

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Identificación de contenido de clorofila en hojas del cultivo de maíz por medio de imágenes multiespectrales adquiridas con drones y su correlación con métodos estándares de laboratorio

PROYECTO INTEGRADOR

Previo la obtención del Título de:

INGENIERO EN COMPUTACIÓN

Presentado por:

Allan Samuel Alarcón Yépez

Joe Sebastián Saverio Alvarado

GUAYAQUIL - ECUADOR

Año: 2019

DEDICATORIA

El presente proyecto está dedicado en primer lugar a Dios y a mis padres por todo el esfuerzo que han realizado y el apoyo que me han dado a lo largo de mi vida.

ALLAN SAMUEL ALARCÓN YÉPEZ

El presente proyecto está dedicado a mis padres y a mi familia por el apoyo incondicional y ayuda que tuve de ellos durante el transcurso de la carrera.

JOE SEBASTIÁN SAVERIO ALVARADO

AGRADECIMIENTOS

Agradezco a Dios, a mis padres por sus inmensos deseos de superación y a nuestros profesores que nos brindaron su apoyo y enseña a lo largo de la carrera.

ALLAN SAMUEL ALARCÓN YÉPEZ

A mi padres y familia que estuvieron dándome su apoyo y consejos desde el inicio hasta el fin de la carrera. A mis amigos y compañeros con los que siempre conté con su ayuda.

JOE SEBASTIÁN SAVERIO ALVARADO

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; y damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

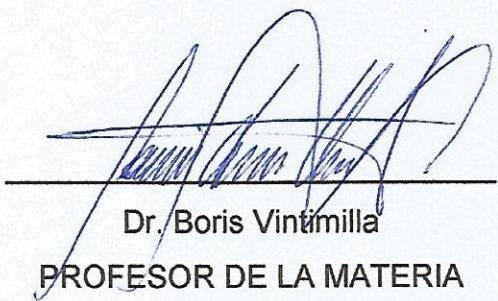


Allan Samuel Alarcón Yépez

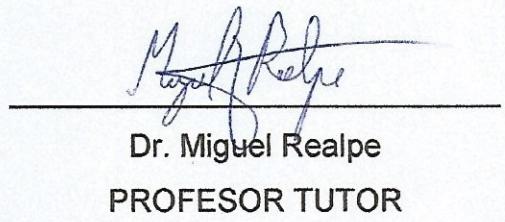


Joe Sebastián Saverio Alvarado

EVALUADORES



Dr. Boris Vintimilla
PROFESOR DE LA MATERIA



Dr. Miguel Realpe
PROFESOR TUTOR

RESUMEN

Es importante para los agricultores conocer el nivel de clorofila de las plantas ya que de esto depende el tratamiento que deben darle a sus cultivos. Existen dos métodos comunes para obtener la clorofila, a partir de análisis en laboratorios y de dispositivos electrónicos. Ambos métodos obtienen el nivel de clorofila de una muestra a la vez, aunque pueden ser destructivos. El objetivo es desarrollar un sistema que permita la obtención del nivel de clorofila de las plantas utilizando imágenes.

Para el desarrollo de la solución se utilizó el lenguaje de programación Python y diferentes librerías de dicho lenguaje. Se decidió realizar un sistema de etiquetado, una regresión lineal simple y un sistema de predicción. El primer sistema se utilizó para crear una base de datos que relacione los valores de las imágenes con los de clorofila, la cual se utilizó para obtener la regresión lineal y determinar la relación que existe entre estas variables. Esta regresión lineal fue utilizada en el sistema de predicción para obtener los valores de clorofila a partir de las imágenes.

La correlación lineal entre las imágenes y la clorofila SPAD fue 0.7054 y la raíz del error cuadrático medio fue 6.581877160210805 unidades SPAD en la primera prueba, en la segunda prueba la correlación lineal fue 0.0186 y la raíz del error cuadrático medio fue 6.477188141888954, mientras que en la tercera prueba la correlación lineal fue 0.1320 y la raíz del error cuadrático medio fue 7.27287620832889 unidades SPAD.

Se concluye que, debido a que la correlación lineal es positiva, los valores de las imágenes son directamente proporcionales a los valores de clorofila SPAD y, como es mayor a 0.7, la relación es significativa; con lo cual el sistema de predicción brinda una precisión del 70.54% utilizando la primera regresión. En los casos de la segunda y tercera regresión se concluye que para obtener una mejor regresión deben usar diferentes estados de salud de las plantas para tener un mayor rango de valores NDVI con su respectivo valor SPAD.

Palabras clave: NDVI, clorofila, plantas, cultivo, imágenes

ABSTRACT

It is important for farmers to know the level of chlorophyll in plants since this depends on the treatment they should give to their crops. There are two common methods to obtain chlorophyll, from laboratory analysis and electronic devices. Both methods obtain the chlorophyll level of one sample at a time, although they can be destructive. The objective is to develop a system that allows obtaining the chlorophyll level of the plants using images.

The Python programming language and different libraries of that language were used to develop the solution. It was decided to carry out a labeling system, a simple linear regression and a prediction system. The first system was used to create a database that links the values of the images with those of chlorophyll, which was used to obtain linear regression and determine the relationship between these variables. This linear regression was used in the prediction system to obtain chlorophyll values from the images.

The linear correlation between the images and the SPAD chlorophyll was 0.7054 and the root-mean-square error was 6.581877160210805 SPAD units in the first test, in the second test the linear correlation was 0.0186 and the root-mean-square error was 6.477188141888954, while in the third test the linear correlation was 0.1320 and the root-mean-square error was 7.27287620832889 SPAD units.

It is concluded that, because the linear correlation is positive, the image values are directly proportional to the SPAD chlorophyll values and, as it is greater than 0.7, the relationship is significant; whereby the prediction system provides an accuracy of 70.54% using the first regression. In the cases of the second and third regression it is concluded that to obtain a better regression they must use different states of health of the plants to have a greater range of NDVI values with their respective SPAD value.

Keywords: NDVI, Chlorophyll, plant, crop, images

ÍNDICE GENERAL

RESUMEN.....	I
ABSTRACT	II
ÍNDICE GENERAL	III
ABREVIATURAS.....	V
ÍNDICE DE FIGURAS	VI
ÍNDICE DE TABLAS.....	VII
CAPÍTULO 1.....	1
1. INTRODUCCIÓN	1
1.1 Descripción del problema.....	1
1.2 Objetivos	2
1.2.1 Objetivo General	2
1.2.2 Objetivos Específicos.....	2
1.3 Marco Teórico	3
1.3.1 Análisis a través de dispositivos electrónicos	3
1.3.2 Análisis de tejidos	3
1.3.3 Sistemas de información geográficos para agricultura.....	4
1.4 Justificación.....	6
CAPÍTULO 2.....	8
2. METODOLOGÍA	8
2.1 Diseño de la solución	8
2.1.1 Recolección de datos.....	9
2.1.2 Relacionamiento entre NDVI y clorofila.....	10
2.1.3 Regresión lineal	11

2.2 Tecnologías de desarrollo	13
2.3 Metodología.....	14
CAPÍTULO 3.....	17
3. IMPLEMENTACIÓN Y RESULTADOS	17
3.1 Implementación de la solución	17
3.1.1 Base de datos	17
3.1.2 Sistema de etiquetado	18
3.1.3 Regresión lineal	22
3.1.4 Sistema de predicción.....	23
3.2 Análisis de resultados	25
3.2.1 Prueba del sistema de etiquetado de imágenes	25
3.2.2 Primera prueba de la regresión lineal simple con SPAD teórico	28
3.2.3 Segunda prueba de la regresión lineal simple con SPAD manual	30
3.2.4 Tercera prueba de la regresión lineal simple con SPAD manual	32
3.2.4 Prueba del sistema de predicción	34
CAPÍTULO 4.....	38
4. CONCLUSIONES Y RECOMENDACIONES	38
4.1 Conclusiones.....	38
4.2 Recomendaciones.....	40
4.1 Trabajos futuros	41
BIBLIOGRAFÍA.....	42
APÉNDICES.....	44

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
GAE	Granja Experimental Agrícola
SPAD	Soil Plant Analysis Development
NDVI	Índice de vegetación diferencial normalizada
NDRE	Índice de borde rojo diferencial normalizado
VANT	Vehículo aéreo no tripulado
RGB	Red, Green, Blue (modelo de color)

ÍNDICE DE FIGURAS

Figura 2.1 Modelo conceptual del sistema	8
Figura 2.2 Modelo planteado de recolección de datos	10
Figura 2.3 Modelo planteado de almacenamiento de datos	11
Figura 2.4 Representación de una regresión lineal	13
Figura 3.1 Ventana principal del sistema de etiquetado	19
Figura 3.2 Ingreso de valores de SPAD y de laboratorio.....	20
Figura 3.3 Procedimiento para identificar la región de interés.....	21
Figura 3.4 Etiquetado de la región de interés	22
Figura 3.5 Ventana principal del sistema de predicción	24
Figura 3.6 Imagen del cultivo de maíz en GEA	25
Figura 3.7 Etiquetado de las plantas del cultivo	26
Figura 3.8 Representación de los datos de la primera prueba con SPAD teórico	28
Figura 3.9 Primera prueba de regresión lineal simple con SPAD teórico	29
Figura 3.10 Representación de los datos de la segunda prueba con SPAD manual	30
Figura 3.11 Segunda prueba de regresión lineal simple con SPAD manual	31
Figura 3.12 Representación de los datos de la tercera prueba con SPAD manual	33
Figura 3.13 Tercera prueba de regresión lineal simple con SPAD manual	33
Figura 3.14 Ventana principal del sistema con imagen del cultivo	35
Figura 3.15 Valor de NDVI del píxel y de clorofila usando la regresión lineal	36

ÍNDICE DE TABLAS

Tabla 1.1 Comparativa de métodos de medición de clorofila	7
Tabla 3.1 Representación de la base de datos	27
Tabla 3.2 Comparación de resultados de la primera prueba de regresión lineal.....	30
Tabla 3.3 Comparación de resultados de la segunda prueba de regresión lineal	32
Tabla 3.4 Comparación de resultados de la tercera prueba de regresión lineal.....	34
Tabla 3.5 Comparación de resultados de sistema de predicción	37

CAPÍTULO 1

1. INTRODUCCIÓN

En este capítulo introductorio se presenta el problema, los objetivos planteados, el marco teórico y la respectiva justificación del proyecto. En la sección del marco teórico se detallan diferentes soluciones que existen actualmente para resolver el problema planteado.

1.1 Descripción del problema

En el área de la agricultura, las condiciones ambientales y el tratamiento del cultivo, como es el uso de fertilizantes, definen el crecimiento adecuado de las plantas ya que estos afectan a los niveles de clorofila. Debido a esto, es fundamental para los agricultores conocer los niveles de clorofila ya que de esta forma se conoce en qué estado se encuentra el cultivo, permitiéndoles tomar la mejor decisión respecto a la cantidad de fertilizantes que deben aplicar para así tener una producción de gran calidad [1].

En la Granja Experimental Agrícola (GEA) de la Escuela Superior Politécnica del Litoral (ESPOL), institución que sirvió como objeto de estudio del presente trabajo de investigación, la obtención del nivel de clorofila de las plantas se realiza a través del dispositivo electrónico llamado SPAD, el cual es un dispositivo que determina la cantidad de clorofila en la planta de forma no destructiva a partir de una lectura en la planta [2], y también a través del análisis de tejidos de las plantas en los laboratorios.

La obtención mediante los dispositivos electrónicos SPAD requiere una gran inversión inicial por parte de las instituciones agrícolas debido al costo que poseen dichos dispositivos, sin embargo, el tiempo de obtención del nivel de clorofila es

inmediato siendo no perjudicial para las plantas, mientras que el análisis de tejidos es destructivo para las plantas y requiere mayor tiempo [3] debido a que la obtención de la clorofila no se realiza de forma inmediata sino que se realiza en un laboratorio.

Por lo antes expuesto, es preciso encontrar un sistema que mejore el proceso de obtención de clorofila de las plantas, tanto en tiempos de obtención como en costos operativos, al igual que no debe ser destructivo para las plantas, para de esta manera mejorar la producción de los cultivos agrícolas.

1.2 Objetivos

1.2.1 Objetivo General

Diseñar, desarrollar e implementar un sistema para predecir el nivel de clorofila que poseen las plantas en un cultivo a partir de imágenes multiespectrales, capturadas por un vehículo aéreo no tripulado (VANT), utilizando procesamiento de imágenes.

1.2.2 Objetivos Específicos

- Diseñar, desarrollar, implementar y probar un sistema que permita segmentar y etiquetar el nivel de clorofila de las plantas en imágenes multiespectrales.
- Determinar la relación que existe entre el nivel de clorofila en las plantas con el valor del índice de vegetación diferencial normalizada obtenido a partir de imágenes multiespectrales usando una regresión lineal simple.
- Comparar los métodos que se utilizan en la GEA para la obtención de clorofila de las plantas con el sistema de visión por computador desarrollado.

1.3 Marco Teórico

En la Granja Experimental Agrícola se obtiene el nivel de clorofila en las plantas por medio de dos métodos: utilizando un dispositivo electrónico y realizando un análisis de tejido en los laboratorios. Ambos métodos requieren que una persona esté presente en el campo, ya sea para poder tomar las muestras necesarias, en el caso del análisis de tejidos, o poder hacer las lecturas, en el caso del dispositivo electrónico.

1.3.1 Análisis a través de dispositivos electrónicos

El dispositivo de análisis de desarrollo de plantas de suelo, o SPAD por sus siglas en inglés, posee dos sensores donde se coloca la hoja y se toma una medida en unidades SPAD, esta medida se da en un rango de -9.9 y 199 [4]. Para poder llegar a dar esta lectura, el dispositivo mide la absorción de la hoja en dos regiones de longitud de onda, el infrarrojo cercano y el rojo, usando estas dos regiones se calcula un valor numérico de SPAD, el cual es proporcional al nivel de clorofila que tiene la planta. Este método se caracteriza por dar una respuesta inmediata, precisa y por no ser perjudicial para las plantas debido a que no es necesario perforarlas [5].

1.3.2 Análisis de tejidos

El método utilizado para el análisis de tejidos es el método de Dumas, el cual es comúnmente aplicado para determinar el nivel de nitrógeno y azufre de sustancias químicas, siendo un método de alta precisión [6]. Este método consiste en una combustión seca en la cual una muestra se somete a temperaturas altas para obtener dióxido de carbono, agua y nitrógeno. Luego de esto, estas sustancias pasan a través de un nitrómetro con el cual se puede aislar el nitrógeno [7]. Debido a que el nivel de nitrógeno y el de clorofila están fuertemente relacionados [8], este método brinda una medida fiable del nivel de clorofila presente en las plantas. Para facilitar el análisis mediante el método Dumas, se han creado dispositivos que permiten realizarlo de forma automática,

como, por ejemplo, el vario MAX Cube, el cual garantiza un tiempo de medición de hasta 4 minutos [9].

1.3.3 Sistemas de información geográficos para agricultura

Una forma no invasiva y automática de poder conseguir el nivel de clorofila de las plantas en un cultivo, y por consiguiente determinar el estado de salud de las plantas, es a través de los sistemas de información geográficos, GIS por sus siglas en inglés, enfocados en la agricultura. Estos sistemas utilizan índices de vegetación para poder determinar el nivel de clorofila de las plantas, los cuales los obtienen mediante el uso de imágenes multiespectrales capturadas por satélites o por VANTs [10], [11].

