PRECISION DIAGNOSTICS AND INNOVATIONS FOR PLANT BREEDING RESEARCH

By

Eli Hugghis

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology – Crop and Soil Sciences – Master of Science

2021

ABSTRACT

PRECISION DIAGNOSTICS AND INNOVATIONS FOR PLANT BREEDING RESEARCH

By

Eli Hugghis

Major technological advances are necessary to reach the goal of feeding our world's growing population. To do this, there is an increasing demand within the agricultural field for rapid diagnostic tools to improve the efficiency of current methods in plant disease and DNA identification. The use of gold nanoparticles has emerged as a promising technology for a range of applications from smart agrochemical delivery systems to pathogen detection. In addition to this, advances in image classification analyses have allowed machine learning approaches to become more accessible to the agricultural field. Here we present the use of gold nanoparticles (AuNPs) for the detection of transgenic gene sequences in maize and the use of machine learning algorithms for the identification and classification of *Fusarium* spp. infected wheat seed. AuNPs show promise in their ability to diagnose the presence of transgenic insertions in DNA samples within 10 minutes through colorimetric response. Image-based analysis with the utilization of logistic regression, support vector machines, and k-nearest neighbors were able to accurately identify and differentiate healthy and diseased wheat kernels within the testing set at an accuracy of 95-98.8%. These technologies act as rapid tools to be used by plant breeders and pathologists to improve their ability to make selection decisions efficiently and objectively.

This thesis is dedicated to my mom, dad, and my fiancée.
Thank you for all of your love, support, and consistent encouragement to push me forward toward my goals.

ACKNOWLEDGEMENTS

I would like to thank the Thompson Lab and all its members for their kindness and support over the years that I have been at MSU. Phong Los, thank you for your assistance in scanning the multitude of wheat seed images and Robert Shrote and Ruijuan Tan for assistance during the struggles of code writing. Linsey Newton for the many questions I have asked you over the years about all the things. To everyone else in the lab who has offered an encouraging word, random act of kindness, or other act of help that made this all possible!

Thank you to the Day Lab and its members for making me feel welcomed during my shared lab space visits and Saroopa Samaradivakara for your assistance early on in my assay development process.

Thank you to the Chilvers lab and Mikaela Breunig for providing wheat seed samples for my image collection and answering my many plant pathology inquiries.

Lastly, thank you to the members of my committee for all your time, guidance, and support throughout this master's degree process.

THANK YOU ALL!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: REVIEW OF PRECISION DIAGNOSTICS AND INNOVATIONS FOR PLANT BREEDING RESEARCH

ABSTRACT:

Major technological advances are necessary to reach the goal of feeding our world's growing population. To do this, there is an increasing demand within the agricultural field for rapid diagnostic tools to improve the efficiency of current methods in plant disease and DNA identification. The use of gold nanoparticles has emerged as a promising technology for a range of applications from smart agrochemical delivery systems to pathogen detection. In addition to this, advances in image classification analyses have allowed machine learning approaches to become more accessible to the agricultural field. Here we present the use of gold nanoparticles (AuNPs) for the detection of transgenic gene sequences in maize and the use of machine learning algorithms for the identification and classification of *Fusarium* spp. infected wheat seed. AuNPs show promise in their ability to diagnose the presence of transgenic insertions in DNA samples within 10 minutes through colorimetric response. Image-based analysis with the utilization of logistic regression, support vector machines, and k-nearest neighbors were able to accurately identify and differentiate healthy and diseased wheat kernels within the testing set at an accuracy of 95-98.8%. Rapid diagnostic tools can be used by plant researchers to accelerate their decision-making ability efficiently and objectively.

PART 1 – THE VERSATILE APPLICATION OF GOLD NANOPARTICLES IN THE SCIENCES

HISTORY AND ORIGIN OF GOLD NANOPARTICLES

Colloidal gold and the use of gold nanoparticles have been extensively observed and studied for several centuries. The use of colloidal gold can be found as far back as the 4th century BC in the Middle East, China, and India where "potable gold" was used for medicinal purposes (Dykman & Khlebtsov, 2019; (Huaizhi & Yuantao, 2001). In Europe, early use of this material was for artistic application as a color stain used by glassworkers in ancient Rome. One of the most famous examples of this is the Lycurgus Cup developed in the 4th century which was known for its very interesting dichromatic color properties (Loos, 2015; Taylor, 2010). These properties were also useful in large cathedral halls in their stained-glass windows. The scientific understanding for these phenomena was not fully understood until 1990 when a study explained the presence of various nanoparticles dispersed in the glass of the cup (Barber & Freestone, 1990). Centuries later the use of colloidal gold for art transitioned to its use within medicine and science.

One of the oldest writings on the medicinal use for this material is in 1618 by Francis Anthony where he discusses his alchemy studies for its curative properties for various diseases and the formation of the colloidal solution (Antonii, 1618; Culpeper, 1657). In the Middle Ages, gold colloids were used as an elixir of life and longevity, with the belief that drinking the gold solution would allow individuals to stay youthful (Charlier et al., 2009). In addition, alchemists would use the liquid elixir to treat several mental illnesses, syphilis, epilepsy, leprosy, and many other diseases (Daraee et al., 2016; Louis & Pluchery, 2012). A German chemist named Johann

Kunckels published a book in 1676 that discussed a “drinkable gold that contains metallic gold in a slightly pink solution that can exert curative properties for several diseases” (Rahman & Rebrov, 2014; Daniel & Astruc, 2004). He was able to conclude that the gold was present in the liquid, yet invisible to the human eye. In 1856, a more scientific evaluation of colloidal gold was seen during an accidental observation by Michael Faraday. Faraday was studying light refraction on different objects and was making thin pieces of gold for his microscope slides (Bean, 2020). To do this, he would wash the gold strips with a phosphorous-based reducing agent to make the gold pieces thin enough to pass light through them (Rahman & Rebrov, 2014). The wash’s runoff would produce a faint ruby red liquid, which when light passed through it generated unique cone-shaped refraction (Tweney, 2006). This discovery, known as the “Faraday-Tyndall effect”, is seen to be one of the main precursors to research within nanoscience and nanotechnology fields.

GOLD NANOPARTICLES IN MODERN MEDICINE

Through extensive study over the years, gold nanoparticles (AuNPs) have become extremely useful within the medical field. For quite some time, gold therapy was used as a main treatment for rheumatoid arthritis and tuberculosis (Garcia, 1981; Davis, 1988; Louis & Pluchery, 2012). Gold nanoparticles have a high surface area to volume ratio and a surface that is frequently conjugated with a variety of ligands used for a multitude of applications. Thomas and Kibanov were able to modify AuNPs with polyethyleneimine chains to improve the delivery of plasmid DNA into mammalian cells (Thomas & Klibanov, 2003). Bowman et. al. used conjugated AuNPs as a potent therapeutic for HIV and Gibson and co-workers used them for targeted, tumor-inhibiting, drug delivery (Bowman et al., 2008; Gibson et al., 2007). In addition to these

innovative applications, AuNPs have been used in a variety of biosensors and diagnostic assays. They have been used for multiplexed detection of cancer markers (Stoeva et al., 2006), detection of target proteins related to prostate and breast cancer (Nam et al., 2003), and more recently in the detection of COVID-19 (Kotz, 2020; Medhi et al., 2020; Ventura et al., 2020). The use of AuNPs has rapidly emerged as a useful technology within the medical field as well as in agriculture.

GOLD NANOPARTICLES IN AGRICULTURE AND PLANT SCIENCES

Advancements in gold nanoparticle technology have allowed this material to become a versatile tool in agriculture research. The gene gun, used in plant transformation, uses gold nanoparticles coated with plasmid DNA to transform crops such as maize (Kao et al., 2008), wheat (Ismagul et al., 2018), peanut (X. Y. Deng et al., 2001), and rice (Mortazavi & Zohrabi, 2018). Torney et. al. used functionalized nanoparticles to deliver DNA and chemicals into isolated plant cells and intact leaves. They found that uncapping the gold nanoparticles released bound chemicals and triggered expression of the green fluorescent protein gene contained within the plasmid attached to the surface of the AuNP (Torney et al., 2007). Also, AuNPs have been seen to improve seed germination (Arora et al., 2012; Shah & Belozerova, 2009), affect vegetative growth (Gopinath et al., 2014; Kumar et al., 2013), enhance total seed yield, and improve plant shoot to root ratios (Shah & Belozerova, 2009). Gold nanoparticles made by an extract from the seeds of *Abelmoschus esculentus* were seen to have antifungal properties against plant pathogens such as *Puccinia graminis* f. sp. *tritici* (stem rust pathogen) and *Aspergillus niger* (black mold) (Jayaseelan et al., 2013). When used on root-knot nematodes in tomato crops, AuNPs act as

management tools to combat the pest with no negative impact on plant growth (Thakur et al., 2018).

Along with these versatile applications for AuNPs, researchers also have examined their use as a rapid and efficient diagnostic tool for pesticide residue on fruit and vegetable products. Bai et. al. was able to develop a sensitive, relatively low-cost optical sensor for screening pymetrozine using unmodified AuNPs. They were able to detect the chemical at a detection limit of 1×10^{-6} M and visually diagnose the presence of the chemical due to the assays colorimetric response (L. Y. Bai et al., 2010). The residual pesticide Kitazine was detected by a visual assay that coupled an enzyme-linked immunoassay (ELISA) with bioconjugated AuNPs (Malarkodi et al., 2017). Kang et. al. made a colorimetric sensor for a pesticide using modified AuNPs. The assay was highly sensitive as it could detect the chemical as low as 10nM with the use of UV-Vis spectroscopy and could detect from water and food samples (Kang et al., 2018).

AuNP sensing has also been investigated for organic compounds such as pathogen DNA. Gold nanoparticles were used to detect the plant pathogen *Xanthomonas campestris*. AuNPs were modified to bind to pathogen DNA, leading to nanoparticle aggregation causing a visible shift in their color (H. Peng & Chen, 2019). Firrao et. al. used oligonucleotide-modified AuNPs to act as a fluorescence signal when hybridized with target DNA for the vineyard pathogen, *Flavescence dorée* (Firrao et al., 2005). Baetsen et. al. were able to detect viral DNA of cucurbit downy mildew in cucumber using unmodified gold nanoparticles to detect very low concentrations of the *Pseudoperonospora cubensis* DNA (Baetsen-Young et al., 2018). Outside of the use of gold nanoparticles, several other diagnostic tools for DNA sequence detection have been developed.

DNA DETECTION METHODS OF THE PAST

Over the last 30 years, the science community has been employing the use of methods such as Polymerase Chain Reaction (PCR), Restriction Fragment Length Polymorphism (RFLP), Short Tandem Repeat (STR) Analysis, and several others for genetic sequence analysis. Though these methods are widely used, each has its own set of drawbacks ranging from processing time and efficiency to overall costs associated (Table 1.1).

Table 1.1: Common DNA Detection Methods Of The Past

Common DNA Detection Methods of the Past		
Method	**Challenges/ Drawbacks**	**Literature**
Polymerase Chain Reaction (PCR) and its variations (qPCR, RT-PCR, RT-qPCR)	Very sensitive to contamination, relatively slow analysis, requires specialized and costly equipment	(Thomson & Dietzgen, 1995; Khan et.al., 2018; Broccanello et. al., 2018; Singh & Kapoor, 2018, Hoy et. al., 2019)
Restriction Fragment Length Polymorphism (RFLP)	Requires a large DNA sample for analysis, results can take weeks	(Powell et. al., 1996; Camele et. al., 2005; Kumar et. al., 2020)
Random amplified polymorphic DNA (RAPD)	Requires standardized laboratory conditions for reproducibility, requires specialized equipment, sensitive to the quality of DNA samples	(Mata et.al., 2009; Lin et. al., 2009; Zheng et. al., 2008)
Amplified fragment length polymorphism (AFLP)	Specialized and costly equipment and reagents, requires very clean template DNA samples	(Coyle et. al.,2005; Bryan et. al., 2017; Smith et. al., 2007)
Short Tandem Repeat (STR) analysis	Costly equipment and slow analysis	(Howard et. al., 2009; Undurraga et. al., 2012; Carlson et. al., 2015)
Sourthern Blot	Requires a large DNA sample for analysis, slow analysis	(McCabe et. al., 1997; Glowacka et. al., 2016; Honda et. al., 2002)
DNA Sequencing	Requires specialised and expensive equipment, data can be difficult to interpret, results can take a long time	(James et. al., 2013; Chandler et. al., 2002; Gill et. al., 2004)

With the progression of knowledge in the study of DNA and genetic sequence analysis, novel and more efficient detection methods have been developed.

MODERN, RAPID DNA DETECTION METHODS

PCR was the gold-standard method for reliable DNA detection in plant material, but since Kary Mullis developed this technology in 1983, several new techniques for the study of DNA have been developed that work to overcome the drawbacks of their predecessors:

- **ISOTHERMAL ASSAYS:** The emerging technology of isothermal assays has opened new opportunities for access and point-of-care use for plant disease diagnostics. These DNA amplification techniques are conducted at a constant temperature, lessening the need for specialized and costly equipment. Rojas et. al. demonstrated the use of recombinase polymerase amplification (RPA) as a rapid, species-specific diagnostic assay for detection of *Phytophthora sojae* and *P. sansomeana* (Rojas et al., 2017)**.** Loop-mediated isothermal amplification (LAMP), which utilizes primers that form hairpin-like structures to induce amplification, was found to be more rapid and sensitive than conventional PCR when detecting *Alternaria solani* (M. Khan et al., 2018). Though the assay was not as sensitive as nested PCR and qPCR, it was simpler, faster, and able to detect disease in young leaves that only showed minimal symptoms of early blight.
- **LATERAL FLOW ASSAYS:** Lateral flow assays are rapid immunological platforms that are typically comprised of a nitrocellulose membrane, sample pad, conjugate pad, and absorbent pad, and are best known for their point of care application. A rapid point of care method for the detection of cauliflower mosaic virus promoter (CaMV 35S) was achieved and coupled with cross-priming amplification technology (Huang et al., 2014). This nucleic acid lateral flow assay could detect as little as 30 copies of the plasmid containing the CaMV 35S gene and was made to monitor the presence of genetic

modifications rapidly and efficiently in products. A nucleic acid lateral flow immunoassay (NALFIA) was combined with PCR to detect *Macrophomina phaseolina* in soil and seed samples. This NALFIA used labeled primers to overcome the timely use of gel electrophoresis, allowing it to be simpler and faster than conventional PCR (Pecchia & Da Lio, 2018).

- **GENE CHIPS AND MICROARRAYS:** Microarray technology allows for multi-parallel analysis of many gene sequences at once. They typically involve separate gene-specific DNA fragments that are attached to a solid support. Detection occurs when fragments hybridize with targeted DNA sequences. Several common potato viruses were simultaneously detected with a microarray assay with comparable sensitivity to ELISA (Boonham et al., 2003). Liebe et. al. developed a microarray assay to successfully identify several sugar beet root diseases. This innovative tool allowed for high-throughput multiplexed detection of pathogens (Liebe et al., 2016).
- **A COMBINATION OF TECHNOLOGIES:** To get the most out of these rapid diagnostic tools, researchers have combined some of the technologies. Lau et. al. developed a nanoparticle-based electrochemical biosensor for rapid detection of *Pseudomonas syringae* using disposable screen-printed carbon electrodes. This assay was coupled with recombinase polymerase amplification (RPA) to produce a method that was 10,000 times more sensitive than conventional PCR and could diagnose the presence of *Pseudomonas syringae* before disease symptoms were visible on the plant (Lau et al., 2017). In another study, Karnal bunt of wheat was detected on sight using a AuNP- based lateral flow immune-dipstick assay at the genus level. AuNPs were conjugated with anti-teliospore antibodies for improved specificity in the detection of *Tilletia indica*, the

fungal pathogen for the disease (Singh et al., 2010). Ghosh et al (2018) combined RPA with a lateral flow assay as a tool for detection of the citrus greening pathogen, *Candidatus Liberibacter asiaticus,* on mandarin oranges (Ghosh et al., 2018).

These other rapid detection technologies have helped push DNA analysis toward high throughput, low cost, and sensitive advances. Just like these methods, AuNPs have steadily grown in popularity due to their unique physical and chemical properties.

SURFACE PLASMON RESONANCE AND GOLD NANOPARTICLE PROPERTIES

AuNPs have a large surface-to-volume ratio, which gives them a platform for surface modification. This surface functionalization of the particles is often what determines the use of the material. The alteration to the particles can be done through physical adsorption or covalent attachment of ligands to their surface (Dykman & Khlebtsov, 2019). Modifications act to provide protection of the particles from aggregation, improve biocompatibility, and allow for targeted hybridizations to be used in assays. As mentioned previously, AuNPs can also be useful as unmodified materials due to their localized surface plasmon resonance. Surface plasmon resonance (SPR) is a result of the electrons on the particles' surface oscillating as they interact with light and other analyte materials (McDonnell, 2001; Pattnaik, 2005; Tang et al., 2010). Nanoparticles can occur in a multitude of shapes including spherical, cube, star, rod, cluster, and shell-shaped (A. K. Khan et al., 2014). SPR is greatly affected by the gold particles' size, shape, and environment. For example, as nanoparticle size increases the wavelength of light that is adsorbed shifts to longer, redder wavelengths (Figure 1.1). This means that larger particles will adsorb red light and reflect blue light leading to a pink, purple, or blue color of the colloidal

solution. The various sizes and shapes of the AuNPs are controlled so that they will have specific optical properties for their intended applications. This red shifting occurrence can also happen when gold particles are in an excess salt solution (Anderson et al., 2011; Baetsen-Young et al., 2018; Han et al., 2015; Li & Rothberg, 2004b; Wang et al., 2016). The surface of AuNPs is usually negatively charged, but in a salt environment, the charge becomes neutral leading to aggregation, and as a result, the gold solution turns from red to blue.

Figure 1.1: Various colors of different sized monodispersed colloidal gold nanoparticles. Particle size increases from left to right. Modified from sigmaaldrich.com.

Our study investigates the use of AuNPs as a diagnostic detection assay for DNA sequences in maize. The red shifting properties of the d-AuNPs as they aggregate or disperse in an ionic salt environment are utilized for this sequence-specific detection assay. The d-AuNPs are stabilized within a complex formed between the single-stranded DNA probe (ssDNAp) and the target dsDNA. This stability causes a color display of red/pink when target DNA is present as the nanoparticles bind to the complex loop and are dispersed. When there is no target DNA, the nanoparticles can freely aggregate as there is no loop complex to stabilize them, thus a blue/purple color is displayed. The assay is used to detect a *Xerico* insertion gene that is known to induce ABA sensitivity and improve water use efficiency in maize. As this gene was inserted into the B73 variety of maize, untransformed samples were used in the assay as a non-target negative control. Though a reproducible assay was not achieved, this study shows promise for further research to be done for a rapid DNA diagnostic tool once challenges with nanoparticle

synthesis can be overcome. Through further research and application, this assay can be used to assist breeders in their selection process with a rapid, simple method of detection of native sequences, transgenic insertions, introgressed regions, and recurrent parent DNA.

PART 2- FUSARIUM DISEASED WHEAT SEED DETECTION WITH MACHINE LEARNING

THE IMPORTANCE OF WHEAT, FUSARIUM HEAD BLIGHT AND ITS IMPACT ON GRAIN VALUE

Since the Fertile Crescent, wheat (*Triticum* spp.), the Middle East originating crop, has been amongst the world's top staple crops. Major improvements in the genetics and resistance in wheat came after World War II through the Green Revolution. Despite these great advances, wheat is still plagued by several impactful diseases. One of the most devastating of these is Fusarium Head Blight (FHB). The scab disease has caused billions of dollars in losses due to its negative effects on the nutritive, physical, and chemical qualities of the grain (Cowger et al., 2020), which lowers the market value. FHB is caused by the *Fusarium* spp. with its dominant pathogen being *Fusarium graminearum*. Symptoms of the infection are seen as bleached spike heads, beginning in a single spikelet, and spreading to the rest of the wheat head. After harvest, infection in wheat is often visualized in the kernels as a tombstone, pink or chalky color and shriveled in appearance (Figure 1.2).

