nas taxas de crescimento demonstram a grande quantidade de COVs gerados como subproduto do processo de maturação.

Basicamente o desenvolvimento do fruto em climas subtropicais se dá em três estágios.

Figura 2 – Curva de crescimento sigmóide duplo do fruto de pêssgo



Adaptado de (KESKE, 2009)

No estágio I não há alteração hormonal significativa. O estágio I, para esse cultivar, compreende os primeiros 35 dias após a floração, caracterizado pelo rápido crescimento do pericarpo (camada externa do fruto que envolve as sementes). A alteração mais significativa ocorre no estágio II, com elevação dos aldeídos, ésteres e cetonas. Com a elevação destes compostos tem-se importantes alterações fisiológicas e anatômicas do fruto: diminuição do crescimento da polpa, endurecimento do caroço e formação total ou parcial da semente. No estágio II, aproximadamente do 36º ao 49º dia da floração, ocorre a parada de crescimento do pericarpo. Após o 51º dia da florada, inicia-se o estágio III: caracterizado por um intenso e contínuo crescimento do pericarpo, perdurando até a maturação do fruto (JUNIOR et al., 2007). No estágio III, após cerca de 20 dias de frutificação, há a redução da concentração de AIA (Ácido indolacético) na semente, e a polpa volta a crescer (LILIEN-KIPNIS; LAVEE, 1971). Na maturação é onde acontecem as principais mudanças físicas e bioquímicas do fruto, como a produção de terpenóides; mudanças na cor, na taxa respiratória, na permeabilidade dos tecidos e na textura; e transformações químicas que atingem os açúcares, ácidos orgânicos, proteínas, pigmentos, entre outras. Durante a fase de amadurecimento, os sabores e odores específicos, juntamente com aumento na doçura e diminuição da acidez, tornam-se mais acentuados. É também nesse período que ocorre o amaciamento do fruto em conjunto com mudanças na coloração (BARBOSA et al., 1993).

O fruto maduro é caracterizado por mudanças na coloração, aumento do teor de açúcares, diminuição da acidez, intensificação no aroma e diminuição da firmeza da polpa, que a torna mais sucosa. Como órgão de reserva, é constituído por 85 a 90% de água, além dos sais minerais e outros sintetizados como carboidratos, lipídeos, proteínas, ácidos orgânicos, compostos fenólicos, flavonóides, carotenóides e substâncias voláteis (BARBOSA et al., 1990). O ácido abscísico provoca senescência (envelhecimento e murchamento) nos frutos, acelerando também a maturação. Portanto é o responsável por haver a supermaturação. O desenvolvimento da casca e do núcleo é contínuo até atingir o tamanho máximo, ao término da nona semana da floração. Após isso, ocorre a abscisão natural do fruto, até que este seja mantido até a completa maturação. Durante esse processo de crescimento, a temperatura exerce papel essencial na qualidade e maturação dos frutos. Se ocorrer, por exemplo, queda brusca e contínua na temperatura, reduzindo o metabolismo da planta, principalmente durante o estágio II, o ciclo total de crescimento dos frutos pode ser bastante afetado, pois o estágio III somente se inicia após o total endurecimento do caroço (COUTO, 2006).

A cultura do pêssgo tem se expandido, com o plantio de mais de 500 mil mudas/ano em todo o sul do Brasil (DUTRA; KERSTEN; FACHINELLO, 2002). Tendo em vista as características físico-químicas do crescimento do fruto do pêssgo, para este trabalho foi realizado o monitoramento da cultivar Eragil na cidade de Ponta Grossa, Paraná. Para isso, foi construído um equipamento não-destrutivo, aplicado próximo à cultura, capaz de detectar os gases exalados pelo pêssgo, a fim de identificar o amadurecimento nos diferentes ciclos, desde a plena floração até o período pós-colheita.

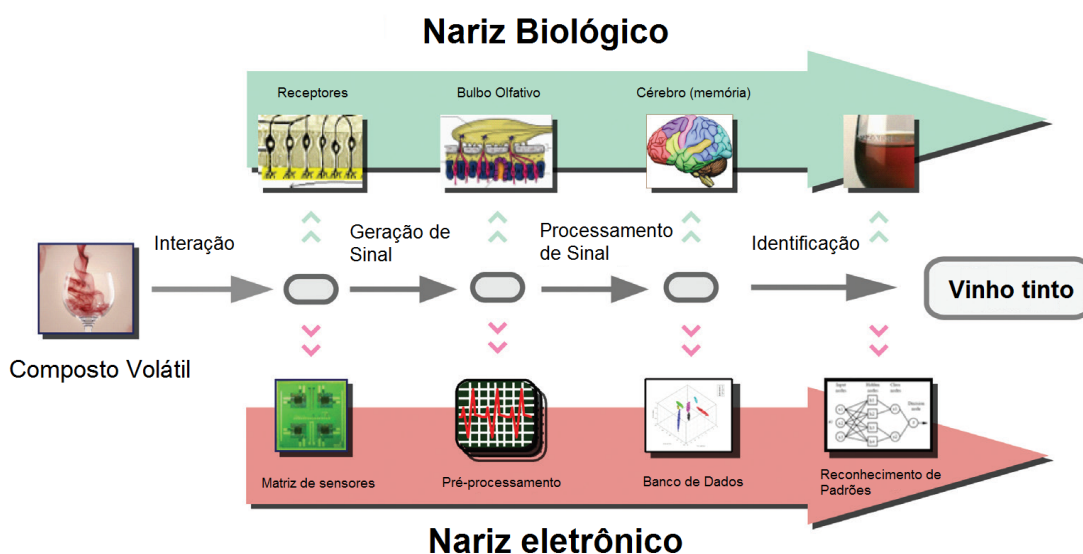
2.4 NARIZES ELETRÔNICOS

O termo nariz eletrônico pode fazer pensar em um dispositivo capaz de simular o olfato humano, onde a maior semelhança deste dispositivo está na função biológica em contrapartida com a sua forma física. A Figura 3 demonstra essa similaridade. Para o Nariz Biológico a amostra é detectada pelos receptores, que por sua vez gera um sinal que é enviado para o bulbo olfativo. O bulbo olfativo processa estes sinais, enviando-os para o cérebro, que por sua vez através da memória identifica a amostra. Para o Nariz Eletrônico o processo é similar: na primeira etapa a interação com a amostra é realizada pela matriz de sensores, que realizará a geração do sinal na qual será pré-processado por algoritmos específicos, e assim geram dados que serão armazenados num banco de dados e, por fim, será feita a identificação do composto através de ferramentas de reconhecimento de padrões.

De acordo com Santonico et al. (2013), termos como matriz de sensores de gás ou sistema multissensorial podem ser usados, mas isto depende do autor em questão. No que tange a literatura científica, o termo nariz eletrônico está mais incorporado do que esses citados. O exemplo mais comum que se tem sobre esses dispositivos seria uma matriz de sensores químicos baseado em materiais de óxido metálico, trabalhando juntamente com um sistema eletrônico

associado à inteligência artificial, como uma RNA (*Artificial Neural Network* - Rede Neural Artificial), *software* com algoritmos e uma base de dados, com bibliotecas de referência, com o propósito de reconhecimento de padrões (PINEDA; PÉREZ, 2017; SALES et al., 2014; WILSON; BAIETTO, 2009; WILSON, 2013). Como receptores do olfato humano esses equipamentos são mais sensíveis a uma determinada gama de substâncias em detrimento de outras. O nariz eletrônico pode ser usado como ferramenta analítica para qualquer sistema emissor de COVs, como por exemplo, os solos (PINEDA; PÉREZ, 2017).

Figura 3 – Semelhança entre o sistema biológico olfativo e um nariz eletrônico



Adaptado de: (SANTOS; LOZANO; ALEIXANDRE, 2017)

Os narizes eletrônicos utilizam uma matriz de sensores. Uma matriz de sensores é um grupo de sensores individuais, dispostos lado a lado, com o propósito de adicionar dimensões na medida/observação. Com isso, pode-se aumentar o desempenho da medida, provendo uma melhora na precisão dessa observação. Um sensor individual é usualmente definido como um dispositivo que recebe um estímulo e responde com um sinal elétrico, como um sensor de temperatura, umidade ou luminosidade (FRADEN, 2004). Um exemplo de sensor matricial é o sensor CCD (*Charge Coupled Device* - Dispositivo de Carga Acoplada), utilizado em câmeras digitais, em que cada sensor prove uma resposta a luz projetada na sua superfície.

A matriz de sensores de resposta cruzada é composta por diferentes sensores, escolhidos para responder a uma ampla gama de classes químicas e discriminar possíveis substâncias em uma amostra composta por uma mistura de diversos elementos. A saída de sensores individuais é coletivamente montada e integrada para produzir um padrão de resposta digital distinto. A identificação e classificação de uma mistura de uma substância é realizada através do reconhecimento desta impressão digital eletrônica vinda das respostas dos vários sensores, utilizando uma RNA desenvolvida para este fim. A identidade de uma mistura simples ou complexa, representada por

um padrão de assinatura exclusivo, pode então ser determinada sem ter que separar a mistura em seus componentes individuais.

2.4.1 TIPOS E CARACTERÍSTICAS DE SENSORES DE GÁS EM NARIZES ELETRÔNICOS

Liu et al. (2012) expõem que a escolha de um tipo de sensor adequado para uma determinada aplicação deve ser ditada por condições em que a medição será realizada. Além disso, deve ser levado em consideração a durabilidade dos sensores e também a composição da mistura gasosa a ser analisada. Vale ressaltar que os sensores de gás disponíveis no mercado possuem valor de sensibilidade ajustável através da resistência de carregamento, que é a resistência equivalente do sensor ligada à saída do circuito.

Como características, informadas por dois dos principais fabricantes de sensores de gás, MQ e TGS, tem-se: a tensão de operação; corrente de aquecimento, que é a corrente necessária para aquecer o dispositivo até a correta faixa de operação; resistência de detecção, que é a faixa de variação da resistência conforme a quantidade de gás amostrado; resistência de carregamento, na qual se refere a faixa de resistência a ser colocada em paralelo ao circuito do sensor a fim de alterar a sensibilidade; tempo de pré-aquecimento, que é o tempo necessário para o dispositivo responder as características informadas pelo fabricante; e os gases que o sensor é capaz de detectar. Esses parâmetros influenciam quando for dimensionar e projetar o dispositivo de nariz-eletrônico e devem ser levados em consideração.

Além disso, esses sensores utilizam um aquecedor interno em conjunto a um sensor eletroquímico. O princípio de funcionamento ocorre da seguinte forma: em condições normais, a resistência do sensor será elevada, e portanto a queda de tensão na carga será baixa e constante. Havendo a detecção de algum gás, a resistência do sensor irá diminuir, portanto, mais corrente irá fluir do resistor de carga e consequentemente haverá o aumento da tensão. Então, essa tensão de saída aumenta com o aumento da concentração de gás no ar.

2.5 PRÉ-PROCESSAMENTO DOS DADOS

Embora vários métodos de pré-processamento de dados tenham sido desenvolvidos, o pré-processamento de dados continua sendo uma área ativa de pesquisa, devido à enorme quantidade de dados inconsistentes e/ou à complexidade dos problemas. Segundo Han, Pei e Kamber (2011) os passos envolvidos no pré-processamento dos dados podem ser divididos em:

- **Qualidade dos dados:** é definida em termos de exatidão, integridade, consistência, pontualidade, credibilidade e interpretabilidade. Essas qualidades são avaliadas com base no uso pretendido dos dados. Como exemplos tem a aplicação de algoritmos de imputação de dados faltantes e a transformação dos atributos para o tipo numérico.

- **Limpeza dos dados:** tentam preencher os valores ausentes, suavizar o ruído ao identificar *outliers* e corrigir inconsistências nos dados. A limpeza de dados é geralmente executada como um processo iterativo de duas etapas, que consiste em detecção de discrepância e alteração desses dados. Um exemplo típico de limpeza dos dados é a aplicação do filtro da média móvel, no qual para cada novo valor do conjunto de dados é realizada a média deste com uma quantidade n de valores anteriores. Esse filtro possibilita a remoção de possíveis picos da curva de respostas geradas pelos sensores devido a ruídos provenientes do circuito.
- **Integração dos dados:** combina dados de várias origens para formar um armazenamento de dados coerente. A resolução da heterogeneidade semântica, metadados, análise de correlação, detecção de duplicação de tuplas e detecção de conflitos de dados contribuem para suavizar a integração de dados.
- **Normalização de dados:** convertem os dados em formas apropriadas para mineração. O método mais comum de normalização de dados é a máximo-mínimo, no qual os dados de atributo são dimensionados de modo a se enquadrarem em um intervalo pequeno, de 0,0 a 1,0. A etapa de normalização máximo-mínimo ($p_{i_{norm}}$), é dada pela equação 4.2:

$$p_{i_{norm}} = \frac{p_i - \min(p)}{\max(p) - \min(p)} \quad (2.1)$$

onde $\min$ é o valor mínimo e $\max$ é o valor máximo do conjunto p .

A normalização é particularmente útil para algoritmos de classificação envolvendo redes neurais ou medições de distância, como classificação e agrupamento de vizinhos mais próximos. Se estiver usando o algoritmo de *backpropagation*¹ da rede neural para classificação, a normalização dos valores de entrada para cada atributo medido na fase de treinamento ajudará a acelerar a fase de aprendizado. Para métodos baseados em distância, a normalização ajuda a evitar que atributos com intervalos inicialmente grandes superem atributos com intervalos inicialmente menores.

- **Redução dos dados:** estas técnicas obtêm uma representação reduzida dos dados enquanto minimizam a perda de conteúdo informativo. Como exemplos tem-se os métodos de redução de dimensionalidade, redução de numerosidade e compressão de dados. A redução de dimensionalidade reduz o número de variáveis aleatórias ou atributos em consideração. Os métodos incluem transformações *wavelet*, PCA (*Principal Component Analysis* - Análise de Componentes Principais) e seleção de subconjunto de atributos (*Features Selection*). Os métodos de compactação de dados aplicam transformações para obter uma representação reduzida ou compactada dos dados originais.

¹O algoritmo de *backpropagation* trata-se de um algoritmo com treinamento supervisionado. Com o erro calculado, o algoritmo corrige os pesos em todas as camadas, partindo da camada de saída indo até a de entrada

2.5.1 ANÁLISE DE REDUNDÂNCIA E CORRELAÇÃO

A análise de redundância e correlação é uma questão importante na integração de dados. Um atributo pode ser redundante se puder ser derivado de algum conjunto de atributos. Inconsistências na atribuição ou nomeação de dimensão também podem causar redundâncias no conjunto de dados resultante. Algumas redundâncias podem ser detectadas por análise de correlação. Dados dois atributos, essa análise pode medir quão fortemente um atributo implica no outro, com base em todo conjunto de dados. Para dados nominais, usamos o teste χ^2 (*chi-quadrado*). Para atributos numéricos, podemos usar a covariância e o coeficiente de correlação, os quais mostram como os valores de um atributo variam em relação aos outros (HAN; PEI; KAMBER, 2011).

2.5.2 SELEÇÃO DE VARIÁVEIS

A seleção de variáveis ou seleção de subconjunto de atributos é uma técnica de pré-processamento de dados que seleciona um subconjunto de características originais e que objetiva reduzir o número de características, remover dados irrelevantes, redundantes ou com ruídos e que traz efeitos imediatos para as aplicações como: aumentar a velocidade de processamento de um algoritmo de mineração de dados, melhorar o desempenho de mineração, com o aumento da precisão preditiva e compreensão do resultado. Outro benefício imposto pela redução da dimensionalidade dos dados é para se evitar o sobre-ajuste ou *overfitting*² do modelo, uma vez que são exigidas mais instâncias de treinamento a medida que a dimensionalidade do conjunto de dados aumenta (LIU; YU, 2005; WANG; TANG; LIU, 2016).

A geração de subconjunto é um procedimento de pesquisa que produz subconjuntos de atributos candidatos para avaliação com base em uma determinada estratégia de pesquisa. Cada subconjunto candidato é avaliado e comparado com o melhor anterior, de acordo com um determinado critério de avaliação. Se o novo subconjunto for melhor, substitui o melhor subconjunto anterior. O processo de geração e avaliação de subconjunto é repetido até que um determinado critério de parada seja satisfeito. Então, o melhor subconjunto selecionado geralmente precisa ser validado por conhecimento prévio ou testes diferentes via conjuntos de dados sintéticos e/ou reais. A seleção de variáveis pode ser encontrada em muitas áreas de mineração de dados, como classificação, agrupamento, regras de associação e regressão. Essa seleção pode ser classificada em: seleção de atributos supervisionada, quando as informações de classe estão disponíveis; e não-supervisionada, quando as informações não estão disponíveis (LIU; YU, 2005; WANG; TANG; LIU, 2016).

A natureza do processo de geração do subconjunto é determinada por dois problemas básicos. Primeiro, deve-se decidir o ponto de partida da pesquisa que, por sua vez, influencia

(TAN; STEINBACH; KUMAR, 2005).

²Termo utilizado em estatística para descrever quando um modelo se ajusta muito bem ao conjunto de dados de treino, mas se mostra ineficaz para prever novos resultados através dos dados de teste.

a direção da busca. A pesquisa pode começar com um conjunto vazio e adicionar recursos sucessivamente, ou iniciar com um conjunto completo e remover sucessivamente os recursos, ou iniciar com ambas as extremidades e adicionar e remover recursos simultaneamente. Após a primeira etapa, é preciso decidir uma estratégia de busca. Para um conjunto de dados com N recursos, existem 2^N subconjuntos candidatos. Esse espaço de busca é exponencialmente proibitivo para uma busca exaustiva, com um moderado N . Portanto, diferentes estratégias podem ser exploradas como: busca completa, sequencial e aleatória (LIU; YU, 2005).

- **Busca Completa:** garante encontrar o melhor resultado de acordo com o critério de avaliação utilizado. Durante uma pesquisa exaustiva nenhum subconjunto ideal é perdido, ou seja, o espaço de busca será 2^N .
- **Busca Sequencial:** não se dá completamente e, portanto, corre o risco de perder os subconjuntos ideais, mas obtém um resultado mais rápido por conta do espaço de busca ser reduzido.
- **Busca Aleatória:** começa com um subconjunto selecionado aleatoriamente e prossegue de duas maneiras diferentes. Uma delas é seguir a busca sequencial, que injeta aleatoriedade nela. A outra é gerar o próximo subconjunto de uma maneira completamente aleatória.

Segundo Wang, Tang e Liu (2016), os métodos de seleção de atributos projetados com diferentes critérios de avaliação se dividem em duas categorias:

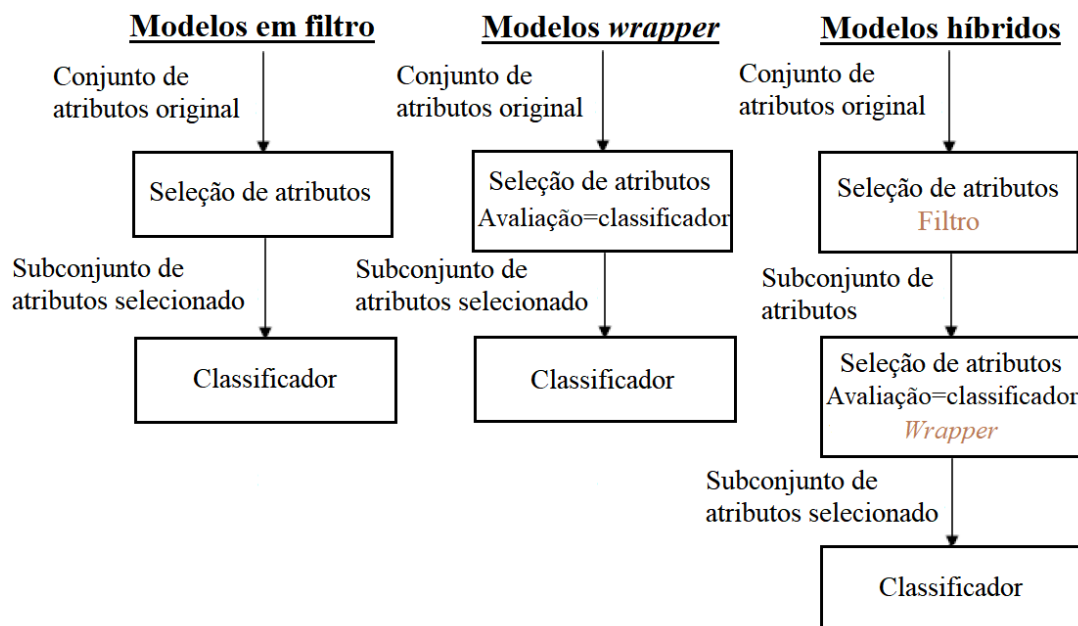
- **Modelos de filtro:** para modelos de filtro, os atributos são selecionados e avaliados com base nas características gerais dos dados sem a utilização de algoritmos de aprendizado. Portanto, modelos em filtro são bastante eficientes. Usualmente utilizam-se ferramentas estatísticas nesse tipo de abordagem. Assim, pode ocorrer de perder características relevantes dependendo do algoritmo de aprendizado utilizado. Um algoritmo de filtro geralmente consiste em duas etapas. Na primeira etapa, as características são classificadas com base em determinado critério. Na segunda etapa, as características com as classificações mais altas são escolhidas (MHAMDI; MHAMDI, 2014; WANG; TANG; LIU, 2016).
- **Modelos *wrapper*:** o subconjunto de atributos ideal deve depender dos vieses e heurísticas específicas dos algoritmos de aprendizado. Com base nessa suposição, os modelos *wrapper* usam um algoritmo de aprendizado específico para avaliar a qualidade das características selecionadas. O componente de pesquisa de características produzirá um conjunto de atributos com base em determinadas estratégias de pesquisa. O componente de avaliação de características usará o algoritmo de aprendizado predefinido para avaliar o desempenho, que será retornado ao componente de pesquisa de característica para a próxima iteração da seleção de subconjunto de atributos. O conjunto de atributos com o melhor desempenho

será escolhido como o conjunto final (MHAMDI; MHAMDI, 2014; WANG; TANG; LIU, 2016).

- **Modelos híbridos:** os modelos híbridos são uma combinação entre os dois modelos citados, incorporando a seleção de atributos na construção do modelo. Assim, os modelos híbridos, que aproveitam os modelos de filtro e os modelos *wrapper*: (1) são menos intensivos em computação do que os métodos *wrapper*, pois não precisam executar os modelos de aprendizado várias vezes para avaliar as características e (2) incluem a interação com o modelo de aprendizagem. A maior diferença entre modelos *wrapper* e modelos híbridos é que os modelos *wrapper* treinam por primeiro modelos de aprendizado usando os atributos candidatos e executam a seleção de atributos usando o modelo de aprendizado, enquanto modelos híbridos selecionam atributos durante o processo de construção do modelo para executar a seleção de atributos sem avaliação adicional das características (MHAMDI; MHAMDI, 2014; WANG; TANG; LIU, 2016).

A Figura 4 ilustra as diferenças entre os três modelos de seleção de atributos explicados anteriormente.

Figura 4 – Comparativo entre as abordagens de seleção de subconjuntos de atributos



Adaptado de: (MHAMDI; MHAMDI, 2014)

2.5.3 ANÁLISE DE DADOS MULTIVARIADOS - MDA

A MDA (*Multivariate Data Analysis* - Análise de Dados Multivariados) compreende técnicas para analisar um conjunto de dados com mais de uma variável, reduzindo altas dimensões

em um problema multivariado, quando as variáveis estão parcialmente correlacionadas, para que possam ser exibidas em duas ou três dimensões (WILSON; BAIETTO, 2009). A análise multivariada pode ser dividida em técnicas treinadas ou não-treinadas. Técnicas de aprendizado treinadas ou supervisionadas classificam amostras desconhecidas com base em características de amostras conhecidas ou conjuntos de amostras com propriedades conhecidas, geralmente mantidas em uma biblioteca de referência acessada durante a análise. Já as técnicas não-treinadas são usadas quando um banco de dados de amostras conhecidas não foi construído anteriormente, portanto, não é necessário e nem se deseja reconhecer a própria amostra, mas se quer fazer comparações entre diferentes amostras desconhecidas para discriminá-las. A técnica MDA não-treinada mais simples e amplamente utilizada é a PCA (WILSON; BAIETTO, 2009).

2.5.3.1 ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

A PCA é um dos mais conhecidos e possivelmente o mais antigo método de análise de dados multivariados. O PCA analisa o conjunto de dados fazendo observações descritas por diversas variáveis dependentes, normalmente correlacionadas entre si. O objetivo desse método é extrair os dados mais importantes do conjunto de dados e expressá-los em um conjunto de variáveis ortogonais denominados componentes principais. Esse método representa o padrão de similaridade das observações e variáveis através de pontos no espaço (WOLD; ESBENSEN; GELADI, 1987; ABDI; WILLIAMS, 2010).

Assim, esta análise multivariada possui como objetivos:

1. Extrair as informações mais importantes do conjunto de dados;
2. Reduzir o tamanho do conjunto de dados, mantendo as informações mais importantes;
3. Analisar a estrutura de variáveis e observações.

A fim de alcançar esses objetivos, o PCA processa as novas variáveis chamadas de componentes principais que são obtidas através de combinações lineares das variáveis originais. É necessário que o primeiro componente principal possua a maior variância entre todas, ou seja, esse componente irá extrair a maior inércia do conjunto de dados, que é a soma dos quadrados dos componentes de cada dimensão. O segundo componente é processado sob a restrição de ter a maior inércia possível e que seja ortogonal ao primeiro componente principal. Os mesmos procedimentos são realizados para as demais dimensões ou variáveis. Os valores dessas novas variáveis são chamados de pontuações do fator, e estas pontuações podem ser interpretadas geometricamente como as projeções das observações para os componentes principais (ABDI; WILLIAMS, 2010).

A PCA é mais útil quando nenhuma amostra conhecida está disponível ou quando as relações ocultas entre amostras ou variáveis são suspeitas (WILSON; BAIETTO, 2009). Isso

fornece uma avaliação preliminar, mas eficiente, da capacidade do sistema de discriminar entre diferentes classes olfativas. Essa técnica ajuda a reduzir a complexidade de todo o conjunto de dados, reduzindo o espaço n-dimensional dos recursos extraídos pelas respostas do sensor para um espaço de dimensões menores, mantendo a maioria das informações contidas no conjunto de dados. Isso é possível porque as informações contidas no conjunto de respostas dos sensores geralmente são redundantes. Esse método também torna mais simples visualizar qualquer eventual agrupamento dos dados, ou seja, o agrupamento de medições relativas à mesma classe olfativa (CAPELLI; SIRONI; ROSSO, 2014)

2.5.3.2 ANÁLISE DISCRIMINANTE LINEAR - LDA

A LDA (*Linear Discriminant Analysis* - Análise Discriminante Linear) é uma técnica que pode ser empregada para classificação linear de dados, mas é mais comum utilizá-la, assim como o PCA, para redução da dimensionalidade de dados antes da etapa de classificação. Essa técnica trabalha facilmente com casos em que as frequências dentro da classe são desiguais e em análise de dados de teste gerados aleatoriamente (BALAKRISHNAMA; GANAPATHIRAJU, 1998; MARTÍNEZ; KAK, 2001).

Esse método maximiza a proporção da variação entre as classes para a variação dentro da classe em qualquer conjunto de dados específico, garantindo assim a máxima capacidade de separação (BALAKRISHNAMA; GANAPATHIRAJU, 1998).

A LDA procura por aqueles vetores no espaço subjacente que melhor discriminam entre as classes, ao invés daquelas que melhor descrevem os dados. Mais formalmente, dado um número de características independentes em relação às quais os dados são descritos, o LDA cria uma combinação linear destes que gera as maiores diferenças médias entre as classes desejadas. Os conjuntos de dados podem ser transformados e os vetores de teste podem ser classificados no espaço com transformação dependente e independente de classe (MARTÍNEZ; KAK, 2001) Portanto, esse método gera como respostas vetores de direção canônica LD1, LD2,..., LDm, em que m é o número de vetores possíveis que gerem 100% da variabilidade dos dados, ou o número de dimensões reduzidas geradas pelo método.

