



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

INSTITUTO DE INVESTIGACION EN CIENCIAS BASICAS Y APLICADAS

CENTRO DE INVESTIGACION EN CIENCIAS

Segmentación Automática de Cultivos de Higos a partir de imágenes tomadas por drones

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA:

Lic. Osvaldo Alejandro Saez Lombira

DIRECTOR DE TESIS:

Dr. Jorge Alberto Fuentes Pacheco

SINODALES:

Dr. Jorge Alberto Fuentes Pacheco, Dr. Juan Manuel Rendón Mancha, Dr. Gabriel Corkidi

CUERNAVACA, MORELOS

MARZO, 2018

Agradecimientos

A mi asesor, el Dr. Jorge Alberto Fuentes Pacheco por su apoyo, guía y amistad a lo largo de los dos años que ha tomado la realización de este trabajo de Maestría.

A mi co-asesor, el Dr. Juan Manuel Rendón Mancha, por sus oportunos consejos y comentarios durante cada etapa del desarrollo de esta tesis.

Al Dr. Gabriel Corkidi Blanco quien, como miembro de mi comité tutorial, apoyó con su experiencia y análisis crítico a perfeccionar este trabajo.

A la Lic. Heidy Peidro Guzmán por su amistad y apoyo emocional.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico brindado a lo largo de mis estudios, dándome la oportunidad de realizar este trabajo de la manera más efectiva posible.

A todos, muchísimas gracias.

Tabla de Contenidos

Abstract.....	V
Resumen.....	VI
1. Introducción.....	1
1.1. Descripción del problema.....	2
1.2. Hipótesis y Objetivos.....	3
2. Estado del arte	3
3. Análisis y diseño del algoritmo.....	7
3.1. Definición del SOM.....	7
3.1.1. Entrenamiento del SOM	8
3.1.2. Vegetación y No Vegetación	11
3.2. Segmentación	12
3.2.1. Casos especiales	14
4. Resultados Experimentales.....	16
4.1. Detalles sobre la experimentación.....	17
4.1.1. Especificaciones del algoritmo	17
4.1.2. Especificaciones del dron y el proceso de captura.....	17
4.1.3. Especificaciones de las imágenes.....	18
4.2. Metodología de testeo	18
4.2.1. Curvas de Precisión y Recuerdo.....	20
4.2.2. Validaciones cruzadas.....	23
5. Conclusiones y trabajos futuros.....	35
6. Referencias.....	37

Abstract

Today, humanity has more than seven billion inhabitants, as a result the food needs of our species have increased tremendously compared to other historical periods. That is why today it is imperative to abandon the inefficient practices of the past and make sure that we make more efficient and rational use of our resources if we want to preserve the future of next generations.

Precision agriculture allows us to increase the efficiency of our crops by using advanced techniques of image recognition and artificial intelligence allowing us to determine the exact amount of water and fertilizers that a portion of the crop or plant needs to grow in a healthy way. Saving huge amounts of resources and generating large economic dividends.

At the heart of these techniques is always required a segmentation mechanism that separates the plants of interest in images and/or videos to then apply the techniques of precision.

This work proposes a solution to this problem for complex scenarios of crop figs, where shadows, the method of capture, lighting and objects outside the crop are uncontrolled variables with which the system must deal to perform a successful segmentation.

For this, a non-supervised neural network known as SOM (Self-organizing Maps) was used along with other stochastic techniques to carry out the process of learning and classification, yielding superior results in precision and specificity to those of already established mechanisms such as color indices.

Resumen

En la actualidad la humanidad cuenta con más de siete billones de habitantes, como resultado las necesidades alimenticias de nuestra especie han incrementado tremendamente respecto a otros períodos históricos. Es por ello que, hoy día es imperativo abandonar las prácticas ineficientes del pasado y asegurarnos de realizar un uso más eficiente y racional de los recursos si queremos preservar el futuro de próximas generaciones.

La agricultura de precisión nos permite incrementar la eficiencia de nuestros cultivos mediante el empleo de avanzadas técnicas de reconocimiento de imágenes e inteligencia artificial que nos permiten determinar la cantidad exacta de agua y fertilizantes que una porción de cultivo o planta necesita para crecer de manera sana. Ahorrándose enormes cantidades de recursos y con grandes dividendos económicos.

