



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
CURSO DE QUÍMICA INDUSTRIAL**

INGRID KELLY TELES DE FARIAS

**DIAGNÓSTICO RÁPIDO PARA DETECÇÃO DE SEMENTES
DE ALGODÃO TRANSGÊNICAS UTILIZANDO IMAGENS HIPERESPECTRAIS**

**CAMPINA GRANDE - PB
2020**

INGRID KELLY TELES DE FARIAS

**DIAGNÓSTICO RÁPIDO PARA DETECÇÃO DE SEMENTES DE ALGODÃO
TRANSGÊNICAS UTILIZANDO IMAGENS HIPERESPECTRAIS**

Trabalho de Conclusão de Curso (TCC) apresentado à Coordenação do Curso de Química Industrial do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Química Industrial.

Orientadora: Profa. Dra. Simone da Silva Simões

**CAMPINA GRANDE - PB
2020**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

F224d Farias, Ingrid Kelly Teles de.

Diagnóstico rápido para detecção de sementes de algodão transgênicas utilizando imagens hiperespectrais [manuscrito] / Ingrid Kelly Teles de Farias. - 2020.

32 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Química Industrial) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2020.

"Orientação : Profa. Dra. Simone da Silva Simões ,
Coordenação do Curso de Química Industrial - CCT."

1. Organismos Geneticamente Modificados. 2.
Quimiometria. 3. Química analítica. I. Título

21. ed. CDD 660

INGRID KELLY TELES DE FARIAS

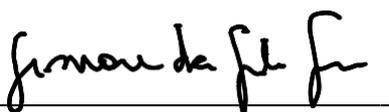
DIAGNÓSTICO RÁPIDO PARA DETECÇÃO DE SEMENTES DE ALGODÃO
TRANSGÊNICAS UTILIZANDO IMAGENS HIPERESPECTRAIS

Trabalho de Conclusão de Curso (TCC)
apresentado à coordenação do Curso de
Química do Centro de Ciências e
Tecnologia da Universidade Estadual da
Paraíba, como requisito parcial à
obtenção do título de bacharel em
Química Industrial.

Área de concentração: Química Analítica.

Aprovado em: _16_/_outubro_/_2020_____.

BANCA EXAMINADORA



Profa. Dra. Simone da Silva Simões (Orientadora)
Universidade Estadual da Paraíba (UEPB)



MSc. Luciano Bernardo Ramo
Universidade Federal da Paraíba (DQ/UFPB)



Prof. Dr. José Germano Veras Neto
Universidade Estadual da Paraíba (DQ/UEPB)

AGRADECIMENTOS

Primeiramente, quero agradecer a Deus e a Virgem Maria por iluminarem o meu caminho e me abençoar durante toda a minha vida.

Agradeço a minha família, em especial aos meus pais, Jailson e Lône pelo apoio, pelas palavras e por sempre me incentivarem a realizar meus sonhos.

Obrigada a minha avó Áurea e aos meus tios Altair e Airton, que desde o começo me apoiaram para eu ter uma ótima educação e motivaram a me tornar uma pessoa independente.

Obrigada a minha tia Silvia e a minhas primas Giovana e Gabriela, que me acolheram em sua casa e participaram de todo o processo de choro, noites sem dormir e desespero para eu estar aqui hoje.

Obrigada a professora Verônica Evangelista, que contribuiu ao longo da graduação de forma tão prestativa e amiga com ensinamentos que vou levar para a vida. E obrigada a todos os docentes que contribuíram para o meu desenvolvimento na UEPB.

Obrigada aos meus colegas de classe, em especial Juliana, Kátia, Rayane, Amanda, Diego e Izabelle pelos momentos de amizade e apoio. Vocês deixaram essa graduação mais leve.

Obrigada a Priscilla pela colaboração para esse trabalho dar certo e a Juliana pela amizade e parceria em todos os momentos de desenvolvimento desse projeto.

Obrigada a Joabson, Edjane e Dr. Everaldo que foram sempre muito prestativos e pacientes na Embrapa Algodão para que tudo fosse bem executado.

Por fim, mais não menos importante quero agradecer a minha orientadora professora Dra. Simone Simões pela oportunidade de trabalhar com ela, pelo apoio, paciência e dedicação para que tudo desse certo. Uma profissional incrível e um exemplo de mulher a ser seguido.

RESUMO

Organismos Geneticamente Modificados (OGM's) são organismos cujo material genético (DNA/RNA) tenha sido modificado por uma técnica de engenharia genética. Assim, para um melhor desenvolvimento de cultivares, as sementes de algodão geneticamente modificadas têm por objetivo uma maior eficiência na produtividade além de resistência a pragas. Por existir ainda uma certa insegurança sobre o uso e efeitos de alguns OGM's no organismo humano e animal, seu uso é regulado por órgãos responsáveis. Os métodos convencionalmente utilizados para identificação de OGM's geralmente são caros, lentos, destrutivos e pouco acessíveis. Uma técnica que pode ser utilizada como alternativa a estes são as imagens hiperespectrais na região do infravermelho próximo (HI-NIR) associada a técnicas de reconhecimento de padrões multivariadas. Essas metodologias são rápidas, precisas e não-destrutivas. Deste modo, este trabalho teve por objetivo desenvolver um método baseado em HI-NIR e métodos quimiométricos para a distinção entre sementes de algodão transgênicas e convencionais. Para isso, foram registradas as imagens de 40 sementes de algodão de dois cultivares, sendo 20 sementes BRS Aroeira (convencional) e 20 sementes BRS 368 RF (transgênica), respectivamente. O registro das imagens foi realizado separadamente para as duas classes. Em seguida, as imagens foram tratadas retirando ruídos ou qualquer outra característica que pudesse prejudicar o modelo utilizando técnicas de pré-processamento. A construção do modelo PLS-DA foi realizada a partir 26 sementes (13 de cada classe) do conjunto de treinamento, selecionadas de forma aleatória. A validação do modelo foi realizada com as 14 amostras do conjunto de teste (7 de cada classe). De acordo com as figuras de mérito, sensibilidade e seletividade, concluiu-se que o modelo HI-PLS-DA apresenta um bom desempenho para a distinção entre as amostras convencionais e transgênicas de forma rápida, preservando as amostras.

Palavras-Chave: Organismos Geneticamente Modificados. Quimiometria. Química Analítica.

ABSTRACT

Genetically Modified Organisms (GMOs) are organisms whose genetic material (DNA/RNA) has been modified by a genetic engineering technique. Thus, for a better development of cultivars, genetically modified cotton seeds aim at greater efficiency in productivity in addition to resistance to pests. As there is still some uncertainty about the use and effects of some GMOs in the human and animal organisms, their use is regulated by responsible bodies. The methods conventionally used to identify GMOs are generally expensive, slow, destructive and inaccessible. One technique that can be used as an alternative to these are hyperspectral images in the near infrared region (HI-NIR) associated with multivariate pattern recognition techniques. These methodologies are fast, accurate and non-destructive. Thus, this work aimed to develop a method based on HI-NIR and chemometric methods to distinguish between transgenic and conventional cotton seeds. For that, images of 40 cotton seeds from two cultivars, BRS Aroeira and BRS 368 RF, were recorded, 20 conventional and 20 transgenic, respectively. The images were registered separately for the two classes. Then, the images were treated by removing noise or any other characteristic that could harm the model using pre-processing techniques. The construction of the PLS-DA model was carried out from 26 seeds (13 of each class) from the training set, selected at random. The validation of the model was performed with the 14 samples from the test set (7 from each class). According to the figures of merit, sensitivity and selectivity, it was concluded that the HI-PLS-DA model performs well to distinguish between conventional and transgenic samples quickly, preserving the samples.