2.6 MÉTODOS DE RECONHECIMENTO DE PADRÕES

Após o pré-processamento das informações provenientes dos sensores, obtém-se um conjunto de dados numéricos, que podem ser processados posteriormente por sistemas adequados de reconhecimento e classificação e/ou quantificação.

2.6.1 K-VIZINHOS MAIS PRÓXIMOS - KNN

O método KNN (*K-Nearest Neighbours* - K-vizinhos mais próximos) é amplamente utilizado porque é conceitualmente simples e relativamente fácil de implementar, fornecendo

resultados satisfatórios (CAPELLI; SIRONI; ROSSO, 2014).

O classificador KNN é baseado em aprendizado por analogia, ou seja, cada registro novo da base a ser classificado é comparado com os registros da base de treinamento que são semelhantes a esta nova tupla. Cada registro da base de dados é representado por n atributos, e portanto cada um desses registros podem ser representados por um ponto no espaço n -dimensional. Para cada novo registro desconhecido, o classificador pesquisa no espaço n -dimensional as tuplas de treinamento (k) que estão mais próximas deste novo registro, e assim, os k registros são denominados de "k-vizinhos mais próximos" do novo registro desconhecido. Assim, esse classificador possui como principal característica exigir um grande poder computacional quando se recebe um elevado conjunto de dados na base de treinamento (HAN; PEI; KAMBER, 2011).

A proximidade entre os pontos no espaço é definida em termos de distância métrica, como a distância euclidiana. A distância euclidiana entre dois pontos $\mathbf{X}_1$ e $\mathbf{X}_2$, em que $\mathbf{X}_1 = (X_{11}, X_{12}, \dots, X_{1n})$ e $\mathbf{X}_2 = (X_{21}, X_{22}, \dots, X_{2n})$ é definida segundo a equação 2.2:

$$dist(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{\sum_{i=1}^n (X_{1i} - X_{2i})^2}. \quad (2.2)$$

Ou seja, para cada atributo numérico do registro, é feita a diferença entre os valores correspondentes do atributo, nas tuplas $\mathbf{X}_1$ e $\mathbf{X}_2$, é feita a somatória do quadrado desta diferença, e então calculada a raiz quadrada desta somatória. Antes do cálculo de distância, segundo a equação 2.2, é aconselhável realizar a normalização dos valores de cada atributo, pois isto ajuda a evitar que atributos com intervalos inicialmente grandes superem atributos com intervalos inicialmente menores (HAN; PEI; KAMBER, 2011).

2.6.2 MÁQUINAS DE VETOR DE SUPORTE - SVM

A SVM (*Support Vector Machine* - Máquina de Vetores de Suporte) é um método aplicável tanto para classificação de conjuntos de dados lineares quanto para conjuntos não-lineares. Em suma, esse método funciona da seguinte maneira: é realizado um mapeamento não-linear para transformar os dados de treinamento originais para uma dimensão mais alta. Dentro dessa nova dimensão, é calculado um hiperplano de separação ideal linear, ou seja, um limitador que seja capaz de separar os registros entre suas classes. Realizando um mapeamento não-linear apropriado para uma dimensão suficientemente alta, os dados entre duas classes sempre podem ser separados através de um único hiperplano. O SVM encontra esse hiperplano usando vetores de suporte (tuplas de treinamento essenciais) e margens (definidas pelos vetores de suporte) (HAN; PEI; KAMBER, 2011).

Mesmo os SVMs mais rápidos ainda exigirem um tempo de treinamento extremamente alto, eles são altamente precisos, devido à sua capacidade de modelar limites complexos de decisão não-lineares. Os SVMs, em comparação a outros métodos de classificação são menos

propensos a *overfitting*.

2.6.2.1 CONJUNTO DE DADOS LINEARMENTE SEPARÁVEIS

Considerando o conjunto de dados D ser dado como $(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_{|D|}, y_{|D|})$, onde $\mathbf{X}_i$ é o conjunto de tuplas de treinamento com rótulos de classe associados, y_i , e cada y_i pode ter um de dois valores, $+1$ ou -1 , podem existir um número infinito de linhas separadoras. Deseja-se encontrar aquela que terá o erro mínimo de classificação em tuplas não vistas anteriormente. Se os dados fossem em 3 dimensões, ou seja, com três atributos, é desejável encontrar o melhor plano de separação. Generalizando para n dimensões, então é necessário encontrar o melhor hiperplano (HAN; PEI; KAMBER, 2011).

Ainda segundo Han, Pei e Kamber (2011), o SVM aborda esse problema pesquisando o MMH (*Maximum Marginal Hyperplane* - Hiperplano Marginal Máximo). Intuitivamente, no entanto, é esperado que o hiperplano com a margem maior seja mais preciso na classificação de tuplas de dados futuros do que o hiperplano com a margem menor. É por isso que (durante a fase de aprendizado ou treinamento) o SVM procura o hiperplano com a maior margem, ou seja, o MMH. A margem associada fornece a maior separação entre as classes.

Chegando a uma definição de margem, pode-se dizer que a menor distância de um hiperplano para um lado de sua margem é igual à menor distância do hiperplano ao outro lado de sua margem, onde os lados da margem são paralelo ao hiperplano. Ao trabalhar com o MMH, essa distância é, na verdade, a distância mais curta do MMH até a tupla de treinamento mais próxima de qualquer classe (HAN; PEI; KAMBER, 2011).

Um hiperplano de separação pode ser escrito como:

$$\mathbf{W} \cdot \mathbf{X} + b = 0 \quad (2.3)$$

onde $\mathbf{W}$ é um vetor de peso, ou seja, $\mathbf{W} = (w_1, w_2, \dots, w_M)$; M é o número de atributos; b é um escalar, às vezes referido como viés. As tuplas de treinamento são em 2 dimensões, por exemplo, $\mathbf{X} = (x_1, x_2)$, onde x_1 e x_2 são os valores dos atributos A_1 e A_2 , respectivamente, para $\mathbf{X}$. Se b for como um peso adicional, w_0 , 2.3 pode ser reescrito como:

$$w_0 + w_1x_1 + w_2x_2 = 0. \quad (2.4)$$

Os pesos podem ser ajustados para que os hiperplanos que definem os lados da margem possam ser escritos segundo as equações 2.5 e 2.6:

$$H_1 : w_0 + w_1x_1 + w_2x_2 \geq 1, \text{ para } y_i = +1. \quad (2.5)$$

$$H_2 : w_0 + w_1x_1 + w_2x_2 \leq 1, \text{ para } y_i = -1. \quad (2.6)$$

Ou seja, qualquer tupla que caia em H_1 ou acima dela pertence à classe +1, e qualquer tupla que caia abaixo de H_2 pertence à classe -1. Combinando as duas desigualdades das equações 2.5 e 2.6, obtêm-se:

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \forall i. \quad (2.7)$$

Quaisquer tuplas de treino que caiam nos hiperplanos H_1 ou H_2 (isto é, os lados que definem a margem) satisfazem a equação 2.7 e são chamados de vetores de suporte. Ou seja, eles estão igualmente próximos MMH. Essencialmente, os vetores de suporte são as tuplas mais difíceis de classificar e dar mais informações sobre classificação (HAN; PEI; KAMBER, 2011).

A partir disso, pode-se obter uma fórmula para o tamanho da margem máxima. A distância do hiperplano de separação a qualquer ponto em H_1 é $\frac{1}{\|\mathbf{W}\|}$, onde $\|\mathbf{W}\|$ é a norma euclidiana de $\mathbf{W}$, isto é, $\sqrt{\mathbf{W} \cdot \mathbf{W}}$. Por definição, isso é igual à distância de qualquer ponto em H_2 ao hiperplano de separação. Portanto, a margem máxima é $\frac{2}{\|\mathbf{W}\|}$.

2.6.3 FLORESTA ALEATÓRIA - RF

O RF (*Random Forest* - Floresta aleatória) é um método de classificação do tipo *ensemble*. Um método de classificação *ensemble* combina uma série de r modelos (ou classificadores de base), $M_1, M_2, \dots, M_r$, com o objetivo de criar um modelo melhorado de classificação composta, M^* . Um dado conjunto de dados, D , é usado para criar s conjuntos de treinamento, $D_1, D_2, \dots, D_s$, onde D_i ($1 \leq i \leq s - 1$) é usado para gerar o classificador M_i . Dado uma nova tupla de dados para classificar, os classificadores base votam retornando uma predição de classe. O conjunto retorna uma previsão de classe com base nos votos dos classificadores de base. Métodos *ensemble* tendem a ser mais precisos do que seus classificadores individuais. Considerando um método *ensemble* que realiza a votação majoritária, os classificadores base podem cometer erros, mas o conjunto irá classificar erroneamente a instância/registo apenas se mais da metade dos classificadores base estiverem errados (HAN; PEI; KAMBER, 2011).

Como cada um dos classificadores no conjunto é um classificador de árvore de decisão, para que a coleção de classificadores seja uma floresta, as árvores de decisão individuais são geradas usando uma seleção aleatória de atributos em cada nó para determinar a divisão. Mais formalmente, cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta. Durante a classificação, cada árvore vota e a classe mais popular é retornada (TAN; STEINBACH; KUMAR, 2005; HAN; PEI; KAMBER, 2011).

O procedimento geral para gerar t árvores de decisão para o conjunto é o seguinte: para

cada iteração i ($i = 1, 2, \dots, t$), um conjunto de treinamento, D_i , de d tuplas é amostrado com substituição de D . Ou seja, cada D_i é uma amostra *bootstrap*³ de D , para que algumas tuplas possam ocorrer mais de uma vez em D_i , enquanto outras podem ser excluídas. Seja F o número de atributos a serem usados para determinar a divisão em cada nó, onde F é muito menor que o número de atributos disponíveis. Para construir um classificador de árvore de decisão M_i , é selecionado aleatoriamente, em cada nó, os F atributos como candidatos para a divisão no nó. A metodologia CART (*Classification And Regression Tree* - Árvore de Classificação e Regressão) é usada para cultivar as árvores. As árvores são cultivadas até o tamanho máximo e não são podadas (TAN; STEINBACH; KUMAR, 2005; HAN; PEI; KAMBER, 2011).

O erro de generalização de uma floresta converge, desde que o número de árvores na floresta seja grande. Assim, *overfitting* não é um problema. A precisão de uma floresta aleatória depende da força dos classificadores individuais e uma medida da dependência entre eles. Como as florestas aleatórias consideram muitos menos atributos para cada divisão, elas são eficientes em bancos de dados muito grandes (HAN; PEI; KAMBER, 2011).

2.6.4 REDES NEURAIS ARTIFICIAIS - RNA

O estudo das RNAs foi inspirada na tentativa de simular os sistemas neurais biológicos. O cérebro humano consiste basicamente de células nervosas denominadas neurônios, que são ligados com outros neurônios via fios de fibras denominados axônios. Os axônios por sua vez são usados para transmitir impulsos nervosos de um neurônio para outro sempre que os neurônios são estimulados. Um neurônio é interligado em axônios de outros neurônios via dendritos, que são extensões do corpo da célula do neurônio. O ponto de contato entre um dendrito e um axônio é chamado de sinapse. Neurologistas descobriram que o cérebro humano aprende mudando a força da conexão sináptica entre os neurônios em estimulação repetida pelo mesmo impulso.

Durante a fase de treinamento/aprendizagem de uma RNA, os valores de peso para as unidades de processamento individuais (neurônios) na rede neural são configurados de forma a garantir a classificação correta das amostras das classes olfativas conhecidas (CAPELLI; SIRONI; ROSSO, 2014)

Análogo a estrutura do cérebro humano, uma RNA é composta por um conjunto interconectado de nodos e ligações diretas. A seguir é mostrado um resumo das principais características de uma RNA:

1. Redes neurais artificiais com pelo menos uma camada escondida são aproximações universais, ou seja, elas podem ser usadas para aproximar qualquer função alvo. Desde que a RNA tem um espaço de hipótese bastante expressivo, é importante selecionar uma

³O método de bootstrap testa as tuplas de treinamento dadas uniformemente com a substituição. Ou seja, cada vez que uma tupla é selecionada, é igualmente provável que seja selecionada novamente e adicionada novamente ao conjunto de treinamento.

topologia de rede adequada para um dado problema a fim de evitar o *overfitting* do modelo.

2. RNAs podem gerenciar características redundantes pois os pesos são automaticamente aprendidos durante a etapa de treinamento. Os pesos para características redundantes tendem a ser muito pequenos.
3. Redes neurais são muito sensíveis a presença de ruídos nos dados de treinamento. Uma abordagem para gerenciar esses ruídos é usar o conjunto de validação para determinar o erro de generalização do modelo. Outra abordagem é decrementar o peso por algum fator em cada iteração.
4. O método descendente gradiente usado para a aprendizagem dos pesos de uma RNA geralmente converge para um mínimo local. Um caminho para escapar do mínimo local é adicionar um termo momentâneo para a fórmula de atualização de peso.
5. Treinar uma RNA é um processo que demanda um tempo considerável, especialmente quando o número de nodos nas camadas escondidas é elevado. Porém, as amostras de teste podem ser classificadas rapidamente.

2.6.4.1 *EXTREME LEARNING MACHINE - ELM*

A ELM (*Extreme Learning Machine* - Máquina de Aprendizagem Extrema) proposta por Huang, Zhu e Siew (2004) é uma rede neural *feedforward*⁴ de camada oculta única e nós ocultos aleatórios, cuja velocidade de aprendizado pode ser milhares de vezes mais rápida do que algoritmos tradicionais de aprendizagem de rede como o algoritmo de *backpropagation*. Diferentemente dos algoritmos de aprendizado tradicionais, a ELM não apenas tende a alcançar o menor erro de treinamento, mas também a menor norma de pesos (regularização).

A teoria sobre o desempenho de generalização de redes neurais *feedforward* indica que estas redes alcançam menor erro de treinamento, quanto menor é a norma de pesos, e portanto melhor o desempenho de generalização estas redes tendem a ter (HUANG; ZHU; SIEW, 2006).

Segundo Huang, Zhu e Siew (2006), a ELM possui várias características interessantes e significativas, diferentemente dos algoritmos tradicionais de aprendizado baseado em gradiente para redes neurais *feedforward*:

1. A velocidade de aprendizagem da ELM é extremamente rápida. A fase de aprendizado da ELM pode ser concluída em segundos ou até menos para muitas aplicações. Usualmente, parece que existe uma barreira de velocidade virtual que a maioria algoritmos clássicos de aprendizado não conseguem ultrapassar e não é incomum levar muito tempo para treinar

⁴Em uma rede *feedforward*, cada camada da rede neural se conecta à próxima camada, porém não há caminho de volta. Portanto, todas as conexões têm a mesma direção, partindo da camada de entrada rumo a camada de saída (TAN; STEINBACH; KUMAR, 2005).

uma rede *feedforward* usando algoritmos de aprendizado clássicos, mesmo para aplicações simples.

2. A ELM apresenta melhor desempenho de generalização do que o aprendizado baseado em gradiente, como *backpropagation* na maioria dos casos.
3. Os algoritmos tradicionais de aprendizado baseados em gradiente clássico podem enfrentar vários problemas, como mínimos locais, taxa de aprendizado inadequada e *overfitting*. Para evitar esses problemas, alguns métodos como redução de peso e método de parada precoce podem ser usados com frequência. A ELM tende a alcançar as soluções sem problemas triviais.

2.7 AVALIAÇÃO DE CLASSIFICADORES

O erro estimado ajuda o algoritmo de aprendizado a fazer a seleção do modelo; isto é, encontrar um modelo da complexidade correta que não seja suscetível a *overfitting*. Uma vez que o modelo tenha sido construído, ele pode ser aplicado ao conjunto de testes para prever as classes de registros anteriormente não vistos. Geralmente, é útil medir o desempenho do modelo no conjunto de testes, pois esta medida fornece uma estimativa imparcial de seu erro de generalização. A precisão ou a taxa de erro calculada a partir do conjunto de testes também pode ser usada para comparar o desempenho relativo de diferentes classificadores no mesmo domínio. No entanto, para fazer isso, as classes das tuplas de teste devem ser conhecidas. Portanto, tem que ser aplicados métodos para avaliar o desempenho dos classificadores (TAN; STEINBACH; KUMAR, 2005).

2.7.1 VALIDAÇÃO CRUZADA *K-FOLD*

Um dos métodos mais aplicados na literatura é a da validação cruzada *k-fold*. Na validação cruzada *k-fold*, os dados iniciais são divididos aleatoriamente em k subconjuntos mutuamente exclusivos chamados de dobras, $D_1, D_2, \dots, D_k$, cada um com tamanho aproximadamente igual. O treinamento e o teste são realizados k vezes. Na iteração i , a partição D_i é reservada como o conjunto de testes e as partições restantes são usadas para treinar o modelo. Portanto, na primeira iteração, os subconjuntos $D_2, \dots, D_k$ servem como o conjunto de treinamento para obter um primeiro modelo, que é testado em D_1 ; a segunda iteração é treinada nos subconjuntos $D_1, D_3, \dots, D_k$ e testada no D_2 ; e assim por diante. Cada amostra é usada o mesmo número de vezes para treinamento e uma única vez para teste. Para classificação, a estimativa de precisão é o número geral de classificações corretas das k iterações, dividido pelo número total de tuplas nos dados iniciais. Usualmente, o k escolhido é 5 ou 10. O caso especial, quando $k = u$, onde u é o tamanho do conjunto de dados, é chamado de validação cruzada *leave-one-out*, onde o conjunto de testes compreende um único ponto e os dados restantes são usados para fins de treinamento (TAN; STEINBACH; KUMAR, 2005; HAN; PEI; KAMBER, 2011; ZAKI; JR; MEIRA, 2014).

3 ESTADO DA ARTE

O primeiro trabalho relacionado ao desenvolvimento de um instrumento para a medição do odores provavelmente deve-se a Moncrieff (1961). Esse instrumento foi uma espécie de Nariz Mecânico. De fato, o primeiros Narizes Eletrônicos foram descritos por Wilkens e Hartman (1964), através de reações *redox* de substâncias odoríferas em um eletrodo; por Buck, Allen e Dalton (1965) através de modulação de condutividade por odorantes; e também por Dravnieks e Trotter (1965) com modulação do potencial de contato por odorantes.

Mas a definição do conceito de Nariz Eletrônico, como um sistema de sensoriamento de matriz química inteligente para a classificação de odores demorou ainda cerca de 20 anos para surgir, após os trabalhos realizados por Persaud e Dodd (1982) na Inglaterra, e também por Ikegami e Kaneyasu (1985) e Kaneyasu et al. (1987) no Japão. Persaud e Dodd (1982) criaram uma ferramenta capaz de simular o sistema olfativo dos mamíferos, reconhecendo diferentes odores e dando respostas repetitivas. O Nariz Eletrônico que eles desenvolveram incluiu uma matriz de sensores, para simular os receptores do sistema olfativo humano; uma unidade de processamento de dados, que executava a mesma função que o bulbo olfativo, e um sistema de reconhecimento de padrões, que reconhecia os padrões olfativos da substância a ser testada.

O Termo Nariz Eletrônico (*Electronic Nose*) apareceu pela primeira vez no final da década de 80, quando foi criado por Gardner et al. (1988). A primeira conferência dedicada ao tema dos narizes eletrônicos foi realizada em 1990. Os primeiros narizes eletrônicos comerciais foram introduzidos em 1993 pela Alpha MOS¹, para testar solventes residuais em papel. Mais tarde foi definido como "*Um instrumento que compreende uma série de sensores químicos eletrônicos com especificidade parcial e um sistema de reconhecimento de padrões apropriado, capaz de reconhecer odores simples ou complexos*"(GARDNER; BARTLETT, 1994).

Desde então, o interesse em sensores desse tipo tem crescido, e é evidente o surgimento de trabalhos e artigos voltados a este tema, tanto para o desenvolvimento, quanto para o aprimoramento de equipamentos nos diversos setores da indústria. Atualmente, a atenção dos pesquisadores é mais focado em reduzir o tempo de análise dos narizes eletrônicos, aumentar a sensibilidade da configuração de medida, reduzir do tamanho do dispositivo e validar os métodos empregados (WOJNOWSKI et al., 2017a).

3.1 INDUSTRIAL/URBANO

Nesta seção será descrito e discutido alguns dos trabalhos presentes na literatura no âmbito industrial e urbano.

¹<https://www.alpha-mos.com>

Abraham e Pandian (2013) propuseram um sistema de monitoramento ambiental urbano móvel de baixo custo utilizando plataforma aberta para o sistema de *software* e *hardware*. Esse sistema móvel pode ser colocado em veículos de transporte público. O sistema permite construir um mapa local e regional contendo os pontos de maior concentração de poluentes, com os dados obtidos pelos sensores, e assim disponibilizá-los através da plataforma *online* desenvolvida. Sensores de gases (CO, CO₂ e CH₄), GPS, sensor de temperatura e Arduino compõem o *hardware* do sistema. Para o *software* foi construída uma aplicação *WebServer* em PHP com Banco de Dados MySQL. Como resultado é ilustrado o mapa de poluição gerado pela aplicação, com base nos valores analógicos dos sensores lidos pelo Arduino, e destaca-se que foi gasto apenas 200 dólares para a construção do protótipo, mas em contrapartida não foi realizado nenhum pré-processamento, validação ou avaliação estatística dos dados obtidos pelos dispositivos. O *software* também não é capaz de reconhecer padrões e nem classificar os dados, não se tratando portanto de um dispositivo de nariz eletrônico.

Al-Dahoud et al. (2013) propuseram uma rede de sensores e uma interface de simulação para redes de sensores sem fio para o monitoramento da qualidade do ar em cidades metropolitanas. Os nodos da rede são equipados com os sensores: MQ-2, que é sensível ao GLP, Iso-butano, Propano, Metano, Álcool, Hidrogênio e fumaça; temperatura e umidade (DHT11); e poeira. Os dados são transmitidos através de módulos *XBee*, utilizando protocolo *ZigBee*. A plataforma Arduino Uno foi utilizado para monitoramento da qualidade do ar na cidade de Annaba, no leste da Argélia. Foram comparados os gases detectados de três regiões diferentes da cidade com os níveis normais de gases (ar limpo), e os resultados obtidos dos testes indicam que há uma grande diferença nos níveis de gases GLP, NO₂ e CO. No entanto, os resultados adquiridos em algumas áreas da cidade de Annaba não mostram nenhuma situação de risco a ser considerada. Neste trabalho, foi realizado o monitoramento de qualidade do ar em campo, o projeto dos nodos eletrônicos e a transmissão sem fio para a fusão de dados. Uma interface para o usuário foi projetada para simulação da rede de sensores sem fio no controle da qualidade do ar do ambiente. O sistema demonstra-se ser flexível a presença de componentes gasosos e a facilidade do projeto facilita a implementação desse sistema. Os sensores foram calibrados para a detecção de gases utilizando outros como base. O sistema falha em não utilizar métodos estatísticos para a validação dos dados/resultados, nem algoritmos para classificação das informações e também só usa um tipo de sensor de gás por nodo de rede, que pode deixar o sistema suscetível a falhas.

Husain et al. (2015) apresentaram um dispositivo baseado no Arduino, de baixo custo e de fácil manuseio a fim de monitorar a qualidade do ar, através da coleta de dados sobre a quantidade de gases e poeira presentes no ar. Esse dispositivo pode ser colocado em qualquer lugar e os dados podem ser transferidos via *Bluetooth* através do protocolo UART (*Universal Asynchronous Receiver-Transmitter* - Receptor-transmissor Universal Assíncrono) para um *smartphone* ou simplesmente conectando o dispositivo a um computador. O *hardware* do dispositivo é composto de: um sensor de poeira óptico GP2Y1010AU0F, com saída do sensor de tensão analógica

proporcional à densidade de poeira medida, com uma sensibilidade de $0,5V/0,1mg/m^3$; dois sensores de gases (MQ-9 e MQ-135); além de um módulo *Bluetooth* e um Arduino Mega 2560. Segundo seus autores, esse equipamento permite que os dados de diferentes lugares possam ser posteriormente examinados para a fim de tomada de decisões e análise, ou seja, o dispositivo só é capaz de mostrar os dados brutos obtidos por cada um dos sensores, não havendo nenhum método computacional ou estatístico para determinar precisamente a qualidade do ar. A principal vantagem em comparação a outros trabalhos correlatos é o custo reduzido para o desenvolvimento do equipamento. Entretanto, os dados obtidos não foram validados e nenhum algoritmo para classificação/processamento foi usado, assim não se pode garantir que esse dispositivo poderá obter uma resposta confiável em potenciais aplicações.

3.2 AGRICULTURA

Nesta seção, com maior enfoque, serão descritos e discutidos alguns dos trabalhos presentes na literatura, que utilizam de sensores de gás no âmbito agrícola.

3.2.1 SOLO

Siyang et al. (2013) usaram um nariz eletrônico para detecção de compostos voláteis no solo e pragas (traça de cera). Buscou-se relacionar as impressões digitais do solo conforme os níveis de profundidade e a quantidade de pragas em tempo real e a céu aberto. O solo foi amostrado nos níveis de 10 cm, 30 cm e 50 cm de profundidade. Foram utilizados os sensores TGS 821, TGS 822, TGS-825, TGS 826, TGS 2600, TGS 2602, TGS 2610 e TGS 2620 para a detecção dos gases oriundos dos solos. As respostas dos sensores foram usadas para realizar análise de padrões utilizando PCA. Os sinais de saída que correspondem aos COVs do solo foram classificados em três grupos (dimensões). As variações nos componentes principais do primeiro ao terceiro grupo, para todas as impressões voláteis do solo, devido as profundidades do solo, foram contabilizadas como 77,6%, 17,6% e 3,5%, respectivamente. Os resultados mostram que o nariz eletrônico detectou com sucesso as diferentes impressões digitais voláteis do solo em cada nível de profundidade e de pragas. Os resultados também mostram que um aumento de gases oxidantes é maior que a redução de gases em determinadas profundidades. Por fim, nariz eletrônico também foi treinado para diferenciar entre as quantidades traça de cera, variando nas quantidades de 0, 2, 4 e 6. O resultado do PCA é mostrado em dois principais grupos, com variância de impressões digitais de 93,8% e 3,6%. Os pontos PCA de 0, 2 e 4 insetos ocorrem na mesma região, enquanto os pontos PCA de 6 insetos se separam claramente uns dos outros. Pode implicar que o dispositivo pode detectar insetos de cera quando o número de mariposas é 6. Embora esse número seja muito alto para o mundo real, o nariz eletrônico pode mostrar a possibilidade de detecção de pragas de insetos por odor.

Sudarmaji e Kitagawa (2016) desenvolveram e testaram uma técnica de modulação de temperatura com um ponto de detecção especificado. Essa técnica pode superar o problema de

seletividade cruzada nos sensores de gás MOS (*Metal-Oxide-Semiconductor* - Semicondutor de óxido metálico) e também aumentar sua sensibilidade. Foram usados os sensores de gases TGS 2444, TGS 2602, TGS 825, FIS12A, FIS30SB e FISAQ1 a fim de detectar os compostos gasosos de solos, em uma mistura complexa. Um nariz eletrônico foi construído para capturar e analisar o perfil gasoso do espaço de amostragem de dois solos (arenoso e areia) com a adição de nutrientes em diferentes doses (sem adição, normal e alta). Os sensores foram conduzidos em uma determinada modulação para uma interface baseada no microcontrolador PSoC CY8C28445-24PVXI. Foram utilizadas as ferramentas de PCA e uma RNA a fim de reconhecimento de padrões. Os compostos gasosos são deixados em uma câmara de gás com termostato e agitação sob condições controladas para otimizar o equilíbrio e a concentração desses gases. Os resultados obtidos demonstraram que o sensor que obteve o maior incremento de sensibilidade foi o TGS 825, com mais do dobro de ganho para solos arenosos com adição normal de nutrientes. Os outros sensores obtiveram ganho de sensibilidade, mas muito abaixo do TGS 825. É mostrado também que a PCA de discriminação de dois solos, ambos sem adição de composto, acumulativamente representam 78,32% da variância dentro do conjunto de dados, sendo um valor aceitável. Finalmente, foi determinado o desempenho de RNA como unidade de decisão do nariz eletrônico para classificar o nível de adição de nutrientes no solo com base no indicador do erro médio quadrado alcançado resultante do processo de treinamento. Os dados mostram que o maior erro quadrado médio alcançado foi aquele independente do tipo do solo e sem utilização de PCA ($4,080 \times 10^{-3}$). Os resultados do PCA indicam que o conjunto de sensores utilizados são capazes de diferenciar o tipo de solo de forma clara e pode fornecer uma discriminação como resposta ao nível de adição de nutrientes no solo. Além disso, o PCA aumenta o desempenho de classificação, de adições de nutrientes pré-determinados, da rede neural. Esse trabalho prova que é possível, utilizando um nariz eletrônico, classificar tipos de solos e, apesar de terem sido classificados apenas dois tipos, é possível desenvolver um equipamento capaz de detectar diversos tipos de solos a ser aplicado a diferentes regiões do planeta.