En el corazón de estas técnicas siempre se requiere de un mecanismo de segmentación que separe las plantas de interés en imágenes y/o videos para luego aplicar las técnicas de precisión.

En este trabajo se propone una solución a dicha problemática, pero para escenarios de cultivos de higos más complejos, donde las sombras, el método de captura, la iluminación y objetos ajenos al cultivo son variables no controladas con las que el sistema debe lidiar para realizar la segmentación.

Para ello se empleó una red neuronal no supervisada conocida como SOM (*Self-Organizing Maps* por sus siglas en inglés) junto a otras técnicas estocásticas y de segmentación simple para realizar el proceso de aprendizaje y clasificación, obteniéndose resultados superiores en precisión y especificidad a los de otros mecanismos ya establecidos como los índices de color.

1. Introducción

El agua dulce constituye uno de los recursos más importantes para el sostenimiento de la vida en la Tierra. En el caso de los humanos no solo la necesitamos para nuestro consumo, sino para el consumo de otras especies que nos sirven de alimento.

Las plantas en especial son una de nuestras fuentes principales de alimentación y también necesitan del agua dulce. De hecho, se estima que casi un 70% (Heraldo, 2015) de ella es empleada solo en cultivos, a pesar de que solo un 2.5% (School, 2016) de la cantidad total disponible es fresca.

Además del agua dulce, existe otro problema asociado con la enorme cantidad de cultivos existentes en la actualidad. Tal problema consiste en la continua desmineralización de los suelos producto de la siembra indiscriminada en ciertas regiones del planeta. Razón por la cual ha sido necesario emplear enormes cantidades de fertilizantes artificiales como nitrógeno, potasio y fosfatos en increíbles cantidades, como son los aproximadamente 208 millones de toneladas producidas el año 2014, según la Organización de las Naciones Unidas para la Alimentación y la Agricultura (Max Roser and Hannah Ritchie, 2017).

Con el objetivo de disminuir el uso desmedido de estos insumos es que nació la agricultura de precisión. En ella, en vez de emplear indiscriminadamente productos agroquímicos o aplicar métodos de riego absolutos, cada planta o zona de cultivo es tratada como una entidad única, con necesidades específicas, permitiendo ahorrar enormes cantidades de agua y fertilizantes en zonas que no la necesitan e incrementando el rendimiento de las cosechas al brindar atención especializada a aquellas que sí.

Para realizar esta actividad es necesario contar con un mecanismo que permita diferenciar de manera eficiente y confiable aquello que corresponde al cultivo de interés de lo que no lo es. En la actualidad, se emplean máquinas para realizar esta labor, por lo que es primordial emplear algoritmos de procesamiento de imágenes

con la finalidad de que interpreten correctamente aquellas capturas que realicen del cultivo.

Es por esta razón que, en la agricultura de precisión, es esencial la capacidad de segmentar el cultivo en entidades individuales, dígase plantas o pequeños conjuntos de ellas. Esta labor se ha realizado a través del desarrollo de métodos computacionales que trabajan en ciertos casos, ya que existe un enorme número de variables que pueden afectar su efectividad. Por ejemplo, elementos como el clima, la iluminación, la forma en que se capturan las imágenes del cultivo, su color, densidad, altura, tamaño y otros, pueden afectar sobremanera la precisión de cualquier algoritmo de segmentación.

En este trabajo se propone un mecanismo de aprendizaje de máquina no supervisado, que emplea redes neuronales auto-organizadas (SOM) para analizar imágenes de cultivos y realizar el proceso de segmentación de plantas de interés en las condiciones naturales del sembrado, independientemente de objetos extraños que puedan encontrarse en la toma y con un alto nivel de precisión. Para realizar las pruebas y el diseño del sistema se han empleado imágenes de cultivos de higos, ya que poseen una gran relevancia económica para el estado de Morelos, siendo el mayor productor de esta planta en México con un 58% del total (Cuautla, 2017).

Es por esta razón que la importancia de esta tecnología no solo es ambiental, sino también económica ya que los agricultores podrían optimizar sus procesos productivos y ahorrar enormes cantidades de capital en fertilizantes y riego mediante el empleo de técnicas de precisión, apoyadas en este algoritmo de segmentación.