Keywords: Genetically Modified Organisms. Chemometry. Analytical chemistry.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Matriz de dados de imagens (a) em escalas de cinza, (b) em RGB e (c) hiperespectrais..... | 14 |
| Figura 2 - Esquema da construção de um modelo PLS-DA..... | 17 |
| Figura 3 - A) Sementes BRS 368 RF e B) Sementes Aroeiras..... | 19 |
| Figura 4 - Espectrômetro sisuChema SWIR - Specim®,..... | 20 |
| Figura 5 - Imagem média da semente de algodão A) transgênica e B) convencional..... | 22 |
| Figura 6 - Seleção da região para montagem da matriz de dados..... | 23 |
| Figura 7 - Organização da matriz de treinamento..... | 23 |
| Figura 8 - Espectros brutos..... | 23 |
| Figura 9 - Espectros pré-processados..... | 24 |
| Figura 10 - a) Curva ROC e b) limiar de separação entre as classes para a etapa de treinamento..... | 24 |
| Figura 11 - Classificação das amostras de treinamento..... | 25 |
| Figura 12 - a) Curva ROC e b) limiar de separação entre as classes para a etapa de teste..... | 26 |
| Figura 13 - Classificação das amostras treinamento e de teste..... | 26 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-----------|---|
| DNA | Ácido Desoxirribonucleico (<i>Deoxyribonucleic acid</i>) |
| ELISA | Ensaio de Imunoabsorção Enzimático (<i>Enzyme-Linked Immunosorbent Assay</i>) |
| EMBRAPA | Empresa Brasileira de Pesquisa Agropecuária (<i>Brazilian Agricultural Research Corporation</i>) |
| HI | Imagens Hiperespectrais (<i>Hyperspectral Images</i>) |
| HI-NIR | Imagens Hiperespectrais no Infravermelho Próximo (<i>Near Infrared Image Hyperspectroscopy</i>) |
| IFL | Imunoensaio de Fluxo Lateral (<i>Lateral Flow Immunoassay</i>) |
| LATECQ | Laboratório Avançado de Tecnologia Química (<i>Advanced Laboratory of Chemical Technology</i>) |
| MCR-ALS | Resolução de Curva Multivariada - Mínimos Quadrados Alternados (<i>Multivariate Curve Resolution - Alternating Least Squares</i>) |
| MIA | Análise de Imagem Multivariada (<i>Multivariate Image Analysis</i>) |
| OGM | Organismo Geneticamente Modificado (<i>Genetically Modified Organism</i>) |
| PCA | Análise de Componentes Principais (<i>Principal Component Analysis</i>) |
| PCR | Reação em Cadeia Polimerase (<i>Polymerase Chain Reaction</i>) |
| PCs | Componentes Principais (<i>Principal Component</i>) |
| PLS | Mínimos Quadrados Parciais (<i>Partial Least Squares</i>) |
| PLS-DA | Análise Discriminante de Mínimos Quadrados Parciais (<i>Partial Least Squares – Discriminant Analysis</i>) |
| RGB | Vermelho, Verde e Azul (<i>Red, Green and Blue</i>) |
| RNA | Ácido Ribonucleico (<i>Ribonucleic acid</i>) |
| ROC | Características Operacionais do Receptor (<i>Receiver Operating Characteristic</i>) |
| SIMPLISMA | <i>Simple-to-use Interactive Self-modeling Mixture Analysis</i> |
| SNV | Variação Normal Padrão (<i>Standard Normal Deviation</i>) |
| VL | Variáveis Latentes |

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO | 9 |
| 2 | OBJETIVOS | 11 |
| 2.1 | Objetivo geral | 11 |
| 2.2 | Objetivos específicos | 11 |
| 3 | REFERENCIAL TEÓRICO | 12 |
| 3.1 | Metodologia convencionais para identificação de transgenia | 12 |
| 3.2 | Imagens Hiperespectrais | 13 |
| 3.3 | Reconhecimento de Padrões | 14 |
| <i>3.3.1</i> | <i>Análise de Componentes Principais (PCA)</i> | <i>15</i> |
| <i>3.3.2</i> | <i>PLS-DA</i> | <i>16</i> |
| <i>3.3.3</i> | <i>Curva ROC</i> | <i>18</i> |
| 4 | MATERIAL E MÉTODOS | 19 |
| 4.1 | Local | 19 |
| 4.2 | Amostras | 19 |
| 4.3 | Preparo das amostras | 19 |
| 4.4 | Equipamento | 20 |
| 4.5 | Registro das imagens | 20 |
| 4.6 | Construção do modelo PLS-DA | 21 |
| 4.7 | Tratamento das imagens | 21 |
| 5 | RESULTADOS E DISCUSSÃO | 22 |
| 6 | CONCLUSÃO | 28 |
| | REFERÊNCIAS | 29 |

1 INTRODUÇÃO

O algodoeiro é uma angiosperma da classe das dicotiledôneas, de ordem Malvales, família das Malvaceae, gênero *Gossypium*. O algodão colhido é denominado em caroço e é composto pela pluma (fibra) e pelo caroço (sementes com “línter”, ou seja, fibras curtas). São utilizados em indústrias de fiação, tecelagem, na indústria de alimentação animal (farelo) e humana (óleo), além de grande número de produtos secundários (PENNA, 2005). O algodão (*Gossypium hirsutum* L.) é uma das culturas de fibras mais importantes do mundo. Todos os anos, em média 35 milhões de hectares de algodão são plantados mundialmente. Atualmente, em 60 países, nos cinco continentes ocorre a produção de algodão (ABRAPA, 2020). O Brasil está na posição de quinto maior produtor de algodão do mundo, além de ser um dos maiores exportadores mundiais de pluma (EMBRAPA, 2019).

A qualidade da semente do algodão é um dos importantes fatores que condicionam o sucesso da cultura no campo. O controle de sua qualidade envolve tanto ações governamentais quanto do setor privado (ALMEIDA, 2012).

Atualmente a transgenia é um dos temas importantes que se destacam por apresentar múltiplas aplicações no setor agrícola. Ela tem se desenvolvido ao longo dos anos e consiste no desenvolvimento de organismo que recebe um gene de outro organismo doador. Este termo pode ser utilizado como sinônimo para a expressão "Organismo Geneticamente Modificado" (OGM). Na agricultura, esta técnica tem sido desenvolvida com o objetivo de plantas mais resistentes a pragas específicas e adaptadas a cultivos em diferentes zonas climáticas (PETRUCCI, 2011).

Com o desenvolvimento de novas tecnologias, os OGMs geram opiniões diversas, pela falta de informação dos danos que ela pode causar aos seres humanos, animais e na natureza. Assim, os órgãos, governos e cientistas vem criando medidas para detecção de OGMs nos produtos, conseqüentemente, gerando maior confiabilidade para o consumidor. (BOCCIA, 2015; VARGAS *et al* 2018).

Devido à importância quarentenária da ramulose no Brasil, é salutar o desenvolvimento de métodos rápidos, baratos e eficientes que permitam a distinção dos agentes etiológicos da antracnose e da ramulose do algodoeiro. A utilização de imagens hiperespectrais associadas a técnicas de reconhecimento padrão e de classificação podem ser uma alternativa para identificação rápida e acurada desses

patógenos. (SIRISOMBOON, C. D.; PUTTHANG, R.; SIRISOMBOON, P., 2013). As imagens hiperespectrais possuem uma vasta aplicação em diversas áreas de conhecimento. Na área de fármacos, pode ser utilizada para analisar a uniformidade da distribuição de diversos comprimidos comerciais (CRUZ, J.; BLANCO, M., 2011; CRUZ *et al.*, 2009). Na indústria alimentícia, pode ser utilizada para controle de qualidade, como a detecção de defeitos e contaminações na superfície da casca de diferentes tipos de maçãs (MEHL *et al.*, 2004).

As técnicas hiperespectrais se baseiam na identificação de ligações químicas em uma molécula e na obtenção de espectros de uma gama de compostos sem a utilização de reagentes e com a preservação da constituição da amostra analisada. A hiperespectroscopia de imagem no infravermelho próximo, utilizada neste estudo, por exemplo, é uma imagem de camadas múltiplas, em que cada camada é um único comprimento de onda espectral, o qual é utilizado para detectar características em função das "impressões digitais", devido às ligações químicas específicas em diferentes materiais (LORENTE *et al.*, 2012). Assim, essas múltiplas camadas espectrais de uma imagem geram informações de alta complexidade (DONEUS *et al.*, 2014), que associadas a métodos de análises multivariadas para reconhecimento de padrão e de classificação podem ser úteis em estudos de identificação/separação de difícil diferenciação por métodos tradicionais.