Los índices de vegetación nos ayudan a realzar las propiedades vegetativas de un campo de observación, indicando si existe o no vegetación. Para poder visualizar estos indicadores se utilizan imágenes multiespectrales, las cuales permiten observar la radiación emitida por las plantas luego de haber absorbido energía solar, que no es visible en el rango de visión del ser humano y en las imágenes captadas por cámaras convencionales [12]. Los índices se obtienen a partir de un cálculo matemático en donde se utilizan diferentes rangos multiespectrales, los cuales dependen de la información que se desea analizar. Debido al comportamiento multiespectral de las plantas, aproximadamente el 90% de la información relacionada con la vegetación se encuentra en los espectros del rojo y del infrarrojo cercano, de los cuales el rojo se relaciona con la clorofila, mientras que el infrarrojo lo hace con la densidad de las plantas [13]. Entre los indicadores más utilizados para aplicaciones de monitoreo de cultivos que incluyen procesamiento de imágenes están el NDVI y el NDRE.

1.3.3.1 NDVI

El índice de vegetación diferencial normalizada, NDVI por sus siglas en inglés, es un índice que nos indica la presencia de biomasa en una imagen procesada [14]. Este indicador es utilizado para estimar la calidad

y el desarrollo de los campos agrícolas, lo cual les facilita a los expertos tomar decisiones para mejorar la producción del campo. En las aplicaciones de monitoreo de campos agrícolas, el NDVI es comúnmente utilizado para identificar áreas problemáticas, identificar anomalías en las temporadas de cultivos, dar estimaciones de rendimientos, identificar malezas, entre otras funciones.

Para calcular el índice se utiliza la siguiente ecuación:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1.1)$$

Donde NIR se refiere a la captura de información en el rangopectral del infrarrojo cercano, mientras que Red se refiere al rangopectral rojo.

Debido al uso del rangopectral rojo, NDVI es un gran indicador para cultivos donde las plantas sean de poca densidad, de bajo dosel, o que se encuentren en épocas tempranas de crecimiento, ya que el rangopectral rojo es muy sensible a la parte alta de las plantas, lo cual significa que no obtiene suficiente información de la parte baja del cultivo y lo hace propenso a saturaciones, es decir, varias partes de la imagen estarán en nivel NDVI de 1.0 [14], [15].

1.3.3.2 NDRE

El índice de borde rojo diferencial normalizado, NDRE por sus siglas en inglés, es similar al NDVI. Es utilizado para detectar vegetación saludable, pero este se encuentra más enfocado en detectar niveles de clorofila de las plantas [12].

Para calcular el índice se utiliza la siguiente ecuación:

$$NDRE = \frac{NIR - RE}{NIR + RE} \quad (1.2)$$

Donde NIR se refiere a la captura de información en el rango espectral del infrarrojo cercano, mientras que RE se refiere al rango espectral rojo cercano.

A diferencia del NDVI, el NDRE soluciona el problema del rango espectral rojo, utilizando el rojo cercano, lo cual le permite no saturarse en presencia de plantas altas y a su vez obtener más información de la parte baja de las plantas. Debido a esto, el NDRE es útil en las mismas situaciones que el NDVI, pero también se lo puede utilizar en cultivos donde las plantas sean muy densas, de alto dosel o que ya se encuentren en etapas maduras de crecimiento [14], [15].

1.4 Justificación

Los métodos de obtención de la clorofila para analizar el estado de salud de las plantas varían en resultados, procedimientos y costos. El análisis de tejidos produce los resultados más fiables, pero se necesitan los equipos de laboratorio adecuados para poder realizarlos además de que se debe de tomar muestras de las plantas. Los dispositivos SPAD dan una lectura, de igual manera, útil, sin embargo, los costos de estos equipos están alrededor de los \$3000 y también se debe de hacer una lectura en cada planta para poder obtener la clorofila total del campo.

Los sistemas de información geográficos brindan mucha información sobre el estado de las plantas, a la vez que dan recomendaciones para obtener una mejor producción del cultivo. Por otro lado, estas aplicaciones forman parte de paquetes, donde se ofrecen servicios extras por parte de las empresas desarrolladoras de estas herramientas, además de planes mensuales para poder utilizarlos. Cabe recalcar que estos sistemas funcionan por medio de imágenes aéreas, por lo que es indispensable el uso de un VANT, imágenes satelitales o de otra forma de poder capturar fotos aéreas del cultivo.

Adicionalmente, estas cámaras deben estar en la capacidad de captar imágenes multiespectrales debido que no cualquier cámara posee las características necesarias para captar las fotos que se utilizan en este tipo de análisis. Existen empresas que ofrecen, como parte de su plan, un VANT y cámaras multiespectrales para poder utilizar las herramientas, pero esto conlleva a un incremento al momento de adquirir los servicios.

A continuación, se muestra una tabla comparativa de los métodos para obtener la clorofila de un campo. Se tomó como ejemplo de un GIS a Atlas de la compañía MicaSense, desarrolladora de la cámara que se utilizó en el proyecto [10].

Tabla 1.1 Comparativa de métodos de medición de clorofila

	Análisis de tejidos	SPAD	Atlas
VANT (\$100 - \$1000)			X
Cámara RedEdge-MX (\$5500)			X
Equipo SPAD (\$3000)		X	
Proceso Invasivo	X		
Plan			\$100 mensual
Alcance	Por planta	Por planta	Hectáreas

Como podemos observar en la tabla 1.1 las soluciones actuales requieren un alto costo de equipos para realizar análisis, como los VANTs, cámaras multiespectrales, o equipos SPAD. Además de esto, los métodos de análisis de tejidos y SPAD, son muestrales; esto quiere decir, que no se toma en cuenta el campo en su totalidad, solo parte de éste. En el caso de los GIS, a pesar de que éstos no son invasivos ni muestrales, se necesita realizar un pago mensual para usar los servicios de análisis. Por consiguiente, los objetivos planteados solucionan los inconvenientes que se presentan con las alternativas actuales.

CAPÍTULO 2

2. METODOLOGÍA

En este capítulo se describe la solución que se planteó para beneficio de la GEA y cómo se planificó su correspondiente desarrollo. Se presenta un esquema general del proyecto; se describen cada uno de los módulos que componen la solución; las tecnologías usadas para poder implementar la solución y la metodología de trabajo que se llevó a cabo durante el desarrollo de ésta. Por otro lado, también se describe la forma en la que se desarrolló la recolección de datos en el cultivo de maíz y las características de sembrío de este.

2.1 Diseño de la solución

Debido a los problemas encontrados en el capítulo uno acerca de las soluciones utilizadas actualmente en la GEA, se desarrolló un sistema para analizar imágenes multiespectrales de cultivos, en las cuales se identifican las áreas donde existe vegetación y se obtiene el nivel de clorofila de dichas áreas. Para poder llevar a cabo estas tareas, se definieron los siguientes módulos.

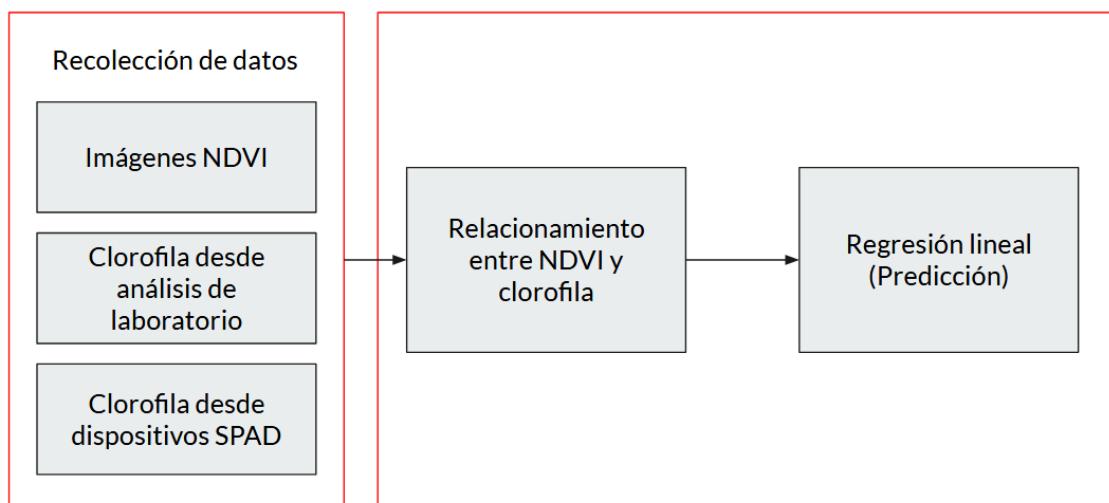


Figura 2.1 Modelo conceptual del sistema

2.1.1 Recolección de datos

Los datos necesarios para el desarrollo del sistema están compuestos por las imágenes multiespectrales aéreas del cultivo de maíz, por los valores del nivel de clorofila, obtenidos mediante estándares de laboratorio, que poseen ciertas plantas del cultivo utilizadas como muestras y mediante la toma de muestras con el dispositivo SPAD.

Para la recolección de imágenes del cultivo de maíz de la GEA se utilizó una cámara de la marca MicaSense, modelo RedEdge-M, con la cual se obtuvieron las imágenes aéreas multiespectrales del mismo. El modelo de cámara utilizado posee 5 lentes, donde cada uno es capaz de capturar imágenes en una banda espectral específica, como lo son el azul, rojo, rojo cercano, verde e infrarrojo cercano [16]. Estas imágenes de las 5 bandas espectrales fueron luego procesadas para obtener imágenes RGB e imágenes que representen valores NDVI del cultivo de maíz, que finalmente son las imágenes que se utilizaron en los siguientes módulos.

La solución se planteó de tal forma que en el proceso de recolección de datos de los valores de niveles de clorofila de las plantas sea a partir de la toma de muestras del cultivo de maíz, las cuales luego fueran analizadas mediante el método Dumas a partir del dispositivo vario MAX Cube y así obtener el nivel de clorofila de las muestras. También se planteó tomar datos con el dispositivo SPAD de diferentes plantas del cultivo sin ser necesariamente las mismas utilizadas en los análisis de laboratorio. Finalmente, debido a que el cultivo de maíz de la GEA se encontraba en etapas tempranas de crecimiento, únicamente se realizaron tomas del nivel de clorofila a partir del dispositivo SPAD. También se utilizó un conjunto de datos correspondientes al artículo *Dataset of chlorophyll content estimation of Coffea Canephora using Red and Near-Infrared consumer-grade camera* [17].

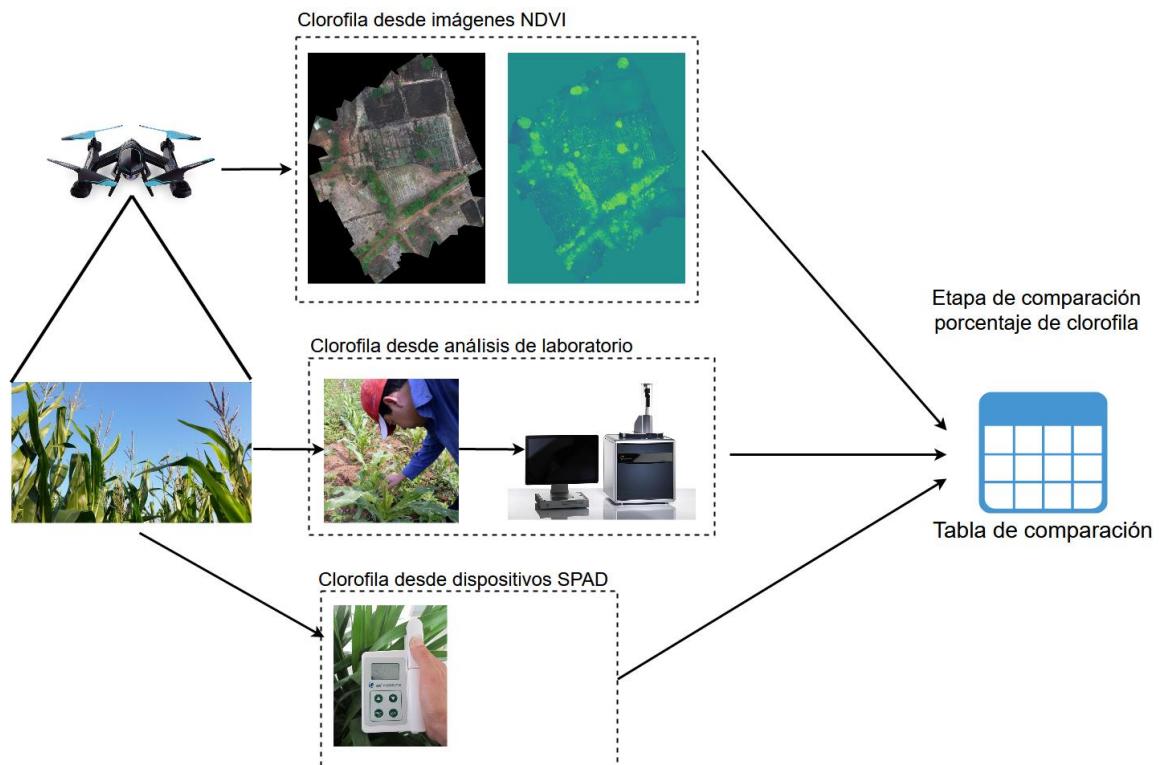


Figura 2.2 Modelo planteado de recolección de datos

2.1.2 Relacionamiento entre NDVI y clorofila

Una vez obtenidas las imágenes RGB y NDVI, se procedió a realizar el sistema de etiquetado de dichas imágenes con sus valores de clorofila correspondientes, por lo cual, se desarrolló un módulo que consiste en asociar los valores NDVI de una región de la imagen con el valor del nivel clorofila obtenido en el laboratorio y con el dispositivo SPAD.

Para realizar esto, el sistema de etiquetado de imágenes multiespectrales permite que el usuario pueda dibujar una región de interés en la imagen NDVI, para luego etiquetarla con un nivel de clorofila correspondiente a dicha región. Los valores NDVI que estén dentro de la región de interés seleccionada serán almacenados en una base de datos, relacionándolos con el nivel de clorofila ingresado por el usuario. Esto se realizó con la finalidad de obtener bases de datos que permitan realizar regresiones lineales entre los valores NDVI y los valores de clorofila de las plantas.