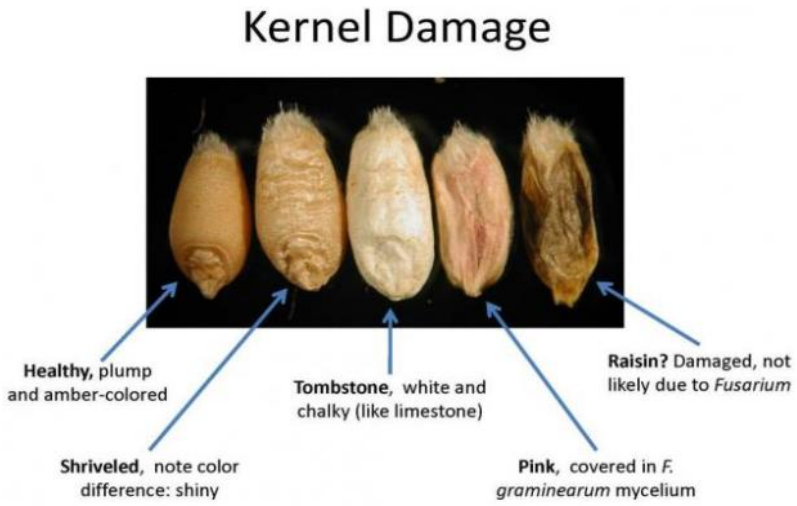

Figure 1.2: Illustration of diseased kernels. Modified from canr.msu.edu and originally from Dr. Pierce Paul, Ohio State University.

The greatest threat of this pathogen is its ability to produce the vomitoxin, deoxynivalenol (DON). DON in grain can be extremely harmful to animals and humans as it disrupts normal cellular function and can lead to nausea, fever, headaches, and vomiting (Chu, 2003). The USDA recommends DON levels not to exceed 1 part per million (ppm) and 2ppm is marked as unacceptable for wheat used in human foods (Food and Drug Administration (FDA), 2010). Due to the dangerous economic and physical effects of this infection, it is important to identify Fusarium infected seed to reduce the possibility of DON contamination.

A BRIEF HISTORY ON MACHINE LEARNING

Machine Learning is a rapidly developing technology that looks to use algorithms to assist computer systems to continually improve their performance for detecting patterns, making predictions, and analyzing data (Awad & Khanna, 2015). The term "machine learning" was coined by an IBM developer named Arthur Samuel that wanted to develop a computer program to play checkers in 1952 (Samuel, 1959)**.** A few years later, a scientist at Cornell built off of Samuel's idea and coupled it with a model of brain cell interaction that was previously published

(Hebb, 1949) to create the “Perceptron” (Rosenblatt, 1960). This software was built for the IBM 704 computer to do image recognition, classification, and simulate progressive learning. Though this invention had great promise, it struggled to successfully recognize complex visual patterns and did more binary classification. To combat this, multilayer perceptrons were developed to significantly increase the detection and classification ability of the technology (Murtagh, 1991; Mondal et al., 2018). As technology progressed and the growth of the internet boomed, machine learning began to investigate more practical problems to provide services focused on probability theory and statistics. These tools have a variety of applications from credit card fraud detection (Awoyemi et al., 2017), to facial recognition in smartphones (Alshamsi et al., 2016), self-driving cars (Stilgoe, 2018), and personalized internet advertisements (Mogaji et al., 2020). The practical application of machine learning algorithms has also been examined in the agricultural field.

MACHINE LEARNING FOR AGRICULTURAL APPLICATION

Advances in high throughput and precision agriculture have created a rapidly emerging sector that utilizes machine learning for innovative research and application. Ramos et. al. used machine learning to measure the number of fruits on a coffee branch through digital image analysis. The machine vision system was able to successfully estimate fruit number, its maturation percentage, and weight with a correlation as high as 90% at early stages of crop development (Ramos et al., 2017). Their method enabled an efficient, low-cost, and non-destructive model for coffee tree fruit counts. Another yield-related model was created for cherry tree harvesting. They developed models that classified images by parts (branch, cherry, leaf, and background), and linked segmented pictures corresponding to whole branches and trees (Amatya et al., 2016). This research shows the potential for automatic harvesting of cherry trees. K-

Nearest Neighbors was used for lettuce growth stage identification based on image analysis (James Loresco et al., 2018). They used KNN to compare color spaces for RGB, HSV, CIELab, and YCbCr. The study found CIELab color space as the most useful to use in growth stage prediction in lettuce. Nari and Yang-Won used SVM, Random Forest, Extremely Randomized Trees, and Deep Learning to predict corn yields based on satellite images and climate data (Kim & Lee, 2016). When compared to data from the USDA, their predictions differed by only 6-8%, thus showcasing machine learning as an option for crop yield modeling. In China, researchers used support vector machines to create a crop modeling system for rice (Su et al., 2017). Their study provided a model that was parametrically simple, regionally applicable, and useful on perennial and one-year rice predictions.

A wide variety of machine learning applications are also seen in disease and pest detection for several crops. The detection of thrips on strawberries grown in greenhouses was facilitated by the analysis of crop canopy images using a support vector machine classification model (Ebrahimi et al., 2017). The model identified the pests in images of strawberry flowers with an error rate of less than 2.5%. Rice blast disease was identified using machine learning algorithms, including multiple regression, neural network, and support vector machine (Kaundal et al., 2006). The models were able to achieve early detection of the disease for different locations and in different seasons. They saw that SVM was the best technique for disease identification and developed a web server for rice blast prediction. This open-access server has helped the plant science community and farmers in their decision-making processes. Logistic regression was used to predict white mold incidence in dry beans from North Dakota. The model used data on rainfall, temperature, and frequency of rain in the growing season (Harikrishnan & Del Río,

2008). It was able to explain 85% of the variability and had a high accuracy of 91%. This gave researchers an additional tool for deciding on fungicide application for mold control. Moshou et. al. utilized neural networks for the detection of yellow rust in wheat plants. The model used hyperspectral image data of wheat plants to distinguish healthy and yellow rust-infected plants during their early developmental stages (Moshou et al., 2004). The identification system was very successful in classification with accuracy ranging from 95-99%. Their research allows for the prospect of a remote sensing device for yellow rust that works in the field. Researchers developed a smartphone app that has integrated machine learning models to detect early signs of disease in bananas (Selvaraj et al., 2019). They used deep convolutional neural networks as an AI-based banana disease and pest detection system to support banana farmers in developing countries. Their model was able to achieve 90% accuracy and was transferred to a mobile app platform that tracks the class and location of various banana diseases and differentiates healthy and diseased plant parts. This classification of healthy and diseased materials is vital for the rapid diagnosis of disease in plants. Machine learning models are very useful for developing low-cost, efficient, and rapid detection tools to be used by researchers.

MODELS USED IN THIS STUDY

The machine learning approaches used for this study are all supervised classification models that include: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression (LR). K-nearest neighbors is a model that estimates how likely a data point is to belong with one group or another based on its proximity to other groups of data points. The “K” in this model represents the number of “nearest neighbor” datapoints to use in the model for association grouping of the testing dataset (Latha Jothi & Sabari, 2020). Support vector machine

is a model that separates data using a hyperplane. This hyperplane can be a linear or multi-dimensional threshold depending on how the dataset is structured (Noble, 2006). In some data sets classification is not always an easy “Yes” or “No”. Separated data are often divided by a soft margin that allows for misclassifications. The specification of the cost function (C-value) is a parameter in SVM that allows for misclassification in the model and prevents overfitting of the model (Lorena & De Carvalho, 2008). The soft margin has observations within it called “support vectors” that act to support the division of the model by the hyperplane.

Logistic Regression is a model that is used to estimate the probability of a binary dependent variable based on a logistic function which is the natural logarithm of an odds ratio. This function gives an “S” shaped curve when modeling predictions of the data (C. Y. J. Peng et al., 2002). A cutoff point can be placed on the logistic prediction curve for binary decision making by using a relative operating characteristic (ROC) curve and choosing the threshold that corresponds to the highest sensitivity and specificity for that dataset (Soureshjani & Kimiagari, 2013).

Visual assessment of wheat kernels is one of the most common ways to non-destructively diagnose Fusarium diseased kernels (FDK) and is most often done by a trained pathologist or another researcher by hand. Despite the reliability of this evaluation process, it can be very time-consuming, subjective, and not ideal for large sample sizes. Therefore, high throughput, low-cost, image-based detection methods are important for pathology and breeding research. We examine three machine learning models for the detection and classification of healthy kernels and FDK. This comparative study investigates Logistic regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) as the models of choice. Each model was

able to identify and classify thousands of healthy or infected kernels with high accuracy from 95-98.6%. The best of the compared models was logistic regression because of its fast processing time when making predictions while maintaining high model accuracy. Utilizing image-based methods for FDK identification will assist researchers to have a faster, more objective method for accurately evaluating disease severity in wheat without expensive image analysis equipment.

BIBLIOGRAPHY

BIBLIOGRAPHY

Alshamsi, H., Meng, H., & Li, M. (2016). Real-time facial expression recognition app development on mobile phones. *2016 12th International Conference on Natural Computation, Fuzzy Systems, and Knowledge Discovery, ICNC-FSKD 2016*, 1750–1755. https://doi.org/10.1109/FSKD.2016.7603442

Amatya, S., Karkee, M., Gongal, A., Zhang, Q., & Whiting, M. D. (2016). Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. *Biosystems Engineering*, *146*, 3–15. https://doi.org/10.1016/j.biosystemseng.2015.10.003

Anderson, M. J., Torres-Chavolla, E., Castro, B. A., & Alocilja, E. C. (2011). One step alkaline synthesis of biocompatible gold nanoparticles using dextrin as capping agent. *Journal of Nanoparticle Research*, *13*(7), 2843–2851. https://doi.org/10.1007/s11051-010-0172-3

Antonii, F. (1618). Panacea Aurea-Auro Potabile. *Hamburg: Ex Bibliopo Lio Frobeniano*, 250.

Arora, S., Sharma, P., Kumar, S., Nayan, R., Khanna, P. K., & Zaidi, M. G. H. (2012). Gold-nanoparticle induced enhancement in growth and seed yield of *Brassica juncea*. *Plant Growth Regulation*, *66*(3), 303–310. https://doi.org/10.1007/s10725-011-9649-z

Awad, M., & Khanna, R. (2015). Machine Learning. In *Efficient Learning Machines* (pp. 1–18). Apress. https://doi.org/10.1007/978-1-4302-5990-9_1

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *Proceedings of the IEEE International Conference on Computing, Networking and Informatics, ICCNI 2017*, *2017-January*, 1–9. https://doi.org/10.1109/ICCNI.2017.8123782

Baetsen-Young, A. M., Vasher, M., Matta, L. L., Colgan, P., Alocilja, E. C., & Day, B. (2018). Direct colorimetric detection of unamplified pathogen DNA by dextrin-capped gold nanoparticles. *Biosensors and Bioelectronics*, *101*(August 2017), 29–36. https://doi.org/10.1016/j.bios.2017.10.011

Bai, L. Y., Zhang, Y. P., Chen, J., Zhou, X. M., & Hu, L. F. (2010). Rapid, sensitive, and selective detection of pymetrozine using gold nanoparticles as colourimetric probes. *Micro and Nano Letters*, *5*(5), 304–308. https://doi.org/10.1049/mnl.2010.0115

Barber, D. J., & Freestone, I. C. (1990). *AN INVESTIGATION OF THE ORIGIN OF THE COLOUR OF THE LYCURGUS CUP BY ANALYTICAL TRANSMISSION ELECTRON MICROSCOPY* (Vol. 32).

Bean, K. (2020). *Michael Faraday*. Young Scientist Journal. https://ysjournal.com/michael-faraday/

Boonham, N., Walsh, K., Smith, P., Madagan, K., Graham, I., & Barker, I. (2003). Detection of potato viruses using microarray technology: Towards a generic method for plant viral disease diagnosis. *Journal of Virological Methods*, *108*(2), 181–187. https://doi.org/10.1016/S0166-0934(02)00284-7

Bowman, M. C., Ballard, T. E., Ackerson, C. J., Feldheim, D. L., Margolis, D. M., & Melander, C. (2008). Inhibition of HIV fusion with multivalent gold nanoparticles. *Journal of the American Chemical Society*, *130*(22), 6896–6897. https://doi.org/10.1021/ja710321g

Charlier, P., Poupon, J., Huynh-Charlier, I., Saliège, J. F., Favier, D., Keyser, C., & Ludes, B. (2009). A gold elixir of youth in the 16th century French court. In *BMJ (Online)* (Vol. 339, Issue 7735, p. 1402). British Medical Journal Publishing Group. https://doi.org/10.1136/bmj.b5311

Chu, F. S. (2003). MYCOTOXINS | Toxicology. In *Encyclopedia of Food Sciences and Nutrition* (pp. 4096–4108). Elsevier. https://doi.org/10.1016/b0-12-227055-x/00823-3

Cowger, C., Smith, J., Boos, D., Bradley, C. A., Ransom, J., & Bergstrom, G. C. (2020). Managing a Destructive, Episodic Crop Disease: A National Survey of Wheat and Barley Growers' Experience with Fusarium Head Blight. *Plant Disease*, *104*(3), 634–648. https://doi.org/10.1094/PDIS-10-18-1803-SR

Culpeper, N. (1657). Mr. Culpepper's Treatise of aurum potabile Being a description of the three-fold world, viz. elementary celestial intellectual containing the knowledge necessary to the study of hermetick philosophy. Faithfully written by him in his lifetime, and since h. *London*.

Daniel, M.-C., & Astruc, D. (2004). *Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications toward Biology, Catalysis, and Nanotechnology*. https://doi.org/10.1021/cr030698

Daraee, H., Eatemadi, A., Abbasi, E., Fekri Aval, S., Kouhi, M., & Akbarzadeh, A. (2016). Application of gold nanoparticles in biomedical and drug delivery. *Artificial Cells, Nanomedicine, and Biotechnology*, *44*(1), 410–422. https://doi.org/10.3109/21691401.2014.955107

Davis, P. (1988). Gold Therapy in the Treatment of Rheumatoid Arthritis SUMMARY. In *Canadian Family Physician* (Vol. 34). College of Family Physicians of Canada. /pmc/articles/PMC2218757/?report=abstract

Deng, X. Y., Wei, Z. M., & An, H. L. (2001). Transgenic peanut plants obtained by particle bombardment via somatic embryogenesis regeneration system. *Cell Research*, *11*(2), 156–160. https://doi.org/10.1038/sj.cr.7290081

Dykman, L. A., & Khlebtsov, N. G. (2019). Methods for chemical synthesis of colloidal gold. *Russian Chemical Reviews*, *88*(3), 229–247. https://doi.org/10.1070/rcr4843

Ebrahimi, M. A., Khoshtaghaza, M. H., Minaei, S., & Jamshidi, B. (2017). Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*, *137*, 52–58. https://doi.org/10.1016/j.compag.2017.03.016

Firrao, G., Moretti, M., Rosquete, M., Gobbi, E., & Locci, R. (2005). *NANOBIOTRANSDUCER FOR DETECTING FLAVESCENCE DORÉE PHYTOPLASMA on JSTOR*. Journal of Plant Pathology. https://www.jstor.org/stable/41998219?seq=1

Food and Drug Administration (FDA). (2010, July). *Guidance for Industry and FDA: Advisory Levels for Deoxynivalenol (DON) in Finished Wheat Products for Human Consumption and Grains and Grain By-Products used for Animal Feed | FDA*. Center for Food Safety and Applied Nutrition. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-and-fda-advisory-levels-deoxynivalenol-don-finished-wheat-products-human

Garcia, C. (1981). Gold therapy in arthritis treatment. *The Nurse Practitioner*, *6*(1), 35. https://pubmed.ncbi.nlm.nih.gov/6450900/

Ghosh, D. K., Kokane, S. B., Kokane, A. D., Warghane, A. J., Motghare, M. R., Bhose, S., Sharma, A. K., & Reddy, M. K. (2018). Development of a recombinase polymerase based isothermal amplification combined with lateral flow assay (HLB-RPA-LFA) for rapid detection of "Candidatus Liberibacter asiaticus." *PLOS ONE*, *13*(12), e0208530. https://doi.org/10.1371/journal.pone.0208530

Gibson, J. D., Khanal, B. P., & Zubarev, E. R. (2007). Paclitaxel-functionalized gold nanoparticles. *Journal of the American Chemical Society*, *129*(37), 11653–11661. https://doi.org/10.1021/ja075181k

Gopinath, K., Gowri, S., Karthika, V., & Arumugam, A. (2014). Green synthesis of gold nanoparticles from fruit extract of Terminalia arjuna, for the enhanced seed germination activity of Gloriosa superba. *Journal of Nanostructure in Chemistry*, *4*(3), 1–11. https://doi.org/10.1007/s40097-014-0115-0

Han, H., Yi, W., Hou, D., Huang, T., & Hao, Z. (2015). AuNPs-based colorimetric assay for identification of chicken tissues in meat and meat products. *Journal of Nanomaterials*, *2015*. https://doi.org/10.1155/2015/469267

Harikrishnan, R., & Del Río, L. E. (2008). A logistic regression model for predicting risk of white mold incidence on dry bean in North Dakota. *Plant Disease*, *92*(1), 42–46. https://doi.org/10.1094/PDIS-92-1-0042

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. https://pdfs.semanticscholar.org/efee/3a0d3e8b34e45188dca4e19c15e6b6029edd.pdf%3C/eref

Huaizhi, Z., & Yuantao, N. (2001). China's Ancient Gold Drugs. *Gold Bulletin*, *34*(1), 24–29.