Portanto, esse trabalho tem como objetivo a utilização de imagens hiperespectrais juntamente com método quimiométrico, como o PLS-DA (Análise Discriminante de Mínimos Quadrados Parciais) para a distinção dos genótipos convencionais e transgênicas das sementes de algodão de forma rápida e precisa.

2 OBJETIVOS

2.1 Objetivo geral

Esse trabalho tem como objetivo geral, desenvolver um método baseado em imagens hiperespectrais e método quimiométrico para reconhecimento de padrões, que serão capazes de diferenciar os genótipos convencionais e transgênicos das sementes de algodão.

2.2 Objetivos específicos

- Registrar de imagens das sementes de algodão convencionais e transgênicos utilizando HSI-NIR;
- Construir um modelo PLS-DA capaz de classificar corretamente as sementes transgênicas e convencionais;
- Escolher para o melhor pré-tratamento espectral;
- Avaliar a capacidade da técnica de imagem hiperespectral aliadas a técnica de reconhecimento de padrões supervisionada (PLS-DA) quanto a distinção entre as sementes convencionais e transgênicas;
- Avaliar a capacidade dos modelos PLS-DA criados.

3 REFERENCIAL TEÓRICO

3.1 Metodologias convencionais para identificação de transgenia

Os OGMs se consolidaram como a tecnologia mais rápida adotada na história recente da agricultura moderna (ISAAA, 2018), com o crescimento de aprovações para o plantio se dando na mesma velocidade, mas para que um organismo transgênico seja liberado para comercialização o mesmo deve passar por rigorosos testes (BRASIL, 2005), para isso mais de 120 instituições públicas e privadas foram credenciadas junto aos órgãos reguladores, para desenvolver pesquisas referentes aos OGMs, e nesses últimos 20 anos de uso de alimentos transgênicos em larga escala, não foram registrados nenhum impacto negativo no meio ambiente ou na saúde humana e animal (EMBRAPA, 2017). Apesar de todos esses fatos a opinião pública sobre o tema é ainda muito divergente. Essa detecção se faz necessária, pois permite verificar algum tipo de contaminação em diversos cultivos expostos a riscos biológicos e com implicações econômicas, sociais e ambientais, também se faz necessários para que o consumidor tenha informações adequadas sobre os produtos que estão adquirindo, e para gerarem confiança nesses alimentos (CONCEIÇÃO, F.R.; MOREIRA, A.N.; BINSFELD, P.C., 2006).

Alguns modos de detecção dos OGMs são caracterizados pela presença de um ou mais segmentos de DNA exógenos, podendo ou não acomodar a expressão de novas proteínas. Sendo assim, a detecção destes organismos é baseada na sequência do DNA exógeno ou na proteína transgênica (PETIT *et al.*, 2003). Outros métodos de detecção são baseados na presença de proteínas como os bioensaios e os imunoenaios. Os bioensaio tem algumas limitações, como a necessidade de um longo tempo para a obtenção dos resultados e a utilização restrita à OGMs resistentes à herbicidas (TORRES *et al.*, 2003). Segundo Stave (1999), os imunoenaios são ideais para a detecção qualitativa e quantitativa de proteínas em misturas complexas. É baseado na concentração típica de proteínas transgênicas em tecidos vegetais (> 10µg. g⁻¹ de tecido). O limite de detecção de um imunoensaio é de aproximadamente 1%, no entanto, nem sempre é possível a detecção de OGMs por este método. Dentre os métodos de imunoenaios pode-se destacar o Ensaio Imunoenzimático (ELISA, do inglês *Enzyme-Linked*

Immunosorbent Assay), o Imunoensaio de Fluxo Lateral (IFL, do inglês *Lateral Flow Immunoassay*) e o *Western blot*, que são bastantes utilizados na detecção e quantificação de alimentos contendo OGMs (MAGIN, 2000).

Existem também métodos de detecção baseados na presença de DNA, já que geralmente pode-se estabelecer uma correlação linear entre a quantidade de OGM e DNA exógeno quando a modificação genética é nuclear. No entanto, em se tratando de modificação genética não nuclear, como é o caso de quando o DNA exógeno é introduzido em uma organela como o cloroplasto, essa correlação não é possível. Até o momento, todos os OGMs comercializados possuem modificação genética nuclear (CONCEIÇÃO *et al.*, 2004). O principal método utilizado na detecção e quantificação de alimentos contendo OGMs é a Reação em Cadeia da Polimerase (PCR, do inglês *Polymerase Chain Reaction*), que consiste na amplificação seletiva de sequências específicas da molécula de DNA. Este método é sensível, específico, seguro, capaz de detectar uma ampla série de eventos (BERTHEAU *et al.*, 2002; GIOVANNINI T.; CONCILLO, L., 2002), e distinguir as variedades geneticamente modificadas que apresentam diferentes construções gênicas, porém expressam a mesma proteína (YAMAGUCHI *et al.*, 2003). As principais limitações do PCR são as dificuldades na construção dos indicadores, pois suas informações sobre a sequência da modificação genética geralmente são confidenciais (HOLST-JENSEN, A. *et al.*, 2003), necessidade de pessoal treinado e de equipamentos especiais, custo elevado, já que cada teste é específico para uma única modificação genética, necessitando também de material de referência certificado, cuja disponibilidade geralmente é limitada (MIRAGLIA *et al.*, 2004).

Outras formas de detecção e quantificação de OGMs vem sendo desenvolvidos, essas metodologias tem sensibilidade, confiabilidade, custos mais baixos, com certa rapidez quando comparadas aos métodos convencionais. Dentre esses novos métodos vem se destacando a cromatografia, a espectrometria de massa, a espectrometria no infravermelho próximo e a eletroforese capilar (OBEID *et al.*, 2004).

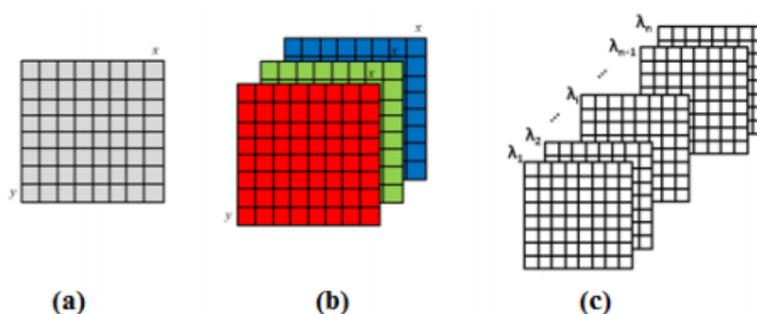
3.2 Imagens Hiperespectrais

Em imagens digitais, o menor elemento formador é o pixel (para imagens de superfície, ou seja, em duas dimensões) ou o voxel (para imagens tridimensionais).

Cada pequena porção da imagem está determinada por coordenadas espaciais – x e y , para os pixels e x , y e z , para os voxels. (PRATS-MONTALBÁN; DE JUAN; FERRER, 2011). As imagens digitais são formadas pelas coordenadas espaciais e valores numéricos para cada pixel que representam a escala de cores utilizada. Quando a imagem estudada está representada na escala de cinza, tem-se uma matriz de dados bidimensional e cada pixel está associado a um valor de intensidade na escala de cinza (Figura 1a). Já para imagens em RGB – formadas a partir dos três canais de cores, Vermelho, Verde e Azul (RGB: Red, Green, Blue), tem-se uma matriz de dados tridimensional, formada a partir de três matrizes de dados concatenadas, em que cada pixel terá três valores de intensidade, um para cada canal de cor (Figura 1b) (PRATS- MONTALBÁN; DE JUAN; FERRER, 2011).

