

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA

CAMILA ANGELA GONZATTI

**USO DE IMAGENS DIGITAIS NO CONTROLE DE QUALIDADE DE SEMENTES DE
MILHO ATRAVÉS DE DISPOSITIVOS MÓVEIS**

Porto Alegre
2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA

CAMILA ANGELA GONZATTI

**USO DE IMAGENS DIGITAIS NO CONTROLE DE QUALIDADE DE SEMENTES DE
MILHO ATRAVÉS DE DISPOSITIVOS MÓVEIS**

Trabalho de conclusão apresentado junto à atividade de ensino “Projeto Tecnológico – Química” do curso de Química Industrial, como requisito parcial para a obtenção do grau de Bacharel em Química Industrial.

Orientador: Profa. Dra. Bruna Tischer
Coorientador: Prof. Dr. Adilson Ben da Costa

Porto Alegre
2021

CIP - Catalogação na Publicação

Gonzatti, Camila Angela
USO DE IMAGENS DIGITAIS NO CONTROLE DE QUALIDADE DE
SEMENTES DE MILHO ATRAVÉS DE DISPOSITIVOS MÓVEIS /
Camila Angela Gonzatti. -- 2021.
53 f.
Orientadora: Bruna Tischer.

Coorientador: Adilson Ben da Costa.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Química, Curso de Química Industrial, Porto Alegre,
BR-RS, 2021.

1. Classificação de sementes de milho. 2.
PhotoMetrix PRO®. 3. Regressão por Mínimos Quadrados
Parciais (PLS). 4. Imagens digitais. 5. Análise
colorimétrica. I. Tischer, Bruna, orient. II. da
Costa, Adilson Ben, coorient. III. Título.

AGRADECIMENTOS

Aos meus pais, Margane Cornelli Gonzatti e Cledio Gonzatti, por todo apoio, incentivo e cuidado ao longo da minha graduação. Agradeço pelo esforço para me proporcionarem essa oportunidade que envolveu tantos desafios, aprendizados e conquistas.

Ao meu companheiro, Ardel Ednor Nunes Junior, por ser meu amigo e namorado, me ajudando, apoiando e incentivando em momentos de felicidade e de dificuldades.

À toda a minha família que neste ano celebra esta conquista comigo.

Aos amigos que a UFRGS me deu, Ana Paula Hoffmann, Ana Paula de Oliveira, Alex Vieira, Daniela Schneider e Taís Klein. Obrigada por todos os momentos de estudo antes das provas, almoços no RU, ajuda, risadas e conteúdos compartilhados, vocês tornaram minha caminhada na UFRGS muito mais fácil e divertida.

À Lilian Luz, por todos os passeios em Porto Alegre, desabafos, risadas, ajuda, companheirismo, confiança e parceria desde o início da nossa amizade.

À minha orientadora Bruna Tischer por toda paciência, incentivos, aprendizados, muita ajuda, conversas e risadas ao longo destes 3 últimos anos.

Ao meu coorientador, Prof. Adilson Ben da Costa, pela ajuda na execução do meu trabalho de conclusão com paciência e dedicação.

Ao Alexandro Souza por sempre se disponibilizar a me ajudar e sanar todas as minhas dúvidas.

Ao Demis Faqui pela ajuda na construção das câmaras e formulação das imagens das mesmas e a todo o pessoal do laboratório 1118 da UNISC que me ajudou de alguma forma. A ajuda de vocês foi muito importante para a finalização e aperfeiçoamento deste trabalho.

Por fim, agradeço à empresa Match Technologies por me proporcionar a oportunidade de realizar meu trabalho de conclusão de curso durante o estágio e à Universidade de Santa Cruz do Sul (UNISC) por me ceder um espaço de trabalho.

RESUMO

O milho é uma das *commodities* mais relevantes no mercado interno e externo do Brasil. Durante o beneficiamento das sementes de milho na indústria são selecionadas as de boa qualidade dentre as demais através de parâmetros como cor e densidade, por exemplo. A mesa de gravidade e a máquina *Color Sorter* são equipamentos utilizados para este fim. Para a redução do desperdício de sementes boas remanescentes no rejeito destas máquinas, os operadores realizam uma avaliação manual de amostras do mesmo determinando a quantidade de sementes de qualidade presentes ali para posterior reajuste das máquinas. Esta análise visual e manual torna-se morosa, subjetiva e imprecisa. Assim, o presente trabalho objetivou avaliar a viabilidade da utilização do aplicativo *PhotoMetrix PRO*® para realizar esta determinação através de imagens digitais e análise multivariada dos elementos de cor. Primeiramente, foi realizada uma otimização das condições experimentais a fim de determinar a condição ideal dentre elas. Foram feitos 18 experimentos variando-se o tamanho do recipiente de análise (Câmaras A e B), a distância da amostra até o celular (altura máxima da câmara, $\frac{1}{3}$ e $\frac{1}{2}$ da altura da câmara) e o tamanho da região de interesse utilizada no aplicativo (384 x 384, 256 x 256 e 192 x 192). Após, construiu-se um modelo de calibração através da regressão por mínimos quadrados parciais na condição experimental previamente determinada e realizou-se a predição de algumas amostras. Na otimização foi possível concluir que a condição que forneceu o menor erro de validação cruzada foi aquela composta pela Câmara A, com ROI de 384 x 384 e máxima distância do celular até a amostra. O modelo de calibração construído nestas condições forneceu um coeficiente de determinação igual a 0,976, o RMSEC igual a 1,23% e o RMSECV igual a 4,76%. Na predição observou-se dois *outliers* que geraram um erro absoluto de 17,7%, quando eles foram retirados, o RMSEP caiu para 2,24%. Os limites de quantificação e detecção também foram calculados, obteve-se 10,20 e 3,06% respectivamente. Assim, foi possível concluir que a metodologia se mostrou promissora, apresentou coerência nos resultados e forneceu erros aceitáveis para a aplicação proposta, ainda que necessite de aperfeiçoamentos e mais estudos.

Palavras-chave: Classificação de sementes de milho; *PhotoMetrix PRO*®; Regressão por Mínimos Quadrados Parciais (PLS); Imagens digitais; Análise colorimétrica.

ABSTRACT

Corn is one of the most important commodities in Brazil's economy. During the processing of corn seeds in the industry, those of good quality are selected among the others through parameters such as color and density, for example. The gravity table and the Color Sorter machine are equipment used for this purpose. To reduce the waste of good seeds remaining in the tailings of these machines, the operators carry out a manual evaluation of samples of the same, determining the quantity of quality seeds present there for later readjustment of the machines. This visual and manual analysis becomes time consuming, subjective and imprecise. Thus, this study aimed to evaluate the feasibility of using the PhotoMetrix PRO® application to perform this determination through digital images and multivariate analysis of color elements. First, an optimization of the experimental conditions was carried out in order to determine the ideal condition. Eighteen experiments were performed varying the size of the analysis recipient (Chambers A and B), the distance from the sample to the cell (maximum height of the chamber, $\frac{1}{3}$ and $\frac{1}{2}$ of the height of the chamber) and the size of the region of interest used in the application (384 x 384, 256 x 256 and 192 x 192). A calibration model was built using partial least squares regression in the previously determined experimental condition and a prediction of some samples was performed. In the optimization, it was possible to conclude that the condition that provided the smallest cross-validation error was that composed by Chamber A, with a ROI of 384 x 384 and maximum distance from the cell phone to the sample. The calibration model built under these conditions provided a coefficient of determination equal to 0.976, the RMSEC equal to 1.23% and the RMSECV equal to 4.76%. In the prediction, two outliers were observed that generated an absolute error of 17.7%, when they were removed, the RMSEP dropped to 2.24%. The limits of quantification and detection were also calculated, obtaining 10.2 and 3.06% respectively. Thus, it was possible to conclude that the methodology proved to be promising, presented coherence in the results and provided acceptable errors for the proposed application, even though it requires improvements and more studies.

Keywords: *Classification of corn seeds; PhotoMetrix PRO®; Partial Least Squares Regression (PLS); Digital images; Colorimetric analysis.*

LISTA DE FIGURAS

- Figura 1 - Interface do aplicativo *PhotoMetrix* PRO®: (A) tela inicial, (B) tela de configurações, (C) escolha do modelo de cores a ser utilizado na análise univariada, e (D) interface para análise univariada..... 20
- Figura 2 - Módulo de opções para análise multivariada no *PhotoMetrix* PRO®: (A) página inicial, (B) opções para análise multivariada (PCA, PLS e HCA), (C) Módulo de opções para PCA, (D) opções de processamento para PCA e PLS, (E) Módulo de opções para PLS, e (F) Módulo de opções para HCA..... 21
- Figura 3 - Regiões de interesse selecionadas para os experimentos de otimização... 26
- Figura 4 - Esquema descritivo das câmaras construídas para análise das sementes.. 28
- Figura 5 - Comparação das tonalidades de amarelo vistas nas sementes de boa (B) e má qualidade (A)..... 30
- Figura 6 - Configuração do aplicativo para a realização dos experimentos: A) Seleção do modo de análise multivariada; B) Escolha da regressão por mínimos quadrados parciais para formulação do modelo quimiométrico; C) Início da calibração do modelo; D) Informe da quantidade de amostras analisadas, do modo de decomposição das imagens a ser utilizado e do local para salvar o modelo..... 31
- Figura 7 - Configuração do aplicativo para a realização dos experimentos de predição: A) Seleção do modo de análise multivariada; B) Escolha da regressão por mínimos quadrados parciais C) Seleção do modo de amostragem; D) Informe da quantidade de amostras, do modo de decomposição das imagens a ser utilizado, do local para salvar os resultados e seleção da opção de concentração conhecida..... 34
- Figura 8 - Diagrama de Pareto para os resultados obtidos com a Câmara B durante a otimização..... 37
- Figura 9 - Modelo de superfície linear construído para a Câmara B..... 38

Figura 10 - Diagrama de Pareto para os resultados obtidos com a Câmara A durante a otimização.....	39
Figura 11 - Modelo de superfície linear construído para a Câmara A.....	40
Figura 12 - Modelo de regressão por mínimos quadrados parciais (PLS) construído no Experimento 1 através de 5 variáveis latentes.....	41
Figura 13 - Modelo de regressão por mínimos quadrados parciais (PLS) construído através de 7 variáveis latentes durante a calibração com 21 amostras.....	42

LISTA DE TABELAS

Tabela 1 - Experimentos realizados e parâmetros variados na Câmara A.....	29
Tabela 2 - Experimentos realizados e parâmetros variados na Câmara B.....	29
Tabela 3 - Resultados gerados pelos melhores modelos construídos a partir dos experimentos com a Câmara A.....	35
Tabela 4 - Resultados gerados pelos melhores modelos construídos a partir dos experimentos com a Câmara B.....	36
Tabela 5 - Valores de concentração x Valores preditos pelo modelo (experimento com 7 amostras).....	43
Tabela 6 - Valores de concentração x Valores preditos pelo modelo (experimento com 5 amostras).....	43

LISTA DE ABREVIATURAS E SIGLAS

bit - Dígito binário - do inglês "*binary digit*"

CNN - Redes Neurais Convolucionais - do inglês "*Convolutional Neural Networks*"

FAO - Organização das Nações Unidas para a Alimentação e Agricultura

HCA - Análise Hierárquica de Cluster - do inglês "*Hierarchical Cluster Analysis*"

HSI - Matiz, Saturação e Intensidade - do inglês "*Hue, Saturation, and Intensity*"

HSL - Matiz, Saturação e Luminosidade - do inglês "*Hue, Saturation, and Lightness*"

HSV - Matiz, Saturação e Valor - do inglês "*Hue, Saturation and Value*"

LD - Limite de detecção

LQ - Limite de quantificação

Mask R-CNN - (Redes Neurais Convolucionais Baseadas em Região - do inglês "*Region-Based Convolutional Neural Networks*")

PLS - Mínimos Quadrados Parciais - do inglês "*Partial Least Squares*"

PCA - Análise do Componente Principal - do inglês "*Principal Component Analysis*"

RGB - Vermelho, Verde, Azul - do inglês "R (*Red*), G (*Green*), B (*Blue*)"

RMSEC - Erro Quadrático Médio de Calibração - do inglês "*Root Mean Square Error of Calibration*"

RMSECV - Erro Quadrático Médio de Validação Cruzada - do inglês "*Root Mean Square Error of Cross Validation*"

RMSEP - Erro Quadrático Médio de Predição - do inglês "*Root Mean Square Error of Prediction*"

ROI - Região de interesse - do inglês "*Region of interest*"

VBP - Valor Bruto da Produção

VL - Variáveis latentes

LISTA DE EQUAÇÕES

Equação 1.	$X = TP^t + E$	23
Equação 2.	$Y = TQ^t + F$	23
Equação 3.	$u_h = bt_h$	24
Equação 4.	$Y_{novo} = T^*Q^t + F$	24
Equação 5.	$\sum_i (c_i - \hat{c}_i)^2$	24
Equação 6.	$LD = 3.S_o$	34
Equação 7.	$LQ = 10.S_o$	34

SUMÁRIO

1.	INTRODUÇÃO	12
2.	OBJETIVOS	13
2.1	Objetivo Geral	13
2.2	Objetivos Específicos	13
3.0	REVISÃO BIBLIOGRÁFICA	14
3.1	Mercado do milho	14
3.2	Classificação das sementes	15
3.3	Imagens digitais	16
3.4	<i>PhotoMetrix</i> PRO®	18
3.5	Análise Multivariada	22
3.5.1	Mínimos quadrados parciais (Partial Least Squares - PLS)	23
4.	METODOLOGIA	25
4.1	Amostras	25
4.2	Definição dos parâmetros de otimização	25
4.3	Otimização das condições experimentais	26
4.3.1	Construção dos recipientes de análise	27
4.3.2	Experimentos de teste	28
4.3.3	Uso do <i>PhotoMetrix</i> PRO®	29
4.3.4	Avaliação dos resultados da otimização	32
4.4	Calibração do modelo e predição das amostras	32
5.	RESULTADOS E DISCUSSÕES	34
5.1	Otimização	34
5.1.1	Resultados Câmara B	36
5.1.2	Resultados Câmara A	38
5.2	Calibração do modelo e predição das amostras	41
6.	CONCLUSÃO	47
	REFERÊNCIAS	49

1. INTRODUÇÃO

O agronegócio é um setor de grande importância na economia do Brasil, sendo o milho uma das *commodities* mais relevantes no mercado interno e externo do país. A cadeia do milho é vasta e diversificada, além de suprir e estar atrelada a outras diversas cadeias produtivas de grande importância econômica como a pecuária, combustíveis e a indústria alimentícia.

