Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física.
Maestría en Ciencia de Datos

DETECCIÓN DE CERCOSPORA EN AGAVE A TRAVÉS DEL ANALISIS DE IMÁGENES MULTIESPECTRALES.

Tesis obtener el grado de maestría.

Presenta: Claudia Patricia Mancera Almeida

Director: Alberto de Obeso Orendain

Tlaquepaque, Jalisco. Mayo 2021

Resumen

El Agave es un cultivo sumamente importante para el país ya que es la materia prima de varios destilados entre ellos el más importante en nuestro país; el Tequila, el cual cuenta con denominación de origen. La demanda del agave ha tenido un aumento contante en los últimos años por lo que su cuidado y supervisión durante la etapa de crecimiento se ha convertido en una tarea crucial para los agricultores. La cercospora es una de las enfermedades principales que ataca al cultivo del agave. Este hongo se manifiesta con manchas negras y si no es detectado y tratado a tiempo puede resultar en la muerte de la planta. Este estudio tiene como objetivo detectar este hongo a través de imágenes multiespectrales, probar diferentes modelos de clasificación y decidir cual es el ideal para encontrar esta enfermedad.

Palabras Claves

Agave, Imágenes Multiespectrales, Modelos de Clasificación

Dedicatoria

A mis hermanos, a mi novio y a mis padres que me dieron su apoyo incondicional para poder dar un paso más en mi preparación académica

1 CONTENIDO

2 INTRODUCCIÓN

2.1 ANTECEDENTES

En México la agricultura es uno de los sectores económicos más importantes históricamente del país. Aun cuando el porcentaje del PIB ha disminuido a través de los años, en 2018, México produjo:

- 56.8 millones de toneladas de caña de azúcar (sexto productor mundial)
- 27.1 millones de toneladas de maíz (octavo productor mundial)
- 4.7 millones de toneladas de naranja (cuarto productor más grande del mundo)
- 4.5 millones de toneladas de tomate (noveno productor más grande del mundo)
- Entre muchos otros cultivos.

FIGURA 1. Porcentaje del PIB que representa la agricultura en México históricamente (Banco Mundial, 2021)

El desafío de la agricultura, según la Organización de las Naciones Unidas para la Alimentación (FAO) en la Agenda 2030, es que los productores aumenten la productividad promoviendo sistemas productivos integrados, haciendo uso de tecnologías, con el fin de reducir las exploraciones de los suelos agrícolas y resistir a los cambios climáticos, como un instrumento para garantizar la demanda de alimentos, debido al crecimiento de la población y al desarrollo económico mundial. (Food and Agriculture Organization of the United Nations, 2020)

En Mexico tenemos 1.972.550 km2 de los que el 29% son tierras cultivables, pero sólo el 1,28% son cosechas permanentes. El agave no solo es la materia prima del segundo producto más exportado en Mexico: El tequila, también de otros destilados que han tomado mucha fuerza los últimos años como el Mezcal. Los agaves se originaron hace aproximadamente 10 millones de

años y México es el centro de diversidad biológica de estos. Existen aproximadamente 200 especies de agave de las cuales 150 se encuentran en México, lo cual representa un 75%. (Garcia Mendoza, 2007). En México existen sembradas 109,568 hectáreas de agave, de estas hectáreas, más del 80% corresponde al cultivo de agave azul y solo 14,460 hectáreas a otras distintas especies utilizados en la industria mezcalera.

No es de extrañarse que la mayoría de las hectáreas sean del agave azul (Agave tequilana W) ya que es la especie que se requiere para la producción del Tequila siendo este la bebida nacional debido a que posee denominación de origen.

Entendemos como denominación de origen, el nombre de una región geográfica del país que sirva para designar un producto originario de la misma, y cuya calidad o características se deban exclusivamente al medio geográfico. (Secretaria de Economía, 2015). Solamente se puede cosechar agave en 5 estados de México para que pueda ser procesado y etiquetado como Tequila y 8 estados como Mezcal. La Denominación de Origen del Tequila (DOT) está conformada por 181 municipios de los estados de Guanajuato (7), Michoacán (30), Nayarit (8), Tamaulipas (11) y la totalidad de Jalisco (125).

FIGURA 2. Estados donde únicamente se pueden producir Tequila y mezcal debido a la denominación de origen.

Jalisco es el estado con mayor superficie sembrada de agave. Cuenta con 67,822 hectáreas de agave lo cual representa el 77% del país aproximadamente. Guanajuato es el segundo estado con mayor superficie. Cuenta con 8,573 hectáreas de agave. La especie Agave tequilana es la de mayor superficie. Oaxaca es el tercer estado con mayor superficie. Cuenta con 8,100

hectáreas. En este estado se siembran diversas especies de agave para la elaboración de mezcal. El mezcal obtenido suele llevar el nombre del tipo de agave utilizado en su elaboración, algunos ejemplos de ello son el mezcal espadín (Agave vivípara), tobalá (Agave potatorum), bacanora (Agave angustifolia), etc.

La derrama económica que la industria del Tequila genera en la región es de 1,600 millones de pesos anuales (SIAP 2015) y la continuación de esta potencia económica depende de la denominación de origen y las implicaciones de mantener la región DOT son la rentabilidad y sustentabilidad de la región.

A partir de la década de los 90´s, los sembradíos de agave en Jalisco han sido fuertemente afectados por factores fitosanitarios entre lo que destacan el picudo (Scyphophorus acupunctatus), la marchitez por Fusarium spp, y la mancha gris causada por Cercospora agavicola (Juan José Coria-Contreras, 2018). Y por eso, el monitoreo de los cultivos se ha convertido en uno de los retos más importantes para optimizar su cosecha.

Actualmente, una de las principales preocupaciones de productores es la dificultad del control y prevención de C. agavicola,. En el estado de Jalisco históricamente se han reportado daños en Los Altos (Acatic, San Juan de los Lagos, Tepatitlán y Yahualica), Valles (Amatitán, Ahualulco, Arenal, Magdalena, San Juanito de Escobedo y Tequila), Sur (Autlán, El Grullo, El Limón y Unión de Tula) y Sierra Occidental (Atenguillo Mascota y Mixtlán) (Rubio, 2007)

El uso de nuevas tecnologías y ciencia de datos como una estrategia paralela de investigación y/o toma de decisiones de negocio cada vez es más utilizada en diferentes industrias. En el caso de la agricultura se está explorando las aplicaciones de esta ciencia utilizando sensores en la tierra, tomando de imágenes con drones, entre otras opciones innovadoras buscando obtener la mayor cantidad de datos significativos que les brinde información sobre el estado de sus cultivos teniendo como objetivo asegurarse que cuentan con los nutrientes necesarios, tener un control de enfermedades, poder predecir su producción por ende la rentabilidad de las parcelas. En México vamos un poco atrás en la implementación y aprovechamiento de estas tecnologías en esta industria sin embargo algunos productores están comenzando, por ejemplo, productores de Café en Veracruz y Agave en Jalisco.

Este estudio pretende analizar a través de imágenes multiespectrales plantas de Agaves que fueros sometidos a diferentes tratamientos y además expuestos a Cercospora Agavicola, identificar si hay diferencias entre los tratamientos y con un algoritmo de clasificación determinar si hay presencia de la enfermedad en el agave.

2.2 PROBLEMA

La detección de enfermedades en agaves en el estado de Jalisco específicamente la Cercospora agavicola es una tarea difícil para los productores de Agave, por lo que el control y la prevención se ha convertido en uno de los mayores retos que tienen en la industria.

2.3 OBJETIVOS

2.3.1 Objetivo General

Detectar a través de imágenes multiespectrales la presencia de la enfermedad (Cercospora agavicola) en plantas de agave azul.

2.3.2 Objetivos Específicos

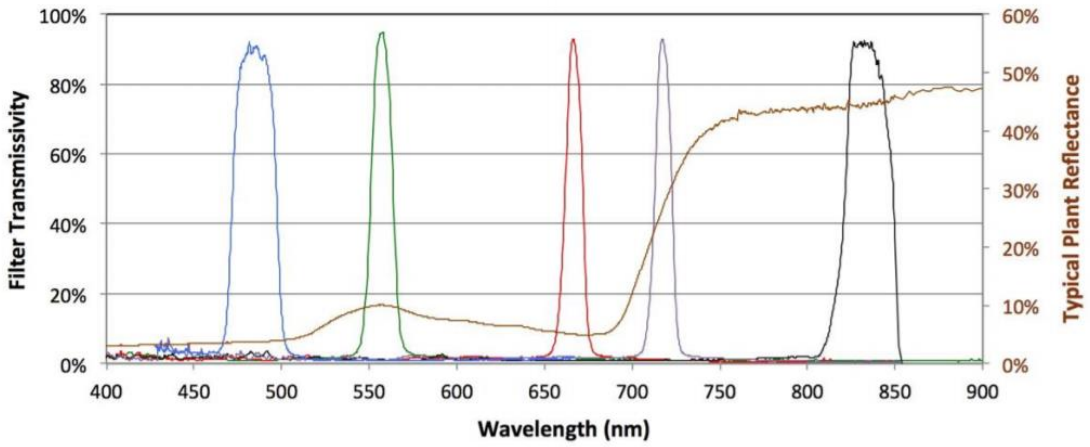
- Monitorear el desarrollo de la enfermedad en los primeros 20 días.
- Desarrollar un algoritmo de clasificación para poder identificar plantas enfermas y sanas
- Determinar si existe diferencia en el desarrollo de la enfermedad entre los diferentes tratamientos a los que fueron sometidos los agaves