1.1. Descripción del problema

El problema tratado en este trabajo es el de segmentar plantas de higo en imágenes aéreas de gran resolución capturadas mediante una cámara RGB alojada en un dron. Estas imágenes poseen variadas condiciones lumínicas y presentan objetos ajenos al cultivo de interés, ya que fueron tomadas en sembrados a campo abierto.

Como ejemplos de estos objetos extraños encontramos: rocas, sombras, hierbas malas y zonas baldías. Esta forma de captura añade un nivel extra de dificultad y realismo a las imágenes analizadas con respecto a las utilizadas en el estado del arte, las cuales son tomadas en invernaderos o condiciones controladas.

Otra característica importante del problema es la naturaleza aérea de las imágenes, lo cual nos permite analizar grandes extensiones de cultivo de manera eficiente, pero también dificulta el proceso de segmentación debido a la pérdida de detalles y al oscurecimiento de las plantas en algunos casos, debido a las sombras generadas por la iluminación natural de la toma.

1.2. Hipótesis y Objetivos

En este trabajo de investigación se plantea la siguiente hipótesis: considerando las condiciones anteriormente descritas de las imágenes se asume que es posible implementar un proceso computacional que, de manera automática, segmente las plantas de higo de manera eficiente y con un alto nivel de correctitud en comparación a varios algoritmos de índice de color.

El objetivo de la tesis consiste en el desarrollo de un algoritmo de aprendizaje de máquina no supervisado basado en características relacionadas con el color, con la finalidad de detectar plantas de higo en imágenes aéreas de un cultivo a campo abierto.

Dicha implementación debe ser no supervisada, debido a la improbabilidad de que en situaciones reales sea posible etiquetar imágenes aéreas de gran resolución con el único propósito de entrenar al algoritmo antes de realizar la segmentación.

2. Estado del arte

El problema de segmentar automáticamente plantas ha sido de gran interés para la comunidad de Visión por Computadora e Inteligencia Artificial durante un largo periodo de tiempo debido a su enorme importancia económica y social. Por esta razón existen una gran cantidad de trabajos e información al respecto, de los cuales,

en este acápite de la tesis, se abordarán aquellos que han resultado de mayor importancia para la solución del problema que se plantea en esta tesis.

Debido a la complejidad del problema, la mayoría de las soluciones propuestas tienden a establecer condiciones muy específicas para el buen funcionamiento del algoritmo como, por ejemplo: el tipo de cultivo, la iluminación, el método de captura, la ausencia de elementos ajenos al cultivo, entre otras.

Guo y Lihong (2016) presentaron un nuevo método para la segmentación de hojas de cultivos de invernaderos con el objetivo de adquirir una estructura tridimensional de la planta y analizar su crecimiento. Este método de segmentación se basa en la técnica de extracción de la nube de puntos que caracteriza a una hoja para luego obtener un modelo tridimensional de la planta. En este trabajo se emplean varios tipos de plantas: *Pachira macrocarpa*, *Scindapsus*, fresas y plantas de tomates, sobre las cuales se aplica el proceso de segmentación obteniendo resultados cualitativamente altos. Sin embargo, solo funciona en condiciones controladas, como un invernadero donde no existen una gran cantidad de elementos ajenos al objeto de interés, además de que no es aplicable para grandes extensiones de cultivos, sino plantas aisladas debido a que la extracción del modelo tridimensional es afectada por la presencia de otras plantas en el mismo espacio, ya que no es capaz de separarlas unas de otras.

Otro mecanismo empleado para lograr la segmentación de plantas es el de los algoritmos genéticos. En 2017, Singh y Misra realizaron una segmentación de hojas de plátano, frijol, limón y rosas empleando algoritmos genéticos (Whitley 1994), con el objetivo de detectar enfermedades sobre ellas. Sin embargo, al igual que en el trabajo anterior las pruebas fueron realizadas sobre imágenes capturadas en condiciones controladas, en especial, estas fueron tomadas enfocando hojas de interés particular, lo cual demostró su efectividad para detectar zonas enfermas, pero no tanto para segmentar la planta en escenarios más complejos.