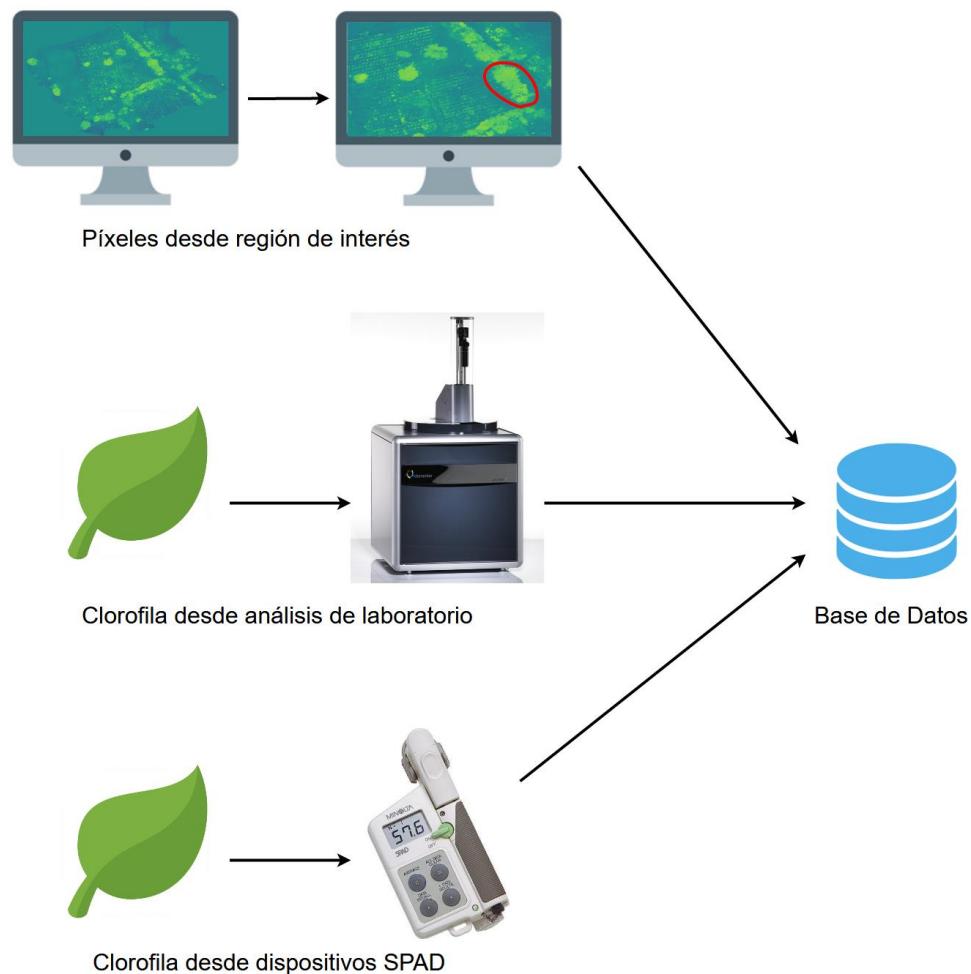


Figura 2.3 Modelo planteado de almacenamiento de datos

2.1.3 Regresión lineal

Con las imágenes obtenidas de las plantas del cultivo de maíz y sus correspondientes valores de SPAD tomados, se procedió a realizar una regresión simple que represente la correlación que existe entre los datos del cultivo de maíz.

También se realizó una segunda representación de la regresión lineal simple con el conjunto de datos del artículo mencionado en la sección de recolección de datos [17].

Este conjunto de datos posee diferentes índices de vegetación obtenidos a partir de imágenes y también a partir de un espectrofotómetro, como lo son NDVI, R, G, B, RRE, entre otros; de todos estos se utilizaron los valores NDVI y el espectrofotómetro. El artículo también brinda las funciones que relacionan los índices mencionados anteriormente con los valores SPAD correspondientes, debido a esto, se utilizó el índice capturado por el espectrofotómetro y su función de relación con el SPAD para obtener los valores teóricos de este último, mientras que los valores de NDVI se utilizaron para la obtención de la función de regresión lineal simple.

Para calcular el valor SPAD teórico se utilizó la siguiente ecuación:

$$y = 20.051 \ln(x) - 3.5718 \quad (3.1)$$

Donde x se refiere al valor obtenido por el espectrofotómetro, mientras que y se refiere al valor NDVI.

Para la regresión lineal, se consideraron como variable x al valor del píxel NDVI brindado por el conjunto de datos y como variable y al valor del nivel de SPAD teórico calculado con la función (3.1), obteniendo de esta forma una función lineal que relaciona ambas variables.

Una vez obtenida las funciones de regresión lineal, se procedió a utilizarlas para realizar la predicción del nivel de clorofila a partir de las imágenes NDVI. Se probaron distintos valores de NDVI pertenecientes a imágenes del cultivo de maíz, para así obtener los valores de nivel de clorofila SPAD relacionados a dichos valores NDVI.

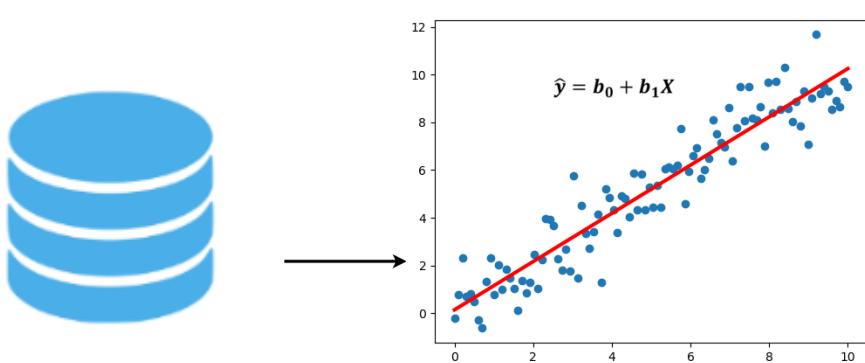


Figura 2.4 Representación de una regresión lineal

2.2 Tecnologías de desarrollo

La tecnología que se utilizó principalmente para el desarrollo de los módulos del proyecto, con excepción del módulo de recolección de datos, fue el lenguaje de programación Python, debido a que este provee librerías eficientes y optimizadas para el procesamiento de imágenes e inteligencia artificial.

A continuación, se listan las librerías que se utilizaron en el desarrollo del proyecto:

- OpenCV: Librería de código abierto utilizado para procesamiento de imágenes y visión por computador.
- PyQt: Librería para el desarrollo e implementación de interfaces gráficas.
- Scikit-learn: Librería para el desarrollo e implementación de aprendizaje automático (machine learning en inglés).
- Matplotlib: Librería para el desarrollo de gráficos a partir de datos almacenados en arreglos o listas.
- NumPy: Librería que da soporte para la utilización de arreglos multidimensionales y matrices, además de funciones matemáticas para el uso de álgebra lineal.

Las librerías de OpenCV, PyQt y NumPy se utilizaron en el módulo de almacenamiento de datos para poder desarrollar la aplicación que permite el etiquetado de imágenes del cultivo con su correspondiente valor de clorofila obtenido en el laboratorio. La librería PyQt fue utilizada para el desarrollo de la

interfaz gráfica de la aplicación, la apertura de las imágenes RGB y NDVI que visualiza el usuario y para la funcionalidad de dibujo de una región de interés por parte del usuario en la imagen NDVI. OpenCV se utilizó para el procesamiento de las regiones de interés dibujadas por el usuario en las imágenes NDVI y, finalmente, la librería NumPy fue utilizada para la obtención de los valores de los píxeles que se encuentran dentro de la región de interés dibujada por el usuario.

Para la base de datos en dónde se almacenarán los datos del sistema de etiquetado de imágenes se utilizó MongoDB, la cual es una base de datos NoSQL orientada a documentos, lo que significa que posee una mayor flexibilidad a cambios en la estructura de los datos y una fácil escalabilidad. Por otro lado, utiliza una sintaxis del tipo JSON para representar los documentos.

La librería Scikit-learn se utilizó para el desarrollo de los modelos de regresión lineal simple a partir de los datos almacenados en la base de datos MongoDB y también se la utilizó para la predicción de los niveles de clorofila de laboratorio y SPAD a partir del valor NDVI, utilizando los modelos de regresión lineal creados; mientras que la librería Matplotlib fue utilizada para la representación gráfica de las funciones lineales obtenidas en los modelos de regresión lineal simple.

2.3 Metodología

Para el desarrollo de los módulos de almacenamiento de datos y regresión lineal del proyecto se utilizó la metodología de Programación en Pareja (Pair Programming en inglés). Esta metodología consiste en tener a dos desarrolladores en una misma estación de trabajo, participando activamente en el desarrollo del sistema [18].

Los desarrolladores pueden tomar e intercambiar los siguientes dos roles:

- Driver: Encargado de escribir el código [19].
- Navigator: Encargado de revisar cada línea de código [19].

Los beneficios que supone esta metodología de trabajo es que al tener dos desarrolladores trabajando sobre la misma tarea, algoritmo, módulo, entre otros, ambos tendrán el conocimiento a fondo la estructura del código y conocerán cómo está implementada cada una de las funciones desarrolladas; también existirán menos errores y la resolución de problemas que se presenten se realizará de manera más fácil [18]. También se considera que, al ser desarrollado por dos personas a la vez, el diseño de la solución será superior a que si desarrollan el sistema con participaciones individuales [18].

La primera parte del proyecto corresponde al desarrollo del software de etiquetado de imágenes multiespectrales correspondiente al módulo de almacenamiento de datos. El desarrollo del software se dividió en tres tareas, las cuales poseen sus propias subtareas, que corresponden a la implementación de la interfaz gráfica de usuario, obtención de valores NDVI dentro de la región de interés dibujada por el usuario, relacionar el valor NDVI con el valor de clorofila obtenido en laboratorio y obtenido con el dispositivo electrónico SPAD, y luego almacenar los datos en la base de datos.

En interfaz gráfica se desarrollaron las funcionalidades relacionadas a la apertura de imágenes RGB y NDVI, la manipulación de estas con acercamiento y alejamiento de la imagen, poder seleccionar libremente una región de interés en la imagen NDVI y finalmente poder ingresar por pantalla los valores de nivel de clorofila de laboratorio y SPAD.

Para la obtención de los valores NDVI de la región de interés, se desarrolló la funcionalidad para crear una máscara a partir de la región de interés seleccionada por el usuario y utilizar la misma para obtener únicamente las posiciones de la imagen NDVI que se encuentren dentro del área de la máscara de la región de interés.

En el relacionamiento entre NDVI y clorofila se desarrolló la funcionalidad para guardar la relación que existe entre los valores NDVI obtenidos a partir de la máscara de la región de interés con los valores de nivel de clorofila obtenidos en las tomas de muestra con el dispositivo SPAD y con los obtenidos a partir del análisis en los laboratorios.

Para la captura de imágenes multiespectrales y valores de clorofila SPAD se utilizaron 100 hojas del cultivo de maíz de la GEA, luego de 5 semanas de haber sido plantado, donde cada una pertenecía a una planta diferente. A estas 100 hojas se le realizaron las tomas de su valor de clorofila con el dispositivo electrónico SPAD y también se realizaron capturas con la cámara multiespectral. Se obtuvieron un total de 500 imágenes, debido a que la cámara posee 5 lentes distintos, que luego fueron procesadas para obtener una sola imagen por hoja y sus valores en los espectros rojo e infrarrojo cercano. Estas imágenes fueron utilizadas en el sistema de etiquetado.

Una vez concluidas las pruebas del sistema de etiquetado, se procedió a obtener dos funciones de regresión lineal. Una de estas basándose en el conjunto de datos con SPAD teórico obtenido del artículo científico y también en la base de datos del sistema de etiquetado.

La última parte del proyecto corresponde al desarrollo del software de predicción de niveles de clorofila. En este software se desarrollaron funcionalidades que permitan utilizar las regresiones lineales obtenidas anteriormente para que al momento de que se seleccione un punto de la hoja, obtener el valor de clorofila predicho de esta última.

CAPÍTULO 3

3. IMPLEMENTACIÓN Y RESULTADOS

En este capítulo se describe la implementación de la solución propuesta en el capítulo anterior, los algoritmos que se utilizaron, se detalla cada uno de los componentes del sistema propuesto, y también se analizan los resultados obtenidos con el uso del sistema.

3.1 Implementación de la solución

3.1.1 Base de datos

La estructura de la base de datos se definió de tal manera que almacene información relevante con respecto de la imagen para facilitar la identificación de las imágenes que han sido etiquetadas, como también información sobre la vegetación presente en la imagen. Como la base de datos se creó usando MongoDB, se tiene una estructura de colección con la siguiente información:

```
{  
    "_id": <ObjectId>,  
    "name": "nombre_del_archivo",  
    "extension": "png",  
    "NDVI": [  
        0: 0.34905,  
        1: 0.32853  
        ...  
        9: 0.39588  
    ],  
    "SPAD": 2,  
    "LAB": 3  
}
```

Cada campo de la colección se detalla a continuación:

- Id: Campo que se genera automáticamente por MongoDB, es único para cada registro.
- Name: Nombre del archivo de la imagen que se utilizó en el sistema de etiquetado.
- Extension: Extensión del archivo de la imagen.
- NDVI: Conjunto de valores NDVI que pertenecen a una región de la imagen.
- SPAD: Valor del dispositivo SPAD asociado a la misma región de la imagen.
- LAB: Valor de nitrógeno medido en el laboratorio asociado a la misma región de la imagen.

3.1.2 Sistema de etiquetado

El sistema está diseñado para que el usuario pueda abrir una imagen que represente los valores NDVI de los cultivos y poder seleccionar las diferentes zonas de vegetación donde se han recolectado datos de clorofila con los dispositivos SPAD y también por métodos de análisis de laboratorio. Opcionalmente, el usuario puede abrir otra imagen del cultivo que le facilite la identificación de las zonas de vegetación donde se tomaron los datos de clorofila.

En la ventana principal del sistema también se muestra una breve información del archivo de imagen que se utilizará para el etiquetado, como es el nombre del archivo en el título de la ventana, la resolución en píxeles de cada una de las imágenes y el tamaño del archivo en memoria.



Figura 3.1 Ventana principal del sistema de etiquetado

La pantalla posee un botón llamado *Seleccionar* que activa la función de selección en el que el usuario puede dibujar una región de interés en la imagen NDVI que se encuentra a la derecha del sistema. Esta región puede ser de forma irregular y es la que encierra los valores NDVI que el usuario quiera asociar con los valores de clorofila obtenidos mediante la utilización del dispositivo SPAD y análisis de laboratorio, los cuales podrán ser ingresados después de haber sido dibujada la región.