O processo pós-colheita das sementes de milho na indústria envolve várias etapas de beneficiamento que removem os materiais indesejáveis e garantem a qualidade das mesmas. Dentre as diversas etapas inseridas no processo de beneficiamento das sementes, existem as de classificação e separação, nas quais além da avaliação de parâmetros como tamanho e forma, são selecionadas as sementes de boa qualidade dentre as demais. A referida seleção pode ser feita através de parâmetros como cor, densidade, textura, etc.

Existem diversos equipamentos nos quais estas etapas são realizadas e, dentre eles, encontram-se a mesa de gravidade e a máquina *Color Sorter*. As mesas de gravidade removem materiais e sementes indesejáveis através da diferença de peso em relação às sementes de qualidade. Já a máquina *Color Sorter*, separa as sementes pela diferença de coloração e textura. Dentro de uma planta industrial, vários exemplares destes equipamentos são utilizados concomitantemente. Ainda que a separação gerada por eles seja eficaz, no rejeito destas máquinas remanescem sementes boas que acabam sendo desperdiçadas. Estes equipamentos têm capacidade para processar toneladas de sementes por hora. Deste total, aproximadamente 3% (varia conforme o tipo de semente) se tornam rejeitos, ou seja, centenas de quilogramas de semente/hora. Neste descarte, há, aproximadamente, 5 a 30% de sementes boas (dezenas de kg de sementes boas/hora).

A fim de diminuir a quantidade de sementes de qualidade desperdiçadas na seleção pelos equipamentos, as empresas empregam uma avaliação manual deste processo, na qual a amostragem é feita no rejeito das máquinas. Os operadores responsáveis coletam com periodicidade (previamente estipulada) amostras do rejeito de cada máquina e realizam a seleção das sementes de boa qualidade que estão em meio às sementes rejeitadas. Após, estas sementes são pesadas e se faz o cálculo da porcentagem das mesmas dentre o rejeito para posterior correção dos parâmetros nos equipamentos a fim de diminuir o desperdício. Como descrito, a inspeção das amostras

de rejeito das máquinas é feita de maneira visual e manual pelos operadores tornando-se assim, morosa, subjetiva e imprecisa.

Desta forma, o objetivo do trabalho é estudar a viabilidade do uso do aplicativo *PhotoMetrix* PRO® para determinar a porcentagem de sementes de qualidade presentes nos rejeitos destas máquinas. Para isso serão utilizadas imagens digitais e análise multivariada aplicando regressão por mínimos quadrados parciais (PLS - do inglês *Partial Least Squares*). Caso seja constatada a viabilidade do uso do *PhotoMetrix*®, o processo que, atualmente é custoso em termos de tempo e dinheiro para a indústria, poderá ser feito em poucos minutos através de um celular de uma forma precisa, diminuindo significativamente a mão de obra necessária para este fim. Portanto, este trabalho visa contribuir para o processo de produção da indústria de sementes, através do desenvolvimento de uma metodologia rápida, precisa e de baixo custo, que facilite a coleta de informações e o ajuste de desempenho dos equipamentos. O parâmetro utilizado para avaliação da qualidade será a intensidade da cor amarela nas sementes, visto que as sementes de baixa qualidade/com defeitos possuem um tom de amarelo escurecido.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Avaliar a viabilidade do uso do aplicativo *PhotoMetrix* PRO® para determinar o percentual de sementes de qualidade em meio ao rejeito das máquinas empregadas durante o processo industrial de seleção de sementes, utilizando análise multivariada de imagens digitais.

2.2 OBJETIVOS ESPECÍFICOS

- Construir um recipiente adequado para a realização das capturas das imagens das sementes e análises das mesmas através de um aplicativo de celular;
- Otimizar as melhores condições experimentais para a realização das análises;
- Construir um modelo de regressão que seja capaz de informar o percentual de sementes de qualidade na amostra apenas através de uma imagem da mesma;

- Contribuir para o processo industrial de seleção de sementes, através do desenvolvimento de uma metodologia rápida, precisa e de baixo custo que possa ser empregada na indústria e que facilite a coleta de informações e o ajuste de desempenho dos equipamentos.

3. REVISÃO BIBLIOGRÁFICA

3.1 MERCADO DO MILHO

A cultura do milho é a maior cultura agrícola do mundo, sendo a única a ter ultrapassado a marca de 1 bilhão de toneladas colhidas por ano. Os maiores produtores mundiais do cereal são os Estados Unidos, seguidos da China e do Brasil. No ano agrícola de 2019/2020, quando a produção mundial alcançou o recorde de 1,11 bilhões de toneladas, estes países produziram 345,9 milhões de toneladas, 260,8 milhões de toneladas e 101 milhões de toneladas, respectivamente, concentrando 63,6% da produção global (DUARTE et al., 2021).

Atualmente, a cultura do milho é atrelada a várias cadeias produtivas como na criação de aves de corte e de postura (ovos), na bovinocultura de corte e de leite, na suinocultura e na piscicultura (DUARTE et al., 2021). Estima-se que cerca de 70% da produção mundial de milho é destinada à alimentação animal (PAES, 2006). Além de ser importante também para as indústrias de alimentos, bebidas e combustíveis. Esta importância econômica do cereal fez sua produção prosperar em todas as partes do globo ao longo das últimas décadas. Segundo a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), o Valor Bruto da Produção (VBP) da safra mundial de milho em 2015/2016 totalizou US\$ 191,9 bilhões, sendo inferior apenas ao da safra de arroz (US\$ 332,5 bilhões) (DUARTE et al., 2021).

No Brasil, o milho é praticamente cultivado em todo o território nacional (CRUZ et al., 2011). Os principais produtores de milho no Brasil são as regiões Centro-Oeste, Sudeste e Sul do país, concentrando cerca de 90% da produção no país. Destaca-se o estado do Paraná como maior produtor nacional, seguido por Minas Gerais (DUARTE et al., 2006). O destino do milho produzido no país consiste principalmente na formulação de ração para alimentação animal ou na produção de silagem para a pecuária leiteira. Assim, é possível afirmar que o milho é fundamental para a manutenção da produção interna de proteína animal, que, por sua vez, é importante

para a alimentação da população brasileira. Adicionalmente, o cereal também origina uma produção com valor agregado para comercialização no mercado externo. A exportação é um meio de escoar o excedente produzido e de garantir os preços recebidos pelos agricultores brasileiros (DUARTE et al., 2021).

O consumo industrial do milho é outro segmento importante dentro das cadeias correlatas do cereal. Além da indústria alimentícia, os derivados do milho também são aplicados nas indústrias química, farmacêutica, de papéis, têxtil, entre outras de aplicação ainda mais nobres (PAES, 2006). O consumo do milho na indústria ocorre através de dois processos dos quais é possível obter subprodutos de interesse econômico. Na moagem via seca são obtidos subprodutos tais como farinhas, floculados e óleo. Já na moagem via úmida, são obtidos amidos, xaropes de glicose, ingredientes para rações e álcool combustível. A produção de etanol a partir do cereal vem se estabelecendo no Brasil, o que tem feito crescer o consumo industrial de milho (DUARTE et al., 2021).

3.2 CLASSIFICAÇÃO DAS SEMENTES

A pós-colheita do milho abrange as etapas de beneficiamento das sementes, que, por sua vez, são de suma importância para garantir a qualidade e a segurança do produto, complementando os cuidados no campo de cultivo. Apesar de possuir variações de acordo com a espécie da planta e das condições de cada lote em questão (CARVALHO & NAKAGAWA, 2000), as etapas de beneficiamento consistem basicamente nas seguintes operações: recepção e acondicionamento, pré-limpeza, secagem, debulha, limpeza, classificação, tratamento e embalagem (VILLELA & PERES, 2004). Estas etapas devem seguir uma sequência específica e possuem o objetivo de separar matérias indesejáveis junto às sementes. Entretanto, por mais detalhados e cuidadosos que sejam o cultivo e a colheita, estes materiais muitas vezes acabam remanescendo (MAZURKIÉVICZ, 2011).

A etapa de classificação das sementes consiste no uso de parâmetros como tamanho (largura, espessura e comprimento), forma, peso, textura e cor para separar as sementes (CARVALHO & NAKAGAWA, 2000). Esta etapa é feita em equipamentos especializados que separam sementes de qualidade daquelas indesejáveis ou ainda de impurezas através de diferenças nos parâmetros citados. Os equipamentos mais comuns para a realização da classificação das sementes de milho são peneiras,

separadores de cilindro, mesas de gravidade e a máquina *Color Sorter*, equipamento com maior tecnologia que ascendeu após a modernização do setor.

As mesas de gravidade são equipamentos constituídos de uma mesa de superfície levemente inclinada e porosa para a passagem de uma corrente de ar. A passagem do ar com fluxo regulado e os movimentos elípticos separam as sementes em camadas nas quais as sementes mais leves se concentram na parte mais baixa da mesa e as mais pesadas, na parte superior da inclinação (TOLEDO & MARCOS FILHO, 1977). Elas possuem como meio de separação o peso específico, removendo sementes mal formadas, deterioradas, atacadas por insetos ou microrganismos, além de outros tipos de impurezas. Em geral, são postas no final da linha de beneficiamento e realizam uma separação mais eficiente quando as sementes apresentam um tamanho uniforme (FERREIRA, 2010).

A máquina *Color Sorter*, por sua vez, é um método moderno para separação e sua finalidade é separar as sementes pela diferença de textura e coloração, já que as sementes de qualidade possuem um tom de amarelo diferente daquelas indesejáveis. A máquina possui câmeras que capturam imagens de alta definição e utilizam filtros de cores para promover a separação das sementes atacadas por fungos e/ou insetos, danificadas e impurezas. O descarte das mesmas é feito por meio de um jato de ar comprimido, permitindo a remanescência apenas das sementes de boa qualidade (VIEIRA, 2013).

A separação utilizando os equipamentos citados acima é de alta eficiência. Entretanto, ainda assim, dentro do montante de grãos de baixa qualidade, ou seja, no rejeito da máquina *Color Sorter* e das mesas de gravidade, há uma porcentagem de sementes boas que acabam sendo desperdiçadas. Para diminuir este desperdício, os operadores responsáveis por esta etapa dentro da empresa coletam periodicamente amostras deste rejeito e realizam a seleção das sementes de alta qualidade que estão dentre as sementes rejeitadas, a fim de calcular a porcentagem das mesmas e reajustar parâmetros nos equipamentos (VIEIRA, 2013).

3.3 IMAGENS DIGITAIS

As imagens digitais são representações de imagens bidimensionais através de um sistema de numeração binária (0,1). Uma imagem digital é uma longa sequência destes números binários que, por sua vez, se transformam em pequenos quadrados,

denominados *pixels* (MACHADO & SOUKI, 2004). Portanto, a menor unidade formadora de uma imagem digital é denominada *pixel* (MACHADO & SOUKI, 2004). A quantidade de *pixels* que formam a imagem dita o tamanho e a resolução da mesma. Quanto mais *pixels* tiver uma imagem, maior será o seu tamanho e melhor será a qualidade de detalhe visível da mesma (LEÃO, 2005).

A menor unidade do código binário é denominada *bit* (abreviatura de *binary digit*) e consiste em dados que podem ser 0 ou 1 (LEÃO, 2005). Um *byte* é composto por 8-bits, que, por sua vez, representa 2 elevado na oitava potência, ou seja, 256 combinações possíveis (LEÃO, 2005). Assim, uma imagem com 8 *bits* terá uma combinação de 256 (2^8) cores ou 256 níveis de intensidade diferentes de tons de cinza (imagens em escala de cinza) (LEÃO, 2005 & BUSCH, 1993).

Para a formação das cores das imagens digitais existem principalmente 3 canais de cores: R (*red*), G (*green*) e B (*blue*). Imagens com esta escala de cores são conhecidas como imagens RGB. A grande maioria das câmeras digitais e dos *scanners* é capaz de gerar imagens com profundidade de 24 *bits*, ou seja, 8 *bits* para cada canal de cor, sendo 256 possibilidades de tons de vermelho, 256 de verde e 256 de azul (TRIGO, 2003). Portanto, a cor de um *pixel* pode ser descrita como, por exemplo: R: 32 G: 56 B: 74 (LEÃO, 2005).

As imagens digitais podem ser definidas por uma função bidimensional da intensidade da luz, refletida ou emitida, por um objeto, ou seja, pela distribuição luminosa da energia. Parte desta energia é absorvida, transmitida e refletida, sendo as duas últimas possíveis de serem capturadas por um aparelho (HELPER et al., 2017). Desta forma, as imagens digitais podem fornecer muitas informações de interesse em diversas áreas, como na medicina, geografia, ciências exatas, agricultura, etc. Entretanto, para que sejam extraídas estas informações muitas vezes são necessários processamentos das imagens geradas. O processamento de imagens digitais consiste nas práticas que abordam a manipulação e análise das imagens através do computador (QUINTANILHA, 1990). Ele possui por objetivo realçar as imagens melhorando sua qualidade e aspecto além de fornecer melhores recursos para a interpretação das mesmas (LAPOLLI, 1994).

O uso do processamento de imagens digitais na agricultura vem sofrendo uma evolução ao longo dos anos. A visão computacional foi utilizada para reconhecimento de ervas daninhas, doenças, etc, através de artifícios como reconhecimento de padrões e *deep learning*. Após, passou a ser utilizada a visão de máquina para seleção

e classificação de sementes, através de extração, segmentação e reconhecimento de imagens, como é feito na máquina *Color Sorter*, explicada anteriormente. Por fim, o processamento de imagens digitais também vem sendo utilizado na agricultura de precisão através de drones, sensores remotos, satélites a fim de monitorar plantações, por exemplo.

