

PROCESAMIENTO DE IMÁGENES APLICADO A CULTIVOS DE CAÑA DE
AZÚCAR PARA DETECCIÓN DE MUERMO ROJO

Ing. NESTOR JAVIER ZAPATA SANCHEZ

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA INDUSTRIAL
MAESTRIA EN INGENIERIA INDUSTRIAL
TRABAJO DE GRADO
BOGOTÁ
2023

PROCESAMIENTO DE IMÁGENES APLICADO AL CULTIVO DE CAÑA DE
AZÚCAR PARA DETECCIÓN DE MUERMO ROJO

Ing. NESTOR JAVIER ZAPATA SANCHEZ

Tutor: Ing. Danilo Abril Hernández

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA INDUSTRIAL
MAESTRIA EN INGENIERIA INDUSTRIAL

TRABAJO DE GRADO

BOGOTÁ

2023

TABLA DE CONTENIDO

RESUMEN.....	7
ABSTRACT	8
1. EL PROBLEMA	8
1.1. PLANTEAMIENTO DEL PROBLEMA.....	8
1.2. FORMULACION DEL PROBLEMA.....	9
2. JUSTIFICACIÓN.....	9
3. OBJETIVOS	11
3.1. OBJETIVO GENERAL.....	11
3.2. OBJETIVOS ESPECÍFICOS.....	11
4. MARCO REFERENCIAL	11
4.1. ANTECEDENTES DE LA INVESTIGACIÓN.....	11
4.2. BASES TEÓRICAS	16
4.2.1. Enfermedades caña de azúcar	16
4.2.2. Proceso caña de azúcar	16
4.2.3. Procesamiento de imágenes	19
4.2.4. Regresión Logística	25
4.2.5. Geolocalización	26
4.3. DEFINICIÓN DE TÉRMINOS	27
5. DISEÑO METODOLÓGICO.....	29
5.1. NIVEL DE INVESTIGACIÓN	29
5.2. DISEÑO DE INVESTIGACIÓN.....	30
5.3. POBLACIÓN Y MUESTRA.....	31
5.4. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	31
5.5. TÉCNICAS DE PROCESAMIENTO Y ANALISIS DE DATOS	32
6. DESARROLLO PROYECTO	33
6.1. TOMA DE DATOS.....	33
6.2. PROCESAMIENTO DE IMÁGENES.....	35
6.3. REGRESIÓN LOGÍSTICA	36
6.4. GEOLOCALIZACIÓN	40

7. RESULTADOS	42
8. CONCLUSIONES Y RECOMENDACIONES	44
9. BIBLIOGRAFIA.....	45
10. ANEXOS	48

LISTADO DE GRÁFICOS

Gráfico 1. Exportaciones de panela colombiana (Fedepanela, 2023)	12
Gráfico 2. Sector panelero en cifras (FedePanela, 2021)	13
Gráfico 3. Documentos por año Scopus.....	14
Gráfico 4. Sistemas de coordenadas(Zhang, 2017)	21
Gráfico 5. Escala de colores, brillo de imagen (Zhang, 2017)	21
Gráfico 6. Procesamiento de imágenes binarias(Zhang, 2017).....	22
Gráfico 7. Categorías Ingeniería de imágenes (Zhang, 2017).....	22
Gráfico 8. Diagrama bloques del sistema (Zhang, 2017)	24
Gráfico 9. Ejemplo mapa libreria Folium (Folium,2023)	27
Gráfico 10. Sistema de Variables	28
Gráfico 11. Imágenes de Referencia Cultivo Puerto López (Meta)	33
Gráfico 12. Imágenes de Referencia Cultivo Maripí (Boyacá).....	33
Gráfico 13. Archivos Json.....	34
Gráfico 14. Estructura archivo Json	34
Gráfico 15. Imagen Original formato RBG.....	35
Gráfico 16. Imagen formato HSV	35
Gráfico 17. Imagen segmentación planta.....	36
Gráfico 18. Imagen segmentación muermo rojo.....	36
Gráfico 19. Diagrama de dispersión Planta vs Muermo Rojo	37
Gráfico 20. Matriz de confusión de sin balanceo	37
Gráfico 21. Gráfica de Correlación	38
Gráfico 22. Datos sin escalar	38
Gráfico 23. Datos escalados	39
Gráfico 24. Porcentaje de predicción sobre nuestros datos de test	39
Gráfico 25. Matriz de confusión.....	39
Gráfico 26. Imagen marcador.....	40
Gráfico 27. Visualización de marcadores	41
Gráfico 28. Diagrama de dispersión para determinar presencia focalizada	42
Gráfico 29. Afectación de variables.....	43
Gráfico 30. Límites mascara de Muermo Rojo	43
Gráfico 31. Límites mascara de planta.....	43
Gráfico 32. Mapa cultivos.....	44

LISTADO DE TABLAS

Tabla 1. Revisión literatura (Gomez-Camperos,2021).	15
Tabla 2. Diseño Investigación	30

RESUMEN

El propósito principal de este documento es presentar una propuesta para la detección del muermo rojo en cultivos de caña de azúcar, utilizando como muestra dos cultivos ubicados en diferentes zonas de Colombia. Este enfoque se basa en el procesamiento de imágenes, considerando la relevancia que tiene la exportación de panela para el país.

Es importante destacar que el procesamiento de imágenes empleado en este documento se basa en el formato HSV. A través de este formato, se realizan pruebas en los componentes de matiz, saturación y brillo. Las imágenes de prueba siguen un estándar y presentan similitudes en cuanto a las características mencionadas.

Además, se ha integrado un componente de regresión logística, el cual se basa en los parámetros derivados del procesamiento de imágenes. Este componente tiene como objetivo predecir la presencia de muermo rojo en las imágenes de muestra en donde se trabajó para evitar el sobreajuste del modelo, logrando una precisión del 77% en la predicción correcta de las imágenes que forman parte de nuestra base de pruebas.

Por último, a partir de los metadatos de cada imagen, se lleva a cabo la geolocalización de la imagen en un mapa donde se muestran los resultados de la detección de muermo rojo junto con las imágenes correspondientes.

Se presentan la imagen original, la imagen con presencia de muermo y la imagen con presencia de la planta. El objetivo es proporcionar información completa y evidencias visuales para respaldar la toma de decisiones.

Palabras Clave: Caña de Azúcar, Geolocalización, Muermo rojo, Procesamiento de imágenes, regresión logística.

ABSTRACT

The main purpose of this document is to present a proposal for detecting red rot in sugar cane crops, using two samples from different regions of Colombia. This approach is based on image processing, considering the relevance of sugar cane (panela) exportation for the country.

It is important to highlight that the image processing used in this document is based on the HSV format. Through this format, tests are conducted on the hue, saturation, and brightness components. The test images follow a standard and exhibit similarities in the mentioned characteristics.

Furthermore, a logistic regression component has been integrated, based on the parameters derived from image processing. This component aims to predict the presence of red rot in the sample images, working to avoid model overfitting and achieving a 77% accuracy in correctly predicting images within our test base.

Finally, based on the metadata of each image, geolocation of the image is carried out on a map, displaying the results of red rot detection along with the corresponding images. The original image, the image with red rot, and the image showing the plant's presence are presented. The objective is to provide comprehensive information and visual evidence to support decision-making.

Keywords: Sugarcane, Geolocation, Red glanders, Image processing, logistic regression.

1. EL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

Para el desarrollo de la propuesta de profundización basada en el énfasis de producción y tecnología se tuvo en cuenta el interés por el aumento de las exportaciones de panela en el país que son uno de los derivados principales de la caña de azúcar, ya que durante el año 2020 llegamos a un pico de 12.360 toneladas por un valor de \$17.672.053 de USD.

Actualmente los cultivos de caña de azúcar se ven afectado por algunas plagas entre estas el muermo rojo la cual buscamos identificar por medio del procesamiento de imágenes ya que se presenta la necesidad de realizar detección de una de las

enfermedades más difundidas en los cultivos de caña de azúcar la cual disminuye la producción al ocasionar la inversión de la sacarosa. (Potes, 2023)

También se debe destacar que el Ministerio de Comercio, Industria y Turismo con base en el informe de exportaciones del DANE destaca que los azúcares de caña representan el 32,4% de las exportaciones agropecuarias. (MINCIT,2022)

Y una de las metodologías es incorporar el término Agricultura 4.0, la cual integra un conjunto de tecnologías, dispositivos, protocolos y paradigmas computacionales para mejorar los procesos agrícolas. La información sobre las condiciones climáticas, suelos, enfermedades, insectos, semillas, fertilizantes, constituye una importante contribución al desarrollo económico y sostenible de este sector. Las técnicas de procesamiento digital de imágenes son una herramienta que permite la identificación temprana de las plagas o enfermedades en los cultivos como: cereales, frutales, raíces, hojas y tubérculos; y de esta forma, mitigar pérdidas económicas en el sector agrícola (Gomés,2021)

1.2. FORMULACION DEL PROBLEMA

Teniendo en cuenta la importancia que tiene para el país los cultivos de caña de azúcar, los empleos ofrecidos y la cantidad de hectáreas sembradas que se extienden por el país representando en la panela uno de sus derivados principales 1,2 millón de toneladas producidas por año destacamos la importancia de atacar una de las enfermedades presentes en el territorio nacional y que afectan los cultivos. (Fedepanela, 2021)

A partir de la anterior descripción del problema surge la siguiente pregunta de investigación:

¿Es posible realizar la detección y predicción de la enfermedad de muermo rojo en los cultivos de caña de azúcar mediante la aplicación de técnicas de procesamiento de imágenes en conjunto con un modelo de regresión logística?

2. JUSTIFICACIÓN

Este trabajo de grado tiene por objetivo realizar el procesamiento de imágenes para la caña de azúcar con el fin de establecer un parámetro que permita realizar la detección de la enfermedad y cuantificar dentro de la muestra el grado de afectación de muermo rojo.

La importancia del estudio se basa en establecer un parámetro de detección y medición que le pueda ofrecer al agricultor de caña la posibilidad de cuantificar y controlar dos de las enfermedades de acuerdo con la muestra procesada e infectada, y el tipo de materia prima que se está utilizando para la elaboración de la panela teniendo en cuenta que según proyecciones de Fedepanela las exportaciones de este producto han incrementado en un 300%. (Fedepanela, 2021)

Se tiene en cuenta también que las proyecciones del país en este sector para el año 2022 aumentarán hasta 18.000 toneladas de panela exportada (Quintero, 2018) y que parte de este objetivo tiene ver con la calidad de la materia prima utilizada ya que dentro del proceso las cañas muy jóvenes o “biches”, las muy viejas, las caídas y las dañadas por cualquier motivo, presentan una menor concentración de azúcares lo que afecta la calidad de la materia prima. (García & Bohórquez, 2000)

La contribución de mi trabajo de grado en modalidad de profundización se basa en otorgar la posibilidad de medir la afectación de la enfermedad a los agricultores que trabajan la caña de azúcar, para que estos puedan tomar medidas correctivas o preventivas en posteriores cultivos, con lo cual pueden mejorar la calidad y aspirar a entrar en diferentes mercados a nivel internacional teniendo en cuenta las diferentes posibilidades que se están abriendo a este producto como panela pulverizada y panelas especiales.