3 MARCO CONCEPTUAL

3.1 IMÁGENES MULTIESPECTRALES

Una imagen multiespectral es la captura de una misma escena en diferentes longitudes de onda, a estos niveles de longitud de onda los llamamos canales, espectros. Nuestro ojo humano solo tiene la capacidad de detectar los espectros rojo, azul y verde; cuando estos se combinan logramos obtener una imagen RGB que nos da como resultado la gama de colores que vemos normalmente. Existen diferentes tipos de espectros:

- Blue: 450–515..520 nm
- Green, 515..520–590..600 nm
- Red, 600..630–680..690 nm,
- Near infrared (NIR), 750–900 nm,
- Mid-infrared (MIR), 1550–1750 nm,
- Far-infrared (FIR), 2080–2350 nm
- Thermal infrared, 10400-12500 nm,

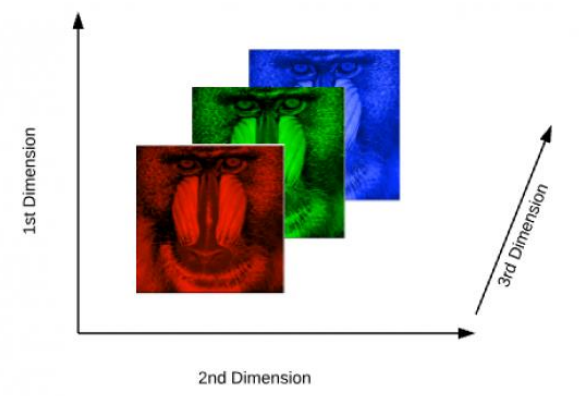


Band Number	Band Name	Center Wavelength (nm)	Bandwidth FWHM (nm)
1	Blue	475	20
2	Green	560	20
3	Red	668	10
4	Near IR	840	40
5	Red Edge	717	10

FIGURA 3. Longitud de onda de imágenes multiespectrales

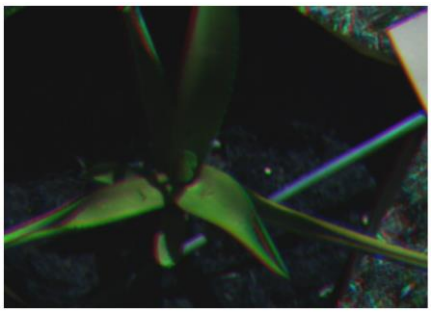
Dependiendo la cámara con la que se tomen las imágenes son los espectros que se podrán obtener. Cuanto mayor sea el número de píxeles en una imagen, mejor es su resolución, las imágenes tomadas por una cámara convencional cuentan con 3 espectros (Red, Blue y Green) por eso su nombre de imágenes RGB. Sin embargo, las cámaras multiespectrales tienen la capacidad de aumentar este número de espectros. Cada imagen se representa como una matriz n x m x s donde n x m es la cantidad de pixeles que la imagen contiene (el tamaño de la imagen) y s el número de espectros.

DIMENSIONES IMAGEN RGB



IMÁGENES EN 5 ESPECTROS UTILIZADAS EN ESTE ESTUDIO,

RGB



BLUE

GREEN

RED

NIR

RED EDGE

3.1.1 Índices Vegetativos

A partir de que tenemos imágenes enviadas por los satélites, se ha hecho un gran esfuerzo por poder hacer análisis de las superficies. Las imágenes multiespectrales han demostrado una gran habilidad para encontrar, agua, vegetación, minerales entre muchas cosas más a través de índices vegetativos. Dependiendo del caso de estudio las imágenes puedes ser capturadas por satélites, drones o simplemente personas que cuentes con una cámara multiespectral.

Al hablar de índices vegetativos nos referirnos a un conjunto de operaciones algebraicas efectuadas sobre los valores numéricos de los pixeles, usando dos o más bandas pertenecientes a la misma escena. Estas fórmulas matemáticas han sido propuestas por diferentes autores a través del tiempo y en las últimas décadas se han desarrollado más de 40 índices vegetativos diferentes.

A. First Generation Indices

Index	Abbreviation	Formula	Author and Year
Ratio Vegetation Index	RVI	$\frac{R}{NIR}$	Pearson and Miller, 1972
Vegetation Index Number	VIN	$\frac{NIR}{R}$	Pearson and Miller, 1972
Transformed Vegetation Index	TVI	$\sqrt{NDVI + 0.5}$	Rouse et al., 1974
Green Vegetation Index	GVI	$(-0.283MSS4 - 0.660MSS5 + 0.577MSS6 + 0.388MSS7)$	Kauth and Thomas, 1976
Soil Brightness Index	SBI	$(0.332MSS4 + 0.603MSS5 + 0.675MSS6 + 0.262MSS7)$	Kauth and Thomas, 1976
Yellow Vegetation Index	YVI	$(-0.899MSS4 + 0.428MSS5 + 0.076MSS6 - 0.041MSS7)$	Kauth and Thomas, 1976
Non Such Index	NSI	$(-0.016MSS4 + 0.131MSS5 - 0.425MSS6 + 0.882MSS7)$	Kauth and Thomas, 1976
Soil Background Line	SBL	$(MSS7 - 2.4MSS5)$	Richardson and Wiegand, 1977
Differenced Vegetation Index	DVI	$(2.4MSS7 - MSS5)$	Richardson and Wiegand, 1977
Misra Soil Brightness Index	MSBI	$(0.406MSS4 + 0.600MSS5 + 0.645MSS6 + 0.243MSS7)$	Misra et al., 1977
Misra Green Vegetation Index	MGVI	$(-0.386MSS4 - 0.530MSS5 + 0.535MSS6 + 0.532MSS7)$	Misra et al., 1977
Misra Yellow Vegetation Index	MYVI	$(0.723MSS4 - 0.597MSS5 + 0.206MSS6 - 0.278MSS7)$	Misra et al., 1977
Misra Non Such Index	MNSI	$(0.404MSS4 - 0.039MSS5 - 0.505MSS6 + 0.762MSS7)$	Misra et al., 1977
Perpendicular Vegetation Index	PVI	$\sqrt{(\rho_{sol} - \rho_{végé})^2_R + (\rho_{sol} - \rho_{végé})^2_{NIR}}$	Richardson and Wiegand, 1977
Ashburn Vegetation Index	AVI	$(2.0MSS7 - MSS5)$	Ashburn, 1978
Greenness Above Bare Soil	GRABS	$(GVI - 0.09178SBI + 5.58959)$	Hay et al., 1979
Multi-Temporal Vegetation Index	MTVI	$(NDVI(date\ 2) - NDVI(date\ 1))$	Yazdani et al., 1981
Greenness Vegetation and Soil Brightness	GVSB	$\frac{GVI}{SBI}$	Badhwar, 1981
Adjusted Soil Brightness Index	ASBI	$(2.0\ YVI)$	Jackson et al., 1983
Adjusted Green Vegetation Index	AGVI	$GVI - (1 + 0.018GVI)YVI - NSI/2$	Jackson et al., 1983
Transformed Vegetation Index	TVI	$\frac{(NDVI + 0.5)}{\lvert NDVI + 0.5\rvert}\sqrt{\lvert NDVI + 0.5\rvert}$	Perry and Lautenschlager, 1984
Differenced Vegetation Index	DVI	$(NIR - R)$	Clevers, 1986
Normalized Difference Greenness Index	NDGI	$\frac{(G - R)}{(G + R)}$	Chamard et al., 1991
Redness Index	RI	$\frac{(R - G)}{(R + G)}$	Escadafal and Huete, 1991
Normalized Difference Index	NDI	$\frac{(NIR - MIR)}{(NIR + MIR)}$	McNairn and Protz, 1993

Index	Abbreviation	Formula	Author and Year
Ashburn Vegetation Index	AVI	$(2.0MSS7 - MSS5)$	Ashburn, 1978
Greenness Above Bare Soil	GRABS	$(GVI - 0.09178SBI + 5.58959)$	Hay et al., 1979
Multi-Temporal Vegetation Index	MTVI	$(NDVI(date\ 2) - NDVI(date\ 1))$	Yazdani et al., 1981
Greenness Vegetation and Soil Brightness	GVSB	$\frac{GVI}{SBI}$	Badhwar, 1981
Adjusted Soil Brightness Index	ASBI	$(2.0\ YVI)$	Jackson et al., 1983
Adjusted Green Vegetation Index	AGVI	$GVI - (1 + 0.018GVI)YVI - NSI/2$	Jackson et al., 1983
Transformed Vegetation Index	TVI	$\frac{(NDVI + 0.5)}{\lvert NDVI + 0.5\rvert}\sqrt{\lvert NDVI + 0.5\rvert}$	Perry and Lautenschlager, 1984
Differenced Vegetation Index	DVI	$(NIR - R)$	Clevers, 1986
Normalized Difference Greenness Index	NDGI	$\frac{(G - R)}{(G + R)}$	Chamard et al., 1991
Redness Index	RI	$\frac{(R - G)}{(R + G)}$	Escadafal and Huete, 1991
Normalized Difference Index	NDI	$\frac{(NIR - MIR)}{(NIR + MIR)}$	McNairn and Protz, 1993