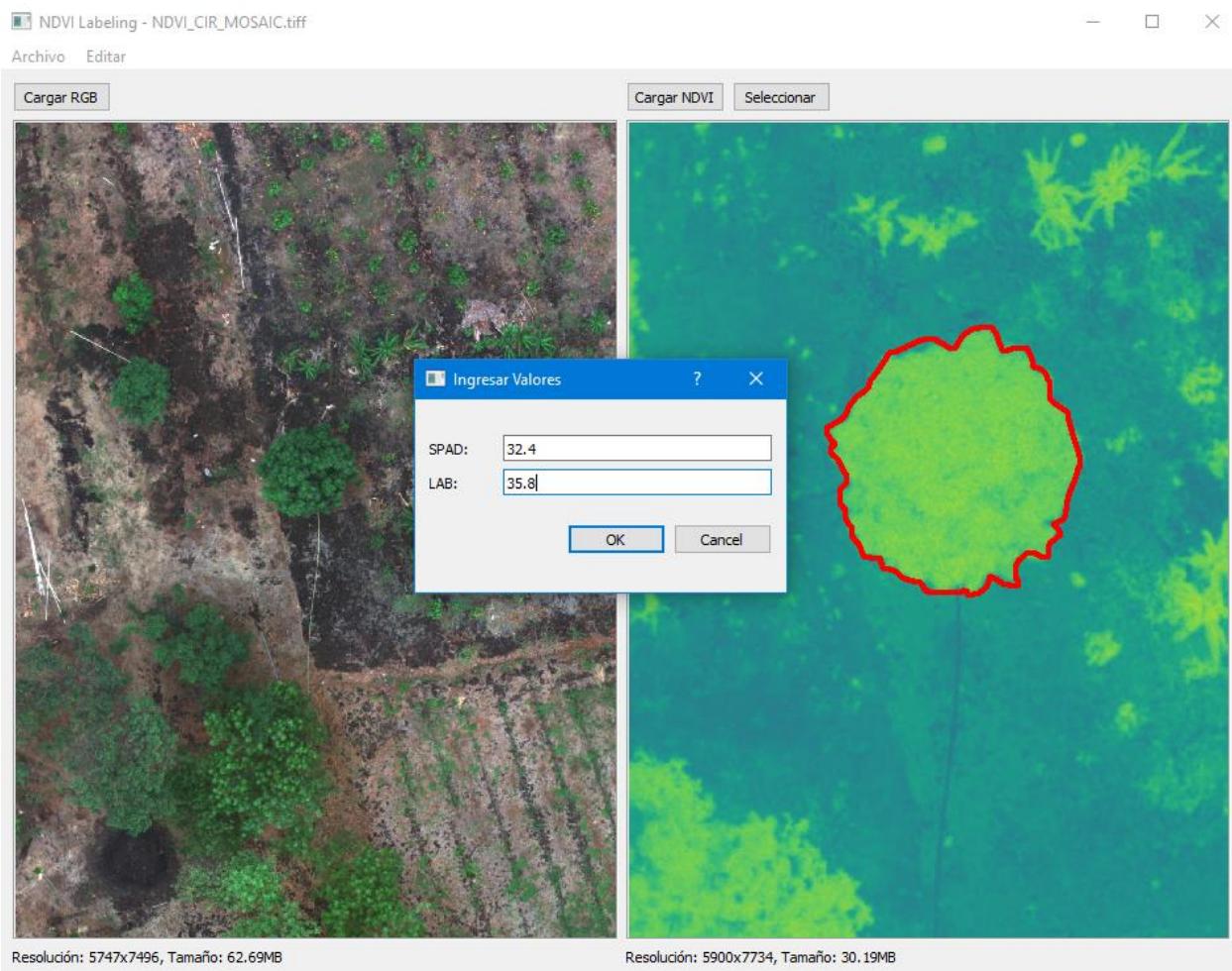


Figura 3.2 Ingreso de valores de SPAD y de laboratorio

La forma en que se obtiene los valores NDVI dentro de la región de interés es por medio de la librería OpenCV. Este proceso conlleva en primer lugar binarizar (blanco y negro) la imagen para poder identificar la zona de la selección. Se define un umbral con el color utilizado al dibujar la región y con esto se genera una máscara de la región dibujada.

La máscara sirve para identificar qué elementos pertenecen a la región seleccionada por el usuario. La región dibujada aparece de color blanco mientras que lo que está fuera, de color negro.

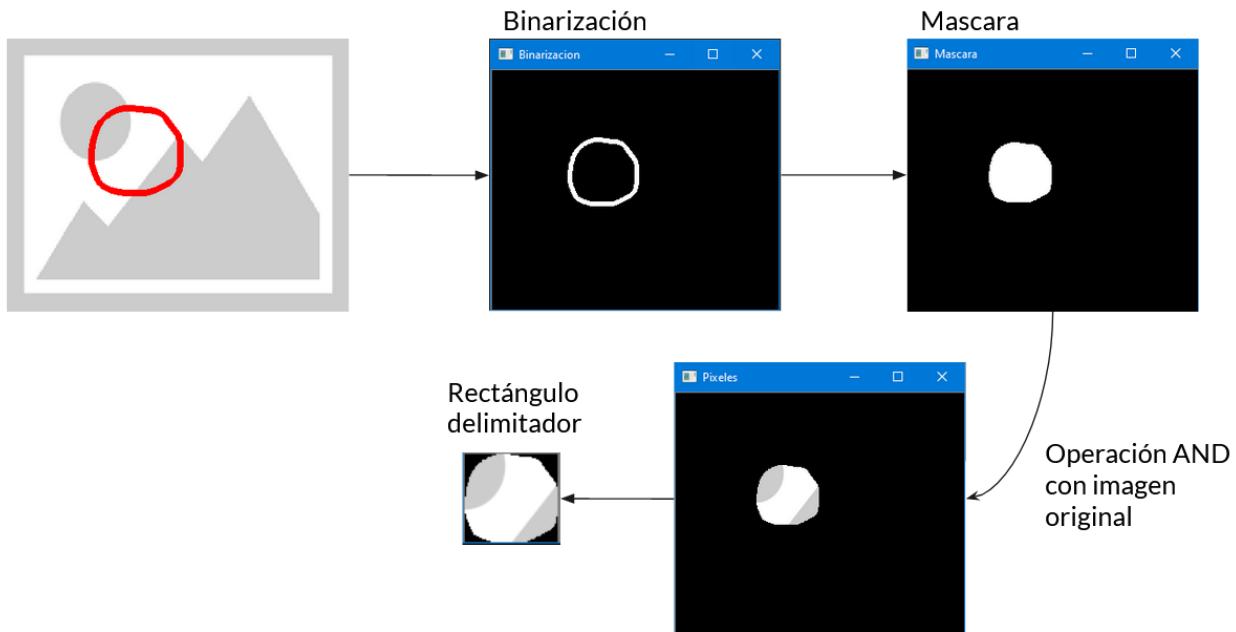


Figura 3.3 Procedimiento para identificar la región de interés

Una vez que el usuario acepta ingresar los valores de clorofila obtenido mediante el dispositivo SPAD y análisis de laboratorio, se calculan los valores de NDVI pertenecientes a la región dibujada, utilizando el rango espectral rojo y el infrarrojo cercano de la ecuación (1.1). Los valores de estos rangos espectrales se encuentran almacenados en archivos de extensión npy, que deben residir en el mismo directorio de la imagen. Utilizando la máscara se obtiene los valores de infrarrojo cercano y rojo pertenecientes a la región dibujada por el usuario.

Una vez que se tiene los valores de NDVI de la región dibujada calculados, se crea un nuevo registro en la base de datos donde se relaciona dichos valores NDVI con las mediciones de clorofila ingresadas por el usuario. Finalmente, se agrega una etiqueta en la imagen en la región que se seleccionó mostrando los valores de clorofila ingresados y se cambia el color del borde de la región por azul, esto para identificar zonas ya etiquetadas en la imagen.

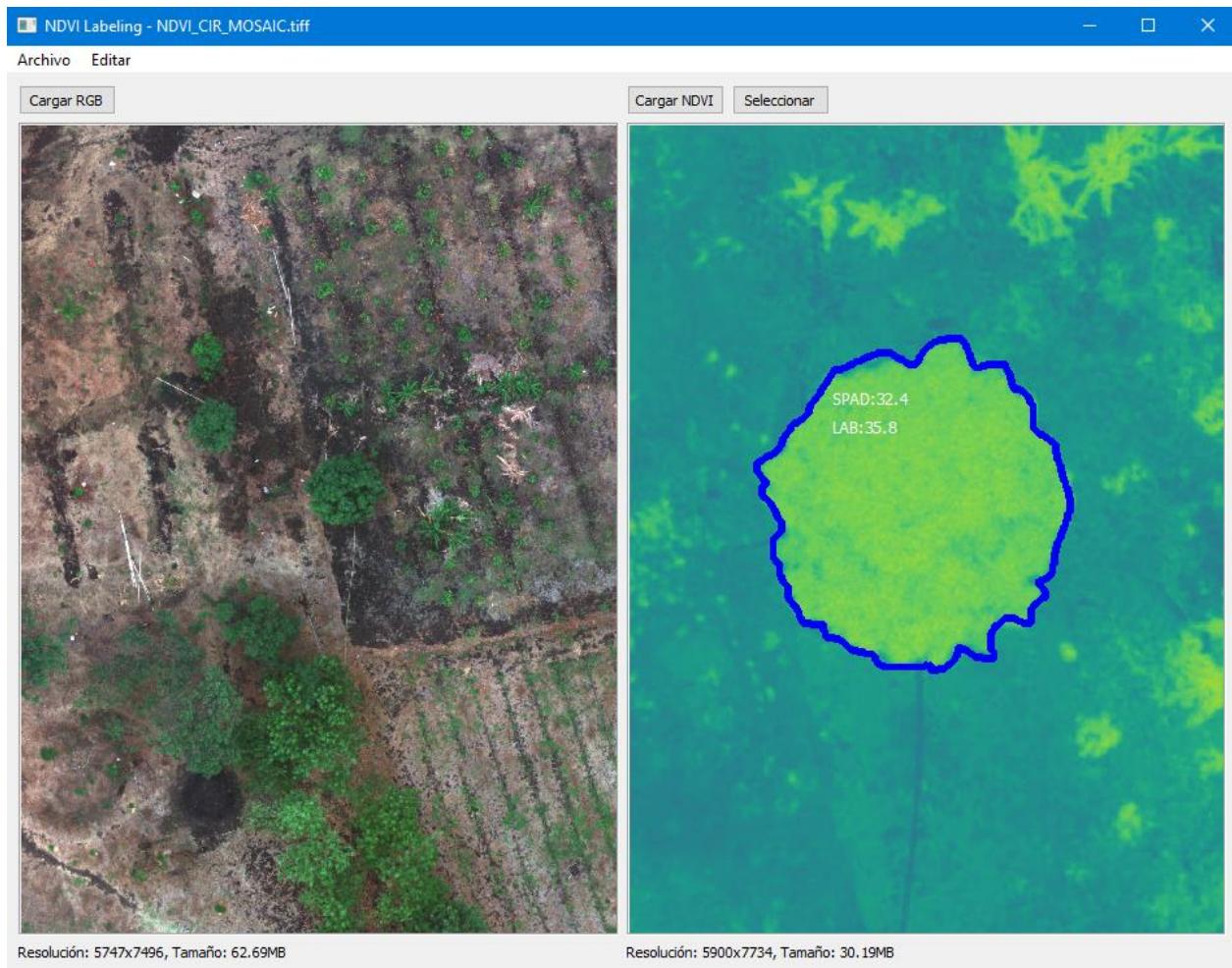


Figura 3.4 Etiquetado de la región de interés

3.1.3 Regresión lineal

Debido a que los valores NDVI están comprendidos en un rango de [-1, 1], se filtraron estos valores de tal manera que sólo se utilicen los valores que representan vegetación. Según el rango donde se ubique el valor de NDVI se pueden tener aproximaciones de sus características.

Según lo que representan [20]

- [-1, -0.1]: Agua
- (-0.1, 0.1]: Rocas, arena o nieve
- (0.1, 0.4]: Plantaciones jóvenes, arbustos, matorrales
- (0.4, 0.9]: Vegetación abundante

Según la densidad de la región [21]:

- [-1, 0]: Sin vegetación
- (0, 0.15]: Densidad muy baja
- (0.15, 0.3]: Densidad baja
- (0.3, 0.45]: Denso
- (0.45, 0.6]: Densidad alta
- (0.6, 1]: Densidad muy alta

Basándose en estas aproximaciones, se decidió tomar los valores de NDVI mayores a 0.1 ya que estos representan desde una vegetación joven hasta una vegetación abundante.

También se realizaron las estructuras de los datos de valores NDVI y de valores SPAD, almacenándolos en arreglos de tal manera que sean compatibles con los tipos de datos utilizados por la librería Scikit-learn de Python. Esta librería fue utilizada para obtener la función de regresión lineal simple entre los valores NDVI y los valores SPAD teóricos, el coeficiente de correlación lineal entre ambas variables y la raíz del error cuadrático medio.

3.1.4 Sistema de predicción

Para poder comprobar que la regresión lineal simple funciona correctamente, se implementó una herramienta de predicción en la cual el usuario puede cargar una imagen perteneciente a algún cultivo o que tenga zonas de vegetación y obtener los valores de clorofila de las plantas que aparecen en la imagen. Esto se realiza seleccionando cualquier parte de la imagen para que el sistema brinde el valor de clorofila en dicha selección.

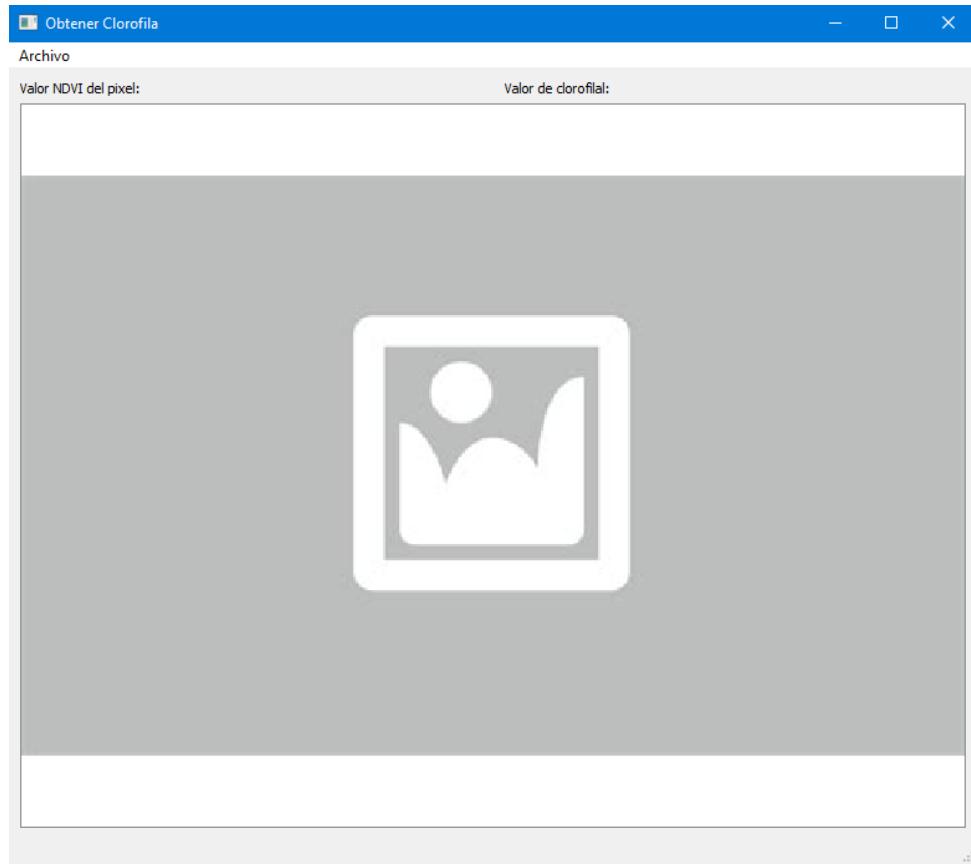


Figura 3.5 Ventana principal del sistema de predicción

Para poder realizar esta funcionalidad, primero se obtienen las coordenadas del lugar que el usuario seleccionó con respecto a la imagen cargada. Luego, se calcula el valor de NDVI en esa posición tomando los valores del rango espectral rojo e infrarrojo cercano de los archivos en la misma posición. Este valor de NDVI se envía a la función *predict* de la regresión lineal simple obtenida previamente, para finalmente presentar en la parte superior la información del valor de clorofila predicho.