Huang, X., Zhai, C., You, Q., & Chen, H. (2014). Potential of cross-priming amplification and DNA-based lateral-flow strip biosensor for rapid on-site GMO screening. *Analytical and Bioanalytical Chemistry*, *406*(17), 4243–4249. https://doi.org/10.1007/s00216-014-7791-y

Ismagul, A., Yang, N., Maltseva, E., Iskakova, G., Mazonka, I., Skiba, Y., Bi, H., Eliby, S., Jatayev, S., Shavrukov, Y., Borisjuk, N., & Langridge, P. (2018). A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. *BMC Plant Biology*, *18*(1). https://doi.org/10.1186/s12870-018-1326-1

James Loresco, P., Dadios, E. P., Valenzuela, I., James Loresco, P. M., & Valenzuela, I. C. (2018). Color Space Analysis Using KNN for Lettuce Crop Stages Identification in Smart Farm Setup. *IEEE Region 10 Conference*. https://doi.org/10.1109/TENCON.2018.8650209

Jayaseelan, C., Ramkumar, R., Rahuman, A. A., & Perumal, P. (2013). Green synthesis of gold nanoparticles using seed aqueous extract of Abelmoschus esculentus and its antifungal activity. *Industrial Crops and Products*, *45*, 423–429. https://doi.org/10.1016/j.indcrop.2012.12.019

Kang, J. Y., Zhang, Y. J., Li, X., Dong, C., Liu, H. Y., Miao, L. J., Low, P. J., Gao, Z. X., Hosmane, N. S., & Wu, A. G. (2018). Rapid and sensitive colorimetric sensing of the insecticide pymetrozine using melamine-modified gold nanoparticles. *Analytical Methods*, *10*(4), 417–421. https://doi.org/10.1039/c7ay02658g

Kao, C. Y., Huang, S. H., & Lin, C. M. (2008). A low-pressure gene gun for genetic transformation of maize (Zea mays L.). *Plant Biotechnology Reports*, *2*(4), 267–270. https://doi.org/10.1007/s11816-008-0067-2

Kaundal, R., Kapoor, A. A., & Raghava, G. P. S. (2006). Machine learning techniques in disease forecasting: A case study on rice blast prediction. *BMC Bioinformatics*, *7*(1), 1–16. https://doi.org/10.1186/1471-2105-7-485

Khan, A. K., Rashid, R., Murtaza, G., & Zahra, A. (2014). Gold nanoparticles: Synthesis and applications in drug delivery. In *Tropical Journal of Pharmaceutical Research* (Vol. 13, Issue 7, pp. 1169–1177). University of Benin. https://doi.org/10.4314/tjpr.v13i7.23

Khan, M., Wang, R., Li, B., Liu, P., Weng, Q., & Chen, Q. (2018). Comparative Evaluation of the LAMP Assay and PCR-Based Assays for the Rapid Detection of Alternaria solani. *Frontiers in Microbiology*, *9*(SEP), 2089. https://doi.org/10.3389/fmicb.2018.02089

Kim, N., & Lee, Y. W. (2016). Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, *34*(4), 383–390. https://doi.org/10.7848/ksgpc.2016.34.4.383

Kotz, D. (2020). *UM School of Medicine Researchers Develop Experimental Rapid COVID-19 Test Using Innovative Nanoparticle Technique*. University of Maryland School of Medicine. https://www.medschool.umaryland.edu/news/2020/UM-School-of-Medicine-Researchers-Develop-Experimental-Rapid-COVID-19-Test-Using-Innovative-Nanoparticle-Technique.html

Kumar, V., Guleria, P., Kumar, V., & Yadav, S. K. (2013). Gold nanoparticle exposure induces growth and yield enhancement in Arabidopsis thaliana. *Science of the Total Environment*, *461–462*, 462–468. https://doi.org/10.1016/j.scitotenv.2013.05.018

Latha Jothi, V., & sabari, N. S. (2020). Crop Yield Prediction using KNN Model . In *International Journal of Engineering Research & Technology* (Vol. 8, Issue 12). IJERT-International Journal of Engineering Research & Technology. www.ijert.org

Lau, H. Y., Wu, H., Wee, E. J. H., Trau, M., Wang, Y., & Botella, J. R. (2017). Specific and sensitive isothermal electrochemical biosensor for plant pathogen DNA detection with colloidal gold nanoparticles as probes. *Scientific Reports*, *7*(1), 1–7. https://doi.org/10.1038/srep38896

Li, H., & Rothberg, L. J. (2004). Label-free colorimetric detection of specific sequences in genomic DNA amplified by the polymerase chain reaction. *Journal of the American Chemical Society*, *126*(35), 10958–10961. https://doi.org/10.1021/ja048749n

Liebe, S., Christ, D. S., Ehricht, R., & Varrelmann, M. (2016). Development of a DNA Microarray-Based Assay for the Detection of Sugar Beet Root Rot Pathogens. *Phytopathology®*, *106*(1), 76–86. https://doi.org/10.1094/PHYTO-07-15-0171-R

Loos, M. (2015). Chapter 1- Nanoscience and Nanotechnology. In *Carbon Nanotube Reinforced Composites: CNR Polymer Science and Technology* (pp. 1–36). Elsevier Inc. https://doi.org/10.1016/B978-1-4557-3195-4.00001-1

Lorena, A. C., & De Carvalho, A. C. P. L. F. (2008). Evolutionary tuning of SVM parameter values in multiclass problems. *Neurocomputing*, *71*(16–18), 3326–3334. https://doi.org/10.1016/j.neucom.2008.01.031

Louis, C., & Pluchery, O. (2012). *Gold Nanoparticles for Physics, Chemistry and Biology*. World Scientific. https://books.google.com/books?id=0HX7Z_A5z2MC&pg=PA9&dq=gold+nanoparticles+in+medicine&lr=&source=gbs_selected_pages&cad=3#v=onepage&q=gold nanoparticles in medicine&f=false

Malarkodi, C., Rajeshkumar, S., & Annadurai, G. (2017). Detection of environmentally hazardous pesticide in fruit and vegetable samples using gold nanoparticles. *Food Control*, *80*, 11–18. https://doi.org/10.1016/j.foodcont.2017.04.023

McDonnell, J. M. (2001). Surface plasmon resonance: Towards an understanding of the mechanisms of biological molecular recognition. In *Current Opinion in Chemical Biology* (Vol. 5, Issue 5, pp. 572–577). Elsevier Ltd. https://doi.org/10.1016/S1367-5931(00)00251-9

Medhi, R., Srinoi, P., Ngo, N., Tran, H. V., & Lee, T. R. (2020). Nanoparticle-Based Strategies to Combat COVID-19. In *ACS Applied Nano Materials* (Vol. 3, Issue 9, pp. 8557–8580). American Chemical Society. https://doi.org/10.1021/acsanm.0c01978

Mogaji, E., Olaleye, S., & Ukpabi, D. (2020). *Using AI to Personalise Emotionally Appealing Advertisement* (pp. 137–150). Springer, Cham. https://doi.org/10.1007/978-3-030-24374-6_10

Mondal, A., Ghosh, A., & Ghosh, S. (2018). Scaled and oriented object tracking using ensemble of multilayer perceptrons. *Applied Soft Computing Journal*, *73*, 1081–1094. https://doi.org/10.1016/j.asoc.2018.09.028

Mortazavi, S., & Zohrabi, Z. (2018). Biolistic co-transformation of rice using gold nanoparticles. *Iran Agricultural Research*, *37*(1), 75–82. https://doi.org/10.22099/iar.2018.4755

Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., & Ramon, H. (2004). Automatic detection of "yellow rust" in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, *44*(3), 173–188. https://doi.org/10.1016/j.compag.2004.04.003

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5–6), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5

Nam, J. M., Thaxton, C. S., & Mirkin, C. A. (2003). Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science*, *301*(5641), 1884–1886. https://doi.org/10.1126/science.1088755

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

Pattnaik, P. (2005). Surface plasmon resonance: Applications in understanding receptor-ligand interaction. In *Applied Biochemistry and Biotechnology* (Vol. 126, Issue 2, pp. 79–92). Humana Press. https://doi.org/10.1385/abab:126:2:079

Pecchia, S., & Da Lio, D. (2018). Development of a rapid PCR-Nucleic Acid Lateral Flow Immunoassay (PCR-NALFIA) based on rDNA IGS sequence analysis for the detection of

Macrophomina phaseolina in soil. *Journal of Microbiological Methods*, *151*, 118–128. https://doi.org/10.1016/j.mimet.2018.06.010

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Peng, H., & Chen, I. A. (2019). Rapid Colorimetric Detection of Bacterial Species through the Capture of Gold Nanoparticles by Chimeric Phages. *ACS Nano*, *13*(2), 1244–1252. https://doi.org/10.1021/acsnano.8b06395

Rahman, M. T., & Rebrov, E. V. (2014). Microreactors for gold nanoparticles synthesis: From faraday to flow. In *Processes* (Vol. 2, Issue 2, pp. 466–493). MDPI AG. https://doi.org/10.3390/pr2020466

Ramos, P. J., Prieto, F. A., Montoya, E. C., & Oliveros, C. E. (2017). Automatic fruit count on coffee branches using computer vision. *Computers and Electronics in Agriculture*, *137*, 9–22. https://doi.org/10.1016/j.compag.2017.03.010

Rojas, J. A., Miles, T. D., Coffey, M. D., Martin, F. N., & Chilvers, M. I. (2017). Development and application of qPCR and RPA genus and species-specific detection of Phytophthora sojae and P. Sansomeana root rot pathogens of soybean. *Plant Disease*, *101*(7), 1171–1181. https://doi.org/10.1094/PDIS-09-16-1225-RE

Rosenblatt, F. (1960). Perceptron Simulation Experiments. *Proceedings of the IRE*, *48*(3), 301–309. https://doi.org/10.1109/JRPROC.1960.287598

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., & Blomme, G. (2019). AI-powered banana diseases and pest detection. *Plant Methods*, *15*(1), 92. https://doi.org/10.1186/s13007-019-0475-z

Shah, V., & Belozerova, I. (2009). Influence of metal nanoparticles on the soil microbial community and germination of lettuce seeds. *Water, Air, and Soil Pollution*, *197*(1–4), 143–148. https://doi.org/10.1007/s11270-008-9797-6

Singh, S., Singh, M., Agrawal, V. V., & Kumar, A. (2010). An attempt to develop surface plasmon resonance based immunosensor for Karnal bunt (Tilletia indica) diagnosis based on the experience of nano-gold based lateral flow immuno-dipstick test. *Thin Solid Films*, *519*(3), 1156–1159. https://doi.org/10.1016/j.tsf.2010.08.061

Soureshjani, M. H., & Kimiagari, A. M. (2013). Calculating the best cut-off point using logistic regression and neural network on credit scoring problem-A case study of a commercial

bank. *African Journal of Business Management*, *7*(16), 1414–1421. https://doi.org/10.5897/AJBM11.394

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science*, *48*(1), 25–56. https://doi.org/10.1177/0306312717741687

Stoeva, S. I., Lee, J. S., Smith, J. E., Rosen, S. T., & Mirkin, C. A. (2006). Multiplexed detection of protein cancer markers with biobarcoded nanoparticle probes. *Journal of the American Chemical Society*, *128*(26), 8378–8379. https://doi.org/10.1021/ja0613106

Su, Y. xue, Xu, H., & Yan, L. jiao. (2017). Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi Journal of Biological Sciences*, *24*(3), 537–547. https://doi.org/10.1016/j.sjbs.2017.01.024

Tang, Y., Zeng, X., & Liang, J. (2010). Surface plasmon resonance: An introduction to a surface spectroscopy technique. *Journal of Chemical Education*, *87*(7), 742–746. https://doi.org/10.1021/ed100186y

Taylor, A. C. (2010). Advances in nanoparticle reinforcement in structural adhesives. In *Advances in Structural Adhesive Bonding* (pp. 151–182). Elsevier Inc. https://doi.org/10.1533/9781845698058.1.151

Thakur, R. K., Dhirta, B., & Shirkot, P. (2018). Studies on effect of gold nanoparticles on Meloidogyne incognita and tomato plants growth and development. In *bioRxiv* (p. 428144). bioRxiv. https://doi.org/10.1101/428144

Thomas, M., & Klibanov, A. M. (2003). Conjugation to gold nanoparticles enhances polyethylenimine's transfer of plasmid DNA into mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9138–9143. https://doi.org/10.1073/pnas.1233634100

Torney, F., Trewyn, B. G., Lin, V. S. Y., & Wang, K. (2007). Mesoporous silica nanoparticles deliver DNA and chemicals into plants. *Nature Nanotechnology*, *2*(5), 295–300. https://doi.org/10.1038/nnano.2007.108

Tweney, R. D. (2006). Discovering discovery: How faraday found the first metallic colloid. *Perspectives on Science*, *14*(1), 97–121. https://doi.org/10.1162/posc.2006.14.1.97

Ventura, B. Della, Cennamo, M., Minopoli, A., Campanile, R., Bollet-Ti Censi, S., Terracciano, D., Portella, G., & Velotta, R. (2020). *Colorimetric Test for Fast Detection of SARS-CoV-2 in Nasal and Throat Swabs*. https://doi.org/10.1101/2020.08.15.20175489

Wang, L., Liu, Z., Xia, X., & Huang, J. (2016). Visual detection of: Maize chlorotic mottle virus by asymmetric polymerase chain reaction with unmodified gold nanoparticles as the colorimetric probe. *Analytical Methods*, *8*(38), 6959–6964. https://doi.org/10.1039/c6ay02116f

CHAPTER 2: THE USE OF DEXTRIN-CAPPED GOLD NANOPARTICLES FOR THE DETECTION OF TRANSGENIC INSERTIONS IN MAIZE

ABSTRACT:

DNA detection techniques are essential to the science community and our everyday lives, with applications in pathogenic and genetic disease diagnosis, forensic analysis, crop breeding, and much more. Through recent advances in technology, rapid, low-cost, and more efficient techniques had been created. Gold nanoparticle (AuNP) technology has emerged as a versatile tool in rapid diagnostic and bio-sensory assays. Here we present a novel colorimetric assay for the detection of DNA sequences in maize using unmodified gold nanoparticles and unamplified DNA. For this sequence-specific assay, we exploit the red shifting properties of the AuNPs that are caused by its surface plasmon resonance. In a salt environment, the AuNPs are stabilized within a loop complex between the single-stranded DNA probe (ssDNAp) and the target dsDNA. This stability results in a colorimetric response that is dependent upon the presence of target DNA in a sample. If the target gene of interest is present, the assay solution will turn red/pink and if there is no gene of interest present in the sample, the assay will turn blue/purple. AuNPs show promise in their ability to accurately diagnose the presence of transgenic insertions in DNA samples within 10 minutes. Through further research and development, this assay can be used to assist breeders in their selection process with a rapid, simple method of detection of native sequences, transgenic insertions, introgressed regions, and recurrent parent DNA.

INTRODUCTION:

Over the last 30 years, the science community has been employing the use of methods such as Polymerase Chain Reaction (PCR), Restriction Fragment Length Polymorphism (RFLP), Short Tandem Repeat (STR) Analysis, and several others for genetic sequence analysis. Though these methods are widely used, each have their own set of drawbacks ranging from processing time and efficiency to overall costs associated. In recent years, gold nanoparticles (AuNPs) have been applied as a DNA detection tool for diagnostic and bio-sensory assays ranging from cancer detection in hospitals, to virus and pathogen detection in the field (X. Bai et al., 2020; Dykman & Khlebtsov, 2011; Giraldo et al., 2019; Vetrone et al., 2012). AuNPs have been extensively used because of their stability, and controlled geometrical, visual, and surface chemical properties. Target DNA-induced aggregation of AuNPs has been shown to result in color changes in gold nanoparticle solutions due to the electrostatic interactions between the negatively charged surface of the AuNPs and the exposed nucleotide bases (Izanloo, 2017). Our goal is to develop and optimize the characteristics of AuNPs as an unamplified genomic DNA biosensor in maize for breeding applications.

ASSAY FOUNDATION

Our study builds from the scheme presented by Baetsen et. al. regarding their experiments for the detection of viral DNA of cucurbit downy mildew in cucumber. They were able to use unmodified gold nanoparticles to detect very low concentrations of viral DNA (Baetsen-Young et al., 2018). In addition to this rapid diagnostic tool, they proposed that in the presence of elevated NaCl conditions and target DNA, AuNPs are stabilized within a complex created by generated ssDNA after the hybridization of genomic dsDNA and ssDNA probe during sample

denaturation and annealing steps (Figure 2.1). In the presence of non-target DNA, this complex would not form and AuNPs would aggregate as they are adsorbed to ssDNA probes that do not hybridize with the non-target DNA. This difference in stability and aggregation of the d-AuNPs is what causes a colorimetric response. This electrostatic interaction between DNA and gold nanoparticles is consistent with studies in the literature where the charged surface of the nanoparticles bind to the nucleotide bases (Arvizo et al., 2010; Brown et al., 2000; Vorobjev et al., 2019). The exploitation of this physical property of AuNP is the basis for our study.

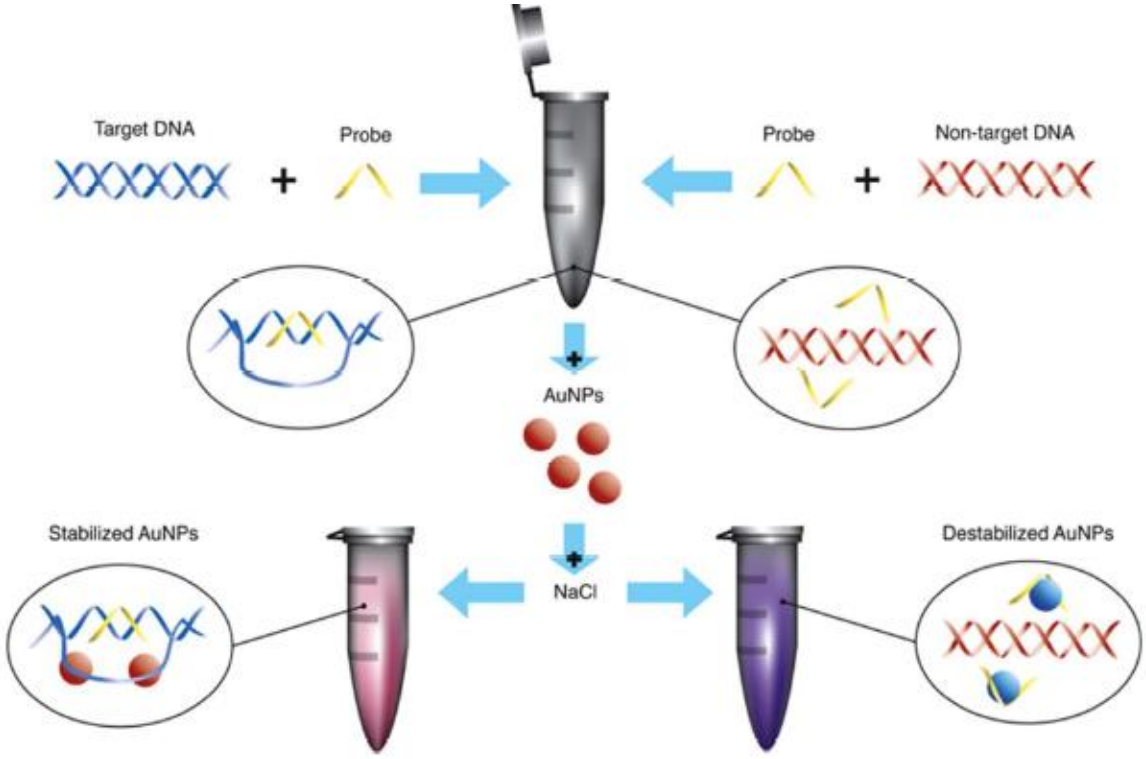

Figure 2.1: The proposed mechanism for the interaction of the target and non-target dsDNA, ssDNA probe, and d-AuNPs in a high salt concentration environment. Modified from Baetsen et. al., 2018.

GOLD NANOPARTICLE PROPERTIES

Gold nanoparticles are largely applied for their optical properties caused by the oscillation of electrons on the surface of the particles called surface plasmon resonance (X. Bai et al., 2020; Bayazit et al., 2016; Zhu & Gao, 2018). This unique physical property is what allows AuNPs to exhibit color changes when interacting with various materials such as DNA. Nanoparticles also have specific size and shape-related electronic properties and excellent biocompatibility (Yeh et al., 2012). The aggregation and stability of nanoparticles cause color shifts of aqueous AuNPs solutions, resulting in blue or red solutions respectively (Dykman & Khlebtsov, 2011). Nanoparticle-based assays for the detection of genomic DNA have been developed previously. Deng et. al. did a study in 2012 showing the usefulness of AuNPs for the detection of *Bacillus anthracis*. They found that when coupled with asymmetric polymerase chain reaction for amplification of the target sample, functionalized AuNPs provided a colorimetric response assay (H. Deng et al., 2012). In several examples in the literature, nanoparticles were "functionalized" with the attachment of a ssDNA probe to the surface of the particle for ensuring the specificity of their assay (Franco et al., 2015; Khaliliazar et al., 2020; Zhou et al., 2016). In addition to this, unmodified AuNPs were also shown to be viable for specific detection assays (H. Deng et al., 2013; Han et al., 2015; Hussain et al., 2013; Li & Rothberg, 2004a; Liu et al., 2011). In Hussain et al.'s study, they even took the method a step further by using unamplified DNA samples to detect *Mycobacterium tuberculosis* by using a restriction digestion of the mycobacterium DNA for rapid, sensitive detection.

This study investigates the use of AuNPs as a diagnostic detection assay for DNA sequences in maize. The aggregation and dispersion characteristics of the d-AuNPs in an ionic salt

environment are utilized for this sequence-specific detection assay. The d-AuNPs form a complex between the single-stranded DNA probe (ssDNAp), the target DNA, and the nanoparticles to achieve stability. This stability causes a color display of red/pink when target DNA is present, but when there is no target DNA, a blue/purple color is displayed. Through further research and application, we hope to use this assay to assist breeders in their selection process with a rapid simple method of detection of native sequences, transgenic insertions, introgressed regions, and recurrent parent DNA.