Cabe destacar que ya existe un estudio que trata el tema de procesamiento de imágenes desde la perspectiva aérea para la enfermedad de la caída de caña en el cual se procesaron imágenes multiespectrales para el análisis de texturas que permitieron la clasificación y cuantificación de la superficie afectada por el fenómeno del quebrado y caída de la caña. (Schneider, Solano, Kemerer, & Hadad,)

El estudio anterior difiere de la actual propuesta ya que se realizará en distintas etapas de la cosecha, justo antes de obtener el jugo de caña y sobre el tallo de la planta aplicando técnicas de procesamiento de imágenes por medio de filtrado por color y contornos.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Diseñar un método de detección de la enfermedad del muermo rojo en la caña de azúcar mediante técnicas de procesamiento de imágenes, predicción y geolocalización.

3.2. OBJETIVOS ESPECÍFICOS

- Recolectar muestras de la enfermedad de muermo rojo en cultivos de caña de azúcar.
- Realizar procesamiento de imágenes en los cultivos de caña de azúcar con el fin de realizar identificación de las características que contiene la enfermedad de muermo rojo y se pueden procesar mediante imágenes.
- Diseñar el algoritmo que permita aplicar la técnica de procesamiento de imágenes para la enfermedad y aplicar modelo de regresión logística con el fin de identificar la presencia en el cultivo.
- Aplicar modelo en imágenes de referencia y realizar análisis de presencia en cultivo por medio de geolocalización.

4. MARCO REFERENCIAL

4.1. ANTECEDENTES DE LA INVESTIGACIÓN

Inicialmente se debe realizar énfasis en la extracción de jugos de la caña de azúcar para la producción de panela mediante métodos de evaporación y horneado que permiten desarrollar un producto que tiene beneficios enmarcados en su valor nutricional y de salud en el cual Colombia es un referente mundial.

Colombia es el segundo productor de este producto a nivel global con 198.857 hectáreas sembradas, un área cosechada de 173.793 hectáreas, un promedio de rendimiento de 6.28 toneladas de panela por hectárea y una producción total de 1.091.502 toneladas de panela. (Fedepanela, 2020a)

El interés principal nace a partir del crecimiento de las exportaciones de panela definidas en el gráfico 1 de la Federación Nacional de productores de Panela (FEDEPANELA) en Colombia y los requerimientos que realiza el gobierno nacional dentro de la legislación vigente mediante la resolución número 000779 de 2006 por la cual se establece el reglamento técnico sobre los requisitos sanitarios que se deben cumplir en la producción y comercialización de la panela para consumo humano y se dictan otras disposiciones.

De acuerdo con la normativa exigida por el gobierno nacional la Federación Nacional de Productores de Panela desarrolló informes en los que destaca la importancia de la calidad y cumplimiento de BPM (Buena prácticas de Manufactura) en la producción de panela. (Fedepanela, 2020b)



Gráfico 1. Exportaciones de panela colombiana (Fedepanela, 2023)

A partir de la necesidad de cumplir con los reglamentos estipulados por el gobierno nacional y la importancia que destaca la Federación Nacional de paneleros se empieza a evaluar la calidad como un factor decisivo en la producción de panela (Fedepanela, 2020); por lo cual se desarrollan diferentes lineamientos asociados a buenas prácticas agrícolas y de manufactura dentro del sector panelero. (Osorio, 2007)

Cabe destacar que el sector panelero actualmente tiene unas características definidas de acuerdo con la distribución de tierras que los productores de caña realizan para la materia prima de la panela, la cantidad de departamentos que lo producen, la cantidad de empleos que se genera, la población que consume panela y la producción anual que se visualiza en el siguiente gráfico:

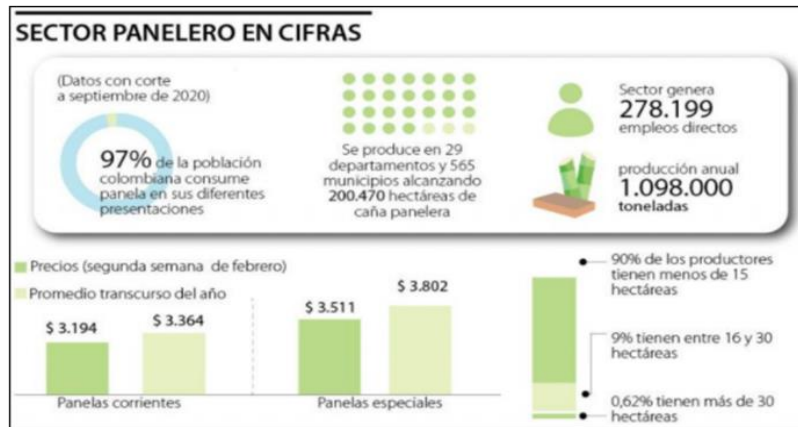


Gráfico 2. Sector panelero en cifras (Fedepanela, 2021)

Dentro de los cultivos de caña se encuentra la enfermedad de muermo rojo la cual genera pudrición en la planta y se extiende de acuerdo a las condiciones de humedad presentes en el entorno por lo cual surge como una nueva tendencia para evaluar la calidad de la planta el procesamiento de imágenes como una alternativa adecuada, eficiente y confiable para la detección de enfermedades con la ayuda de imágenes de hojas de plantas (Kumar & Raghavendra, 2019).

Mediante las imágenes podemos identificar y clasificar realizando un diagnóstico temprano de enfermedades mediante el procesamiento de imágenes digitales donde finalmente esto puede ayudar a los agricultores a analizar el tipo de enfermedad y tomar decisiones tempranas al respecto. (Ahmad Supian, Madzin, & Albahari, 2019)

A partir de la anterior introducción inicial se definió la ecuación de búsqueda (Anexo A) en tres bases de datos en donde se obtuvieron los siguientes resultados: en la base de datos Scopus, IEEE, Ebscohost inicialmente no se obtuvieron resultados por la ecuación de búsqueda completa por lo cual se dividió la ecuación de búsqueda en dos:

- (“image processing” OR “image digitalization”) AND (“diseas*” OR “mushroom*”)
- (“sugar cane” OR “sugar cane industr*” OR “sugarmill”) AND (“panela process”)

Arrojando en Scopus para la primera ecuación 97 documento los cuales se detallan de acuerdo con su año de publicación (gráfico 3), y para la segunda ecuación un total de 3 documentos.

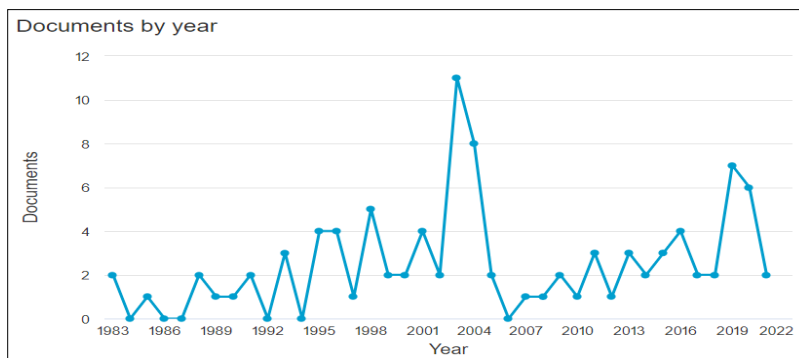


Gráfico 3. Documentos por año Scopus

En la IEEE a partir de la primera ecuación de búsqueda se obtuvieron 18.625 resultados entre los cuales se destacan artículos como “Identificación de enfermedades de las plantas mediante la técnica de procesamiento de imágenes” (Devaraj, Rathan, Jaahnavi Sarvepalli, & Indira, 2019) y “Técnica de procesamiento de imágenes para la detección automática de enfermedades de las plantas y sistema de alerta en granjas agrícolas” (Pradeep Kumar Mugithe, Rohit Varma Mudunuri, B Rajasekar, & S Karthikeyan, 2020) y finalmente en la búsqueda realizada en Ebscohost se obtuvieron 260 resultados.

En contexto se encontró que el procesamiento de imágenes para la detección de enfermedades asociadas a las plantas se ha utilizado en procesos agrícolas de diferentes cultivos como el de espinaca en el cual el estudio se basó en examinar las plagas de los cultivos de hortalizas que se ven afectados por distintas plagas y enfermedades afectando el color, tamaño y la forma de la planta con lo cual se realizó una segmentación, extracción de características y reconocimiento para cada imagen con el fin de poder analizar si la planta estaba afectada.(Padilla Jiménez, Rojas, & Guzmán, 2016)

Así mismo se han realizado estudios que han destacado la importancia de la identificación de enfermedades de las plantas mediante las técnicas de procesamiento de imágenes ya que gran parte de las enfermedades se observan en las hojas de la planta, los frutos y los tallos; y debido a la dificultad de los patrones visuales la cuantificación exacta de esas enfermedades, plagas o rasgos diagnosticado visualmente se hace compleja. (Devaraj et al., 2019)

De acuerdo con lo anterior las técnicas de procesamiento de imágenes para la detección de enfermedades de plantas se realiza teniendo en cuenta que a partir de una imagen se realiza una segmentación por grupos para extraer las características de la enfermedad buscando detectarla con el fin de alertar a agricultor y que este

pueda evitar la propagación de la enfermedad en el cultivo. (Pradeep Kumar Mugithe et al., 2020)

En la revisión de antecedentes también se verificó que se han realizado estudios para la detección de enfermedades de manchas foliares en los cultivos de algodón por medio del procesamiento de imágenes en donde se destaca que la identificación de las enfermedades del algodón en las primeras etapas del diagnóstico es importante ya que para la época se evidenciaba una tremenda pérdida en la calidad y cantidad de la producción de algodón debido a diversas enfermedades que afectan a la planta. (Bhimte & Thool, 2018)

A continuación, presentamos estudios preliminares sobre procesamiento de imágenes en diversos cultivos:

Técnicas de Procesamiento	Precisión
Cultivo: Trigo Técnicas: Normalización de la constancia del color, segmentación SLIC automática y recorte manual de la región de interés (RoI) mediante la pantalla táctil del dispositivo móvil y el algoritmo Chan-Vese	78%
Cultivo: Arroz Técnicas: Intersección de histogramas, agrupación de K-medias basadas en umbrales.	La proporción de probabilidad de enfermedades de mancha marrón y escaldadura de hojas fue de 25 y 50%, respectivamente.
Cultivos: Manzana, Maíz, Uva, Patata, Tomate, Tabaco Técnicas: Algoritmo de segmentación automática de imágenes y aprendizaje profundo con un conjunto de datos ampliado.	Manzana: 83,3% Maíz: 87,2% Uva: 86,8% Patata: 81,9% Tomate: 87,3% Tabaco: 82,5%
Cultivos: Frijoles y Patatas Técnicas: Clasificadores neuronales RSC (Sub-space Random Classifier) y LIRA (Limited Received Area)	LIRA, una tasa de reconocimiento del 88%. Con CSR, o 89%
Cultivos: Maíz Técnica: Redes Neuronales convolucionales.	88,4%
Cultivo: Plátano Técnica: Coeficiente de correlación	98,6%
Cultivo: Cebolla Técnica: Redes neuronales	87,2%
Cultivo: Manzana, Plátano, Repollo, Yuca, Apio, Maíz, Pepino, Uva Técnica: Redes neuronales convolucionales	99,5%

Tabla 1. Revisión literatura (Gomez-Camperos,2021).