Es importante aclarar que el tamaño de la imagen y el tamaño de los arreglos de rangos espectrales rojo y el infrarrojo cercano deben de ser iguales para poder tomar los valores correspondientes al lugar que el usuario seleccionó.

3.2 Análisis de resultados

Finalizada la implementación del sistema de etiquetado de imágenes, la regresión lineal simple y del sistema de predicción, se procedió a realizar una prueba con los datos que se tenían previamente. En el sistema de etiquetado de imágenes y de predicción se realizaron pruebas prácticas y en la regresión lineal simple una prueba teórica.

3.2.1 Prueba del sistema de etiquetado de imágenes

Para validar que el sistema de etiquetado de imágenes funciona correctamente se utilizaron 100 imágenes que se capturaron del cultivo de maíz en la GEA, donde cada imagen contiene una hoja de una planta distinta del cultivo. La siguiente imagen de muestra fue capturada cuando las semillas tenían una semana de haber sido plantadas.



Figura 3.6 Imagen del cultivo de maíz en GEA

El etiquetado de imágenes se hizo en una computadora con 8GB de RAM y esto limita la cantidad de datos que pueden ser procesados al momento de hacer la regresión lineal. De las pruebas realizadas, se observó que en cada una de las hojas etiquetadas existen alrededor de 150000 valores NDVI y debido a que son 100 hojas, una en cada imagen, se tenía más de 15 millones de datos. Debido a eso, se decidió que, al momento de etiquetar las hojas que aparecen en las imágenes, la región de interés sea menor al área total de la hoja, como se observa en la figura 3.7. De esta forma se reduce la cantidad de datos que finalmente son procesados en la regresión lineal.

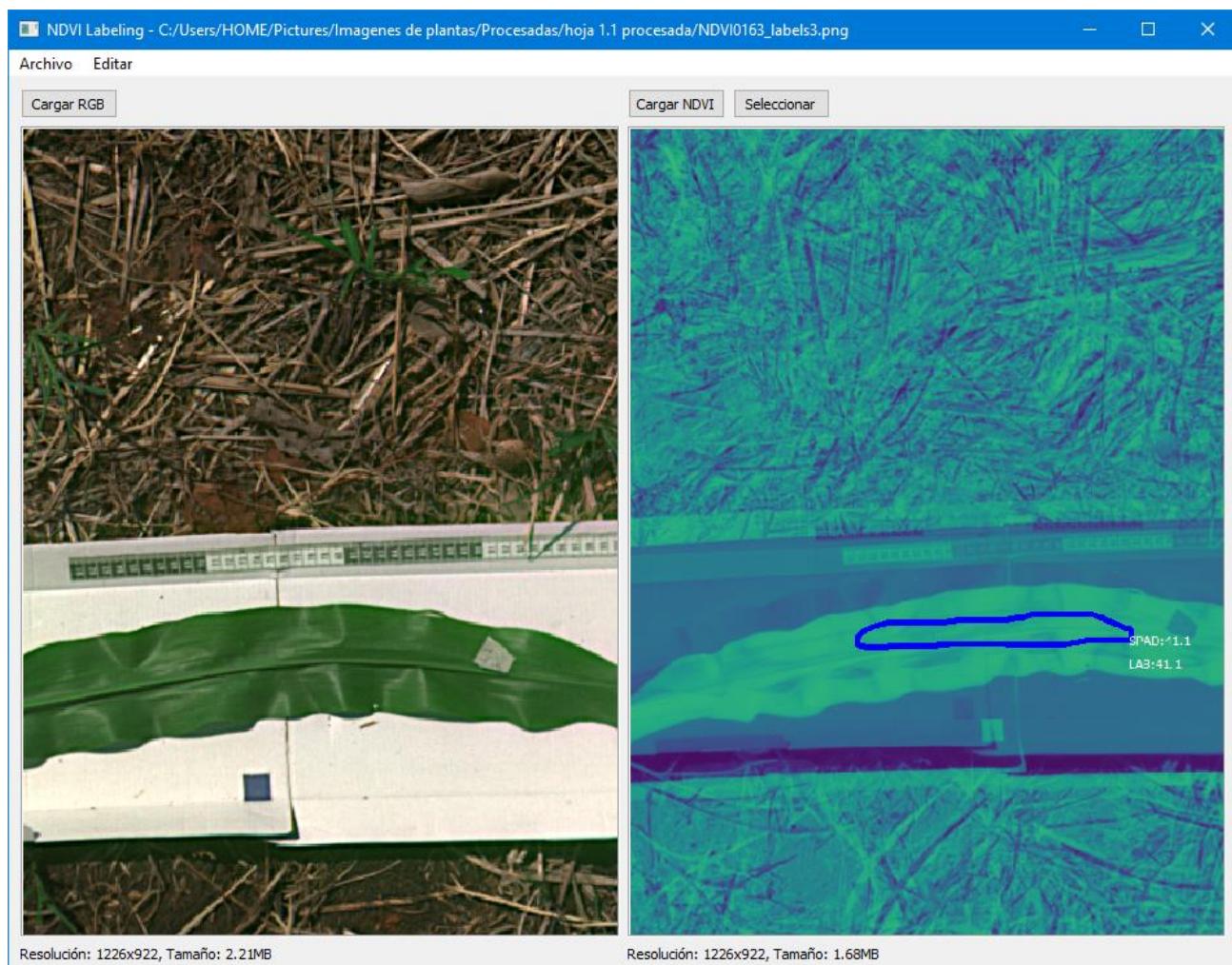


Figura 3.7 Etiquetado de las plantas del cultivo

La información que se almacenó en la base de datos a partir de esta imagen se muestra en la siguiente tabla. Los campos NDVI MIN y NDVI MAX fueron agregados solo para visualización de los datos ya que el campo NDVI tiene una cantidad muy grande de elementos para mostrar. Los campos correspondientes a la clorofila obtenida mediante el análisis de laboratorio fueron omitidos ya que, por motivos de retraso en el inicio del cultivo de maíz y porque el cultivo se encontraba en una etapa muy joven de crecimiento, estos valores no pudieron ser obtenidos y se consideraron como 0.

Tabla 3.1 Representación de la base de datos

Nombre	Extensión	NDVI	NDVI MIN	NDVI MAX	NDVI AVG	SPAD
NDVI0434	png	1844 elementos	0.5737901329994 202	0.77697020769 11926	0.69942983086 35224	32.2
NDVI0437	png	2708 elementos	0.6066501140594 482	0.80950927734 375	0.72107878956 87222	37.3
NDVI0440	png	2052 elementos	0.5420783758163 452	0.76918220520 01953	0.65276326770 67093	37.4
NDVI0443	png	1928 elementos	0.7805064916610 718	0.92617207765 57922	0.87607061071 51459	36.6
NDVI0446	png	900 elementos	0.8665105700492 859	0.92999970912 93335	0.89974493318 13389	38.3
NDVI0449	png	2656 elementos	0.9469475150108 337	0.97080206871 03271	0.96001299865 6336	31.6
NDVI0452	png	2072 elementos	0.5885360240936 279	0.78045779466 62903	0.70666326864 34565	31.7
NDVI0455	png	860 elementos	0.6065247654914 856	0.80987644195 55664	0.70965491934 69845	33.2

3.2.2 Primera prueba de la regresión lineal simple con SPAD teórico

En la primera prueba realizada para la obtención de la regresión lineal simple se decidió utilizar un conjunto de datos correspondientes al artículo *Dataset of chlorophyll content estimation of Coffea Canephora using Red and Near-Infrared consumer-grade camera* [17]. Con este artículo se obtuvo un valor de SPAD teórico a partir de la función (3.1) para poder realizar las relaciones con el valor NDVI de las plantas.

Una vez filtrada la información quedó la siguiente cantidad de datos para utilizar.

Datos NDVI:

- 600 datos NDVI en total en el conjunto de datos.
- 585 datos NDVI utilizados una vez filtrados los datos.

Datos SPAD:

- 600 datos SPAD teóricos en total.
- 585 datos SPAD teóricos utilizados una vez filtrado los datos.

Con el conjunto de datos filtrado, se procedió a representarlos gráficamente.

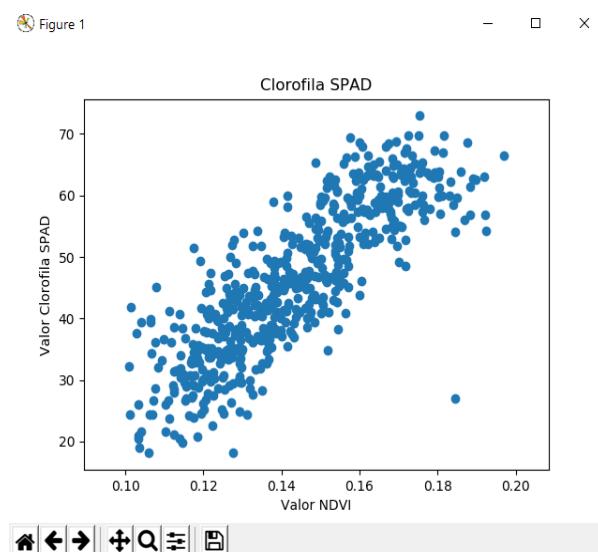


Figura 3.8 Representación de los datos de la primera prueba con SPAD teórico

Se observa que los datos en la gráfica están distribuidos linealmente, por lo cual se procedió a realizar la regresión lineal simple en la que se obtuvo la siguiente ecuación:

$$y = 44.75355778 x - 25.16372078443378 \quad (3.1)$$

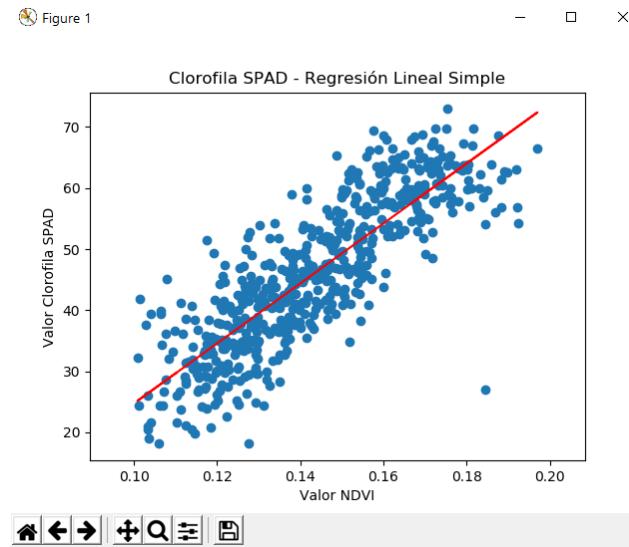


Figura 3.9 Primera prueba de regresión lineal simple con SPAD teórico

Una vez obtenida la función de regresión lineal simple, se procedieron a calcular el coeficiente de correlación lineal y la raíz del error cuadrático medio, a partir de las funciones implementadas en la librería Scikit-learn.

El coeficiente de correlación lineal fue igual a:

$$0.7054842041300065$$

Lo que indica que existe una relación del 70.54% entre los valores NDVI y los valores medidos mediante el dispositivo SPAD.

La raíz del error cuadrático medio fue igual a:

$$6.581877160210804 \text{ unidades SPAD}$$

Lo que indica que los valores predichos de SPAD obtenidos con la función de regresión lineal tienen un error promedio de ese valor con respecto a los valores reales SPAD utilizados para entrenar el modelo de la regresión lineal simple.

Tabla 2.2 Comparación de resultados de la primera prueba de regresión lineal

Valor NDVI	SPAD	SPAD Predicho
0.725708323	58.00058626978684	57.64175016
0.904705921	62.28469041802716	65.65252951
0.74350862	62.90889918374017	58.43837678
0.798948626	62.06036415601634	60.91951429

3.2.3 Segunda prueba de la regresión lineal simple con SPAD manual

En la segunda prueba realizada para la obtención de la regresión lineal simple se utilizó la base de datos obtenida con el sistema de etiquetado de imágenes, donde las plantas pertenecen al cultivo de maíz de la GEA y el valor de SPAD fue capturado manualmente.

Con el conjunto de datos filtrado, se procedió a representarlos gráficamente.

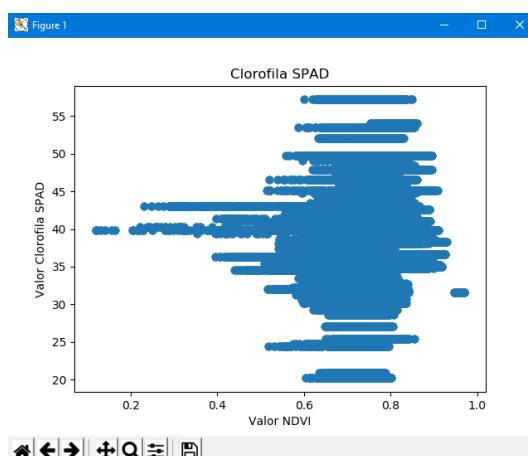


Figura 3.10 Representación de los datos de la segunda prueba con SPAD manual

Se procedió a realizar la regresión lineal simple en la que se obtuvo la siguiente ecuación:

$$y = 13.12331649 x - 28.481211104893845 \quad (3.2)$$

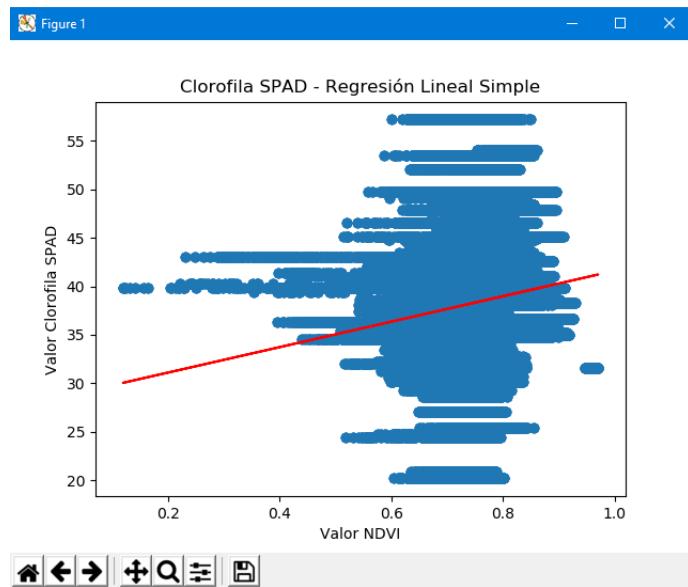


Figura 3.11 Segunda prueba de regresión lineal simple con SPAD manual

Una vez obtenida la función de regresión lineal simple, se procedieron a calcular el coeficiente de correlación lineal y la raíz del error cuadrático medio, a partir de las funciones implementadas en la librería Scikit-learn.

El coeficiente de correlación lineal fue igual a:

$$0.01861993856531241$$

Lo que indica que existe una relación del 1.08% entre los valores NDVI y los valores medidos mediante el dispositivo SPAD.

La raíz del error cuadrático medio fue igual a:

$$6.477188141888954 \text{ unidades SPAD}$$

Lo que indica que los valores predichos de SPAD obtenidos con la función de regresión lineal tienen un error promedio de ese valor con respecto a los valores reales SPAD utilizados para entrenar el modelo de la regresión lineal simple.