MATERIALS AND METHODS:

MATERIALS

This experiment used genomic DNA from maize plants grown in plant breeding research fields at Michigan State University. Maize varieties used include B73 and transgenic Xerico lines introgressed into B73 via backcrossing. Maize plants were previously transformed with an inserted *Xerico* gene patented by Han and Ko in 2011 that originated in *Arabidopsis thaliana* (Han, Kyung-Hwan & Ko, 2011; Ko et al., 2006). The *Xerico* gene encodes a small protein with an N-terminal transmembrane domain and a RING-H2 zinc finger motif located on the C-terminus. Over expression of this RING domain in maize has been seen to induce ABA hypersensitivity and improved water use efficiency, enhancing yield performance in drought conditions (Brugière et al., 2017).

PRIMER DESIGN

DNA primers were developed using the Integrated DNA Technologies (IDT) Primer Quest and Oligo Analyzer (Found at idtdna.com). These tools enable the design of oligonucleotides with unique predicted biophysical, chemical and hybridization properties (Owczarzy et al., 2008). The ssDNA oligonucleotide 5'- GTGCAAGAAACAGGCAGACA-3'was synthesized by Integrated DNA Technologies (Coralville, IA). The target sequence for the ssDNA probe was 5'-TGTCTGCCTGTTTCTTGCAC-3' which is within the genomic DNA of the *Xerico* insertion. Sequences were analyzed using the NCBI Basic Local Alignment Search Tool (BLAST) to identify any regions of similarity within the *Xerico* gene and native B73 v3. maize DNA. The *Xerico* gene was compared to the B73 v.3 DNA sequence to identify any regions of similarity. Once identified, these regions were excluded from the possible locations for primer

development. From the possible primer results, the Oligo Analyzer tool was used to examine the sequence for any predicted hairpin loops, self-dimers, and heterodimers. Primers were chosen based on recommendations from IDT protocol.

DNA EXTRACTION AND SAMPLE VERIFICATION

Genomic DNA was extracted from *Zea mays* plants using the DNeasy® Plant Mini Kit from Qiagen (Venlo, Netherlands). Four-centimeter-long leaf samples were flash-frozen using liquid nitrogen and homogenized in a Qiagen Tissue Lyser for 2 minutes at a frequency of 30 Hz (1800 oscillations/minute) using pre-chilled sample holder plates that were stored in a -80° C freezer. DNA samples were purified according to manufacturer protocols listed for the kit. DNA concentration and purity were quantified by Qubit (Thermofisher, Waltham, MA). After extraction, DNA was stored in a -20°C freezer for later use. Sample sequences were verified with PCR and Gel electrophoresis (Figure 2.2) using the following PCR protocol: initial denaturation at 95°C for 2 mins; 35 cycles of 95°C for 30 seconds, annealing at 63°C for 30 seconds, and extension at 72°C for 1 minute; followed finally by 72°C for 5 minutes and held at 10°C until ready for storage.

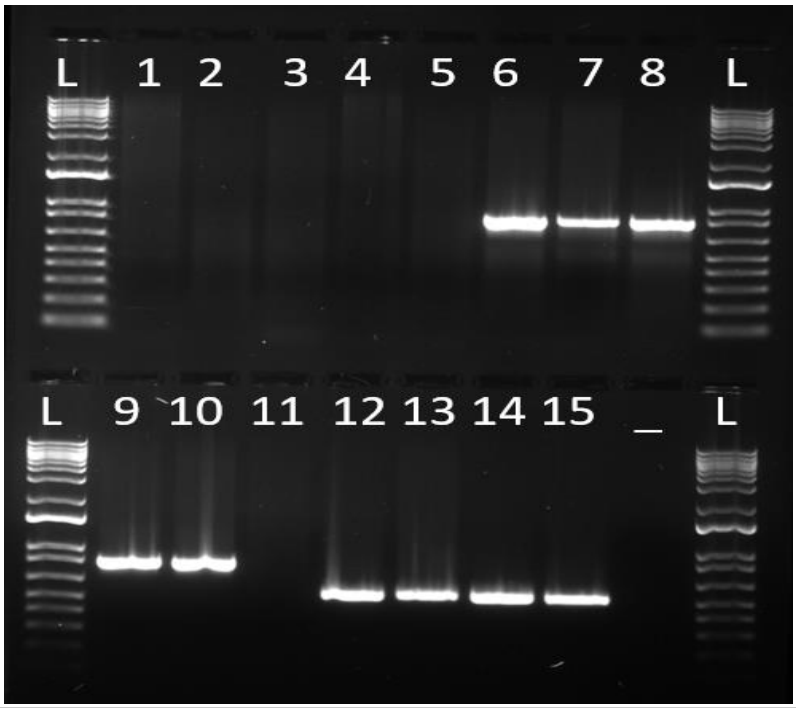

Figure 2.2: Gel analysis of PCR done on B73 (1-5) and Xerico DNA (6-15) samples collected from leaf tissue. The difference in band size represents maize lines that have an inducible promoter (6-10) and those with a constitutive promoter (11-15). None of the B73 samples showed any presence of the *Xerico* insertion. All Xerico samples showed the presence of the *Xerico* gene except one sample (11).

GOLD NANOPARTICLE AND REAGENT SYNTHESIS

Dextrin-capped gold nanoparticles (approximately 13nm in diameter) were synthesized utilizing methods demonstrated by Anderson in 2010 and once again by Baetsen-Young in 2018 (Anderson et al., 2011; Baetsen-Young et al., 2018). A gold chloride ($HAuCl_4$) stock solution was prepared with distilled sterile water for a 20mM concentration and stored under refrigeration. 5mL of the $HAuCl_4$ solution was added to 20mL of dextrin stock prepared at a 25g/L concentration and was added to a 250mL flask. The pH of the solution was adjusted to 9.0 using 1% filter sterile sodium carbonate (Na_2CO_3). The final reaction volume was adjusted to 50mL by adding 25mL of pH 9.0 sterile distilled water. The solution was wrapped in tinfoil and

incubated on a stir plate at 50°C at 250rpm for 8 hours. The solution was checked regularly to evaluate particle formation through color change stages as exhibited by Anderson. The solution went from clear, to light purple, dark purple, bright red, and wine red within the 8hr period. Once complete, nanoparticles were evaluated by TEM to evaluate the average size, shape, and uniformity of particles (Figure 2.3). AuNP absorption was also evaluated with a SpectraMax ABS Plus Microplate Reader (Molecular Devices, Sunnyvale, CA).

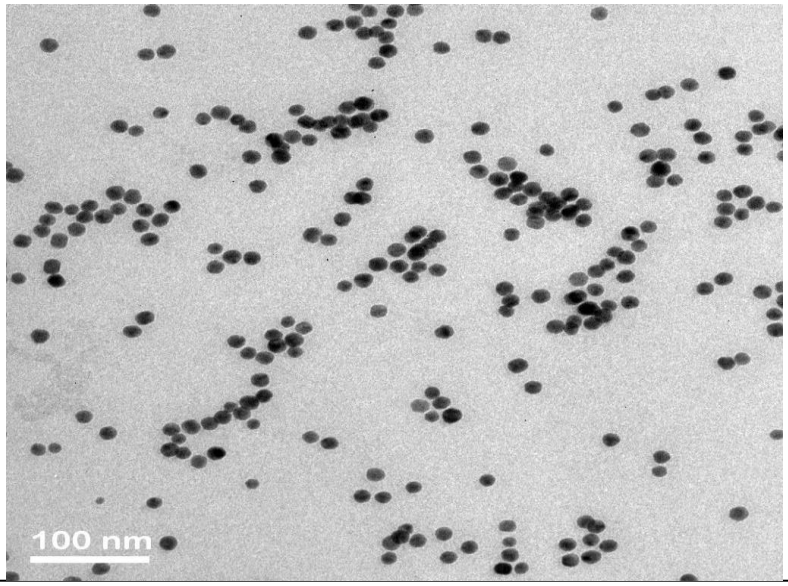

Figure 2.3: Gold nanoparticle batch under Transmission Electron Microscope (TEM). This batch of gold nanoparticles was seen to have uniform round shape and 13nm diameter.

A phosphate-buffered saline (PBS) solution was prepared using the protocol from Chazotte for a 10mM PBS stock solution (Chazotte, 2012). Briefly, 8g of NaCl, 0.2g of KCl, 1.44g of Na_2HPO_4, and 0.24g of KH_2PO_4 were added to 800mL of sterile distilled water and the pH was adjusted to 7.4 with HCl. Afterward, the volume of the solution was adjusted to 1L with sterile distilled water and was autoclaved. A 1.5mM NaCl stock solution was made with 4.38g NaCl dissolved in 50mL sterile distilled water and was autoclaved.

AUNP ASSAY DEVELOPMENT

This diagnostic assay has five components including PBS buffer, ssDNA probe, dsDNA sample, dextrin-capped gold nanoparticles (d-AuNPs), and salt solution (Figure 2.4). The B73 dsDNA was used as a non-target negative control for assay testing, while the Xerico dsDNA was the target material. For assay testing, each experiment was done in 5 different control wells. Each well had a differing mix of DNA and probe samples based on if it was a positive or negative control (Table 2.1). Briefly, control well 1 contained a probe and target DNA sample. Well 2 did not have the probe or target DNA. Well 3 contained the probe but not target DNA. Well 4 contained target DNA, but not the probe. Well 5 contained the probe and non-target B73 DNA. The volumes used for each test were based on the final reaction concentrations of each component within a 100µL reaction except the d-AuNP which was always 20µL. The reaction concentration of the target and non-target dsDNA samples was 1.5ng/µL, the Xerico probe was 0.05µM, the NaCl concentration was based on results from the salt series dilution mentioned below, and the volume of reaction was adjusted to 100µL by using 50mM PBS buffer. Reactions were denatured at 95°C for 5min, followed by annealing at 64°C for 1min, and cooled at 23°C for 10min before adding 20µL of AuNPs, followed by the appropriate volume of NaCl. Immediately after, reactions were aliquoted to a clear plate and measured at 520 and 620nm absorbance values at 1-minute intervals over 10 min.

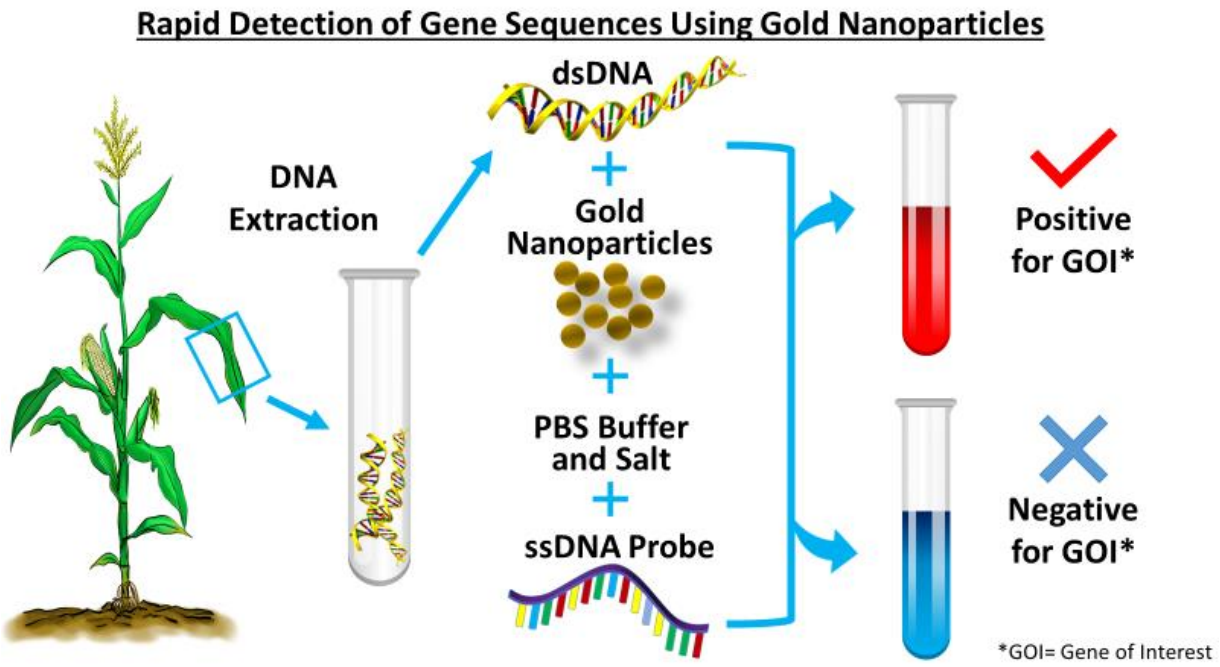

Figure 2.4: Infographic of AuNP assay for rapid detection.

Experimental Design	Control 1	Control 2	Control 3	Control 4	Control 5
50mM PBS	Yes	Yes	Yes	Yes	Yes
Xerico probe	Yes	No	Yes	No	Yes
Xerico dsDNA sample	Yes	No	No	Yes	No
B73 dsDNA sample	No	No	No	No	Yes
d-AuNPs	Yes	Yes	Yes	Yes	Yes
1.5M NaCl	Yes	Yes	Yes	Yes	Yes

Table 2.1: The experimental design for each assay development test. Each control represented a different reaction well on a plate.

Prior to experiments with the dsDNA, the stability of the d-AuNPs was evaluated with a salt series dilution study where 5µL of our 1µM ssDNA probe was added to 10µL of d-AuNPs. Then varying volumes of NaCl and PBS buffer were used to achieve salt concentrations of 0, 50, 100, 150, 200, …450mM in a final reaction volume of 50µL per reaction. The visible absorption spectrum of the AuNP aggregation was measured by the SpectraMax plate reader mentioned above to determine the ideal salt concentration to use for assay development. Plate readings were

taken at 1-minute intervals over 10 minutes. Absorption measurements were taken at the 520 and 620nm absorbance values as described by Baetsen et. al., 2018.

SPECTRAL ANALYSIS FOR AUNP AGGREGATION

Spectral data was formatted, analyzed, and plotted using R (Figure 2.5) (R Team, 2020). The aggregation and stability of the AuNPs are seen visibly in samples, as control solutions without the presence of target DNA should turn blue, while control solutions with the presence of target DNA will turn red (Figure 2.6). The rate of aggregation of the AuNPs was calculated by dividing the absorbance measurement at 620nm by the measurement at 520nm.

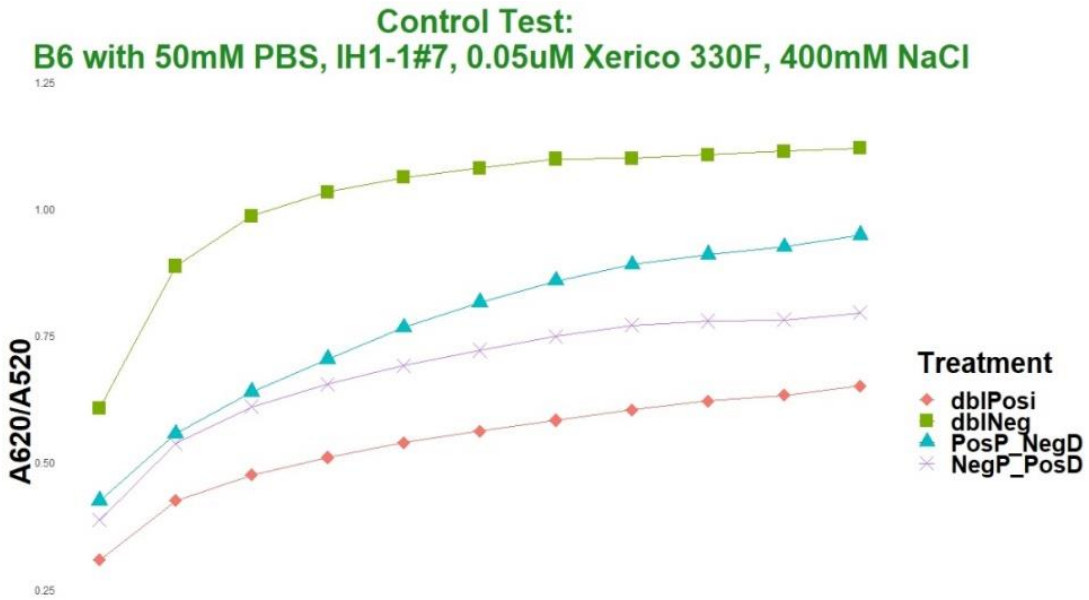

Figure 2.5: Spectral results for an ideal AuNP assay test. Control 1 ("dblPosi") shows the lowest rate of aggregation, Control 2 ("dblNeg") shows the highest rate of aggregation, and controls 3 ("PosP_NegD") and 4 ("NegP_PosD") show rates in between the other two. The colorimetric response related to these results are seen in Figure 2.5. Quantification of color intensity was done in R, after measurements taken with a SpectraMax Plus.

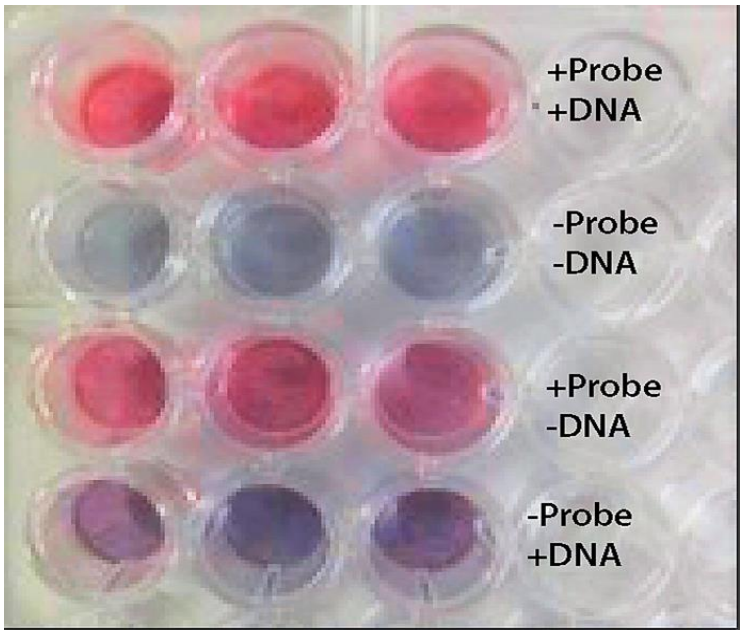

Figure 2.6: Ideal colorimetric response to the 10-minute assay. The colorimetric response of results seen in Figure 2.4. The wells for control 2 show a blue color indicating no presence of target DNA. The wells for control 1 show a red color indication the presence of target DNA. The wells for controls 3 and 4 show a purple color, indicating partial stability of the AuNPs from the probe or DNA.

RESULTS AND DISCUSSION:

ASSAY DEVELOPMENT AND TROUBLESHOOTING

We started our assay development using a XERICO reverse probe (5'-GAATTTCGACAAACACACAGAAC-3') but ran into issues with control 3 (+P -Target DNA) showing lower and similar rates of aggregation as control 1 (+P +Target DNA) (Figure 2.7). Ideally, we would want to see an assay test where control 2 (-P-DNA) shows the highest rate of aggregation and control 1 shows the lowest rate of aggregation. The 5 components of the d-AuNP assay - PBS buffer, dsDNA sample, ssDNA probe, NaCl concentration, and a batch of d-AuNPs - were changed one by one to optimize results for the assay. Improvements in the assay were not seen until the probe sequence was changed to a forward probe (5'-CCAAGGGGATTCAGAGATCA-3'). This probe was selected because it had fewer self-dimers with higher Delta G values and a higher GC content. This change would reduce the probe's ability of binding to itself and give it a stronger bond to target sequences. When this was done, the assay's testing concentrations were optimized for the NaCl and probe concentration to 200mMol and 0.05 µM, respectively. Reproducibility was seen in the assay as we tested these parameters with multiple target dsDNA samples (Figure 2.8). These results were good, but we wanted to increase the separation rate of aggregation curves between the controls. This would allow for a clearer distinction of controls and better confidence in the assay.