FIGURA 4. Primera generación de índices vegetativos (Bannari, 1995)

B. Second Generation Indices

Index	Abbreviation	Formula	Author and Year
Normalized Difference Vegetation Index	NDVI	$\frac{(NIR - R)}{(NIR + R)}$	Rouse et al., 1974
Perpendicular Vegetation Index	PVI	$\frac{(NIR - aR - b)}{\sqrt{a^2 + 1}}$	Jackson et al., 1980
Soil Adjusted Vegetation Index	SAVI	$\frac{(NIR - R)}{(NIR + R + L)}(1 + L)$	Huete, 1988
Transformed SAVI	TSAVI	$\frac{[a(NIR - aR - b)]}{(R + aNIR - ab)}$	Baret et al., 1989
Transformed SAVI	TSAVI	$\frac{[a(NIR - aR - b)]}{[R + aNIR - ab + X(1 + a^2)]}$	Baret and Guyot, 1991
Atmospherically Resistant Vegetation Index	ARVI	$\frac{(NIR - RB)}{(NIR + RB)}$ $RB = R - \gamma(B - R)$	Kaufman and Tanré, 1992
Global Environment Monitoring Index	GEMI	$GEMI = \eta(1 - 0.25\eta) - \frac{(R - 0.125)}{(1 - R)}$ $\eta = \frac{[2(NIR^2 - R^2) + 1.5NIR + 0.5R]}{(NIR + R + 0.5)}$	Pinty and Verstraete, 1992
Transformed Soil Atmospherically Resistant Vegetation Index	TSARVI	$\frac{[a_{rb}(NIR - a_{rb}RB - b_{rb})]}{[RB + a_{rb}NIR - a_{rb}b_{rb} + X(1 + a_{rb}^2)]}$	Bannari et al., 1994
Modified SAVI	MSAVI	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}}{2}$	Qi et al., 1994
Angular Vegetation Index	AVI	$\tan^{-1}\left\{\frac{\lambda_3 - \lambda_2}{\lambda_2}[NIR - R]^{-1}\right\} + \tan^{-1}\left\{\frac{\lambda_2 - \lambda_1}{\lambda_2}[G - R]^{-1}\right\}$	Plummer et al., 1994

FIGURA 5. Segunda generación de índices vegetativos (Bannari, 1995)

Los valores bajos de los índices de vegetación usualmente indican vegetación poco vigorosa, mientras que los valores altos, indican vegetación muy vigorosa. Sin embargo, en algunos casos (como los índices RVI y NRVI) el valor del índice de vegetación es inversamente proporcional a la cantidad de vegetación presente en el área.

Dependiendo de la cámara con la que se cuente y el objetivo del proyecto se debe explorar entro los índices y decidir cuál es el ideal.

3.2 CERCOSPORA AVICOLA

La cercospora avicola es un hongo que tiene como consecuencia manchas grises en los agaves cuando están infectados, es un hongo que está bajo control oficial de los estados que tienen Denominación de Origen.

Este hongo genera lesiones negras y ovaladas de 1-3 cm, la enfermedad se desarrolla rápidamente y ocasiona la muerte de la planta. A continuación, encontramos algunos ejemplos de cómo se ven las manchas ocasionadas por este hongo dentro de los rectángulos naranja.

Cuando la enfermedad llega a la piña, es muy difícil lograr la recuperación de la planta. Si no se aplican estrategias de control al presentarse dicha enfermedad, puede causar la muerte de las plantas de dos a seis meses, según la edad del cultivo y la intensidad del daño. Existen campañas contra Plagas Reglamentadas del Agave a partir del 2013 implementadas por el SENASICA (Servicio Nacional de Sanidad Inocuidad y Calidad Agroalimentaria)

3.3 LIBRERIAS DE PYTHON PARA ANALISIS DE IMAGENES

La limpieza y el preprocesamiento de datos es una parte esencial para cualquier proyecto basado en datos, es una de las partes que en algunas ocasiones toma más tiempo y además el hacerlo de forma correcta es crucial para el éxito del resultado. Para el tema de imágenes existen algunas librerías que nos ayudan con este proceso.

La librería Open CV es una biblioteca libre desarrollada originalmente por Intel en 1990. es una de las más utilizadas. Los algoritmos de esta librería permiten identificar objetos, caras, clasificar acciones humanas en vídeo, hacer tracking de movimientos de objetos, extraer modelos 3D, encontrar imágenes similares, eliminar ojos rojos, seguir el movimiento de los ojos, reconocer escenarios entre otras cosas.

También existe la librería de Scikit-image que de igual forma que Open CV es libre y cuenta con una gran cantidad de algoritmos. Ambas librerías usan como base imágenes RGB que son las imágenes que normalmente obtenemos de las cámaras comunes e incluso de celular.

Además de estas librerías, la marca Micasense que es una marca que vende cámaras multiespectrales desarrollo su propia librería donde se apalanca del trabajo realizado por las 2 librerías antes mencionadas para manipular y transformar las imágenes.

3.4 Modelos Predictivos y estadísticos

3.4.1 ANOVA

El análisis de varianza (ANOVA) es una prueba estadística que pone a prueba la siguiente hipótesis nula:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

Donde $\mu_i = 1, \ldots, k$ son las medias de k diferentes grupos de datos (tratamientos), bajo el supuesto de independencia de los datos, normalidad y homocedasticidad. La hipótesis alternativa es la siguiente:

$$H_a: \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$$

ANOVA utiliza la prueba F para determinar si la variación es lo suficientemente grande como para ser considerada estadísticamente significativa.

3.4.2 Regresión logística

La regresión logística modela la probabilidad de que cada dato de entrada pertenezca a una categoría en particular. Para generar dichas probabilidades este modelo utiliza la función sigmoide donde los datos siempre estarán entre 0 y 1

$$sigmoid = \frac{1}{1 + \exp(-n)}$$

La función tiene pesos representados como theta en nuestra notación, y queremos encontrar los mejores valores para esos pesos. Para comenzar, elegimos números aleatorios, y necesitamos una manera de medir el desempeño del algoritmo usando esos pesos. Eso lo conseguimos usando la siguiente función de pérdida.

$$h = g(X\theta)$$

$$J(\theta) = \frac{1}{m}.(-y^T \log(h) - (1-y)^T \log(1-h))$$

Nuestro objetivo es minimizar la función de pérdida, ajustando los valores en los pesos. La derivada de la función de perdida nos indica si los pesos deben incrementar o disminuir.

$$gradient = \frac{\delta J(\theta)}{\delta \theta_j} = \frac{1}{m} X^T (g(X\theta) - y)$$

Después actualizamos los pesos restándole la derivada por el valor de aprendizaje. Repetimos este proceso hasta encontrar la solución óptima.

$$theta = learning_{rate} * gradient$$

Por último, llamamos a la función sigmoide para obtener la probabilidad de que un dato de entrada x pertenezca a la clase 1. Podemos tomar las probabilidades <0.5= clase 1 y el resto clase 0. Este criterio puede ser ajustado y definido basado en el problema que se está intentando resolver.

$$\hat{y} = \frac{1}{1 + \exp(theta)}$$

3.4.3 Random Forest and bagging

Un Random Forest es un conjunto (ensemble) de árboles de decisión combinados con bagging. Al usar bagging, lo que en realidad está pasando, es que distintos árboles usan distintas porciones de los datos. Ningún árbol ve todos los datos de entrenamiento. Esto hace que cada árbol se entrene con distintas muestras de datos para un mismo problema. De esta forma, al combinar sus resultados, unos errores se compensan con otros, se reduce la varianza y tenemos una predicción que generaliza mejor.

Este modelo suele ser mejor para clasificación que para regresión ya que las predicciones no son de naturaleza continua.

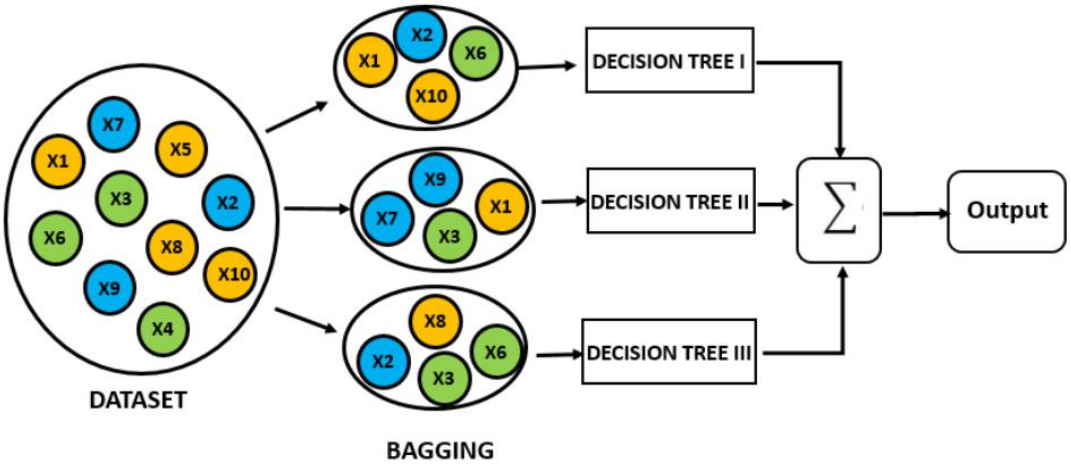


FIGURA 6. Ensambladores: Random Forest (Alvear, 2018)

3.4.4 Maquinas de Soporte Vectorial (SVM)

La metodología de SVM se puede resumir de la siguiente manera: Si se requiere clasificar un set de datos (representados en un plano n-dimensional) no separable linealmente y esos datos se pueden transformar utilizando una función Kernel a otro espacio donde es posible hacer dicha separación lineal.