4.2. BASES TEÓRICAS

4.2.1. Enfermedades caña de azúcar

Algunas de las enfermedades que afectan este cultivo han ocasionado grandes pérdidas a la producción en todo el mundo y motivado intensos estudios y controversias. Las enfermedades revisten gran importancia económica y quizás son las que menos atención han recibido por parte de los cultivadores de caña en el país. (Osorio, 2007)

Se han logrado establecer diferentes tipos de afecciones asociadas con las distintas etapas del ciclo vegetativo de la caña de azúcar:

- Muermo o pudrición rojos (*Physalospora tucumanensis*)

Se caracteriza por presentar pudriciones rojizas en el interior de la caña (gráfico 4), las cuales se extienden de acuerdo con las condiciones de humedad. Esta afección se encuentra distribuida en la mayoría de los países productores de caña de azúcar. (Regional Siete de Corpoica y el SENA Regional Santander, 1998)



Gráfico 1. Muermo o pudrición roja (Osorio, 2007)

4.2.2. Proceso caña de azúcar

En el proceso de tratamiento de caña de azúcar se debe tener en cuenta las siguientes etapas:

- Labores de adecuación del suelo

En la cual intervienen la edad y las condiciones físicas (suelo, clima y luminosidad) en que se desarrolla el cultivo, cumplen una función fundamental en la producción de tallos y en la concentración de los azúcares. (Osorio, 2007)

En donde intervienen varias etapas, destaca la adecuación y preparación del suelo, cuya importancia radica en la contribución en el incremento de la productividad, mejorando el sistema de producción del cultivo. La adecuación del terreno abarca diversas labores como la planificación de los lotes de caña, definición de sus dimensiones, construcción de canales de riego, drenajes (en caso de suelos inundables) y caminos para movilizar la caña cortada y preparación el suelo para la siembra. (García & Bohórquez, 2000)

- Selección de semillas

La selección de semillas según García & Bohórquez se realiza mediante la propagación de tallos o estacas que reciben el nombre de "semilla" y que normalmente tienen entre 25 o 30 cm de longitud. Al seleccionar un material para semilla debe tenerse en cuenta que reúna las siguientes características:

- Este libre de plagas y enfermedades
- Tener un estado nutricional adecuado
- Tener la edad de corte y el tamaño recomendados
- Sea una semilla pura (libre de mezclas con otras variedades)
- Yemas funcionales(García & Bohórquez, 2000)

La semilla para siembra puede obtenerse del cogollo, de bretones de plantaciones maduras, de plantilla o primeras socas y de semilleros. (García & Bohórquez, 2000)

- Sistema de siembra

El sistema de siembra se debe realizar al inicio de la temporada de lluvias especialmente cuando no se dispone de riego o la pendiente del terreno no permite instalar diseños funcionales de sistemas de riego.(García & Bohórquez, 2000)

Finalmente, es importante tener en cuenta en cuanto a los métodos y distancias de siembra lo siguiente:

- Para cultivos mecanizados y en zonas planas, la distancia de siembra varía entre 1,30 y 1,50 m.
- En los sistemas a chorrillo y mateado, para cultivos en suelos de ladera, la distancia entre surcos de 1,20 m ofrece los mejores resultados.

- El mateado, con distancias entre 1,0 a 1,30 m entre surcos y con distancias entre plantas de 0,25 a 0,50 m entre mates, en suelos con buena estructura, permite obtener rendimientos muy similares a los del sistema en chorrillo.
- La falta de agua en la época de siembra, la desigualdad en el crecimiento de los tallos (primarios, secundarios y terciarios) y el mayor riesgo de vuelco, hacen que el sistema de mateado sea menos eficaz que el chorrillo.
- En zonas de ladera y con inadecuada preparación de suelos, en la medida en que la distancia de siembra es mayor que 1,20 m se reducen los rendimientos, debido ante todo a la mayor competencia de malezas con el cultivo, por el mayor acceso de estas a la luz. En función de esta última variable podría establecerse que entre mayor sea la pendiente menor debería ser la distancia de siembra.
- Cabe recordar que los rendimientos están directamente relacionados con una buena densidad de población. En suelos de ladera, no se puede cuidar esa población con una óptima densidad de yemas por metro lineal, debido a los obstáculos (como piedras, árboles, etc.) que impiden una óptima densidad. Esta restricción física se disminuye, aumentando la población mediante la disminución de distancias entre surcos. (García & Bohórquez, 2000)

- Control de malezas

Finalizada la siembra se debe ejercer un control sobre malezas las cuales pueden disminuir los rendimientos hasta en un 60%, el control que sobre ellas se ejerce en las zonas paneleras es, por lo general, deficiente. Así mismo, se ha determinado que el periodo más crítico de competencia por agua, luz y nutrientes entre las malezas y el cultivo ocurre en la etapa de macollamiento. (García & Bohórquez, 2000)

- Maduración y cosecha

El momento de cosecha debería establecerse cuando se alcance el punto de máximo rendimiento, el cual debería coincidir con el punto de madurez. Algunos productores, de acuerdo con su experiencia, establecen este punto por la edad de la caña, desde la siembra o la cosecha anterior y por el cambio de apariencia (color de los tallos, reducción de la longitud de los entrenudos y del tamaño de las hojas), entre otras variables. (García & Bohórquez, 2000)

4.2.3. Procesamiento de imágenes

A partir de que el muermo rojo contiene características distinguibles en el cultivo de caña de azúcar que se pueden apreciar durante el corte de la planta debido al color que las hace distinguibles.

El procesamiento de imágenes surge como una alternativa para tratar imágenes con el fin de obtener información (Zhang, 2017) en este caso a partir del color que contiene cada una de las plantas que posteriormente se va a utilizar en el proceso de elaboración de la panela.

A partir de la anterior idea el procesamiento de imágenes se traduce en un sentido más amplio a un término adecuado utilizado como es la “Ingeniería de imágenes” ya que comprende toda una disciplina relacionada en donde se deben revisar conceptos tales como:

- La imagen y su expresión

Las imágenes son un medio importante por el cual los seres humanos observan la mayoría de la información que reciben del mundo real las cuales se pueden obtener utilizando diferentes sistemas de observación y captura del mundo real en diversas formas y maneras en donde pueden actuar directa y/o indirectamente sobre los ojos humanos y producir percepción visual. (Zhang, 2017)

El mundo real es tridimensional (3-D) en el espacio sin embargo las imágenes obtenidas son generalmente bidimensionales (2-D) para $f(x, y)$, donde cada una de las coordenadas son espaciales y f es la amplitud en cualquier par de coordenadas (x, y) que representan ciertas propiedades de la escena real proyectada en la imagen en ese punto en particular. (Zhang, 2017)

En donde matemáticamente, una imagen se puede explicar cómo una función $f(x, y)$ con dos variables. Para el procesamiento por computadora, una imagen analógica $f(x, y)$ debe digitalizarse a una imagen digital $l(r, c)$, en la que r (fila) y c (columna) se refieren a la posición discreta de cualquier punto en la imagen digital y la amplitud l se refiere a la discreta magnitud de la imagen en el punto (r, c) . (Zhang, 2017)

En donde $f(x, y)$ se usa para representar una imagen digital en cualquier punto (x, y) y f toma un número entero valores solamente. Una imagen digital puede considerarse una matriz cuyas filas y columnas se refieren a la posición de cualquier

punto de la imagen y el valor de matriz correspondiente se refiere a la intensidad en ese punto. (Zhang, 2017)

La digitalización de imágenes utiliza un proceso de muestreo para extraer de la imagen un conjunto discreto de números reales, las muestras se cuantifican en valores de nivel de gris discretos, que a menudo se consideran igualmente espaciados. (Zhang, 2017)

El resultado del muestreo y la cuantificación es una imagen digital definida mediante una matriz rectangular de valores enteros en donde cada elemento se denomina elemento de imagen a menudo abreviado como "Pixel". (Zhang, 2017)

Una imagen 2 – D de $M * N$ (donde M y N) son el número total de filas y el total número de columnas de la imagen) se puede representar mediante una matriz 2-D en $f(x, y)$ o mediante una matriz $f(2 - D)$ cómo se relaciona a continuación según Zhang:

$$f = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1N} \\ f_{21} & f_{22} & \dots & f_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1} & f_{M2} & \dots & f_{MN} \end{bmatrix}$$

La representación matricial anterior también se puede convertir en una representación vectorial. Para ejemplo, la ecuación anterior se puede reescribir como:

$$F = [f_1 f_2 \dots f_N]$$

Donde:

$$f_i = [f_{11} f_{21} \dots f_{M1}]^T \quad i = 1, 2, \dots, N$$

Tenga en cuenta que la operación de matriz y las operaciones de matriz son diferentes. Considerando dos 2×2 imágenes $f(x, y)$ y $g(x, y)$ como un ejemplo, el resultado del producto es:

$$f(x, y)g(x, y) = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} f_{11}g_{11} & f_{12}g_{12} \\ f_{21}g_{21} & f_{22}g_{22} \end{bmatrix}$$

Y su producto de matriz es:

$$FG = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} f_{11}g_{11} + f_{12}g_{21} & f_{12}g_{12} + f_{12}g_{22} \\ f_{21}g_{11} + f_{22}g_{21} & f_{21}g_{12} + f_{22}g_{22} \end{bmatrix}$$

La visualización de la imagen significa poner la imagen en una forma visible que se muestra en la pantalla o equipo. En un equipo de visualización normal, una imagen se muestra como un patrón $2 - D$ con diferentes tonos o colores. (Zhang, 2017)

Un ejemplo se muestra a continuación (gráfico 11) en el cual para la visualización de la imagen se toma un sistema de coordenadas en el cual el eje de la fila va de izquierda a derecha y el eje de la columna de arriba hacia abajo (a) o toma las coordenadas desde la esquina inferior izquierda como origen del sistema (b).(Zhang, 2017)



Gráfico 4. Sistemas de coordenadas (Zhang, 2017)

A partir del sistema de coordenada se puede realizar la manipulación de la imagen, siguiendo el caso anterior a partir de la escala de grises en las imágenes representadas por el brillo de la imagen en donde los niveles más altos representan pixeles más brillantes mientras que los pixeles más bajos representan pixeles más oscuros como se muestra a continuación:

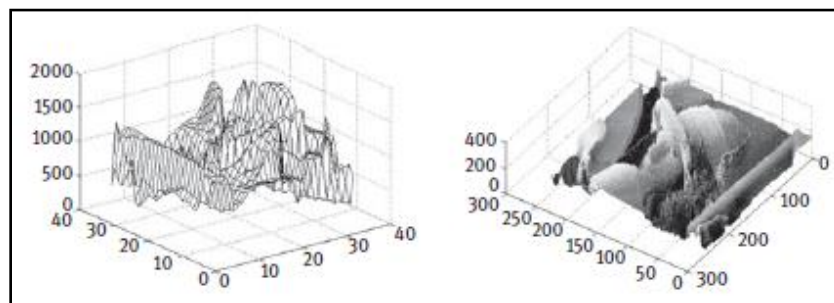


Gráfico 5. Escala de colores, brillo de imagen (Zhang, 2017)

Así mismo otro tratamiento que se puede realizar a las imágenes es a través de una imagen binaria de acuerdo con el (gráfico 14) en el que el modelo matemático para la representación de imágenes el área del píxel está representada por su centro en

la cual forma un conjunto de puntos (a) de la figura (b) mediante la representación matricial de la figura (c). (Zhang, 2017)