Dorji, Pobkrut e Kerdcharoen (2017) desenvolveram um dispositivo de nariz eletrônico baseado em uma rede de sensores sem fio (usando protocolo *ZigBee*) a fim de resolver o problema de análise de nutrientes do solo, que corriqueiramente exigem uma análise laboratorial que pode tornar-se cara e demorada. Baseado na análise de COVs, foram instaladas estações de narizes eletrônicos com os sensores de gás TGS 821, TGS 825 (2x), MQ 3, MQ 5, MQ 6, MQ 135, MQ 137, para análise do solo de uma fazenda, a fim de gerar dados *online* em tempo real e realizar o monitoramento rápido e preciso. Um nariz eletrônico portátil, composto por uma matriz de oito sensores de gás, sendo eles MQ 8, TGS 2444, TGS 823, TGS 2600, TGS 2603, TGS 2610, TGS 825 e TGS 2620 foi também empregado para a determinação dos COV, com o solo anteriormente tratado com fertilizantes orgânicos e com fertilizantes compostos por NPK (Nitrogênio, Fósforo e Potássio). A utilização do PCA foi empregada com sucesso na classificação de COVs, indicando diferenças significativas de fertilidade para os solos baseados em matéria orgânica. Os resultados obtidos pelos narizes eletrônicos dos COVs não diferem dos resultados obtidos em laboratório, o

que demonstra o potencial desses equipamentos para esses fins. Quatro diferentes amostras do solo foram classificadas. A variação do primeiro, segundo e terceiro componentes principais é de 86,2%, 12,4% e 1,4%, respectivamente.

Pineda e Pérez (2017) expõem que os solos tropicais são os maiores emissores de CO₂. Com base nisso foi construído um nariz eletrônico com custo de 50 dólares para o monitoramento de gases lançados na atmosfera por alguns tipos de solos. O sistema é composto por uma matriz de três sensores de gás (MQ-4, MQ-8 e MQ-135), um microcontrolador PIC16F886 e um *software* construído em Java para a aquisição e visualização dos dados em um computador. Aos sensores do sistema foi incorporado uma multiplexação modificada para permitir uma análise em paralelo de até oito amostras. Foram realizados testes demonstrando que esses sensores obtiveram resposta proporcional à concentração de gás (conforme injeção de ar na câmara de gás), com o MQ-135 obtendo uma correlação de 0,993 para o CO₂ e o MQ-4 com uma correlação de 0,9841 para o CH₄. Primeiro foi testado o conjunto dos três sensores de gás em amostras de solos da Colômbia e do Equador, para o estudo dos padrões de emissão de gases, em testes de respiração induzida por glicose D(+), como fonte de carbono orgânico. Os resultados (em gráficos) demonstraram que haviam diferenças dos padrões de gases emitidos por essas duas amostras e um aumento diferencial para o tempo de produção de gás, após o aumento de glicose D(+), pois, conforme havia o aumento da dose de glicose, a resposta analógica do sensor também aumentava e, se permanecesse constante, a resposta do sensor também se mantinha no padrão. No segundo experimento, usando um conjunto de sensores multiplexados em um conjunto diferente de amostras de solo colombianas, os padrões de emissão de gases foram obtidos a partir de solo autoclavado com atividade microbiana reduzida, que foram contaminados com óleo diesel como fonte de carbono. Os resultados (em gráficos) do segundo estudo mostraram um padrão claramente diferente de emissões de gases em amostras contaminadas com óleo diesel, quando a atividade microbiana foi reduzida pelo autoclave no solo. Além disso, dois grupos de amostras de solo não autoclavadas, que diferem apenas na quantidade de óleo diesel, também puderam ser diferenciados, com base em seus padrões de respiração, após algumas horas de monitoramento dos gases. Assim como no primeiro experimento, os valores analógicos dos sensores aumentaram conforme o incremento da dose do composto. Segundo seus autores, com base nos resultados, pode-se prever um grande potencial com o uso desse sistema nas áreas de fertilidade, microbiologia do solo, biotecnologia e ciências ambientais, onde os processos como a disponibilidade de nutrientes das plantas, a atividade microbiana ou o volume de poluentes do solo são monitorados através de seus padrões de emissão de gases. Não foram aplicadas ferramentas para a classificação dos dados oriundos desses sensores.

3.2.2 FRUTICULTURA

Nesta subseção, serão descritos e discutidos alguns dos trabalhos presentes na literatura, que utilizam de sensores de gás focados na fruticultura.

3.2.2.1 BANANAS

Sanaeifar et al. (2014) aplicaram um nariz eletrônico para detectar a alteração de COVs da banana durante o processo de amadurecimento. Foram utilizados os sensores MQ-3, MQ-5, MQ-9, MQ-131, MQ-135 e MQ-136 para detecção dos odores. Os dados do nariz eletrônico foram pré-processados através das etapas de identificação, manipulação e determinação da linha de base; compressão; normalização; e, após isso, as técnicas PCA, LDA, SIMCA (*Independent Modelling of Class Analogy* - Modelagem Suave Independente de Analogias de Classe) e SVM analisaram esses dados pré-processados. Os resultados mostraram que o nariz eletrônico pode distinguir entre diferentes estágios de amadurecimento. Os resultados obtidos pelo PCA, categorizados pela maturação da banana, mostram que os dois primeiros componentes, PC1 e PC2, contêm 82% da variação de dados. O primeiro componente principal, PC1, explica 58% da variação total, enquanto 24% da variância total é explicada pelo PC2. Os resultados obtidos pela LDA, também categorizados pela maturação da banana, proporcionaram uma classificação, com 97,6% da variância total dos dados obtidos por LD1 e LD2 que representaram 85,7% e 11,9% da variância, respectivamente. A precisão da classificação obtida pelo método LDA com validação cruzada foi de 97,3%. O método SIMCA, obteve uma precisão de classificação de 92%, e a análise SVM teve uma taxa de sucesso alta na classificação de amadurecimento da banana, com precisões de treinamento e validação de 100 e 98,66%, respectivamente.

Chen et al. (2018) construíram um nariz eletrônico que tem potencial para monitorar a variação dos COVs produzidos pela banana durante o processo de maturação. Sensores MOS da família TGS foram utilizados para a construção do dispositivo, que segue: TGS2600, TGS2602, TGS2603, TGS2610, TGS2611, TGS2612 e TGS2620. Além do nariz eletrônico, foi incorporado um sistema de câmera para monitorar a cor da casca da fruta como outro recurso para identificação. O sistema duplo nariz/câmera apresenta melhor medida de separabilidade de classe nos 4 estágios de maturação da banana: imaturos, meio maduros, maduros e super maduros. O PCA e o LDA foram aplicados as respostas dos sensores do nariz eletrônico para redução dos dados, e verificação de eventuais agrupamentos dos diferentes estágios de maturação da fruta. O PCA apresentou variabilidade de 93,92% para os 2 componentes principais e o LDA 94,11% de variação. Os 2 métodos de redução de dimensionalidade foram utilizados como entrada para 2 algoritmos de classificação, KNN e SVM. As respostas dos sensores MOS e as respostas do sistema de câmera foram agregados e utilizados no métodos de classificação. Isso permitiu uma perfeita classificação para todos os classificadores independente da imputação (PCA ou LDA). Portanto, comparativamente a outros trabalhos, os dois sistemas agregados incrementam a precisão do sistema como um todo.

3.2.2.2 PÊSSEGOS

Brezmes et al. (2000) apresentaram o uso de um Nariz Eletrônico para monitoramento não destrutivo do processo de amadurecimento do pêssego. As amostras dos frutos foram

armazenadas a temperatura ambiente durante todo o processo de amadurecimento. O dispositivo utilizou-se dos seguintes sensores: TGS800, TGS822, TGS825, TGS826, TGS882, TGS2610, TGS2611 e TGS2620. Utilizou-se uma rede neural de três camadas (entrada, oculta e saída) para classificação entre os três estágios de maturação (verde, maduro e super maduro). Medidas feitas com pêssegos mostram uma taxa de sucesso de 93%. A análise da rede neural demonstrou que apenas os sensores TGS882 e TGS2610 são capazes de prever o estágio de maturação do pêssego, e portanto os únicos necessários para esse tipo de aplicação. Uma característica adicional do sistema é a capacidade de prever com precisão o número de dias que a fruta está armazenada desde a colheita. Medidas feitas com pêssegos apresentaram um erro máximo de 1 dia.

Benedetti et al. (2008) utilizou um nariz eletrônico comercial (PEN2) para classificar quatro cultivares de pêssego e avaliar o estágio de maturação pós-colheita. PCA e LDA foram usados para verificar se o nariz eletrônico foi capaz de distinguir entre quatro cultivares diferentes. PCA aplicado nos dados do nariz eletrônico destacou a capacidade do instrumento em avaliar o estágio de maturação dos pêssegos. Os 2 componentes principais do PCA aplicado para cada um dos cultivares separadamente, mostram que o método foi capaz de obter uma variabilidade de 99,99% de todo conjunto de dados. Ademais, o PCA também foi capaz de agrupar claramente os 4 cultivares, aplicado para o todo conjunto original, sem separação por cultivar, sendo que os 2 componentes principais representaram 99,43% da variância dos dados originais. As respostas dos sensores foram ajustadas com uma função de transferência sigmoide, permitindo a definição de três diferentes estágios de maturação (imaturos, maduros, super maduros). A análise CART foi aplicada para classificar as amostras de pêssego entre as três classes. A análise com árvore de decisão classificou amostras em cada grupo respectivo com uma taxa de erro de validação cruzada de 4,87% e de teste de 2,44%.

Guohua et al. (2012) aplicaram um nariz eletrônico para prever o frescor de pêssego. A análise se baseia no tempo de armazenamento dos pêssegos pós-colheita, ou seja, em classificar pelo número de dias que passaram após as amostras de pêssegos maduras serem colhidas. O dispositivo foi construído com uma matriz contendo oito sensores de gás MOS, sendo eles: MQ-2, MQ-3, MQ-6, MQ-7, MQ-8, TGS825, TGS842 e TGS2201. PCA e SR (*Stochastic Resonance* - Ressonância estocástica) são usados para análise de dados de medição. O PCA foi capaz de agrupar as amostras entre as idades (0 a 4 dias), com variabilidade nos 2 componentes principais de 94,92% do conjunto original de dados. Os resultados com o SR mostram que o *e-nose* pode distinguir pêssegos entre condições frescas e super maduro. A validação dos resultados das experiências demonstra que a precisão de previsão do modelo de SR é de 85%.

Zhang et al. (2012) desenvolveram um nariz eletrônico capaz de classificar amostras de pêssego em três tipos: verdes, meio maduros e maduros, além de descrever os três índices qualidade para esses tipos: força penetrante (CF), teor de açúcar (SC) e acidez (pH). O protótipo utiliza-se dos seguintes sensores de gás MOS - MQ-3, MQ-7, TGS800, TGS813, TGS822, TGS824, TGS825 e TGS880. Os modelos PCR (*Principal Component Regression* - Regres-

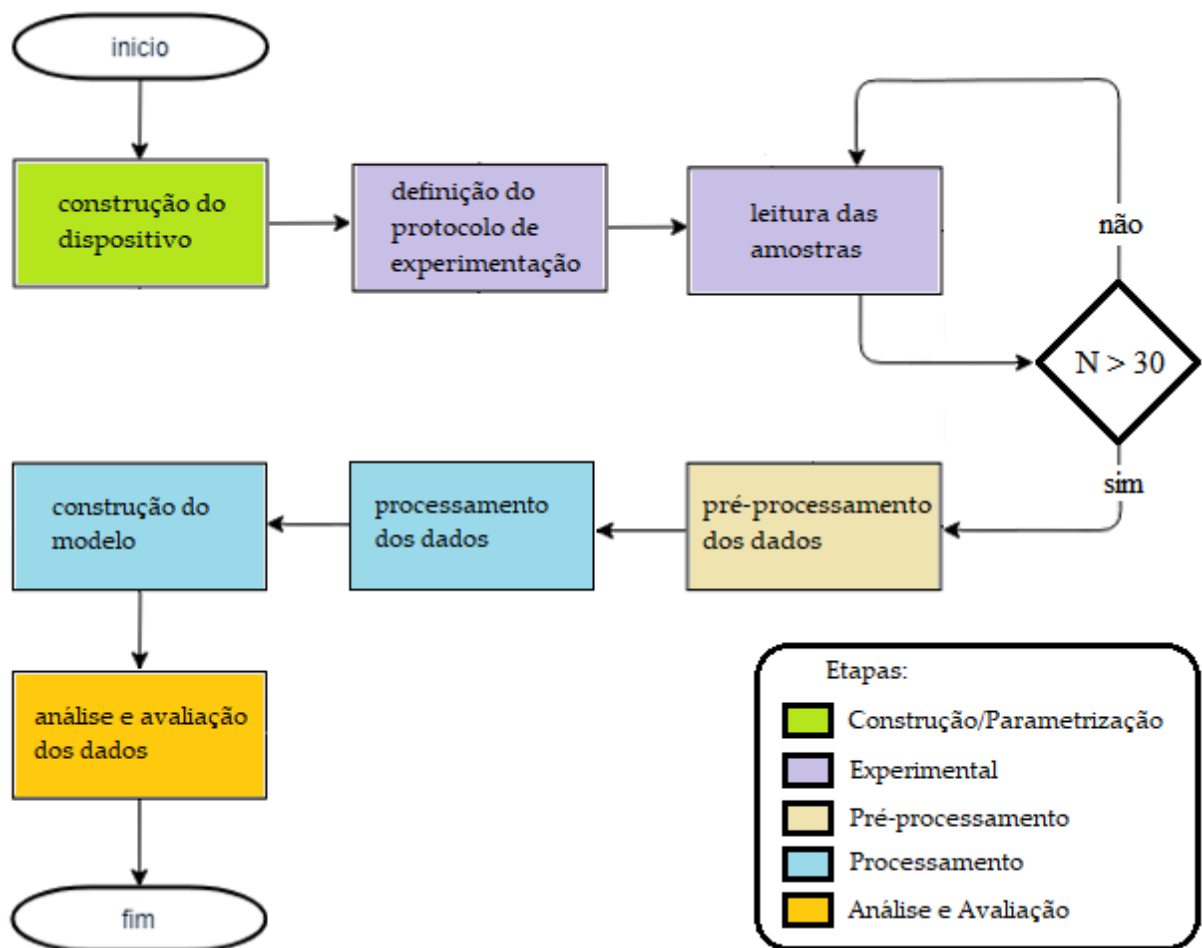
ção de Componentes Principais - Regressão de Componentes Principais) e PLS (*Partial Least Squares Interpolation* - Interpolação Parcial de Mínimos Quadrados) representaram uma boa capacidade para descrever os índices de qualidade dos três conjuntos de pêssegos selecionados. Os resultados mostraram que o PCA e o LDA, em duas dimensões, foram capazes de agrupar os dados entre os três estágios de maturação das amostras. Os 2 componentes principais do PCA representaram 98,38% da variabilidade dos dados, para as 3 classes de maturação (verdes, meio maduros e maduros), enquanto os 2 componentes do LDA foram capazes de representar 99% da variabilidade do conjunto de dados originais, para as classes de maturação. Além do mais, o modelo PLS foi capaz de representar boa capacidade em prever índices de qualidade, com $R=0,86$ e SEP (*Standard Error of Prediction* - Erro padrão de previsão)=8,77 para CF; $R=0,83$ e SEP=0,297 para SC; $R=0,83$ e SEP=0,2 para pH. O modelo de PCR apresentou $R=0,84$ e SEP=7,33; $R=0,82$ e SEP=0,44; $R=0,78$ e SEP=0,21; para CF, SC e pH, respectivamente.

Como pode-se observar, as aplicações tendem a realizar a medida do fruto pós-colheita e não a avaliação do fruto durante o seu crescimento, no pomar. Portanto, o trabalho busca construir um nariz eletrônico que monitora todo o período desde a pós-floração até a pós-colheita, com o monitoramento a campo. As metodologias de redução de dados (PCA e LDA) e classificação (KNN, SVM, RF e RNA), foram escolhidas com base nos trabalhos correlatos feitos por Brezmes et al. (2000), Benedetti et al. (2008), Guohua et al. (2012), Zhang et al. (2012).

4 MATERIAIS E MÉTODOS

A metodologia executada neste trabalho se divide basicamente em cinco etapas: construção/parametrização do equipamento (seção 4.1), estabelecimento do protocolo experimental (seção 4.2), pré-processamento (seção 4.3), processamento (seção 4.4) e análise e avaliação dos dados (seção 4.5). A Figura 5 ilustra o fluxograma referente a metodologia utilizada, que será descrita com mais detalhes nas seções subsequentes, no qual N é o número de amostras para cada classe/estágio de maturação.

Figura 5 – Fluxograma ilustrativo da metodologia adotada



4.1 CONSTRUÇÃO/PARAMETRIZAÇÃO DO EQUIPAMENTO

Esta etapa pode ser subdividida em seleção dos sensores a serem utilizados na matriz de sensores e construção do *hardware* do dispositivo.

Foram utilizados ao todo 13 sensores de gás feitos de semicondutores de óxido metálico (MOS) no dispositivo de nariz eletrônico construído, sendo 9 sensores da família MQ fabricados

pela Hanwei Eletronics Co., Ltd: MQ-2, MQ-3, MQ-4, MQ-5, MQ-6, MQ-7, MQ-8, MQ-9 e MQ-135 e mais 4 sensores fabricados pela TGS Figaro, sendo eles: TGS 822, TGS 2600, TGS2602 e TGS 2603. As reações químicas entre as moléculas da superfície dos sensores e os gases fornecem uma mudança na resposta do sensor conforme a alteração na concentração de gás no ambiente. Mais especificamente, a resistividade dos sensores aumenta na presença de ar e diminui na presença dos gases sensíveis. A Tabela 4 mostra a faixa de detecção de partículas, conforme os *datasheets* dos fabricantes Figaro® e Hanwei®, e os preços dos principais sensores MQ e TGS. Mais detalhes de caracterização dos sensores MOS utilizados neste estudo, podem ser vistas no apêndice A.

Tabela 4 – Faixa de detecção de partículas e preços dos sensores MQ e TGS utilizados no estudo

Sensor	Escopo de Detecção de Concentração	Custo*
MQ-2	200-10000ppm GLP, CH ₄ , H ₂ , CO, Álcool, fumaça e Propano	R\$ 16,90 ^a
MQ-3	0,1-10mg/L GLP, CH ₄ , CO, Álcool, Benzeno e Hexano	R\$ 22,90 ^a
MQ-4	200-10000ppm GLP, CH ₄ , H ₂ , CO, Álcool e fumaça	R\$ 16,90 ^a
MQ-5	200-10000ppm GLP, CH ₄ , H ₂ , CO e Álcool	R\$ 19,90 ^a
MQ-6	200-10000ppm GLP, CH ₄ , H ₂ , CO e Álcool	R\$ 24,90 ^a
MQ-7	50-4000ppm GLP, CH ₄ , H ₂ , CO e Álcool	R\$ 19,90 ^a
MQ-8	200-10000ppm GLP, CH ₄ , H ₂ , CO e Álcool	R\$ 19,90 ^a
MQ-9	200-1000ppm CO 200-10000ppm GLP e CH ₄	R\$ 22,90 ^a
MQ-135	10-200ppm GLP, CO, CO ₂ , NH ₄ e Álcool	R\$ 19,90 ^b
TGS 822	50-5000ppm Metano, etanol, iso-butano, CO, acetona, benzeno e n-hexano	R\$ 16,00 ^c
TGS 2600	1-100ppm H ₂ , metano, etanol, iso-butano e CO	R\$ 21,00 ^c
TGS 2602	0,1-3ppm H ₂ S 1-30ppm H ₂ , NH ₃ , etanol e tolueno	R\$ 8,00 ^c
TGS 2603	0,1-3ppm Trimetilamina 0,3-3ppm H ₂ S e metil mercaptano 1-30ppm Etanol e H ₂	R\$ 71,00 ^d

*Preços obtidos no dia 16 de dezembro de 2018

^a<https://www.filipeflop.com>

^b<https://www.usinainfo.com.br>

^c<https://www.seekic.com>

^d<https://www.aliexpress.com>

Essas famílias de sensores de gás (MQ e TGS) foram escolhidas por serem as mais utilizadas para a construção de narizes eletrônicos, devido ao menor custo, confiabilidade e a fácil aquisição em comparação as demais. Foram selecionadas uma grande quantidade de sensores para ser possível a criação de uma impressão digital de cada amostra, uma vez que cada sensor tem uma resposta diferente para uma determinada amostra e o conjunto das respostas gera uma saída quase única para cada tipo de gás, aliado ao fato de que cada sensor trabalha com um espectro diferente de gases em comparação aos demais. Isso para obter todas as possibilidades, identificar os sensores mais importantes e, se necessário, retirar os sensores redundantes. Esses sensores foram selecionados com base na análise do escopo de detecção de concentração dos

principais gases e custo, conforme mostrado na Tabela 4, além da análise da literatura que mostrou que estes sensores e famílias são os mais utilizados em detrimento dos demais.

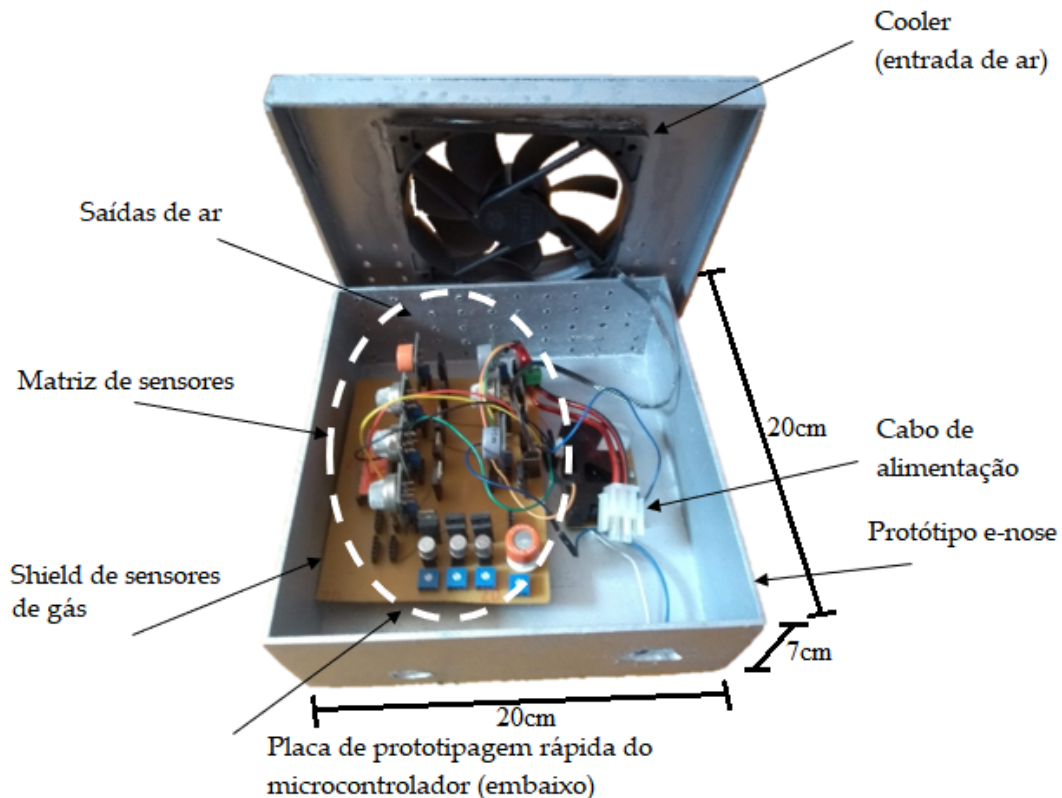
No que tange a construção do *hardware* do sistema foi utilizado o microcontrolador Atmel ATMega 2560. Uma vez que microcontrolador não possui todas as entradas analógicas necessárias para a leitura dos sensores foi utilizado um multiplexador analógico Cd74hc4067 de 16 canais e uma saída para realização da leitura dos dados oriundos dos sensores. Dessa forma, foi usada apenas uma porta analógica do microcontrolador. Também foram usados no sistema o sensor de pressão atmosférica e temperatura BMP180 e o sensor de umidade relativa do ar HIH-4030, utilizados para aquisição dos dados necessários para o ajuste/compensação dos sensores de gás, conforme variação da temperatura e umidade relativa do ar ambiente.

Todo o sistema é alimentado com uma bateria automotiva. Essa bateria gera uma tensão de saída de 12V, e capacidade nominal de 40Ah. Conforme será descrito com mais detalhes, o circuito consome uma corrente total na ordem de 2A. Portanto, em condições de total carregamento, a bateria é capaz de alimentar o circuito durante 20 horas (40Ah/2A), trabalhando em temperatura ambiente (25°C). A Figura 6 ilustra o equipamento desenvolvido, sem a bateria, e a Figura 7 o diagrama de blocos do circuito.

Conforme pode ser observado na Figura 6, o equipamento possui um cooler que possui a função de circulação do ar ambiente, que possibilita o melhor contato da amostra a ser medida com os sensores. O ar é puxado para dentro do dispositivo, na matriz de sensores, e os furos laterais possibilitam o fluxo de saída de ar do dispositivo. Análises iniciais mostram que esse fluxo constante possibilita uma resposta mais suavizada por parte dos sensores de gás. Vale ressaltar que os sensores de gás possuem a necessidade de um período de pré-aquecimento (cerca de 600 segundos) até estes atingirem uma temperatura constante ideal de funcionamento. Portanto, a estabilização de resposta dos sensores será atingida após esse tempo e esses dados transitórios são descartados da etapa de análise das respostas. Foi criada uma *shield* a fim de facilitar o processo de prototipagem do dispositivo, no qual é possível remover rapidamente os sensores e a placa de prototipagem rápida do microcontrolador (Arduino). A matriz de sensores foi posicionada de maneira que todos os sensores de gás fiquem embaixo do cooler de entrada de ar. O cabo de alimentação permite a conexão por parte de uma fonte externa.