En este nuevo plano, buscamos el hiperplano que es capaz de separar estos datos en dos clases; los puntos más cercamos a este hiperplano son los vectores de soporte.

Vemos el modelo de SVM como el siguiente problema de clasificación:

$$\min J(w, \varepsilon, b) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{n} \varepsilon_i^2$$

Sujeto a:

$$y_i[w^T \varphi(x_i) + b] = 1 - \varepsilon_i, \qquad i = 1, \dots, n$$

$$y_i \in \{-1,1\}$$

Definimos el Lagrangeano:

$$L(w, b, \varepsilon; \alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} \alpha_i (y_i[w^T \varphi(x_i) + b] - 1 + \varepsilon_i)$$

Donde αk son los multiplicadores de Lagrange (los cuales pueden ser positivos o negativos debido a a las condiciones de Kuhn-Tucker(Fletcher, 1987)).

Las condiciones para encontrar el óptimo:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i}^{n} \alpha_i y_i \varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \\ \dfrac{\partial L}{\partial \varepsilon_i} = 0 \rightarrow \alpha_i = \gamma \varepsilon_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i[w^T \varphi(x_i) + b] - 1 + \varepsilon_i = 0 \end{cases}$$

Se puede escribir como la solución del siguiente sistema de ecuaciones lineales:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1v \end{bmatrix}$$

Donde:

$$Y = [y_1, y_2, \dots, y_m]^T, 1_V = [1,1, \dots, 1]^T, \alpha = [\alpha_1, \alpha_2, \dots, \alpha_m]^T,$$
$$Z = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, \dots, \varphi(x_m)^T y_m]^T$$

Basado en las condiciones de Mercer:

$$\Omega = ZZ^T$$

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$$

$$\Omega_{ij} = y_i y_j \, K(x_i, x_j)$$

RBF Kernel:

$$K(x_i, x_j) = exp\left\{\frac{-\|x_j - x_i\|^2}{2\sigma^2}\right\}$$

Como consecuencia es necesario seleccionar 2 parámetros: σ y γ .

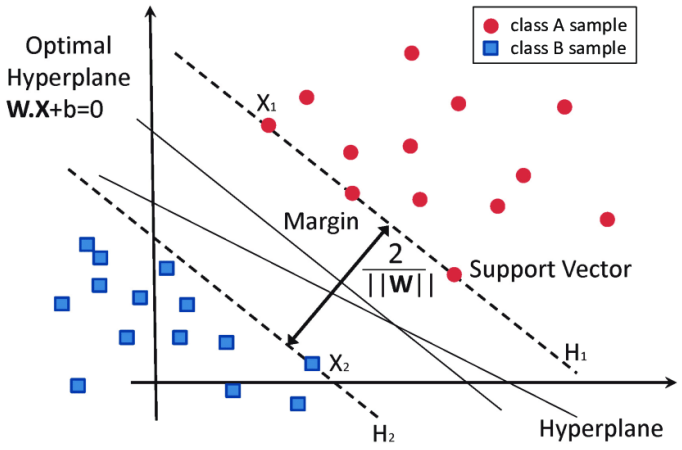


FIGURA 7. Maquinas de soporte vectorial

4 DESARROLLO

4.1 MATERIALES Y MÉTODOS

Las plantas de agave estudiadas se encuentran físicamente en las instalaciones del Centro Universitario de Ciencias Biológicas y Agropecuarias (CUCBA) el cual pertenece a la Universidad de Guadalajara.

El estudio completo cuenta con 45 plantas divididas en 9 tratamientos (5 plantas por tratamiento. Todas las plantas fueron expuestas de manera intencional con el hongo de la Cercospora Agavicola y a su vez cada tratamiento implica la inoculación de algún microorganismo con el objetivo de encontrar ventajas o desventajas entre ellos en cuestión del desarrollo de defensas y de la misma enfermedad.

Tratamientos	
T1	Bacterias
T2	Trichodermas
T3	Combinacion bacterias y trichoderma
T4	Bacterias
T5	Algas marinas
T6	silicio
T7	Algas marinas
T8	Algas marias
T9	testigo

FIGURA 8. Tratamientos utilizados para el estudio

Para este este trabajo se consideraron los tratamientos 1, 7 y 8. La diferencia entre T7 y T8 que tienen Algas Marinas es la fuente de donde se obtuvieron dichas algas más no es otro microorganismo diferente.

La primera toma de imágenes se realizó el día 4 de Marzo del 2020, ese mismo día se realizó la inocuización tanto del tratamiento como de la enfermedad. Posteriormente se tomaron imágenes por los siguientes 24 días, siendo la última imagen capturada el día 28 de Marzo 2020. No contamos con la información diaria, hay algunos días faltantes.

Para tomar las fotos se usó una cámara multiespectral profesional. El modelo es RedEdge-M de la marca Micasense. Esta cuenta con 5 espectros

Representación del espectro de luz visible al no visible (rango 400 nm - 900 nm)

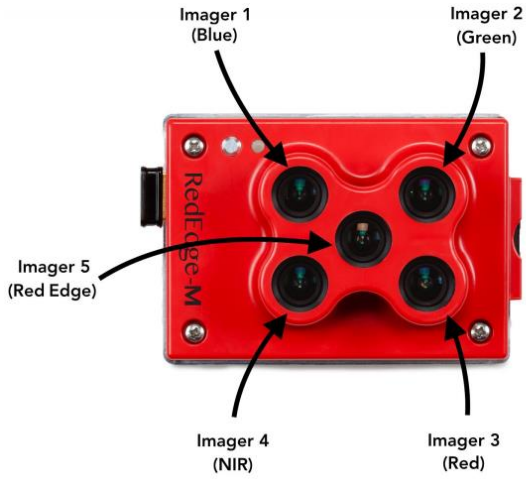


FIGURA 9. Cámara RedEdge-M de Micasense y sus lentes

Cada imagen tiene metada. Esta la podemos obtener con la librería de Micasense y la información que podemos encontrar es la siguiente:

```
MicaSense RedEdge-M firmware version: v5.1.8
Exposure Time: 0.00054 seconds
Imager Gain: 1.0
Size: 1280x960 pixels
Band Name: Blue
Center Wavelength: 475 nm
Bandwidth: 20 nm
Capture ID: 778c46lKBdl5SqpUnSRs
Flight ID: 9eDMyHM8XhmdwbDrZzMZ
Focal Length: None
```

FIGURA 10. Meta data de cada imagen disponible en la librería de Micasense

Para este estudio en donde se utilizarán algoritmos supervisados es necesario tener imágenes con plantas enfermas y con plantas sanas para poder entrenar los modelos. Por ese motivo, basando en la información obtenido de la investigadora utilizamos 48 plantas, XX enfermas y XX sanas. Consideramos como plantas sanas las imágenes de 4 y del 6 de Marzo y consideramos plantas enfermas imágenes del 24,26 y 28 de Marzo.

4.1.1 Preprocesamiento de datos

Aunque generalmente visualizamos las imágenes multiespectrales como índices o compuestos coloridos, la cámara realmente captura las imágenes en escala de grises, que son esencialmente matrices de números digitales.

Cada píxel o celda dentro de un espectro contiene un número digital correspondiente a la intensidad de la radiación dentro de una determinada longitud de onda. Cada una de las 5 imágenes de los agaves tiene unas dimensiones de 1280 x 960, lo que significa que el número total de píxeles en cada imagen es 1.228.800 en cada uno de los espectros.

Los valores de píxeles son relativos a las condiciones en las que se recopilaron los datos y no son absolutos. Esto se debe principalmente a los cambios en las condiciones de luz y del clima. Aun cuando las imágenes fuesen tomadas en la misma hora del día podemos encontrar diferencias significativas. (por ejemplo, soleado frente a nublado, sol en diferentes puntos del cielo durante el día, sombra en algunas plantas etc.).

Por lo tanto, lo primero que hicimos fue hacer una corrección radiométrica, es decir, calibrar las imágenes. Esto implica calcular la reflectancia. La radiancia es cuanta luz tiene la imagen y la reflectancia es la proporción de la cantidad de energía reflejada en un objeto.