Tabla 3.3 Comparación de resultados de la segunda prueba de regresión lineal

Valor NDVI	SPAD	SPAD Predicho
0.7012959	34.8	37.68453932
0.8696189	43	39.89349506
0.8162994	41.1	39.19376621
0.69903266	36.3	37.654838

3.2.4 Tercera prueba de la regresión lineal simple con SPAD manual

En la tercera prueba realizada para la obtención de la regresión lineal simple se volvió a utilizar la base de datos obtenida con el sistema de etiquetado de imágenes, pero en esta ocasión, se promediaron los valores NDVI pertenecientes a la región de interés, obteniendo 92 valores NDVI en total.

Con el conjunto de datos filtrado, se procedió a representarlos gráficamente.

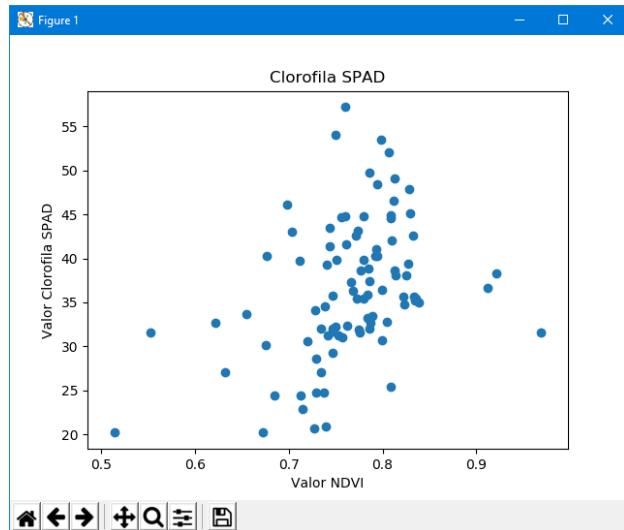


Figura 3.12 Representación de los datos de la tercera prueba con SPAD manual

Se procedió a realizar la regresión lineal simple en la que se obtuvo la siguiente ecuación:

$$y = 43.73495324x - 2.9578003944927787 \quad (3.3)$$

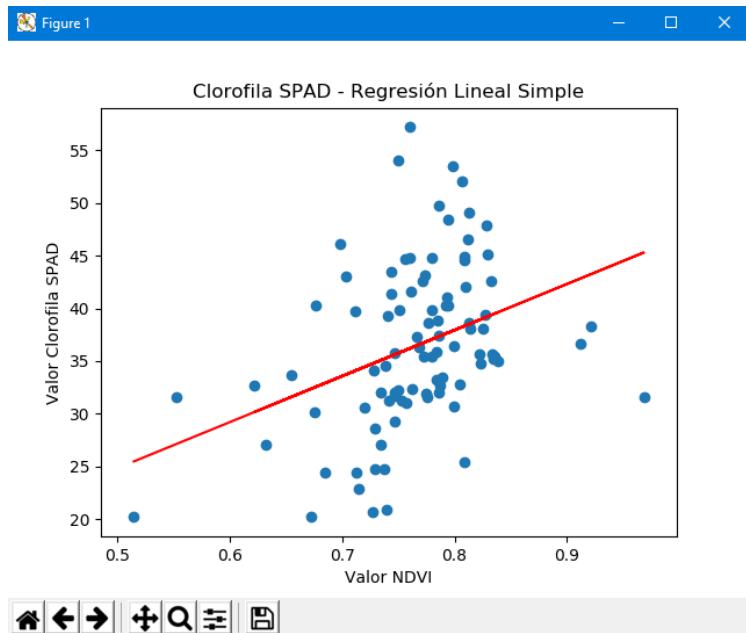


Figura 3.13 Tercera prueba de regresión lineal simple con SPAD manual

Una vez obtenida la función de regresión lineal simple, se procedieron a calcular el coeficiente de correlación lineal y la raíz del error cuadrático medio, a partir de las funciones implementadas en la librería Scikit-learn.

El coeficiente de correlación lineal fue igual a:

0.13196709865097833

Lo que indica que existe una relación del 13.20% entre los valores NDVI y los valores medidos mediante el dispositivo SPAD.

La raíz del error cuadrático medio fue igual a:

7.27287620832889 unidades SPAD

Lo que indica que los valores predichos de SPAD obtenidos con la función de regresión lineal tienen un error promedio de ese valor con respecto a los valores reales SPAD utilizados para entrenar el modelo de la regresión lineal simple.

Tabla 3.4 Comparación de resultados de la tercera prueba de regresión lineal

Valor NDVI	SPAD	SPAD Predicho
0.7012959	34.8	36.78035163
0.8696189	43	40.48576784
0.8162994	41.1	40.35309443
0.69903266	36.3	35.53141767

3.2.4 Prueba del sistema de predicción

Una vez que se obtuvieron las dos regresiones lineales simples, basada en SPAD teórico y en SPAD manual, se pasó a probar el sistema de predicción. Para esto se utilizaron las 100 imágenes pertenecientes al cultivo de maíz de la

GEA debido a que, de las 100 hojas, que se encuentran una en cada imagen, se tiene el valor real de clorofila SPAD, lo que permite realizar una comparación entre dicho valor y los predichos que se obtuvieron con las regresiones lineales. También se utilizaron 200 archivos que representan los arreglos con los valores del espectro infrarrojo y del espectro rojo, necesarios para calcular el valor del NDVI, según la fórmula (1.1). Estas pruebas fueron realizadas con la regresión lineal simple con el SPAD teórico y manual.

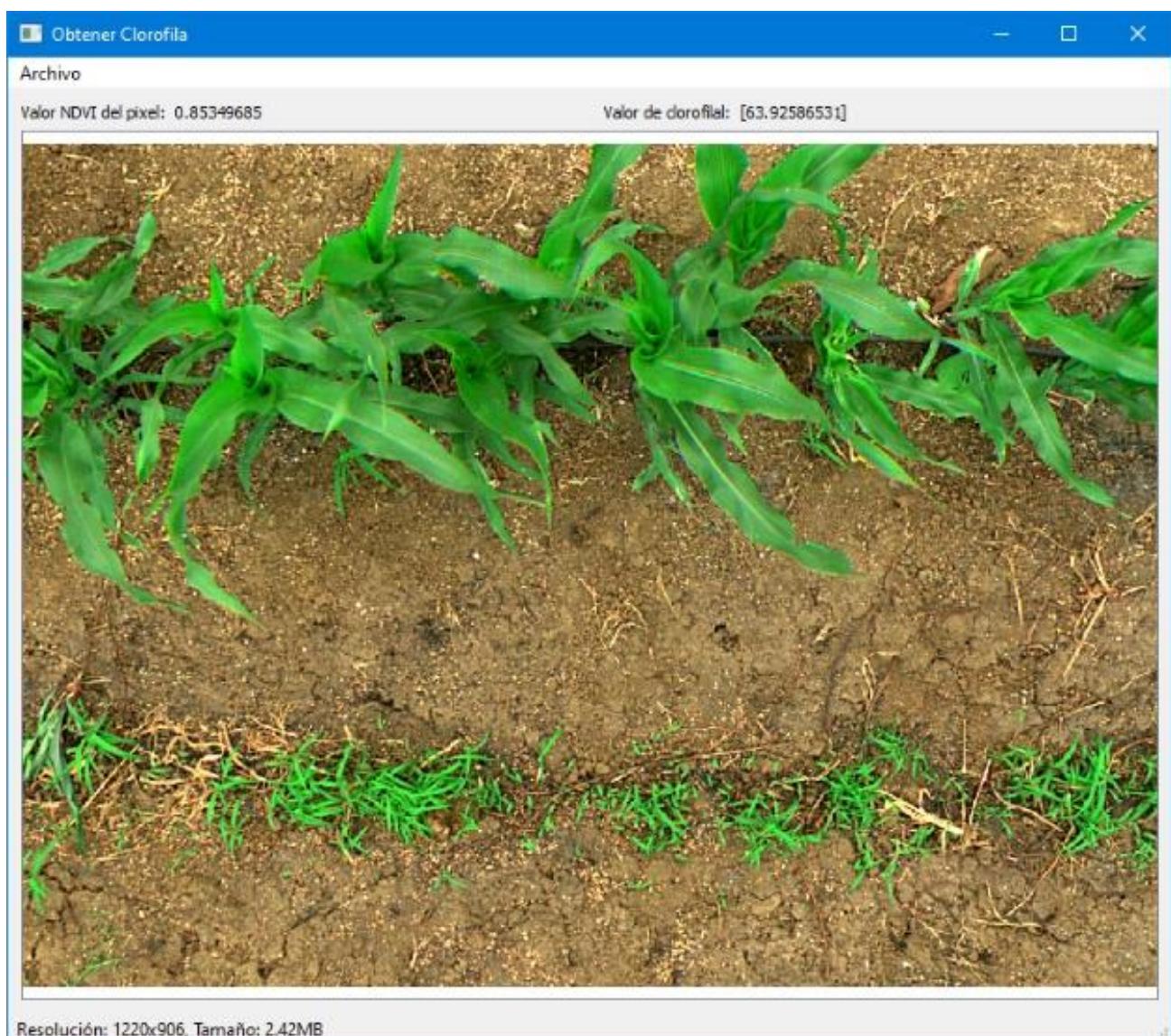


Figura 3.14 Ventana principal del sistema con imagen del cultivo

Luego de que se abre la imagen, se puede seleccionar cualquier posición de la imagen y se muestra el valor de NDVI de esa posición en particular y luego el valor de clorofila utilizando la regresión lineal calculada anteriormente.

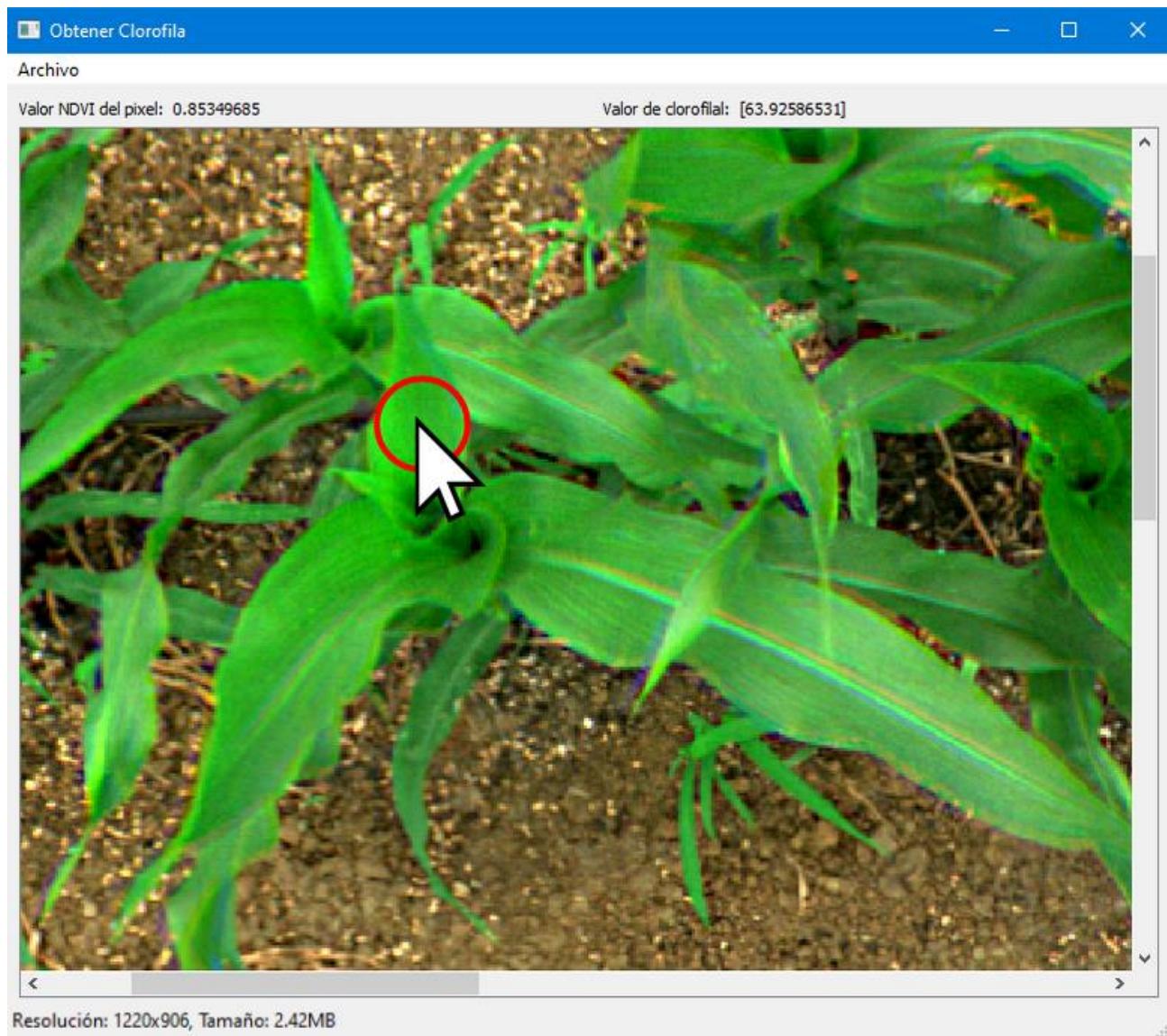


Figura 3.15 Valor de NDVI del píxel y de clorofila usando la regresión lineal

En el apartado de apéndices se encuentran distintas pruebas realizadas con el sistema de predicción de nivel de clorofila y sus resultados.

En la siguiente tabla se representan 10 valores de nivel de clorofila obtenidos utilizando las regresiones lineales simples mencionadas previamente y su respectiva relación con el valor NDVI y el valor real de clorofila SPAD.

Tabla 3.5 Comparación de resultados de sistema de predicción

Valor NDVI	SPAD real	SPAD predicho (Regresión 1 Teórica)	SPAD predicho (Regresión 2 Manual)	SPAD predicho (Regresión 3 Manual)
0.6682494	38.1	55.07025892	37.25085985	38.44229234
0.5511283	38.1	49.828673	38.61825083	37.69348461
0.82342315	42.6	62.0148363	39.28725367	40.89074824
0.864155	47.9	63.8377315	39.8217906	41.14669231
0.6238252	35.7	53.08211791	36.66786654	41.24439288
0.5811929	35	51.17417081	36.10838962	35.64828601
0.5918739	35.4	51.65218356	36.24855941	35.20947708
0.61275303	35.2	52.58659891	36.5225631	36.07730117
0.6876595	35.7	55.93892995	37.50558439	35.95191117
0.9160053	45.1	66.1582169	40.50223874	42.59750256

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

En el desarrollo de este proyecto, se creó un sistema de etiquetado de imágenes con los valores de clorofila medidos por el usuario, el cual facilita la captura de datos de NDVI y de clorofila mediante dispositivos SPAD y análisis de laboratorio, almacenando en la base de datos su relación. También se desarrolló un código que permite la lectura de esta base de datos para obtener una regresión lineal simple entre los valores de índice de vegetación y valores de clorofila, la correlación entre estas dos variables y la raíz del error cuadrático medio de la regresión lineal simple. Por último, se desarrolló un sistema de predicción que utiliza la regresión lineal simple para obtener fácilmente el valor de clorofila de las plantas a partir de imágenes pertenecientes a cultivos, de las cuales se tienen sus valores espectrales rojo e infrarrojo cercano.