Existe um tipo particular de imagem que, além de fornecer as informações espaciais da amostra, também fornece informações a respeito da composição química dela. Estas imagens possuem um espectro de alta resolução ou alguns comprimentos de onda associados a cada pixel da imagem e são chamadas de imagens hiperespectrais (HI, do inglês *Hyperspectral Images*) ou multiespectrais, respectivamente (Figura 1c). Esses espectros podem ser obtidos por diversas técnicas analíticas, como Raman, Infravermelho, Fluorescência, UV-vis, etc., sendo as técnicas de Raman e de Infravermelho, geralmente, as mais informativas (DE JUAN, *et al.*, 2009).

Figura 1 - Matriz de dados de imagens (a) em escalas de cinza, (b) em RGB e (c) hiperespectrais.



Fonte: DE JUAN, 2009.

3.3 Reconhecimento de Padrões

O reconhecimento de padrões é uma das principais vertentes do uso da estatística multivariada em química analítica, para a obtenção de mais informações quando comparado com os procedimentos univariados que são usualmente adotados (CORREIA; FERREIRA, 2007). Ela pode ser supervisionada ou não supervisionada. Sendo assim, uma análise multivariada supervisionada deve ter um conhecimento prévio sobre identidade das amostras, um exemplo seria o PLS-DA, pois o agrupamento observado segue uma tendência (um direcionamento) previamente informado na análise. Já uma análise multivariada não supervisionada não é necessário ter um conhecimento prévio, um exemplo seria a PCA (do inglês *Principal Component Analysis*), pois elas são úteis para o reconhecimento de padrões, tendências e agrupamentos não direcionados das amostras. (YI *et al.*, 2013) (SANTOS, 2014). Quando há informações anteriores sobre as classes ou objetos na imagem, a tecnologia de reconhecimento de padrões estatísticos pode ser aplicada diretamente à ela. Sem essas informações, pode primeiro segmentar a imagem e, em seguida, as categorias detectadas podem ser utilizadas como ferramenta de reconhecimento de padrões. (PRATS-MONTALBÁN; DE JUAN; FERRER, 2011).

3.3.1 Análise de Componentes Principais (PCA)

Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) são técnicas de reconhecimento de padrões usadas para interpretação de resultados. A PCA é utilizada para projetar os dados multivariados em um espaço que reduz a dimensionalidade da matriz de dados, empregando combinação linear das variáveis originais. Assim, esse novo modelo de variáveis existentes é representado na forma de componentes principais (PCs, do inglês *Principal Component*) (FERNANDES, 2016). A aplicação da PCA em imagens hiperespectrais são usadas para extração de informações a respeito de características importantes da amostra, já que é possível comprimir os dados, e assim, preservar informações fundamentais e interpretar a relação entre as variáveis utilizadas e a estrutura da amostra. A PCA pode ser usada como ferramenta para seleção de características da imagem. A regressão por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*) pode ser utilizada como outra técnica de projeção para a Análise Multivariada de Imagens (MIA, do inglês *Multivariate Image Analysis*). Técnicas

como SIMPLISMA (SIMPLISMA, do inglês *Simple-to-use Interactive Self-modeling Mixture Analysis*) e MCR-ALS (MCR-ALS, do inglês *Multivariate Curve Resolution – Alternating Least Squares*) são aplicadas para o aprimoramento da resolução (PRATS-MONTALBÁN, J. M.; DE JUAN, A.; FERRER, A., 2011).

3.3.2 PLS-DA

Segundo Ballabio e Consonni (2013), a Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Squares - Discriminant Analysis*) é um método de classificação linear onde as propriedades da regressão de mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*) tem a aptidão de distinguir uma técnica de classificação. O PLS-DA é baseado no algoritmo de regressão PLS, que analisa as variáveis latentes com covariância máxima com as variáveis da matriz **Y**.

O PLS-DA é classificado como um método supervisionado, pois é desenvolvido mediante o conhecimento prévio das classes do conjunto de calibração. Neste tipo de análise não são consideradas as informações iniciais sobre a identidade das amostras e, assim, apenas identifica ou encontra a tendência dos dados de se separarem, formando subgrupos (LOPES, 2015)

A construção de um método por PLS-DA é desenvolvido pelo modelo PLS que utiliza uma regressão entre os scores das matrizes **X** e **Y**. Nele se tem uma relação interna dos dados de cada matriz (matriz **X** e matriz **Y**) estabelecendo um vínculo entre as duas (GELADI, P.; KOWALSKI, B. R., 1986).

A PCA calcula os componentes com a mínima correlação entre as variáveis originais, com a variância máxima **X** e o PLS procura variável latente que expliquem a máxima correlação entre as matrizes **X** e **Y**. Assim, as informações podem ser explicadas geometricamente pelas variáveis latentes como uma leve rotação das PCs de forma a aumentar a relação entre os dois conjuntos de dados (WOLD, 2001).

As equações (1) e (2) mostram matematicamente que o PLS consiste na decomposição das matrizes **X** e **Y**, respectivamente:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

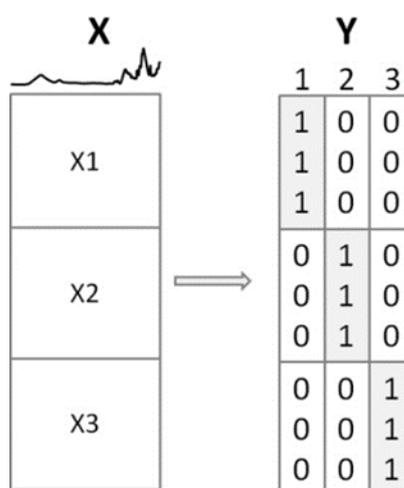
$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (2)$$

Portanto, **T** e **U** são semelhantes às matrizes dos escores e **P** e **Q** são semelhantes às matrizes dos pesos (ver equação 2). Assim, o PLS ajusta as equações (1) e (2) de forma que **E** e **F** sejam aproximadamente zero e constrói uma relação entre **U** e **T** (equação 3) (KALIVAS, 2006):

$$\mathbf{U} = \mathbf{T}\mathbf{W} \quad (3)$$

Na verdade, ao usar a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) em uma matriz **Y** com apenas 0 e 1 para classificação, e supondo que um sistema contenha 3 classes pertencentes a matriz **X**, conforme é mostrado na figura 2, essas matrizes vão ser relacionadas aos espectros que contêm as amostras envolvendo as classes 1, 2 e 3. Portanto, ao contruir o modelo para classificação das amostras, o valor para se determinar a categoria na matriz **Y** será no valor de 1, caso contrário, o resultado será no valor 0. Geralmente, o limiar é usado para a classificação de amostras que apresentem valor superior ou igual a 0,5, o que indica à que classe ela representa.

Figura 2 - Esquema da construção de um modelo PLS-DA



Fonte: Silva, 2013.

A vantagem do PLS-DA são as fontes relevantes de variabilidade de dados chamadas de Variáveis Latentes (VLs), que são combinações lineares das variáveis. Ela permite a visualização gráfica e interpretação dos diferentes padrões de dados e relaciona os escores e loadings das variáveis. Os escores são as coordenadas das

amostras no hiperespaço de projeção da VL e os loadings são os coeficientes das variáveis nas combinações lineares para determinar as variáveis latentes e assim, podem ser interpretadas como influência de cada variável em cada VL (BALLABIO; CONSONNI, 2013).

3.3.3 Curva ROC

Análise ROC (*do inglês Receiver Operating Characteristic*) é um método gráfico de avaliar, organizar e selecionar um sistema de diagnóstico de detecção. Esses gráficos são utilizados para detecção de sinais, avaliar a qualidade de transmissão em um canal com ruído. A análise ROC também é utilizada para a construção e refinamento de modelos (PRATI *et al.* 2008).

Gráficos ROC são usados também na psicologia para avaliar se o indivíduo consegue distinguir os estímulos e não estímulos, na medicina, na qualidade de um teste clínico, em economia, para avaliar a desigualdade de renda e também na previsão do tempo, avaliando a qualidade das previsões de eventos raros (PRATI *et al.* 2008).