Para eso utilizamos un panel de calibración, este panel tiene valores de reflectancia medido previamente que actúa como un “control”. Cada vez que se realiza una toma de imágenes, en nuestro caso es por día, se debe tomar una del panel. (Micasense, 2020)

Debido a que la reflectancia de la planta puede ser un indicador de salud, estrés, enfermedad, entro otras cosas los valores de reflectancia precisos son claves para comprender la fisiología de la planta y comparar imágenes de un día a otro o de un tratamiento a otro. Los análisis temporales no son posibles sin tener en cuenta las condiciones de iluminación y, por lo que se requiere una calibración radiométrica de calidad.

FIGURA 11. Panel de calibración

En la imagen anterior podemos observar el panel y señalo la parte del panel que se toma como base para calibrar las imágenes.

En este mismo ejemplo la librería de Micasense calcula la radiancia del cuadrado gris del panel además de otros valores como el conteo de pixeles la desviación estándar entre otros datos. Y con estos datos se calibran las imágenes para que todas estas estén en igualdad de condiciones de luz.

FIGURA 12. Lectura del panel de calibración

```
Panel found: True
Panel serial: RP04-1826246-SC
Panel mean raw radiance value: 26064.762666666666
Panel raw pixel standard deviation: 1857.2192282621713
Panel region pixel count: 19500
Panel region saturated pixel count: 19500
```

FIGURA 12. Datos obtenidos a través de la lectura del panel de calibración

Una vez calibradas las imágenes tenemos que alinearlas, este paso es necesario ya que las imágenes fueron tomadas a una corta distancia y la separación que tienen los lentes de la cámara genera una diferencia en cada espectro.

Esta alineación se logró con ayuda de la librería OpenCV utilizando el método de alineación por puntos función, en el que se busca identificar puntos estables o esquinas de cada imagen los cuales llamamos “keypoints” y después la homografía relaciona estos puntos entre cada uno de los espectros y los alinea. Esto lo hacemos de dos en dos imágenes siempre alineándonos con el lente central de la cámara. Nuestro lente central es Red Edge por lo que cada uno de los espectros se buscan keypoints y los alinean con este lente.

IMAGEN RGB SIN ALINEAR

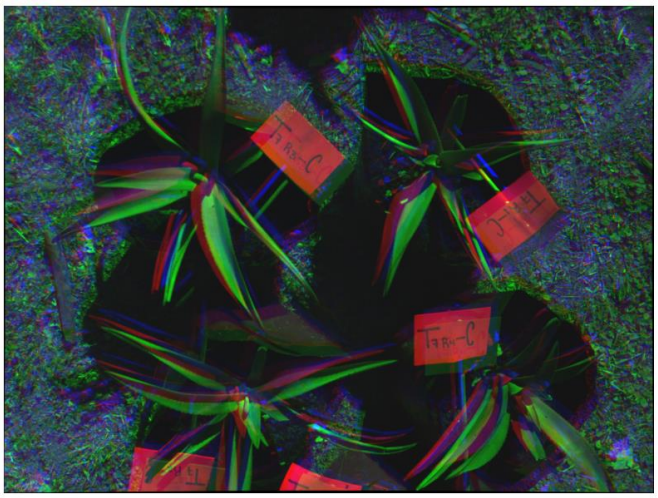


IMAGEN RGB ALINEADA

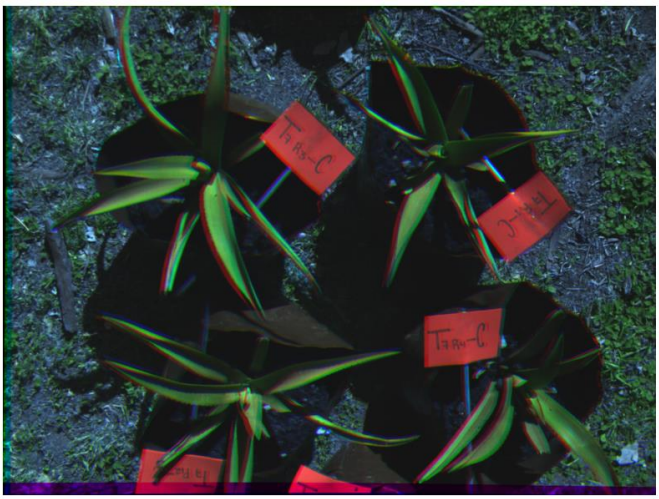


FIGURA 13. Alineación de imágenes.

Una vez alineadas las imágenes separamos las plantas dentro de la imagen, esto debido a que no todas las imágenes contaban con el mismo número de plantas y con el mismo acomodo. Por lo que seleccionamos la parte de la imagen donde se encuentra cada planta, pero todas las mantuvimos del mismo tamaño 250x 350 pixeles.

Ahora si es momento de calcular los índices vegetativos, como lo vimos en la sección anterior existen un gran número de índices vegetativos, sin embargo, basado en la investigación que realice utilizamos inicialmente los siguiente 4 índices:

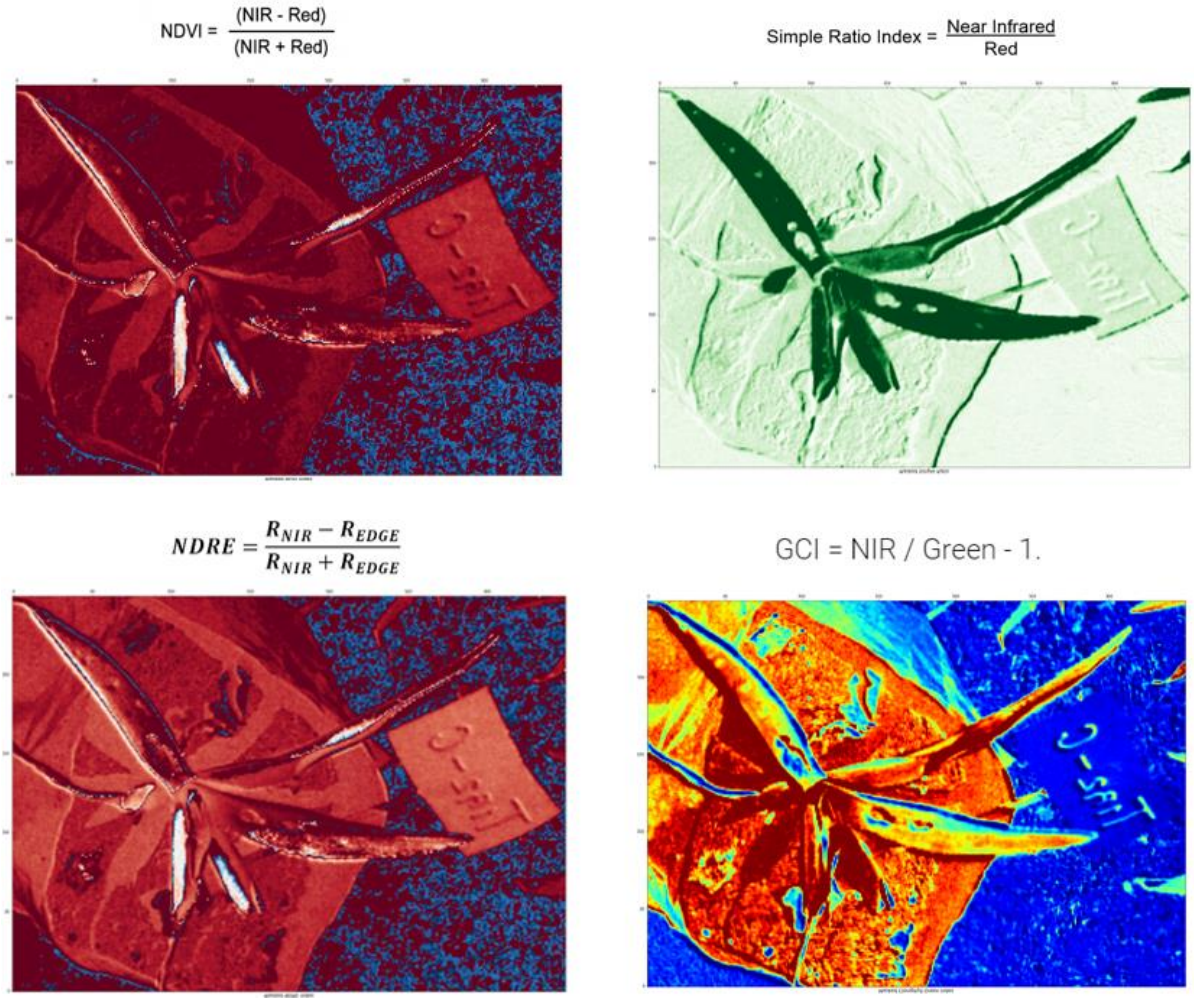


FIGURA 14. Comparación 4 índices vegetativos en una toma de agave

FIGURA 15. Imagen en RGB de la toma anterior

Visualmente podemos percibir que los índices vegetativos NDVI, GCI y NDRE no logran identificar contundentemente las manchas que vemos en la imagen RGB. Si vemos el índice Simple Ratio (SR) en ese índice si podemos encontrar diferencias por lo que ese es el índice que se utilizará para la detección de Cercospora.

Se probo de diferentes maneras utilizar algoritmos de clasificación con los pixeles de ese índice. Al ser imágenes de 250 x 350 pixeles cada imagen cuenta con 87,500 datos y por ende el tiempo de entrenamiento era muy alto y además los resultados no eran favorables.