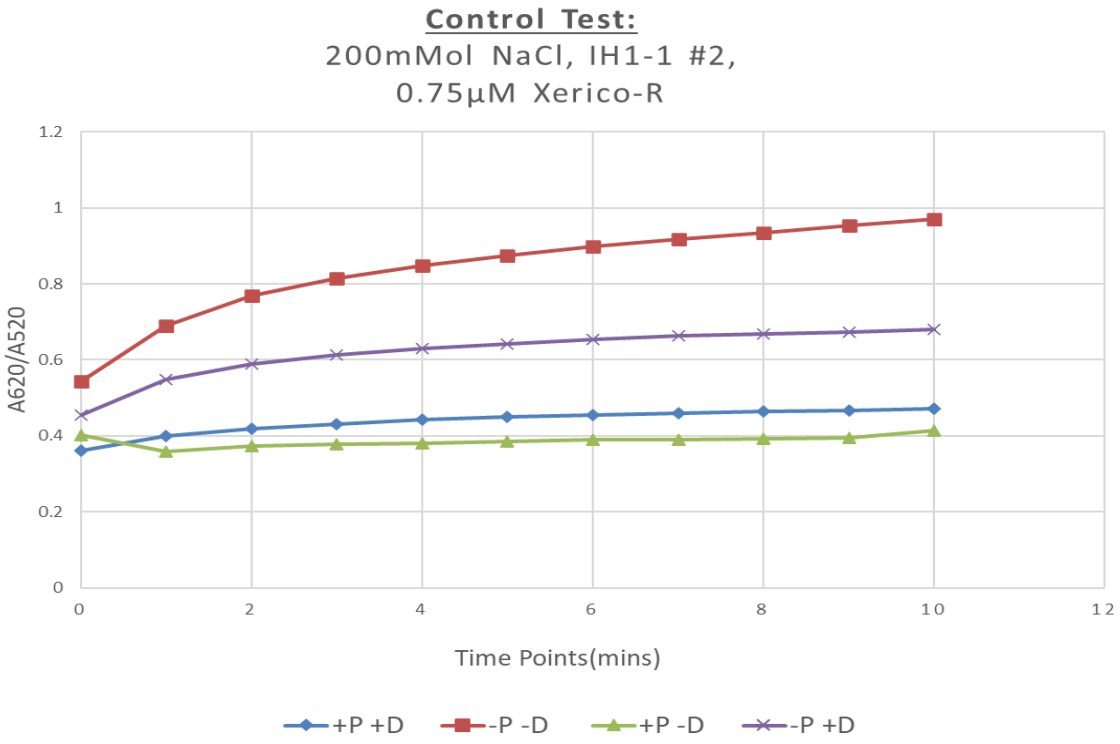

Figure 2.7: d-AuNP assay test showing rate of aggregations for the controls with not ideal results. We do not see ideal conditions with this test as the control 3 (positive for ssDNA probe, and negative for target DNA) showed a lower rate of aggregation than control 1 (positive for ssDNA probe, and positive for target DNA).

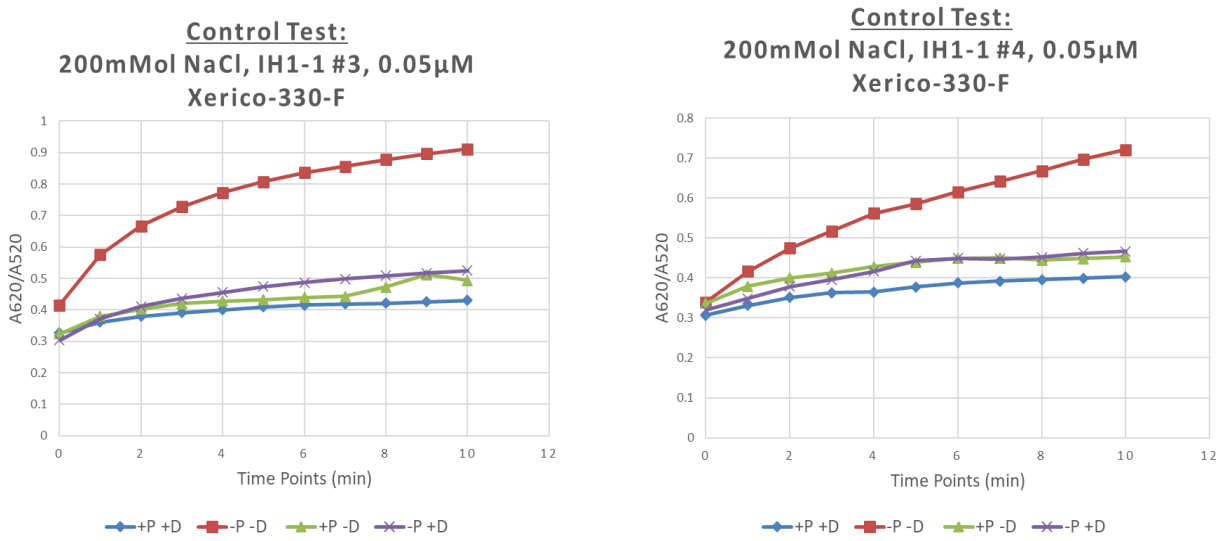

Figure 2.8: d-AuNP assay test showing rate of aggregations for the controls with reproducibility. Here we see reproducibility of the assay with different target dsDNA samples with optimized testing concentrations.

To do this, we changed the d-AuNP batch along with optimizing the assay's testing concentrations for this specific batch of gold nanoparticles. The results of this optimization can be seen in Figures 2.5 and 2.6 where there is a clear distinction between controls and the colorimetric response of the assay. After achieving these ideal results, an additional 5th control was added to the assay to ensure the experiment was not falsely detecting non-target DNA (Table 2.1). When introduced in the assay, it showed good results as control 1 had the lowest absorption measurement after 10 minutes, and control 2 had the largest, with the other controls in between (Figure 2.9). However, as seen in the previous testing, there was not a clear distinction between all controls (control 1 and 5 showed a similar measurement).

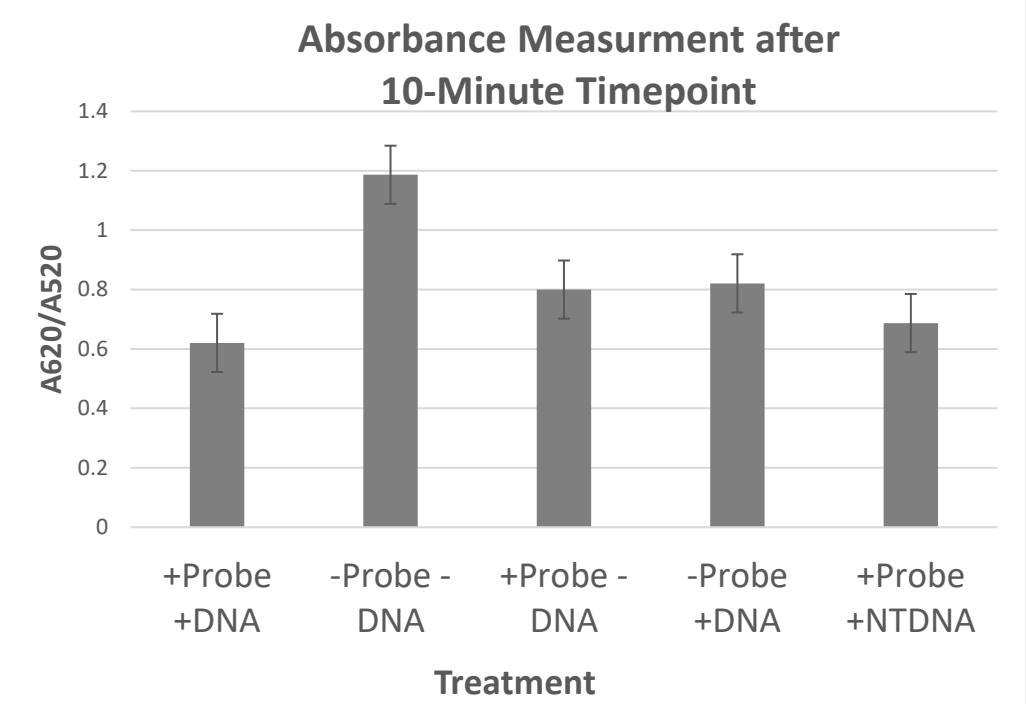

Figure 2.9: d-AuNP assay test showing absorbance measurements after 10 minutes. This shows the introduction of the 5th control using the first Xerico forward probe, a different d-AuNP batch, with 50mM PBS, Xerico target dsDNA, non-target B73 DNA, and 350mM NaCl.

To examine this further, the coding sequence for the *Xerico* insertion was Blast against the B73 v4. coding sequence within the MaizeGDB database (Portwood et al., 2019). This allowed us to find six regions of similarity between the *Xerico* insertion and the *Zea mays* reference genome for B73 (Figure 2.10). Due to the forward probe being within one of the regions of similarity, there was no clear differentiation between the absorbance measurements of control 1 and 5 (Figure 2.9). Future probes were designed to exclude targeting within these regions of similarity. A new forward Xerico probe was utilized for future experiments (5'-GTGCAAGAAACAGGCAGACA-3'). With this new forward probe, we were able to see good results (Figure 2.11), but when attempting to repeat the experiment, a new PBS buffer had to be made due to stock contamination. With the new PBS buffer, we did not achieve similar results as when the second probe was originally introduced as there was less distinction between the controls even though the assay's testing concentrations were the same. This lack of reproducibility in the assay testing made it difficult to assess the viability of the assay.

XERICO-cds
ATGGGTCTATCAAGTCTTCCTGGTCCATCAGAAGGAATGTTATGTGTGATATTAGTTAATACAGCA
TTATCGATCTCCATTGTCAAAGGCATTGTAAGATCATTCCTTGGCATAGTAGGAATCAGTCTCTCG
CCGTCTTCATCCTCGCCTTCTTCGGTGACGGTATCTTCAGAGAATTCATCAACTTCAGAGTCATTT
GATTTCCGGGTCTGCCAACCAGAGAGTTACCTTGAGGAGTTCAGGAACCGGACTCCGACACTG
AGGTTTGAGAGCTTGTGCAGGTGCAAGAAACAGGCAGACAATGAGTGTTCTGTGTGTTTGTCG
AAATTCCAAGGGGATTCAGAGATCAACAAGCTCAAGTGCGGCCATTTGTTTCACAAAACATGCT
TGGAGAAATGGATAGACTATTGGAACATCACTTGCCCATTGTGTAGGACTCCTCTTGTTGTTGTG
CCAGAAGACCATCAGCTTTCTTCTAATGTTTGGTGA

Figure 2.10: The coding sequence for the *Xerico* insertion. Highlighted in yellow are regions of similarity between the insertion and the B73 v.4 reference gene from MaizeGDB. There was a total of 6 regions of similarity.

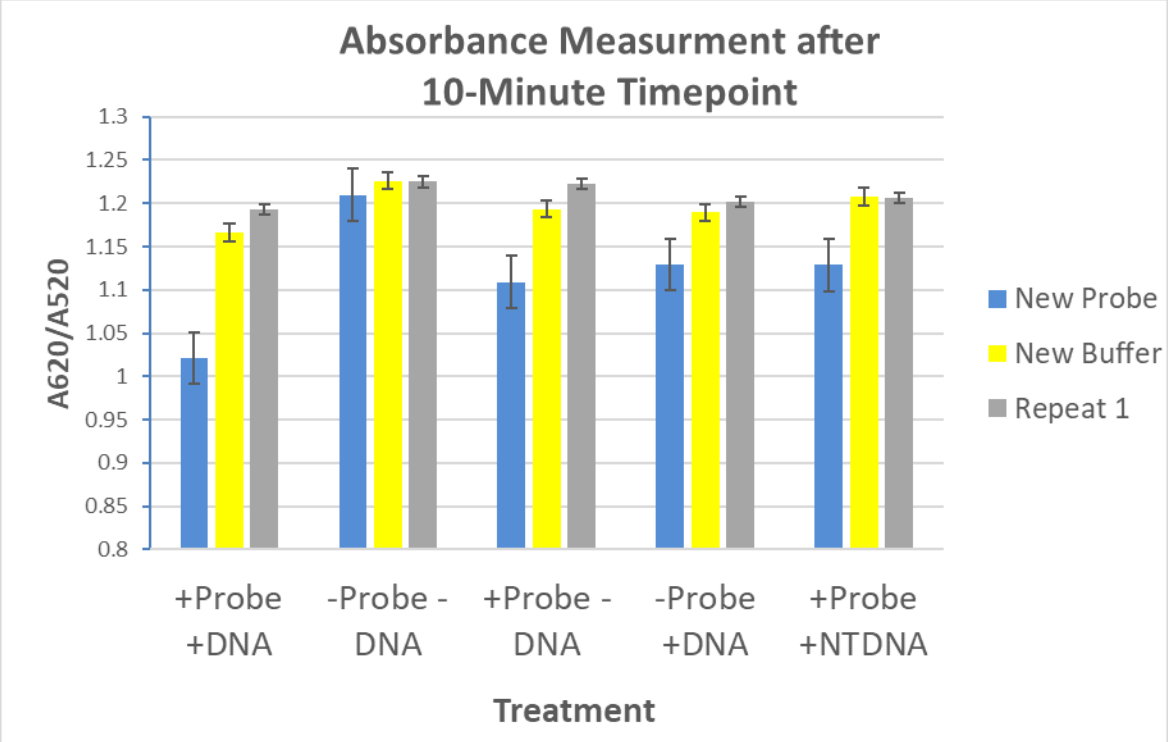

Figure 2.11: Absorbance measurements after 10 minutes for assay test. Blue bars represent the assay test when the second forward probe was tested. The yellow bars represent when the second forward probe was tested with a new PBS buffer. The grey bars represent a repeat of the experiment.

After attempting to troubleshoot the assay, new AuNP batches were made to see if the age of the d-AuNPs was affecting the consistency of the assay. New batches of d-AuNPs were synthesized and analyzed under a TEM and a full spectrum analysis of the batches was done to test absorbance. Full-spectrum analysis shows a lower absorbance measurement at the 520nm value in the old batches of d-AuNPs as compared to the new batches (Figure 2.12).

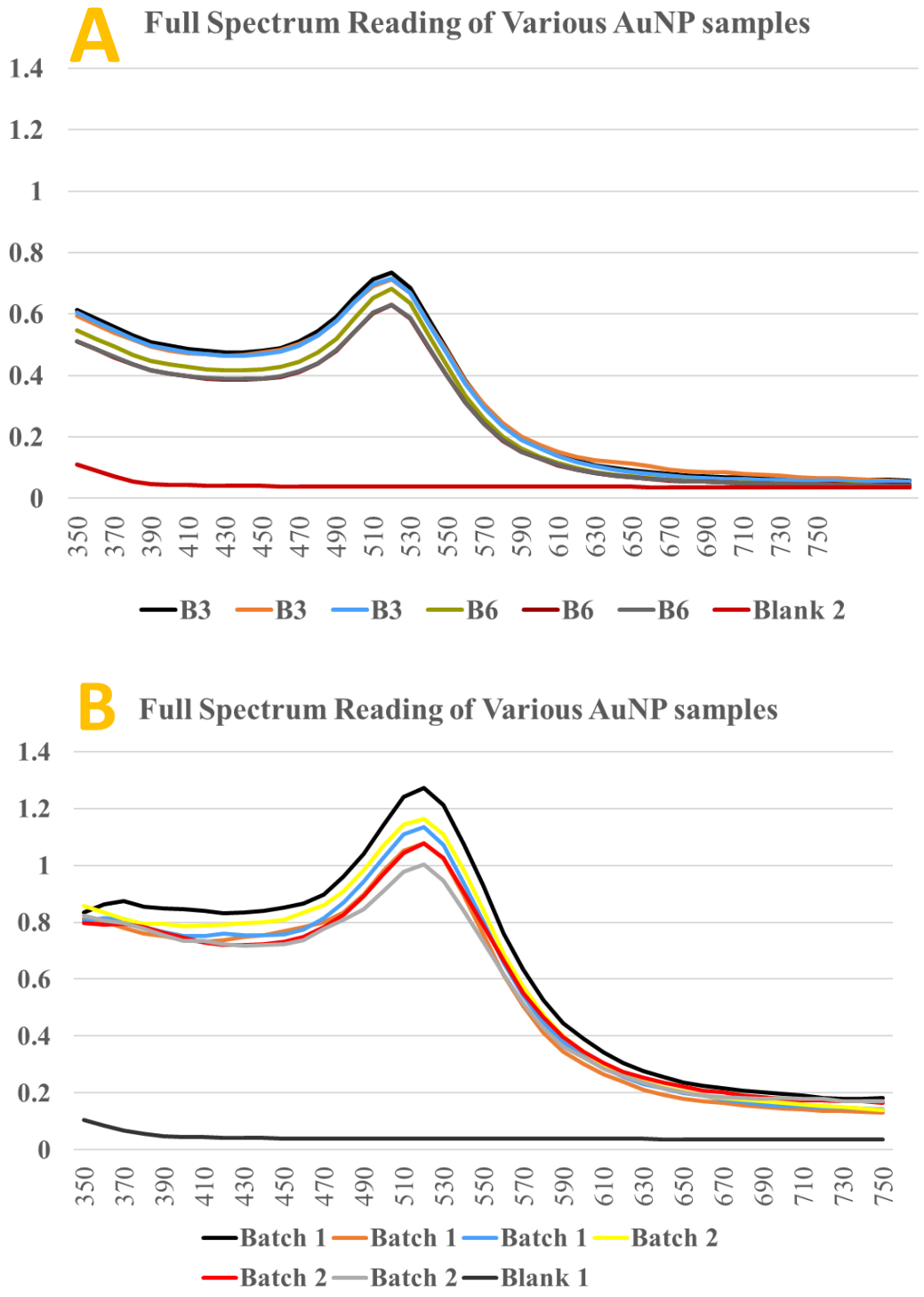

Figure 2.12: Full spectrum analysis of d-AuNP batches. (A) represents the old batches of particles and **(B)** represent new batches. New batches show higher absorbance peaks at the 520nm value as compared to the old batches.

In addition to this, a salt series dilution test was done at varying NaCl concentrations to examine the old batches' responsiveness to ionic environments. The old batches showed a lack of response to ionic environments and were less reactive to the concentrations as compared to earlier dilution tests (Figure 2.13). This indicates that age may be affecting the responsiveness of the nanoparticles to the assay components.

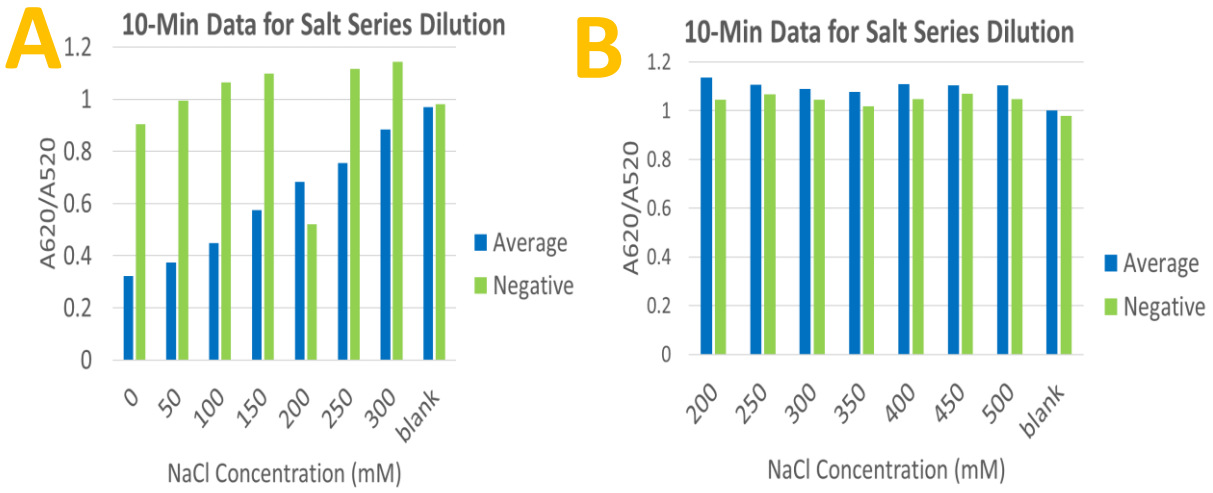

Figure 2.13: Salt series dilution of old nanoparticle batch in the early stages of assay development (A) and a salt series dilution of the same batch 2 years later with fresh reagents (B).