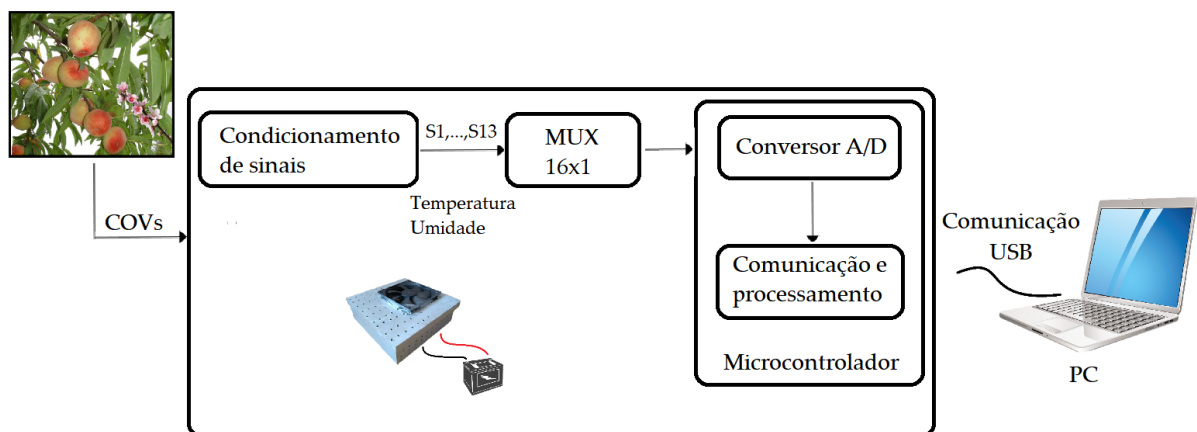
A sequência ilustrada no diagrama de blocos da Figura 7 é dada da seguinte forma: os sinais de respostas dos sensores são condicionados de tal maneira que permita a leitura por parte das entradas analógicas e digitais do microcontrolador. Para os sensores de gás, o circuito irá gerar uma resposta entre 0 e 5V, proporcional a concentração de COVs no ambiente. Uma vez que o número de entradas analógicas no microcontrolador é inferior ao número de sensores, foi necessário a utilização de um multiplexador 16x1, a fim de obtenção dos dados provenientes desses periféricos. As últimas etapas são efetuadas pelo microcontrolador, que é responsável por converter os sinais analógicos em digitais, processamento dos sinais obtidos pelos sensores e comunicação, a fim de transmissão dos dados através de comunicação serial, via USB, ao

Figura 6 – Protótipo do nariz eletrônico



computador.

Figura 7 – Diagrama de blocos do circuito do protótipo desenvolvido



O sensor de pressão atmosférica e temperatura BMP180 utiliza-se do protocolo I²C (*Inter-Integrated Circuit* - Circuito Inter-Integrado) para comunicação com o microcontrolador. A interface I²C utiliza-se de dois fios digitais denominados de SCL e SDA para realização da comunicação. A frequência de *clock* do protocolo supera a taxa de 3,4 Megabits/segundo. O SCL e o SDA precisam cada um deles de um resistor de *pull-up*, tipicamente na ordem de 4,7k Ω para V_{DD} . O barramento I²C é usado para controlar o sensor, ler os dados de calibração

do EEPROM (*Electrically-Erasable Programmable Read-Only Memory* - Memória de apenas leitura programável e apagável eletricamente) e ler os dados de medição quando a conversão analógico/digital estiver concluída. Por padrão, o microcontrolador utilizado no estudo possui suporte a comunicação I²C. A Tabela 5 mostra as principais características do sensor BMP180. Maiores detalhes de caracterização do sensor BMP180 podem ser vistas no apêndice A.

Tabela 5 – Principais características do sensor de temperatura e pressão atmosférica BMP180

Parâmetro	Condição	Mínimo	Típico	Máximo	Unidade
Tensão de operação		1,62	2,5	3,6	V
Corrente de alimentação	Baixo consumo de energia		3		mV
	Padrão		5		uA
	Resolução alta		7		uA
	Resolução ultra alta		12		uA
	Resolução avançada		32		uA

Para obter corretamente o valor de umidade relativa do ar do sensor HIH-4030, é necessário realizar a compensação do valor do resposta do sensor conforme a temperatura. Essa informação é disponibilizada pelo fabricante, e está descrita na equação 4.1:

$$RH = \frac{RH}{1,0546 - 0,0021 * T}, \text{ T em } ^\circ\text{C} \quad (4.1)$$

As principais características do sensor de umidade do ar HIH-4030 podem ser visualizadas na Tabela 6. Mais detalhes de caracterização do sensor HIH-4030 também podem ser vistas no apêndice A.

Tabela 6 – Principais características do sensor analógico de umidade do ar HIH-4030

Parâmetro	Mínimo	Típico	Máximo	Unidade
Tensão de operação	4		5,8	V
Corrente de alimentação		200	500	uA

Como descrito anteriormente, juntamente com os sensores, foi utilizado o microcontrolador Atmel ATmega2560, com a placa de prototipagem rápida Arduino Mega 2560, para aquisição dos dados. A placa de prototipagem auxilia no processo de montagem e desmontagem do protótipo, a fim de possíveis alterações de atualização do protótipo. As principais características do microcontrolador são mostradas na Tabela 7. Mais informações das características do microcontrolador Atmel ATmega2560 podem ser vistas no apêndice A.

Considerando o consumo máximo de corrente para todos os sensores utilizados no protótipo, que ocorre no período de pré-aquecimento dos sensores de gás, que é quando estes demandam uma maior quantidade de corrente, o valor total exigido pelo dispositivo fica próximo a 2A. Assim, foi necessário utilizar uma fonte de corrente contínua que suportasse pelo menos 3A, considerando uma eventual margem de erro. Ademais, para a construção da fonte é necessário

Tabela 7 – Principais características do microcontrolador Atmel ATmega2560

Parâmetro	Valor	Unidade
Tensão de operação	5	V
Entradas analógicas	16	
Corrente de saída por pino de I/O	40	mA
Velocidade de clock	16	MHz
Conversor A/D	10	bits

haver a redução da tensão para 5V, que é a tensão que alimenta os sensores MOS. Para isso, utilizou-se do regulador de tensão LM7805, para cada um dos sensores MOS, uma vez que este regulador possui corrente máxima de trabalho de 1A, e, portanto, apenas um deste componente em todo o circuito não suportaria devido a alta corrente demandada pelo circuito.

O diagrama esquemático do circuito é mostrado com detalhes no apêndice B. Pode-se perceber as 13 entradas para os sensores de gás, além dos 13 reguladores de tensão LM7805 (um para cada sensor de gás). Para os sensores de gás da família TGS foi necessário colocar em paralelo à entrada dos sensores um resistor de regulação de sensibilidade de 10 K Ω , conforme especificado pelo fabricante. O diagrama esquemático também mostra as entradas para os sensores de temperatura, pressão e umidade relativa do ar. Além do mais, tem-se a utilização do multiplexador, que reduz a quantidade de portas necessárias do microcontrolador, além de auxiliar na redução da corrente total utilizada pelo microcontrolador, uma vez que as 13 portas trabalhando em paralelo demandam mais energia que apenas uma. O circuito foi construído conforme disposto no apêndice B, pois o posicionamento dos componentes permitiu a diminuição da área de trabalho da placa, além de melhorar o espaço utilizado. O *layout* deste circuito que foi usado como base para confecção da placa de circuito impresso está ilustrado no apêndice C.

4.2 PROTOCOLO DE EXPERIMENTAÇÃO

Os experimentos foram conduzidos na cidade de Ponta Grossa no estado do Paraná, com clima *Cfb*¹, onde apresenta precipitação significativa em todos os meses do ano e temperaturas amenas nos meses mais quentes, situada na latitude de 25°05'49"S e longitude de 50°03'11"W.

As medidas compreenderam os três períodos de crescimento do fruto e mais um quarto período de pós-colheita do fruto. Considerando os períodos descritos anteriormente, uma vez que a florada ocorreu no fim de agosto/início de setembro, o estágio I compreendeu até o dia 05/10/2018. O estágio II ocorreu até o dia 26/10/2018. O estágio III compreendeu até a data que houve a colheita do fruto, no dia 17/11/2018. Após isso, até a data do último experimento (13/12/2018), tem-se o estágio IV. A Tabela 8 resume cada estágio de crescimento, além do período abrangido, em dias, após a plena floração, além das principais classes de COVs envolvidas nesse processo.

¹Clima temperado com verão ameno, segundo a classificação climática de Köppen-Geige

Tabela 8 – Estágios de maturação do pêssgo e seus respectivos dias e classes de gases

Maturação	Dias após plena floração	Principais classes de COVs
Estágio 1	0 - 35	Álcoois
Estágio 2	36 - 49	Áldeídos, Ésteres e Cetonas
Estágio 3	51 - 70	Terpenóides
Estágio 4	71 - 100	Lactonas

Os experimentos foram realizados nos dias 23/08/2018 (23°C de média), 06/09/2018 (21°C de média), 21/09/2018 (24°C de média) – Estágio I; 05/10/2018 (25°C de média), 11/10/2018 (17°C de média), 19/10/2018 (24°C de média) – Estágio II; 26/10/2018 (21°C de média), 08/11/2018 (19°C de média), 15/11/2018 (27°C de média) – Estágio III; 30/11/2018 (27°C de média), 07/12/2018 (24°C de média) e 13/12/2018 (32°C de média) – Estágio IV. Todos os experimentos foram realizados na mesma hora, aproximadamente se iniciando às 15:00 horas, a fim de se obter uma padronização no processo experimental.

A fim de validar a repetibilidade do equipamento, foi realizada uma avaliação em laboratório utilizando amostras predeterminadas contendo álcool. Um total de 20 amostras contendo diferentes proporções de água com álcool foram produzidas. Durante 7 dias essas amostras foram lidas pelo equipamento, e então foram analisadas as curvas de respostas (já compensadas pela temperatura e umidade relativa do ar). Os resultados mostraram que o equipamento foi capaz de gerar respostas equivalentes para as mesmas amostras, mesmo em períodos distintos com variações ambientais.

O equipamento foi posicionado na mesma posição para todos os experimentos, ao pé de um pessegueiro, no meio do pomar, como está ilustrado na Figura 8.

Cada dia envolveu a medida de 15 amostras durante 120 segundos. Como foram desconsiderados o tempo de pré-aquecimento e de estado transitório (cerca de 600 segundos), foram necessários 2400 segundos de leitura de dados. Como foram realizadas medidas a cada segundo, um total de 1800 registros foram obtidos para cada dia. Para cada amostra foi realizada a média aritmética das respostas de cada um dos sensores, e estas respostas foram usadas como valores de entrada nos algoritmos. Então, a base de dados compreende um total de 180 (15x12) registros ou amostras e o modelo do algoritmo foi construído com base nesta base de dados.

Para a etapa de treinamento, 70% dos registros foram usados (128 amostras), e para a etapa de teste foram utilizadas as outras 52 amostras. Vale ressaltar, que o *k-fold* aplicado com 10 partições realiza o processo de validação apenas com as amostras utilizadas para treino, ou seja, para cada uma das 10 iterações, 13 dos 128 registros são testados. As etapas de treinamento e teste, para cada modelo, foram repetidas 10 vezes, considerando assim a mudança do estado aleatório na divisão do conjunto de dados, que eventualmente pode alterar os resultados obtidos em cada execução. No estudo foram consideradas as médias dos resultados obtidos nas 10 repetições.

Figura 8 – Equipamento montado no cultivar



4.3 PROCEDIMENTOS DE PRÉ-PROCESSAMENTO

As saídas digitais geradas pelos sensores de narizes eletrônicos devem ser analisadas e interpretadas a fim de fornecer o padrão olfativo da amostra. Esse procedimento envolve a extração de certas características significativas das curvas de resposta do sensor, para produzir um conjunto de dados numéricos que podem ser processados pelo sistema de reconhecimento e classificação do nariz eletrônico.

Uma das etapas de pré-processamento dos sinais dos sensores foi a aplicação do filtro da média móvel modificado, a fim de reduzir os possíveis ruídos gerados pelos sensores. Esse filtro foi usado, pois é o mais simples, rápido, robusto e de fácil implementação (ALESSIO et al., 2002). Uma média móvel simples sobre n elementos é constituída das médias não ponderadas dos subconjuntos de n elementos em um conjunto de dados. O valor adotado de n é 50. A diferença entre o filtro modificado e o original é que o primeiro só leva em consideração os valores que estejam em torno da média com diferença máxima de três vezes o desvio padrão, ou seja, $|p_i - \mu| \leq 3\sigma$, em que p_i é o elemento i do conjunto de dados p , μ é a média do conjunto e σ o desvio padrão. Portanto, o valor de n na equação 1 será: $1 \leq n \leq 50$, e dentro da curva da distribuição normal, 99,74% dos dados estarão em torno da média mais 3 desvios padrões. Essa alteração proporciona maior suavização no conjunto de dados, removendo possíveis picos de sinais devido a ruídos provenientes do circuito, pois desconsidera os valores discrepantes do conjunto de

dados.

Os sensores de gás possuem um comportamento que varia conforme a temperatura e a umidade relativa do ar. Os *datasheets* dos fabricantes mostram algumas curvas características dos comportamentos dos sensores, conforme essas duas variáveis mudam. Essas curvas usam como base a resistência dos sensores em temperatura ambiente e ar limpo (R_0) e em condições variantes (R_s). A resistência de carga (R_L) altera a sensibilidade dos sensores, e para o estudo aqui, todas foram ajustadas para 10 k Ω .

Usando os dados das curvas informados pelos fabricantes, foi realizada a compensação de resposta como uma função $R_{\text{ajust}}(R_s, x_0, x_1)$, em que R_{ajust} é a resistência ajustada, R_s é a resistência medida, x_0 é a temperatura medida e x_1 é a umidade relativa do ar. O modelo de ajuste linear múltiplo para todos os sensores tem a forma da equação 4.2:

$$R_{\text{ajust}} = \frac{R_s}{\alpha_0 x_0 + \alpha_1 x_0^2 + \beta_0 x_1 + \beta_1 x_1^2 + \delta x_0 x_1 + \varepsilon} \quad (4.2)$$

Aqui fica explícito a necessidade de utilização dos sensores de temperatura e umidade relativa do ar, uma vez que esses dois parâmetros são usados para o ajuste de respostas geradas pelos sensores de gás. A diferença entre as equações das curvas dos sensores está nas constantes α_0 , α_1 , β_0 , β_1 , δ e ε . A Tabela 9 mostra esses valores para os 13 sensores.

Tabela 9 – Valores das constantes utilizadas nas equações de ajustes dos sensores

Sensor	$\alpha_0(10^{-3})$	$\alpha_1(10^{-3})$	$\beta_0(10^{-3})$	$\beta_1(10^{-3})$	$\delta(10^{-3})$	ε	R^2
MQ-2	-26,6000	0,3000	-2,3000	0,0000	0,0000	1,4700	0,9989
MQ-3	-22,9000	0,2000	-3,6000	0,0000	0,0000	1,4640	0,9960
MQ-4	-10,3000	0,1000	-3,3000	0,0000	0,0000	1,2760	0,9970
MQ-5	-14,5000	0,1000	-4,0000	0,0000	0,0000	1,3350	0,9881
MQ-6	-12,6000	0,1000	-3,4000	0,0000	0,0000	1,2920	0,9941
MQ-7	-15,7000	0,1000	-3,6000	0,0000	0,0000	1,3580	0,9876
MQ-8	-10,6000	0,1000	-1,2000	0,0000	0,0000	1,0880	0,9962
MQ-9	-15,8000	0,1000	-3,7000	0,0000	0,0000	1,3590	0,9875
MQ-135	-25,8000	0,3000	-2,3000	0,0000	0,0000	1,4660	0,9978
TGS822	-57,6000	0,5000	-1,7900	0,1000	0,1000	2,7390	0,9910
TGS2600	-82,5000	0,8000	-18,7000	0,0000	0,3000	3,1140	0,9907
TGS2602	-56,6000	0,4000	-7,0000	0,0000	0,1000	2,2480	0,9944
TGS2603	-40,0000	0,2000	0,0000	0,0000	0,0000	1,7310	0,9980

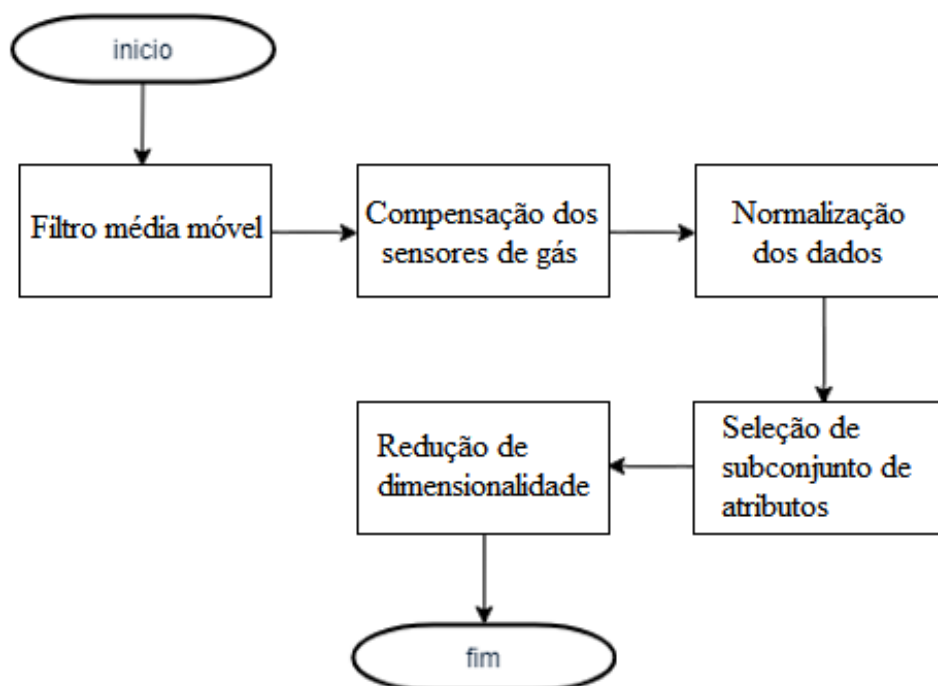
Com os dados normalizados através do método de normalização máximo-mínimo e com o objetivo de reduzir a quantidade de sensores ou de seleção de subconjunto de atributos, foi aplicado o teste *Chi-quadrado* de Pearson, a fim de verificar a associação entre cada uma das variáveis de entrada (sensores de gás) e a variável de saída (estágio de maturação). Esse teste fornece como saída o coeficiente V de Cramer, que é uma medida de associação entre duas variáveis nominais, dando um valor entre 0 (nenhuma associação entre as variáveis) e 1 (completa associação). Para este estudo foi adotado que caso a variável tenha o coeficiente maior

que 0,75 ela será mantida, e caso ela seja menor 0,55 ela será removida. Variáveis com os valores no intervalo entre 0,55 e 0,75 passaram pelo método de busca exaustiva (outro método de seleção de subconjunto de atributos) para os algoritmos de redução de dados e classificação, a fim de confirmar se a presença de cada uma delas influenciariam na precisão dos modelos, e, portanto, se seriam mantidas ou não na análise dos dados. Assim, no final, foi construído um modelo híbrido para seleção de subconjunto de atributos, combinando as duas técnicas de seleção de atributos descritas anteriormente. Portanto, neste trabalho foi estudado 2 comportamentos nas respostas dos sensores: os dados com os 13 sensores e os dados com os sensores reduzidos.

Por fim, os métodos PCA e LDA, para redução de dimensionalidade de dados, foram aplicados para os 2 conjuntos de dados - o primeiro contendo os 13 sensores de gás e o segundo reduzido - antes da etapa de construção dos modelos de classificação aplicados neste estudo.

A Figura 9 ilustra o fluxograma da etapa de pré-processamento do conjunto dos dados, que foi descrito com detalhes anteriormente.

Figura 9 – Fluxograma ilustrativo da etapa de pré-processamento dos dados



4.4 PROCEDIMENTOS DE PROCESSAMENTO DOS DADOS

Como dito na seção 4.2, para a implementação dos algoritmos de mineração de dados, foram desconsiderados os dados do regime transitório, que é o período necessário para estabilização da leitura dos sensores depois da amostra ser colocada para a leitura. Esse período de regime transitório leva cerca de 600 segundos.

Foram utilizados quatro algoritmos de classificação supervisionados: o KNN, que é o algoritmo mais simples, intuitivo e de fácil implementação para classificação (MUCHERINO;

PAPAJORGJI; PARDALOS, 2009); o SVM, que possui as vantagens de gerar fronteiras lineares ou não-lineares e o uso da representação em espaço de estados (RUIZ et al., 2018); o RF, que tem recebido crescente interesse por causa de sua previsão precisa e robustez ao ruído em comparação com classificadores únicos, além de ser computacionalmente mais leve que outros métodos de conjunto de árvores (LI; GU; WANG, 2017); e a RNA com uma simples camada oculta denominada ELM.

O KNN e SVM foram selecionados pois, como mostrado em Chen et al. (2018), bons resultados foram alcançados na classificação da maturação dos frutos utilizando-se estes métodos. Com relação aos demais algoritmos Brezmes et al. (2000) utilizou uma rede neural de única camada oculta e Benedetti et al. (2008) utilizou uma árvore de decisão (CART). Neste trabalho preferiu-se utilizar o RF em detrimento a CART, pois o RF, sendo um método *ensemble*, tende a gerar um resultado mais satisfatório.

Para o RF não foi necessário a aplicação do método da poda em alguns nodos a fim de evitar o *overfitting*. Alguns trabalhos recentes mostram que o uso de RF em narizes eletrônicos pode melhorar a taxa de precisão na regressão/classificação (LI; GU; WANG, 2017; MEN et al., 2018a).

A rede neural ELM apresenta melhor desempenho de treinamento comparado aos algoritmos tradicionais, além de facilitar a construção do modelo *online* (HUANG et al., 2012). Diferentemente dos algoritmos de aprendizado tradicionais, a ELM não apenas tende a alcançar o menor erro de treinamento, mas também menores valores de pesos. Os únicos parâmetros a serem ajustados é o tamanho da camada oculta e a função de aprendizado.

A etapa de construção do modelo envolve fazer o algoritmo de classificação, utilizar a base de dados e treinar o algoritmo com os dados coletados. Assim, é construído um modelo ideal (classificador) que é responsável por prever os novos dados inseridos na base de dados.

Todos os algoritmos foram implementados utilizando as linguagens R e Java, que possuem uma elevada gama de bibliotecas e funções que auxiliam no processo de construção dos modelos. Para o armazenamento dos dados foi utilizado o banco de dados relacional PostgreSQL.

4.5 PROCEDIMENTOS DE AVALIAÇÃO DOS DADOS

Para validação dos resultados obtidos com os classificadores foi utilizado o método *k-fold cross-validation*. A validação cruzada *k-fold* é uma técnica comumente empregada que, a partir de um conjunto de m exemplos, realiza a partição em K conjuntos (*folds*) de tamanhos iguais a m/K . Para cada conjunto, um classificador é treinado nos outros conjuntos (LANGFORD, 2005).

Portanto, a validação do modelo antes da etapa de teste irá auxiliar no processo de verificar quais os parâmetros que se ajustam melhor na construção de um classificador que gere uma resposta mais satisfatória para o tipo de conjunto de dados que envolve o problema. Assim,

o objetivo é construir um modelo ótimo de classificação para a base de dados.

Após isso, o modelo é testado com o conjunto de teste, ou seja, dados desconhecidos para o determinado modelo. Se o resultado com a base de teste é satisfatória, o modelo pode ser considerado aceito, caso contrário tem-se possivelmente um caso de sobre-ajuste ou *overfitting*, que é quando o modelo se ajusta demais e obtém respostas satisfatórias para a base de treino, mas para dados novos é ineficiente.

5 RESULTADOS E DISCUSSÕES

Neste capítulo serão expostas as curvas de respostas dados pelos sensores em diferentes etapas do ciclo de crescimento do fruto (seção 5.1). Os resultados do processo de remoção de atributos equivalentes serão abordados (seção 5.2), além das respostas em três dimensões geradas pelos métodos de redução de dimensionalidade utilizados no estudo (seção 5.3). Os resultados obtidos com os métodos de classificação, tanto para validação, quanto para teste, além das matrizes de confusão serão descritas neste capítulo (seção 5.4). Esses resultados também serão discutidos na seção 5.4, e por último será apresentada uma comparação com trabalhos correlatos, com análise dos resultados e metodologia (seção 5.5).

5.1 RESPOSTAS GERADAS PELOS SENSORES

A Figura 10(a) ilustra as respostas típicas dos sensores para um experimento medido no pessegueiro com pêssago imaturo. Lembrando que os primeiros 600 segundos compreendem o tempo de pré-aquecimento e o período de estado transitório. A Figura mostra a resposta obtida pelo primeiro dos 180 conjunto de dados. A Figura 10(b) mostra em detalhes as respostas dos sensores às curvas da Figura 10(a), com valores até 0,2 volts.

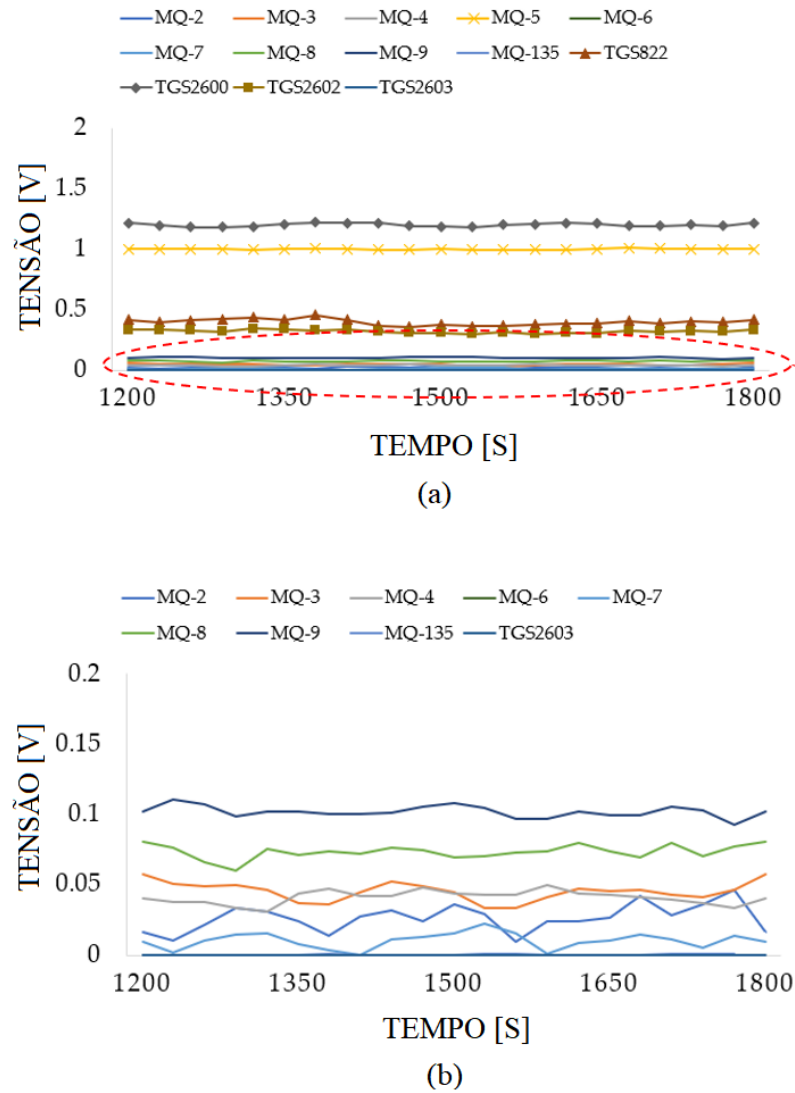
A Figura 11(a) ilustra as respostas típicas dos sensores para um experimento medido no pessegueiro com pêssago maduro. A Figura 11(b) mostra em detalhes as respostas dos sensores às curvas da 11(a), com valores inferiores a 0,2 volts. Da mesma maneira, a Figura 12(a) ilustra as respostas típicas dos sensores para um experimento medido no pessegueiro com pêssago super maduro e a Figura 12(b) ilustra em detalhes as curvas dos sensores da Figura 12(a), com valores abaixo de 0,2 volts.

Os dados obtidos mostraram que os estágios iniciais possibilitam uma estabilização anterior dos sensores, uma vez que a curva de resposta considerando o período de pré-aquecimento tem um comportamento exponencial decrescente, ou seja, para os estágios iniciais a curva é mais abrupta. As maiores concentrações de gases nos estágios mais avançados de maturação, fazem com que os sensores demorem um tempo maior para estabilização. Através das Figuras 10, 11 e 12 é visível a diferença de resposta dos sensores entre essas três amostras de estágios de maturação. Os dados obtidos, com mais detalhes e medidas estatísticas descritivas, podem ser visto no apêndice D.