Yumang et al. utilizaram rede neural convolucional que emprega e integra algoritmos de processamento de imagens digitais para identificar danos físicos em amostras de grãos de milho. Já Velesaca et al. fizeram uso de *deep learning* para detecção e classificação de grãos de milho baseadas em visão computacional. A segmentação da imagem foi realizada através da arquitetura Mask R-CNN (Redes Neurais Convolucionais Baseadas em Região - do inglês "*Region-Based Convolutional Neural Networks*"), enquanto a classificação foi realizada por meio da arquitetura CNN leve (Redes Neurais Convolucionais - do inglês "*Convolutional Neural Networks*"), projetada para este fim. Finalmente, López et al. desenvolveram um software de processamento de imagens para determinar alguns índices de vegetação e gerar mapas tridimensionais de culturas agrícolas.

3.4 *PhotoMetrix PRO*®

O aplicativo *PhotoMetrix*® foi desenvolvido em 2015, sendo uma ferramenta para análises colorimétricas cujos resultados advém da decomposição e processamento de imagens digitais adquiridas pela câmera do smartphone ou, na última versão, através de uma câmera externa. Todas as operações são realizadas no mesmo dispositivo permitindo uso do aplicativo para análise *in situ*, além de ser de fácil utilização e baixo custo (BÖCK et al., 2020).

O *PhotoMetrix*® permite realizar a decomposição da imagem de duas formas, uma através do modelo RGB, baseado na adição das cores primárias vermelho (do inglês *Red*), verde (do inglês *Green*) e azul (do inglês *Blue*). A outra maneira é a partir da decomposição independente dos canais RGB e modelos de cores derivados do RGB, como matiz, saturação e valor (do inglês *Hue, Saturation and Value* - HSV); matiz, saturação e luminosidade (do inglês *Hue, Saturation, and Lightness* - HSL); e matiz, saturação e intensidade (do inglês *Hue, Saturation, and Intensity* - HSI). O matiz determina a cor conforme descrito pelo comprimento de onda, por exemplo, a distinção entre vermelho e amarelo. A medida de saturação é a quantidade de cor

(diferenciação) que é apresentada, por exemplo, a distinção entre vermelho e rosa. Já o parâmetro luminosidade, intensidade ou valor é a quantidade de luz (diferenciação de intensidade de luminosidade), a distinção entre vermelho escuro e vermelho claro ou entre cinza escuro e cinza claro (HELPER et al., 2017).

A segunda versão do aplicativo, denominada *PhotoMetrix* PRO®, permite a realização de análises univariadas e algumas análises multivariadas como análise de componentes principais (do inglês *Principal component analysis - PCA*), análise de cluster (do inglês *Hierarchical cluster analysis - HCA*) e regressão por mínimos quadrados parciais (do inglês *Partial least squares - PLS*) (BÖCK et al., 2020). Nesta versão do aplicativo, as imagens são capturadas através da câmera do próprio *smartphone*.

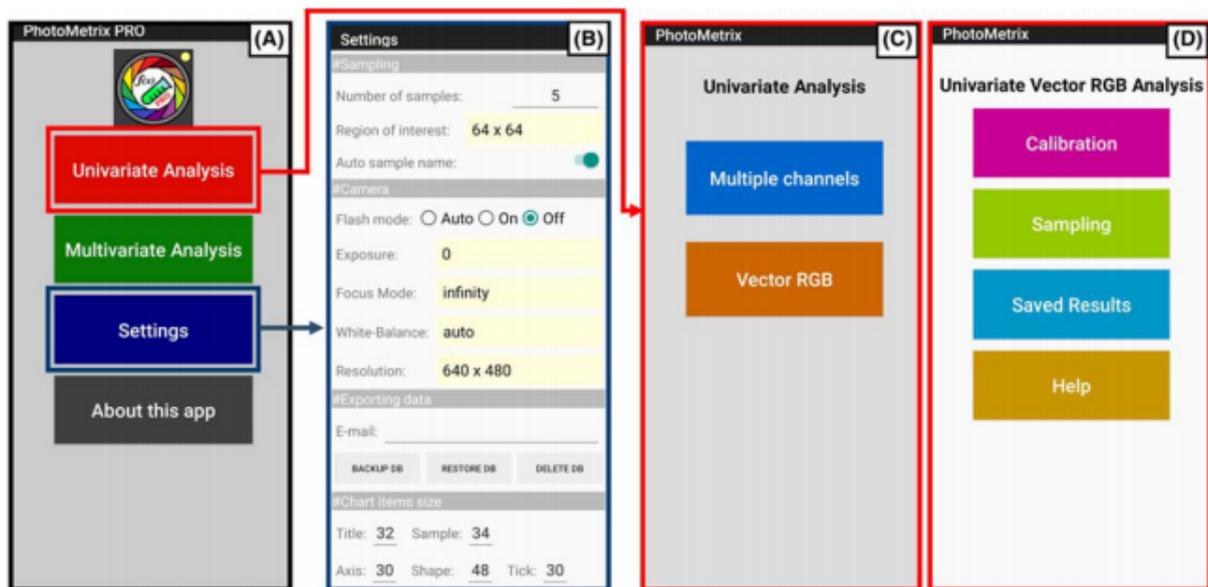
O *PhotoMetrix*® analisa os *pixels* da imagem e a partir de cada um deles, são gerados 3 vetores (R,G e B), cada um com um respectivo valor que varia de 0 a 255 conforme a cor da imagem. Pode-se utilizar o modelo de cores R,G, e B para decompor as imagens, sendo então gerados histogramas indicando a frequência dos valores da intensidade para cada vetor. Porém, se o modelo de decomposição escolhido for o modelo derivado do RGB (HSV, HSI e HSL), a partir dos vetores gerados serão derivadas outras 5 variáveis (H,S,V,L e I) que irão compor o novo modelo de cores.

Na calibração são informadas ao aplicativo as respectivas concentrações dos pontos de calibração e capturadas imagens de cada um deles. Assim, através dos valores para cada componente de cor gerados a partir do modelo escolhido, o *PhotoMetrix*® associará a cor com a concentração, gerando uma curva de calibração e, ao ler as amostras, fornece suas concentrações a partir dela. Uma das grandes vantagens de usar o *PhotoMetrix*® é que todo o processamento de imagem, os cálculos e os gráficos são feitos pelo próprio aplicativo, devolvendo o resultado final de uma maneira rápida e simples. O usuário também tem acesso aos gráficos plotados na sua análise, podendo enviá-los por e-mail.

A Figura 1A ilustra a tela inicial do aplicativo na qual é possível escolher entre análise univariada ou multivariada. Já a Figura 1B mostra a interface de configurações, onde o usuário pode definir parâmetros como número de amostras, tamanho da região de interesse (ROI), uso de *flash* do *smartphone*, configurações para edição dos gráficos e opções para salvar os dados. Ao selecionar a opção "análise univariada" na Figura 1A, a próxima etapa é definir a forma de decomposição das imagens (Figura

1C), que possui as opções “*Multiple channels*” e “*Vector RGB*”. Depois de selecionada a forma de decomposição das imagens, é apresentada a interface de análise univariada (Figura 1D), na qual é possível construir as curvas de calibração, realizar a amostragem e salvar os resultados.

Figura 1 - Interface do aplicativo *PhotoMetrix* PRO®: (A) tela inicial, (B) tela de configurações, (C) escolha do modelo de cores a ser utilizado na análise univariada, e (D) interface para análise univariada.



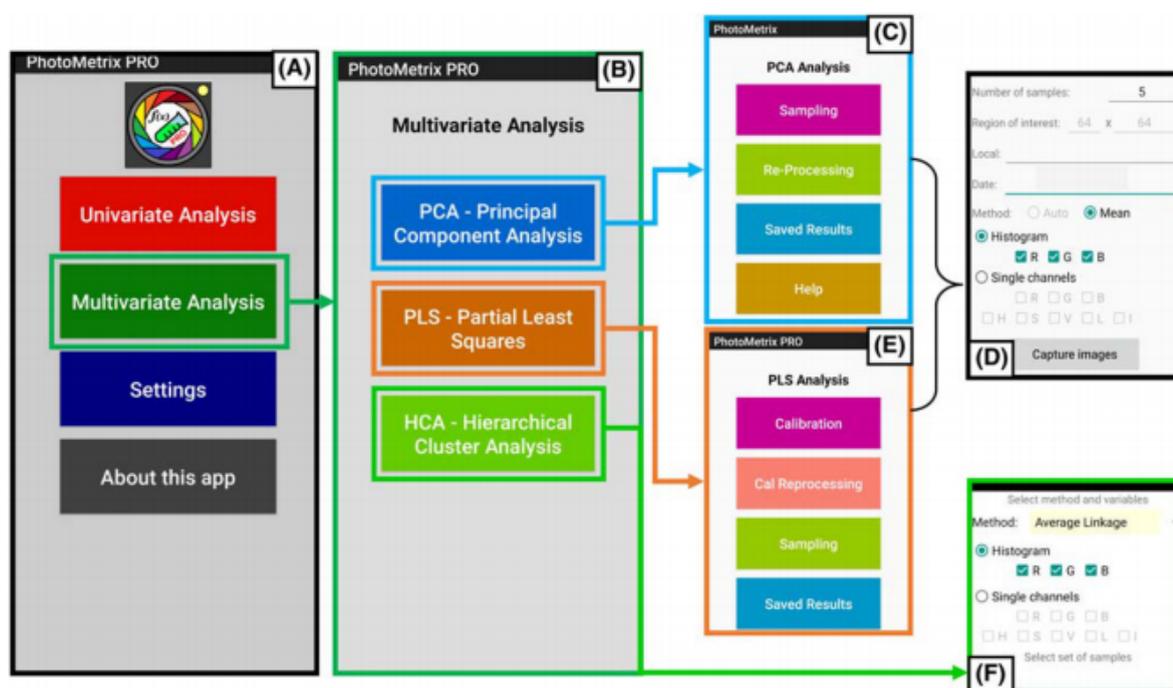
Fonte: BÖCK et al., 2020

Como mostrado na Figura 2A, é possível visualizar as opções para a realização da análise multivariada. O aplicativo oferece três opções de análises multivariadas (Figura 2B): PCA, PLS e HCA. Ao selecionar a opção PCA, uma nova tela é exibida (Figura 2C), onde é possível selecionar entre as opções de "amostragem", "reprocessamento", visualização de "resultados salvos" e "ajuda." Para realizar a amostragem, alguns parâmetros devem ser definidos (Figura 2D), como número de amostras, método de decomposição das imagens e o nome do modelo para o qual as imagens serão adquiridas posteriormente. O reprocessamento de dados é executado na mesma tela (Figura 2D).

Para construir o modelo PLS, deve-se selecionar a opção “calibração” mostrada na Figura 2E, que também mostrará a tela da Figura 2D na qual os parâmetros de análise citados anteriormente serão definidos. Assim como para o PCA, opções de

"reprocessamento", "amostragem" e "resultados salvos" são mostradas na Figura 2E. Para realizar o HCA, o aplicativo leva diretamente à Figura 2F que permite selecionar o modelo de cor que será utilizado na decomposição das imagens.

Figura 2 - Módulo de opções para análise multivariada no *PhotoMetrix* PRO®: (A) página inicial, (B) opções para análise multivariada (PCA, PLS e HCA), (C) Módulo de opções para PCA, (D) opções de processamento para PCA e PLS, (E) Módulo de opções para PLS, e (F) Módulo de opções para HCA.



Fonte: BÖCK et al., 2020

Estudos mostram a aplicação do *PhotoMetrix*® nas mais diversas análises químicas de interesse industrial. Böck et al. determinaram etanol em cachaça através de análise de imagens digitais obtidas com um *smartphone* e analisadas pelo *PhotoMetrix*® em modo multivariado com uso do modelo de calibração PLS. Para isso foi utilizado um recipiente com brilho controlado por duas fitas de LED com seis lâmpadas cada, resultando em uma iluminação de 9 W / m². Os resultados obtidos foram comparados com aqueles encontrados pelo método usado como referência (UV-VIS). Testes estatísticos mostraram que não houve diferença significativa entre os resultados da metodologia proposta e a de referência.

Helfer et al. propuseram uma metodologia alternativa para interpretar a estabilidade térmica do leite cru por reação colorimétrica com alizarol. Os autores construíram um modelo de calibração PLS no *PhotoMetrix*® variando o pH de 3 a 12. As soluções foram dispostas em uma caixa preta com brilho controlado por uma lâmpada LED de 6 W, na qual as imagens foram capturadas para construir o modelo de calibração para posterior previsão de amostras externas. Os resultados obtidos pelo método colorimétrico através do *PhotoMetrix*® foram comparados com o método de referência (potenciométrico). Testes estatísticos mostraram que os resultados obtidos com o aplicativo não apresentaram diferenças estatísticas significativas em relação aos resultados encontrados com o método de referência.

3.5 ANÁLISE MULTIVARIADA

A análise simultânea de duas ou mais variáveis é definida como análise multivariada. O objetivo deste tipo de análise é explicar, prever e medir a relação entre estas variáveis que, por sua vez, devem ser aleatórias e inter-relacionadas (HAIR et al., 2005). A análise multivariada faz uso de elementos estatísticos para descrever e analisar dados de muitas variáveis simultaneamente descrevendo o relacionamento entre elas (JOHNSON & WICHERN, 1988).

Segundo Souza et al., 2014: “O método multivariado consiste em estabelecer e conduzir o menor número de experimentos necessários para extrair o máximo de informação dos dados coletados de modo a avaliar e/ou otimizar um produto ou processo”. Para a realização destes experimentos é feito um planejamento experimental no qual os parâmetros relevantes são considerados. Através deste conjunto de experimentos é possível observar o comportamento das variáveis de interesse frente aos referidos parâmetros e a influência dos mesmos sobre elas. São construídos modelos matemáticos que descrevem o sistema no domínio experimental investigado.