4 MATERIAL E MÉTODOS

4.1 Local

As análises instrumentais deste trabalho foram realizadas no Laboratório Avançado de Tecnologia Química (LATECQ) na Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) em Campina Grande - Paraíba e o tratamento dos dados foi realizado nas dependências da Universidade Estadual da Paraíba, campus I.

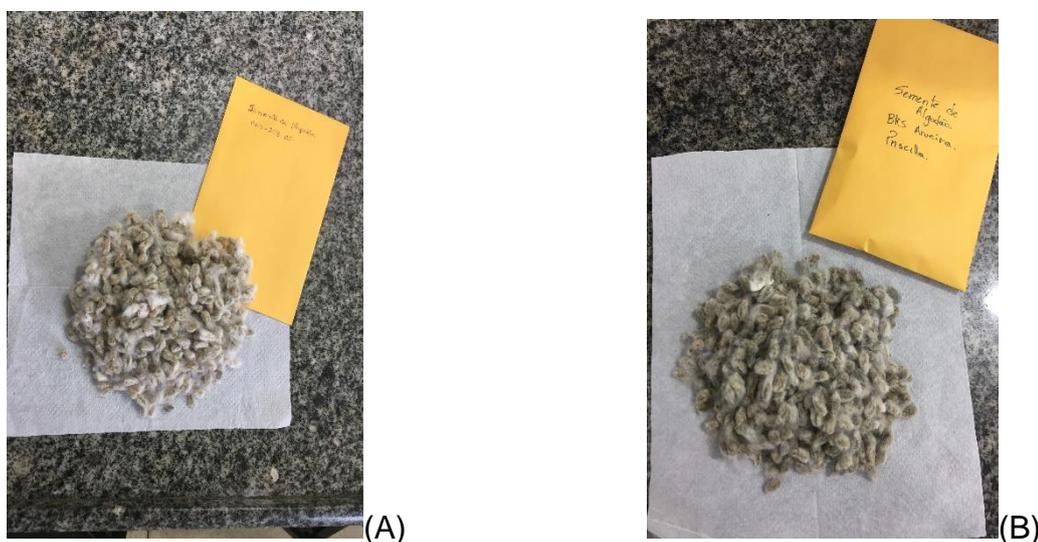
4.2 Amostras

Foram utilizadas para análise de semente de algodão, dois tipos de cultivares BRS 368 RF (transgênicas) e BRS Aroeiras (convencionais), sendo 20 amostras de cada uma.

4.3 Preparo das amostras

Primeiro, foi realizado o descaroçamento do algodão, o qual separa as fibras do algodão da semente (caroço) (Figuras 3a e 3b). Esse processo foi realizado manualmente.

Figura 3 - A) Sementes BRS 368 RF e B) Sementes Aroeiras



Fonte: Própria, 2019.

Em seguida, as sementes passaram pelo processo de deslincamento químico por via úmida. Esse processo utiliza ácido sulfúrico, na proporção de 7 kg de sementes para 1 L de ácido concentrado (p.a.) durante 5 minutos e depois ocorre a lavagem em água corrente para retirar o excesso do ácido e segue para a secagem natural por 12 horas.

4.4 Equipamento

A obtenção das imagens foi adquirida pelo espectrômetro de imagens hiperespectrais na região do infravermelho próximo (HSI-NIR) do fabricante *Specim* (Figura 4), *Spectral Imaging Ltd.* Oulu, Finlândia, modelo *Sisuchema-SWIR*, software de aquisição de imagens *ChemaDaq*, versão 3.621.992.6-R, faixa espectral: 900-2500 nm. Resolução espectral: 6,25 nm. Resolução espacial: Lente de 50 mm (0,156mm = 156 micrômetros). Frequência: 60 Hz. Exposição: 1,6 ms. A altura entre a lente e a amostra foi definida em 50cm, a velocidade do transportador foi fixada em 15.60mm/s.

Figura 4 - Espectrômetro *sisuChema SWIR* - *Specim*®



Fonte: Rocha, 2019.

4.5 Registro das imagens

Para o registro das 2 imagens, as sementes de algodão foram organizadas em fileiras na bandeja disponível pelo equipamento HSI-NIR.

A primeira imagem foi enfileirada as 20 sementes de algodão transgênicas e obtido o registro. O processo foi repetido para obtenção da imagem das 20 sementes de algodão convencionais.

4.6 Construção do modelo PLS-DA

Para a construção do modelo de reconhecimento de padrões supervisionado, PLS-DA, todas as sementes foram selecionadas de forma randômica (aleatória). O conjunto de treinamento foi composto por 13 sementes transgênicas e 13 sementes convencionais, totalizando 26 amostras. Já o conjunto de teste foi composto por 7 sementes transgênicas e 7 convencionais, totalizando 14 amostras.

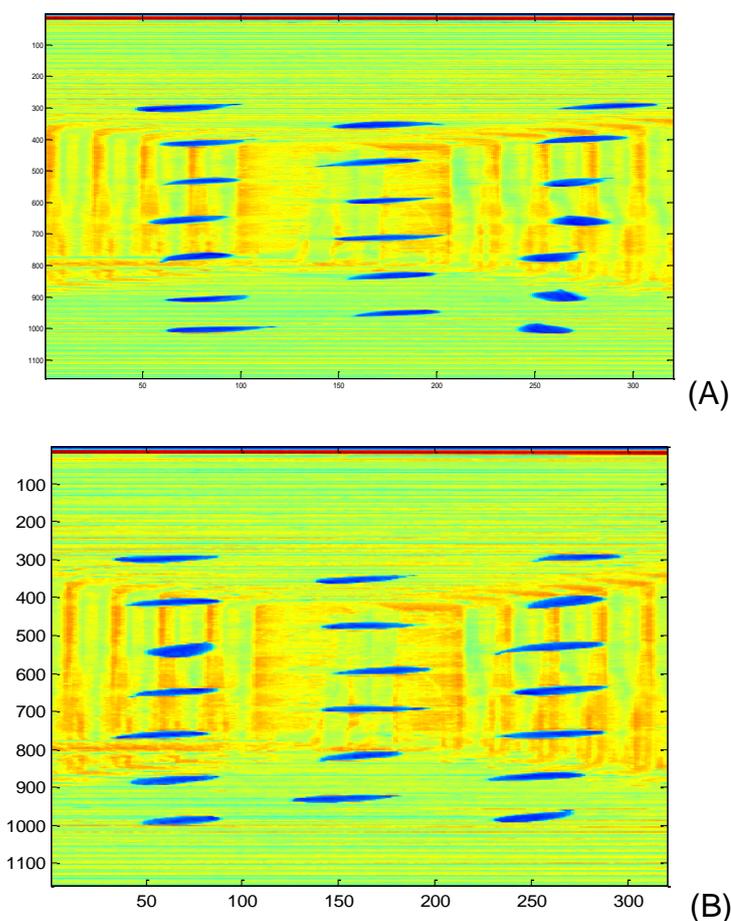
4.7 Tratamento das Imagens

O tratamento das imagens e construção dos modelos de reconhecimento de padrões foi realizado pelos programas de PLS *Toolbox* em ambiente *Matlab*.

5 RESULTADOS E DISCUSSÃO

As imagens médias das sementes transgênicas e convencionais são mostradas na figura 5a e 5b, respectivamente. As manchas azuis representam as sementes e as partes ciano, amarelo e laranja representam a base na qual elas estavam sobrepostas.

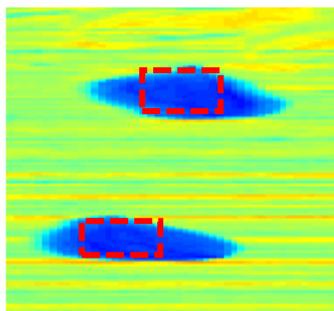
Figura 5 - Imagem média da semente de algodão A) transgênica e B) convencional



Fonte: Algoritmo do Hypertools®

Cada semente foi tratada, inicialmente, de forma individual. Deste modo, na imagem foi selecionada uma região central de cada semente, na forma de um cubo, que foi utilizada para posteriormente compor a matriz de dados, figura 6. Cada cubo contém informações espaciais e espectrais sobre a semente.