5.2 REMOÇÃO DOS ATRIBUTOS EQUIVALENTES

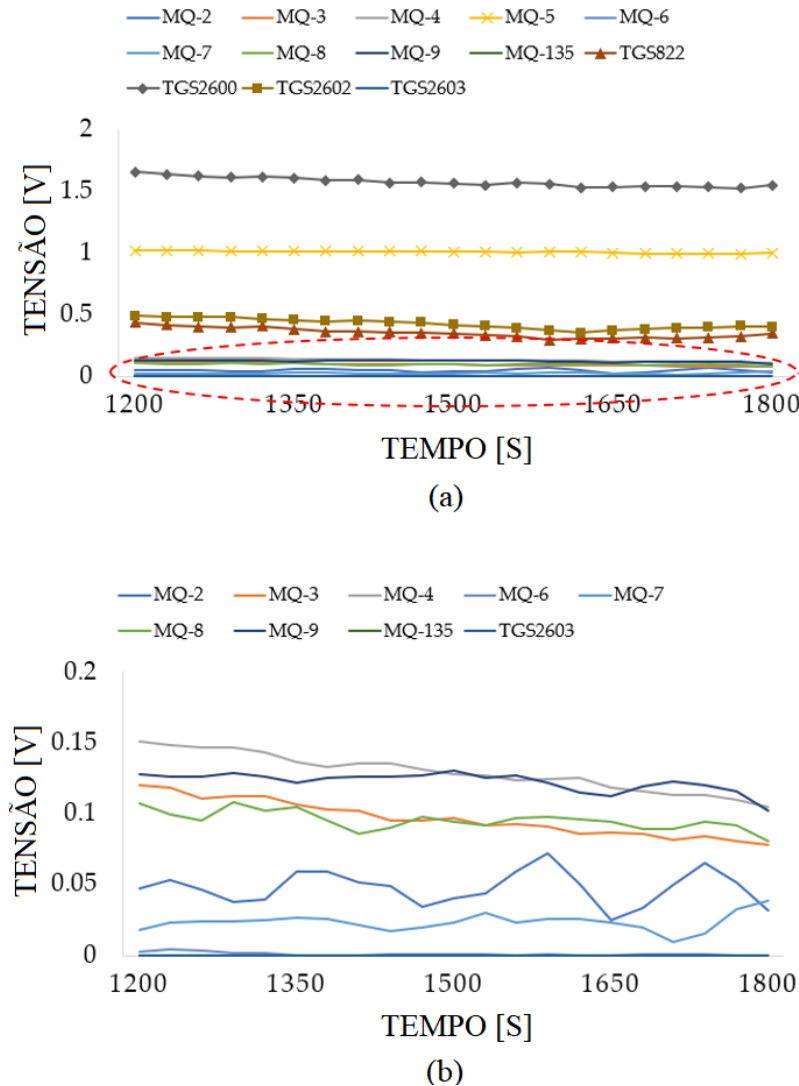
A fim de reduzir a quantidade de sensores, o teste *Chi-quadrado* mostrou que as variáveis TGS2603 e MQ-5 obtiveram um coeficiente V de Cramer maior que 0,75, e, portanto, consideradas importantes para o modelo. As variáveis MQ-6, MQ-9, TGS-822 e MQ-135 foram retiradas

Figura 10 – A resposta típica dada pelos sensores com amostras de pêssgo imaturo



da construção do modelo, pois obtiveram um coeficiente inferior ao limiar de 0,55. Para os sensores restantes, com valores do coeficiente V de Cramer entre 0,55 e 0,75, através da busca exhaustiva nos métodos de redução de dados e classificação, verificou-se que os atributos MQ-3, MQ-4, TGS2600 e TGS2602 resultaram em uma variância maior nos 3 primeiros componentes principais em comparação ao MQ-2, MQ-7 e MQ-8, e, portanto, foram mantidos no modelo de análise reduzido. Assim, no modelo de análise reduzido foi utilizado 6 sensores (MQ-3, MQ-4, MQ-5, TGS2600, TGS2602 e TGS2603) e no modelo de análise completo foi usado os 13 sensores iniciais. A Tabela 10 mostra os valores do coeficiente de Cramer V obtidos por cada sensor.

Figura 11 – A resposta típica dada pelos sensores com amostras de pêssgo maduro



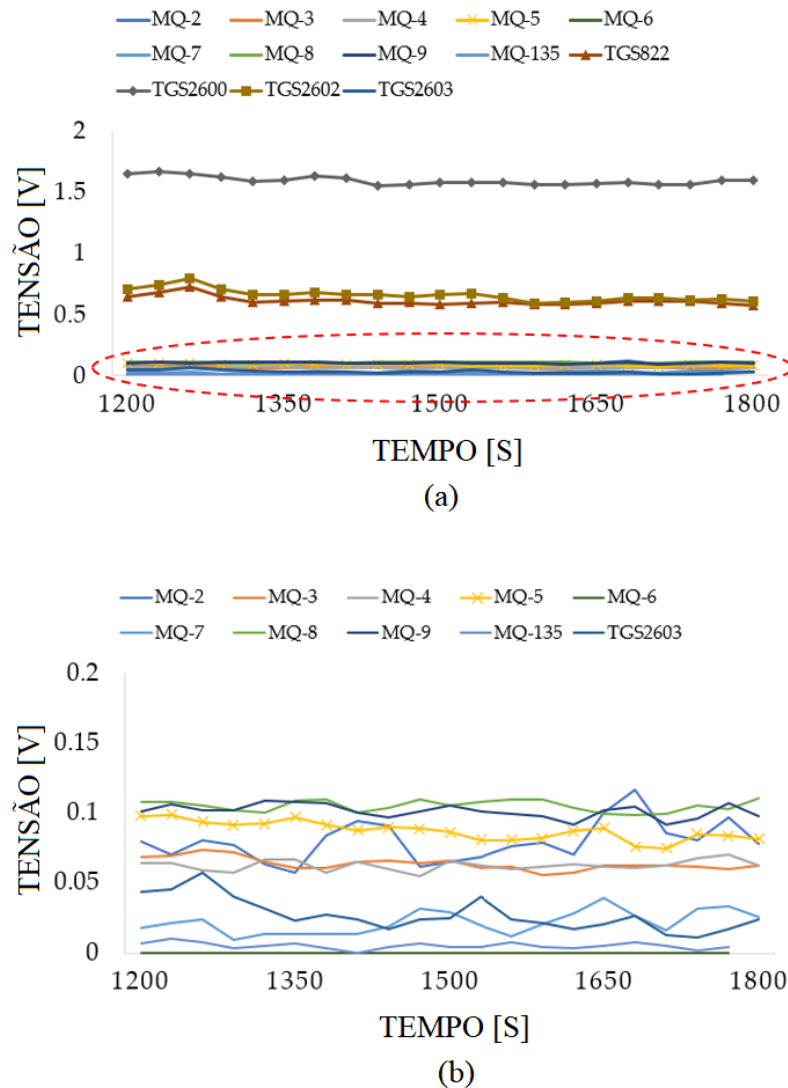
5.3 REDUÇÃO DA DIMENSIONALIDADE DOS DADOS

Com os dados coletados, o primeiro passo foi aplicar os métodos de redução de dimensionalidade. Uma vez que se tem respostas para 13 sensores, foi realizado a redução para três dimensões utilizando o PCA. Como ilustrado na Figura 13, os três componentes principais possuem uma variância total de 84,1% do conjunto original. Pode-se verificar que o estágio de pós-colheita (laranja) obteve uma maior discriminação comparado as outras fases. O mesmo procedimento para três dimensões foi realizado com a ferramenta LDA.

Através da Figura 14 é visível que o segundo método (LDA) conseguiu discriminar melhor, comparado ao PCA. A variância total obtida com o LD1, LD2 e LD3 foi de 100%, ou seja, com apenas três variáveis é possível representar o conjunto de dados obtidos por todos os sensores de gás.

O mesmo procedimento foi realizado para o conjunto de entrada reduzido. Como mos-

Figura 12 – A resposta típica dada pelos sensores com amostras de pêssego super maduro



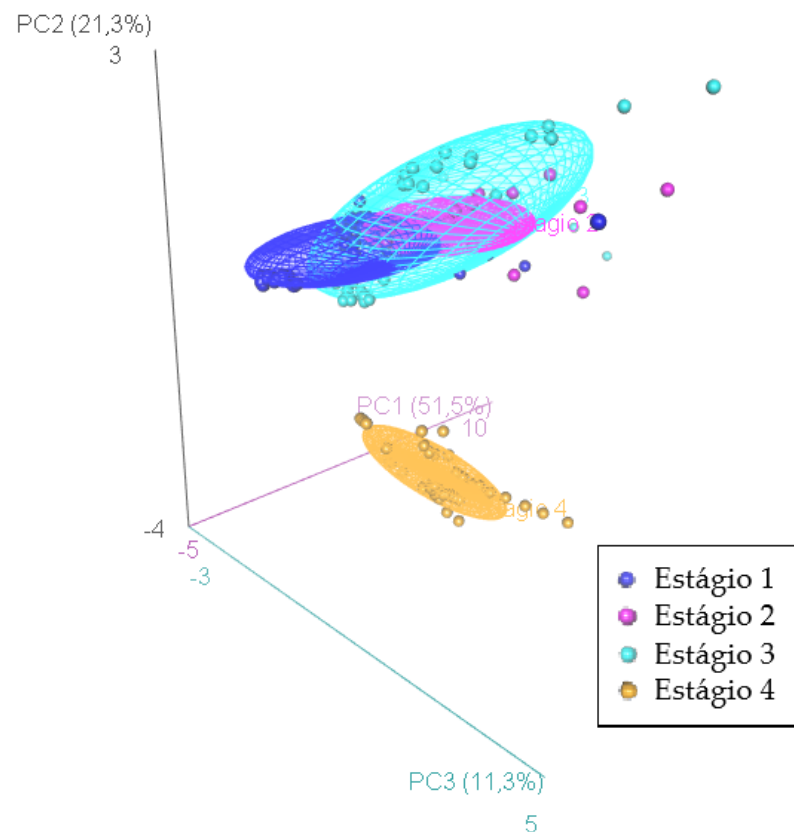
trado na Figura 15, os três componentes principais possuem uma variância total de 98,01% do conjunto original. Comparado com a Figura 13, existe uma maior variância e discriminação dos estágios de maturação.

Analisando a Figura 16, pode-se notar que o LDA, comparado ao PCA, também conseguiu agrupar com mais clareza as classes, e isto é provado com a variância nos LD1, LD2 e LD3 de 100%. Comparado ao conjunto de dados original também houve uma ligeira melhora no desempenho do algoritmo LDA, uma vez que o LD1 e LD2 são responsáveis por 99,32% da variabilidade dos dados contra 95,07% do conjunto total. Desse modo, para os 6 sensores, 2 componentes do LDA seria suficiente, mas a fim de comparação e padronização adotou-se utilizar 3 componentes para todos os casos.

Tabela 10 – Coeficiente V de Cramer para cada um dos sensores utilizados nos experimento

Sensor	Coeficiente V de Cramer
MQ-2	0,6905
MQ-3	0,6171
MQ-4	0,6369
MQ-5	0,7794
MQ-6	0,0000
MQ-7	0,6982
MQ-8	0,7324
MQ-9	0,4584
MQ-135	0,5297
TGS822	0,5340
TGS2600	0,7001
TGS2602	0,6163
TGS2603	0,8653

Figura 13 – Dados tridimensionais obtidos com 13 sensores pelo PCA



5.4 RESULTADOS DAS CLASSIFICAÇÕES E DISCUSSÕES

Além disso, foi realizada a classificação supervisionada com os quatro algoritmos (KNN, SVM, RF e ELM). Os dados obtidos pelos 13 sensores e pelos 6 sensores selecionados foram utilizados como entrada para os quatro métodos. Para cada algoritmo, também foi colocado como entrada os dados reduzidos pelo PCA e LDA. As Tabelas 11 e 12 resumem as porcentagens de

Figura 14 – Dados tridimensionais obtidos com 13 sensores pelo LDA

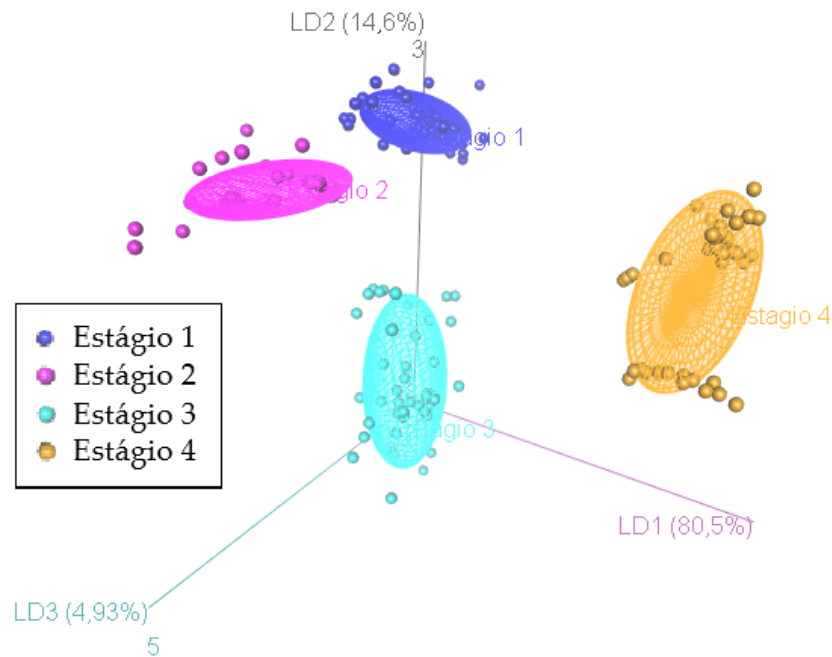
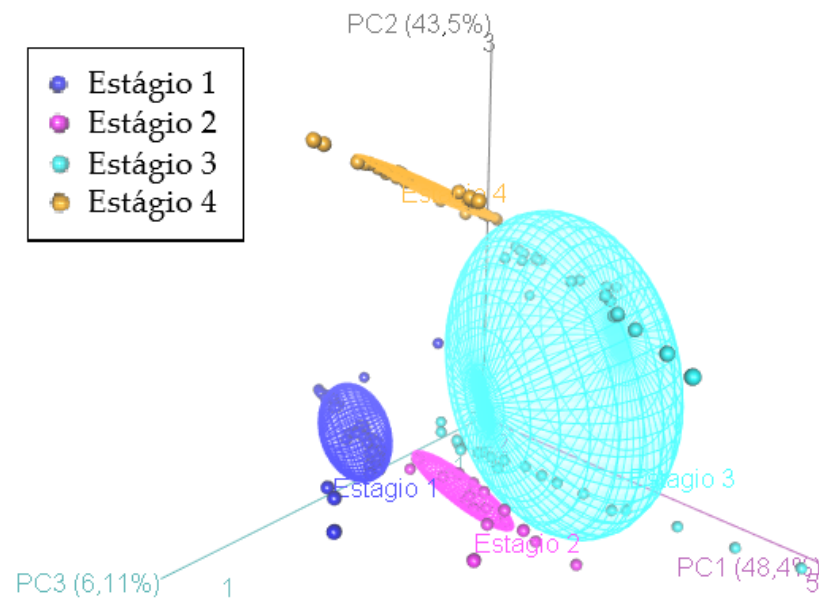
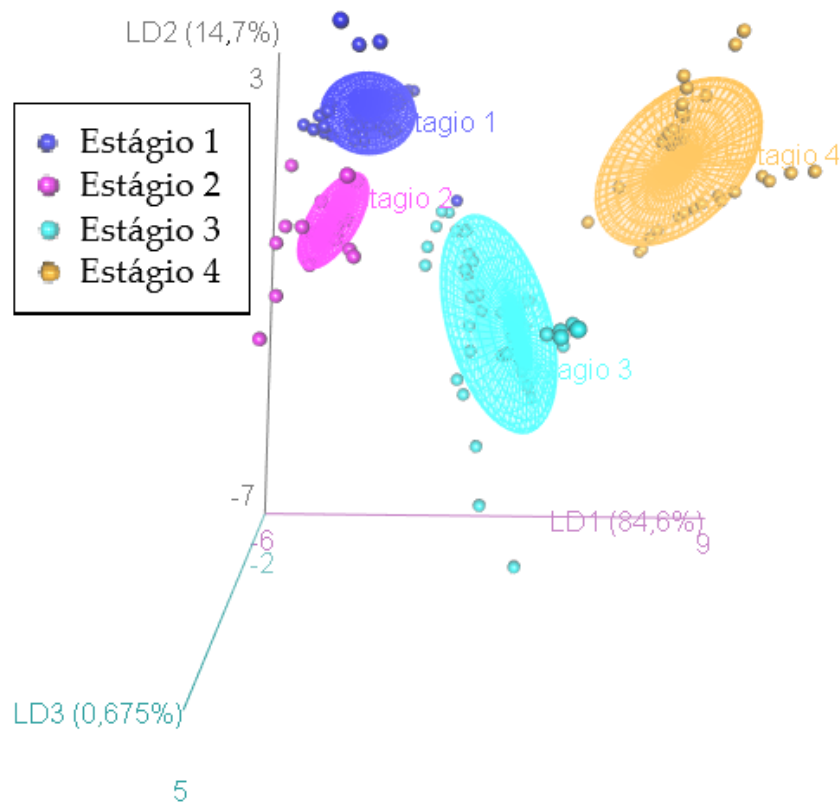


Figura 15 – Dados tridimensionais obtidos com 6 sensores pelo PCA



acerto na fase de validação (teste com os dados treinados) e teste (caixa-preta), para os dados totais e reduzidos, respectivamente, onde *mtry* é o número de variáveis amostradas aleatoriamente como candidatas em cada divisão no método RF; *nhid* é a quantidade de neurônios na camada oculta da ELM; *actfun* é a função de ativação utilizada nos neurônios; e *k* é quantidade máxima de vizinhos mais próximos. Além disso, as Tabelas 11 e 12 mostram os índices Kappa para as etapas de validação e teste, que é uma medida de concordância usada em escalas nominais que fornece um valor referente ao quanto as observações se afastam daquelas esperadas, fruto do

Figura 16 – Dados tridimensionais obtidos com 6 sensores pelo LDA



acaso, indicando assim o quão legítimas as interpretações são. O valor próximo de 0 significa baixa concordância e próximo de 1 alta concordância.

Esses resultados mostram que utilizando 6 sensores e sem nenhum método redutor de dados como entrada, todos os classificadores melhoraram seus desempenhos em relação aos resultados com 13 sensores, com exceção do SVM que obteve uma ligeira piora para a validação e manteve o índice de acerto para o conjunto de testes. Utilizando o PCA como entrada dos classificadores, o conjunto com 6 sensores obteve melhor taxa de acerto para todos os classificadores comparado ao conjunto com 13 sensores. Com o LDA como entrada, ocorreu uma melhora em todos os classificadores para o conjunto de 6 sensores em comparação aos 13 (com exceção do LDA-KNN na etapa de teste), e isso se deve ao fato que o LDA para os 6 sensores conseguiu agrupar melhor que para os 13 sensores, conforme as Figuras 14 e 16.

Comparando os classificadores para os mesmos conjuntos de sensores (6 e 13), a imputação com LDA foi superior ao PCA e sem imputação para praticamente todos os classificadores, tanto para a validação quanto para o teste, pois como mostrado nas Figuras 13 a 16 o LDA além de reduzir a dimensionalidade, facilita o trabalho dos classificadores, caso sejam detectados possíveis agrupamentos. Com isso, proporcionou melhora na acurácia para quase todos os classificadores, em ambas as etapas. Assim, o LDA, além de melhorar classificação possibilita uma redução considerável do conjunto de dados dos sensores, o que é interessante quando se trabalha com dispositivos embarcados e para classificação *online*.

Tabela 11 – Resultados obtidos com os classificadores para o conjunto de dados com 13 sensores

Método	Acurácia validação (10-fold)	Índice Kappa validação (10-fold)	Acurácia teste	Índice Kappa teste
KNN ($k=5$)	93,32%	0,9112	92,31%	0,8974
SVM	98,54%	0,9805	98,08%	0,9744
RF ($mtry=2$)	93,87%	0,9181	96,15%	0,9487
ELM ($nhid=5$, $actfun=purelin$)	78,80%	0,7171	61,54%	0,4872
PCA - KNN ($k=5$)	91,02%	0,8801	88,46%	0,8462
PCA - SVM	78,61%	0,7139	76,92%	0,6923
PCA - RF ($mtry=2$)	91,24%	0,8824	90,38%	0,8718
PCA - ELM ($nhid=5$, $actfun=tansig$)	78,08%	0,7073	73,08%	0,6410
LDA - KNN ($k=5$)	97,92%	0,9721	98,08%	0,9744
LDA - SVM	98,42%	0,9790	98,08%	0,9744
LDA - RF ($mtry=2$)	98,47%	0,9796	98,08%	0,9744
LDA - ELM ($nhid=5$, $actfun=purelin$)	89,80%	0,8633	86,54%	0,8205

Tabela 12 – Resultados obtidos com os classificadores para o conjunto de dados com 6 sensores

Método	Acurácia validação (10-fold)	Índice Kappa validação (10-fold)	Acurácia teste	Índice Kappa teste
KNN ($k=5$)	95,27%	0,9367	96,15%	0,9487
SVM	97,36%	0,9647	98,08%	0,9744
RF ($mtry=4$)	93,90%	0,9186	90,38%	0,8718
ELM ($nhid=5$, $actfun=purelin$)	88,90%	0,8515	90,38%	0,8718
PCA - KNN ($k=5$)	96,55%	0,9539	96,15%	0,9487
PCA - SVM	93,91%	0,9185	98,08%	0,9744
PCA - RF ($mtry=3$)	96,58%	0,9542	96,15%	0,9487
PCA - ELM ($nhid=5$, $actfun=tansig$)	87,82%	0,8372	90,38%	0,8718
LDA - KNN ($k=5$)	98,41%	0,9789	96,15%	0,9487
LDA - SVM	99,20%	0,9894	98,08%	0,9744
LDA - RF ($mtry=2$)	100,00%	1,0000	98,08%	0,9744
LDA - ELM ($nhid=5$, $actfun=purelin$)	90,15%	0,8681	88,46%	0,8462

Em suma, o classificador RF com imputação do LDA para o conjunto dos 6 sensores foi o único a atingir uma taxa de acerto total, na etapa de validação, e uma taxa de acerto de 98,08% na etapa de teste (Tabela 12). Isso demonstra que a redução inicial com o teste *Chi-quadrado* para 6 sensores, a subsequente redução de dimensionalidade com o LDA para 3 dimensões e a utilização do algoritmo correto (RF), melhora a taxa de classificação do modelo e contribui para a construção de um dispositivo ideal para a aplicação estudada.

Como descrito anteriormente, 30% do conjunto de amostras totais foram utilizadas para teste. Com o total de 52 amostras, cada um dos quatro estágios ficou com 13 amostras para teste. As Tabelas 13 a 18 mostram as matrizes de confusão obtidas através do conjunto de amostras de teste classificadas entre os 4 estágios para cada algoritmo classificador em cada conjunto de dados.

Tabela 13 – Matriz de confusão para todos os classificadores sem metodo de redução como entrada utilizando 13 sensores

KNN					
Referência \ Predição		Estágio 1	Estágio 2	Estágio 3	Estágio 4
		Estágio 1	Estágio 2	Estágio 3	Estágio 4
Estágio 1		13	1	0	0
Estágio 2		0	10	1	0
Estágio 3		0	2	12	0
Estágio 4		0	0	0	13
SVM					
Referência \ Predição		Estágio 1	Estágio 2	Estágio 3	Estágio 4
		Estágio 1	Estágio 2	Estágio 3	Estágio 4
Estágio 1		13	0	0	0
Estágio 2		0	12	0	0
Estágio 3		0	1	13	0
Estágio 4		0	0	0	13
RF					
Referência \ Predição		Estágio 1	Estágio 2	Estágio 3	Estágio 4
		Estágio 1	Estágio 2	Estágio 3	Estágio 4
Estágio 1		13	1	0	0
Estágio 2		0	11	0	0
Estágio 3		0	1	13	0
Estágio 4		0	0	0	13
ELM					
Referência \ Predição		Estágio 1	Estágio 2	Estágio 3	Estágio 4
		Estágio 1	Estágio 2	Estágio 3	Estágio 4
Estágio 1		8	5	2	0
Estágio 2		3	8	3	0
Estágio 3		0	0	5	0
Estágio 4		2	0	3	13

A Tabela 13 mostra que o estágio 2, para todos os classificadores, teve a maior quantidade de amostras classificadas de forma incorreta. O algoritmo ELM teve a maior quantidade de amostras preditas erroneamente, nos quais apenas 5 das 13 amostras do estágio 3 foram classificadas corretamente. As Tabelas 14 e 15 mostram que todos os classificadores imputados com o PCA e LDA, respectivamente, tiveram maior dificuldade em classificar amostras do estágio 3, em comparação as demais. Além disso, a imputação com LDA juntamente com o ELM foi o único modelo a ter uma amostra incorretamente classificada no último estágio.

Considerando o conjunto de dados reduzido, a Tabela 16 mostra que o estágio 2 teve

Tabela 14 – Matriz de confusão para todos os classificadores com o PCA como entrada utilizando 13 sensores

PCA - KNN					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	12	0	1	0	
Estágio 2	1	13	4	0	
Estágio 3	0	0	8	0	
Estágio 4	0	0	0	13	
PCA - SVM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	10	3	5	0	
Estágio 2	0	9	0	0	
Estágio 3	3	1	8	0	
Estágio 4	0	0	0	13	
PCA - RF					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	10	0	1	0	
Estágio 2	1	13	1	0	
Estágio 3	2	0	11	0	
Estágio 4	0	0	0	13	
PCA - ELM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	10	2	3	0	
Estágio 2	0	11	6	0	
Estágio 3	3	0	4	0	
Estágio 4	0	0	0	13	

Tabela 15 – Matriz de confusão para todos os classificadores com o LDA como entrada utilizando 13 sensores

LDA - KNN					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
LDA - SVM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
LDA - RF					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
LDA - ELM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	10	1	0	1	
Estágio 2	3	12	2	0	
Estágio 3	0	0	11	0	
Estágio 4	0	0	0	12	

Tabela 16 – Matriz de confusão para todos os classificadores sem metodo de redução como entrada utilizando 6 sensores

KNN					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	11	0	0	0	
Estágio 2	2	13	0	0	
Estágio 3	0	0	13	0	
Estágio 4	0	0	0	13	
SVM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	1	0	0	
Estágio 2	0	12	0	0	
Estágio 3	0	0	13	0	
Estágio 4	0	0	0	13	
RF					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	11	0	0	0	
Estágio 2	2	10	0	0	
Estágio 3	0	3	13	0	
Estágio 4	0	0	0	13	
ELM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	4	0	0	
Estágio 2	0	9	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	

Tabela 17 – Matriz de confusão para todos os classificadores com o PCA como entrada utilizando 6 sensores

PCA - KNN					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	0	0	
Estágio 3	0	0	11	0	
Estágio 4	0	0	2	13	
PCA - SVM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
PCA - RF					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	1	0	0	
Estágio 2	0	12	0	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	1	13	
PCA - ELM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	12	0	0	0	
Estágio 2	1	13	4	0	
Estágio 3	0	0	9	0	
Estágio 4	0	0	0	13	

Tabela 18 – Matriz de confusão para todos os classificadores com o LDA como entrada utilizando 6 sensores

LDA - KNN					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	12	1	0	0	
Estágio 2	1	12	0	0	
Estágio 3	0	0	13	0	
Estágio 4	0	0	0	13	
LDA - SVM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
LDA - RF					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	13	0	0	0	
Estágio 2	0	13	1	0	
Estágio 3	0	0	12	0	
Estágio 4	0	0	0	13	
LDA - ELM					
Predição \ Referência	Estágio 1	Estágio 2	Estágio 3	Estágio 4	
Estágio 1	12	2	1	0	
Estágio 2	1	11	2	0	
Estágio 3	0	0	10	0	
Estágio 4	0	0	0	13	

maior parcela de amostras classificadas erroneamente, com exceção do algoritmo KNN. As Tabelas 17 e 18 mostram maior dificuldade em predição de amostras do estágio 3, nos quais o ELM teve a maior quantidade de amostras incorretas para ambos os métodos de redução de dimensionalidade (PCA e LDA). As Tabelas 13 a 18 mostram que todos os modelos tiveram pelo menos uma amostra classificada incorretamente. Comparativamente, o conjunto de dados reduzido teve uma quantidade total menor de amostras classificadas incorretamente (33 amostras) em comparação ao conjunto de dados original com os 13 sensores (72 amostras), considerando todos os modelos estudados.