El principal motivo de la falta de eficiencia de los modelos es que el color de las manchas es muy parecido al del fondo por lo que los pixeles de ambos se encontraban dentro del mismo rango numérico, es ahí donde tuvimos que agregar un algoritmo determinístico el cual es capaz de detectar si un pixel que por su valor sospechamos que es enfermedad, se encuentra dentro de la planta (evaluando proximidad con materia orgánica y materia orgánica en el perímetro), de ser así lo determina como mancha y remplaza su valor por 15 que sería el valor máximo dentro de la matriz, si el pixel esta fuera de la planta lo considera fondo y no cambia su valor. De esa manera los pixeles de las manchas ahora tienen un valor mucho más grande que el resto. El rango ahora es de 0 a 15. Así es como se visualiza la imagen después de este paso:

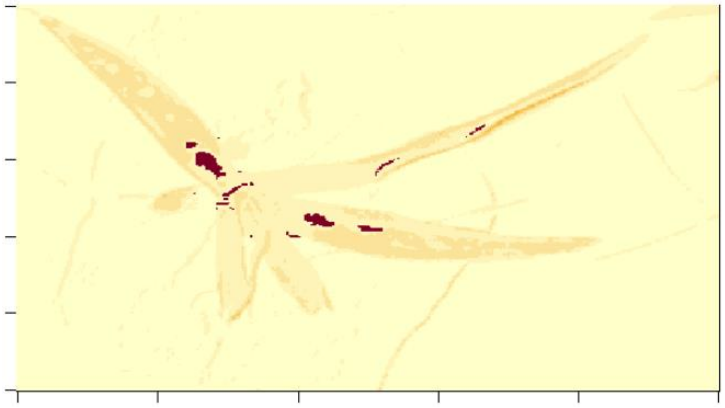


FIGURA 16. Resultado de algoritmo determinístico para encontrar las manchas

Este algoritmo al evaluar uno por uno los pixeles de la planta tarda entre 20 y 25 segundos por planta en realizar el proceso. Evaluar las 48 plantas tomó 10 min aproximadamente. Una vez concluido este paso, con el objetivo de reducir el tiempo de entrenamiento de los modelos se realiza un conteo de pixeles, creando 3 variables: “Numero de pixeles menor o igual a 5”, “Numero de pixeles mayor a 5 & menor a 10” y “Numero de pixeles mayor o igual a 10”. De esta forma estamos reduciendo de 87k datos por planta a 3 y esas serán las nuevas variables que utilizaremos para el algoritmo de clasificación.

4.1.2 MODELOS PREDICTIVOS

Se entrenaron 3 modelos de clasificación: Regresión logística, SVM (Maquina de soporte vectorial), y Árboles de decisión. Se utilizó un Split de 80% de entrenamiento y 20% de prueba. Los resultados fueron los siguientes:

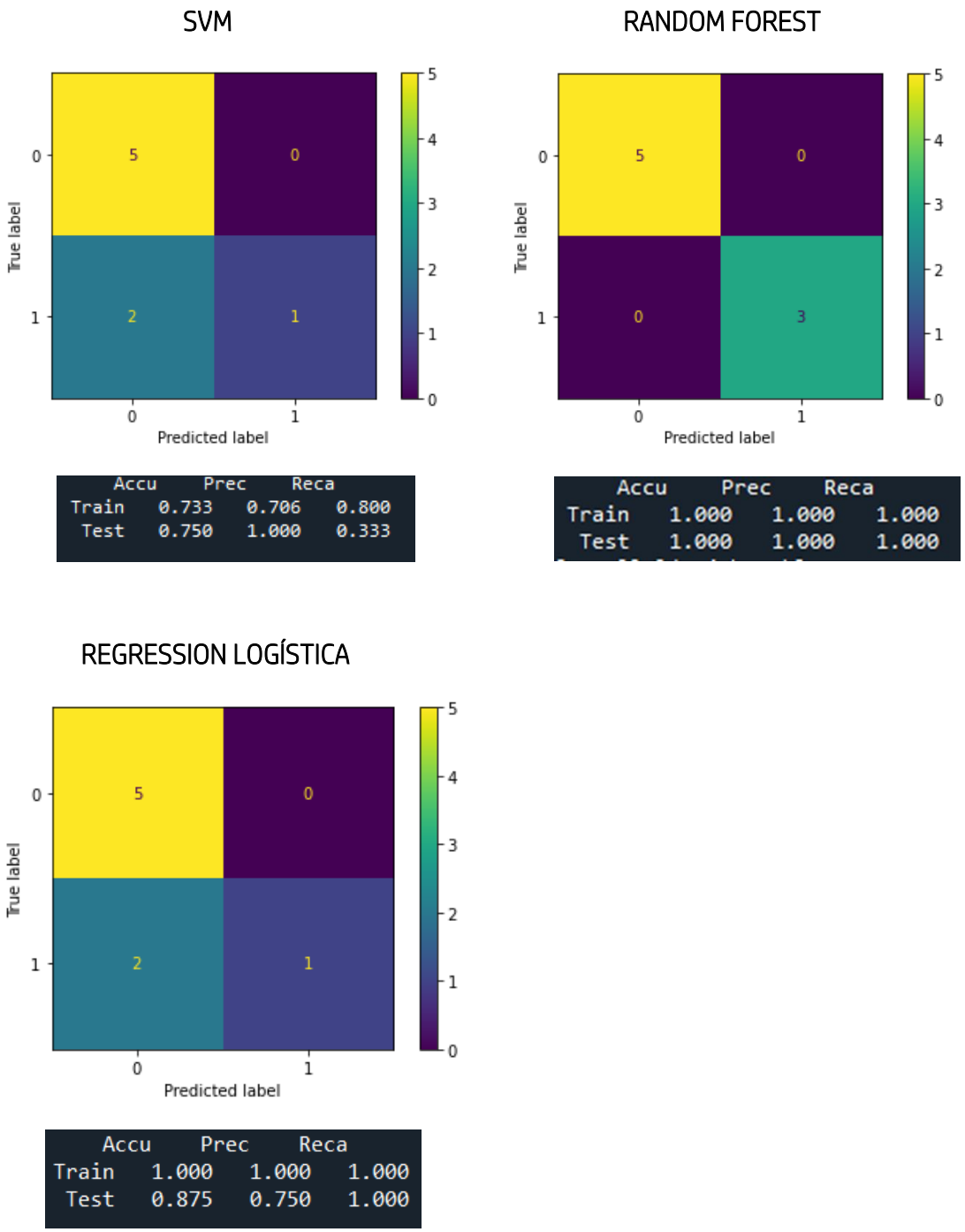
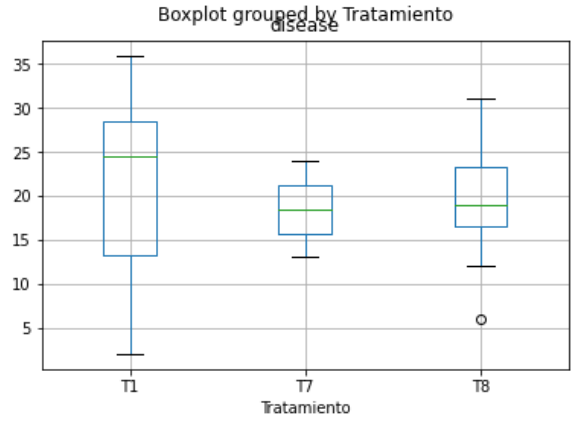


FIGURA 17. Comparación de resultados de modelos de clasificación.

4.1.3 ANOVA

A simple vista revisando las imágenes RGB podemos ver que todos los agaves presentan las manchas grises que caracterizan la Cercospora Agavicola sin embargo no es posible definir si realmente existe una diferencia significativa en los pixeles. Es por eso que se realizó un ANOVA. Se dividieron en dos los datos para no comparar plantas sanas con enfermas. Se realizo un ANOVA para cada grupo.

Para las plantas sanas es de esperarse que se rechace la hipótesis nula concluyendo que no hay diferencias significativas entre los tratamientos ya que aún no existe enfermedad; y así como se esperaba podemos ver en los resultados que en efecto no existe diferencia:



	sum_sq	df	F	PR(>F)
Tratamiento	69583.926676	2.0	1.775916	0.184248
Residual	685685.047009	35.0	NaN	NaN

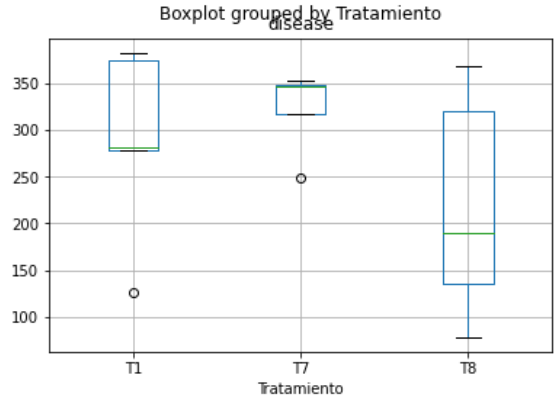
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1	T7	111.6923	0.2188	-48.8696	272.2542	False
T1	T8	2.3034	0.9	-122.3549	126.9617	False
T7	T8	-109.3889	0.1999	-261.9462	43.1684	False

FIGURA 18. Resultados ANOVA de plantas sanas

Después se realizó el ANOVA para las plantas enfermas, es ahí donde queríamos verificar si realmente hay una diferencia estadísticamente hablando ya que visualmente todas las plantas presentaban síntomas del hongo.