Figura 6 - Seleção da região para montagem da matriz de dados



Fonte: Própria, 2019.

Todos os cubos possuíam as mesmas dimensões (11x6x256), relativas as espaciais (11x6) e espectrais (256 comprimentos de onda) sobre a semente. Em seguida, a matriz de dados foi montada, organizando-se os cubos referentes as amostras na forma de linha, de acordo com a figura 7.

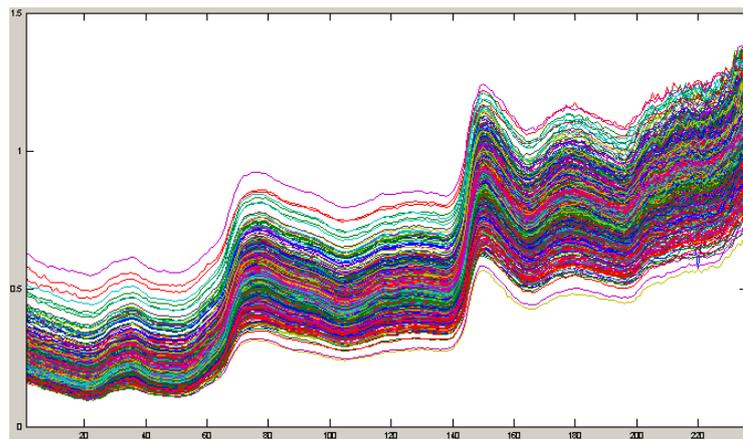
Figura 7 - Organização da matriz de treinamento

| | | | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|-----|-----|----|----|----|----|-----|-----|----|----|----|-----|----|----|----|----|-----|----|
| T1 | T2 | T3 | T4 | ... | T13 | C1 | C2 | C3 | C4 | ... | C13 | T1 | T2 | T3 | ... | T7 | C1 | C2 | C3 | ... | C7 |
|----|----|----|----|-----|-----|----|----|----|----|-----|-----|----|----|----|-----|----|----|----|----|-----|----|

Fonte: Própria, 2019.

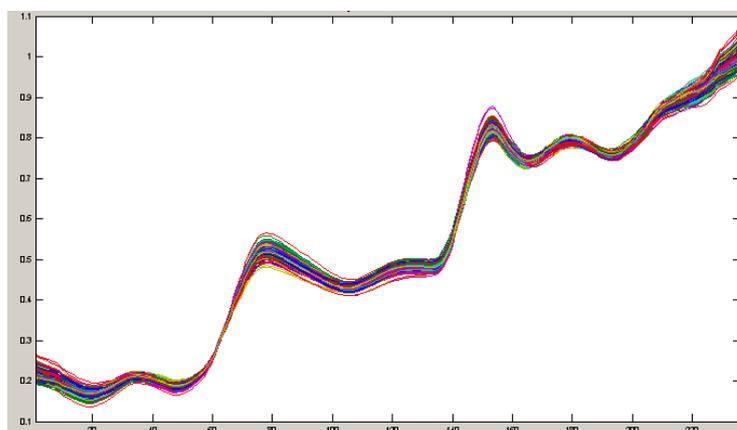
Posteriormente as imagens foram submetidas a diversos pré-processamentos espectrais, a fim de remover ruídos ou outras características que viessem a prejudicar a qualidade dos modelos de reconhecimento de padrão a serem construídos. Diversos pré-processamentos foram testados para o tratamento de dados, entre eles, SNV, MSC, derivadas e suavização com diversos tamanhos de janelas e a combinação destes. A figura 8 mostra os dados brutos e a figura 9 mostra os dados pré-tratados com o pré-processamento que apresentou a melhor performance na distinção entre amostras transgênicas e convencionais, foi a suavização com janela de 9 pontos e polinômio de 2ª ordem associada ao SNV (SNV, do inglês *Standard Normal Deviation*).

Figura 8 - Espectros brutos



Fonte: Própria, 2019.

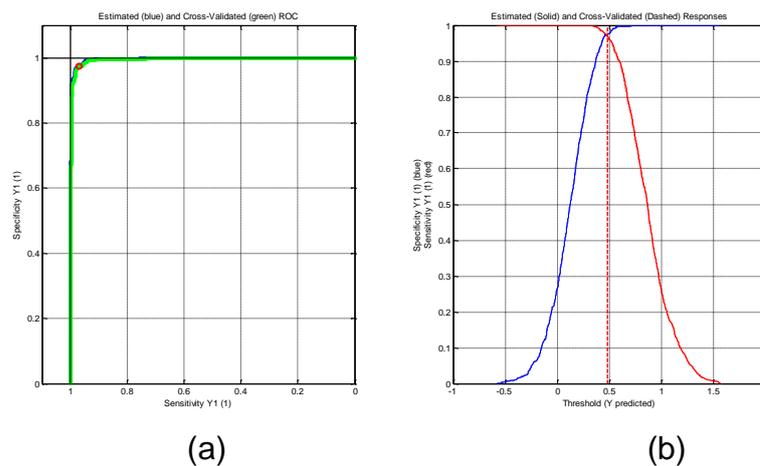
Figura 9 - Espectros pré-processados



Fonte: Própria, 2019.

O modelo PLS-DA foi construído utilizando as amostras de calibração. O modelo mostrou um bom desempenho na etapa de calibração apresentando especificidade e sensibilidade de 0,97 tanto na etapa de validação cruzada como nos valores estimados para a curva ROC como mostrada na figura 10a. O limiar de separação para as duas classes foi de aproximadamente 0,5 como mostra o gráfico da figura 10b.

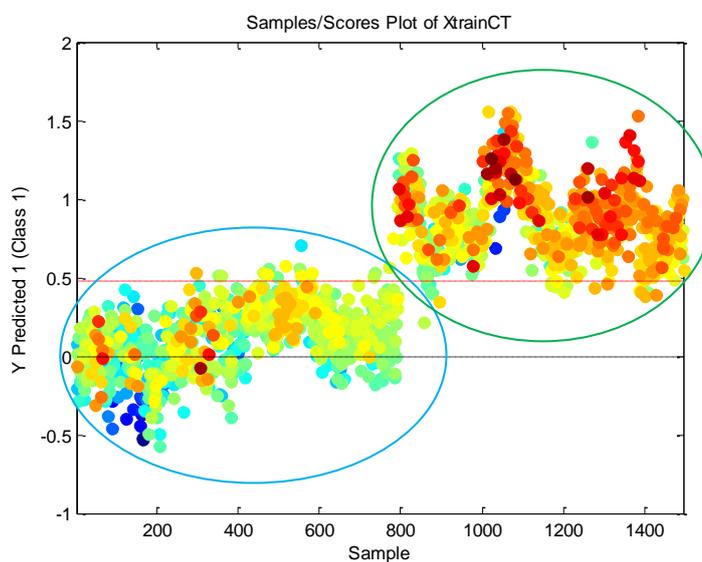
Figura 10 - a) Curva ROC e b) limiar de separação entre as classes para a etapa de treinamento



Fonte: Algoritmo do PLS Toolbox

Estes dados indicam a boa qualidade do modelo de reconhecimento de padrões construído, de modo que o mesmo foi utilizado para a classificação das amostras externas ao conjunto de treinamento, as amostras do conjunto de teste. A figura 11 mostra a classificação das amostras convencionais (ciano) e transgênicas (verde) na etapa de treinamento.