Como mostrado em (GUOHUA et al., 2012; VISAI; VANOLI, 1997), e com base na Tabela 1, durante o estágio 1 (primeiros dias após a plena floração), os álcoois (1-Hexanol, (E)-2-hexenol e (Z)-3-hexenol) são os principais componentes exalados pelos pêssgo. Os aldeídos (Hexanal, (E)-2-hexenal e Benzaldeído), os ésteres (Acetato de hexila e Acetato de (E)-2-hexenilo) e a Acetoína possuem um pico de emissão estágio 2 e que decresce após o início do estágio 3 (dia 51 após a plena floração). Os Terpenóides (α -terpineno, γ -terpineno, Terpinoleno e Linalol), conforme o estudo em (VISAI; VANOLI, 1997) são exalados em uma taxa mais abundante no dia 61 após a plena floração, que já compreende o terceiro estágio de crescimento do fruto. Por fim, os lactonas (γ -hexalactona, γ -heptalactona, γ -octalactona, γ -nonalactona, γ -dodecalactona, γ -decalactona, δ -decalactona, δ -dodecalactona) serão identificados somente no estágio 4, após 80 dias da plena floração. Portanto, o dispositivo é capaz, através da detecção

do estágio de crescimento do pêssgo, identificar quais COVs estão sendo exalados com maior abundância no determinado período.

5.5 COMPARAÇÃO COM OUTROS TRABALHOS

A Tabela 19 faz uma comparação dos resultados dos trabalhos correlatos com o desenvolvido neste trabalho. Pode-se perceber que a melhor técnica desta pesquisa (LDA-RF) possibilita a redução do conjunto de dados gerados pelos sensores, além da taxa de classificação superar a maioria dos trabalhos com o propósito de medir a maturação do fruto do pêssgo.

Tabela 19 – Resumo dos resultados dos trabalhos relacionados de maturação do pêssgo

Objetivo	Técnicas utilizadas	Parâmetros	Referência
Estágio de maturação pós-colheita	RNA	Erro = 7% (Teste)	Brezmes et al. (2000)
Estágio de maturação pós-colheita	PCA e CART	Erro = 2,44% (Teste)	Benedetti et al. (2008)
Estágio de maturação pós-colheita	PCA e SR	Erro = 15% (Teste)	Guohua et al. (2012)
Estágio de maturação na colheita	PCA e PLS	SEP = 8,77 (Teste)	Zhang et al. (2012)
Estágio de maturação / monitoramento <i>in loco</i>	LDA - RF	Erro = 1,92% (Teste)	Este trabalho

O trabalho desenvolvido por Brezmes et al. (2000) utilizou uma RNA de três camadas para classificação do estado de maturação do pêssgo pós-colheita. O erro de 7% foi alcançado, pois a rede foi alimentada, além dos valores de resposta dos sensores TGS882 e TGS2610, por outros 2 parâmetros medidos: peso do fruto e área de superfície média. O trabalho destaca-se pela utilização de poucos sensores, mas ainda foi necessário trabalhar com outros parâmetros, que no estado da arte exigiria um sistema de aquisição e processamento de imagens a fim de obter um resultado mais satisfatório.

Benedetti et al. (2008) aplicaram um nariz eletrônico a fim de classificar 4 tipos de cultivares entre os estágios de maturação pós-colheita. Apesar de ser utilizado um equipamento comercial, que pode agregar um custo mais alto em detrimento a um dispositivo construído utilizando sensores descritos no capítulo 2, o trabalho destaca-se por trabalhar com diferentes cultivares, o que demonstra uma maior aplicabilidade do aparelho. Futuramente, é desejável também testar o desempenho do dispositivo elaborado nesta pesquisa entre diferentes cultivares. A árvore de decisão mostrou-se ter um desempenho satisfatório, uma vez que o erro de validação

foi de 4,87% e de teste foi de 2,44%. A utilização do classificador de árvore de decisão *ensemble* - RF, poderia diminuir ainda mais o erro, uma vez que algoritmos conjuntos tendem a ter uma melhor taxa de acerto.

Guohua et al. (2012) construíram um nariz eletrônico para determinar o frescor do pêssogo, ou seja, determinar a 'idade' de determinada amostra pós-colheita. Apesar de serem testadas amostras de apenas um cultivar, o dispositivo destaca-se por classificar as amostras em um período mais reduzido de tempo (entre dias), o que tende a ser mais desafiador. Com o estudo, conseguiu-se verificar que o sensor TGS842 não obteve resposta para nenhuma medida, e portanto pode ser eliminado do protótipo. Considerando a aplicabilidade do sistema, o erro de 15% pode ser considerado aceitável, uma vez que o SR gera uma resposta quantitativa referente a idade da amostra. Das 20 amostras, apenas 3 tiveram um erro na ordem de 20%. O restante foram menores que 7%. Baseando-se nesse estudo, um equipamento de monitoramento no pomar, classificando amostras em um range mais específico é desejável.

Zhang et al. (2012) desenvolveram um equipamento capaz de detectar o frescor do pêssogo, assim como Guohua et al. (2012). Apenas amostras pertencentes ao cultivar '*Dabai*' foram testadas. O equipamento além de ser capaz de discriminar entre os estados de maturação da fruta, é capaz de identificar 3 características (força penetrante, quantidade de açúcar e índice de acidez). O equipamento destaca-se por gerar uma resposta quantitativa desses 3 índices, apesar da aplicabilidade pós-colheita. Assim, uma potencial aplicação seria identificar esses índices no fruto, com um equipamento semelhante ao que foi construído neste trabalho. Com a análise quantitativa, pode-se testar métodos mais simples, como regressão linear, a fim de se verificar a precisão do modelo.

A fim de comparação, o dispositivo apresentado nesta pesquisa, além de possibilitar as medidas em campo aberto (*in loco*), devido à estrutura diferenciada, possui uma taxa de erro bastante reduzida, em comparação aos outros trabalhos com narizes eletrônicos para a medição da maturação dos pêssogos encontrados na literatura, já que o protótipo construído obteve uma taxa de 98,08% no melhor dos casos, contra 93% de Brezmes et al. (2000), 97,56% de Benedetti et al. (2008) e 85% de Guohua et al. (2012). A metodologia adotada por Zhang et al. (2012) mostra um considerável SEP de 8,77.

As aplicações dos sensores de gás apresentam vários desafios, pois estes apresentam uma sensibilidade grande para variações de temperatura e umidade relativa do ar, bem como a interferência do gás ambiente na medida e na análise dos resultados. Isso pode ser amenizado utilizando um sensor de espectroscopia de massa, uma vez que estes sensores possuem uma boa estabilidade e sensibilidade em comparação com uma matriz de sensores químicos. Mas como mostrado na seção 4.3 do capítulo 4 isso pode ser compensado através de equações geradas por curvas disponibilizadas por fabricantes, reduzindo os erros gerados por essas variáveis. Por outro lado, os sensores de espectroscopia de massa mostram-se mais custosos que sensores químicos, o que acaba inibindo a sua aplicação em potencial. O desenvolvimento de narizes eletrônicos

surge como uma solução para esses problemas de instabilidade de medição, necessidade de calibração frequente (deriva do sensor) e alto consumo de energia desses sensores químicos, além de serem relativamente mais baratos quando comparados a outros tipos de equipamentos. Os narizes eletrônicos mostram-se portáteis, altamente versáteis e facilmente treinados para serem empregados em uma grande variedade de aplicações de detecção de gases no ambiente.

6 CONCLUSÕES

Com o objetivo da identificação do amadurecimento dos pêssegos Eragil, *in loco*, de forma não-invasiva, através da presença dos COVs, durante o ciclo de crescimento, começando no período de plena floração até o período pós-colheita, o protótipo de nariz eletrônico construído com os 13 sensores de gás MOS, o sensor de temperatura e pressão atmosférica BMP180, o de umidade relativa do ar HIH-4030 e o sistema de ventilação com o ar ambiente obteve êxito na classificação dos estágios de maturação do pêssego. Assim, os produtores rurais têm a possibilidade de obter um dispositivo robusto capaz de monitorar os gases emitidos nos pessegueiros e que custa cerca de R\$ 485,00.

A matriz de sensores projetada para o uso em dispositivos microcontroladores de prototipagem rápida facilitou o processo de montagem do dispositivo e possíveis alterações provenientes na estrutura do protótipo.

Os *firmwares* auxiliaram no processo de construção da base de dados, uma vez que foram responsáveis pela aquisição dos dados e transmissão via comunicação serial. A base de dados construída, por sua vez, poderá ser utilizada juntamente com novos dados obtidos em diferentes áreas cultivadas e períodos de tempo entre safras.

As 180 amostras dos 4 estágios de crescimento (desde a floração até a pós-colheita), com o resultado de acerto total na validação do modelo e de 98,08% no teste (51 das 52 amostras classificadas corretamente), utilizando classificador e redutor de dados combinados (LDA-RF), para os sensores reduzidos à 6, demonstra a adequação do nariz eletrônico para monitoramento de áreas com uma intensa emissão de COVs. Isso demonstra que a aplicação do método correto para a redução de dimensionalidade e a implementação do algoritmo adequado, melhora o desempenho do classificador.

Além disso, a etapa de remoção dos sensores através do teste *Chi-quadrado* auxiliou no processo de remoção de atributos que prejudicariam o desempenho dos classificadores, pois para a maioria dos métodos aplicados houve uma melhora tanto na validação quanto no teste, além de permitir a redução do número de sensores. O LDA, além de melhorar o taxa de acerto dos algoritmos de classificação, comparado a imputação com o PCA e sem imputação, possibilitou a redução do conjunto de dados dos sensores, que é importante quando se trabalha com dispositivos embarcados monitorando de forma *online*. Comparativamente aos trabalhos correlatos encontrados na literatura, o dispositivo apresentado além de possibilitar a aplicação em campo, obteve uma taxa de erro inferior aos demais.

O equipamento desenvolvido mostrou-se promissor, demonstrando ter sensibilidade para a detecção de COVs emitidos no ambiente de pessegueiros Eragil. Portanto, os produtores rurais poderão identificar o estágio de maturação de forma clara e precisa, diminuindo o risco de

perdas de frutos em pontos que a maturação não seja a desejável para a comercialização, além de permitir o medição de forma não-destrutiva da qualidade do fruto.

7 TRABALHOS FUTUROS

Alternativas de trabalhos futuros foram identificadas, com o propósito de aprimorar o dispositivo e os resultados obtidos neste trabalho, nas quais se destacam:

1. **Realizar a reconstrução do modelo:** desta vez treinando-o com amostras de diversas safras, além de aplicar o equipamento para o monitoramento de outras espécies de pêssego, frutas e culturas;
2. **Construir uma topologia em rede sem fio:** identificando aquela mais apropriada para o objetivo em questão, com diversos narizes eletrônicos conectados, realizando a coleta dos dados continuamente e em diversos pontos da cultura tratada, pois pode ocorrer de áreas cultivadas estarem com os frutos mais ou menos maduros;
3. **Realizar o monitoramento *online* do cultivar:** a detecção *online*, contínua e em tempo real permite detectar precisamente quais os períodos do dia em que ocorre maior produção de gases, além de permitir a detecção do ponto ótimo de colheita;
4. **Classificar as amostras em maiores espectros:** além das quatro fases, obter um espectro mais preciso de grau de maturação, como, por exemplo, no estágio final, onde o fruto pode estar imaturo, maduro ou super maduro, com a aplicação de novas ferramentas e algoritmos a fim de ter acurácias similares as obtidas neste trabalho;
5. **Utilizar novos sensores de gás:** neste trabalho foram utilizados ao todo 13 sensores da família MQ e TGS. Sugere-se aplicar sensores de outros modelos e/ou fabricantes, a fim de incrementar a capacidade de detectar os diversos gases exalados nas culturas. Vale ressaltar que cada modelo de sensor é mais sensível a um determinado tipo de gás e concentração específica, assim é necessário um estudo mais aprofundado para identificar os melhores sensores disponíveis para o propósito estudado. Além do mais, a utilização de outros sensores além de auxiliar no processo de miniaturização do protótipo, pode diminuir a quantidade de dados, o custo de fabricação, aumentar a velocidade de processamento e incrementar o desempenho do equipamento na classificação de novas amostras de dados, uma vez que o espectro de sensores necessários para a aplicação pode ser menor que o encontrado neste estudo, que foi de 6.

REFERÊNCIAS

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, Wiley Online Library, v. 2, n. 4, p. 433–459, 2010.
- ABRAHAM, K.; PANDIAN, S. A low-cost mobile urban environmental monitoring system. In: IEEE. *Intelligent Systems Modelling & Simulation (ISMS), 2013 4th International Conference on*. [S.l.], 2013. p. 659–664.
- AL-DAHOUD, A.; FEZARI, M.; JANNOUD, I.; AL-RAWASHDEH, T. Monitoring metropolitan city air-quality using wireless sensor nodes based on arduino and xbee. p. 121–125, 2013.
- ALEIXANDRE, M.; CABELLOS, J. M.; ARROYO, T.; HORRILLO, M. Quantification of wine mixtures with an electronic nose and a human panel. *Frontiers in bioengineering and biotechnology*, Frontiers, v. 6, p. 14, 2018.
- ALEIXANDRE, M.; SANTOS, J. P.; SAYAGO, I.; CABELLOS, J. M.; ARROYO, T.; HORRILLO, M. C. A wireless and portable electronic nose to differentiate musts of different ripeness degree and grape varieties. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 15, n. 4, p. 8429–8443, 2015.
- ALESSIO, E.; CARBONE, A.; CASTELLI, G.; FRAPPIETRO, V. Second-order moving average and scaling of stochastic time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, Springer, v. 27, n. 2, p. 197–200, 2002.
- BAGULA, A.; ZENNARO, M.; INGGIS, G.; SCOTT, S.; GASCON, D. Ubiquitous sensor networking for development (usn4d): An application to pollution monitoring. *Sensors*, Molecular Diversity Preservation International, v. 12, n. 1, p. 391–414, 2012.
- BAIETTO, M.; WILSON, A. Electronic-nose applications for fruit identification, ripeness and quality grading. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 15, n. 1, p. 899–931, 2015.
- BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, v. 18, p. 1–8, 1998.
- BALKE, M. E. *Desidratação de pêsego em diferentes processos: modelagem matemática e análise empreendedora*. Tese (Doutorado) — Tese (Doutorado em Engenharia de Alimentos) - Universidade Regional Integrada do Alto Uruguai e das Missões, Erechin, 2016.
- BARALDI, R.; RAPPARINI, F.; ROSSI, F.; LATELLA, A.; CICCIOI, P. Volatile organic compound emissions from flowers of the most occurring and economically important species of fruit trees. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, Elsevier, v. 24, n. 6, p. 729–732, 1999.
- BARBOSA, W.; CAMPO-DALL'ORTO, F.; OJIMA, M.; SAMPAIO, V.; BANDEL, G. *Ecofisiologia do desenvolvimento vegetativo e reprodutivo do pessegueiro em região subtropical. Campinas: Instituto Agrônômico*, 1990.

- BARBOSA, W.; OJIMA, M.; DALL'ORTO, F. A. C.; MARTINS, F. P.; LOVATE, A. A. Desenvolvimento dos frutos e das sementes de pêssegos subtropicais de diferentes ciclos de maturação. *Pesquisa Agropecuária Brasileira*, v. 28, n. 6, p. 701–707, 1993.
- BENEDETTI, S.; BURATTI, S.; SPINARDI, A.; MANNINO, S.; MIGNANI, I. Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life. *Postharvest biology and technology*, Elsevier, v. 47, n. 2, p. 181–188, 2008.
- BIEGANOWSKI, A.; JÓZEFACIUK, G.; BANDURA, L.; GUZ, Ł.; ŁAGÓD, G.; FRANUS, W. Evaluation of hydrocarbon soil pollution using e-nose. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 8, p. 2463, 2018.
- BLANCO-NOVOA, O.; FERNÁNDEZ-CARAMÉS, T.; FRAGA-LAMAS, P.; CASTEDO, L. A cost-effective iot system for monitoring indoor radon gas concentration. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 7, p. 2198, 2018.
- BRACKMANN, A.; PAVANELLO, E. P.; BOTH, V.; WEBER, A.; PINTO, J. A. V. Atmosfera refrigerada e controlada para pêssegos 'eragil'. *Ciência Rural*, SciELO Brasil, v. 39, n. 7, 2009.
- BREZMES, J.; LLOBET, E.; VILANOVA, X.; SAIZ, G.; CORREIG, X. Fruit ripeness monitoring using an electronic nose. *Sensors and Actuators B: Chemical*, Elsevier, v. 69, n. 3, p. 223–229, 2000.
- BUCK, T.; ALLEN, F.; DALTON, J. *Detection of chemical species by surface effects on metals and semiconductors*. [S.l.]: Bell Telephone Laboratories, 1965.
- CAMILLO, M. F. *Danos mecânicos e seus efeitos na qualidade pós-colheita de frutos de caroço*. Dissertação (Mestrado) — Dissertação (Mestrado em Fitotecnia) - Universidade Estadual do Rio Grande do Sul, Porto Alegre, 2009.
- CAPELLI, L.; SIRONI, S.; ROSSO, R. D. Electronic noses for environmental monitoring applications. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 14, n. 11, p. 19979–20007, 2014.
- CHEN, L.-Y.; WU, C.-C.; CHOU, T.-I.; CHIU, S.-W.; TANG, K.-T. Development of a dual mos electronic nose/camera system for improving fruit ripeness classification. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 10, p. 3256, 2018.
- COUTO, M. Efeito da temperatura durante a diferenciação de gemas, floração, crescimento e desenvolvimento de frutos em pessegueiro na região de pelotas, rs. Universidade Federal de Pelotas, 2006.
- DESHMUKH, S.; BANDYOPADHYAY, R.; BHATTACHARYYA, N.; PANDEY, R.; JANA, A. Application of electronic nose for industrial odors and gaseous emissions measurement and monitoring—an overview. *Talanta*, Elsevier, v. 144, p. 329–340, 2015.
- DORJI, U.; POBKURUT, T.; KERDCHAROEN, T. Electronic nose based wireless sensor network for soil monitoring in precision farming system. In: IEEE. *Knowledge and Smart Technology (KST), 2017 9th International Conference on*. [S.l.], 2017. p. 182–186.
- DRAVNIEKS, A.; TROTTER, P. Polar vapour detector based on thermal modulation of contact potential. *Journal of Scientific Instruments*, IOP Publishing, v. 42, n. 8, p. 624, 1965.

- DUTRA, L. F.; KERSTEN, E.; FACHINELLO, J. C. Cutting time, indolebutyric acid and tryptophan in rooting of peach tree cuttings. *Scientia Agricola*, SciELO Brasil, v. 59, n. 2, p. 327–333, 2002.
- FRADEN, J. *Handbook of modern sensors: physics, designs, and applications*. [S.l.]: Springer Science & Business Media, 2004.
- GARDNER, J. W.; BARTLETT, P.; DODD, G.; SHURMER, H. Pattern recognition in the warwick electronic nose. In: *8th Int. Congress of the European Chemoreception Research Organisation*. [S.l.: s.n.], 1988. p. 9.
- GARDNER, J. W.; BARTLETT, P. N. A brief history of electronic noses. *Sensors and Actuators B: Chemical*, Elsevier, v. 18, n. 1-3, p. 210–211, 1994.
- GUOHUA, H.; YULING, W.; DANDAN, Y.; WENWEN, D.; LINSHAN, Z.; LVYE, W. Study of peach freshness predictive method based on electronic nose. *Food Control*, Elsevier, v. 28, n. 1, p. 25–32, 2012.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.
- HORVAT, R.; JR, G. C.; ROBERTSON, J.; MEREDITH, F.; SCORZA, R.; CALLAHAN, A.; MORGENS, P. Comparison of the volatile compounds from several commercial peach cultivars. *Journal of Agricultural and Food Chemistry*, ACS Publications, v. 38, n. 1, p. 234–237, 1990.
- HUANG, C.-H.; ZENG, C.; WANG, Y.-C.; PENG, H.-Y.; LIN, C.-S.; CHANG, C.-J.; YANG, H.-Y. A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 9, p. 2845, 2018.
- HUANG, G.-B.; ZHOU, H.; DING, X.; ZHANG, R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, IEEE, v. 42, n. 2, p. 513–529, 2012.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: *IEEE. Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. [S.l.], 2004. v. 2, p. 985–990.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: theory and applications. *Neurocomputing*, Elsevier, v. 70, n. 1-3, p. 489–501, 2006.
- HUSAIN, A. M.; RINI, T. H.; HAQUE, M. I.; ALAM, M. R. Air quality monitoring with arduino and android. 2015.
- IKEGAMI, A.; KANEYASU, M. Olfactory detection using integrated sensors. *Digest of Technical Papers, Transducers*, v. 85, p. 136–139, 1985.
- IPARDES. *Caderno estatístico do município de Ponta Grossa*. [S.l.], 2019. Disponível em: <<http://www.ipardes.gov.br/cadernos/MontaCadPdf1.php?Municipio=84000>>. Acesso em: 07 fev. 2019.
- JIANG, S.; WANG, J. Internal quality detection of chinese pecans (*carya cathayensis*) during storage using electronic nose responses combined with physicochemical methods. *Postharvest Biology and Technology*, Elsevier, v. 118, p. 17–25, 2016.

- JUNIOR, L. C. C.; DURIGAN, M. F. B.; MATTIUZ, B.-H.; MARTINS, R. N.; DURIGAN, J. F. Caracterização da curva de maturação de pêssegos 'aurora-1', na região de jaboticabal-sp. *Revista Brasileira de Fruticultura*, Sociedade Brasileira de Fruticultura, p. 661–665, 2007.
- KANEYASU, M.; IKEGAMI, A.; ARIMA, H.; IWANAGA, S. Smell identification using a thick-film hybrid gas sensor. *IEEE transactions on components, hybrids, and manufacturing technology*, IEEE, v. 10, n. 2, p. 267–273, 1987.
- KESKE, C. Epidemiologia da podridão parda em pessegueiros conduzidos em sistema de produção orgânico no alto vale do itajaí-sc. 2009.
- KIM, J.; HWANGBO, H. Sensor-based optimization model for air quality improvement in home iot. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 4, p. 959, 2018.
- LANGFORD, J. The cross validation problem. In: SPRINGER. *International Conference on Computational Learning Theory*. [S.l.], 2005. p. 687–688.
- LAREF, R.; LOSSON, E.; SAVA, A.; SIADAT, M. Support vector machine regression for calibration transfer between electronic noses dedicated to air pollution monitoring. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 11, p. 3716, 2018.
- LI, Q.; GU, Y.; WANG, N.-F. Application of random forest classifier by means of a qcm-based e-nose in the identification of chinese liquor flavors. *IEEE Sensors Journal*, IEEE, v. 17, n. 6, p. 1788–1794, 2017.
- LI, W.; LIU, H.; XIE, D.; HE, Z.; PI, X. Lung cancer screening based on type-different sensor arrays. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1969, 2017.
- LILIEN-KIPNIS, H.; LAVEE, S. Anatomical changes during the development of “ventura” peach fruitst. *Journal of Horticultural Science*, Taylor & Francis, v. 46, n. 1, p. 103–110, 1971.
- LIN, S.; ZHANG, X. A rapid and novel method for predicting nicotine alkaloids in tobacco through electronic nose and partial least-squares regression analysis. *Analytical Methods*, Royal Society of Chemistry, v. 8, n. 7, p. 1609–1617, 2016.
- LIU, H.; YU, L. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 17, n. 4, p. 491–502, 2005.
- LIU, X.; CHENG, S.; LIU, H.; HU, S.; ZHANG, D.; NING, H. A survey on gas sensing technology. *Sensors*, Molecular Diversity Preservation International, v. 12, n. 7, p. 9635–9665, 2012.
- LORWONGTRAGOOL, P.; SOWADE, E.; WATTHANAWISUTH, N.; BAUMANN, R. R.; KERDCHAROEN, T. A novel wearable electronic nose for healthcare based on flexible printed chemical sensor array. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 14, n. 10, p. 19700–19712, 2014.
- MARTÍNEZ, A. M.; KAK, A. C. Pca versus lda. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 2, p. 228–233, 2001.
- MEN, H.; JIAO, Y.; SHI, Y.; GONG, F.; CHEN, Y.; FANG, H.; LIU, J. Odor fingerprint analysis using feature mining method based on olfactory sensory evaluation. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 10, p. 3387, 2018.

- MEN, H.; SHI, Y.; JIAO, Y.; GONG, F.; LIU, J. Electronic nose sensors data feature mining: a synergetic strategy for the classification of beer. *Analytical Methods*, Royal Society of Chemistry, v. 10, n. 17, p. 2016–2025, 2018.
- MHAMDI, H.; MHAMDI, F. Feature selection methods on biological knowledge discovery and data mining: A survey. In: IEEE. *2014 25th International Workshop on Database and Expert Systems Applications*. [S.l.], 2014. p. 46–50.
- MONCRIEFF, R. An instrument for measuring and classifying odors. *Journal of applied physiology*, Am Physiological Soc, v. 16, n. 4, p. 742–749, 1961.
- MUCHERINO, A.; PAPAJORGJI, P.; PARDALOS, P. M. *Data mining in agriculture*. [S.l.]: Springer Science & Business Media, 2009.
- MUTALIB, N. A. A.; JASWIR, I.; AKMELIAWATI, R. Iium-fabricated portable electronic nose for halal authentication in beverages. In: IEEE. *Information and Communication Technology for the Muslim World (ICT4M), 2013 5th International Conference on*. [S.l.], 2013. p. 1–4.
- NURUL, A.; JASWIR, I.; AKMELIAWATI, R.; IBRAHIM, A.; ASLAM, M.; OCTAVIANI, F. et al. Rapid detection of ethanol in beverages using iium-fabricated electronic nose. *International Food Research Journal*, Faculty of Food Science and Technology, v. 24, n. Suppl., 2017.
- PERSAUD, K.; DODD, G. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature*, Springer, v. 299, n. 5881, p. 352–355, 1982.
- PETERSON, P. J.; AUJLA, A.; GRANT, K. H.; BRUNDLE, A. G.; THOMPSON, M. R.; HEY, J. V.; LEIGH, R. J. Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 17, n. 7, p. 1653, 2017.
- PINEDA, D. M.; PÉREZ, J. C. Senose: An under u\$50 electronic nose for the monitoring of soil gas emissions. *Computers and Electronics in Agriculture*, Elsevier, v. 133, p. 15–21, 2017.
- RAGAZZO-SANCHEZ, J.; CHALIER, P.; CHEVALIER, D.; CALDERON-SANTOYO, M.; GHOMMIDH, C. Identification of different alcoholic beverages by electronic nose coupled to gc. *Sensors and Actuators B: Chemical*, Elsevier, v. 134, n. 1, p. 43–48, 2008.
- RAJKUMAR, P.; WANG, N.; EIMASRY, G.; RAGHAVAN, G.; GARIEPY, Y. Studies on banana fruit quality and maturity stages using hyperspectral imaging. *Journal of Food Engineering*, Elsevier, v. 108, n. 1, p. 194–200, 2012.
- RUIZ, L. I.; BECCARO, W.; EVARISTO, B. G.; FERNANDEZ, F. J. R. Tactile sensing glove-based system for objects classification using support vector machine. *IEEE Latin America Transactions*, IEEE, v. 16, n. 6, p. 1658–1663, 2018.
- SALES, E.; SANTOS, J. d.; COSTA, A. d.; LISBOA, V.; MEDEIROS, A. d. Odores provenientes de graves problemas ambientais. In: *XX Congresso Brasileiro de Engenharia Química, Florianópolis*. [S.l.: s.n.], 2014.
- SANAEIFAR, A.; MOHTASEBI, S.; GHASEMI-VARNAMKHAHI, M.; AHMADI, H.; ROGADO, J. S. L. Development and application of a new low cost electronic nose for the ripeness monitoring of banana using computational techniques (pca, lda, simca, and svm). Czech Academy of Agricultural Sciences, 2014.