Considerando aquellos trabajos que se enfocan en imágenes de cultivos en condiciones menos controladas. Haul, *et al.* (2014) emplearon un clasificador de *Random Forest* (Kam Ho 1995) para estimar la certeza en píxeles de que estos

formaran parte del cultivo o la hierba mala, basándose en características extraídas a partir de una vecindad donde ambas se sobrelapan. Para esto emplearon un conjunto de imágenes capturadas en una granja de zanahorias tomadas por un robot terrestre en las condiciones naturales del cultivo. Sus resultados los obtuvieron empleando una validación cruzada en el esquema *leave one out* y la precisión (es la única estadística reportada) del sistema fue superior al 90% en varios casos. Ver la figura 1 para un ejemplo de sus resultados. Observe que el suelo es distintivo y contrastante respecto al cultivo y que las zonas de interés son pequeñas.



Fig. 1. Resultados de Haul, *et al.* (2014): *a*) Imagen inicial y *b*) Imagen procesada.

De igual manera, en el trabajo de Ronghua, *et al.* (2007) se relajan las condiciones relacionadas con la iluminación y se implementa un algoritmo que permite segmentar plantas en condiciones: soleadas, y nubladas con y sin sombra. El mecanismo empleado se basa en convertir la imagen a color a una representación en escala de grises, para luego realizar una segmentación de esta a partir de un umbral calculado mediante un método iterativo. Sin embargo, este mecanismo asume implícitamente que el histograma de la imagen es bimodal, lo cual limita su utilidad a escenarios más complejos donde se pueden encontrar elementos ajenos al cultivo de interés. Además, al igual que en los trabajos anteriores las imágenes comprenden solo una pequeña porción del cultivo, generalmente solo una planta.

En otros trabajos, como el publicado por Hamuda, *et al.* (2017) se emplean aproximaciones más sencillas como considerar un umbral sobre el espacio de color HSV (del inglés: *Hue*, *Saturation* y *Value*) junto a operaciones morfológicas, para determinar la presencia de plantas. El rendimiento del algoritmo fue de una alta sensibilidad y precisión, 98.91% y 99.04% respectivamente. Sin embargo, una rápida implementación de esta idea en nuestro conjunto de imágenes demuestra que la diferencia de altitud, detalle y condiciones lumínicas afectan negativamente los resultados, obteniendo resultados muy inferiores a los mencionados. Esto no es una sorpresa, ya que emplear rangos fijos (de cualquier espacio de color) para imágenes de cultivos tan variables es una aproximación demasiado rígida para resolver el problema planteado en nuestro trabajo. Ver la figura 2 para un ejemplo de los resultados obtenidos por Hamuda, *et al.* (2017)



Fig. 2. Resultados del trabajo de Hamuda, *et al.* (2017): *a)* Imagen inicial y *b)* Imagen procesada.

Existen enfoques más complejos como el análisis y segmentación de texturas empleado en Aksoy, *et al.* (2012), donde a partir de imágenes satelitales de cultivos de orquídeas, y bajo la asunción de que éstos se encuentran sembrados siguiendo ciertos patrones de plantación, crearon un algoritmo no supervisado para la identificación de estos cultivos. Curiosamente, en este trabajo se emplearon imágenes provenientes de Ikonos, QuickBird y Google Earth, por lo que la segmentación individual de cada planta era inviable debido a la altitud y falta de detalle de las imágenes analizadas. Es por ello que su enfoque fue el de clasificar

cultivos enteros, lo cual realizaron exitosamente acorde a los resultados presentados.

Por último, es importante mencionar el gran número de trabajos basados en los diferentes índices de color (*ExG*, *CIVE*, *NDI*, etc.), los cuales constituyen una forma simple y poco costosa de realizar el proceso de segmentación con relativamente buenos resultados, dependiendo de la iluminación y factores externos (Campos, et al. 2016). Por esta razón los índices de color se han convertido en el “Gold Standard” de comparación a la hora de proponer algún algoritmo de segmentación. Como veremos, en la sección de resultados experimentales de este trabajo, también se compara el desempeño de estos índices de color con respecto al sistema que se propone.