Figura 11 - Classificação das amostras de treinamento

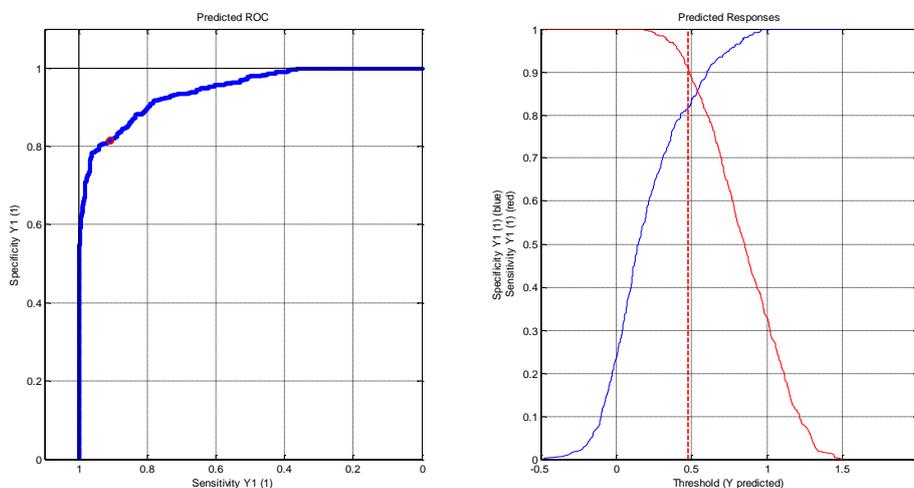


Fonte: Própria, 2019.

Pode-se observar que a maioria das amostras foram classificadas de forma correta a um limiar de 05.

A figura 12 ilustra o desempenho do modelo construído na etapa com as amostras de treinamento para a classificação das amostras do conjunto de teste.

Figura 12 - a) Curva ROC e b) limiar de separação entre as classes para a etapa de teste

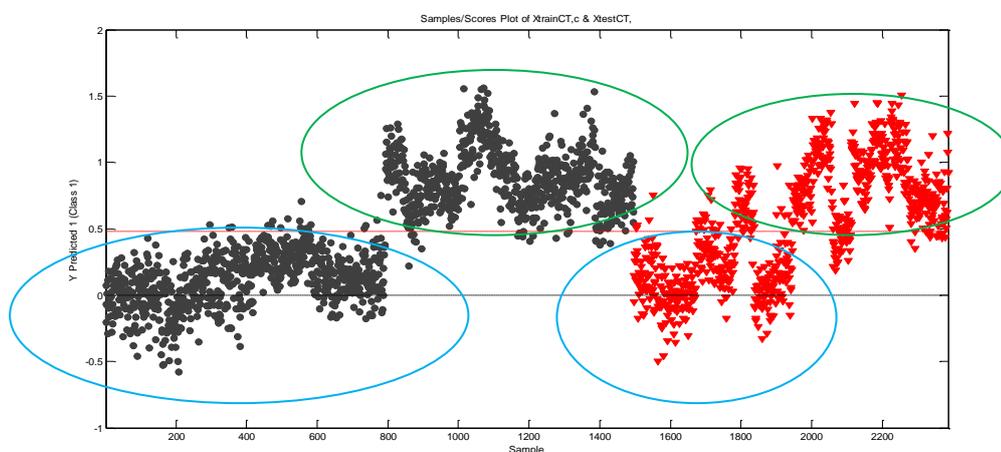


Fonte: Própria, 2019.

As figuras 12a e 12b mostram uma sensibilidade de 0,9 e uma especificidade de 0,82, e um limiar de separação próximo a 0,5 na etapa de teste, respectivamente. O que representa um bom desempenho do modelo PLS-DA na etapa de teste.

A figura 13 ilustra a classificação das amostras dos conjuntos de treinamento e de teste nas classes convencional (ciano) e transgênica (verde).

Figura 13 - Classificação das amostras treinamento e de teste



Fonte: Própria, 2019.

Na imagem podemos observar a distinção entre as classes no limiar de separação 0,5. As imagens de cor pretas são as amostras do conjunto de treinamento e as de cor vermelha são as amostras do conjunto teste. As figuras circuladas na coloração de ciano, abaixo do limiar 0,5, são as sementes classificadas convencionais e as imagens circuladas na cor verde, acima do limiar 0,5, são as sementes classificadas transgênicas.

6 CONCLUSÃO

Os métodos convencionais normalmente usados para identificar organismos geneticamente modificados (OGM's) costumam ser caros e não estão facilmente disponíveis. Deste modo, existe a necessidade do desenvolvimento de métodos alternativos para a detecção e quantificação de OGM's, com o objetivo de serem mais rápidos, menos onerosos e acessíveis.

As HI-NIR foram registradas e observou-se que a distinção entre as classes só era possível com o uso dos métodos de reconhecimento de padrões, para isso foi construído um modelo PLS-DA, para classificar as classes das sementes transgênicas e convencionais. Em relação ao tratamento dos dados, o pré-processamento que forneceu o melhor modelo foi a suavização com janela de 9 pontos e polinômio de 2ª ordem associada ao SNV. Na etapa de treinamento, obteve-se valores de especificidade e sensibilidade de 0,97 e limiar de separação de 0,5, mostrando a capacidade do modelo em distinguir entre as duas classes. Na etapa de teste, foram obtidos valores de especificidade de 0,82 e sensibilidade 0,9, um limiar de 0,5.

Diante dos resultados obtidos, pode-se observar que a utilização das HI-NIR, aliada a métodos de reconhecimento de padrões supervisionados é capaz de distinguir entre sementes convencionais e transgênicas, e que pode ser utilizada como método alternativo aos métodos convencionais. Mesmo que a HI ainda não seja tão popular no Brasil e que, a princípio, o tratamento de dados pareça complexo, esta técnica é bastante rica já que é composta por informações espaciais e espectrais. Deste modo, o método desenvolvido mostrou ser eficaz para o diagnóstico de sementes, a fim de distinguir entre as convencionais e transgênicas. O método desenvolvido é bastante rápido para a predição da classe das sementes, realizando a classificação num tempo entre 3 e 5 minutos, diferentemente dos métodos convencionais que são bastante laboriosos e demorados. Além disso, não destrói as amostras durante a análise.

REFERÊNCIAS

ABRAPA, Associação Brasileira dos Produtos de Algodão. **Algodão no Mundo**. Disponível em: < <https://www.abrapa.com.br/Paginas/dados/algodao-no-mundo.aspx>>. Acesso em 11 nov. 2020.

ALMEIDA, Mirella Figueiró De. **Deteção e Transmissão de *Colletotrichum gossypii* var. *cephalosporioides* Pelas Sementes de Algodão de Algodoeiro e Uso de RFP para o Estudo da Interação Patógeno-Hospedeiro**. Tese de (Doutorado em Agronomia/Fitopatologia) – Universidade Federal de Lavras. Lavras. 2012.

BALLABIO, D.; CONSONNI, V. Classification tools in Chemistry. Part 1: Linear models. PLS-DA. **Analytical Methods**, v.5, 2013.

BERTHEAU, Y.; DIOLEZ, A.; KOBILINSKY, A.; MAGIN, K. Detection methods and performance criteria for genetically modified organisms. **Journal of AOAC International**, 2002. v.85, p. 801-808.

BRASIL. Ministério de Ciência e Tecnologia. Comissão Técnica Nacional de Biossegurança. Parecer Técnico n. 513/2005. **Liberação comercial de algodão geneticamente modificado resistente às principais pragas da Ordem Lepidoptera**. 2005. Disponível em: <<http://www.cntbio.mcti.gov.br>> Acesso em: 09 jan. 2020.

BOCCIA, F. Os alimentos geneticamente modificados e perspectiva do consumidor. **Revista Nutricional Agricultura**, São Paulo, v.20, n.1, p.10-22, 2015.

CONCEIÇÃO, F.R.; MOREIRA, A.N.; BINSFELD, P.C; Deteção e quantificação de organismos geneticamente modificados em alimentos e ingredientes alimentares. **Ciência Rural, Santa Mariam** v.36, n.1, p.315-324, jan-fev, 2006.

CONCEIÇÃO, F.R. et al. Deteção de organismos geneticamente modificados. In: BINSFELD, P.C. **Biossegurança em biotecnologia**. Rio de Janeiro: Interciência. 2004.145-169 p.

CORREIA, P.R.M.; FERREIRA, M.M.C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. **Quím. Nova [online]**. 2007. v.30, n.2, 481-487 p.