De igual forma que en la prueba anterior rechazamos la hipótesis nula y podemos decir que no hay evidencia estadística que demuestre que hay diferencia entre las medias de los tratamientos.




```
               sum_sq    df         F    PR(>F)
Tratamiento   46355.3   2.0  2.384779  0.122206
Residual     165222.9  17.0       NaN       NaN
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj    lower    upper   reject
-----------------------------------------------------
    T1     T7     34.2 0.8369 -125.7099 194.1099  False
    T1     T8    -76.1 0.3591 -214.5861  62.3861  False
    T7     T8   -110.3 0.1321 -248.7861  28.1861  False
-----------------------------------------------------
```

FIGURA 19. Resultados ANOVA de plantas enfermas

5 ESTADO DEL ARTE

5.1 AGRICULTURA DE PRECISIÓN

El aumento en la demanda agrícola como consecuencia de la sobrepoblación mundial y el cambio climático han obligado a introducción de nuevas tecnologías que sean más eficientes para satisfacer dicha demanda. La Agricultura de Precisión (AP) se define como el conjunto de tecnologías que se aplican al trabajo en el campo con satélites, sensores, imágenes y datos geográficos, que reúnen la información necesaria para entender las variaciones del suelo y los cultivos. Aun cuando no es un concepto nuevo, el avance en las tecnologías, los equipos y la facilidad procesar los datos ha despertado el interés de muchos productores agrícolas ya que colabora en una parte fundamental: el uso de los recursos.

Entre los beneficios que se han encontrado son control y monitoreo de enfermedades, monitoreo de desarrollo vegetativo, sensores que regulan variables cruciales para el desarrollo vegetativo como la humedad, el PH del suelo y la temperatura. En otras palabras, la agricultura de precisión ayuda a los agricultores a tomar mejores decisiones que eficienticen los recursos, maximicen su cosecha y por ende el rendimiento de cada porción de su sembradío. No solo representa un gran beneficio para la industria sino también para el medio ambiente.

La agricultura de precisión sebe ser adoptada gradualmente:

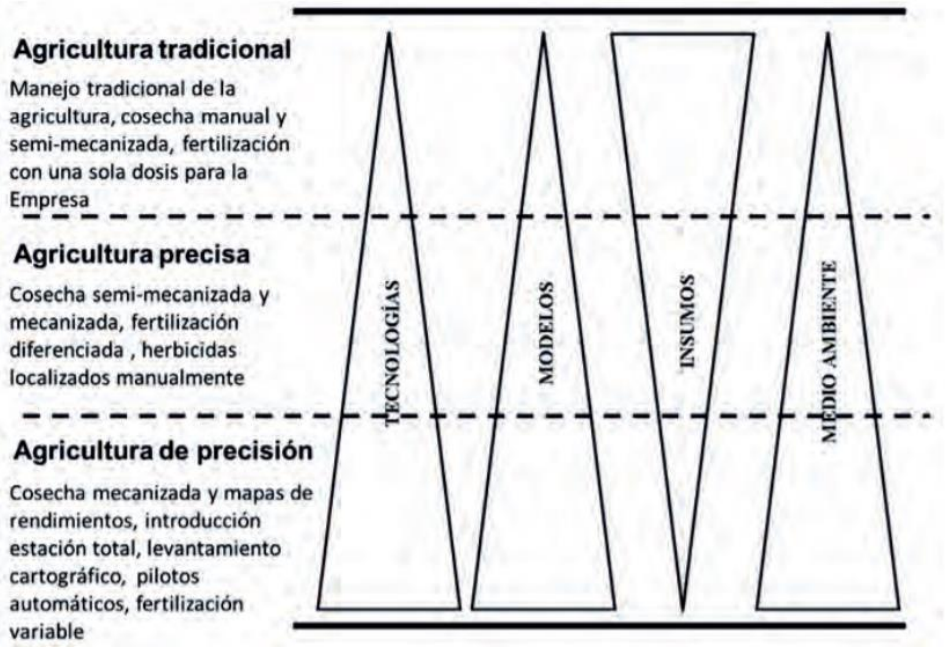


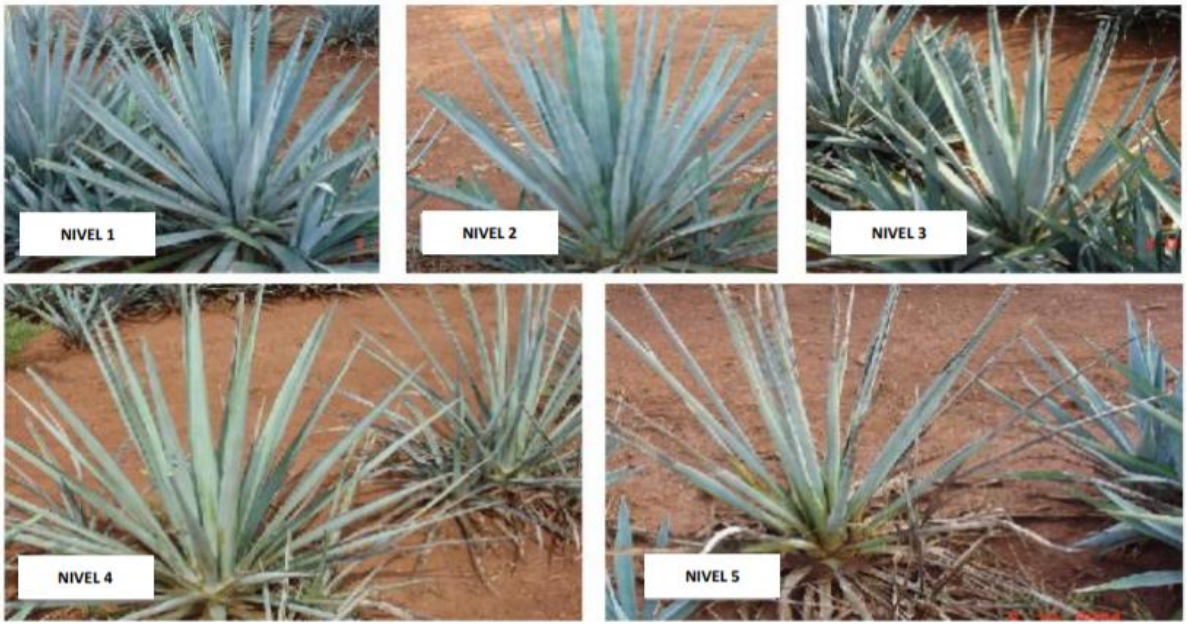
FIGURA 20. Grado de participación de los principales componentes de los costos de producción agrícola, considerando nivel de desarrollo tecnológico de su gestión (Arcia Porrúa, 2020)

5.2 AP En mexico

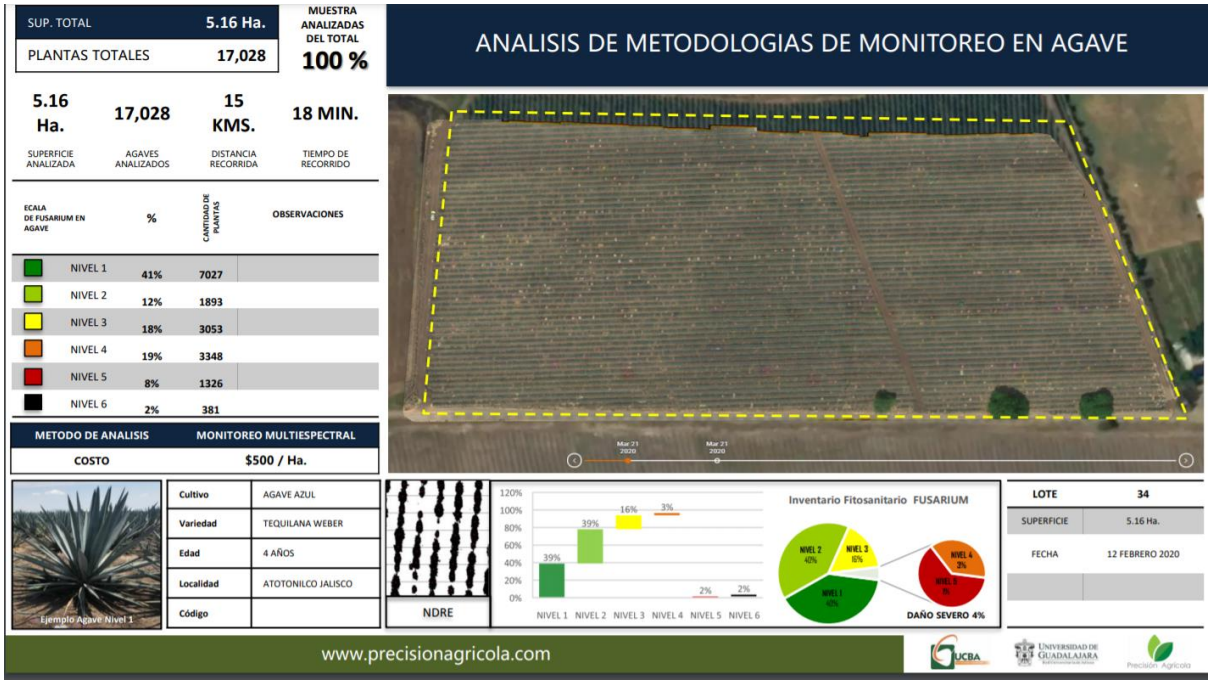
En México también se están trabajando diferentes investigaciones y proyectos usando las tecnologías antes mencionadas.