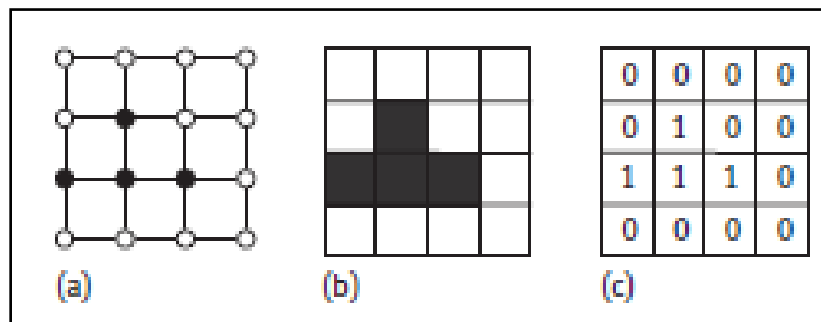


Gráfico 6. Procesamiento de imágenes binarias (Zhang, 2017)

- Ingeniería de imágenes

La ingeniería de imagen es un tema amplio que abarca y se basa en temas tales como matemáticas, física, biología, fisiología, psicología, ingeniería eléctrica, informática y automatización.(Zhang, 1996) Sus avances también están estrechamente relacionados con el desarrollo de telecomunicaciones, ingeniería biomédica, teledetección, documento de procesamiento, aplicaciones industriales, etc.(Zhang, 1996)

A partir de lo anterior la imagen actual de la ingeniería de imágenes desde una perspectiva orientada a las técnicas de imágenes relacionadas según Zhang se divide en tres categorías:

- Procesamiento de Imágenes (IP)
- Análisis de imágenes (IA)
- Comprensión de imágenes (IU)

Estas categorías se encuentran estrechamente relacionadas como se muestra a continuación:

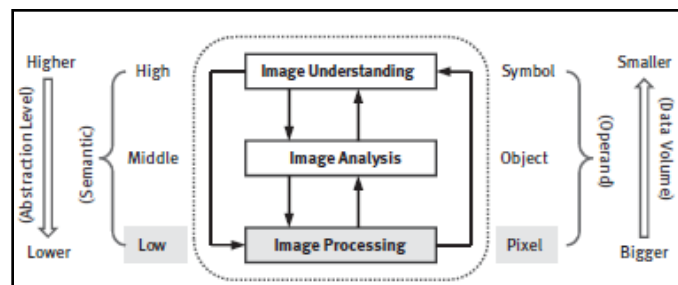


Gráfico 7. Categorías Ingeniería de imágenes (Zhang, 2017)

A partir de la anterior imagen el procesamiento de imágenes incluye principalmente la adquisición, representación, comprensión, mejora, restauración y reconstrucción de imágenes; aunque el procesamiento de imágenes se refiere a la manipulación de una imagen para producir otra imagen (mejorada), el análisis de imágenes se refiere a la extracción de información de una imagen.

Basado en el análisis de imágenes, la comprensión de imágenes se refiere a un conjunto de conocimientos utilizado para transformar estos datos extraídos en ciertas descripciones comúnmente entendidas, tomar decisiones y acciones posteriores de acuerdo con la interpretación realizada. (Zhang, 2017)

De la misma forma Zhang define operaciones computarizadas dentro de cada categoría:

- Procesos de nivel bajo en el cual se implican operaciones como el procesamiento de imágenes para reducir el ruido, el contraste, la mejora y nitidez de la imagen. Este proceso se caracteriza por que tanto su entrada como su salida son imágenes.
 - Procesos de nivel medio implica tareas como segmentación, descripción de objetos para reducirlos a una forma adecuada para procesamiento de computadora y clasificación de objetos visuales. Este proceso se caracteriza por que sus entradas son generalmente imágenes, pero sus salidas zona tributos extraídos de esas imágenes.
 - Procesos de nivel superior implica “tener sentido” de un conjunto de objetos reconocidos, como en el análisis de imágenes. (Zhang, 2017)
- Módulos y sistemas para el procesamiento de imágenes

Se pueden construir sistemas de procesamiento de imágenes combinando técnicas para resolver problemas prácticos en los que los módulos integrados tienen una función específica, adquisición (imágenes), síntesis (generación de imágenes), procesamiento, visualización, impresión, comunicaciones y almacenamiento. La entrada del sistema incluye la adquisición y generación de imágenes, mientras que la salida del sistema incluye tanto la visualización como la impresión. (Zhang, 2017)

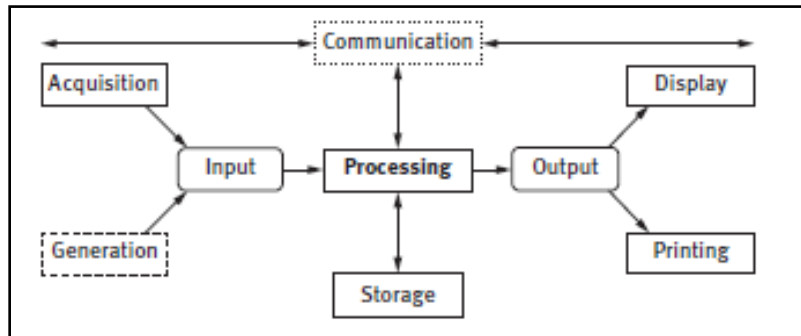


Gráfico 8. Diagrama bloques del sistema (Zhang, 2017)

- La adquisición de imágenes la cual se realiza con cierto equipo conocido como imagen equipo, que transforma la escena del objetivo en una imagen que podría procesarse por computadora.

Para el sistema de procesamiento el propósito de ambos es obtener la imagen discreta que se puede poner en computadora para su posterior procesamiento. (Zhang, 2017)

- Visualización e impresión de imágenes, usada principalmente en pantalla para visualizar a humanos. La visualización de la imagen es un paso importante entre la comunicación entre el sistema y los usuarios.

Los dispositivos de visualización comúnmente utilizados en los sistemas de procesamiento de imágenes incluyen tubos de rayos catódicos (CRT), monitores de pantalla de cristal líquido (LCD) y televisores.

Las imágenes que ingresan al monitor también se pueden convertir en copias impresas de diapositivas, fotografías o transparencias. Además del monitor, varios dispositivos de impresión, como una variedad de impresoras, también se pueden considerar como un dispositivo de visualización de imágenes. (Zhang, 2017)

- Almacenamiento de imágenes, que requiere memoria para almacenar la imagen de un formato determinado. (Zhang, 2017)
- Procesamiento de imágenes para el cual se puede utilizar una serie de diferentes tecnologías en donde el propósito principal y la tecnología para el procesamiento de imágenes incluyen aspectos como mejorar la calidad visual de la imagen, la restauración de degradado, eliminar los efectos de varios tipos de interferencia, la reconstrucción de escenas basada en

diferentes proyecciones, la codificación de la imagen para reducir la cantidad de datos de representación para facilitar el almacenamiento y la transmisión, la marca de agua de la imagen para proteger la propiedad de las imágenes y la extensión de estas tecnologías para procesar imágenes en color, video o imágenes de múltiples escalas.(Zhang, 2017)

En general, se pueden describir procedimientos de procesamiento de imágenes con algoritmos disponibles implementados en software de las cuales muchas tareas de procesamiento de imágenes se pueden completar con solo una computadora ordinaria o en muchas aplicaciones en línea, para mejorar el procesamiento, velocidad o superar las limitaciones de la computadora, y se puede tomar con hardware especial o dedicado. (Zhang, 2017)

- Procesamiento de imágenes con OPENCV

OpenCV es una biblioteca nativa y multiplataforma para visión por computadora, aprendizaje automático y procesamiento de imágenes que se adapta en Python para el desarrollo de aplicaciones para procesar datos visuales; esto puede incluir datos de transmisión en vivo, como fotografías o videos, desde un dispositivo como una cámara.

Los proyectos que se pueden desarrollar a través de OpenCV van desde manipulación de imágenes, realidad aumentada, seguimiento de objetos, reconstrucción de escenas en 3D, aprendizaje estadístico y categorización de objetos. (Beyer, 2015)

Teniendo en cuenta la anterior descripción el lenguaje de programación python va a ser el utilizado para diseñar por medio de la biblioteca de OpenCv el algoritmo para hacer los diferentes filtros y segmentaciones a cada una de las imágenes de muestra.

4.2.4. Regresión Logística

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica en este caso muermo rojo en la planta o resultado según los valores de un conjunto de predictores que son la cantidad de pixeles en cada una de las imágenes a partir de un análisis de datos que utiliza las matemáticas para encontrar las relaciones entre dos factores de datos.

Los coeficientes de regresión logística pueden utilizarse para estimar la razón de probabilidad de cada variable independiente del modelo en donde también se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante.

Se debe tener en cuenta que la variable dependiente debe ser categóricas en este caso 0 o 1 por lo que es una regresión logística binaria; el procedimiento de regresión logística genera todas las predicciones, residuos, estadísticos de influencia y pruebas de bondad de ajuste utilizando los datos a nivel de los casos. (IBM,2023)

Aunque la función logística calcula un rango de valores entre 0 y 1, el modelo de regresión binaria redondea la respuesta a los valores más cercanos. Por lo general, las respuestas por debajo de 0,5 se redondean a 0 y las respuestas por encima de 0,5 se redondean a 1, de modo que la función logística devuelve un resultado binario.

- Procedimiento de regresión logística

Inicialmente se debe plantear una pregunta concreta que en este trabajo de grado sería:

“¿La cantidad de pixeles de muermo rojo y/o planta segmentados en la imagen de referencia afectan la detección de muermo rojo en cada imagen?”

Seguidamente debemos recopilar datos históricos de los dos factores que intervienen:

- Variables independientes: Cantidad de pixeles con una característica
- Variable dependiente: Cantidad de imágenes con presencia de muermo rojo.

Finalmente debemos entrenar el modelo de regresión logística por medio de la librería scikit-learn en python la cual nos otorga el método de “LogisticRegression”. (Scikit-learn,2023)

4.2.5. Geolocalización

En este caso para descripción visual de los resultados de procesamiento de imágenes y de regresión logística se utilizará la georreferenciación por medio de la

librería Folium en python la cual facilita la visualización de datos manipulados en Python en un mapa interactivo.