O modo multivariado de análise apresenta uma série de vantagens frente ao modo univariado. Dentre elas estão a possibilidade de verificação da relação entre variáveis, na análise univariada as variáveis são estudadas de forma independente. Adicionalmente, também é possível fazer previsões ao longo de todo o domínio experimental descrevendo como a propriedade de interesse varia em função da variação dos níveis dos fatores (na análise univariada são feitas conclusões pontuais

do processo). Por fim, a otimização multivariada de experimentos permite ao pesquisador tratar todas as variáveis com igual importância sem correr o risco de realizar pré-julgamentos equivocados (SOUZA et al., 2014).

3.5.1 Mínimos Quadrados Parciais (*Partial Least Squares - PLS*)

A regressão por mínimos quadrados parciais é uma técnica de calibração multivariada que permite a avaliação da influência de cada matriz de dados (SHARAF et al., 1986). O PLS relaciona a matriz de dados instrumentais X (variáveis independentes) com o vetor y ou matriz Y (variáveis dependentes) que contém a propriedade de interesse (BERNARDES, 2015). As informações contidas nestas matrizes são coletadas de amostras utilizadas para calibrar o modelo de PLS, ou seja, para treiná-lo de forma a possibilitar que ele faça a predição de amostras externas ao modelo (WOLD et al., 2001).

Em alguns casos, para corrigir fatores experimentais que podem gerar erros no modelo, como espalhamento de luz, se faz necessária a aplicação de alguns pré-processamentos nas matrizes X e Y. Na regressão por PLS, as matrizes X e Y são decompostas simultaneamente através de variáveis latentes (VL), criadas a partir de combinações lineares entre as variáveis originais que eliminam informações não relevantes ao sistema (BERNARDES, 2015 & FIDÊNCIO, 2001). Estas variáveis latentes descrevem simultaneamente a máxima variância de um bloco e promovem o melhor ajuste para o modelo (OLIVEIRA, 2012).

Para encontrar a relação entre a matriz X e a matriz Y, elas são decompostas usando uma matriz de escores (T) comum a ambas, ou seja, calcula-se uma Análise de Componentes Principais (PCA) para X e uma para Y:

$$X = TP^t + E \quad (1)$$

$$Y = TQ^t + F \quad (2)$$

Sendo T a matriz de escores, P e Q, as matrizes de pesos (*loadings*) para X e Y, e E e F, as matrizes de resíduos. O sobrescrito “t” indica uma matriz transposta. O produto de T e P^t, mostrado na Equação 1, é uma aproximação das variáveis independentes (ex: espectros). Já o produto T e Q^t, mostrado na Equação 2, é uma

aproximação das variáveis dependentes (ex: concentrações; classes) (BERNARDES, 2015). O objetivo do modelo é encontrar a melhor relação entre X e Y e que simultaneamente resulte num valor do resíduo F o mais próximo de zero possível (NUNES, 2007).

A relação entre as duas matrizes é encontrada correlacionando-se os escores para cada componente (variáveis dependentes e independentes) de cada vez através de um modelo linear mostrado na Equação 3 (NUNES, 2007).

$$u_h = bt_h \quad (3)$$

Sendo b, o vetor de coeficientes de regressão entre a matriz de escores T para X e Y. A matriz Y pode ser calculada de u_h . Este algoritmo realiza pequenas rotações das variáveis latentes de X e Y para que seja obtida a melhor relação linear possível entre os escores das matrizes X e Y (NUNES, 2007).

A predição da propriedade de interesse (ex: concentração) de novas amostras é estimada a partir dos novos escores chamados T^* , e dos pesos do modelo (Q) de acordo com a equação 3 (BERNARDES, 2015):

$$Y_{novo} = T^* Q^t + F \quad (4)$$

Para a construção de um bom modelo preditivo, é necessário escolher o número correto de variáveis latentes (VL) (BERNARDES, 2015). Esta escolha é feita através da validação cruzada que utiliza a magnitude dos erros de previsão para determinar o número ideal de variáveis latentes a serem utilizadas no modelo (OLIVEIRA, 2012). Na validação cruzada, primeiramente, remove-se uma ou um bloco de amostras do conjunto de calibração e constrói-se o modelo novamente. Então, através do novo modelo construído é realizada a predição das amostras removidas. Por fim, calcula-se o erro de previsão e a soma dos quadrados deste erro através da Equação 4 (FERREIRA et al., 1999). Este processo é repetido para todas as amostras.

$$\sum_i (c_i - \hat{c}_i)^2 \quad (5)$$

Para cada número de VL calcula-se a média dos erros de previsão e, aquele que apresentar o menor erro será o número de VL sugerido para o modelo. O erro de previsão é dado pela raiz quadrada do erro médio da validação cruzada (RMSECV, *root mean square error of cross validation*) (BERNARDES, 2015).

4. METODOLOGIA

4.1 AMOSTRAS

As amostras de sementes de milho utilizadas nos experimentos foram fornecidas por uma empresa do ramo localizada na Região dos Vales do Rio Grande do Sul. Elas consistiram de amostras de rejeito retiradas das máquinas *Color Sorter* presentes na referida indústria. A empresa forneceu as mesmas com a porcentagem de sementes boas presentes em meio ao rejeito previamente calculadas. Cada amostra foi enviada com um recipiente contendo o rejeito da máquina bruto e dois recipientes plásticos contendo as sementes boas e ruins separadas. O total de amostras fornecido foi de 28 amostras, sendo 21 selecionadas para os modelos de calibração e 7 para a predição. A porcentagem de sementes boas contidas nestas amostras variou de 6,0 a 36,5%.

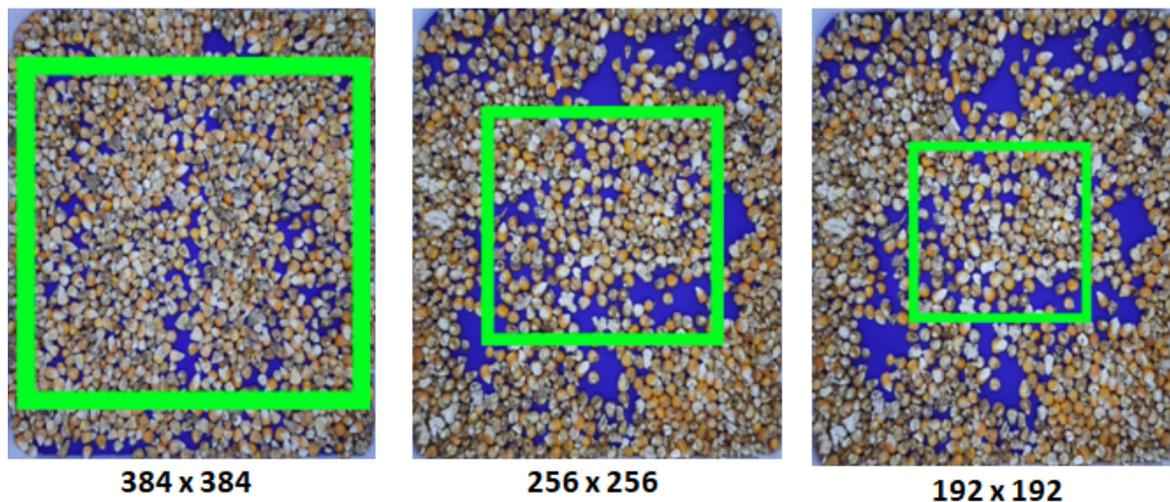
4.2 DEFINIÇÃO DOS PARÂMETROS DE OTIMIZAÇÃO

Alguns testes foram feitos com o objetivo de definir os parâmetros que seriam testados na otimização do aplicativo *PhotoMetrix PRO®* para aquisição das imagens das sementes. Primeiramente, foi decidida a forma ideal da superfície na qual as sementes seriam dispostas. Para isso, pegou-se uma amostra de sementes e elas foram postas sobre uma superfície de forma retangular (31,0 cm x 14,5 cm) e outra, de forma quadrada (16,0 cm x 16,0 cm). Então, capturou-se imagens das sementes em cada uma das superfícies com o auxílio do aplicativo. A partir disso, foi possível perceber que na superfície quadrada conseguia-se abranger uma maior quantidade de sementes na imagem capturada tornando-a o mais representativa possível.

Adicionalmente, foram escolhidos os tamanhos das regiões de interesse (ROI) que seriam testados. Para isso, pegou-se uma caixa de fundo quadrado na qual as sementes foram dispostas e, uma imagem das mesmas com as ROIs de 512 x 512 e 384 x 384 foram capturadas. Assim, foi possível verificar que a maior ROI

disponibilizada no aplicativo poderia ser utilizada sem capturar as bordas da caixa. A partir da ROI de 384 x 384, foram selecionadas as duas menores ROIs precedentes. Na Figura 3 estão mostradas as regiões de interesse escolhidas para realizar a otimização.

Figura 3 - Regiões de interesse selecionadas para os experimentos de otimização.



Fonte: a autora.

Por fim, foram capturadas imagens com diferentes distâncias do celular até a amostra. Para isso, o celular foi posicionado no topo da caixa, 19,5 cm até a amostra, e variou-se a posição das sementes com o auxílio de um suporte construído. As sementes foram postas no fundo caixa (19,5 cm) depois, em $\frac{1}{3}$ da altura da caixa (6,5 cm), em aproximadamente $\frac{1}{2}$ da altura (10,0 cm) e, finalmente, em $\frac{2}{3}$ da altura (13,0 cm). Notou-se que, quando as sementes foram posicionadas em $\frac{2}{3}$ da altura da caixa, a imagem ficava desfocada, pois a amostra estava demasiadamente perto do celular. Assim, resolveu-se testar as distâncias correspondentes à distância máxima, à $\frac{1}{3}$ da altura e à metade dela.

4.3 OTIMIZAÇÃO DAS CONDIÇÕES EXPERIMENTAIS

A partir da definição dos parâmetros que seriam testados, foi feita a otimização das condições experimentais a fim de selecionar aquela que apresentaria os melhores resultados. Nestes experimentos foram variados o tamanho do recipiente de análise (Câmara A e Câmara B), tamanho da ROI (384 x 384, 256 x 256 e 192 x 192) e distância do celular até a amostra. Na Câmara A foram testadas as seguintes

distâncias do celular até a amostra: 19,5 cm, 13,0 cm e 10,0 cm. Já na Câmara B, foram as seguintes: 29,5 cm, 20,0 cm e 15,0 cm.

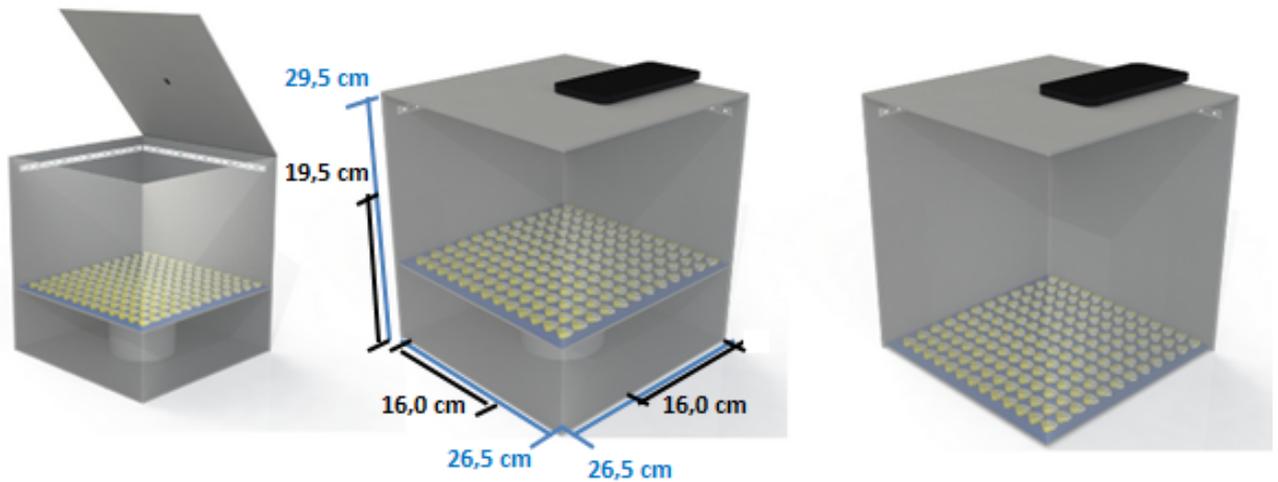
4.3.1 Construção dos recipientes de análise

Para que se pudesse fixar parâmetros como distância e luminosidade foi necessário construir um recipiente de análise no qual estes parâmetros pudessem ser padronizados. A importância de se fazer esta etapa está associada ao erro na captura das imagens que a variação na iluminação, promovida pela luminosidade ambiente, e na distância, se não fixada, podem gerar.

Foram construídas duas câmaras quadradas feitas de papelão com tamanhos diferentes a fim de testar qual delas seria a ideal para a captura das imagens das sementes de milho. A principal variação percebida é o modo com o qual as sementes ficaram dispersas no fundo da câmara (muito próximas ou mais distantes entre si) conforme seu tamanho. Uma das câmaras, nomeada “Câmara B”, foi construída contendo 29,5 cm de altura, 26,5 cm de comprimento e 26,5 cm de profundidade. Já a câmara nomeada “Câmara A”, foi construída contendo 19,5 cm de altura, 16,0 cm de comprimento e 16,0 cm de profundidade.

Na tampa de ambas as câmaras foi recortado um orifício para posicionar a câmera do celular. O fundo delas foi recoberto com papel azul a fim de evitar reflexos nas imagens e, suas paredes, revestidas com papel branco com o objetivo de proporcionar uma melhor dispersão da luz no interior das câmaras. A luminosidade foi promovida por uma unidade de fita *led* (luz branca, 4,6 W/m, 12 V, LED Line - M-2835, 8 mm) colada nas paredes das mesmas. No interior da Câmara A foram utilizados 0,6 m da fita, contendo 38 *leds* que fornecem 1500 LUX. Já no interior da Câmara B, foi utilizado 1 m da fita, cujo número de *leds* é 62, fornecendo 2300 LUX. Por fim, também foi construído um suporte removível de papelão com superfície da cor azul (idêntico ao fundo da câmara) para tornar viável a variação da distância das sementes até o celular. Um esquema das câmaras de análise construídas é mostrado na Figura 4.