4.1 Conclusiones

Para la utilización del sistema de etiquetado de imágenes y clorofila, es importante que el usuario pueda dibujar libremente las regiones de interés que quiera seleccionar con formas irregulares para así evitar, en la mayor posibilidad, la captura de datos de valores NDVI de la imagen que no correspondan a los valores de clorofila que el usuario finalmente ingresará, como son, zonas que representen agua, tierra, piedras, fuego, entre otras.

Con respecto a la obtención de la función de regresión lineal simple, se concluye que es importante realizar un filtrado en los datos que se encuentran almacenados en la base de datos al momento de realizar las regresiones lineales simples, esto se realiza como un segundo filtro para evitar el uso de cualquier tipo de dato de valores NDVI de las imágenes que no correspondan a zonas de vegetación, en caso de que el usuario hubiese cometido errores al momento de seleccionar las regiones de interés en las imágenes que fueron etiquetadas.

Los resultados obtenidos previamente con la regresión lineal simple usando el SPAD teórico indican que los datos utilizados de NDVI y SPAD se distribuyen de forma lineal entre ellos y que, debido a que el coeficiente de correlación lineal entre los valores NDVI y SPAD obtenidos en las regresiones lineales simples son mayores a 0, existe una correlación lineal positiva entre estas dos variables, con lo que se concluye que son directamente proporcionales entre sí, es decir, si el valor NDVI aumenta, el valor SPAD también, lo mismo si disminuye. También, se concluye que existe una correlación significativa entre los valores NDVI y los valores SPAD debido a que el valor absoluto del coeficiente de correlación entre ambas variables es mayor a 0.7 [22], lo cual lo coloca en el rango de relaciones significativas.

Con respecto a los resultados obtenidos con la segunda y tercera prueba de la regresión lineal simple usando el SPAD manual, se concluye que, debido a que las hojas se encuentran en un estado de salud parecido ya que se encontraban en una etapa de 5 semanas de haber sido plantadas, la mayoría de los datos se distribuyeron en los mismos rangos de NDVI en donde los valores SPAD son muy cercanos entre sí, como se observa en la figura 3.10 y 3.12. Esto dificulta el entrenamiento debido a que no se obtiene un gran rango de valores diferentes de NDVI al igual que de valores SPAD, por lo que la regresión lineal simple obtenida, garantiza una buena predicción en un mismo rango de valores, es decir, en un mismo estado de salud de las plantas.

Una vez obtenidas las regresiones, se las comparó utilizando el cultivo de maíz de la GEA. En la tabla 3.4 se puede observar cómo los valores predichos por la regresión lineal simple entrenada con el conjunto de datos del artículo de café tienen valores que no guardan ninguna relación con respecto al valor real de clorofila, mientras que las regresiones lineales entrenadas con el mismo cultivo de maíz sí tienen valores muy cercanos. Se concluye que es importante que el cultivo con el cual está entrenado la regresión lineal sea el mismo que los valores de clorofila que se desea predecir, igualmente que las condiciones del cultivo.

4.2 Recomendaciones

Se recomienda que el sistema de etiquetado de imágenes permita al usuario guardar más de un valor de clorofila del mismo tipo, obtenido por métodos estándares de laboratorio, en las regiones de interés elegidas por el usuario. Esto, para facilitar el ingreso de diferentes valores en una misma región de interés, debido a que en la actualidad se realiza un promedio de los datos tomados en la región de interés dibujada, con esto se puede realizar más de una aproximación por región.

Se recomienda que el sistema detecte automáticamente las regiones que pertenecen a zonas de vegetación dentro de la región de interés dibujada por el usuario, esto para mejorar la precisión con la que se seleccionan los valores de NDVI pertenecientes a la imagen ya que en las pruebas realizadas se obtenían valores no pertenecientes a la planta si se trataba de cubrir manualmente toda el área de la hoja.

Debido a que sólo se realizó una toma de valores de clorofila mediante el dispositivo electrónico SPAD y a que los datos teóricos de SPAD a partir de una función de relación entre espectrofotómetro y SPAD ya poseen un error en su predicción, se recomienda realizar más tomas de niveles de clorofila reales de las plantas, esto para obtener una regresión lineal más robusta, con más datos de entrenamiento.

También, se recomienda que las tomas realizadas de nivel de clorofila en las plantas y las capturas de imágenes multiespectrales sean en diferentes etapas de crecimiento de las plantas y en diferentes condiciones climáticas, para poder comparar diferentes aproximaciones en los modelos de regresión lineal y que el sistema brinde una mejor precisión al momento de obtener un valor de clorofila.

4.1 Trabajos futuros

Debido a la culminación del tiempo disponible para la realización de este documento sólo se realizó una prueba con respecto a la toma de valores NDVI y SPAD, por lo tanto, se continuarán realizando más pruebas en diferentes etapas de crecimiento de las plantas del cultivo de maíz para el desarrollo de un artículo científico en dónde se pueda cubrir un mayor rango de valores NDVI y SPAD.

Como trabajos futuros, se puede escalar todo el proyecto para la utilización de diferentes índices de vegetación y valores de clorofila para validar el aprendizaje automático entre estos y no solamente validar valores de NDVI y SPAD. El sistema de etiquetado se implementaría con otros tipos de valores, por ejemplo, NDRE, mencionado en el primer capítulo, entre otros [23], con lo cual la regresión lineal simple también se crearía a partir de estos nuevos valores.

Otro aspecto es que se podría utilizar la base de datos que brinda el sistema de etiquetado de imágenes para desarrollar un código que encuentre modelos de regresión no lineales entre los índices de vegetación y los valores de clorofila, y también utilizarla para realizar un aprendizaje automático supervisado a partir de una red neuronal.

BIBLIOGRAFÍA

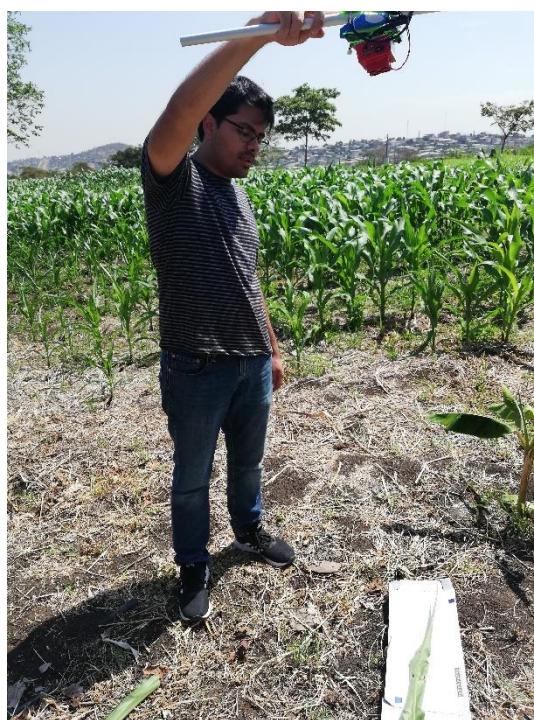
- [1] M. Mendoza-Elos, C. Mosqueda-Villagómez, J. A. Rangel-Lucio, A. López-Benítez, S. A. Rodríguez-Herrera, L. Latournerie-Moreno and E. Moreno-Martinez, "Densidad de población y fertilización nitrogenada en la clorofila, materia seca y rendimiento de maíz normal y QPM," *Agricultura técnica en México*, vol. I, no. 32, pp. 89-99, 2006.
- [2] H. Sainz Rozas and H. E. Echeverría, "Sainz Rozas, Hernán, and Hernán E. Echeverría. "Relación entre las lecturas del medidor de clorofila (Minolta SPAD 502) en distintos estadios del ciclo del cultivo de maíz y el rendimiento en grano.," *Revista de la Facultad de Agronomía* , p. 103 , 1998.
- [3] R. F. Muñoz-Huerta, R. G. Guevara-Gonzalez, L. M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez and R. V. Ocampo-Velazquez, "A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances," *Sensors*, vol. 8, no. 13, 2013.
- [4] "The SPAD," [Online]. Available: <http://www.gisiberica.com/MEDIDOR%20DE%20CLOROFILA/MCL502.htm>.
- [5] R. Novoa S.-A. and N. Villagrán A., "EVALUACIÓN DE UN INSTRUMENTO MEDIDOR DE CLOROFILA EN LA DETERMINACIÓN DE NIVELES DE NITRÓGENO FOLIAR EN MAÍZ," *Agricultura Técnica*, vol. 62, no. 1, pp. 166-171, 2002.
- [6] N. Reussi Calvo, H. Echeverría and H. Sainz Rozas, "COMPARACIÓN DE MÉTODOS DE DIAGNÓSTICO DE DEFICIENCIA DE AZUFRE EN TRIGO," 2019.
- [7] C. Cori, A. de V. C, M. Ruiz, Z. M, L. Castillo, E. J, A. E, C. C, A. B, G. O, D. L, P. A, A. I and S. L, "Definición de los métodos para analizar nitrógeno total en fertilizantes," *VENESUELOS*, no. 6, pp. 33-38, 1998.
- [8] B. Bojović and A. Marković, "Correlation between nitrogen and chlorophyll content in wheat (Triticum aestivum L.)," *Kragujevac Journal of Science*, no. 31, pp. 69-74, 2009.
- [9] "vario MAX cube," [Online]. Available: <https://www.elementar.de/en/products/organic-elemental-analysis/vario-max-cube.html>.
- [10] "Cuando los datos se convierten en información," [Online]. Available: <https://www.micasense.com/es/atlas>.
- [11] "Solución GIS," [Online]. Available: <http://site.geoagro.com/es/producto/soluci%C3%B3n-gis>.

- [12] "Soluciones inteligentes para el manejo de cultivos," [Online]. Available: <https://www.micasense.com/es/altum>.
- [13] M. Gilabert, J. Gonzalez-Piqueras and J. García-Haro, "Acerca de los índices de vegetación," *Revista de teledetección: Revista de la Asociación Española de Teledetección*, no. 8, pp. 1133-0953, 1997.
- [14] "An overview of the available layers and indices in Atlas," 16 Agosto 2019 . [Online]. Available: <https://support.micasense.com/hc/en-us/articles/227837307>.
- [15] E. Taipale, "NDVI vs. NDRE: What's the Difference?," 1 Mayo 2018. [Online]. Available: [https://sentera.com/ndvi-vs-ndre-whats-difference/..](https://sentera.com/ndvi-vs-ndre-whats-difference/)
- [16] "Compare nuestros sensores," [Online]. Available: <https://www.micasense.com/es/comparar-sensores>.
- [17] P. S. Bayu Taruna Widjaja Putra, "Dataset of chlorophyll content estimation of Coffea Canephora using Red and Near-Infrared consumer-grade camera," *Data in Brief*, vol. 21, pp. 736-741, 2018.
- [18] A. Cockburn and L. Williams, "The Costs and Benefits of Pair Programming," in *In eXtreme Programming and Flexible Processes in Software Engineering XP2000*, Addison-Wesley, 2000, pp. 223-247.
- [19] L. Williams, R. R. Kessler, W. Cunningham and R. Jeffries, "Strengthening the case for pair programming," *IEEE Software*, vol. 17, no. 4, pp. 19-25, Agosto 2000.
- [20] "Why does NDVI, NDBI, NDWI Ranges From -1 to 1?," 6 Marzo 2018. [Online]. Available: www.gisresources.com/ndvi-ndbi-ndwi-ranges-1-1/.
- [21] A. López-Pérez, M. Ménez and D. Fernández-Reynoso, "Prioritization of Intervention Areas using a Morphometric Analysis and Vegetation Index," *Water Technology and Sciences*, no. 6, pp. 121-137, 2015.
- [22] C. Laguna, *Correlación y regresión lineal*, Instituto Aragonés de Ciencias de la Salud.
- [23] "Índices de vegetación," AV3 Aerovisual, [Online]. Available: <https://av3aerovisual.com/indices-de-vegetacion/>.

APÉNDICES

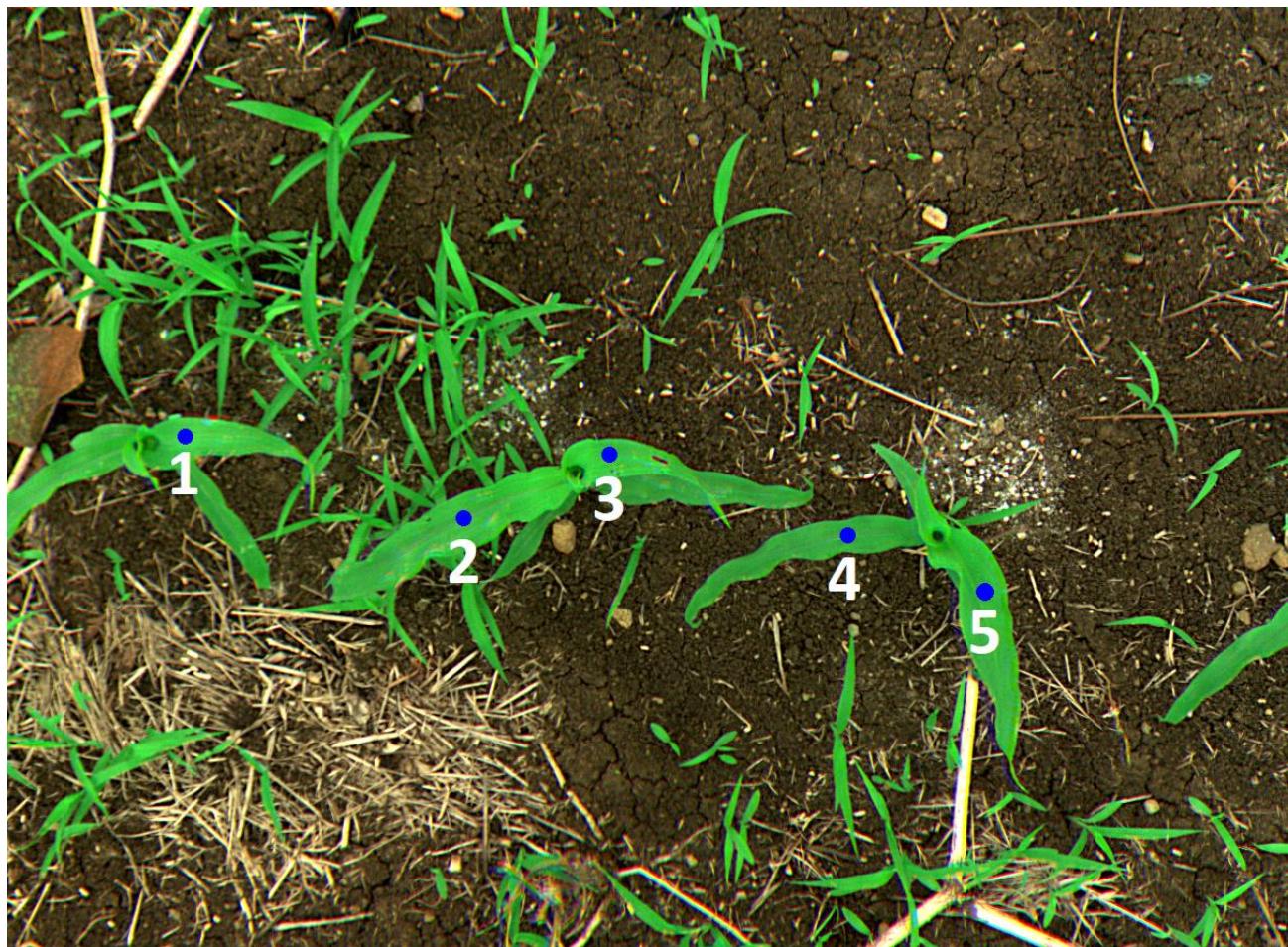
APÉNDICE A

Captura de imágenes de las plantas en el cultivo de maíz de la GEA.