Donde se vincularon los datos, marcadores e imágenes que mostrara los resultados finales para cada uno de los cultivos. (Folium,2023)

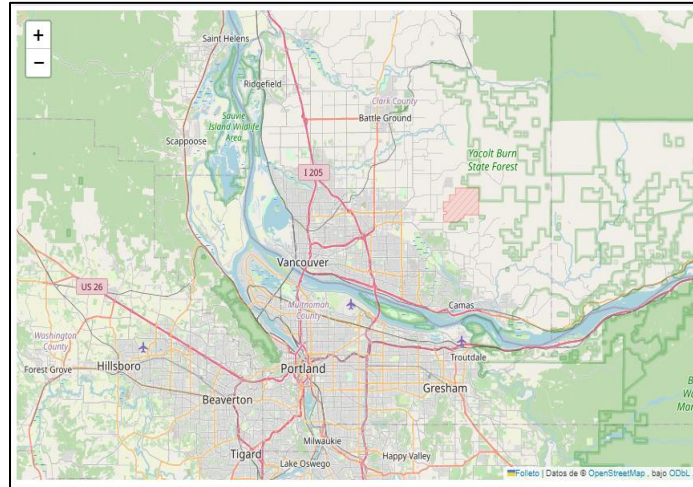


Gráfico 9. Ejemplo mapa librería Folium (Folium,2023)

4.3. DEFINICIÓN DE TÉRMINOS

Archivo JSON: JavaScript Object Notation es un formato de datos estructurado que a menudo se utiliza para hacer intercambios de información la cual o es relacional por si sola. (Oracle,2023)

BPA: Buenas Prácticas Agrícolas; conjunto de prácticas que buscan garantizar la inocuidad de los productos agrícolas, la protección del ambiente, la seguridad y el bienestar de los trabajadores y la sanidad agropecuaria. (Osorio, 2007)

BPM: Buenas Prácticas de Manufactura; principios básicos y prácticas generales de higiene en la manipulación, preparación, elaboración, envasado, almacenamiento, transporte y distribución de alimentos para consumo humano. (Osorio, 2007)

CIMPA: Centro de investigación para el mejoramiento de la panela.

Fungicidas: Cualquier sustancia que se usa para destruir hongos (organismos parecidos a las plantas que no elaboran clorofila), como la levadura y el moho. (Instituto Nacional del Cáncer, s.f)

Macollamiento: Etapa del cultivo de la caña, en la cual se incrementa el número de brotes o tallos después de la germinación o corte. (Osorio, 2007)

Mateado: Sistema de siembra de caña. (Osorio, 2007)

Metadatos: Información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen la conceptualización, calidad, generación, cálculo y características de un conjunto de datos estadísticos (NTCPE 1000:2020).

Panela: Producto obtenido de la extracción y evaporación de los jugos de la caña de azúcar, elaborado en los establecimientos denominados trapiches paneleros o en las centrales de acopio de mieles vírgenes, en cualquiera de sus formas y presentaciones. (Betancourt, 2006)

Sacarosa: Compuesta por glucosa y fructosa, es el edulcorante más conocido y distribuido en la naturaleza, se encuentra en la caña de azúcar, la remolacha y el jarabe de maple. En la industria se extrae y purifica para obtener azúcar morena, blanca o panela. (Minsalud,2019)

Trapiche panelero: Establecimiento donde se extrae y evapora el jugo de la caña de azúcar y se elabora la panela. (Betancourt, 2006)

4.4. SISTEMA DE VARIABLES

Para nuestro sistema de variables tenemos en cuenta como variables independientes la cantidad de pixeles en cada imagen que hacen parte de muermo rojo y de la planta; y como variable dependiente la predicción de la presencia de muermo rojo en cada imagen, como se describe en el siguiente diagrama:

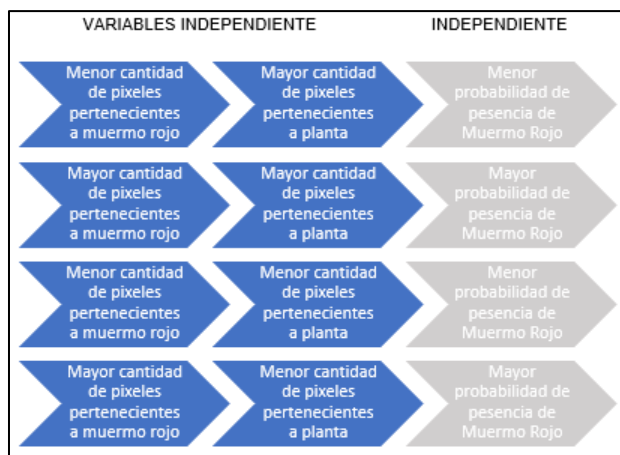


Gráfico 10. Sistema de Variables

5. DISEÑO METODOLÓGICO

5.1. NIVEL DE INVESTIGACIÓN

El nivel de investigación corresponde a descriptivo porque se busca analizar el grado de afectación de un tipo de enfermedades presentes en los cultivos de caña de azúcar de un determinado espacio geográfico ubicado en el municipio de Pachaquiario (Meta) y Maripí (Boyacá) teniendo en cuenta que son regiones que nos proporcionan la facilidad de contar con los cultivos de caña de azúcar y los accesos a la toma de imágenes.

Teniendo en cuenta lo anterior el propósito principal es medir el grado de relación que existe entre cada una de las enfermedades y la afectación de la planta; en donde la variable independiente en este estudio se constituye como las plantas del cultivo y la variable dependiente como el tipo de enfermedad distinguible en la planta.

Se destaca que el proyecto utiliza elementos cuantitativos ya que busca expresar relaciones entre cada una de las variables definidas y hacer utilización de técnicas estadísticas con el fin de analizar la información proporcionada por el procesamiento de imágenes que nos permite definir la presencia y/o ausencia de la enfermedad en la planta, su distribución dependiendo la muestra y el grado de afectación en términos porcentuales.

El trabajo de campo y el análisis de datos hacen parte fundamental de la investigación ya que se deben tomar las imágenes correspondientes al proceso directamente desde el cultivo que contemplen la calidad necesaria para su posterior procesamiento y se deben analizar estas para determinar el grado de afectación en la muestra.

5.2. DISEÑO DE INVESTIGACIÓN

Objetivos específicos	Secuencia de actividades	Mecanismo de obtención de información	Espacios académicos utilizados y/o temas relacionados
1. Recolectar muestras de la enfermedad de muermo rojo en cultivos de caña de azúcar	1.Toma de imágenes con características optimas en cultivos de caña de azúcar. 2.Identificación y análisis de características de color y forma que se presenta en los cultivos de estudio.	Información otorgada por el ICA (Instituto Colombiano Agropecuario) Muestra en campo.	Metodología de la investigación.
2. Realizar procesamiento de imágenes en los cultivos de caña de azúcar con el fin de realizar identificación de las características que contiene la enfermedad de muermo rojo y se pueden procesar mediante imágenes.	1.Generar código en Python utilizando librería OpenCv para identificar características en la planta que pueden definir la enfermedad por medio del color. 2.Análisis descriptivo de los cultivos y las imágenes a partir de Python y R-	Imágenes de la muestra	Metodología de la investigación Estadística Procesamiento de imágenes
3. Diseñar el algoritmo que permita aplicar la técnica de procesamiento de imágenes para la enfermedad y aplicar modelo de regresión logística con el fin de identificar la presencia en el cultivo.	1. Procesamiento de imágenes con el lenguaje de programación Python utilizando la librería Open CV. 2.Establecer parámetros de color para la enfermedad. 3.Regresión logística por medio de librería scikit-learn.	Imágenes de la muestra Python	Procesamiento de imágenes Gestión de Datos Métodos y aplicaciones de analítica
4. Aplicar modelo en imágenes de referencia y realizar análisis de presencia en cultivo por medio de geolocalización.	1.Análisis de resultados de muestras iniciales. 2.Construcción de herramienta visual por medio de geolocalización utilizando la librería Folium.	Información otorgada por los metadatos y el procesamiento de las imágenes, y por el resultado de la regresión logística	Estadística Métodos y aplicaciones de analítica

Tabla 2. Diseño Investigación

5.3. POBLACIÓN Y MUESTRA

La población base que se toma para el estudio comprende fincas ubicadas en el departamento de Meta en el municipio de Puerto López y en Boyacá en zona rural del municipio de Maripí debido a que se cuenta con los cultivos de caña de azúcar de forma constante durante todas las épocas del año, así como con los permisos respectivos para la realización del estudio.

La muestra principal se va a constituir de las imágenes correspondientes al cultivo que se realice próximo al mes de acuerdo con el cronograma establecido en el informe esto con objeto:

- Recolectar la información inicial para verificar la existencia de las enfermedades estudiadas
- Organizar la información de acuerdo con la calidad de imágenes y la necesidad del estudio.
- Identificación de las imágenes obtenidas para verificar las características establecidas.
- Analizar la presencia de las enfermedades y su comportamiento en el cultivo el cual fue objeto de estudio.

5.4. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

En la investigación las técnicas e instrumentos de recolección de datos se realizará a partir de la combinación de distintas estrategias que van desde la observación hasta el análisis de documentos y reportes.

La observación inicial es nuestra principal fuente para verificar que las plantas están siendo afectadas por las enfermedades ya que estas contemplan ciertas características asociadas con el color para finalmente capturar esta información por medio de una cámara fotográfica estándar que tenga la calidad deseada y que nos aporte la suficiente información respecto a las características establecidas.

El análisis de documentos y reportes se asocia a los resultados que generemos en los informes del procesamiento por medio del software utilizado, para posteriormente realizar un estudio estadístico, para determinar el grado de afectación y distintas medidas que nos otorguen información con varianza,

promedio de afectación de la planta, promedios de afectación en los cultivos, para finalmente expresar esta información a través del análisis de resultados totales.

5.5. TÉCNICAS DE PROCESAMIENTO Y ANALISIS DE DATOS

Para el procesamiento de imágenes se utilizará la librería Open CV en el lenguaje de programación python con el fin de obtener los resultados individualizados de cada fotografía, realizar la transformación de la imagen a formato HSV y establecer límites para segmentar las imágenes todo a través de los datos que contiene cada imagen, como lo son:

- Metadatos
- Información de los pixeles

Luego de este punto utilizamos la librería panda para manipulación y análisis de datos la cual nos sirve para generar los archivos correspondientes por medio de tablas, en este punto también se hace necesario la librería "Os" la cual permite realizar operaciones relacionadas con la gestión de archivos, directorios, rutas, variables de entorno y otras funcionalidades del sistema operativo con el fin de guardar nuestros resultados.

Para la descripción de los datos obtenidos en el procesamiento de imágenes utilizaremos las librerías Numpy la cual nos permite realizar cálculos lógicos y matemáticos sobre los cuadros y matrices resultantes; la librería Matplotlib la cual nos permitirá realizar gráficos en dos dimensiones para describir los datos.

Luego de describir los datos para la regresión logística, utilizamos la librería Scikit-learn respectivamente sus modelos de preprocesamiento, escalado de datos, regresión logística, métricas, predicción de probabilidad con el fin de entrenar y realizar pruebas sobre nuestras bases de datos de imágenes.

Finalmente, a partir de nuestros resultados en el procesamiento de imágenes y nuestra regresión logística mostramos nuestros datos en un mapa interactivo por medio de la librería Folium.

6. DESARROLLO PROYECTO

6.1. TOMA DE DATOS

Inicialmente se tomaron 462 imágenes de dos cultivos de caña de azúcar ubicados en el municipio de Puerto López en el departamento del Meta (95 imágenes) y en Maripí el departamento de Boyacá (367 imágenes) las cuales contenían imágenes de referencia de plantas, como se muestra a continuación.