Figura 4 - Esquema descritivo das câmaras construídas para análise das sementes.



Fonte: a autora.

4.3.2 Experimentos de teste

Através de um planejamento fatorial foram elaborados 18 experimentos variando-se a distância do celular até a amostra e o tamanho da região de interesse (ROI) em três diferentes níveis dessas variáveis. Além de variar também o tamanho da câmara, através das Câmaras A e B previamente construídas. Estes experimentos foram feitos em duplicata com 12 amostras de calibração, selecionadas dentre as 21 amostras reservadas para este fim. Selecionou-se as amostras com as concentrações máxima e mínima, e com valores de concentrações intermediárias de forma que a variação da concentração entre duas amostras fosse de aproximadamente 2. As Tabelas 1 e 2 mostram as condições experimentais testadas nos experimentos com as Câmaras A e B, respectivamente.

Tabela 1 - Experimentos realizados e parâmetros variados na Câmara A.

Experimento	ROI	Distância (cm)
1	384 x 384	19,5
2	384 x 384	13,0
3	384 x 384	10,0
4	256 x 256	19,5
5	256 x 256	13,0
6	256 x 256	10,0
7	192 x 192	19,5
8	192 x 192	13,0
9	192 x 192	10,0

Tabela 2 - Experimentos realizados e parâmetros variados na Câmara B.

Experimento	ROI	Distância (cm)
10	384 x 384	29,5
11	384 x 384	20,0
12	384 x 384	15,0
13	256 x 256	29,5
14	256 x 256	20,0
15	256 x 256	15,0
16	192 x 192	29,5
17	192 x 192	20,0
18	192 x 192	15,0

4.3.3 Uso do *PhotoMetrix* PRO®

O *PhotoMetrix* PRO® foi utilizado como ferramenta para construir um modelo quimiométrico que pudesse relacionar a intensidade da cor amarela nas amostras de

rejeito com a quantidade de sementes boas presentes ali. Isso porque, como explicado anteriormente e mostrado na Figura 5, as sementes boas e ruins possuem diferença na tonalidade da cor amarela. Para isso, foi utilizada análise multivariada através do modelo de regressão por mínimos quadrados parciais (PLS). Na otimização foi construído um modelo de PLS para cada experimento a fim de verificar quais condições experimentais forneceria o melhor modelo de calibração.

O arranjo experimental consistiu da câmara de análise construída com o celular posicionado na parte superior da mesma, tendo sua câmera encaixada onde foi feita a abertura destinada à ela. Para a realização de todas as análises, foi utilizado o *smartphone* Samsung Galaxy M21s.

Figura 5 - Comparação das tonalidades de amarelo vistas nas sementes de boa (B) e má qualidade (A).

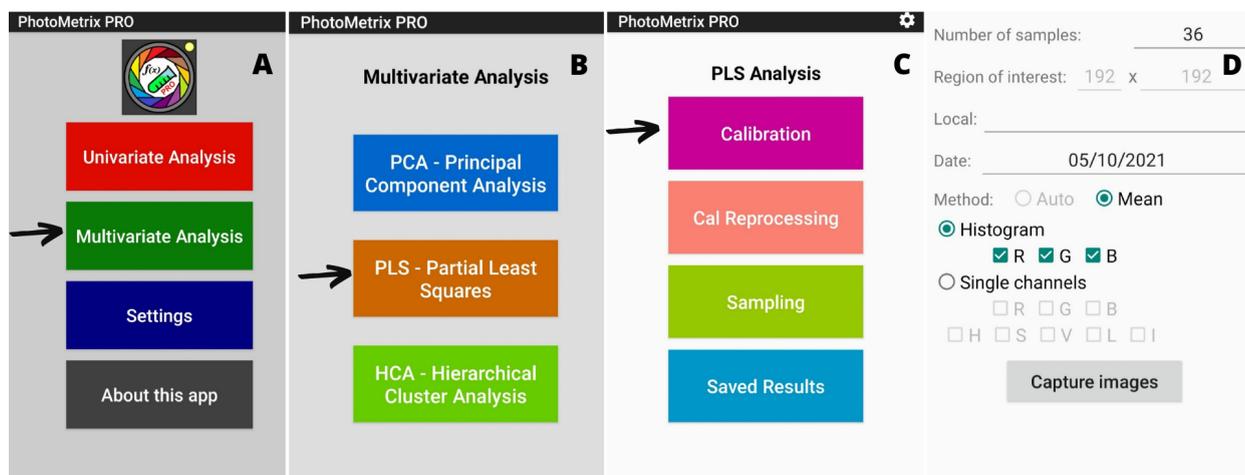


Fonte: a autora.

No *PhotoMetrix PRO®* a ROI utilizada em cada experimento foi selecionada na opção “*Settings*” como mostrado na Figura 1B. Adicionalmente, o aplicativo foi configurado seguindo os botões mostrados na Figura 6. Primeiramente, foi selecionada a opção de “Análise Multivariada” e foi escolhida a análise por PLS. Para a construção do modelo, foi escolhida a opção de “Calibração”. As imagens foram capturadas em triplicata sendo a câmara de análise agitada a cada captura para garantir que as sementes seriam analisadas em ambos os lados, visto que uma semente ruim pode estar danificada em apenas uma de suas faces. Assim, na aba de configurações para calibração colocou-se 36 amostras, pois as imagens das 12 amostras selecionadas

seriam capturadas três vezes. Além disso, nesta aba também foi escolhido o local no qual o modelo criado seria salvo e o modo de decomposição das imagens utilizado (Histograma RGB).

Figura 6 - Configuração do aplicativo para a realização dos experimentos: A) Seleção do modo de análise multivariada; B) Escolha da regressão por mínimos quadrados parciais para formulação do modelo quimiométrico; C) Início da calibração do modelo; D) Informe da quantidade de amostras analisadas, do modo de decomposição das imagens a ser utilizado e do local para salvar o modelo.



Fonte: a autora.

Tendo o aplicativo configurado, iniciaram-se as análises das amostras de calibração. A leitura das mesmas foi feita na ordem crescente de concentração de sementes boas: 6,0%, 8,54%, 10,2%, 13,7%, 15,8%, 17,64%, 19,1%, 21,1%, 23,0%, 25,17%, 31,8% e 36,5%. Cada amostra foi posta na câmara de análise, que, por sua vez, teve sua tampa fechada e o celular posicionado como explicado anteriormente. No aplicativo foi informada a concentração da amostra em questão e foi capturada uma imagem da mesma, cujo tamanho corresponde a ROI utilizada no experimento.

Ao fim da leitura de todas as amostras em triplicata, o aplicativo forneceu o valor de R^2 da reta de regressão e dos erros de calibração (RMSEC - *Mean Square Error of Calibration*) e de validação cruzada (RMSECV - *Root Mean Square Error of Cross Validation*) conforme o número de variáveis latentes utilizadas para construção do modelo. Além disso, o *PhotoMetrix PRO*® também forneceu o gráfico da reta de

regressão dado pelo valor de concentração medido por ele em função do valor de concentração informado na calibração.

4.3.4 Avaliação dos resultados da otimização

A fim de analisar os modelos gerados por cada condição experimental e escolher o melhor dentre eles, foi feito um estudo de como os erros RMSEC e RMSECV variavam em função do número de fatores utilizados para a construção de cada modelo. Foram analisados os erros gerados pelo uso de até 10 variáveis latentes. O modelo escolhido para cada condição experimental foi aquele gerado por um número de variáveis latentes cujo resultado forneceu concomitantemente o menor valor de RMSECV e um valor de coeficiente de determinação (R^2) maior ou igual a 0,90.

De posse destes resultados, foi utilizado o software Chemoface 1.65 com o intuito de determinar a condição experimental que forneceu o melhor modelo de calibração. Através desta ferramenta foram feitos testes estatísticos e construídos diferentes modelos para representar a variação do erro de validação cruzada em função do tamanho da ROI e da distância da amostra até o celular. Estes modelos foram construídos para as duas câmaras testadas e, foram escolhidos a distância, a ROI e o tamanho de câmara que forneceram o menor erro. Para a escolha do tamanho da câmara de análise também consideraram-se parâmetros como portabilidade e facilidade de manuseio.

4.4 CALIBRAÇÃO DO MODELO E PREDIÇÃO DAS AMOSTRAS

A partir da determinação da condição experimental ideal, construiu-se um modelo de calibração através da regressão por mínimos quadrados parciais (PLS) para prever a porcentagem de sementes de qualidade em meio à uma amostra de rejeito. Para a construção deste modelo foram utilizadas as 21 amostras selecionadas para calibração contendo as seguintes porcentagens de sementes boas: 6,00%, 7,00%, 8,54%, 9,90%, 10,20%, 11,40%, 13,70%, 14,00%, 15,80%, 16,30%, 17,64%, 19,10%, 20,30%, 21,10%, 22,89%, 23,00%, 24,90%, 25,17%, 26,63%, 31,80% e 36,5%. O arranjo experimental, a configuração do *PhotoMetrix* PRO® e os procedimentos utilizados nesta etapa foram análogos àqueles descritos na etapa de otimização.

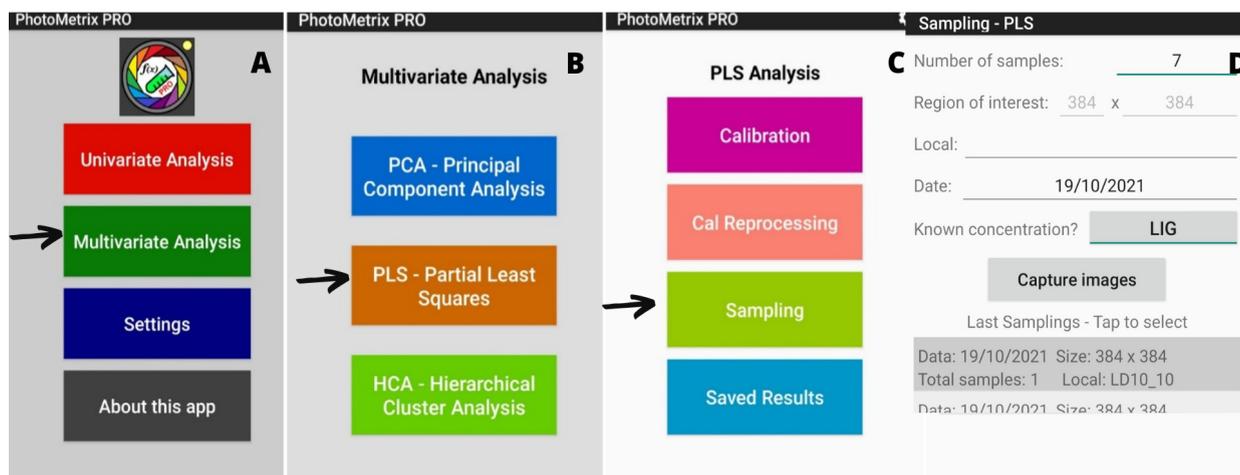
Entretanto, na aba de configurações para calibração colocou-se 63 amostras, pois as imagens das 21 amostras selecionadas foram capturadas três vezes.

Após construir o modelo de calibração, ele foi utilizado para a predição das 7 amostras selecionadas para este fim. Para isso, utilizando o mesmo arranjo e a mesma condição experimental da etapa de calibração, foram capturadas imagens de cada uma destas amostras no interior da câmara de análise. A partir destas imagens e utilizando o modelo de calibração construído anteriormente, o *PhotoMetrix* PRO® realizou a predição da porcentagem de sementes boas presentes em cada uma das amostras.

Neste experimento de predição, o aplicativo foi configurado seguindo os botões mostrados na Figura 7. Primeiramente, foi selecionada a opção de “Análise Multivariada” e foi escolhida a análise por PLS. Para realizar a predição das amostras, foi escolhida a opção “*Sampling*”. Analogamente, as imagens foram capturadas em triplicata sendo a câmara agitada após cada captura. Na aba de configurações para amostragem colocou-se 7 amostras. Da mesma maneira, nesta aba também foi escolhido o nome com o qual estas amostras seriam salvas e o modo de decomposição das imagens utilizado (histograma RGB). Adicionalmente, na mesma tela também foi informado ao aplicativo que a concentração de cada amostra era conhecida, pois assim, ao final do experimento ele fornece o erro de predição (RMSEP - *Root Root Mean Square Error of Prediction*). Este erro é dado em função da diferença entre o valor de concentração fornecido e aquele calculado através do modelo de calibração.

Tendo o aplicativo configurado, iniciou-se as análises das amostras de predição. Analogamente, a leitura das mesmas foi feita na ordem crescente de concentração de sementes boas. Os procedimentos experimentais para leitura das amostras foram os mesmos utilizados na calibração. Ao fim do experimento, o aplicativo gerou o valor de R^2 e a reta de predição, dada pelo valor fornecido ao aplicativo em função do calculado por ele através do modelo de calibração. Além disso, como dito anteriormente, o *PhotoMetrix* PRO® também forneceu o valor do RMSEP.

Figura 7 - Configuração do aplicativo para a realização dos experimentos de predição: A) Seleção do modo de análise multivariada; B) Escolha da regressão por mínimos quadrados parciais C) Seleção do modo de amostragem; D) Informe da quantidade de amostras, do modo de decomposição das imagens a ser utilizado, do local para salvar os resultados e seleção da opção de concentração conhecida.



Fonte: a autora.