Fresh reagents were made for the assay components (PBS buffer, target, and non-target DNA extracts, NaCl solution), except the ssDNA probe, as it had not reached its company recommended expiration date. As done previously, a salt serious dilution test was conducted to find the optimal salt concentration for assay development. When done, the new batches of AuNPs did not show a reaction to varying salt concentrations (Figure 2.14). This may have been due to the non-uniformity in the size of the particles (Figure 2.15). The varying sizes of the particles counteract the effects of one another.

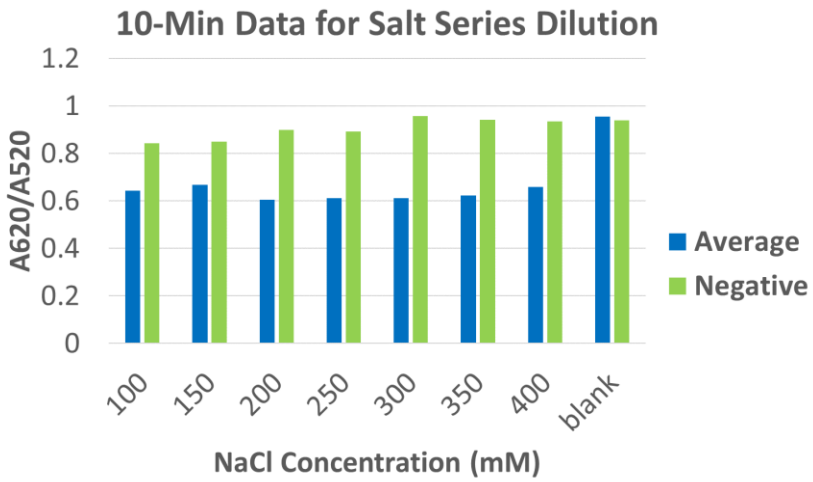

Figure 2.14: Salt series dilution of a new nanoparticle batch. No response was seen from the nanoparticles with the varying NaCl concentrations.

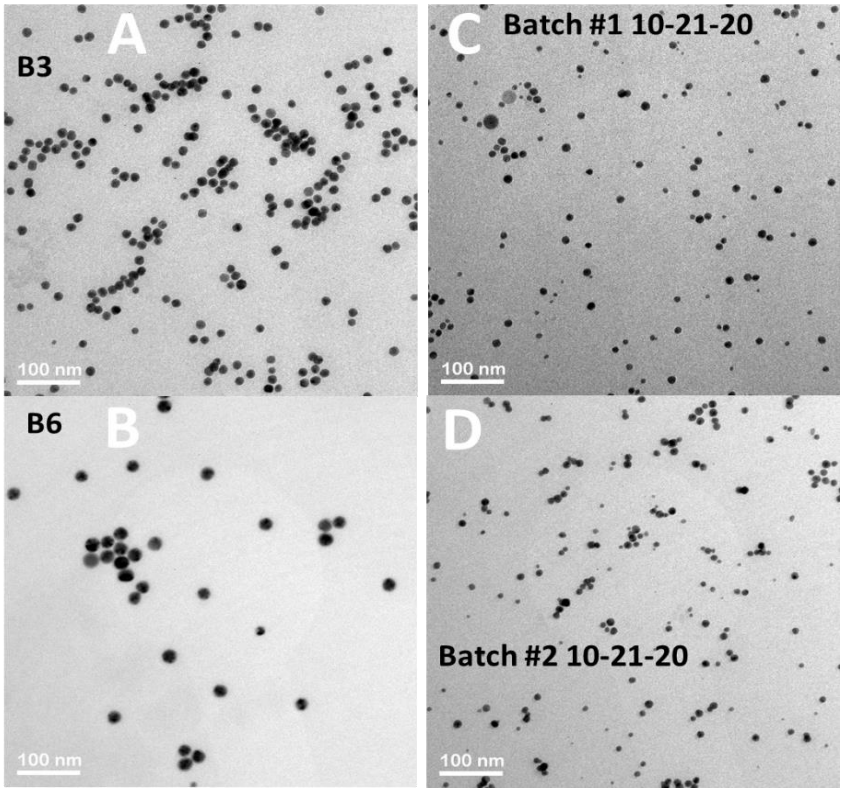

Figure 2.15: d-AuNP batches examined under a TEM. A and **B** represent old gold nanoparticle batches and **C** and **D** represent new batches of the particles. Uniform spherical shaped particles were seen for all batches. **A** and **B** have an average size of 13-15nm and 17-20nm diameters, respectively. **C** and **D** have an average size of 11-13nm diameters. However, the new batches did not have uniform shape across the entire solution.

DISCUSSION AND FUTURE IMPLICATIONS:

Future studies will need to be done to examine the consistency and reproducibility of this assay for gene sequence detection. Though there were successes in assay testing, reproducibility was never achieved. This lack of reproducibility may have been due to batch-to-batch variability and stability of the nanoparticles. This issue has been noted in the literature. Zhang et. al. had seen a depreciation in the quality of lab synthesized nanoparticles within one month of storage (Zhang et al., 2008). Tso et. al. also found difficulty with maintaining the stability of nanoparticles in aqueous conditions from commercially available nanoparticle materials (Tso et al., 2010). In the literature, it is noted that the shelf life of nanoparticle dispersions can range from a few months to 2 years and there is a need for more literature on how representative a single batch of nanoparticles is across multiple batches (Mülhopt et al., 2018). Synthesis of nanoparticles has noted challenges that may inhibit the reproducibility of studies for labs that don't have efficient synthesis or quality control technology and resources (Rahman & Rebrov, 2014). Though there have been several methods claiming highly reproducible synthesis of nanoparticles (Bayazit et al., 2016; Dong et al., 2020; Keijok et al., 2019; Panariello et al., 2020), lab resource limitations and varied environments will induce synthesis variability from batch to batch. Therefore, we would suggest nanoparticles used for future studies should be obtained from a third-party group or company that is known for consistency in nanoparticle synthesis. In addition to this, we suggest that future research be done within a short time frame to limit the degradation of nanoparticles that will occur over time.

This study was able to provide insight into the challenges of d-AuNP assay development, but also successes. This study shows the use of d-AuNPs as a diagnostic detection assay for DNA

sequences in maize. Reproducibility of the assay was limited due to batch-to-batch variation of gold nanoparticles and nanoparticle degradation. If the reproducibility of nanoparticle batches can be increased, this technology would provide a rapid detection tool for plant breeders for making breeding decisions.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, M. J., Torres-Chavolla, E., Castro, B. A., & Alocilja, E. C. (2011). One step alkaline synthesis of biocompatible gold nanoparticles using dextrin as capping agent. *Journal of Nanoparticle Research*, *13*(7), 2843–2851. https://doi.org/10.1007/s11051-010-0172-3

Arvizo, R., Bhattacharya, R., & Mukherjee, P. (2010). Gold nanoparticles: Opportunities and challenges in nanomedicine. *Expert Opinion on Drug Delivery*, *7*(6), 753–763. https://doi.org/10.1517/17425241003777010

Baetsen-Young, A. M., Vasher, M., Matta, L. L., Colgan, P., Alocilja, E. C., & Day, B. (2018). Direct colorimetric detection of unamplified pathogen DNA by dextrin-capped gold nanoparticles. *Biosensors and Bioelectronics*, *101*(August 2017), 29–36. https://doi.org/10.1016/j.bios.2017.10.011

Bai, X., Wang, Y., Song, Z., Feng, Y., Chen, Y., Zhang, D., & Feng, L. (2020). The basic properties of gold nanoparticles and their applications in tumor diagnosis and treatment. In *International Journal of Molecular Sciences* (Vol. 21, Issue 7). MDPI AG. https://doi.org/10.3390/ijms21072480

Bayazit, M. K., Yue, J., Cao, E., Gavriilidis, A., & Tang, J. (2016). Controllable Synthesis of Gold Nanoparticles in Aqueous Solution by Microwave Assisted Flow Chemistry. *ACS Sustainable Chemistry and Engineering*, *4*(12), 6435–6442. https://doi.org/10.1021/acssuschemeng.6b01149

Brown, K. A., Chemistry, B. A., & Hamad-Schifferli, K. (2000). *Noncovalent Adsorption of Nucleotides in Gold Nanoparticle DNA Conjugates: Bioavailability at the Bio-Nano Interface*. Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/44866

Brugière, N., Zhang, W., Xu, Q., Scolaro, E. J., Lu, C., Kahsay, R. Y., Kise, R., Trecker, L., Williams, R. W., Hakimi, S., Niu, X., Lafitte, R., & Habben, J. E. (2017). Overexpression of RING domain E3 ligase ZmXerico1 confers drought tolerance through regulation of ABA homeostasis. *Plant Physiology*, *175*(3), 1350–1369. https://doi.org/10.1104/pp.17.01072

Chazotte, B. (2012). Labeling Golgi with fluorescent ceramides. *Cold Spring Harbor Protocols*, *7*(8), 913–915. https://doi.org/10.1101/pdb.prot070599

Deng, H., Xu, Y., Liu, Y., Che, Z., Guo, H., Shan, S., Sun, Y., Liu, X., Huang, K., Ma, X., Wu, Y., & Liang, X. J. (2012). Gold nanoparticles with asymmetric polymerase chain reaction for colorimetric detection of DNA sequence. *Analytical Chemistry*, *84*(3), 1253–1258. https://doi.org/10.1021/ac201713t

Deng, H., Zhang, X., Kumar, A., Zou, G., Zhang, X., & Liang, X. J. (2013). Long genomic DNA amplicons adsorption onto unmodified gold nanoparticles for colorimetric detection of bacillus anthracis. *Chemical Communications*, *49*(1), 51–53. https://doi.org/10.1039/c2cc37037a

Dong, J., Carpinone, P. L., Pyrgiotakis, G., Demokritou, P., & Moudgil, B. M. (2020). Synthesis of precision gold nanoparticles using Turkevich method. *KONA Powder and Particle Journal*, *37*, 224–232. https://doi.org/10.14356/kona.2020011

Dykman, L. A., & Khlebtsov, N. G. (2011). *Gold Nanoparticles in Biology and Medicine: Recent Advances and Prospects*. *3*(9), 2011. https://doi.org/10.1039/b711490g

Franco, R., Pedrosa, P., Carlos, F. F., Veigas, B., & Baptista, P. V. (2015). Gold nanoparticles for DNA/RNA-based diagnostics. In *Handbook of Nanoparticles* (pp. 1339–1370). Springer International Publishing. https://doi.org/10.1007/978-3-319-15338-4_31

Giraldo, J. P., Wu, H., Newkirk, G. M., & Kruss, S. (2019). Nanobiotechnology approaches for engineering smart plant sensors. *Nature Nanotechnology*, *14*(6), 541–553. https://doi.org/10.1038/s41565-019-0470-6

Han, H., Yi, W., Hou, D., Huang, T., & Hao, Z. (2015). AuNPs-based colorimetric assay for identification of chicken tissues in meat and meat products. *Journal of Nanomaterials*, *2015*. https://doi.org/10.1155/2015/469267

Han, Kyung-Hwan & Ko, J.-H. (2011). *DNA ENCODING RING ZINC-FINGER PROTEIN AND THE USE OF THE DNA IN VECTORS AND BACTERIA AND IN PLANTS* (Patent No. US7977535 B2). United States Patent and Trademark Office.

Hussain, M. M., Samir, T. M., & Azzazy, H. M. E. (2013). Unmodified gold nanoparticles for direct and rapid detection of *Mycobacterium tuberculosis* complex. *Clinical Biochemistry*, *46*(7–8), 633–637. https://doi.org/10.1016/j.clinbiochem.2012.12.020

Izanloo, C. (2017). Effect of gold nanoparticle on stability of the DNA molecule: A study of molecular dynamics simulation. *Nucleosides, Nucleotides and Nucleic Acids*, *36*(9), 571–582. https://doi.org/10.1080/15257770.2017.1353697

Keijok, W. J., Pereira, R. H. A., Alvarez, L. A. C., Prado, A. R., da Silva, A. R., Ribeiro, J., de Oliveira, J. P., & Guimarães, M. C. C. (2019). Controlled biosynthesis of gold nanoparticles with Coffea arabica using factorial design. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-52496-9

Khaliliazar, S., Ouyang, L., Piper, A., Chondrogiannis, G., Hanze, M., Herland, A., & Hamedi, M. M. (2020). Electrochemical Detection of Genomic DNA Utilizing Recombinase Polymerase Amplification and Stem-Loop Probe. *ACS Omega*, *5*(21), 12103–12109. https://doi.org/10.1021/acsomega.0c00341

Ko, J. H., Yang, S. H., & Han, K. H. (2006). Upregulation of an arabidopsis RING-H2 gene, XERICO, confers drought tolerance through increased abscisic acid biosynthesis. *Plant Journal*, *47*(3), 343–355. https://doi.org/10.1111/j.1365-313X.2006.02782.x

Li, H., & Rothberg, L. (2004). Colorimetric detection of DNA sequences based on electrostatic interactions with unmodified gold nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(39), 14036–14039. https://doi.org/10.1073/pnas.0406115101

Liu, M., Yuan, M., Lou, X., Mao, H., Zheng, D., Zou, R., Zou, N., Tang, X., & Zhao, J. (2011). Label-free optical detection of single-base mismatches by the combination of nuclease and gold nanoparticles. *Biosensors and Bioelectronics*, *26*(11), 4294–4300. https://doi.org/10.1016/j.bios.2011.04.014

Mülhopt, S., Diabaté, S., Dilger, M., Adelhelm, C., Anderlohr, C., Bergfeldt, T., de la Torre, J. G., Jiang, Y., Valsami-Jones, E., Langevin, D., Lynch, I., Mahon, E., Nelissen, I., Piella, J., Puntes, V., Ray, S., Schneider, R., Wilkins, T., Weiss, C., & Paur, H. R. (2018). Characterization of nanoparticle batch-to-batch variability. *Nanomaterials*, *8*(5). https://doi.org/10.3390/nano8050311

Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J. A. McQuisten, K. A. Almabrazi, H. G., & Peek, A. S. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, *36 (Web Se*, W136-9. https://doi.org/10.1093/nar/gkn198

Panariello, L., Damilos, S., Du Toit, H., Wu, G., Radhakrishnan, A. N. P., Parkin, I. P., & Gavriilidis, A. (2020). Highly reproducible, high-yield flow synthesis of gold nanoparticles based on a rational reactor design exploiting the reduction of passivated Au(iii). *Reaction Chemistry and Engineering*, *5*(4), 663–676. https://doi.org/10.1039/c9re00469f

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., Braun, B. L., Dietze, M., Dunfee, B., Elsik, C. G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., … Andorf, C. M. (2019). Maizegdb 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Research*, *47*(D1), D1146–D1154. https://doi.org/10.1093/nar/gky1046

Rahman, M. T., & Rebrov, E. V. (2014). Microreactors for gold nanoparticles synthesis: From faraday to flow. In *Processes* (Vol. 2, Issue 2, pp. 466–493). MDPI AG. https://doi.org/10.3390/pr2020466

Team, R. C. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.r-project.org/.

Tso, C. P., Zhung, C. M., Shih, Y. H., Tseng, Y. M., Wu, S. C., & Doong, R. A. (2010). Stability of metal oxide nanoparticles in aqueous solutions. *Water Science and Technology*, *61*(1), 127–133. https://doi.org/10.2166/wst.2010.787

Vetrone, S. A., Huarng, M. C., & Alocilja, E. C. (2012). Detection of Non-PCR amplified S. enteritidis genomic DNA from food matrices using a gold-nanoparticle DNA biosensor: A proof-of-concept study. *Sensors (Switzerland)*, *12*(8), 10487–10499. https://doi.org/10.3390/s120810487

Vorobjev, P., Epanchintseva, A., Lomzov, A., Tupikin, A., Kabilov, M., Pyshnaya, I., & Pyshnyi, D. (2019). DNA Binding to Gold Nanoparticles through the Prism of Molecular Selection: Sequence-Affinity Relation. *Langmuir*, *35*(24), 7916–7928. https://doi.org/10.1021/acs.langmuir.9b00661

Yeh, Y. C., Creran, B., & Rotello, V. M. (2012). Gold nanoparticles: Preparation, properties, and applications in bionanotechnology. In *Nanoscale* (Vol. 4, Issue 6, pp. 1871–1880). Royal Society of Chemistry. https://doi.org/10.1039/c1nr11188d

Zhang, D., Huarng, M. C., & Alocilja, E. C. (2010). A multiplex nanoparticle-based bio-barcoded DNA sensor for the simultaneous detection of multiple pathogens. *Biosensors and Bioelectronics*, *26*(4), 1736–1742. https://doi.org/10.1016/j.bios.2010.08.012

Zhou, Y., Tang, L., Zeng, G., Zhang, C., Xie, X., Liu, Y., Wang, J., Tang, J., Zhang, Y., & Deng, Y. (2016). Label free detection of lead using impedimetric sensor based on ordered mesoporous carbon-gold nanoparticles and DNAzyme catalytic beacons. *Talanta*, *146*, 641–647. https://doi.org/10.1016/j.talanta.2015.06.063

Zhu, X., & Gao, T. (2018). Spectrometry. In *Nano-inspired Biosensors for Protein Assay with Clinical Applications* (pp. 237–264). Elsevier. https://doi.org/10.1016/B978-0-12-815053-5.00010-6

CHAPTER 3: UTILIZING MACHINE LEARNING ALGORITHMS FOR IDENTIFICATION AND CLASSIFICATION OF FUSARIUM INFECTED WHEAT SEED VIA IMAGE-BASED ANALYSIS

ABSTRACT:

Fusarium Head Blight (FHB) is a devastating plant disease that is caused by the *Fusarium* spp. with its dominant pathogen being *Fusarium graminearum*. FHB, or scab infection, has led to several billion dollars in losses due to its degenerative effects on the nutritive, physical, and chemical qualities of infected grains. Infection in wheat (*Triticum* spp.) is often visualized as bleaching of the spike where kernels are a ghostly pink color and shriveled in appearance. This disease also produces a harmful vomitoxin called deoxynivalenol (DON) that causes nausea, fever, headaches, vomiting, and disruption of normal cell function in humans and animals. The damaging effects of this infection, cause a need for diagnostic tools to prevent DON contamination. This study aimed to develop an image-based identification model for the detection of and differentiation between healthy and diseased wheat seeds. We compared the accuracy of FHB detection in multiple machine learning models including Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). All methods were extremely accurate: 95 to 98.8% accuracy in the withheld testing set. Utilizing image-based methods for disease identification can help researchers to improve the efficiency of detecting Fusarium diseased kernels (FDK), which is typically done by hand. This would also provide a more objective and accurate method for evaluating disease severity.

INTRODUCTION:

Fusarium Head Blight (FHB) is one of the most devastating plant diseases in the world. The scab disease has caused billions of dollars in losses due to its degenerative effect on the nutritive, physical, and chemical qualities in the grain (Cowger et al., 2020; McMullen et al., 1997), which lowers the market value of the grain. FHB, or scab infection, is caused by the *Fusarium* spp. with its dominant pathogen being *Fusarium graminearum*. Scab infection in wheat (*Triticum* spp.) is shown by the bleaching of the spike head, beginning in one of its spikelets, and spreading to the rest of the spike. After harvest, infection in wheat is often visualized in the kernels as a tombstone, pink or chalky color and shriveled in appearance (Figure 3.1).