Gráfico 11. Imágenes de Referencia Cultivo Puerto López (Meta)

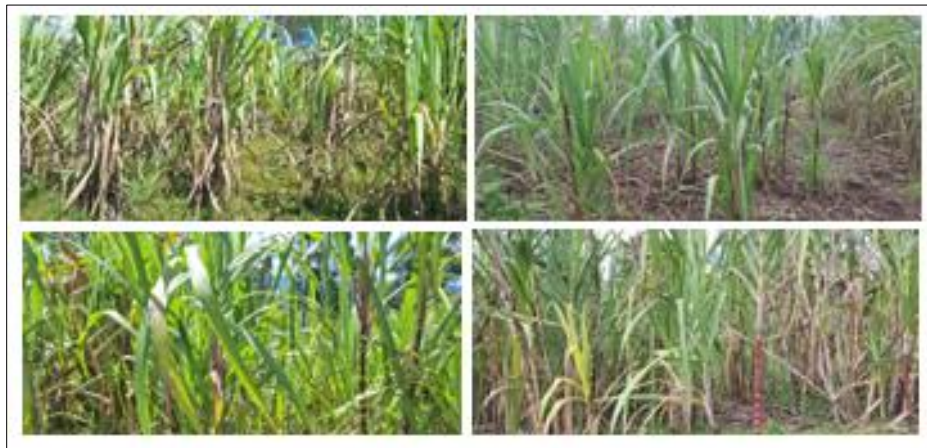


Gráfico 12. Imágenes de Referencia Cultivo Maripí (Boyacá)

Las características principales de las imágenes debían ser que se pudiera visualizar el tallo de la planta y que se tuvieran buenas condiciones de luz para visualizar los colores que son el principal componente por el cual se determina si hay presencia de muermo rojo o si por el contrario no hay presencia y estamos frente a una planta la cual no contiene las características físicas asociadas a la enfermedad.

Los cuales tuvimos en cuenta para nombrar las imágenes, georreferenciarlas y ubicarlas entre nuestros archivos.

6.2. PROCESAMIENTO DE IMÁGENES

Para procesar las imágenes inicialmente las redimensionamos para trabajar en todas las imágenes con un tamaño uniforme a una cantidad de píxeles (512*512).



Gráfico 15. Imagen Original formato RBG

Seguidamente para cada imagen la cual viene en el sistema RBG en su formato original el cual se encuentra definido por los componentes de color rojo, verde y azul para cada píxel individual realizamos una transformación al sistema HSV el cual utiliza componentes de tono, saturación y valor.

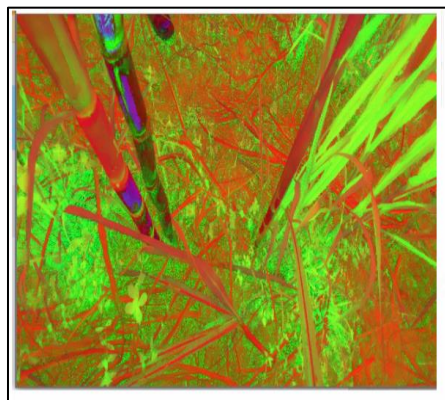


Gráfico 16. Imagen formato HSV

A partir de este punto empezamos con las pruebas en las imágenes para definir parámetros de tono, saturación y brillo, para lograr segmentar por medio de límites inferiores y superiores de color en la imagen a continuación planta segmentada:



Gráfico 17. Imagen segmentación planta

Luego de segmentar los componentes de cada planta en cada una de las imágenes, segmentamos la enfermedad por medio de una máscara con unos límites superiores e inferiores con el fin de identificarla de la siguiente forma:

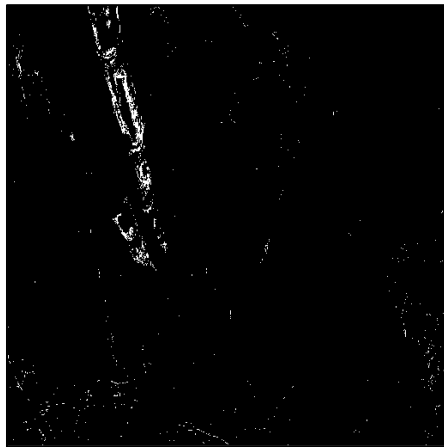


Gráfico 18. Imagen segmentación muermo rojo

Para finalizar el código de procesamiento de imágenes realizamos un conteo de los pixeles que se segmentaron en cada una de las imágenes para determinar cuántos hacen parte de la planta y cuantos se identificaron como muermo rojo según los parámetros establecidos de HSV, esto para posteriormente catalogarlos.

6.3. REGRESIÓN LOGÍSTICA

A partir de las imágenes originales realizamos la creación de una variable binaria independiente (Presencia de muermo) en donde catalogamos cada una de las 462 imágenes individualmente en:

- 1 si hay presencia de muermo rojo
- 0 si no tenemos presencia de la enfermedad

Luego graficamos nuestros resultados de forma visual con el fin de identificar si en base a estas dos variables dependientes tenemos algunos grupos definidos en donde podemos identificar una presencia de muermo rojo.

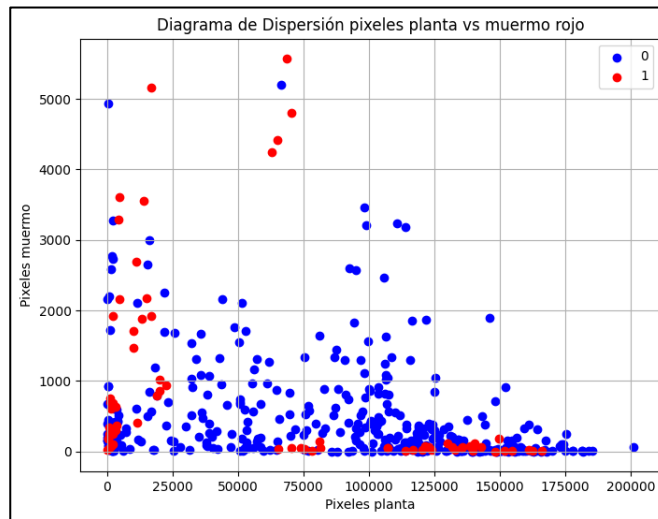


Gráfico 19. Diagrama de dispersión Planta vs Muermo Rojo

A partir de estos datos realizamos nuestra regresión logística, teniendo en cuenta trabajar con la misma cantidad de registros que cuentan con presencia de muermo rojo y los que no tienen presencia en total 151 registros ya que si trabajamos con la base categorizada original se va a sesgar a categorizar todos los datos sin presencia de la enfermedad ya que en la categorización visual nos indicada que el 84,6% de las imágenes no cuentan con la enfermedad por lo cual al predecir valores se va a balancear a categorizar nuestras imágenes como “Sin presencia de la enfermedad”.

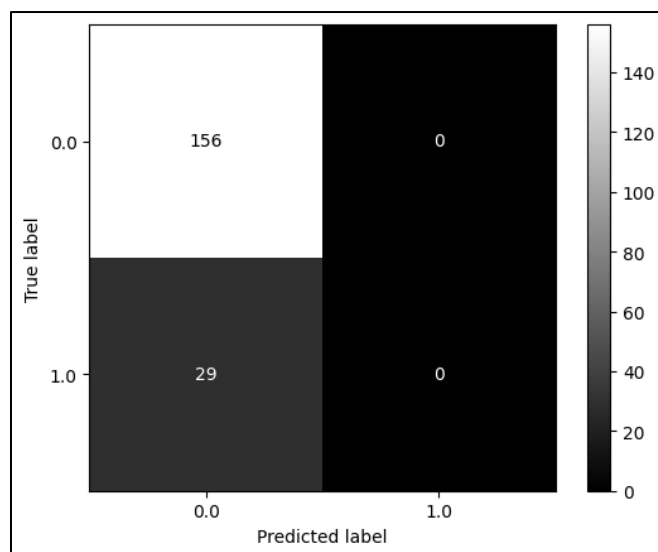


Gráfico 20. Matriz de confusión sin balanceo

A partir de esta matriz inicial aplicamos nuestro balanceo aleatorio suprimiendo imágenes donde no tenemos presencia de muermo en los datos y nos quedamos con un total de 151 registros a los cuales el 47% de las imágenes contaban con presencia de muermo rojo, en donde definimos nuestra variable dependiente (Presencia de muermo) y nuestras 2 variables independientes y que se desprendieron del procesamiento de imágenes como la cantidad de pixeles de la planta y de la enfermedad.

En este caso generamos nuestro gráfico de correlación con el cual se identificó el grado de correlación de cada una de las variables con nuestra variable objetivo

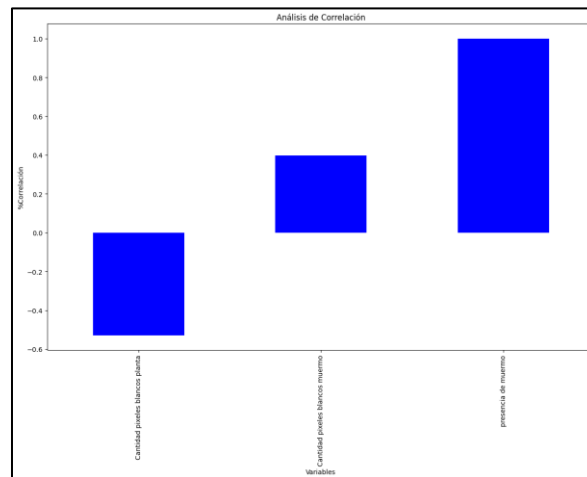


Gráfico 21. Gráfica de Correlación

Seguidamente escalamos los datos para las variables independientes de nuestra predicción:

- Variables sin escalar

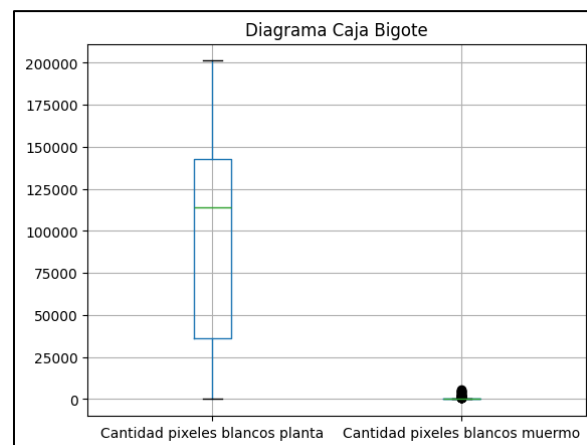


Gráfico 22. Datos sin escalar

- Variables escaladas

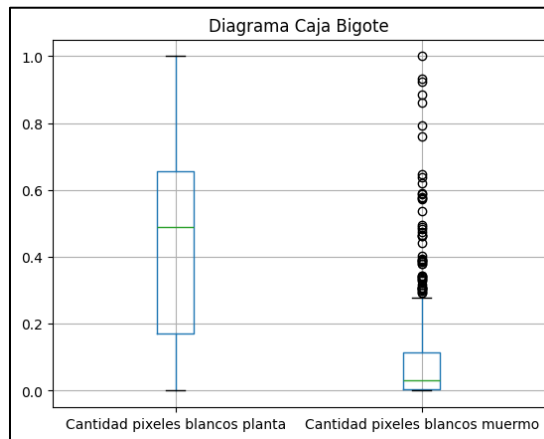


Gráfico 23. Datos escalados

En este caso con los datos escalados y utilizando la librería sklearn tomamos el 60% de los datos para entrenar el modelo y el 40% para evaluarlo, en donde nos quedan 90 datos para entrenamiento y 60 datos para test logrando una predicción del 77% sobre nuestros datos de test como se visualiza a continuación:

```
from sklearn import metrics
prediction_test = model.predict(X_test)
print(metrics.accuracy_score(y_test, prediction_test))
#Buena predicción en el 77% de los casos logro predecir la presencia de muermo
✓ 0.0s
0.7704918032786885
```

Gráfico 24. Porcentaje de predicción sobre nuestros datos de test

En este caso generamos nuestra matriz de confusión a partir de nuestra base de test como se visualiza a continuación:

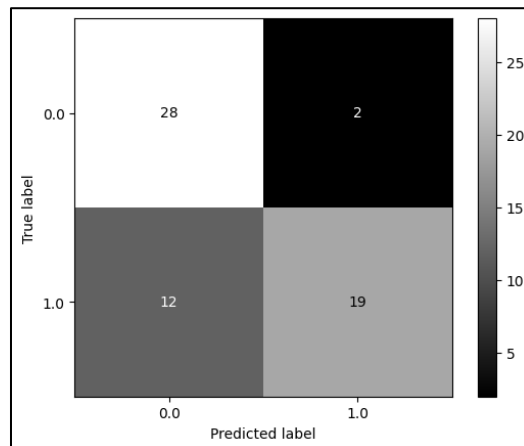


Gráfico 25. Matriz de confusión

En donde dividimos nuestros resultados en:

- Imágenes donde no hay presencia de muermo rojo y se catalogó correctamente.
- Imágenes donde no hay presencia de muermo rojo y se catalogó con presencia.
- Imágenes donde tenemos presencia de muermo rojo y se catalogaron correctamente.
- Imágenes donde tenemos presencia de muermo rojo y se catalogaron sin presencia.