Por fim, também foram calculados os limites de quantificação (LQ) e limite de detecção (LD). Para isso, foi feita a leitura de 10 amostras cuja concentração de sementes de qualidade foi de 6,0 %, ou seja, a menor concentração dentre todas as trabalhadas. Então, foi feita a predição destas amostras utilizando o modelo de calibração construído. Com os resultados de predição, foi calculado o desvio padrão amostral e através das normas da EURACHEM (MAGNUSSON & ÖRNEMARK, 2014) obteve-se o LD e o LQ utilizando-se as equações 6 e 7.

$$LD = 3.S_o \quad (6)$$

$$LQ = 10.S_o \quad (7)$$

5. RESULTADOS E DISCUSSÕES

5.1 OTIMIZAÇÃO

Para cada experimento realizado foi feita uma análise dos modelos construídos a partir do número de variáveis latentes empregadas. Assim, para um número de variáveis latentes compreendido entre 1 e 10, foram construídos diferentes modelos nos quais os valores de coeficiente de determinação e dos erros de calibração

(RMSEC) e validação cruzada (RMSECV) foram analisados. Percebeu-se assim que o menor erro de validação cruzada foi obtido no experimento 1 através de um modelo construído a partir de 5 variáveis latentes. Assim, para cada experimento foi selecionado o melhor modelo em função do erro. Cuidou-se para não utilizar um número de variáveis latentes muito alto, pois o modelo fornecido seria ruim, visto que, ele passa a ter influência do erro de sobreposição ocasionando imprecisão ao avaliar amostras externas.

Portanto, o critério utilizado para selecionar o melhor modelo dentre aqueles gerados foi o de menor erro de validação cruzada concomitantemente com um coeficiente de determinação maior que 0,90. Assim, foi escolhido aquele que forneceu o menor RMSECV juntamente com um R^2 superior a 0,90. Este critério foi utilizado visto que é importante o modelo não só explicar bem os dados como também gerar um baixo erro de validação cruzada, uma vez que ele é um indicativo da magnitude do erro de predição. As Tabelas 3 e 4 fornecem os resultados gerados pelo modelo escolhido para cada experimento.

Tabela 3 - Resultados gerados pelos melhores modelos construídos a partir dos experimentos com a Câmara A.

Experimento	nº de VL	RMSEC (%)	RMSECV (%)	R²
1	5	0,8	5,1	0,991
2	6	1,4	5,3	0,973
3	6	2,2	8,9	0,936
4	5	1,1	6,9	0,984
5	6	1,0	7,8	0,986
6	5	2,6	9,4	0,912
7	5	1,0	8,6	0,987
8	6	1,2	7,0	0,981
9	6	2,1	11,6	0,942

Tabela 4 - Resultados gerados pelos melhores modelos construídos a partir dos experimentos com a Câmara B.

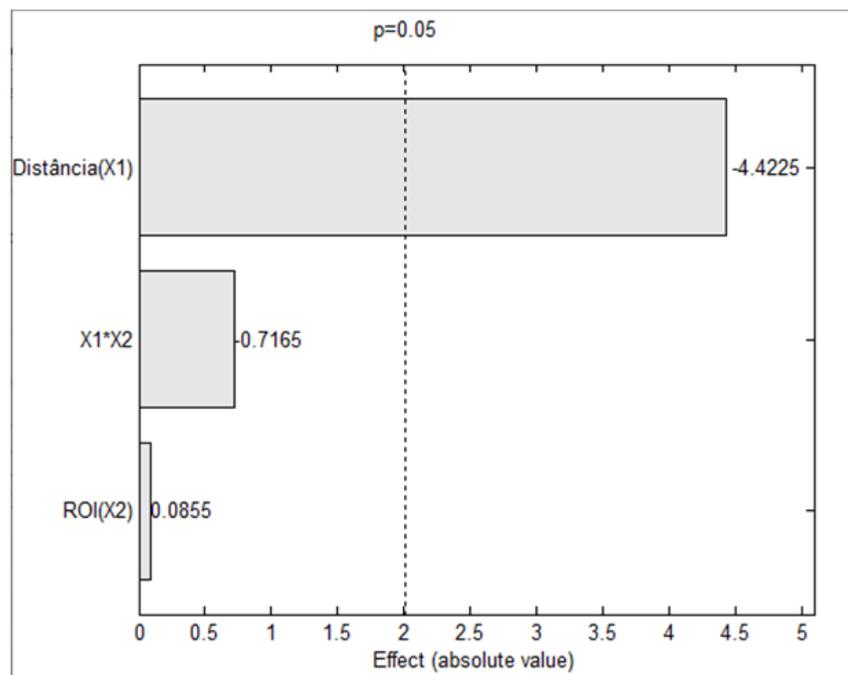
Experimento	nº de VL	RMSEC (%)	RMSECV (%)	R ²
10	6	1,0	6,5	0,987
11	6	2,8	8,8	0,900
12	6	2,4	12,1	0,925
13	6	0,8	5,1	0,992
14	6	1,9	9,5	0,951
15	6	1,9	8,9	0,953
16	6	0,9	6,4	0,989
17	5	2,0	9,2	0,949
18	6	1,5	11,3	0,974

Para verificar a significância dos erros obtidos nas condições experimentais avaliadas e determinar a condição ideal de trabalho, utilizou-se o software Chemoface. Os valores de RMSECV para as duplicatas foram colocados no software juntamente com as variações da ROI e da distância. A partir destes dados foi feito o Teste t de Student e foram gerados modelos lineares de superfície que mostraram a relação entre as variáveis ROI e distância e o erro de validação cruzada (RMSECV) obtido.

5.1.1 Resultados Câmara B

Na Figura 8 podemos observar o Diagrama de Pareto construído a partir do Teste t de Student feito para os resultados obtidos nos experimentos de otimização com a Câmara B.

Figura 8 - Diagrama de Pareto para os resultados obtidos com a Câmara B durante a otimização.



Fonte: Software Chemoface 1.65.

Neste diagrama é possível verificar o efeito de cada variável sobre o erro de validação cruzada. Pode-se perceber que o efeito da distância sobre o erro é significativo, pois o valor absoluto de p_{calc} obtido pelo Teste t de Student é maior do que p (0,05). Além disso, o efeito da distância sobre o erro também é inversamente proporcional, já que o valor obtido no teste estatístico é negativo. Portanto, à medida que se diminui a distância da amostra até o celular, o erro aumenta significativamente.

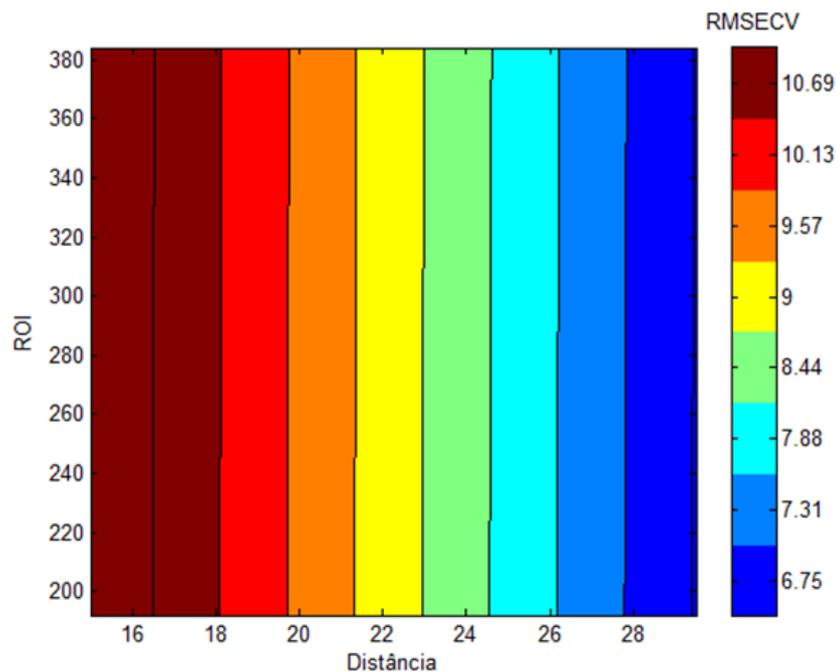
Esse resultado era esperado, visto que, quanto maior a distância, mais sementes são incluídas na imagem, tornando-a mais representativa. Além disso, também supõem-se que uma distância menor das sementes até a borda da câmara possa gerar erros na captura da imagem devido à influência da luminosidade excessiva. Isso porque a fita led foi posta na borda da câmara.

Já para o tamanho da região de interesse (ROI), pode-se perceber pelo Diagrama de Pareto que variando o mesmo não se obtém uma mudança significativa no erro ($p_{\text{calc}} < p$). Uma hipótese formulada para isso é que, como a câmara é grande, em qualquer tamanho da ROI consegue-se abranger uma quantidade razoável de

sementes na imagem. Por fim, também percebe-se que as variáveis distância e tamanho da ROI não proporcionam efeitos significativos entre si (p_{calc} para $X1 \cdot X2 < p$).

Na Figura 9 é observado o modelo de superfície linear construído com base nas variáveis avaliadas e sua influência para os valores de RMSECV.

Figura 9 - Modelo de superfície linear construído para a Câmara B.



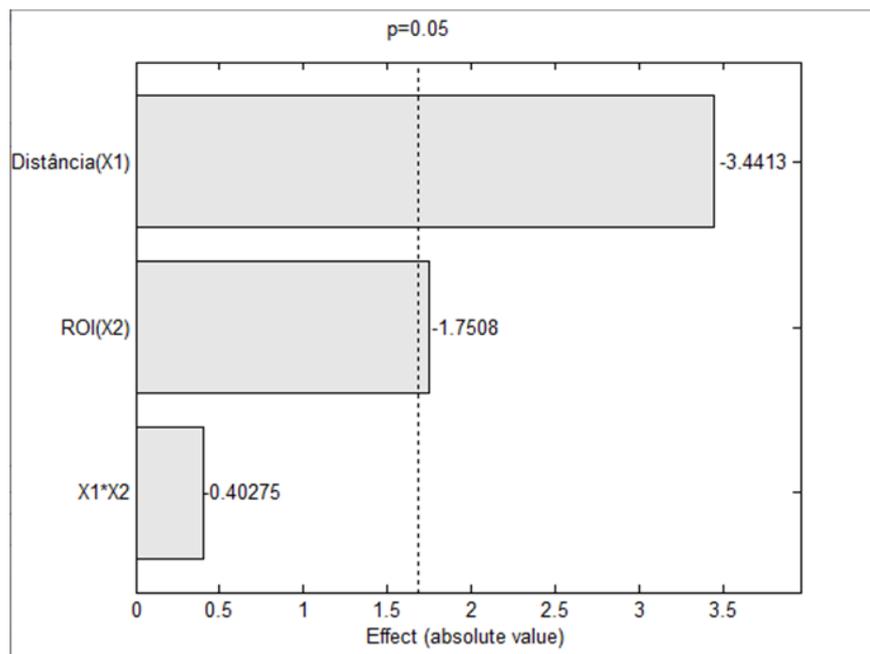
Fonte: Software Chemoface 1.65.

A partir deste modelo nota-se que a única variável cuja mudança interfere nos valores de erro de validação cruzada (RMSECV) é a distância (modelo composto por retas verticais). Na zona colorida de azul escuro, é possível encontrar a condição na qual é obtido o menor valor de RMSECV (em torno de 6,5), sendo esta, a que contém uma distância maior do que 28 cm.

5.1.2 Resultados Câmara A

Na Figura 10 é mostrado o Diagrama de Pareto construído a partir do Teste t de Student feito para os resultados obtidos nos experimentos de otimização com a Câmara A.

Figura 10 - Diagrama de Pareto para os resultados obtidos com a Câmara A durante a otimização.



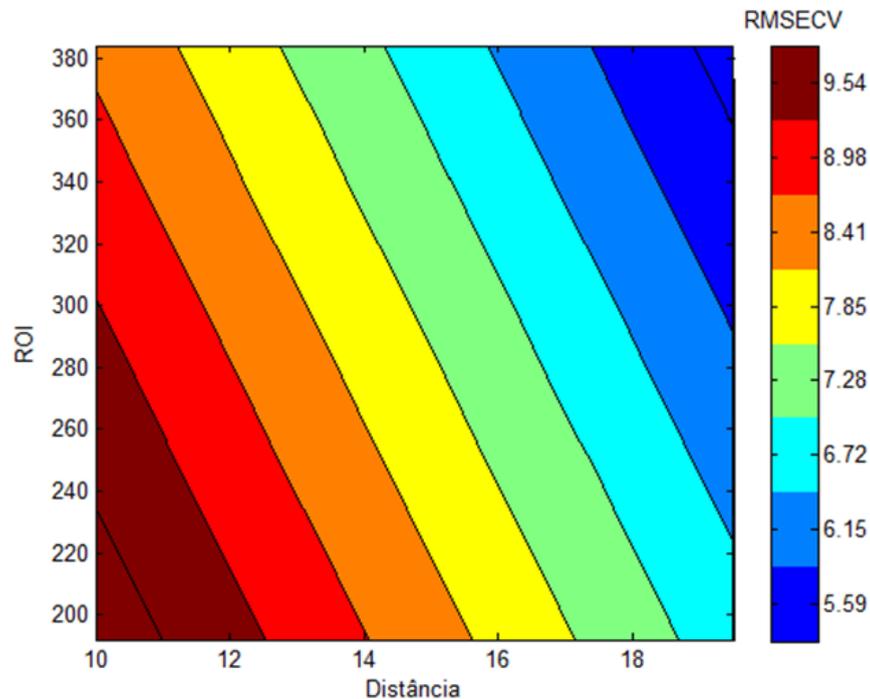
Fonte: Software Chemoface 1.65.

Como explicado anteriormente, a partir do Diagrama de Pareto é possível inferir sobre a influência de cada variável no erro de validação cruzada. O diagrama mostrado na Figura 10 informa que, para a Câmara A, tanto a distância como a ROI possuem influência significativa sobre o valor de RMSECV ($p_{\text{calc}} < p$). Além disso, conclui-se também que a referida influência é inversamente proporcional (p_{calc} com sinal negativo) para ambas as variáveis. Desta forma, quanto menor a região de interesse e a distância do celular até a amostra, maior o erro de validação cruzada.

Pela magnitude dos valores absolutos de p_{calc} para as variáveis estudadas é possível concluir que a distância do celular até a amostra possui uma maior influência sobre o valor de RMSECV do que a ROI, visto que, p_{calc} para a distância é maior. Neste caso, acredita-se que a ROI possua uma influência significativa no erro, pois, como a câmara de análise é consideravelmente menor, a imagem perde representatividade quando capturada através de uma ROI menor, visto que a quantidade de sementes abrangidas nela diminui substancialmente. Por fim, como também pôde ser concluído para a Câmara B, as variáveis distância e tamanho da ROI não proporcionam efeitos significativos entre si (p_{calc} para $X1*X2 < p$).