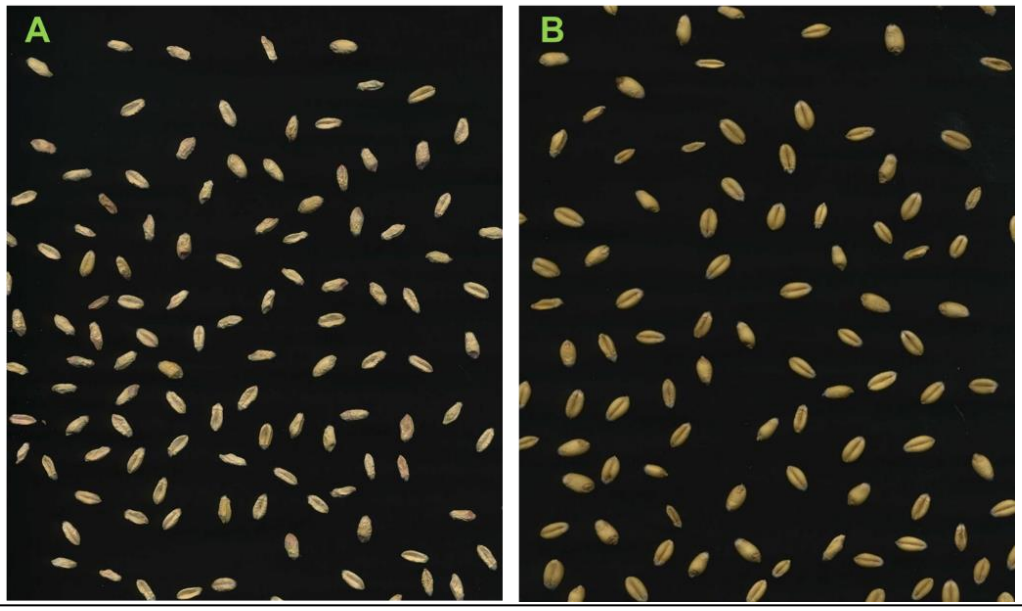

Figure 3.1: Scanned images of diseased and healthy wheat seeds. Images of wheat seeds where seeds in (A) are infected with *Fusarium graminearum* and those in (B) are

This well-documented disease is most impactful due to the pathogen's creation of a mycotoxin called deoxynivalenol (DON) upon infection of wheat. DON in grain can be very harmful to animals and humans as it disrupts normal cellular function and can lead to nausea, fever, headaches, and vomiting (Chu, 2003). DON contaminated grain can cause extreme discounts as

the USDA recommends DON levels not to exceed 1 part per million (ppm) and 2ppm is marked as unacceptable for wheat used in human foods (Food and Drug Administration (FDA), 2010; Xia et al., 2020). DON's impact of lowering the value of grain production leads to the need for rapid detection tools.

Visual assessment of wheat seed can be one of the best ways to evaluate samples in a non-destructive way. In many pathology and breeding research labs, Fusarium diseased kernels (FDK) are often identified by hand using standards set by the USDA Grain Inspection, Stockyards, and Packers Administration (USDA, 2016; Dowell et al., 1999). Though this has been a reliable diagnostic method for researchers, visually detecting FDK by hand can be time-consuming, subjective, and not well suited for large samples. As machine learning technologies in image recognition have advanced, image-based detection for diagnostic assessment has grown in popularity. One approach to more rapid and objective image-based detection of post-harvest FDK measures whitened kernel surface in seed photographs with a correlation around 0.8 with FDK (Saccon et al., 2017). In recent experiments, FDK has been assessed using hyperspectral imaging, Fourier Transform Infrared (FTIR) spectroscopy, and Near Infrared (NIR) spectroscopy (Alisaac et al., 2019; Barbedo et al., 2015; Kautzman et al., 2015; Lahlali et al., 2015). Despite these innovative applications' ability to detect FDK, these methods often require expensive analysis equipment. This study looks to utilize low-cost, easy-to-use equipment for rapid detection of FDK.

This study aimed to develop an image-based identification model for the detection of and differentiation between healthy and *Fusarium* spp. infected wheat seeds. We compared the

accuracy of FHB detection in multiple machine learning models including Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Utilizing image-based methods for disease identification would help researchers to improve the efficiency of detecting FDK, which is typically done by hand. This would also provide a more objective and accurate method for evaluating disease severity while using inexpensive equipment for diagnostic analysis.

MATERIALS AND METHODS:

MATERIALS

This experiment used the infected grains of several Michigan wheat varieties. The grain was comprised of several red and white soft wheat varieties. The samples used for image analysis were obtained from infected plants in fields with natural or grain spawn inoculum. These infected plants were grown in plant pathology research fields at Michigan State University and the Saginaw Valley Research and Extension Center in Frankenmuth, MI. . Presence of infection was confirmed for each field by isolating *Fusarium* from small samples of symptomatic kernels and wheat heads. In addition, seed samples were evaluated by trained pathologists for the incidence of FDK. A "healthy" kernel is one that visually has no *Fusarium* spp. present while diseased kernels were visually confirmed using USDA-GISPA standards (USDA, 2016).

IMAGES COLLECTION

Images of seed were collected using an Epson Perfection 4180 Photo flatbed scanner. Healthy and diseased kernels were placed on the scanner and spread apart so that no seeds were touching (Figure 3.2). Seeds were left in the position that they fell on the scanner (i.e., a mix of the ventral, dorsal, and side profiles of the kernels were scanned). Images were produced using a 24-bit color setting and a 720dpi resolution. A total of 150 images were taken, resulting in nearly 38,000 kernels scanned. Seeds were scanned in 3 different sets. Images within the scan 1 and 2 sets had a mix of healthy and diseased kernels. Images within the scan 3 set had solely healthy or diseased kernels. Images in scan 3 were used to develop and train machine learning algorithms. Scan 3 images contained a total of 11,351 kernels, where 3,760 kernels were FDK and 7,591 were healthy kernels. For each aliquot of seed used for an image, notes were recorded on the

sample's weight, the total number of diseased seeds in the sample, the image number, and the seed variety. Images were taken with a ruler and color panel for calibration and scale setting during image processing.

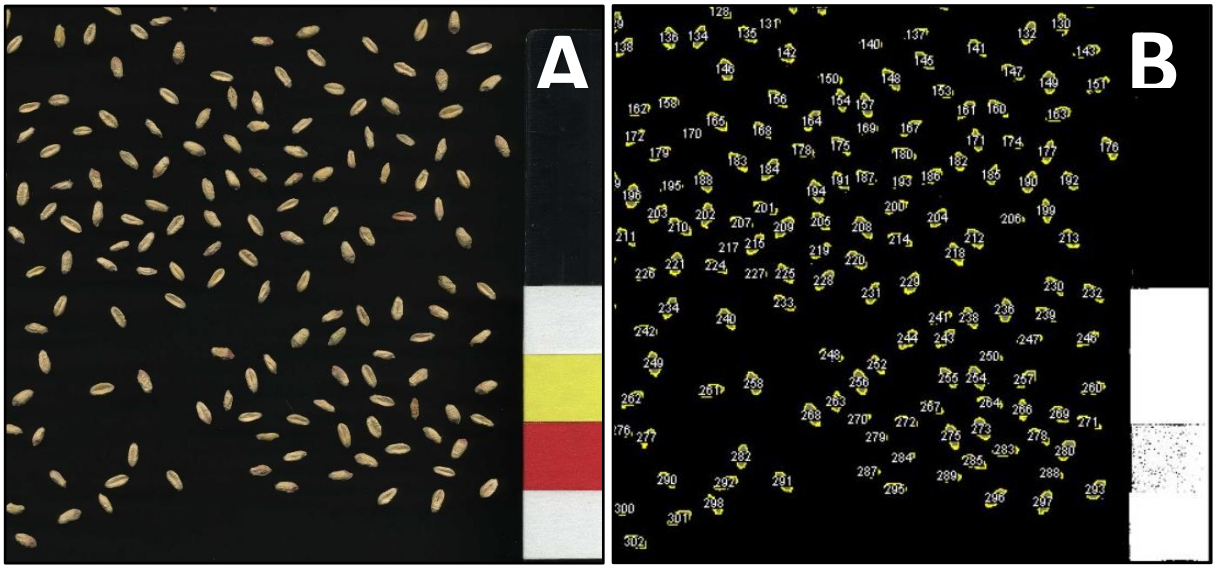

Figure 3.2: Images of FDK collected from the flatbed scanner (A) and the labelled image after ROI detection and measurement via ImageJ software (B).

IMAGE PROCESSING

Scanned images were processed using the ImageJ 1.x software (Schneider et al., 2012) to collect several size, shape, and color parameters. An ImageJ macro program was made to batch-process all images. The software took measurements of the area, mean color, perimeter, circularity, aspect ratio, roundness, solidity, and minimum feret (Table 3.1). There were a total of 10 shape, size, and color parameters. The three color measurements were "Mean Red", "Mean Blue", and "Mean Green", which is the amount of red, blue, or green intensity within the region of interest (ROI) for each seed. This color measurement is based on the additive RGB color model.

Trait	Description
Area	Area of selection in square pixels.
Mean "Color"	The amount of color intensity for the given selection.
Perim. (perimeter)	The length of the outside boundary of the selection.
Circ. (circularity)	4π*area/perimeter^2. A value of 1.0 indicates a perfect circle. As the value approaches 0.0, it indicates an increasingly elongated shape.
AR (aspect ratio)	The ratio of the width to the height of an image or screen.
Round (roundness)	4*area/(π*major_axis^2), or the inverse of the aspect ratio.
Solidity	The area of your features divided by their convex area (the imaginary convex hull around it). How filled in the image selection is.
MinFeret	The minimum seed diameter.

Table 3.1: This table shows a list and description of the various size, shape, and color measurement collected by the ImageJ software.

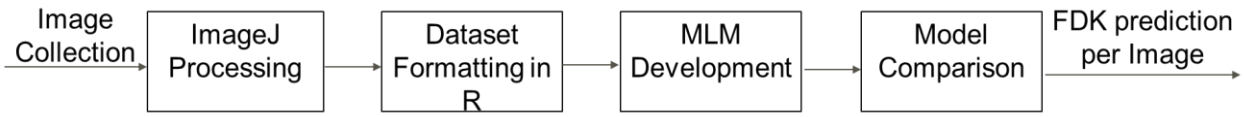

Figure 3.3: Workflow diagram of image processing for determining FDK per image. Logistic regression, support vector machines, and k-nearest neighbors were the machine learning models (MLM) used for analysis and prediction.

MODEL DEVELOPMENT

Data collected by the ImageJ software was formatted and analyzed using R programming software (R Core Team, 2020). Within R, machine learning algorithms were developed and tuned utilizing the caret package (Kuhn, 2020) for model development. Data collected from the scan 3 set of images were used for model development. Training and testing data were divided via a 70-30 split, with 70% of the data (representing 7,946 kernels) used as the training set and the other 30% (representing 3,405 kernels) as the testing dataset. Three different classification methods were used to identify and classify the wheat seed as health or FDK including: support vector machines (SVM), logistic regression (LR), and k-nearest neighbors (KNN). The workflow for image processing and analysis can be found in Figure 3.3. Once data was properly formatted, the scan 3 training data was used to build and tune the models. The optimized models were then

used to predict the classification of each kernel as FDK or healthy. Comparison of the models was based on accuracy, the area under the ROC curve (AUC), and the predictive processing times when used on the testing dataset. These model comparison criteria were used in other experiments to differentiate machine learning models for image classification (Al Zorgani & Ugail, 2018; Saberioon et al., 2018). In each of these studies, support vector machines, logistic regression, and k-nearest neighbors were among the models compared. Each model was optimized using a ten-fold cross-validation, repeated three times. The best performing model was applied to scan 1 and scan 2 datasets to obtain the FDK incidence prediction per image.

RESULTS AND DISCUSSION:

TUNING MODELS FOR OPTIMIZATION

To optimize the model for support vector machines, the type of support vector machine model and cost value (C-value) for SVM must be examined. C-value is a parameter within SVM that allows for misclassifications within the model, preventing overfitting. The higher the cost value, the more misclassifications that are allowed within the model. In our analysis, accuracy across various cost values was examined to find an optimal C-value to use for the model (Figure 3.4). The optimal C-value for SVM was 2.01. SVM models utilize a kernel function that finds the support vector classifier in higher dimensions to separate the data via a hyperplane. The three types of kernel functions for SVM are linear, polynomial, and radial functions. Each of these instances differentiate in how they make decisions for hyperplane boundaries of classes. When comparing these three kernel functions, it was found that the SVM linear kernel function had the highest accuracy of 98.7%. With these tuned parameters, SVM was able to achieve an accuracy of 98.7%, with a 97.9% sensitivity and 99.1% specificity (Figure 3.6).

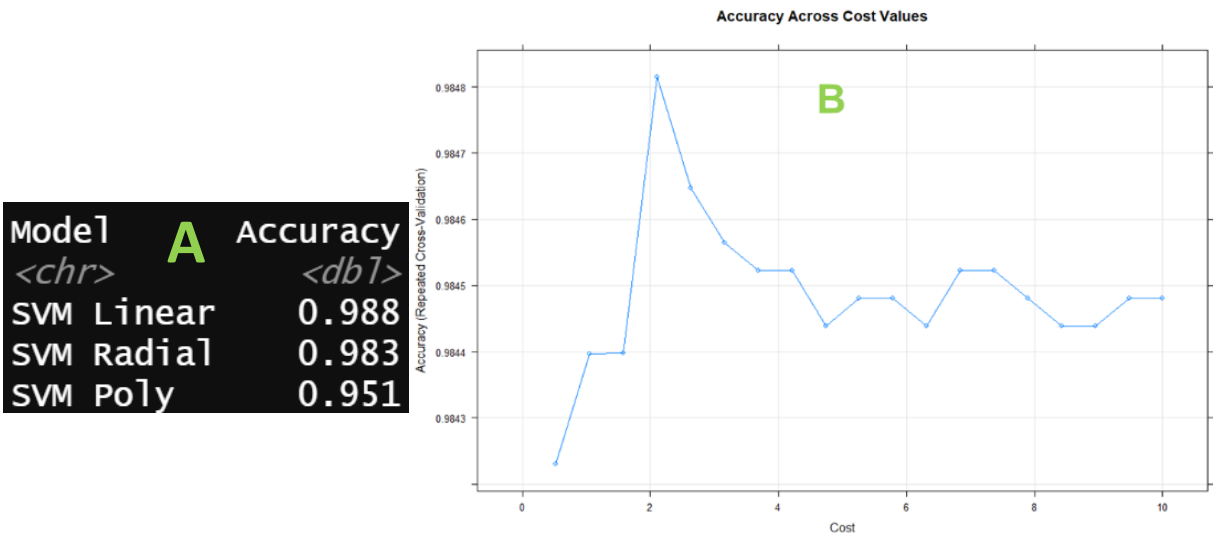

Figure 3.4: Tuning for Support Vector Machine (SVM) model. The SVM Linear model showed the highest accuracy (A). The SVM model was tuned by analyzing the accuracy of varying SVM model across cost values (B). The accuracy of the SVM Linear model was then measured across cost values on the training data set. The optimal cost value found and used in later analysis was 2.01.

Though there are no tuning parameters for logistic regression, we were able to look at the importance of each variable in the model. Within the caret package, the "varImp" function can be used to look at the importance of each parameter. This importance is based on the absolute value of the t-statistic for each parameter in the model (Dalpiaz, 2020). Based on this analysis, we see that "Mean Blue" is the most important parameter in the model followed by Mean Green and circularity **(**Figure 3.7). This is consistent with the standard used to rate FDK by hand set by the USDA, as color and a shriveled shape are the most obvious indicators of fusarium infection (Bauriegel et al., 2010; USDA, 2016; West et al., 2017). Upon further examination, it was seen that the distribution of the Mean Blue parameter had a bimodal distribution (Figure 3.5). This may be the cause of its high importance in the model. The logistic regression model performed strongly with its accuracy (98.6%), sensitivity (97.6%), and specificity (99%) (Figure 3.6).

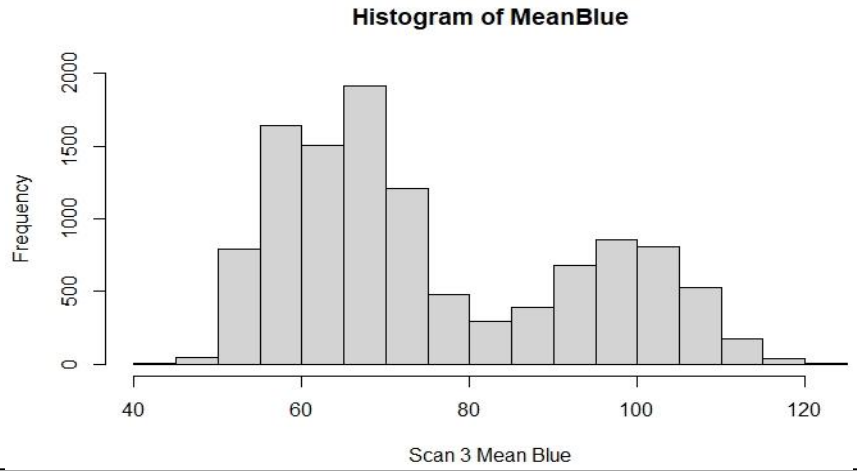

Figure 3.5: Bimodal distribution for the "Mean Blue" parameter in the model. This plot shows the distribution of the dataset used for training the model. Mean Blue was the most important variable for the classification model.

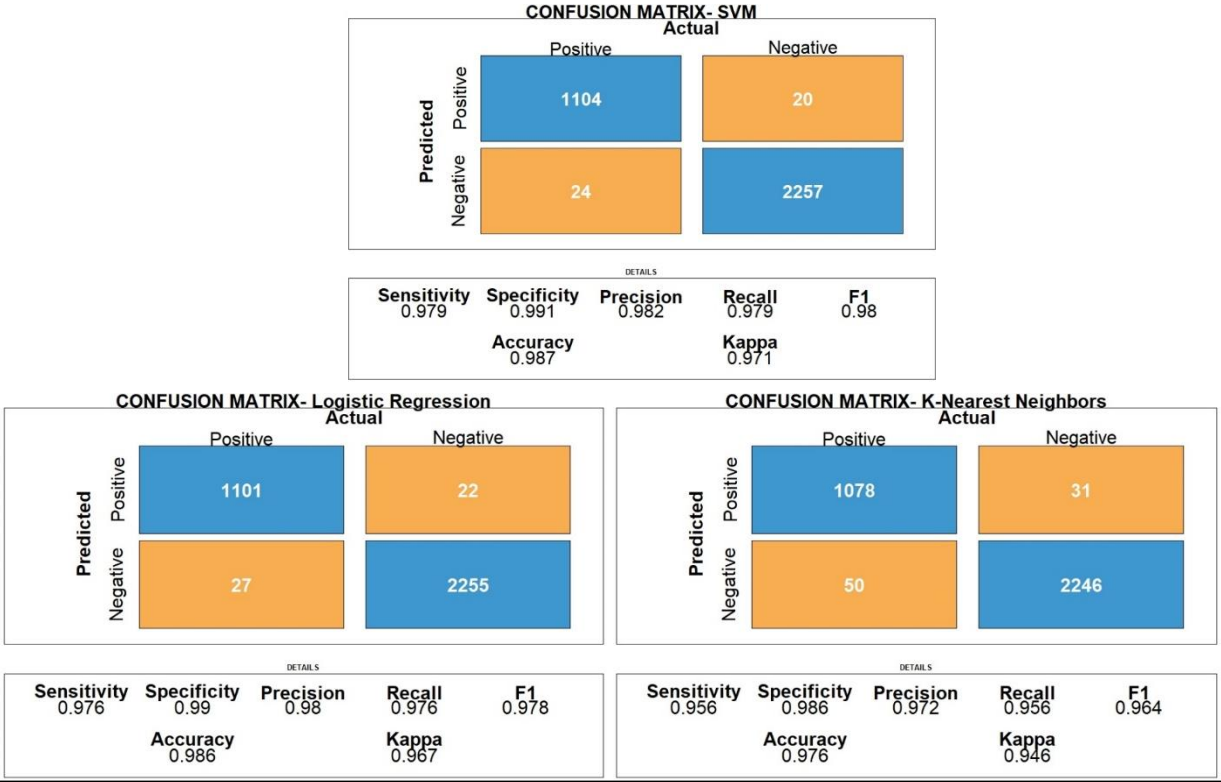

Figure 3.6: Confusion matrices for tuned models. Confusion matrixes were used to compare the accuracy of the machine learning models when classifying wheat seeds as healthy or diseased using the testing dataset. Support Vector Machine model had the highest true positive (sensitivity and recall), true negative rate (specificity), positive predictive value (precision), accuracy, and F1 values. The Logistic Regression model showed only slightly lower performance in each of these categories (0.4% > difference in each category). The K-Nearest Neighbor model showed the worst performance in all categories.