6.4. GEOLOCALIZACIÓN

A partir de los metadatos obtenidos específicamente de los campos “Nombre, latitud, longitud” cuando se realizó el cargue de las imágenes y del procesamiento de las imágenes y la predicción realizada por el modelo de regresión logística graficamos nuestras imágenes en un mapa por medio de la latitud y longitud obtenidas las imágenes con la máscara de detección de la enfermedad, el muermo y la imagen original.

A continuación, se realiza la descripción de cada marcador con ayuda de las bibliotecas Folium, Os y base 64 en python respectivamente.



Gráfico 26. Imágenes de cada marcador

En donde podemos visualizar para cada marcador en el mapa:

- La proyección
- Latitud y longitud de la imagen original
- La imagen original
- La máscara de muermo rojo
- La máscara de planta

De la misma forma el marcador agranda su tamaño si la cantidad de pixeles que representan el muermo rojo es mayor en cada uno de los puntos de referencia y toma color rojo si hay presencia de muermo rojo o verde si no hay presencia.

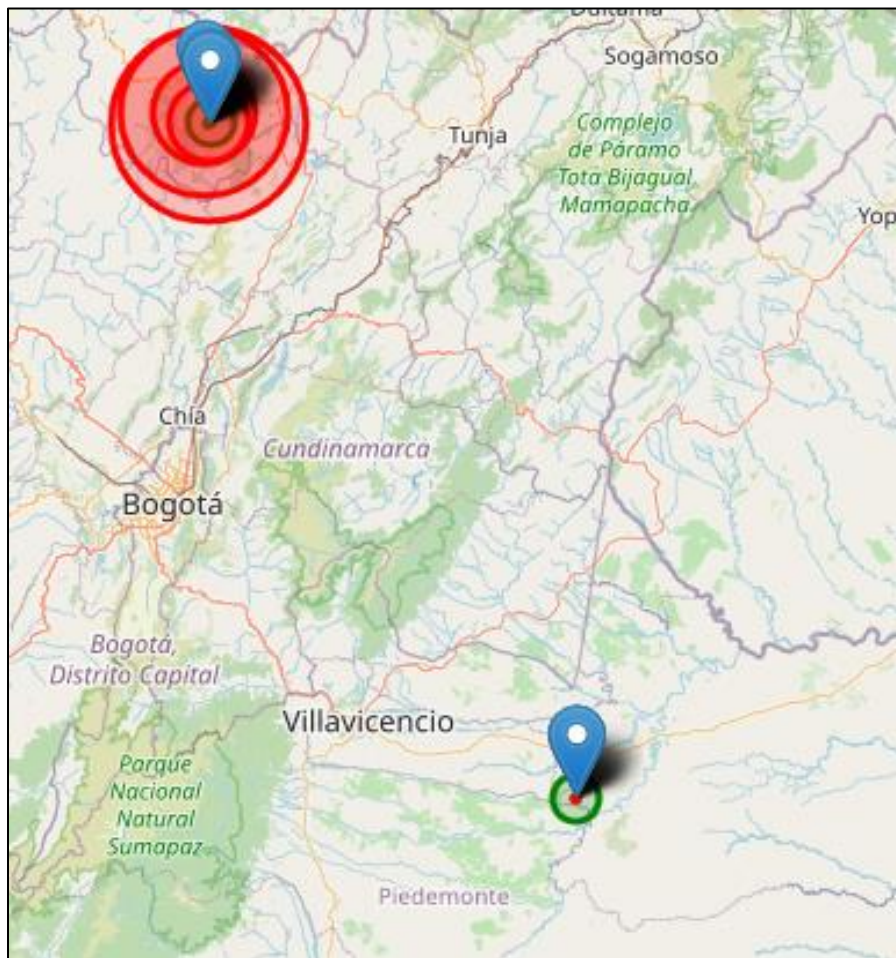


Gráfico 27. Visualización de marcadores

7. RESULTADOS

De acuerdo a la caracterización inicial de forma visual para cada uno de los dos espacios donde se tomaron muestras, se caracterizó la información de la siguiente manera.

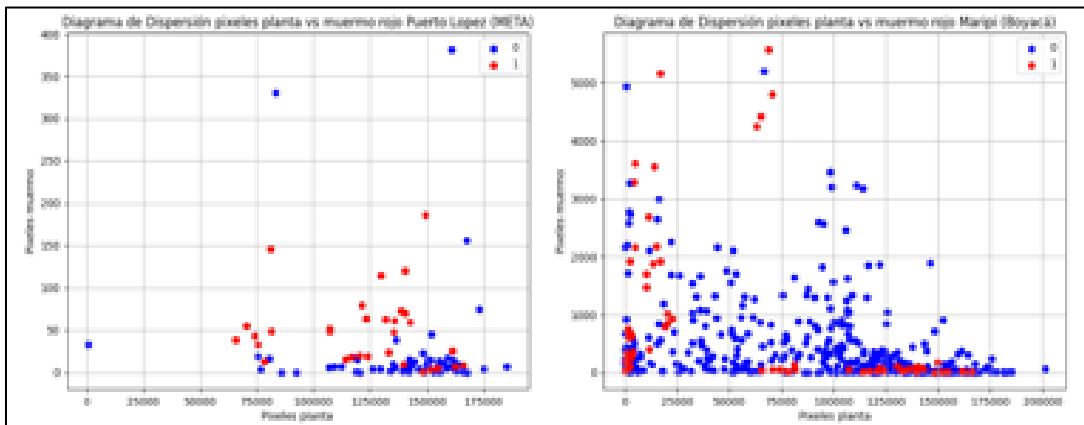


Gráfico 28. Diagrama de dispersión para determinar presencia focalizada

En donde se identificó inicialmente que tenemos una presencia de muermo Rojo en el 32% de las imágenes de Puerto López en el departamento del Meta y del 15% de las imágenes procesadas en Maripí en el departamento de Boyacá de acuerdo a nuestra caracterización inicial.

A partir de esta información con todas las imágenes realizamos balanceo de los datos para evitar un sobreajuste y que la predicción siempre nos arroje “ausencia de muermo” como se visualizó en el gráfico 20; teniendo en cuenta el sobreajuste no quedamos con 151 imágenes las cuales se encuentran balanceadas en una proporción de 47% con presencia de muermo rojo y el restante sin presencia.

Teniendo en cuenta nuestros datos para la regresión logística en el total de datos tomamos el 40% para test obteniendo un accuracy el cual es nuestra medida de predicción correcta del 77% donde identifica correctamente la presencia de muermo rojo, siendo parte fundamental lo que aporta la cantidad de píxeles que pertenecen a muermo y a planta de la siguiente forma.

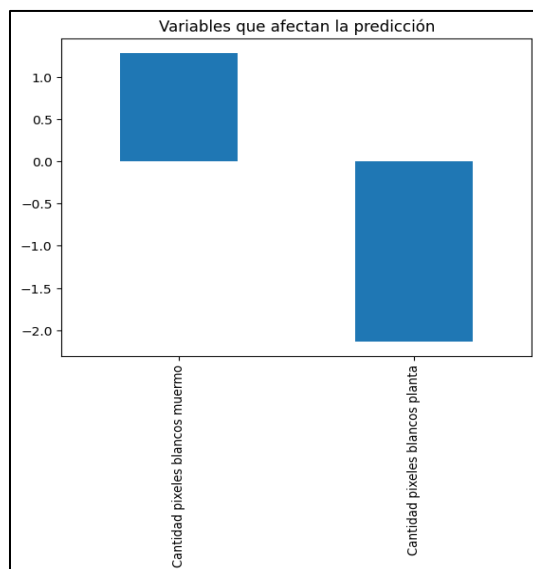


Gráfico 29. Afectación de variables

En donde como resultado tenemos que nuestra variable que afecta positivamente la predicción para detección de muermo rojo es la “Cantidad de pixeles blancos de muermo” y la que afecta negativamente es la “Cantidad de pixeles que pertenecen a la planta” teniendo en cuenta que los límites establecidos para nuestras imágenes en formato HSV son los siguientes:

```
lim_inf_muermo = (min_pos, 0, 0)
lim_sup_muermo = (min_pos + 10, 100, 230)
mask_muermo = cv2.inRange(image_hsv, lim_inf_muermo, lim_sup_muermo)
```

Gráfico 30. Límites mascara de Muermo Rojo

Y para detección de planta los límites son:

```
lim_inf_planta = (max_pos - 10, 50, 50)
lim_sup_planta = (max_pos + 10, 255, 255)
mask_planta = cv2.inRange(image_hsv, lim_inf_planta, lim_sup_planta)
```

Gráfico 31. Límites mascara de planta

Finalmente, para toma de decisiones de los agricultores se geolocaliza la información obtenida junto con la latitud y longitud, la imagen original, la imagen donde se evalúa la presencia de muermo rojo y la imagen de la planta segmentada para que de acuerdo a las Buenas Prácticas agropecuarias para el control de la enfermedad tomen la decisión de hacer un buen manejo de la semilla, no causar heridas sobre las yemas y tratar la semilla con fungicidas. (Osorio, 2007)

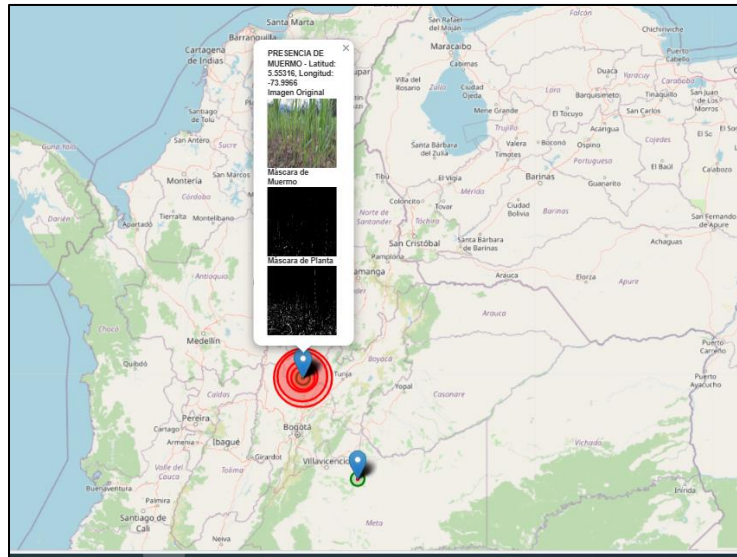


Gráfico 32. Mapa cultivos

8. CONCLUSIONES Y RECOMENDACIONES

El procesamiento de imágenes funciona como un medio de detección de muermo rojo debido a las características físicas que contiene la enfermedad primordialmente los colores que se asocian.