3. Análisis y diseño del algoritmo

El algoritmo propuesto emplea un tipo de red neuronal conocida como *Self-Organizing Maps* (SOM) propuestas por el profesor finlandés Teuvo Kohonen (Kohonen 1982) en la década de 1980. Esta red neuronal difiere del resto en que es entrenada empleando aprendizaje no supervisado, lo cual facilita el proceso de entrenamiento del sistema que se propone, así como también hace más accesible su empleo en escenarios donde no se cuenta con las herramientas para entrenar un sistema supervisado.

3.1. Definición del SOM

Un SOM está compuesto de pequeñas entidades conocidas como neuronas. Cada una de ellas posee una ubicación espacial que permite posicionarla con respecto a otras neuronas en una red bidimensional y un vector de pesos que caracteriza el valor que cada una de ellas representa.

Como toda red neuronal, el SOM cuenta con dos modos: entrenamiento y mapeo. Durante el entrenamiento se emplean los vectores de entrada (estos representan la información sobre la que se desea entrenar a la red) para modificar los pesos de las

neuronas. Mientras que durante el mapeo se compara cada uno de los vectores de entrada con los pesos de cada neurona y se decide a través de la utilización de alguna función de distancia cual es la neurona que mejor lo representa.

Además, un SOM posee un conjunto de variables que deben ser establecidas antes de realizar cualquiera de estas acciones:

- **Número de neuronas:** Es el número de neuronas que componen la red. En general se define estableciendo el número de ellas horizontal y verticalmente, ya que como se dijo antes éstas poseen una posición en un espacio bidimensional. Resultando el número total de neuronas en el producto de ambos. Por ejemplo, en nuestro problema se definió una red de 5x5, por lo que el sistema posee 25 neuronas.
- **Razón de aprendizaje (A_R):** Determina cuánto aprende cada neurona a partir de un vector de entrada. Este valor decrece a medida que más valores son presentados a la red. En nuestro sistema el valor inicial es de 0.1.
- **Radio de vecindad (V_R):** Determina qué neuronas serán modificadas durante el proceso de entrenamiento. Este valor decrece a medida que más valores son presentados a la red. El valor inicial es 2.5 que es el radio de la circunferencia ubicada en el centro de la red formada por las 25 neuronas citadas antes.

3.1.1. Entrenamiento del SOM

Luego de determinar las imágenes que serán empleadas para el entrenamiento, se crea un SOM para cada una de ellas. El primer paso es convertir el espacio de color de la imagen de *RGB* a *CIE L * a * b*, el razonamiento detrás de esta conversión se basa en el significado de las componentes en cada espacio de color. En el caso de *RGB* no existe componente para la luminosidad, esto quiere decir que esta se encuentra codificada en la combinación de las tres componentes de color: rojo (*R*), verde (*G*) y azul (*B*), o sea que para imágenes con iluminación no controlada como las que se utilizan en este trabajo un mismo color podría estar representado por diferentes valores debido a que su luminosidad cambia de pixel a pixel dificultando

su identificación. En el caso de $CIE L * a * b$ la luminosidad de cada pixel se codifica en la componente L , por lo tanto, los colores con tonos parecidos presentan similitud en sus componentes a y b , inclusive si la luminosidad cambia. Además, en este espacio de color los colores se encuentran uniformemente distribuidos y muy cercanos a la percepción humana (León, 2006).

A continuación, por cada pixel se genera un vector con tres componentes $p = (p_1, p_2, p_3)$ que corresponden a su representación en $CIE L * a * b$, normalizadas en el rango de $[0,1]$. Cada vector se compara con los pesos de cada neurona del SOM para determinar aquella neurona que produzca la distancia más cercana a éste a través del cálculo de la distancia euclidiana. Esta neurona es definida como $n \rightarrow (l, w)$, donde $l = (x, y)$ es un vector con las coordenadas de posición de la neurona y $w = (w_1, w_2, w_3)$ es un vector que contiene los pesos. Estos pesos se inicializan aleatoriamente en el rango de $[0,1]$. A esta neurona se le conoce como BMU (*Best Matching Neuron*).