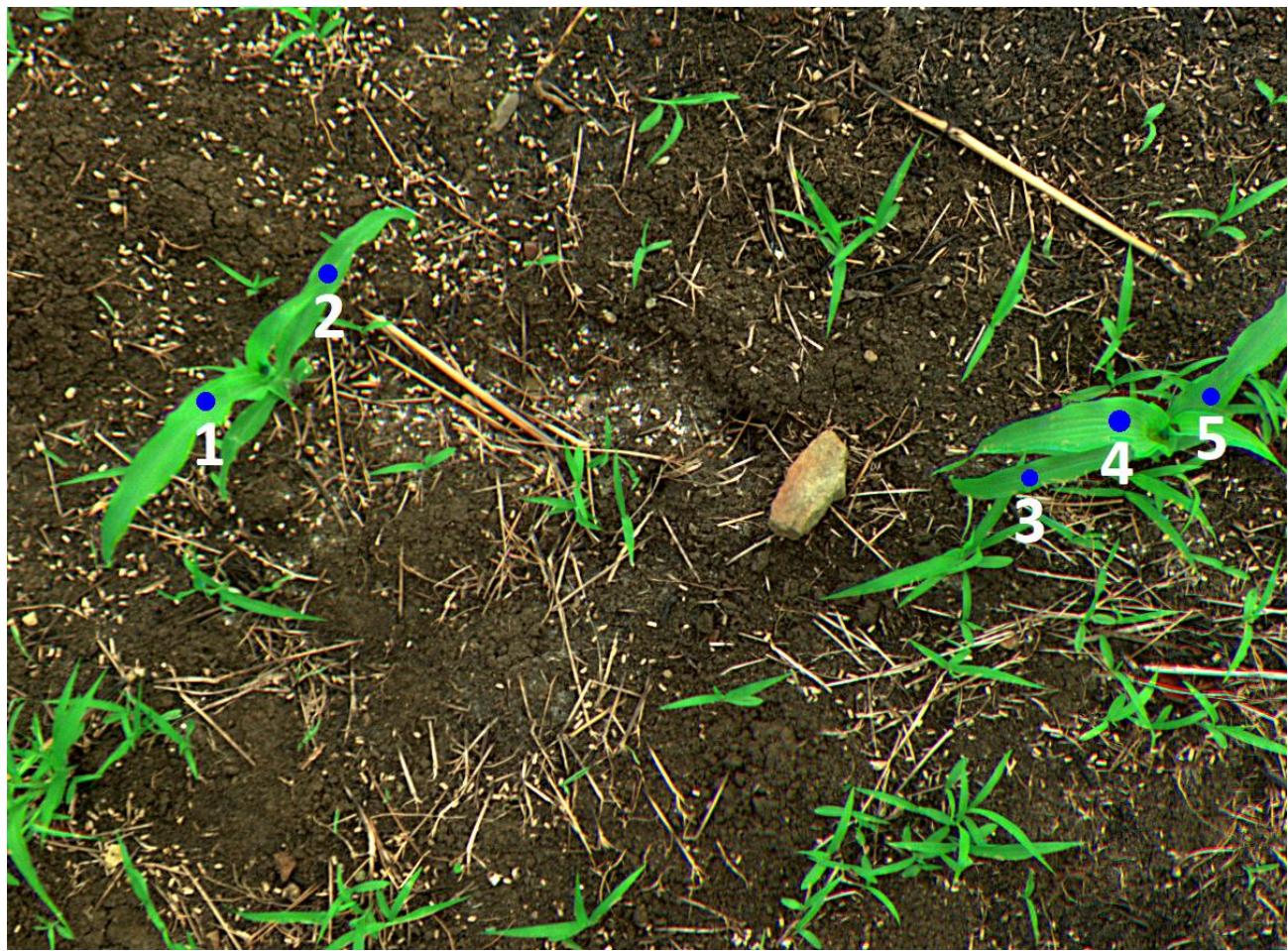


APÉNDICE B

Ejemplos de ejecución del sistema de predicción utilizando la regresión lineal basada en SPAD teórico.



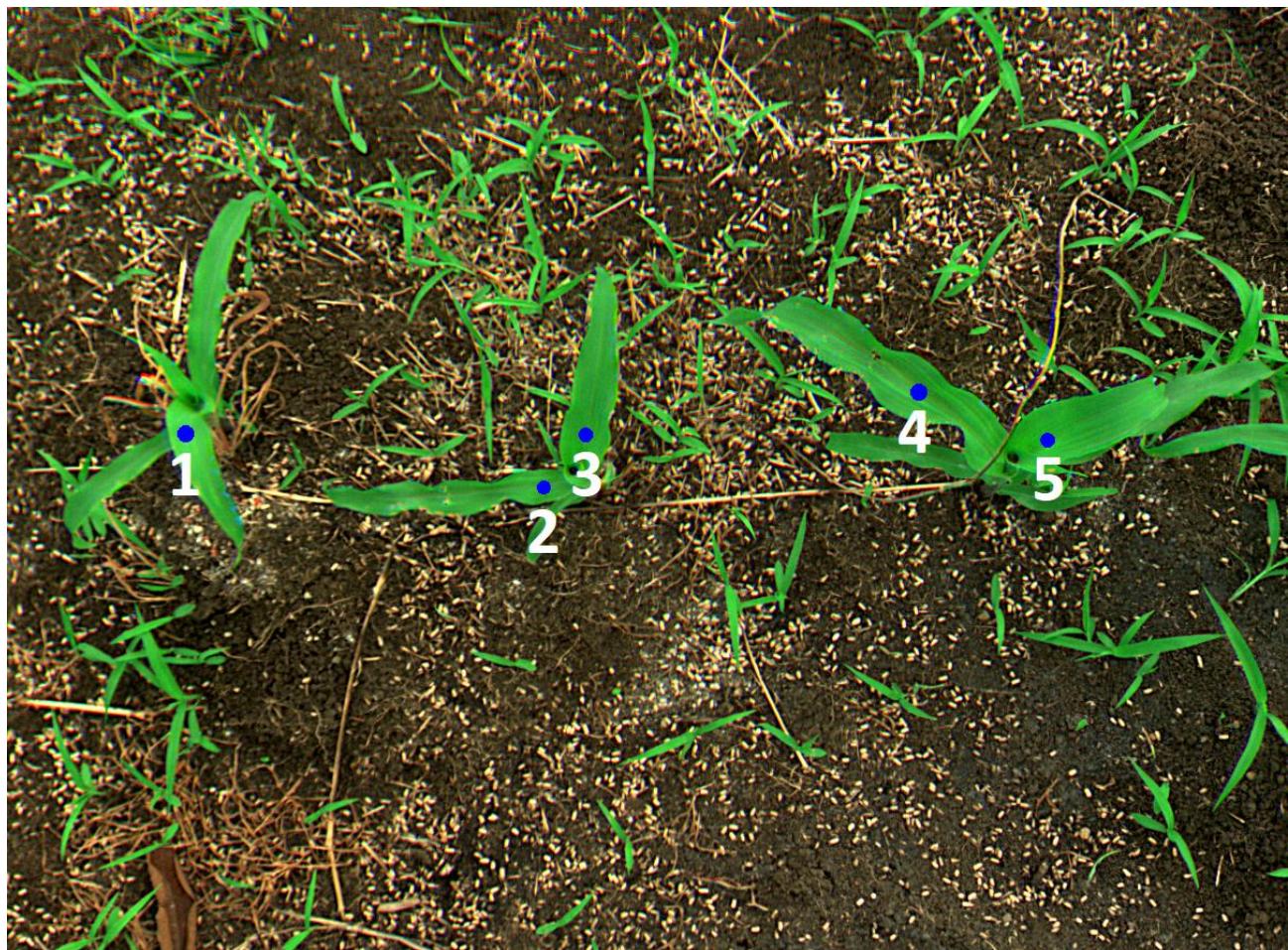
Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8228221	64.02707391
2	0.8230077	64.03584055
3	0.8041814	65.00129674
4	0.8328928	63.14664059
5	0.8388385	64.78355813



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8641816	65.98055602
2	0.84107834	64.88934898
3	0.8328689	64.50160123
4	0.8552278	65.55765446
5	0.8618302	65.86949506



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8492367	65.27468407
2	0.83623016	64.6603607
3	0.8624415	65.8983681
4	0.8947694	67.42527194
5	0.8676937	66.14644084



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8067438	63.26766748
2	0.82863295	64.30153103
3	0.8624415	65.8983681
4	0.88016367	66.73541867
5	0.85438615	65.51790054



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.83749425	64.72006617
2	0.8147233	63.64455406
3	0.8146507	63.6411251
4	0.8282326	64.28262111
5	0.8935775	67.36897853



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8526036	65.43370817
2	0.8365553	64.6757178
3	0.81193596	63.51290248
4	0.8390957	64.79570587
5	0.83868414	64.77626668



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.83786803	64.7377205
2	0.91501576	68.38154484
3	0.84835845	65.23320161
4	0.8667684	66.10273716
5	0.83541715	64.62196091



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.87481534	66.48280776
2	0.8670319	66.11518049
3	0.8684827	66.18370609
4	0.7850401	62.24256212
5	0.8254794	64.15258266

APÉNDICE C

Ejemplos de ejecución del sistema de predicción con imágenes de un cultivo externo en una mayor etapa de crecimiento.



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8690004	66.20815639
2	0.9477249	69.92645492
3	0.8773904	66.60443148
4	0.8328928	64.50273014
5	0.90992844	68.14126184



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.9217159	68.69800524
2	0.94443864	69.77123984
3	0.8994003	67.6439987
4	0.94213253	69.66231845
5	0.823927	64.0792599



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.88661695	67.04021837
2	0.9101077	68.14972725
3	0.72738767	59.51953592
4	0.881065	66.77799063
5	0.8803564	66.74452032



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.9376621	0.6945117312
2	0.9477075	69.92563287
3	0.85342443	65.47247675
4	0.9379553	69.46502125
5	0.9063047	67.9701069



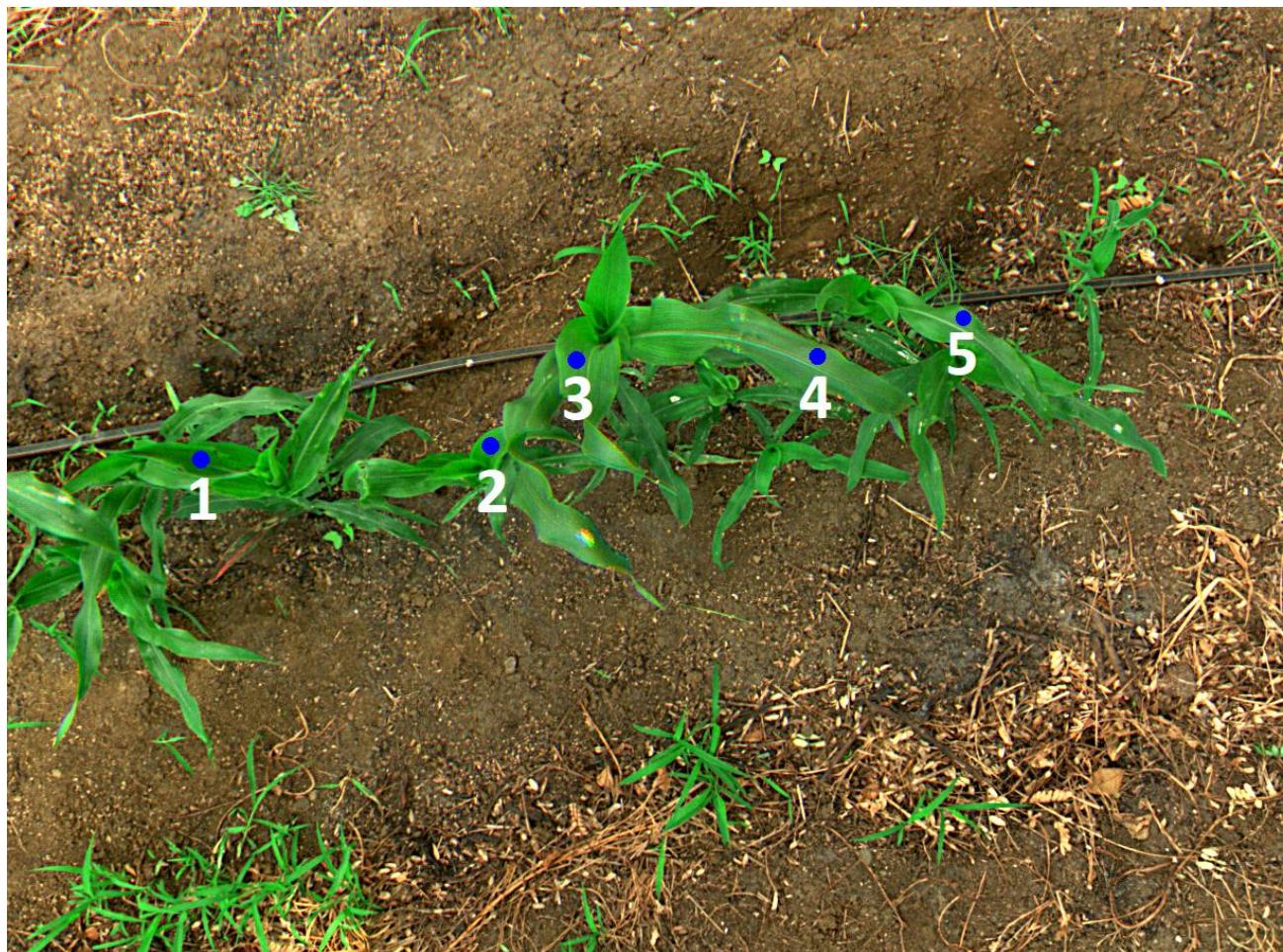
Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.91002536	68.14583941
2	0.9221184	68.71701652
3	0.9128194	68.2778063
4	0.9359129	69.36855446
5	0.8329795	64.50682631



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.8358537	64.64257968
2	0.8781003	66.63796091
3	0.94231814	69.67108509
4	0.83479017	64.59234747
5	0.91511256	68.38611678



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.95733744	70.38047327
2	0.93059105	69.11719352
3	0.92883724	69.03435809
4	0.9454148	69.81734492
5	0.928139	69.00137763



Punto	Valor de NDVI del píxel	Valor de clorofila
1	0.96085215	70.54647914
2	0.9176988	68.50826976
3	0.90445286	67.88264041
4	0.8247689	64.11902507
5	0.8075863	63.30733412

APÉNDICE D

Imágenes NDVI con etiquetado.