Específicamente con agave que es el cultivo por estudiar en este proyecto podemos encontrar un trabajo realizado con el analista de datos Eduardo Santos en colaboración con el Centro Universitario de Ciencias Biológicas y Agropecuarias (CUCBA) de la Universidad de Guadalajara.

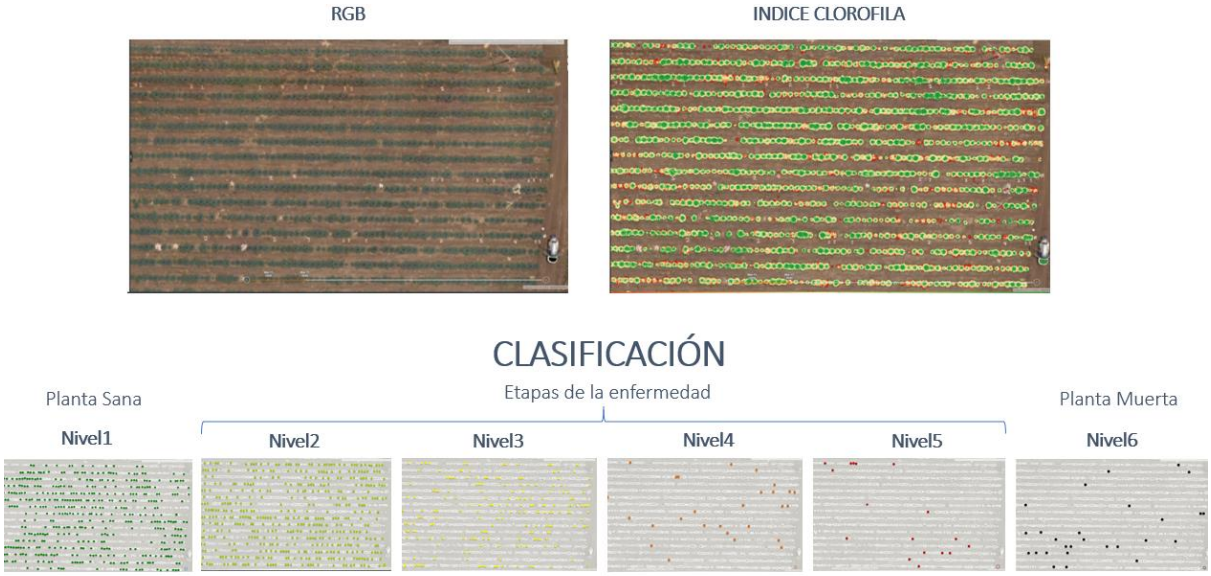
En este estudio ellos buscan clasificar la enfermedad Fusarium oxyporum el cual es un hongo con origen en el suelo, que causa marchitez y pudrición en diferentes especies de plantas. Se determinaron 6 estados posibles de la planta siendo 1 una planta sana y 6 una planta muerta. Del 2 al 4 son diferentes niveles de avance de la enfermedad.



La metodología de este proyecto fue la siguientes. Se contaba con una superficie total de 5.16 hectáreas, lo que equivalía a 17,028 plantas. Con un dron y una cámara multiespectral se realizó un vuelo de 18 donde se capturo imágenes de todo el terreno.



Usando índices vegetativos como el NDRE se logró hacer una contabilización de las plantas y usando el índice de clorofila, tomando un 8% de agaves etiquetados, se logró hacer un algoritmo de clasificación supervisado que clasifico de la siguiente forma los agaves:



La muestra que se utiliza de agaves etiquetados es muy pequeña debido a que se tiene que evaluar personalmente cada planta para asignarle una clasificación. Tomando en cuenta el tamaño de la superficie de siembra incrementar ese porcentaje es un poco complicado, pero basado en esa muestra la precisión del modelo oscilaba entre 90% y 93%.

6 RESULTADOS Y CONCLUSIONES

Encontrar Cercospora en etapa temprana en agaves mediante toma de imágenes multiespectrales es posible mediante la metodología que se diseñó e implemento. El preprocesamiento de las imágenes es la parte más importante del proceso siendo el índice vegetativo SR el mejor índice para detectar las manchas grises que provoca el hongo de la Cercospora Agavicola.

Aun cuando los agaves sean sometidos a diferentes tratamientos de bacterias y algas, en términos del desarrollo del hongo no hay evidencia que encuentre una diferencia significativa. Los árboles de decisión es el mejor modelo para poder clasificar las plantas de agave enfermas de las sanas. Los primeros síntomas del hongo son pequeñas manchas grises que van creciendo hacia dentro de la planta. Estos primeros síntomas se detectan 12 días aproximadamente después de que la planta es expuesta al hongo siendo el mayor riesgo que el hongo llegue a la piña ya que este daño seria irreversible.

7 TRABAJOS FUTUROS

Precisión Agrícola es una empresa que cuenta con equipo de cámaras y drones para su uso en la agricultura, actualmente se están apoyando la investigación y la ciencia de datos para poder brindarle un servicio de control y monitoreo a sus clientes productores. Ésta empresa fue la que nos brindó los datos que se utilizaron en este estudio y podrán usar lo aprendido para intentar encontrar Cercospora en Jalisco.

Para poderlo implementar como un proceso de control con cierta periodicidad a mayor escala se requieren mejores prácticas para la recolección de datos. La metodología de la captura de imágenes debe estar bien definida en términos de periodicidad, orden en la que se captura los datos, mejorar el etiquetado de las plantas (en caso de contar con uno) esto sobre todo es muy útil cuando se estudian diferentes tratamientos y se quiere llevar el control de cada tratamiento, la altura/distancia de la captura de imágenes y en caso de que los agaves no estén plantados (como lo fue en este estudio) definir el lugar donde se harán las tomas.

Además de esto es importante asegurarse que se toma la imagen de los paneles de calibración y que estos están completos dentro de la imagen para poder ajustar la reflectancia de cada toma.

La intención de probar diferentes tratamientos en agaves es encontrar alguna solución orgánica que ayude a prevenir las enfermedades o por lo menos genere defensas en los agaves para que sean menos propensas a adquirir el hongo.

Con base en lo encontrado en nuestro estudio, Precisión Agrícola podrá continuar con otros estudios de otros tratamientos para lograr el objetivo antes mencionado.

8 REFERENCIAS

Alvear, J. O. (2018, 11 16). *Arboles de decision y Random Forest.* Retrieved from https://bookdown.org/content/2031/ensambladores-random-forest-parte-i.html

Arcia Porrúa, J. (2020). De la agricultura precisa a la agricultura de precisión. *Revista Ingeniería Agrícola*.

Banco Mundial. (2021, 01 28). Retrieved from https://datos.bancomundial.org/indicador/NV.AGR.TOTL.ZS?locations=MX

Bannari, A. M. (1995). A review of vegetation indices. *Remote Sensing*, 95 — 120.

Boyd, S. D. (2016). *Machine Learning: Lasso Regression*. Retrieved from CVXPY: A Python-Embedded Modeling Language for Convex Optimization: https://www.cvxpy.org/examples/machine_learning/lasso_regression.html

Boyd, S. D. (2016). *Machine Learning: Ridge Regression*. Retrieved from CVXPY: A Python-Embedded Modeling Language for Convex Optimization: https://www.cvxpy.org/examples/machine_learning/ridge_regression.html

Boyd, S. D. (2016). *Support vector machine classifier with ℓ1-regularization*. Retrieved from CVXPY: A Python-Embedded Modeling Language for Convex Optimization: https://www.cvxpy.org/examples/machine_learning/svm.html

Food and Agriculture Organization of the United Nations. (2020, December 22). Retrieved from FAOSTAT: http://www.fao.org/faostat/en/#data/QC/

Garcia Mendoza, A. J. (2007). Los agaves de México. *CIENCIA 87*, 14-23.

Juan José Coria-Contreras, G. M.-A.-M. (2018). Applied regional epidemiology to inductive characterization and forecasting of blue agave gray spot (Cercospora agavicola) in Jalisco, Mexico. *Mexican Journal of Phytopathology*, 71-94.

Micasense. (2020, 09 2). Retrieved from Hablemos de calibración: https://micasense.com/es/hablemos-de-calibracion/

Secretaria de Economía. (2015, 09 14). *Gobierno de Mexico.* Retrieved from Denominaciones de Origen: https://www.gob.mx/se/articulos/denominaciones-de-origen-orgullodemexico