Na Figura 11 é observado o modelo de superfície linear construído com base nas variáveis avaliadas e sua influência para os valores de RMSECV.

Figura 11 - Modelo de superfície linear construído para a Câmara A.



Fonte: Software Chemoface 1.65.

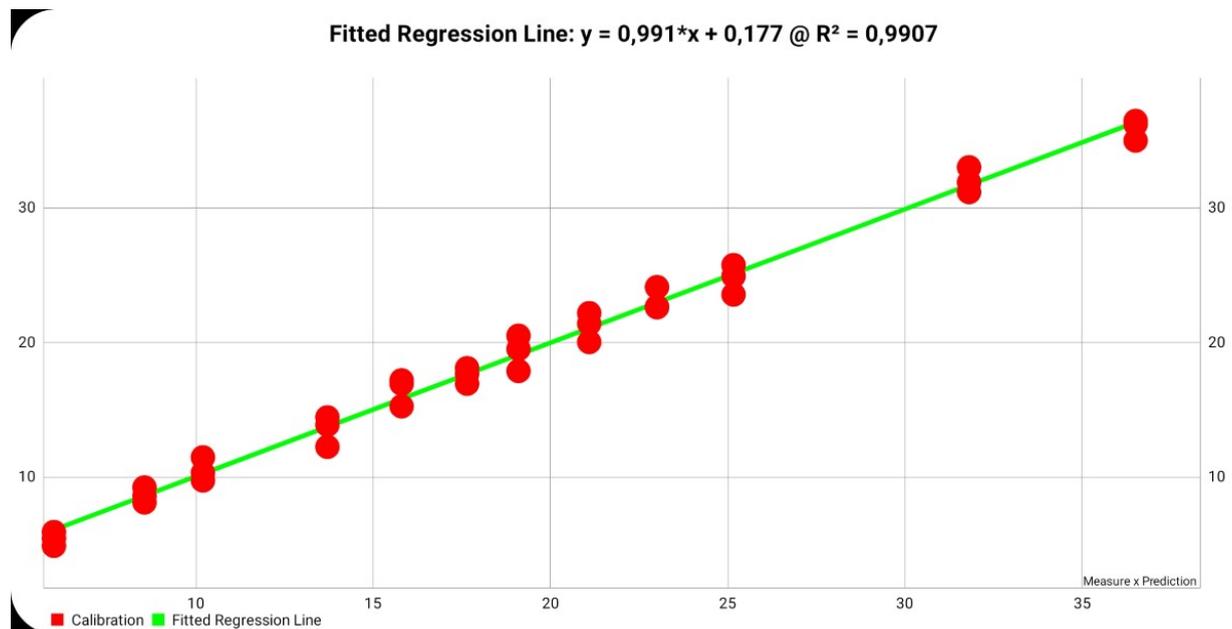
Através do modelo de superfície mostrado acima pode-se notar que, como mostrado no Teste t de Student, o erro sofre influência de ambas as variáveis. Além disso, pela escala de cores, conclui-se que de fato o erro diminui com o aumento da ROI e da distância do celular até a amostra. É possível concluir também que o menor erro de validação cruzada (5,1%) é obtido na região colorida de azul forte, que, por sua vez, corresponde a uma ROI acima de 340 e uma distância acima de 17,5 cm. A possibilidade de uma pequena variação na distância sem interferir no valor de erro é importante, pois, em certas condições cotidianas na indústria pode não ser possível fixar com precisão a distância do celular até a amostra durante o experimento.

Pelos resultados obtidos através dos experimentos nas Câmaras A e B, é possível notar que a Câmara A forneceu um menor valor de erro de validação cruzada. Ainda que esta diferença no erro não foi muito discrepante, ela foi levada em consideração juntamente com o fato de que a Câmara A é portátil e mais fácil de se

trabalhar. Portanto, a Câmara A foi escolhida como recipiente ideal para realizar a captura das imagens das sementes de milho.

Adicionalmente, a condição experimental escolhida foi a composta pela distância máxima da amostra até o celular (19,5 cm) e por um tamanho da ROI igual a 384 x 384. Isso porque, como provado anteriormente, esta condição testada no Experimento 1 foi a que forneceu o menor valor de RMSECV. A reta de regressão gerada no aplicativo pelo modelo com 5 variáveis latentes a partir dos resultados do Experimento 1 está mostrada na Figura 12. Este gráfico é dado pelo valor medido através do modelo em função do valor predito na calibração. Nota-se que com 5 variáveis latentes o modelo já foi capaz de prever de uma maneira efetiva a concentração de cada amostra de calibração, se assemelhando bastante com o valor fornecido e gerando um valor de R^2 muito próximo de 1,0.

Figura 12 - Modelo de regressão por mínimos quadrados parciais (PLS) construído no Experimento 1 através de 5 variáveis latentes.



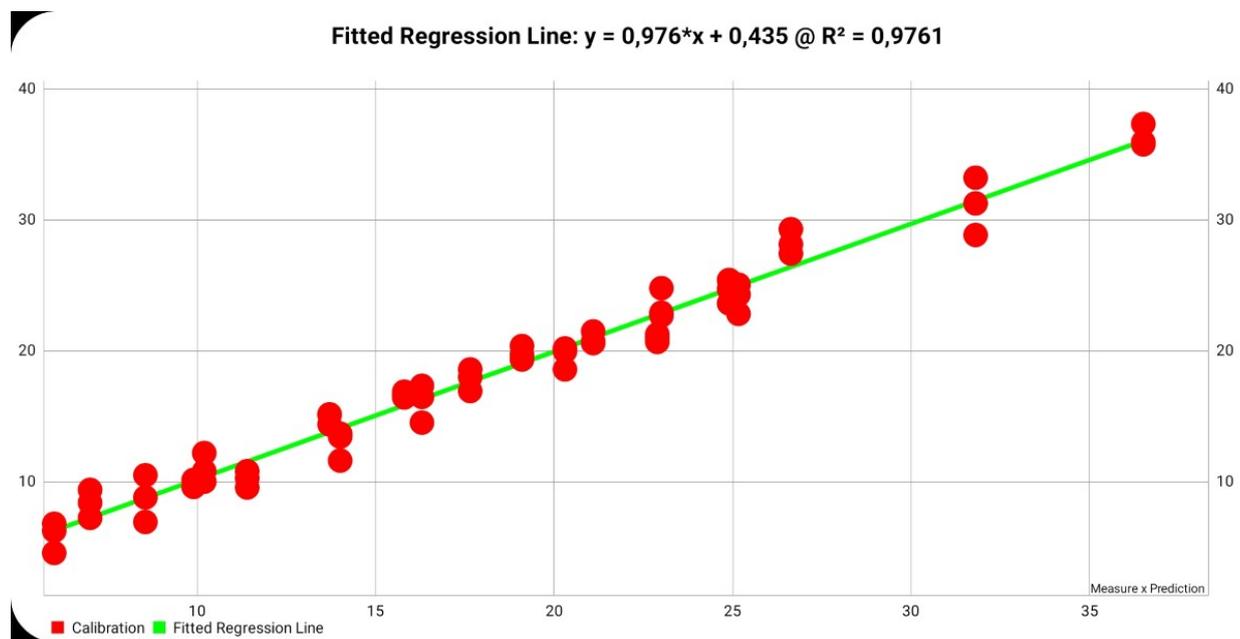
Fonte: Aplicativo *PhotoMetrix* PRO®.

5.2 CALIBRAÇÃO DO MODELO E PREDIÇÃO DAS AMOSTRAS

Na Câmara A e nas condições de maior ROI (384 x 384) e maior distância (19,5 cm) foi construído o modelo de calibração mostrado na figura abaixo. Este modelo foi construído com 7 variáveis latentes, que, por sua vez, forneceram o menor RMSECV.

O erro de validação cruzada obtido foi de 4,76%, já o erro de calibração foi de 1,23%. Além disso, ele apresentou um coeficiente de determinação igual a 0,976 indicando que, estas 7 variáveis escolhidas conseguiram gerar um modelo cuja leitura das amostras se aproximou muito ao valor real. A Figura 13 mostra a reta de regressão plotada no aplicativo que é dada pelo valor medido pelo modelo em função daquele informado durante a calibração.

Figura 13 - Modelo de regressão por mínimos quadrados parciais (PLS) construído através de 7 variáveis latentes durante a calibração com 21 amostras.



Fonte: Aplicativo *PhotoMetrix* PRO®.

Com o intuito de testar o modelo de calibração construído, foram selecionadas 7 amostras nas quais realizou-se a predição da concentração de sementes boas através do modelo. Estas amostras tinham as seguintes porcentagens de sementes de qualidade em meio ao rejeito: 6,50%, 10,51%, 10,90%, 16,50%, 20,60%, 21,20%, 21,40% e 23,27%.

Foi observado, entretanto, que o erro de predição (RMSEP), ou seja, a diferença entre o valor calculado através do modelo construído e o valor real, foi de 17,7%. Este erro alto foi gerado principalmente pelas amostras dos extremos de concentração que se comportaram como *outliers*, ou seja, as amostras de 6,50% e de 23,27%, como mostrado na Tabela 5.

Tabela 5 - Valores de concentração x Valores preditos pelo modelo (experimento com 7 amostras).

Amostra	Concentração (%)	Predição (%)	Erro Absoluto (%)
1	6,50	24,17	17,67
2	10,90	14,59	3,69
3	10,51	12,65	2,14
4	16,50	19,83	3,33
5	21,20	20,59	0,61
6	21,40	12,74	8,66
7	23,27	9,61	13,66

Assim, o experimento foi refeito, retirando-se estas duas amostras a fim de verificar se o modelo seria capaz de prever corretamente a concentração das demais. Com o novo experimento, obteve-se um valor de erro de predição (RMSEP) igual a 2,24%. A comparação entre os valores preditos e os valores reais de concentração, bem como o erro para cada amostra podem ser encontrados na Tabela 6.

Tabela 6 - Valores de concentração x Valores preditos pelo modelo (experimento com 5 amostras).

Amostra	Concentração (%)	Predição (%)	Erro Absoluto
1	10,51	12,75	2,24
2	10,90	14,89	3,99
3	16,50	14,35	2,15
4	21,20	19,97	1,23
5	21,40	17,95	3,45

Como conclusão dos resultados de predição mostrados, percebe-se que o modelo conseguiu traçar a relação esperada entre a intensidade da cor amarela da amostra com a concentração de sementes de qualidade contidas nela. Esta inferência

pode ser feita, pois, para uma amostra cuja concentração real de sementes boas era significativamente maior, o modelo conseguiu prever uma concentração mais alta.

A hipótese formulada para fontes de erro nos experimentos como um todo envolve um erro originado pela diferença dos parâmetros utilizados para a determinação de sementes de boa qualidade na indústria, a massa, e no *PhotoMetrix PRO®*, a cor. Estes valores de concentração utilizados para calibrar o modelo foram obtidos na indústria através de uma análise mássica. Assim, como explicado anteriormente, as sementes boas foram selecionadas visualmente em meio ao rejeito e pesadas, calculando-se a porcentagem das mesmas na amostra retirada.

Observou-se, entretanto, que as amostras não eram compostas apenas de sementes boas e ruins, mas também de outros materiais mais leves como pedaços de sabugos de milho secos. A amostra cuja concentração era de 23,27%, por exemplo, continha uma quantidade expressiva de sabugos. A presença de outros tipos de materiais mais leves é relevante, pois, ela implica que a suposição de que todas as amostras ditas com maior porcentagem de sementes de alta qualidade tenham de fato uma maior quantidade nominal das mesmas não seja totalmente verdade. Isso porque, caso a amostra seja composta por estes materiais mais leves com uma abundância significativa, o valor obtido para a quantidade de sementes boas através do cálculo mássico será incoerente. Ele será maior não porque há uma quantidade nominal maior de sementes boas, mas sim, porque estas sementes vão ter um peso muito significativo frente à amostra como um todo.

Portanto, é possível dizer que, através do cálculo mássico, uma amostra com maior porcentagem de sementes boas pode ter uma menor quantidade nominal das mesmas se comparada a uma amostra de porcentagem inferior. Esta observação gera erros nos experimentos realizados através do *PhotoMetrix PRO®*, pois, com esta ferramenta se faz uma análise superficial apenas em função da cor sendo inviável a noção de densidade. Ademais, supõem-se que uma amostra com maior concentração de sementes boas será mais amarelada. Entretanto, tendo em vista o que foi colocado acima, nem sempre essa suposição será verdade, gerando assim, uma fonte de erro.

Os resultados também mostram que, retirando as duas amostras que se comportaram como *outliers* obteve-se um erro absoluto de predição variando de aproximadamente 1,2 até 4%, sendo o erro médio (RMSEP) igual a 2,24%.

A fim de verificar se os valores de concentração medidos pelo modelo e aqueles preditos (valores fornecidos pela empresa) eram estatisticamente diferentes, foi feito o Teste t de Student Pareado no software GraphPad InStat 3.1. Obteve-se um valor de p_{calc} bicaudal igual a 0,9356 sendo, portanto, considerado não significativo quando $t = 0,08604$ com 4 graus de liberdade e com 95% de confiança. Assim, ainda que as diferenças entre os valores preditos e aqueles medidos pelo modelo sejam percentualmente altas, elas não são significativas. Portanto, é possível afirmar que os valores de concentração fornecidos pelo modelo de calibração construído são estatisticamente equivalentes aqueles preditos (fornecidos pela empresa).

Assim, estes valores de erro se mostram promissores para o fim no qual se deseja empregar a análise. Pois, além de os valores de concentração preditos e aqueles medidos pelo modelo de calibração não serem estatisticamente diferentes, a intenção é aplicar o modelo para avaliar o desempenho de classificação e separação das sementes realizadas pela máquina *Color Sorter* e pela Mesa de Gravidade. Para este fim, não é necessária uma grande precisão nos dados das análises visto que, o objetivo é descobrir se está saindo uma porcentagem maior ou menor de sementes boas em meio ao rejeito para que se possa ajustar os parâmetros das máquinas.