- SANTONICO, M.; PENNAZZA, G.; GRASSO, S.; D'AMICO, A.; BIZZARRI, M. Design and test of a biosensor-based multisensorial system: A proof of concept study. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 13, n. 12, p. 16625–16640, 2013.
- SANTOS, C. A. A. *Uso de quitosana e embalagem plástica associado à refrigeração na conservação da qualidade de pêssegos' Douradão'e'Aurora-1'*. 2007. 52 f. Dissertação (Mestrado) — Dissertação (Mestrado em Agronomia) - Instituto Agronômico de Campinas, Campinas, 2007.
- SANTOS, J. P.; LOZANO, J. Real time detection of beer defects with a hand held electronic nose. In: IEEE. *Electron Devices (CDE), 2015 10th Spanish Conference on*. [S.l.], 2015. p. 1–4.
- SANTOS, J. P.; LOZANO, J.; ALEIXANDRE, M. Electronic noses applications in beer technology. In: *Brewing Technology*. [S.l.]: InTech, 2017.
- SIYANG, S.; LORWONGTRAGOOL, P.; NOOSIDUM, A.; WONGCHOOSUK, C.; KERDCHAROEN, T. Development and application of electronic nose for agricultural robot. In: IEEE. *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on*. [S.l.], 2013. p. 1–4.
- SUDARMAJI, A.; KITAGAWA, A. Application of temperature modulation-sdp on mos gas sensors: Capturing soil gaseous profile for discrimination of soil under different nutrient addition. *Journal of Sensors*, Hindawi Publishing Corporation, v. 2016, 2016.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to data mining. 1st*. [S.l.]: Boston: Pearson Addison Wesley. xxi, 2005.
- VISAI, C.; VANOLI, M. Volatile compound production during growth and ripening of peaches and nectarines. *Scientia Horticulturae*, Elsevier, v. 70, n. 1, p. 15–24, 1997.
- VOSS, A.; WITT, K.; FISCHER, C.; REULECKE, S.; POITZ, W.; KECHAGIAS, V.; SURBER, R.; FIGULLA, H. R. Smelling heart failure from human skin odor with an electronic nose. In: IEEE. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.], 2012. p. 4034–4037.
- WANG, S.; TANG, J.; LIU, H. Feature selection. *Encyclopedia of Machine Learning and Data Mining*, Springer, p. 1–9, 2016.
- WEI, Z.; XIAO, X.; WANG, J.; WANG, H. Identification of the rice wines with different marked ages by electronic nose coupled with smartphone and cloud storage platform. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 17, n. 11, p. 2500, 2017.
- WILKENS, W. F.; HARTMAN, J. D. An electronic analog for the olfactory processes. *Journal of Food Science*, Wiley Online Library, v. 29, n. 3, p. 372–378, 1964.
- WILSON, A. D. Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 13, n. 2, p. 2295–2348, 2013.
- WILSON, A. D.; BAIETTO, M. Applications and advances in electronic-nose technologies. *Sensors*, Molecular Diversity Preservation International, v. 9, n. 7, p. 5099–5148, 2009.
- WOJNOWSKI, W.; MAJCHRZAK, T.; DYMERSKI, T.; GĘBICKI, J.; NAMIEŚNIK, J. Electronic noses: Powerful tools in meat quality assessment. *Meat Science*, Elsevier, 2017.

- WOJNOWSKI, W.; MAJCHRZAK, T.; DYMERSKI, T.; GĘBICKI, J.; NAMIEŚNIK, J. Portable electronic nose based on electrochemical sensors for food quality assessment. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 17, n. 12, p. 2715, 2017.
- WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 2, n. 1-3, p. 37–52, 1987.
- WU, H.; YUE, T.; XU, Z.; ZHANG, C. Sensor array optimization and discrimination of apple juices according to variety by an electronic nose. *Analytical Methods*, Royal Society of Chemistry, v. 9, n. 6, p. 921–928, 2017.
- ZAKI, M. J.; JR, W. M.; MEIRA, W. *Data mining and analysis: fundamental concepts and algorithms*. [S.l.]: Cambridge University Press, 2014.
- ZHANG, H.; WANG, J.; YE, S.; CHANG, M. Application of electronic nose and statistical analysis to predict quality indices of peach. *Food and Bioprocess Technology*, Springer, v. 5, n. 1, p. 65–72, 2012.
- ZHU, J.; CHEN, F.; WANG, L.; NIU, Y.; XIAO, Z. Evaluation of the synergism among volatile compounds in oolong tea infusion by odour threshold with sensory analysis and e-nose. *Food chemistry*, Elsevier, v. 221, p. 1484–1490, 2017.

Apêndices

APÊNDICE A – CARACTERÍSTICAS DOS SENSORES UTILIZADOS NESTE TRABALHO

Tabela 20 – Características dos sensores da família MQ e TGS utilizados neste trabalho

Sensor	Parâmetro	Valor
MQ-2	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 160\text{mA}$
	Resistência de Detecção	$3\text{K}\Omega\text{-}30\text{K}\Omega$
	Resistência de Carregamento	$5\text{K}\Omega\text{-}47\text{K}\Omega$
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	H ₂ , GLP, CH ₄ , CO, álcool, propano, butano e metano
MQ-3	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 150\text{mA}$
	Resistência de Detecção	$1\text{M}\Omega\text{-}8\text{M}\Omega$
	Resistência de Carregamento	$100\text{K}\Omega\text{-}470\text{K}\Omega$
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	Álcool, benzeno, CH ₄ , hexano, GLP e CO
MQ-4	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 150\text{mA}$
	Resistência de Detecção	$10\text{K}\Omega\text{-}60\text{K}\Omega$
	Resistência de Carregamento	$10\text{K}\Omega\text{-}47\text{K}\Omega$
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	GLP, CH ₄ , H ₂ , CO e álcool
MQ-5	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 160\text{mA}$
	Resistência de Detecção	$10\text{K}\Omega\text{-}60\text{K}\Omega$
	Resistência de Carregamento	$10\text{K}\Omega\text{-}47\text{K}\Omega$
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	H ₂ , GLP, CH ₄ , CO, álcool, iso-butano e propano
MQ-6	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 150\text{mA}$
	Resistência de Detecção	$10\text{K}\Omega\text{-}60\text{K}\Omega$
	Resistência de Carregamento	$10\text{K}\Omega\text{-}47\text{K}\Omega$
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	GLP, H ₂ , CH ₄ , CO, álcool, iso-butano e propano
MQ-7	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 70\text{mA}$
	Resistência de Detecção	$2\text{K}\Omega\text{-}20\text{K}\Omega$
	Resistência de Carregamento	$5\text{K}\Omega\text{-}47\text{K}\Omega$
	Tempo de Pré Aquecimento	48 horas
	Sensível aos gases	CO, H ₂ , GLP, CH ₄ e álcool

Sensor	Parâmetro	Valor
MQ-8	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 160\text{mA}$
	Resistência de Detecção	10K Ω -60K Ω
	Resistência de Carregamento	5K Ω -33K Ω
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	H ₂ , GLP, CH ₄ , CO e álcool
MQ-9	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 68\text{mA}$
	Resistência de Detecção	2K Ω -20K Ω
	Resistência de Carregamento	10K Ω -47K Ω
	Tempo de Pré-Aquecimento	48 horas
	Sensível aos gases	CO, CH ₄ e GLP
MQ-135	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 160\text{mA}$
	Resistência de Detecção	30K Ω -200K Ω
	Resistência de Carregamento	10K Ω -47K Ω
	Tempo de Pré-Aquecimento	24 horas
	Sensível aos gases	NH ₃ , benzeno, álcool, CO ₂ , CO e NH ₄
TGS 822	Tensão de Operação	$\leq 24\text{V}$
	Corrente de Aquecimento	$\leq 132\text{mA}$
	Resistência de Detecção	1K Ω -10K Ω
	Resistência de Carregamento	$\geq 0,45\text{K}\Omega$
	Tempo de Pré-Aquecimento	7 dias
	Sensível aos gases	Acetona, n-hexano, benzeno, álcool, iso-butano, CO e metano
TGS 2600	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 46\text{mA}$
	Resistência de Detecção	10K Ω -90K Ω
	Resistência de Carregamento	$\geq 0,45\text{K}\Omega$
	Tempo de Pré-Aquecimento	7 dias
	Sensível aos gases	H ₂ , CO, metano, iso-butano e álcool
TGS 2602	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 61\text{mA}$
	Resistência de Detecção	10K Ω -100K Ω
	Resistência de Carregamento	$\geq 0,45\text{K}\Omega$
	Tempo de Pré-Aquecimento	7 dias
	Sensível aos gases	H ₂ , NH ₃ , álcool, H ₂ S e tolueno
TGS 2603	Tensão de Operação	5V
	Corrente de Aquecimento	$\leq 48\text{mA}$
	Resistência de Detecção	20K Ω -200K Ω
	Resistência de Carregamento	$\geq 0,45\text{K}\Omega$
	Tempo de Pré-Aquecimento	96 horas
	Sensível aos gases	H ₂ , H ₂ S, etanol, metil mercaptano e trimetilamina

Tabela 21 – Características do sensor de temperatura e pressão atmosférica BMP180

Parâmetro	Condição	Mínimo	Típico	Máximo	Unidade
Temperatura de operação	Operational	-40		85	°C
Tensão de operação	Ondulação máxima	1,8	2,5	3,6	V
		1,62	2,5	3,6	V
Corrente de alimentação	Baixo consumo de energia		3		mV
	Padrão		5		uA
	Resolução alta		7		uA
	Resolução ultra alta		12		uA
	Resolução avançada		32		uA
Corrente de pico	Durante conversão		650	1000	uA
Corrente de <i>standby</i>	25°C		0,1	4	uA
Precisão da pressão relativa (V _{DD} =3,3V)	700 - 900hPa		±0,12		hPa
	25 - 40 °C		±1,0		m
Precisão da pressão absoluta (V _{DD} =3,3V)	300 - 1100 hPa	-6,0	-1,0	+4,5	hPa
	0 - +65 °C				
Precisão da temperatura absoluta (V _{DD} =3,3V)	0 - +65 °C	-2,0	±1,0	+2,0	°C
Resolução dos dados de saída	Pressão		0,01		hPa
	Temperatura		0,1		°C

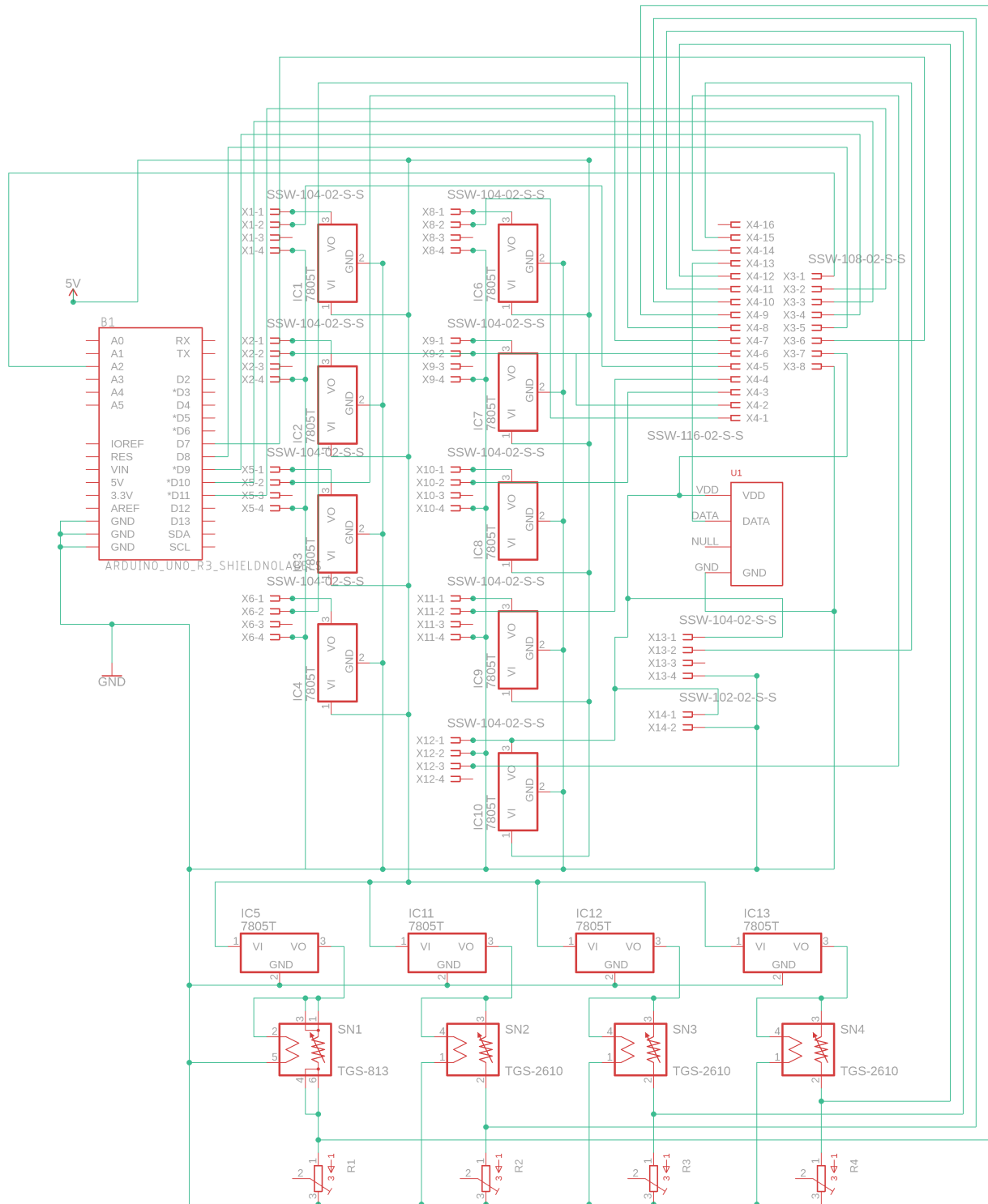
Tabela 22 – Características do sensor analógico de umidade do ar HIH-4030

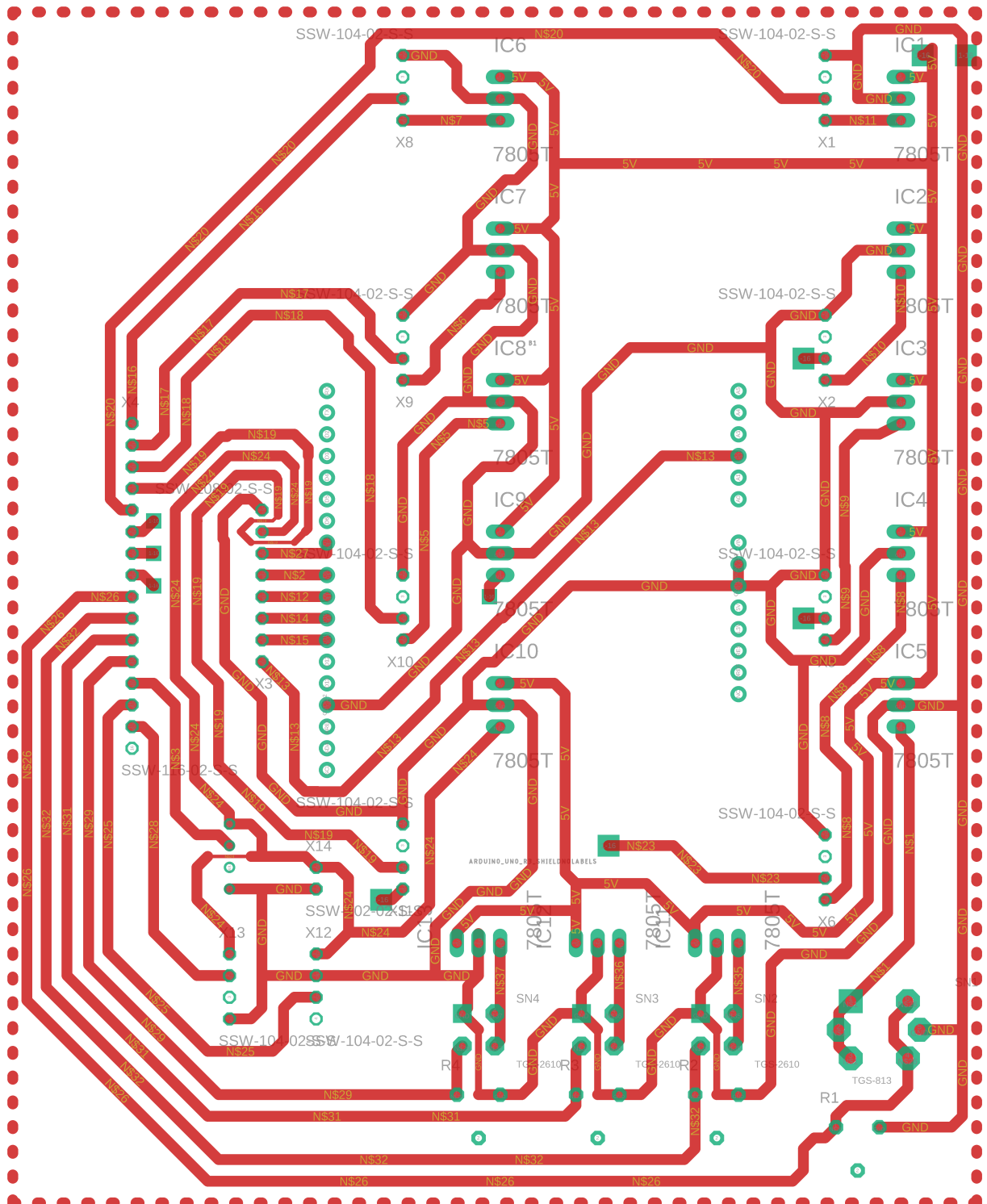
Parâmetro	Condição	Mínimo	Típico	Máximo	Unidade
Tensão de operação		4		5,8	V
Corrente de alimentação			200	500	uA
Temperatura de operação		-40		+85	°C
Umidade de operação		0		100	%RH
Intercambiabilidade	0% a 59% RH	-5		5	%RH
	60% a 100% RH	-8		8	%RH
Precisão		-3,5		+3,5	%RH
Histerese			3		%RH
Repetibilidade			±0,5		%RH
Tempo de configuração				70	ms
Tempo de resposta (1/e)			5		s
Estabilidade			±0,5		%RH

Tabela 23 – Características do microcontrolador Atmel ATmega2560

Parâmetro	Valor	Unidade
Tensão de operação	5	V
Tensão de entrada (recomendado)	7 - 12	V
Tensão de entrada (limite)	6 - 20	V
Pinos digitais de I/O	54 (sendo 14 de saída PWM)	
Entradas analógicas	16	
Corrente de saída por pino de I/O	40	mA
Corrente de saída do pino 3,3V	50	mA
Memória flash	256 (8 utilizado pelo gerenciador de boot)	KB
SRAM	8	KB
EEPROM	4	KB
Velocidade de clock	16	MHz
Conversor A/D	10	bits

APÊNDICE B – DIAGRAMA ESQUEMÁTICO DO CIRCUITO DO NARIZ ELETRÔNICO





APÊNDICE D – RESUMO DOS DADOS OBTIDOS NOS EXPERIMENTOS

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
3,5268	13,8061	9,4558	204,7660	0,0000	2,5086	16,4533	22,9173	0,1589	93,4932	252,3120	68,4678	0,0000	52,9128	83,5250	Estagio 1
6,8733	10,6468	9,6932	205,3035	0,0000	3,0793	15,4105	21,9907	0,0536	91,6362	249,3362	64,8697	0,0000	52,2225	82,8930	Estagio 1
6,4207	10,9702	9,9195	203,6843	0,0000	4,8178	16,6204	22,5733	0,2246	82,5285	248,0302	63,9697	0,0000	51,5199	81,8904	Estagio 1
5,0500	10,7325	8,9418	202,6679	0,0000	1,3316	15,4605	20,5039	0,1537	81,6838	244,6207	59,6503	0,0000	50,0613	81,3665	Estagio 1
6,4498	10,0450	8,7845	202,5253	0,0000	2,9895	15,8910	20,9282	0,1340	76,1415	243,7695	64,8077	0,0000	50,1897	80,9371	Estagio 1
4,0980	10,4514	7,3773	204,1185	0,0000	1,7397	14,0633	21,6147	0,0000	84,7357	243,0400	66,9263	0,0000	50,6281	81,6083	Estagio 1
4,9729	8,8585	8,7277	204,4151	0,0000	0,9523	15,0590	20,6957	0,0200	85,8278	247,0042	67,9700	0,0000	51,1156	82,4883	Estagio 1
5,2270	8,7465	9,1692	203,5625	0,0000	3,7215	15,0565	21,2150	0,0740	74,7190	244,5278	62,0238	0,0000	49,8494	81,2585	Estagio 1
4,8949	8,7713	9,0314	204,1092	0,0000	1,8945	15,6050	20,1777	0,0380	78,6122	245,9675	62,3248	0,0000	50,1097	81,8631	Estagio 1
7,3572	9,1975	7,8898	204,8357	0,0000	2,2479	15,3991	20,3282	0,1278	80,0978	244,6840	64,2783	0,0000	50,4956	81,6926	Estagio 1
5,3127	8,5735	9,4970	204,4459	0,0000	3,6727	15,6214	20,3921	0,0813	81,5168	246,0349	61,0792	0,0000	50,4790	81,8772	Estagio 1
7,4480	9,6706	8,5927	203,8029	0,0000	1,9545	13,8935	19,9946	0,1435	78,7207	245,0186	61,0323	0,0000	50,0209	81,5393	Estagio 1
6,2025	8,2402	8,8663	202,8132	0,0000	2,0605	14,8219	19,4981	0,0962	78,1693	244,8838	61,0444	0,0000	49,7459	81,4231	Estagio 1
6,6193	8,3467	9,2944	202,9247	0,0000	1,7398	14,5544	20,6770	0,0603	79,4033	243,4492	60,2489	0,0000	49,7937	81,1330	Estagio 1
3,8648	8,1718	9,1660	203,7280	0,0000	1,7180	14,9124	20,3747	0,1969	81,7140	245,8128	60,3997	0,0000	50,0045	81,9322	Estagio 1
0,3438	12,6392	13,5202	123,3310	2,5621	3,3868	7,3738	13,7274	0,0000	107,2401	218,4912	123,4788	9,7132	48,9083	70,2958	Estagio 1
0,2247	7,0303	5,7988	202,3408	0,0475	0,0000	10,1703	8,9775	0,0000	56,2983	210,8560	55,9800	0,0000	42,9019	75,3022	Estagio 1
0,2402	6,7318	5,8638	204,2113	0,0000	0,0000	10,6243	9,1237	0,0000	56,2448	210,5915	55,8435	0,0000	43,0365	75,5706	Estagio 1
0,2052	6,8947	5,6080	203,5700	0,0000	0,0136	10,4614	9,0953	0,0000	55,4945	210,1784	55,7383	0,0000	42,8661	75,3800	Estagio 1
0,0702	6,9187	6,1679	170,9482	0,0000	1,6243	8,4198	9,4576	0,0000	56,8612	175,0197	55,5863	0,0000	37,7749	63,1644	Estagio 1
0,4298	6,2755	6,1707	203,7649	0,0000	0,0000	7,4265	8,8725	0,0000	51,3412	206,8615	52,1323	0,0000	41,7904	74,8183	Estagio 1
0,5952	5,9999	5,8318	204,0069	0,0000	0,0000	8,8220	7,8345	0,0000	51,6923	207,3055	52,3210	0,0000	41,8776	74,9527	Estagio 1
0,4447	5,0252	5,7558	203,4440	0,0000	0,0000	7,8690	8,1007	0,0000	51,0093	208,2148	48,6122	0,0000	41,4212	75,0499	Estagio 1
0,3060	4,7100	4,8085	204,1318	0,0000	0,0000	8,5022	7,2663	0,0000	47,3120	203,2170	47,5402	0,0000	40,5995	74,2786	Estagio 1
1,2975	4,5423	4,6758	203,2902	0,0000	0,1373	7,6668	6,6713	0,0000	44,9433	181,5124	44,3595	0,0000	38,3920	70,2658	Estagio 1
0,5393	3,6560	4,8323	203,0028	0,0000	0,0000	7,5865	6,2505	0,0000	41,4613	195,3217	44,4655	0,0000	39,0089	72,6914	Estagio 1
0,6046	3,2352	4,1772	201,5438	0,0000	0,0000	7,5798	5,8195	0,0000	38,8968	194,4017	43,4395	0,0000	38,4383	72,2979	Estagio 1
1,6081	3,5332	4,3215	203,0893	0,0000	0,0000	7,4277	5,7873	0,0000	43,0492	195,0083	45,3843	0,0000	39,1699	72,6655	Estagio 1
0,4674	2,4798	4,0735	203,2085	0,0000	0,0000	8,3887	6,0405	0,0000	43,8220	193,2398	44,2365	0,0000	38,9197	72,4285	Estagio 1