CRUZ, J.; BAUTISTA, M.; AMIGO, J. M.; BLANCO, M. Nir-chemical imaging study of acetylsalicylic acid in commercial tablets. **Talanta**. 2009. v. 80, n. 2, 473–478 p.

CRUZ, J.; BLANCO, M. Content uniformity studies in tablets by NIR-CI. **Journal of pharmaceutical and biomedical analysis**. 2011. v. 56, n. 2, 408–412 p.

DE JUAN, A; MAEDER, M; HANCEWICZ, T; DUPONCHEL, L; TAULER, R. Chemometrics tools for image analysis. In: **H. W. SALZER, R.; SIESLER (Ed.); Infrared and Raman Spectroscopic Imaging**. 2009. 65–106 p.

DONEUS, M.; VERHOEVEN, G.; ATZBERGER, C.; WESS, M.; RUŠ, M. New ways to extract archaeological information from hyperspectral pixels. **Journal of Archaeological Science**. 2014. v. 52, 84–96 p.

EMBRAPA. **Algodão resistente a herbicida será apresentada a produtores cearenses**, 2017. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/25276303/algodao-resistente-a-herbicida-sera-apresentado-a-produtores-cearenses>> Acesso em: 05 mar. 2019.

EMBRAPA. **Aumento da produção de algodão no Brasil traz novos desafios para a pesquisa, aponta documento da Embrapa**, 2019. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/43931817/aumento-da-producao-de-algodao-no-brasil-traz-novos-desafios-para-a-pesquisa-aponta-documento-da-embrapa>>. Acesso em: 06 nov. 2020.

FERNANDES, David Douglas de Sousa. **Novas estratégias para seleção de variáveis por intervalos em problemas de classificação**. Tese (Doutorado em Química) – Universidade Federal da Paraíba, João Pessoa, 2016.

GELADI, P.; KOWALSKI, B. R. **Partial least-squares regression: a tutorial**. *Analytica Chimica Acta*, v. 185, p. 1–17, 1986.

GIOVANNINI, T.; CONCILLO, L. PCR detection of genetically modified organisms: a review. **Starch**. v.54, p.321-327, 2002.

HOLST-JENSEN, A; Ronning, S. B.; Lovseth, A; Berdal, K. G. PCR technology for screening and quantification of genetically modified organisms (GMOs). **Analytical and Bioanalytical Chemistry**. v.375, p.985- 993, 2003.

ISAAA. Global Status of Commercialized Biotech/GM Crops in 2018: Biotech Crops Continue to Help Meet the Challenges of Increased Population and Climate Change. **ISAAA Brief No. 54**. ISAAA: Ithaca, NY. 2018. Disponível em:

<<https://d335luupugsy2.cloudfront.net/cms/files/50569/1566823794Resumo-executivo.pdf>>. Acesso em: 12 jan. 2020.

KALIVAS, J. H.; GEMPERLINE, P. J. Calibration. *In*: P. GEMPERLINE (Ed.); **Practical Guide to Chemometrics**. 2nd ed., p.105–165, 2006.

LOPES, R. E. C. **Discriminação de madeiras similares por NIRS e PLS-DA considerando variações de temperatura e umidade**. Dissertação (Mestrado em Química) – Universidade de Brasília, Brasília, 2015.

LORENTE, D.; ALEIXOS, N.; GÓMEZ-SANCHIS, J.; CUBERO, S.; GARCÍA-NAVARRETE, O. L.; BLASCO, J. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. **Food and Bioprocess Technology**. v.5, p.1121–1142, 2012.

MAGIN, K.; MIHALIAK, C.; SOMERVILLE, L.; PRIVALLE, L.; CHARLTON, S.; PORTER, L. Methods for detection of GMO grain in commerce. **American Crop Protection Agency**. 2000.

MEHL, P. M.; CHEN, Y.-R.; KIM, M. S.; CHAN, D. E. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. **Journal of Food Engineering**. v.61, n.1, p. 67–81, 2004.

MIRAGLIA, M.; Berdal, K. G.; Brera, C.; Corbisier, P.; Holst-Jensen, A.; Kok, E. J.; Marvin, H. J. P.; Schimmel, H.; Rentsch, J.; van Rie, J. P. P. F.; Zagon, J. Detection and traceability of genetically modified organisms in the food production chain. **Food Chemical Toxicology**. v.42, p.1157-1180, 2004.

OBEID, P.J.; Christopoulos, T. K.; Loannou, P. C. Rapid analysis of genetically modified organisms by in-house developed capillary electrophoresis chip and laser-induced fluorescence system. **Electrophoresis**. v.25, p.922-930, 2004.

PENNA, J.C.V. Melhoramento do algodão. *In*: **BORÉM, A. (Ed.). Melhoramento de espécies cultivadas**. 2. ed. Viçosa, UFV. 2005. Cap.1, 15-53 p.

PETIT, L. et al. Screening of genetically modified organisms and specific detection of Bt176 maize in flours and starches by PCR-enzyme linked immunosorbent assay. **European Food Research and Technology**. v.217, p.83-89, 2003.

PETRUCCI, Fabrizio. **Algodão transgênico na Agricultura: Benefícios e Controvérsias**. Trabalho de Conclusão de Curso (Graduação) – Instituto de Economia. Universidade Estadual de Campinas, Campinas, 2011.

PRATI, R. C.; BATISTA, G. E. D. A. P. A.; MONARD, M. C. Curvas ROC para avaliação de classificadores. **IEEE Latin America Transactions**[S.l.], v. 6, n. 2, p. 215-222, 2008. Disponível em: <https://sites.icmc.usp.br/gbatista/files/ieee_la2008.pdf>. Acesso em 30 out. 2020.

PRATS-MONTALBÁN, J. M.; DE JUAN, A.; FERRER, A. Multivariate image analysis: A review with applications. **Chemometrics and Intelligent Laboratory Systems**. v.107, n.1, p.1–23, 2011.

ROCHA, P.D. **Metodologia para distinção entre genótipos convencionais e transgênicos de algodão utilizando espectroscopia NIR e imagens hiperespectrais**. Dissertação (Mestrado em Química) – Universidade Estadual da Paraíba, Campina Grande, 2020.

SANTOS, F.A.D **Quimiometria aplicada a metabolômica de Aldama La Llave: Contribuições quimiotaconômicas direcionada baseada em inibição de cicloxigase-1 e 5-lipoxigase**. Dissertação (Mestrado em Ciências Farmacêuticas) – Universidade de São Paulo, Ribeirão Preto, 2014.

SILVA, C.S. **Uso de imagens hiperespectrais na região do infravermelho próximo para identificar fraudes em documentos**. Dissertação (Mestrado em Química) – Universidade Federal de Pernambuco, Recife, 2013.

SIRISOMBOON, C. D.; PUTTHANG, R.; SIRISOMBOON, P. Application of near infrared spectroscopy to detect aflatoxigenic fungal contamination in rice. **Food Control**. v.33, p.207–214, 2013.

STAVE, J.W. Detection of new or modified proteins in novel foods derived from GMO – future needs. **Food Control**, v.10, p.361-374, 1999.

TORRES, A.C.; Nascimento, W. M.; Paiva, S. A. V.; Aragão, F. A. S. Bioassay for detection of transgenic soybean seeds tolerant to glyphosate. **Pesquisa Agropecuária Brasileira**. v.38, p.1053-1057, 2003

VARGAS, B., BASSO A., RODRIGUES T., SILVA L., GATZKE M., & FRIZZO M. Biotecnologia e Alimentos Geneticamente Modificados: Uma Revisão. **Revista Contexto & Saúde**, v.18, n.35, p.19-26. 2018.

WOLD, S. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**. v. 58, n. 2, p. 109–130, 2001.

YAMAGUCHI, H. et al. Two detection methods of genetically modified maize and the state of its import into Japan. **Food Control**. v.14, p.201-206, 2003.

Yi, W. S.; Cui, D. S.; Li, Z.; Wu, L. L.; Shen, A. G.; Hu, J. M. Gastric cancer differentiation using Fourier transform near-infrared spectroscopy with unsupervised pattern recognition. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v.101, p.127-131, 2013.