Desta forma, esta pequena diferença nominal entre os valores preditos pelo modelo e os valores reais, na prática da tomada de decisão para o ajuste dos equipamentos, não irá proporcionar mudança significativa. Uma prova desta afirmação pode ser inferida pelo modo com o qual a análise é feita atualmente. Como já mencionado, ela é visual e manual, realizada por operadores responsáveis, tornando-a subjetiva e imprecisa. Portanto, caso fosse necessária uma alta precisão, esta análise não poderia ser feita desta maneira, visto que a mesma gera muitos erros dependentes de fatores operacionais.

Na determinação dos limites de quantificação e detecção, obteve-se um desvio padrão de 1,02 entre as 10 predições da amostra com concentração de 6,0% de sementes boas. A partir deste desvio, é possível notar que o modelo de calibração construído é coerente em seus resultados, pois, para uma mesma amostra, gerou um desvio baixo entre os valores de concentração calculados. O limite de detecção obtido foi de 3,06% e o de quantificação foi de 10,20%.

Por fim, conclui-se que a metodologia proposta tem o potencial de contribuir para o processo industrial de produção de sementes. Isso porque ela é rápida, permitindo uma alta frequência analítica e uma reprogramação ágil das máquinas, e precisa, fornecendo valores muito similares para uma mesma amostra e dando fim à subjetividade existente atualmente. A análise que leva, atualmente, em média 15 minutos para ser realizada (por amostra), poderá ser feita em aproximadamente 1 minuto.

Além disso, também possui baixo custo, pois o maior investimento que a empresa necessitará fazer será apenas na aquisição de celulares e recipientes de análises, visto que o aplicativo é gratuito. Para uma planta industrial, sugere-se que a empresa adquira 5 conjuntos a fim de garantir equipamentos reserva em caso de imprevistos. O Quadro 1 fornece uma estimativa quanto ao investimento para aquisição do material necessário já considerando a contratação de um profissional para ministrar treinamentos para a equipe. Considerou-se um valor de R\$ 200,00 por hora de remuneração ao profissional e, à este valor devem ser acrescidas despesas de locomoção, estadia e alimentação

Quadro 1 - Estimativa dos custos de aquisição do material para implementação da metodologia proposta

Item	Valor unitário	Quantidade	Valor total
Câmara de análise + <i>smartphone</i>	R\$ 1300,00	5	R\$ 6500,00
Treinamento de 8h	R\$ 1600,00	1	R\$ 1600,00

Caso o produto se tornasse comercial e fosse vendido por uma empresa especializada, ele iria adquirir um valor agregado dado por melhorias e customizações do software. Entretanto, percebe-se que o custo total que a empresa produtora de sementes teria para implementar a metodologia seria ínfimo se comparado aos gastos usuais de compras de equipamentos e execuções de projetos.

É relevante salientar também que a empresa necessitará de uma menor mão de obra para a realização da análise do rejeito das máquinas já que ela será feita pelo aplicativo. Baseando-se na empresa concedente das amostras utilizadas neste estudo,

na planta industrial é alocado um funcionário por turno para analisar os rejeitos das máquinas *Color Sorter* e dois funcionários por turno, para as mesas de gravidade. Caso a implementação da metodologia proposta seja realizada, estima-se que seria necessário apenas um operador por turno para analisar o rejeito de todas as máquinas. Assim, a empresa deixaria de investir 2 operadores por turno nesta tarefa, diminuindo os custos destinados à ela e possibilitando a realocação destes funcionários dentro da indústria.

Finalmente, a quantidade de sementes de milho de alta qualidade desperdiçadas também será menor, pois, como dito, a análise do rejeito será rápida e a máquina poderá ser prontamente reprogramada. Portanto, a metodologia proposta é uma ferramenta de facilitação na coleta de informações e no ajuste de desempenho dos equipamentos.

6. CONCLUSÃO

A partir do trabalho desenvolvido conclui-se que os objetivos propostos foram atingidos. As condições experimentais ideais foram determinadas através da etapa de otimização e um recipiente de análise adequado pôde ser construído e validado. Adicionalmente, a metodologia proposta para avaliação da quantidade de sementes de qualidade em meio ao rejeito de máquinas industriais utilizando o aplicativo *PhotoMetrix PRO®* mostrou ter um grande potencial de aplicação.

Este modelo de calibração construído através da regressão por PLS conseguiu relacionar a cor amarela das amostras com suas respectivas concentrações de sementes boas gerando uma alta correlação entre os dados. O modelo também foi capaz de prever a concentração das sementes de qualidade em meio às amostras de rejeito externas à ele apenas com uma imagem das mesmas gerando um erro aceitável para a finalidade de aplicação. Adicionalmente, os valores de concentração fornecidos pelo modelo calibração mostraram-se estatisticamente equivalentes àqueles obtidos pela empresa.

Como perspectiva futura deste trabalho pode-se realizar estudos a partir dos resultados obtidos a fim de corrigir as limitações do modelo para determinação em amostras com baixas concentrações. Sugere-se também que seja corrigido o erro originado pela diferença dos parâmetros utilizados para a determinação de sementes

de boa qualidade na indústria, a massa, e no *PhotoMetrix* PRO®, a cor. Além disso, os resultados obtidos no aplicativo também podem ser rodados em softwares específicos como o ChemoStat a fim de realizar alguns testes manipulando os histogramas RGB e aplicando pré-processamentos dos dados com o intuito de verificar se ocorrem melhoras no modelo de calibração construído. Estas melhorias poderão servir como ponto de partida para novos testes experimentais e parâmetros de avaliação.

REFERÊNCIAS:

BERNARDES, C. D. **Métodos Analíticos para Cachaças Utilizando Técnicas Espectrométricas, Cromatográficas e Quimiométricas**. 2015. 144 p. Tese (Doutorado em Ciências – Química) - Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

BÖCK, F. C.; HELFER, G.A; COSTA, A. B. da; DESSUY, M. B.; FERRÃO, M. F. **PhotoMetrix and colorimetric image analysis using smartphones**. *Journal of Chemometrics*. 2020;e3251. <https://doi.org/10.1002/cem.3251>

BÖCK, F. C.; HELFER, G.A; COSTA, A. B. da; DESSUY, M. B.; FERRÃO, M. F. **Rapid determination of ethanol in sugarcane spirit using partial least squares regression embedded in smartphone**. *Food Anal Methods*. 2018;11(7):1951–1957. <https://doi.org/10.1007/s12161-018-1167-4>

BUSCH, D. D. **Scanner para IBM - PC**. Rio de Janeiro: Livros Técnicos e Científicos, 1993.

CARVALHO, N. M. de; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. 4. ed. rev. e ampl. Jaboticabal: FUNEP, 2000. 588 p.

CRUZ, J. C. et al. **Boas Práticas Agrícolas: Milho**. Embrapa Milho e Sorgo. Sete Lagoas, Documentos, 119, nov. 2011, 48 p. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/67645/1/doc-119.pdf>. Acesso em: 7 out. 2021.

DUARTE, J. de O. et al. **Sustentabilidade da cadeia produtiva do milho**. Embrapa Milho e Sorgo. Sete Lagoas, Documentos, 261, maio. 2021, 27 p. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/223515/1/Doc-261.pdf> . Acesso em: 2 set. 2021.

DUARTE, J. de O.; GARCIA, J. C; MATTOSO, M.J. **Benefícios Econômicos do Uso da Cultivar de Milho Híbrido BR 201**. Embrapa Milho e Sorgo. Sete Lagoas, Documentos, 54, dez. 2006, 14 p. Disponível em: http://ainfo.cnptia.embrapa.br/digital/bitstream/CNPMS/19650/1/Doc_54.pdf . Acesso em: 2 set. 2021.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. **Quimiometria I: Calibração Multivariada, um Tutorial**. Química Nova, v. 22, nº 5, p. 724-731, 1999.

FERREIRA, R. L. **Etapas do beneficiamento na qualidade física e fisiológica de sementes de milho**. 2010. 49 p. Dissertação (Mestrado em Agronomia) - Universidade Estadual Paulista, Ilha Solteira, 2010.

FIDÊNCIO, P.H. **Análise de solos por espectroscopia no infravermelho próximo e aplicação de métodos quimiométricos**. 2001. Tese (Doutorado em Ciências) Universidade de Campinas, Campinas, SP, 2001.

HAIR JR., J. F.; ANDERSON, R.; TATHAM, R. L.; BLACK, W. C., **Análise Multivariada de Dados – 5ª Edição – Bookman – 2005 – 593 p.**

HELPER, G. A.; MAGNUS, V. S.; BÖCK, F. C.; TEICHMANN, A.; FERRÃO, M. F.; COSTA, A. B. da. **PhotoMetrix: an application for univariate calibration and principal components analysis using colorimetry on mobile devices**. *Journal Of The Brazilian Chemical Society*, v. 28, n.2, p. 328-335, 2017.

HELPER, G. A.; TISCHER, B.; FILODA, P. F.; et al. **A new tool for interpretation of thermal stability of raw milk by means of the alizarol test using a PLS model on a mobile device**. *Food Anal Methods*. 2018; 11(7):2022-2028. <https://doi.org/10.1007/s12161-018-1190-5>

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. Prentice Hall, New Jersey, USA, 1998, 816 p.

LAPOLLI, E. M. **Processamento de imagens digitais: uma abordagem utilizando conjuntos difusos**. 1994. 206 p. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 1994.

LEÃO, A. C. **Gerenciamento de cores para imagens digitais**. 2005. 135 p. Dissertação (Mestrado em Artes Visuais) - Escola de Belas Artes da Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

LÓPEZ, A. F. J. ; SALAMANCA, J. M.; MEDINA, M. J. Q.; PÉREZ, O. E. A. **Diagnóstico de Cultivos Utilizando Procesamiento Digital de Imágenes y Tecnologías de Agricultura de Precisión**, INGE CUC, vol. 11, no. 1, pp. 63-71, 2015. DOI: <http://dx.doi.org/10.17981/ingecuc.11.1.2015.06>

MAGNUSSON, B.; ÖRNEMARK, U. (eds.) **Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics**. 2ª ed. 2014. ISBN 978-91-87461-59-0. Disponível em: www.eurachem.org. Acesso em: 19 out. 2021

MACHADO, A. W.; SOUKI, B. Q. **Simplificando a obtenção e a utilização de imagens digitais - scanners e câmeras digitais**. Revista Dental Press de Ortodontia e Ortopedia Facial, Maringá, v. 9, n. 4, p. 133-156, jul./ago. 2004.

MAZURKIÉVICZ, J. **Influência das etapas do beneficiamento na qualidade fisiológica e física de sementes de milho (Zea Mayz L.)**. 2011. 43 p. Trabalho de conclusão de curso (Bacharelado em Agronomia) – Universidade Federal de Santa Catarina, Florianópolis, 2011.

NUNES, L. C. **Chemometrics methods applied in the spectrophotometric determination of amino acid mixtures**. 2007. 128 p. Dissertação (Mestrado em Agroquímica analítica; Agroquímica inorgânica e Físico-química; Agroquímica orgânica) - Universidade Federal de Viçosa, Viçosa, 2007.

OLIVEIRA, L. F. de. **Uso de imagens digitais e ferramentas quimiométricas para identificação e determinação de Cr (IV) em solos contaminados**. 2012. 72 p. Dissertação (Mestrado em Química) – Universidade Federal de São Carlos, São Carlos, 2012.

PAES, M. C. D. **Aspectos Físicos, Químicos e Tecnológicos do Grão de Milho. Embrapa Milho e Sorgo**. Sete Lagoas, Circular Técnica 75, dez. 2006, 6 p. Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/489376/1/Circ75.pdf>. Acesso em: 7 out. 2021.

QUINTANILHA, J. A. **Processamento de Imagens Digitais**. Anais do Simpósio Brasileiro de Geoprocessamento. São Paulo, 23-25, Maio de 1990. p. 37-52.

SHARAF, M. A.; IIIMAN, D. L.; KOWALSKI, B. R. **Chemometrics**. Nova Iorque, John Wiley & Sons, 1986.

SOUZA, A. M. de; BREITKREITZ, M.C.; POPPI, R. J. **Experimento didático de quimiometria para planejamento de experimentos: avaliação das condições experimentais na determinação espectrofotométrica de ferro II com o-fenantrolina**. Um tutorial, parte III. *Quim. Nova*, Vol. 37, nº. 3, 564-573, 2014.

TOLEDO, F. F. ; MARCOS FILHO, J. **Manual de sementes - tecnologia da produção**. São Paulo: Agronômica Ceres, 1977. 224 p.

TRIGO, T. **Equipamento fotográfico: teoria e prática**. 2. ed. São Paulo: Senac, 2003.

VELESACA, H. O.; MIRA, R.; SUAREZ, P. L.; LARREA, C. X.; SAPPA, A. D. **Deep learning based corn kernel classification**. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June*. <https://doi.org/10.1109/CVPRW50498.2020.00041>

VIEIRA, R. A. **Avaliação de perdas no processo de beneficiamento de milho para semente**. 2013. 46 p. Trabalho final de estágio supervisionado - Faculdade de Agronomia e Medicina Veterinária - Universidade de Brasília, Brasília, 2013.

VILLELA, F.A ; PERES, B. W. Coleta, beneficiamento e armazenamento. In: FERREIRA, A.G; BORGHETTI, F. **Germinação: do básico ao aplicado**. Porto Alegre: ARTMED, 2004. 323 p.

WOLD, S.; SJÖSTRÖM, M; ERIKSSON, L. **PLS-regression: a basic tool of chemometrics**. *Chemometrics and Intelligent Laboratory Systems*, 58 (2001) 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

YUMANG, A.; MAGWILI, G.; MONTOYA, S. K. C.; ZALDARRIAGA, C. J. G. **Determination of Shelled Corn Damages using Colored Image Edge Detection with Convolutional Neural Network**. *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2020*. <https://doi.org/10.1109/HNICEM51456.2020.9400023>