Se destaca que los metadatos en la recolección de imágenes son parte fundamental para realizar la ubicación de la imagen y otorgar información correcta al agricultor de donde se tiene presencia de la enfermedad.

Para realizar la geolocalización se debe tener en cuenta las capacidades de cómputo la cantidad de imágenes a visualizar ya que cada aumento en el número de imágenes visualizar afecta el tiempo de procesamiento.

A partir del siguiente trabajo se puede empezar a mapear distintas enfermedades o plagas que afectan plantas por medio del procesamiento de imágenes, ofreciendo al agricultor la posibilidad de conocer a nivel nacional donde se determinó presencia y de qué tipo de enfermedad lo cual podría ofrecer información para la contención y toma de decisiones.

Para trabajo futuro si se realizara un control estricto sobre las plantaciones a nivel nacional o regional podríamos combatir también plantas no autóctonas de ciertas regiones que desplazan la flora y los cultivos; así mismo se podría segmentar todo un conjunto de plantas por sus características físicas que se podrían procesar mediante imágenes aéreas y/o satelitales lo que eventualmente ayudaría a combatir la deforestación y los cultivos ilegales.

9. BIBLIOGRAFIA

- Ahmad Supian, M. B., Madzin, H., & Albahari, E. (2019). (2019). Plant disease detection and classification using image processing techniques: A review. Paper presented at the 1-4. doi:10.1109/ICAE47758.2019.9221712
- Beyeler, M. (2015). OpenCV with Python Blueprints. Packt Publishing.
- Resolución número 000779 de 2006, Resolución U.S.C. (2006). Retrieved from <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/Resolucion-779-de-2006.pdf>
- Bhimte, N. R., & Thool, V. R. (2018). (2018). Diseases detection of cotton leaf spot using image processing and SVM classifier. Paper presented at the 340-344. doi:10.1109/ICCONS.2018.8662906
- Devaraj, A., Rathan, K., Jaahnavi Sarvepalli, & Indira, K. (2019). Identification of plant disease using image processing technique. *Conferencia Internacional De Procesamiento De Señales Y Comunicaciones De 2019 (ICCSP)*, Retrieved from <https://ieeexplore-ieee-org.ezproxy.javeriana.edu.co/document/8698056>
- Departamento Administrativo Nacional de Estadística - DANE. (2020). Lineamientos para el proceso estadístico en el Sistema Estadístico Nacional versión 2.0. Retrieved from: http://www.dane.gov.co/files/sen/normatividad/NTC_Proceso_Estadistico.pdf
- Fedepanela. (2020a). Áreas, rendimiento y producción proyección para 2020., 1-4. Retrieved from http://www.sipa.org.co/wp/wp-content/uploads/CIFRAS_2020_FEDEPANELA.pdf
- Fedepanela. (2020b). *Importancia de la calidad y cumplimiento de bpm en la producción de panela.* (). Retrieved from http://www.sipa.org.co/wp/wp-content/uploads/Tema2_BPM.pdf
- Fedepanela. (2021). *Mercado internacional de panela cierre 2020.* ().
- Fedepanela. (2021). *Ventas externas de panela tuvieron un crecimiento de 300% en los últimos cinco años.* (). Retrieved from <https://fedepanela.org.co/gremio/ventas-externas-de-panela-tuvieron-un-crecimiento-de-300-en-los-ultimos-cinco-anos/>
- Fedepanela. (2022). INFORME DE INTELIGENCIA DE MERCADOS. Retrieved from: <https://fedepanela.org.co/gremio/wp-content/uploads/2023/04/Informe-Inteligencia-de-mercados-2022.pdf>

Folium. (2023). Retrieved from:
<https://pythonvisualization.github.io/folium/latest/index.html>

García, H. R., & Bohórquez, J. (2000). *Oportunidades de producción más limpia en la industria panelera*. Colombia: Retrieved from
<http://sie.car.gov.co/bitstream/handle/20.500.11786/36901/02835.pdf?sequence=1&isAllowed=y>

Gómez-Camperos, J., Jaramillo, H., & Guerrero-Gómez, G. (2021). Técnicas de procesamiento digital de imágenes para detección de plagas y enfermedades en cultivos: una revisión. *Ingeniería Y Competitividad*, 24(1).
<https://doi.org/10.25100/iyc.v24i1.10973>

IBM. (2023). Regresión Logística. Retrieved from:
<https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-logistic>

Instituto Nacional del Cáncer. (s.f.). Fungicida.
<https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/fungicida>

Kumar, S. S., & Raghavendra, B. K. (2019). (2019). Diseases detection of various plant leaf using image processing techniques: A review. Paper presented at the 313-316. doi:10.1109/ICACCS.2019.8728325

Mincit (2022). El 2022 fue un año histórico para las exportaciones de bienes no minero energéticos. Retrieved from
[https://www.mincit.gov.co/prensa/noticias/comercio/2022-ano-historico-para-exportaciones-nme#:~:text=Alcanzaron%20los%20US%248.442%20millones,%2C7%20%25\)%2C%20entre%20otros.](https://www.mincit.gov.co/prensa/noticias/comercio/2022-ano-historico-para-exportaciones-nme#:~:text=Alcanzaron%20los%20US%248.442%20millones,%2C7%20%25)%2C%20entre%20otros.)

Minsalud(2019). Módulo – Azúcares Capacitación equipos básicos de salud - APS Retrieved from
<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ENT/modulo-azucares.pdf>

Oracle.(2023). Definición de JSON. Retrieved from:
<https://www.oracle.com/co/database/what-is-json/#json-defined>

Osorio, G. C. (2007). *Manual técnico: Buenas prácticas agrícolas -BPA- y buenas prácticas de manufactura -BPM-en la producción de caña y panela* (Primera Edición ed.). Medellín, Colombia: CTP Print Ltda.

Padilla Jiménez, A. V., Rojas, C. A. P., & Guzmán, L. E. S. (2016). (2016). Procesamiento de imágenes para la identificación de plagas en los cultivos de espinaca. Paper presented at the 1-7. doi:10.1109/CCRA.2016.7811412

Potes, Alberto (2023), Caña de azúcar en ladera, Retrieved from https://repository.agrosavia.co/bitstream/handle/20.500.12324/21826/23182_4671.pdf?sequence=1&isAllowed=y#:~:text=Putridi%C3%B3n%20roja%20o%20muermo%20rojo,del%20%22barrenador%22%20del%20tallo.

Pradeep Kumar Mugithe, Rohit Varma Mudunuri, B Rajasekar, & S Karthikeyan. (2020). Image processing technique for automatic detection of plant diseases and alerting system in agricultural farms. *International Conference on Communication and Signal Processing*, Retrieved from <https://ieeexplore-ieee-org.ezproxy.javeriana.edu.co/document/9182065>

Quintero, A. F. (2018, septiembre 20,). Al 2022, exportaciones de panela llegarían a unas 18.000 toneladas. *Portafolio.Co* Retrieved from <https://www.portafolio.co/economia/al-2022-exportaciones-de-panela-llegarian-a-unas-18-000-toneladas-521390>

Regional Siete de Corpoica y el SENA Regional Santander (Ed.). (1998). *Manual de caña de azúcar para la producción de panela*. Bucaramanga, Colombia:

Scikit-learn. (2023). Available documentation for Scikit-learn. Retrieved from: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

Schneider, G. E., Solano, A., Kemerer, A., & Hadad, A. Gestión y procesamiento de imágenes aéreas de caña de azúcar. *Universidad Nacional De Entre Ríos*,

Zhang, Y. (1996). Image engineering and bibliography in China. Technical Digest of International Symposium on Information Science and Technology:

Zhang, Y. (2017). Image processing. Berlin, Germany]: De Gruyter. Retrieved from <https://login.ezproxy.javeriana.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=1576211&lang=es&site=ehost-live>

10. ANEXOS

ANEXO A. RECURSOS

CANTIDAD	RECURSOS
	HUMANOS
1	Estudiante
1	Tutor tesis
	MATERIALES
1	Software de procesamiento de imágenes
1	Software para tratamiento de datos
1	Cámara definición estándar
1	Computador
1	Memoria extraíble para almacenamiento de imágenes

ANEXO B. ECUACIÓN DE BUSQUEDA

Tema	Palabras clave	Agrupamientos	Búsqueda
Procesamiento de imágenes	Image processing Image digitalization	"image processing" OR "image digitalization"	("image processing" OR "image digitalization")
AND	AND	AND	AND
Caña de azúcar	Sugar cane Sugar cane industry Sugarmill	"sugar cane" OR "sugar cane industr*" OR "sugarmill"	("sugar cane" OR "sugar cane industr*" OR "sugarmill")
AND	AND	AND	AND
Enfermedades	Diseases Mushrooms	"diseas*" OR "mushroom*"	("diseas*" OR "mushroom*")
AND	AND	AND	AND
Proceso panela	Panela process	"panela process"	("panela process")

ANEXO C. EVALUACIÓN DEL MODELO PROBABILIDADES

```
#Evaluación del modelo
model.predict_proba(X_test)
✓ 0.0s

array([[0.32371988, 0.67628092],
       [0.32746487, 0.67253513],
       [0.47899117, 0.52100883],
       [0.58828144, 0.41971856],
       [0.66614968, 0.33385032],
       [0.62238533, 0.37761467],
       [0.7416289 , 0.2583711 ],
       [0.71608993, 0.28391007],
       [0.67073575, 0.32926425],
       [0.25792815, 0.74207985],
       [0.66582582, 0.33417418],
       [0.63756929, 0.36243871],
       [0.3342216 , 0.66577784],
       [0.69649113, 0.30350887],
       [0.59539849, 0.40460951],
       [0.70374965, 0.29625835],
       [0.26755831, 0.73244969],
       [0.38061801, 0.69938199],
       [0.52787468, 0.47212532],
       [0.58823531, 0.41176469],
       [0.70992267, 0.29007733],
       [0.70197568, 0.29802432],
       [0.31729962, 0.68278938],
       [0.30391113, 0.69688887],
       [0.27936706, 0.72063294],
       ...
       [0.76969237, 0.23030763],
       [0.62454412, 0.37545588],
       [0.71092529, 0.28907471],
       [0.68045765, 0.31954235],
       [0.61399423, 0.38600577]])

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings.
```

ANEXO D. CURSOS DE ANÁLISIS ADICIONALES

- Curso de Regresión logística



- Curso de manejo librería de regresión logística para machine learning



- Curso de prompt engineering para apoyarme con consultas en Chat GPT