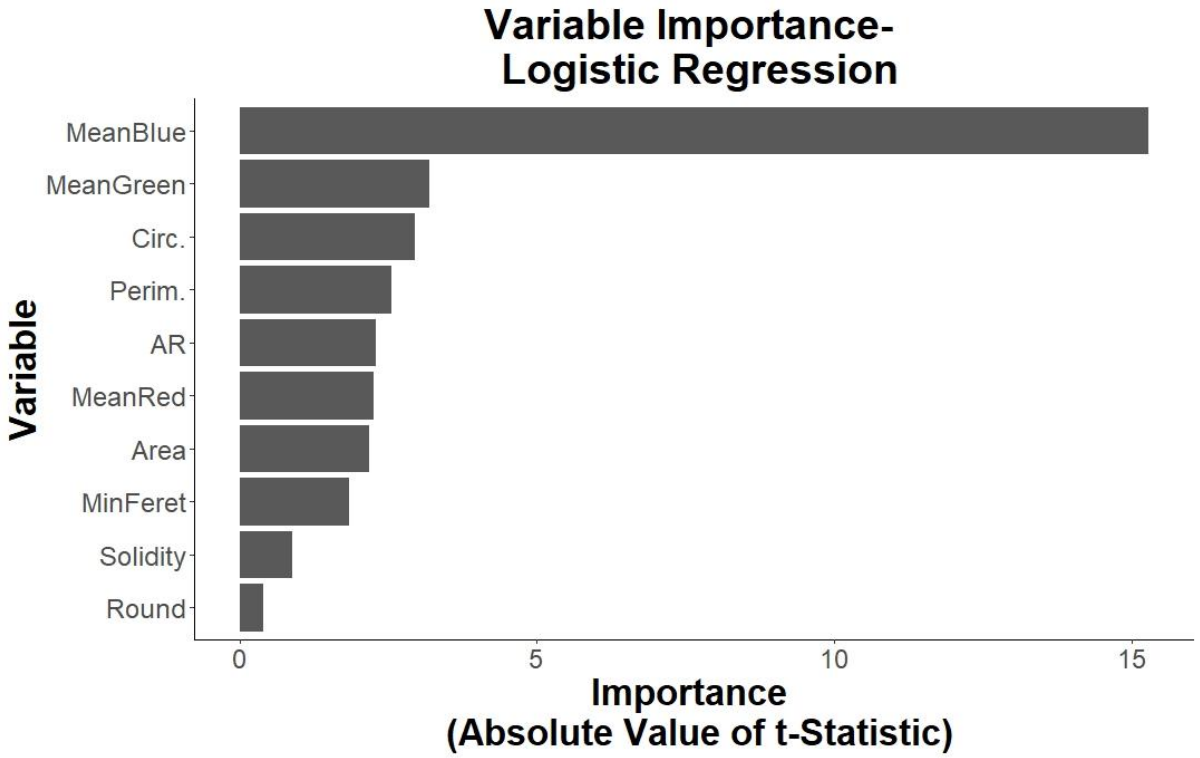

Figure 3.7: The importance of each parameter utilized within the Logistic Regression model. The parameter "Mean Blue", which is the amount of blue color intensity in each seed, was found to be the most important variable within the model.

To optimize the model for k-nearest neighbors, various k-values should be examined. K-value in this model is a parameter that refers to the number of "nearest neighbors" used by the model when making classification predictions. The higher the K-value, the more neighboring data points that are used for voting whether a kernel is diseased or healthy. We examined the accuracy of the model across k-values and found the optimal k to be 5 **(**Figure 3.8**)**. With these tuned parameters, KNN achieved an accuracy of 97.6%, with a 95.6% sensitivity and 98.6% specificity (Figure 3.6).

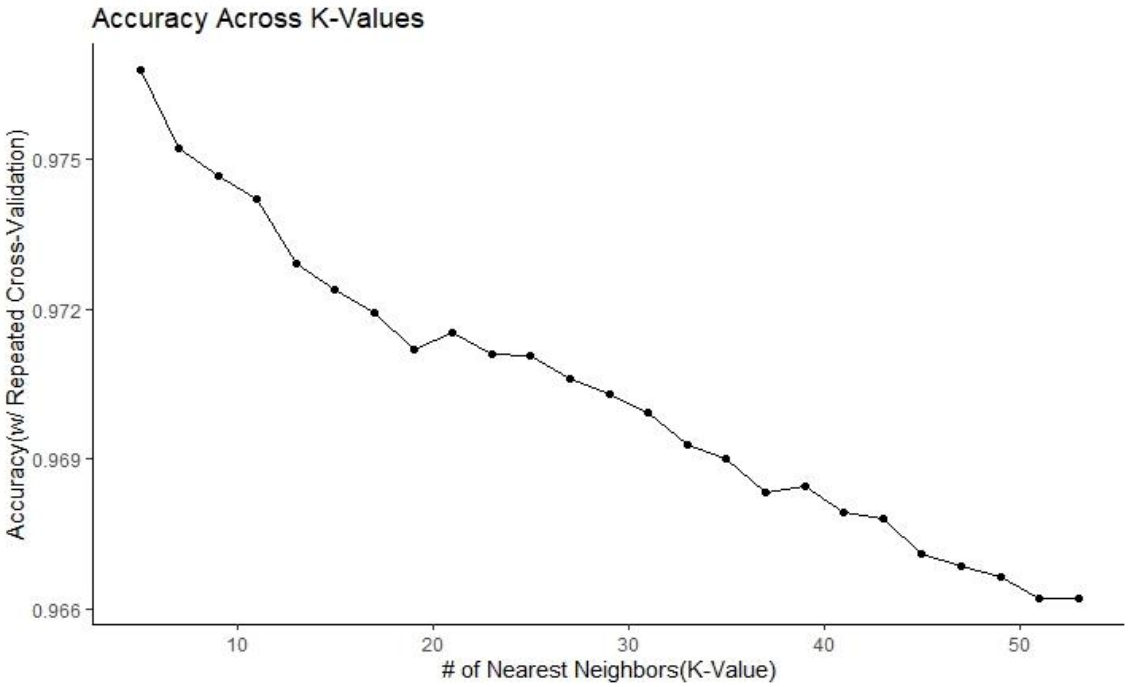

Figure 3.8: Identifying the optimal K value for the K-Nearest Neighbor model. The optimal K value for the K-Nearest Neighbor model was identified by comparing model accuracy across a range of K values on the training dataset. The optimal K value used for further analysis was 5, where K is the number of nearest neighbors used to appropriately classify a data point.

MODEL COMPARISON AND SELECTION

In previous comparative studies, support vector machines showed the highest accuracy amongst classification models (Al Zorgani & Ugail, 2018; Saberioon et al., 2018). This is consistent with our study as SVM showed the highest accuracy (98.8%) of the three algorithms (Table 3.2). The LR model showed only a slightly lower performance (less than 0.4% difference in each category) when compared to SVM. The KNN model performed the worst of the classification algorithms. Though SVM showed the highest accuracy, when comparing the algorithm's processing speed when making predictions on testing sets, LR was seven times faster than SVM (Table 3.2). Due to the minimal gain in accuracy, specificity, sensitivity, and area under the ROC curve, logistic regression was deemed to be the best choice for an accurate yet rapid detection algorithm.

Once logistic regression was chosen as the best model to use, we applied this classification model to the other sets of scanned seed (1 and 2) to examine the applicability of the model on mixed seed images. As mentioned previously, these scans had a mix of both healthy and diseased seeds within each scan. When applied, we compared the predicted number of FDK per image to the actual number of FDK per image. We saw a correlation of 81.8% for the model and it had a significant p-value (<0.001) and an adjusted R-squared of 66.7% (Figure 3.9).

Model	Accuracy	AUC	Threshold	ProcessTime
SVM Linear	0.9883376	0.9988277	0.7595711	0.22
Logistic Regression	0.9856094	0.9987304	0.7871333	0.03
K-Nearest Neighbors	0.9753337	0.9903753	0.4500000	0.28

Table 3.2: The accuracy, area under the curve ROC (AUC), and predictive processing times when analyzing the testing dataset were used to compare the optimized machine learning models. The Support Vector Machine model had the highest accuracy and AUC. The Logistic Regression model was only slightly lower in accuracy and AUC but had the fastest processing time. The K-Nearest Neighbor model performed the worst amongst the compared models.

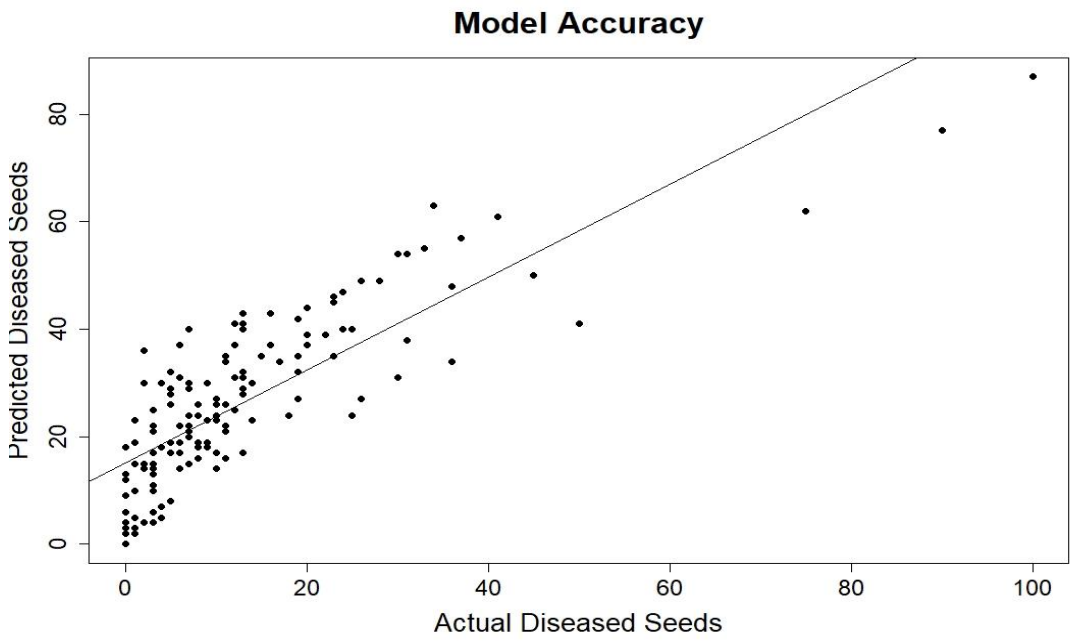

Figure 3.9: Correlation between the predicted number of diseased seed per image and the actual when logistic regression is applied to additional images. Showing the correlation between the predicted number of diseased seed per image and the actual, when the LR model is applied to scan 1 and 2 sets of seed images. The P value for this plot was significant <0.001 and had a correlation of 81.8% and an adjusted R-Squared of 66.74%.

Very few studies have looked at the classification of FDK utilizing relatively inexpensive equipment. Building upon previous work, this study uses a larger dataset for model development (over 11,000 kernels used) and alternative machine learning models (LR, SVM, and KNN) for FDK classification. In a study published in 2018, a comparative analysis examined hyperspectral images versus the use of flatbed scanner images for the classification of FDK (Ropelewska & Zapotoczny, 2018). Ropelewska and Zapotoczny were able to achieve high classification accuracy using a flatbed scanner (94-100%) by examining 120 kernels that were laid on either their dorsal or ventral sides. The accuracy of their model was influenced by the positioning of the analyzed wheat kernels and wheat variety. The following year, Ropelewska did another study using flatbed images and was able to achieve a classification accuracy range of 58.12%-73.37%

(Ropelewska, 2019). This time they used 1,800 kernels and 59 geometric parameters for classification. The highest accuracy was found using a 10-fold cross-validation procedure and various attribute selection methods to lower the processing time for model application. In addition to color, shape, and size parameters, researchers have also utilized textural parameters to classify FDK (Guevara-Hernandez & Gomez Gil, 2011; Zapotoczny, 2011). In future work, textural parameters may also be useful for increasing the classification accuracy of the model.

DISCUSSION AND FUTURE IMPLICATIONS:

This study provides a rapid and low-cost method for discrimination between healthy kernels and FDK based on color, size, and shape parameters. FDK was classified with the greatest accuracy of 98.7% utilizing the SVM model, but this model lacked the speed of processing time when making predictions on the testing set. Logistic regression was the fastest of compared models by at least sevenfold and maintained a high accuracy, AUC, sensitivity, and specificity. Mean Blue was the most important parameter within the LR classification model. Upon model application to other data sets, LR achieved an 81.8% correlation between predicted FDK and actual FDK. Utilizing image-based methods for disease identification would help researchers to improve the efficiency of detecting FDK without the use of expensive equipment in a rapid, non-destructive, and objective manner. In future work, models should also include textural parameters to increase model accuracy. This method could be integrated into an application for smartphone use. Currently, many smartphones have the capability of exporting photos at a resolution of more than 300dpi. Further experimentation should be done to test the efficiency of the model using lower-resolution images. In addition to this, separating the kernels to ensure none of them were touching added significantly to the time used for data collection. Thus, further research should also be done to examine the accuracy of these models using samples with various spacing levels, ranging from no contact to touching on all sides of the seed. This would increase the usability and practicality of this method for plant pathologists and breeders. As a limitation of this study, individual kernel infection was not examined via qPCR or another diagnostic testing. This may limit the ability of the model to be used on visually asymptomatic *Fusarium* spp. infected kernel. This model was developed to align with methods for visually symptomatic infection.

BIBLIOGRAPHY

BIBLIOGRAPHY

Al Zorgani, M., & Ugail, H. (2018). *Comparative Study of Image Classification using Machine Learning Algorithms*. EasyChair. https://doi.org/10.29007/4VBP

Alisaac, E., Behmann, J., Rathgeb, A., Karlovsky, P., Dehne, H.-W., & Mahlein, A.-K. (2019). Assessment of Fusarium Infection and Mycotoxin Contamination of Wheat Kernels and Flour Using Hyperspectral Imaging. *Toxins*, *11*(10), 556. https://doi.org/10.3390/toxins11100556

Barbedo, J. G. A., Tibola, C. S., & Fernandes, J. M. C. (2015). Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. *Biosystems Engineering*, *131*, 65–76. https://doi.org/10.1016/j.biosystemseng.2015.01.003

Bauriegel, E., Giebel, A., & Herppich, W. B. (2010). Rapid Fusarium head blight detection on winter wheat ears using chlorophyll fluorescence imaging. *Journal of Applied Botany and Food Quality*, *83*(2), 196–203.

Chu, F. S. (2003). MYCOTOXINS | Toxicology. In *Encyclopedia of Food Sciences and Nutrition* (pp. 4096–4108). Elsevier. https://doi.org/10.1016/b0-12-227055-x/00823-3

Cowger, C., Smith, J., Boos, D., Bradley, C. A., Ransom, J., & Bergstrom, G. C. (2020). Managing a Destructive, Episodic Crop Disease: A National Survey of Wheat and Barley Growers' Experience with Fusarium Head Blight. *Plant Disease*, *104*(3), 634–648. https://doi.org/10.1094/PDIS-10-18-1803-SR

Dalpiaz, D. (2020). *Chapter 21 The caret Package | R for Statistical Learning*. https://daviddalpiaz.github.io/r4sl/the-caret-package.html

Dowell, F. E., Ram, M. S., & Seitz, L. M. (1999). Predicting Scab, Vomitoxin, and Ergosterol in Single Wheat Kernels Using Near-Infrared Spectroscopy. *Cereal Chemistry Journal*, *76*(4), 573–576. https://doi.org/10.1094/CCHEM.1999.76.4.573

Food and Drug Administration (FDA). (2010, July). *Guidance for Industry and FDA: Advisory Levels for Deoxynivalenol (DON) in Finished Wheat Products for Human Consumption and Grains and Grain By-Products used for Animal Feed | FDA*. Center for Food Safety and Applied Nutrition. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-and-fda-advisory-levels-deoxynivalenol-don-finished-wheat-products-human

Guevara-Hernandez, F., & Gomez Gil, J. (2011). A machine vision system for classification of wheat and barley grain kernels - Dialnet. *Spanish Journal of Agricultural Research*, 672–680. https://dialnet.unirioja.es/servlet/articulo?codigo=3738296

Kautzman, M. E., Wickstrom, M. L., & Scott, T. A. (2015). The use of near-infrared transmittance kernel sorting technology to salvage high-quality grain from grain downgraded due to Fusarium damage. *Animal Nutrition*, *1*(1), 41–46. https://doi.org/10.1016/j.aninu.2015.02.007

Kuhn, M. (2020). *caret: Classification and Regression Training. R package version 6.0-86.* https://cran.r-project.org/package=caret

Lahlali, R., Karunakaran, C., Wang, L., Willick, I., Schmidt, M., Liu, X., Borondics, F., Forseille, L., Fobert, P. R., Tanino, K., Peng, G., & Hallin, E. (2015). Synchrotron-based phase contrast X-ray imaging combined with FTIR spectroscopy reveals structural and biomolecular differences in spikelets play a significant role in resistance to Fusarium in wheat. *BMC Plant Biology*, *15*(1), 24. https://doi.org/10.1186/s12870-014-0357-5

McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., & Sanford, D. (2012). A Unified Effort to fight an enemy of Wheat and barley: Fusarium Head Blight. *The American Phytopathological Society*, *96*(12). https://apsjournals.apsnet.org/doi/pdf/10.1094/PDIS-03-12-0291-FE

Ropelewska, E. (2019). Evaluation of wheat kernels infected by fungi of the genus *Fusarium* based on morphological features. *Journal of Food Safety*, *39*(3), e12623. https://doi.org/10.1111/jfs.12623

Ropelewska, E., & Zapotoczny, P. (2018). Classification of *Fusarium*-infected and healthy wheat kernels based on features from hyperspectral images and flatbed scanner images: a comparative analysis. *European Food Research and Technology*, *244*(8), 1453–1462. https://doi.org/10.1007/s00217-018-3059-7

Saberioon, M., Císař, P., Labbé, L., Souček, P., Pelissier, P., & Kerneis, T. (2018). Comparative Performance Analysis of Support Vector Machine, Random Forest, Logistic Regression and k-Nearest Neighbours in Rainbow Trout (Oncorhynchus Mykiss) Classification Using Image-Based Features. *Sensors*, *18*(4), 1027. https://doi.org/10.3390/s18041027

Saccon, F. A. M., Parcey, D., Paliwal, J., & Sherif, S. S. (2017). Assessment of Fusarium and Deoxynivalenol Using Optical Methods. *Food and Bioprocess Technology*, *10*(1), 34–50. https://doi.org/10.1007/s11947-016-1788-9

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). *NIH Image to ImageJ: 25 years of image analysis* (pp. 671–675). Nature methods.

Team, R. C. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.r-project.org/.

USDA, U. S. D. of A. (2016). *Grain Fungal Diseases and Mycotoxin Reference*

West, J. S., Canning, G. G. M., Perryman, S. A., & King, K. (2017). Novel Technologies for the detection of Fusarium head blight disease and airborne inoculum. *Tropical Plant Pathology*, *42*(3), 203–209. https://doi.org/10.1007/s40858-017-0138-4

Xia, R., Schaafsma, A. W., Wu, F., & Hooker, D. C. (2020). Impact of the improvements in Fusarium head blight and agronomic management on economics of winter wheat. *World Mycotoxin Journal*, *13*(3), 423–440. https://doi.org/10.3920/WMJ2019.2518

Zapotoczny, P. (2011). Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture. *Journal of Cereal Science*, *54*(1), 60–68. https://doi.org/10.1016/j.jcs.2011.02.012