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
0,5042	2,8755	4,0462	202,7778	0,0000	0,0000	7,7110	6,0455	0,0000	39,7438	190,4958	42,8408	0,0000	38,2339	71,8445	Estagio 1
5,9235	30,3424	20,9533	213,1425	0,0000	7,3433	23,8185	28,1511	0,0363	236,1961	414,4962	209,8118	37,1287	94,4111	130,8175	Estagio 1
6,4253	25,3239	18,7142	208,6270	0,0000	7,0203	22,0230	25,4363	0,0467	195,1868	389,5957	183,6254	31,9162	85,6878	120,7297	Estagio 1
8,9110	22,5003	16,0428	210,1381	0,0000	3,4136	20,6297	23,0003	0,0000	177,2490	368,9338	170,0172	26,0367	80,5287	115,0280	Estagio 1
8,0410	18,5213	13,6630	210,6880	0,0000	4,4028	17,6350	20,3304	0,0000	151,1162	346,3560	149,6297	10,4131	73,1382	108,5833	Estagio 1
3,9785	17,5427	13,1958	210,1138	0,0000	4,8008	17,2168	19,4637	0,0387	145,7063	336,6145	138,9863	6,7512	70,3392	106,0375	Estagio 1
5,0538	16,8402	12,1272	208,8557	0,0000	2,0235	16,2294	19,3568	0,0000	141,6972	327,9884	129,5866	6,7147	68,1903	103,5867	Estagio 1
6,4208	16,4593	11,9930	208,6166	0,0000	3,0526	15,0658	18,1119	0,0578	131,1793	321,5178	121,3865	3,9174	65,9830	101,3550	Estagio 1
4,0123	15,1038	11,4903	208,3007	0,0000	2,1109	15,6150	17,4255	0,0417	120,1945	314,9404	118,2220	2,4644	63,8401	99,5623	Estagio 1
5,9153	13,2955	11,5559	207,5288	0,0000	3,0248	15,4738	17,5982	0,0070	112,6069	307,3182	112,3225	1,6663	62,1779	97,2353	Estagio 1
4,4211	14,3641	11,2418	207,3703	0,0000	3,1003	15,9342	17,6740	0,0501	117,0979	303,3828	109,7355	1,6670	62,0030	96,5019	Estagio 1
7,3338	13,6957	11,3969	206,5467	0,0000	2,7872	16,0017	18,7909	0,2355	113,8578	306,7596	115,8971	3,4679	62,8285	96,9718	Estagio 1
7,1447	12,2530	10,3402	203,2076	0,0000	2,1569	14,9434	18,7463	0,0106	107,1630	299,8075	110,4273	2,5967	60,6767	94,8194	Estagio 1
5,5827	11,6068	10,5850	204,6641	0,0000	3,9678	14,8830	18,1745	0,0230	100,7962	299,1559	101,1450	0,1377	59,2863	94,3946	Estagio 1
4,2595	11,1912	10,0644	205,4933	0,0000	3,1992	14,3455	18,4555	0,1251	102,3945	298,7736	94,5123	0,0163	58,6793	94,4058	Estagio 1
4,4437	9,7518	9,0562	204,8833	0,0000	2,8296	15,0885	16,8359	0,0885	97,0379	290,6562	89,2847	0,0000	56,9197	92,3960	Estagio 1
3,7644	35,0073	30,0005	228,9506	0,0753	3,3112	30,0465	31,4138	0,0000	278,9375	461,8732	175,7675	27,3765	100,5019	143,7380	Estagio 2
4,0237	29,0897	25,0994	224,7587	0,0000	1,3455	26,1104	25,2393	0,0000	227,5958	420,3216	143,3411	9,7810	87,4389	130,0799	Estagio 2
6,7355	25,6953	21,8788	221,4464	0,0000	0,8737	24,0293	22,8192	0,0000	193,4769	394,0109	115,0735	0,7299	78,9823	121,1560	Estagio 2
5,4985	23,6089	20,0182	219,2305	0,0000	2,0981	22,3767	21,3158	0,0000	172,6389	379,7812	108,5165	0,2647	75,0268	116,4211	Estagio 2
6,2176	19,7103	17,6040	216,7060	0,0000	1,0831	19,6864	19,7461	0,0000	158,3762	362,5222	90,1027	0,3343	70,1607	111,4706	Estagio 2
5,0343	18,0344	13,9040	213,8271	0,0000	0,8731	16,4432	15,2512	0,0000	127,9721	314,0052	75,9606	0,0000	61,6389	99,2154	Estagio 2
6,2245	15,6695	15,1362	215,3843	0,0000	2,5278	17,5520	15,5929	0,0000	150,8208	340,1565	89,3400	0,0000	66,8004	106,4138	Estagio 2
6,3988	15,5547	15,2808	215,4512	0,0000	2,6970	17,8920	15,4820	0,0000	151,7138	340,4727	89,9743	0,0000	66,9936	106,5328	Estagio 2
6,3987	15,5965	15,2817	215,4223	0,0000	2,7225	17,7465	15,8712	0,0000	151,7423	341,0852	90,0263	0,0000	67,0687	106,6504	Estagio 2
7,0117	15,5627	15,2463	215,3947	0,0000	2,8483	17,8720	15,5363	0,0000	151,6783	341,1315	89,6675	0,0000	67,0730	106,6226	Estagio 2
7,5868	15,6077	15,3227	215,4047	0,0000	2,6852	18,1812	15,7273	0,0000	151,5607	341,3890	89,8313	0,0000	67,1767	106,6309	Estagio 2
7,6847	15,5880	15,2952	215,4163	0,0000	2,6280	17,9037	15,9453	0,0000	152,0802	340,8760	89,9208	0,0000	67,1799	106,5605	Estagio 2
6,5628	15,5728	15,4030	215,4420	0,0000	2,7630	17,7310	15,7713	0,0000	151,5827	341,1033	89,7110	0,0000	67,0495	106,6312	Estagio 2
6,6732	15,5150	15,3691	215,3200	0,0000	2,8193	17,8798	15,8265	0,0000	151,7933	341,0882	90,0697	0,0000	67,1042	106,6218	Estagio 2
7,0750	15,6020	15,4345	215,3953	0,0000	2,7395	17,8028	15,7300	0,0000	151,7090	340,7458	89,9793	0,0000	67,0933	106,5358	Estagio 2
7,0864	34,1655	29,4973	225,0092	1,1953	6,6563	29,6201	28,2047	0,0000	215,8473	400,9429	177,5825	19,8711	90,4368	124,9621	Estagio 2
6,1294	29,8375	26,1175	222,0747	0,7362	5,3750	27,0570	25,8732	0,0000	182,0901	377,3013	158,4126	12,4715	82,5751	117,0751	Estagio 2
7,2272	25,6800	22,2883	220,1148	0,1336	5,2395	22,9687	22,5730	0,0000	141,3191	349,7296	141,6603	4,5836	74,1167	108,6590	Estagio 2

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
7,2535	23,2640	20,4754	218,9521	0,0397	3,2952	22,9320	21,7627	0,0000	130,9986	334,3826	131,6829	1,2371	70,4827	104,7601	Estagio 2
4,9665	21,1231	18,8333	218,1310	0,0000	3,1383	21,4243	21,1690	0,0000	120,8444	325,2279	121,8308	0,0000	67,4376	102,2339	Estagio 2
5,0898	20,8722	18,7865	218,1415	0,0000	3,1547	21,1255	21,2327	0,0000	120,5390	324,7608	121,7338	0,0000	67,3413	102,1325	Estagio 2
4,8553	20,8875	18,8068	218,0828	0,0000	2,8732	21,4085	21,2512	0,0000	119,8850	324,9550	121,5917	0,0000	67,2767	102,1456	Estagio 2
5,1618	21,1675	18,7435	218,1940	0,0000	2,9747	21,2285	21,1648	0,0000	120,4545	325,2022	122,0412	0,0000	67,4102	102,2367	Estagio 2
4,9240	21,0937	18,7748	218,1650	0,0000	3,1938	21,2647	21,2220	0,0000	120,7665	325,1503	121,7940	0,0000	67,4114	102,2234	Estagio 2
4,8537	21,0575	18,7985	218,1290	0,0000	3,1150	21,1917	21,2007	0,0000	120,4242	324,6752	121,7477	0,0000	67,3225	102,1140	Estagio 2
4,7088	20,8815	18,7498	218,2477	0,0000	3,0593	21,1995	21,1108	0,0000	120,7323	324,6575	121,7767	0,0000	67,3172	102,1615	Estagio 2
4,9655	20,8568	18,7842	218,2473	0,0000	3,1430	21,3320	21,2637	0,0000	120,4203	324,7052	121,5648	0,0000	67,3294	102,1199	Estagio 2
4,9835	21,1022	18,8548	218,2615	0,0000	3,0558	21,4115	21,2637	0,0000	120,8783	325,1985	121,8647	0,0000	67,4519	102,2470	Estagio 2
4,9982	21,0435	18,7143	218,1668	0,0000	2,7685	21,3315	21,1380	0,0000	121,0733	324,9563	121,5443	0,0000	67,3642	102,2087	Estagio 2
4,8875	20,9970	18,7893	218,1775	0,0000	2,9495	21,4435	21,1703	0,0000	121,0045	324,9890	121,7323	0,0000	67,3954	102,2117	Estagio 2
4,8303	33,2589	28,1329	221,2513	0,0108	4,9140	26,6035	26,9475	0,0000	153,1903	387,4011	122,1328	0,2832	77,6120	116,3641	Estagio 2
4,9327	29,1547	24,5348	216,4422	0,0000	3,5926	25,1788	24,2580	0,0000	125,0055	363,1322	105,9473	0,0000	70,9368	109,1095	Estagio 2
4,8626	25,2400	22,1173	213,9137	0,0000	3,3702	22,6208	21,9680	0,0000	101,1075	346,0759	93,6439	0,0000	65,7631	104,2644	Estagio 2
4,1157	22,5003	19,3625	212,0740	0,0000	2,2135	21,2067	20,4032	0,0000	86,7041	334,3817	84,7906	0,0000	62,1348	101,2723	Estagio 2
2,6929	20,7333	18,8368	211,5575	0,0000	1,5500	20,6635	20,0750	0,0000	81,3878	328,6523	81,8010	0,0000	60,6115	99,9791	Estagio 2
2,7223	20,6912	18,8993	211,5777	0,0000	1,6647	20,7090	20,1190	0,0000	81,5108	328,3707	82,0310	0,0000	60,6381	99,9141	Estagio 2
2,6252	20,9557	18,9005	211,5478	0,0000	1,8542	20,7485	20,1678	0,0000	81,4033	328,3332	81,8930	0,0000	60,6484	99,8812	Estagio 2
2,5873	20,9115	18,8478	211,5168	0,0000	1,5255	20,5897	19,9882	0,0000	81,0305	328,5412	81,7387	0,0000	60,5598	99,9473	Estagio 2
2,4558	20,9375	18,8992	211,5757	0,0000	1,5172	20,7308	20,1448	0,0000	81,4195	328,3460	81,8428	0,0000	60,6053	99,9138	Estagio 2
2,6893	20,8045	18,8723	211,5317	0,0000	1,7705	20,7300	20,1543	0,0000	81,8997	328,6022	81,9322	0,0000	60,6913	99,9567	Estagio 2
2,7085	20,7523	18,8687	211,5765	0,0000	1,7680	20,7207	20,0345	0,0000	81,6912	328,6300	81,8085	0,0000	60,6584	99,9681	Estagio 2
2,6587	20,8258	18,8383	211,4833	0,0000	1,7102	20,6430	20,1827	0,0000	81,7205	328,6020	81,8150	0,0000	60,6523	99,9522	Estagio 2
3,0253	20,9622	18,8908	211,5665	0,0000	1,6182	20,7637	20,1240	0,0000	81,6520	328,3995	81,7363	0,0000	60,6722	99,8933	Estagio 2
2,5720	20,7657	18,9035	211,5912	0,0000	1,4368	20,7787	20,1647	0,0000	81,4948	328,5017	81,8630	0,0000	60,6209	99,9538	Estagio 2
2,6383	20,7777	18,9000	211,5595	0,0000	1,6153	20,7725	20,0318	0,0000	81,7570	328,5408	81,6410	0,0000	60,6334	99,9516	Estagio 2
9,5555	40,4457	50,1140	218,9608	1,2478	4,2292	26,2982	34,2265	0,0000	160,9615	397,9522	144,7835	1,9719	83,9036	118,2469	Estagio 3
10,6943	35,9262	43,5901	214,3782	1,0268	5,4575	23,7784	30,8087	0,0000	133,0977	375,8426	127,3247	0,0087	77,0718	111,3448	Estagio 3
9,7655	31,2936	38,6205	212,1958	0,6755	5,7558	22,1375	28,2915	0,0000	111,3689	357,8995	112,2683	0,0000	71,5594	106,2414	Estagio 3
9,1875	27,6139	33,9490	209,9061	0,1654	5,5373	20,9680	27,1315	0,0000	91,4051	342,6116	100,6449	0,0000	66,8554	102,0735	Estagio 3
12,8373	26,4992	32,2967	209,0816	0,3861	4,5469	21,2305	28,6503	0,0000	94,0504	346,4528	101,9490	0,0000	67,5370	102,8308	Estagio 3
9,3665	23,4640	30,0925	207,8145	0,5957	4,6675	20,7348	25,9238	0,0000	83,6832	333,0310	97,5743	0,0000	64,3806	99,7517	Estagio 3
9,5750	21,2918	27,8140	206,8618	0,0431	4,8852	19,4167	25,4633	0,0000	76,0560	326,5703	92,1774	0,0000	62,3196	98,1498	Estagio 3

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
8,8713	19,2667	26,5502	206,5294	0,1200	4,5623	19,4276	25,9882	0,0000	70,1447	319,3966	86,0979	0,0000	60,5350	96,4330	Estagio 3
10,3517	17,9884	25,0411	205,4453	0,0588	5,2910	19,4481	24,0922	0,0000	61,7500	315,1140	74,9388	0,0000	58,4246	95,1330	Estagio 3
10,5020	16,6966	22,8691	203,1105	0,0790	4,4307	18,6872	24,4272	0,0000	65,3554	313,7986	81,2739	0,0000	58,5562	94,8258	Estagio 3
7,5148	15,2481	21,7873	204,0131	0,0000	3,7328	17,3534	23,4035	0,0000	66,9745	310,2058	80,0805	0,0000	57,7165	94,4583	Estagio 3
7,1022	14,8020	20,9562	202,0732	0,0000	5,5472	17,5655	21,8154	0,0000	56,4676	301,6215	71,6923	0,0000	55,3572	92,1159	Estagio 3
11,4633	13,5325	18,7998	202,9547	0,0000	3,4390	18,2982	23,0545	0,0000	58,4865	304,7007	79,5067	0,0000	56,4797	93,0286	Estagio 3
8,4342	13,0166	18,8246	202,1108	0,0000	4,2959	17,9493	23,5027	0,0000	63,5664	309,2834	79,7985	0,0000	56,9833	94,0697	Estagio 3
8,9252	12,5906	18,5431	200,8960	0,0000	5,2635	18,2672	24,0252	0,0000	63,2608	305,0384	80,6660	0,0000	56,7289	92,9162	Estagio 3
13,7328	18,9222	22,0008	203,8670	0,1359	5,3840	33,4612	25,6598	0,0000	187,0186	370,0201	115,2884	0,2388	76,5946	113,0318	Estagio 3
9,4098	16,7313	19,7510	197,8150	0,0000	3,5736	30,3623	22,6859	0,0000	160,5053	345,9815	98,2817	0,0000	69,6229	105,4544	Estagio 3
10,2343	14,1810	17,6803	195,3333	0,0000	3,1723	26,7295	19,9160	0,0000	144,3370	327,2001	87,7485	0,0000	65,1179	100,1512	Estagio 3
10,5912	12,3572	16,1420	193,7502	0,0000	3,3142	25,9052	18,1202	0,0000	126,9928	309,3348	81,1056	0,0000	61,3549	95,1012	Estagio 3
9,5132	11,6020	15,6663	186,9308	0,0000	3,2990	25,3408	17,4820	0,0000	120,8351	302,5117	76,0710	0,0000	59,1732	92,5358	Estagio 3
10,8125	10,9929	14,8515	57,0381	0,0000	2,2778	24,0722	16,5817	0,0000	115,2413	296,9303	71,4478	0,0000	47,7112	82,4275	Estagio 3
12,4090	9,1289	13,9458	24,1203	0,0000	1,4571	23,0865	15,4090	0,0000	111,9006	288,7650	67,2657	0,0000	43,6529	80,3354	Estagio 3
8,4427	9,5100	13,3530	22,7233	0,0000	2,2253	22,7290	15,2585	0,0000	109,5958	287,7011	65,1908	0,0000	42,8254	80,0008	Estagio 3
9,9850	9,1923	13,0402	22,0988	0,0000	1,3871	22,0582	15,3325	0,0000	107,0057	286,5916	65,6414	0,0000	42,4871	79,5785	Estagio 3
9,2958	8,8388	13,0450	21,3934	0,0000	2,3126	22,4492	14,9980	0,0000	104,6170	284,4825	64,9520	0,0000	42,0296	78,8755	Estagio 3
8,7746	7,9133	11,9137	20,4687	0,0000	1,8158	20,1033	14,8390	0,0000	103,0548	277,6467	60,8824	0,0000	40,5702	77,1068	Estagio 3
10,5283	7,4943	11,6138	20,0522	0,0000	2,0458	19,9501	15,3428	0,0000	102,5247	275,3805	59,8066	0,0000	40,3646	76,4217	Estagio 3
8,7413	7,3207	11,3460	19,0088	0,0000	2,1398	19,9622	14,4426	0,0000	99,6607	270,4127	58,7318	0,0000	39,3667	75,0499	Estagio 3
7,1748	6,3460	10,9958	18,2265	0,0000	2,4178	19,5750	14,9130	0,0000	96,7065	268,2065	56,3371	0,0000	38,5307	74,3399	Estagio 3
12,0632	6,0622	10,9368	18,0353	0,0000	2,3917	19,4385	14,2542	0,0000	96,5238	266,8418	55,4037	0,0000	38,6116	73,8375	Estagio 3
17,6612	32,0728	23,7310	38,3842	0,0988	4,3807	30,7440	33,1664	1,0750	255,3402	468,1790	238,5640	52,5038	91,9924	141,2706	Estagio 3
14,7295	28,6164	21,4925	34,0388	0,0018	6,7208	27,8196	30,1145	0,8547	223,1855	440,4637	220,5130	43,8545	84,0312	131,3254	Estagio 3
18,2820	24,7712	19,2227	30,4752	0,0000	6,5898	25,8669	27,6582	1,3328	200,2173	415,6030	205,0617	36,3260	77,8005	122,8889	Estagio 3
17,8942	22,3740	18,0344	28,0418	0,0000	6,5895	24,9562	25,7972	1,1117	183,7514	400,4448	193,2278	32,9675	73,4762	117,4826	Estagio 3
18,1995	21,9486	17,3098	26,6363	0,0000	5,7960	24,0217	24,3749	0,9149	181,5963	387,2926	191,2330	29,0103	71,4103	114,4259	Estagio 3
16,7543	19,7452	15,6714	25,7062	0,0000	5,7500	23,5262	23,7705	1,0290	155,9362	364,7365	172,9398	20,1767	65,0571	106,2121	Estagio 3
18,6957	17,3380	15,5835	23,4168	0,0000	4,4303	23,2550	20,5272	0,9613	147,3692	356,8427	166,8305	20,5058	62,7505	103,4981	Estagio 3
15,2297	17,4593	14,8273	21,3515	0,0000	4,8705	22,7138	22,3381	1,0689	142,9192	351,8335	160,3032	16,4122	60,8713	101,7605	Estagio 3
17,8392	14,7902	13,7897	20,1958	0,0000	6,2015	21,9726	20,9286	0,6173	130,1692	340,5680	149,7315	10,3397	57,4725	97,7290	Estagio 3
16,3931	14,1595	13,3382	20,3333	0,0000	4,8546	21,7452	21,3019	0,7444	133,2116	338,0672	145,8120	9,7062	56,8975	97,1752	Estagio 3
16,1987	14,5700	12,7415	19,4802	0,0000	3,8791	21,3189	21,2308	1,2934	135,4174	335,3732	148,1863	9,6002	56,8684	96,9367	Estagio 3

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
15,3697	13,0058	13,0485	18,8582	0,0000	2,9015	21,4775	21,3292	0,7621	123,6828	327,3562	135,1515	5,0553	53,6922	93,5995	Estagio 3
14,5647	13,1257	12,6442	17,3125	0,0000	4,4643	22,0266	20,9431	1,0057	120,8340	321,0897	133,6805	5,6790	52,8746	91,7635	Estagio 3
17,6207	12,1770	13,0325	17,2445	0,0000	5,2303	21,5967	19,8640	1,0743	120,6483	321,0293	124,7537	4,1870	52,1891	91,1063	Estagio 3
18,2752	12,4438	13,5478	16,5282	0,0000	5,5222	21,0025	20,1075	0,8163	121,6908	321,8860	126,8570	3,6365	52,4857	91,5312	Estagio 3
0,0000	15,4576	15,1192	15,0023	0,0000	0,0000	23,3970	25,8622	0,0000	165,9764	316,5864	200,7918	45,6717	63,3742	99,9481	Estagio 4
0,0000	14,2875	15,0000	14,9645	0,0000	0,0000	23,6485	25,0007	0,0000	176,7097	320,3787	201,0101	44,0612	64,2354	101,8218	Estagio 4
0,0000	13,5850	14,9313	14,4113	0,0000	0,0000	23,2640	24,3090	0,0000	171,8213	315,9848	204,1992	41,4367	63,3802	100,9575	Estagio 4
0,0000	12,9537	14,8373	14,0863	0,0000	0,0000	23,0820	23,4535	0,0000	163,6770	308,7077	203,0446	39,7405	61,8141	98,6972	Estagio 4
0,0000	12,4832	14,7367	13,9617	0,0000	0,0000	22,9503	22,8442	0,0000	169,6515	307,5139	204,3106	39,6974	62,1653	99,1811	Estagio 4
0,0000	11,7773	14,4208	13,5465	0,0000	0,0000	22,8023	22,0238	0,0000	164,3559	304,0182	195,4433	36,6269	60,3858	97,0871	Estagio 4
0,0000	11,4505	14,2730	13,5978	0,0000	0,0000	22,9215	22,0023	0,0000	164,5554	303,7524	199,1685	35,6614	60,5679	97,5205	Estagio 4
0,1185	11,3303	14,5582	13,6467	0,0000	0,0000	23,1025	22,2672	0,0000	174,7888	309,8800	204,2623	36,9455	62,3769	100,2661	Estagio 4
0,0573	11,0065	14,2003	13,2665	0,0000	0,0000	23,0000	21,5820	0,0000	170,3076	305,8875	203,0642	35,5320	61,3772	98,9907	Estagio 4
0,0835	10,9767	14,3777	13,3100	0,0000	0,0000	23,0000	21,2003	0,0000	169,2713	304,4742	205,6633	34,6282	61,3066	98,9428	Estagio 4
0,1958	10,8163	14,0982	13,0977	0,0000	0,0000	22,6673	20,6358	0,0000	163,7682	299,3319	200,9216	33,9922	59,9635	96,8799	Estagio 4
0,1644	10,8780	14,2173	13,0000	0,0000	0,0000	22,9145	20,9095	0,0000	171,1522	301,9088	206,5488	34,9068	61,2769	98,7258	Estagio 4
0,6236	10,7435	14,3555	13,1593	0,0000	0,0000	23,0650	21,0850	0,0000	171,8829	304,3668	199,6491	33,9457	60,9905	98,4499	Estagio 4
1,1145	10,4856	14,4427	13,3417	0,0000	0,0000	23,0165	21,1687	0,0000	175,7650	305,8691	209,1173	34,4750	62,2151	100,2182	Estagio 4
0,5308	10,4919	14,3600	13,1933	0,0000	0,0000	23,0912	21,0774	0,0000	170,9730	303,9005	210,9613	34,6956	61,7904	99,6365	Estagio 4
2,6488	21,9952	30,3298	42,8625	0,0000	0,0113	40,6165	33,9233	0,0000	219,4235	394,7666	251,6474	59,4925	84,4398	123,8986	Estagio 4
2,8780	19,7324	27,8458	37,8872	0,0000	0,0000	37,9933	31,1786	0,0000	206,3392	377,0097	238,7297	54,0887	79,5140	118,1318	Estagio 4
2,7765	18,0407	26,1230	34,8282	0,0000	0,0000	36,1163	28,8665	0,0000	189,4613	357,2713	225,5235	48,2738	74,4062	111,4765	Estagio 4
2,5954	16,0643	24,2648	30,7816	0,0000	0,6349	33,8980	26,4475	0,0000	182,0928	343,0938	223,7854	45,5018	71,4739	108,1129	Estagio 4
2,6132	15,0467	23,0000	28,5722	0,0000	0,0000	32,2602	24,7482	0,0000	175,6593	330,6177	213,2494	40,9966	68,2126	104,1088	Estagio 4
2,9713	14,0162	22,1002	27,3713	0,0000	0,0000	31,6625	23,6937	0,0000	168,7725	321,1701	209,7838	38,0638	66,1235	101,3489	Estagio 4
3,1772	13,3488	21,3020	26,4988	0,0000	0,0000	30,8045	23,0034	0,0000	167,4748	315,2527	207,9403	37,0769	65,0676	99,9287	Estagio 4
3,3760	12,7474	20,4305	25,2842	0,0000	0,0000	30,1018	22,0373	0,0000	164,4705	307,5615	202,6428	34,3894	63,3109	97,6414	Estagio 4
3,5661	12,2072	19,9403	24,4525	0,0000	0,0000	29,4812	21,5178	0,0000	164,0228	303,3230	198,7108	32,6152	62,2951	96,3927	Estagio 4
3,4087	11,6808	19,2885	23,3567	0,0000	1,3394	28,9333	20,5207	0,0000	159,5207	295,8151	197,6795	30,8226	60,9512	94,4328	Estagio 4
3,3988	11,2683	18,9355	22,6388	0,0000	3,1713	28,4295	20,2203	0,0000	159,2076	292,5485	193,9848	29,5623	60,2589	93,3040	Estagio 4
3,4898	10,8881	18,2238	21,9400	0,0000	0,0702	27,8933	19,5367	0,0000	153,7388	285,6197	186,3842	26,8253	58,0469	90,8176	Estagio 4
2,4537	10,5110	18,0000	21,6003	0,0000	0,0000	27,1827	19,0465	0,0000	155,1824	284,0053	180,0078	25,2322	57,1709	90,0454	Estagio 4
2,9675	10,0642	17,7633	20,4298	0,0000	1,0501	26,9998	18,8008	0,0000	152,8040	279,0552	175,3508	22,8727	56,0122	88,3331	Estagio 4
2,2632	10,0000	17,7283	20,0322	0,0000	0,2478	26,8600	18,9672	0,0000	152,8613	279,9274	178,8717	23,0845	56,2187	89,0079	Estagio 4

MQ2	MQ3	MQ4	MQ5	MQ6	MQ7	MQ8	MQ9	MQ135	TGS822	TGS2600	TGS2602	TGS2603	Média	Desvio	Classe
0,6582	15,1508	16,6400	19,7900	0,0000	0,0000	23,9335	25,1173	0,2700	160,8303	341,2055	251,6010	51,4176	69,7396	110,7210	Estagio 4
1,1920	14,0167	16,0058	18,9176	0,0000	0,0000	23,5320	23,2557	0,2632	154,9030	323,8357	234,6116	48,7188	66,0963	104,6398	Estagio 4
1,8361	13,0172	15,9167	17,4327	0,0000	0,0000	23,1145	22,3047	0,1770	148,9352	315,0817	234,0576	42,8148	64,2068	102,5685	Estagio 4
1,4416	12,0212	15,0885	16,8017	0,0000	0,0000	22,7488	20,7730	0,3277	144,3297	301,3928	225,9142	38,8955	61,5180	98,5843	Estagio 4
0,9347	11,5487	14,7957	15,6938	0,0000	0,0000	22,0298	19,9823	0,7637	139,3735	290,5188	216,7473	36,4197	59,1391	94,9342	Estagio 4
0,5372	11,0000	14,6200	15,1210	0,0000	0,0000	21,6160	19,8707	1,0000	134,9672	283,5047	210,0247	34,6035	57,4511	92,3839	Estagio 4
0,5733	10,9657	14,3772	15,0209	0,0000	0,0000	21,3282	19,6552	0,9700	133,6183	279,6642	207,2745	33,8672	56,7165	91,1742	Estagio 4
0,3645	10,8910	14,0217	14,1620	0,0000	0,0000	19,8026	19,1127	0,3285	131,3110	276,0705	204,0093	31,8555	55,5330	90,0345	Estagio 4
0,6190	11,0000	14,0000	14,4128	0,0000	0,0000	20,1942	19,1790	0,4558	131,8283	276,4183	204,3752	32,2687	55,7501	90,1352	Estagio 4
0,4345	11,0000	14,0000	14,2812	0,0000	0,0000	20,0693	19,1780	0,3378	131,7893	277,0497	204,4138	32,4633	55,7705	90,2871	Estagio 4
0,5113	11,0000	14,0000	14,2693	0,0000	0,0000	20,0473	19,1233	0,4613	131,7662	276,9668	204,6940	32,3402	55,7831	90,3024	Estagio 4
0,7265	11,0000	14,0000	14,3723	0,0000	0,0000	20,1527	19,0833	0,5088	131,6523	276,7330	204,8612	32,4175	55,8083	90,2486	Estagio 4
0,4858	11,0000	14,0000	14,4697	0,0000	0,0000	20,1325	19,1190	0,4215	131,8765	276,9658	204,6902	32,3403	55,8078	90,3025	Estagio 4
0,5705	11,0000	14,0000	14,4275	0,0000	0,0000	20,2853	19,2107	0,5748	132,0952	276,8403	204,5698	32,1043	55,8214	90,2622	Estagio 4
0,5867	11,0000	14,0000	14,3840	0,0000	0,0000	19,9090	19,0230	0,4932	131,9997	277,0569	204,7155	32,3160	55,8065	90,3388	Estagio 4