Después se ubican aquellas neuronas cuya distancia a la BMU es menor o igual al radio de vecindad (V_R). Esto se realiza calculando la distancia euclidiana entre los vectores de cada neurona que representan su localización en la red (no sus pesos, recuérdese que cada neurona posee ambos datos), o sea $l \rightarrow (x, y)$. Sea entonces, $N = \{ u \mid D(u, n) \leq V_R \}$ el conjunto de neuronas en la vecindad del BMU (referido como n en adelante). Luego, los pesos de cada neurona en N van a cambiar acorde a la siguiente fórmula:

$$w_i = w_i + I * A_R * (p_i - w_i), \quad i = \{1, 2, 3\} \quad (1)$$

Donde A_R es la razón de aprendizaje actual e I es la influencia que el vector de entrada (el pixel p) ejerce sobre cada neurona en particular. La idea detrás de I , es que mientras más cerca se encuentre una neurona u de n , entonces sus pesos serán más parecidos. El valor de I se encuentra dado por:

$$I = e^{-\frac{D(u, n)}{2 * V_R}} \quad (2)$$

Observe que a medida que la distancia entre u y n incrementa, mas pequeño se vuelve I , mientras que si u es igual a n , entonces $D(u, n) = D(n, n) = 0$ y por tanto $I = 1$ el mayor valor posible, o sea la mayor influencia.

Una vez realizado este proceso para todo N , los pesos de n y toda su vecindad de neuronas serán un poco más parecidos a p . Intuitivamente es mediante este proceso que las neuronas “aprenden” ya que si aparece otro pixel similar a p entonces una neurona de esta vecindad será elegida como el BMU realizando el proceso una vez más sobre este grupo de neuronas y haciéndolas “aprender” aún más de pixeles como p .

Sin embargo, como mencionamos antes, los valores de A_R y V_R disminuyen con el tiempo. Esto permite que la red alcance un estado de estabilidad luego de un gran número de iteraciones, ya que así (1) y (2) también disminuirían y se aprendería cada vez menos de cada vector de entrada, hasta que prácticamente no ocurren cambios significativos en el SOM y se puede detener el aprendizaje (ver Kohonen, 1982). Los cambios que experimentan A_R y V_R , luego de calcular los nuevos pesos de cada neurona en N pueden quedar representados de la siguiente manera:

$$A_R = 0.1 * e^{-\frac{iActual}{iTotal - A_R}} \quad (3)$$

donde $iActual$ es el número de pixeles procesados hasta un determinado momento e $iTotal$ es el número de pixeles que posee toda la imagen. El valor 0.1 se debe al valor con el que fue inicializado A_R .

$$V_R = 2.5 * e^{-\frac{iActual}{iTotal / \log 2.5}} \quad (4)$$

Donde el número 2.5 es el valor inicial de V_R . Con ambas fórmulas se observa que a medida que $iActual$ incrementa, los valores de A_R y V_R decrecen proporcionalmente.

Estos cálculos se realizan para cada pixel de la imagen elegido aleatoriamente, hasta que todos los pixeles contenidos en la imagen se hayan utilizado para entrenar al SOM de esta imagen específica.

3.1.2. Vegetación y No Vegetación

Una vez concluido el proceso de entrenamiento del SOM, para cada imagen se realiza un agrupamiento basado en el algoritmo *K-Means* sobre las neuronas, empleando sus pesos como elementos de comparación. Como resultado de esto se clasificará a cada neurona en “vegetal” o “no vegetal” con base en el clúster al que sean asignadas.

Sin embargo, hay un par de interrogantes sobre este proceso que deben ser respondidas antes de poder realizar dicha clasificación. Primero, ¿Cuál es el número de grupos K para realizar el *K-Means*? y segundo, ¿Cómo determinar los clústeres de neuronas que caracterizan a los pixeles como vegetación?

En este trabajo se empleó un $k = 2$ para denotar los dos grandes grupos que aparecen en las imágenes empleadas: vegetación y no vegetación (dígase, suelo, sombras, rocas y otros objetos ajenos al cultivo). La elección del clúster que contiene las neuronas pertenecientes a la vegetación se determina convirtiendo los centroides de cada clúster al espacio de color *HSV* eligiendo aquel con el *Hue* (H), más cercano al intervalo $[50, 70]$ que corresponde al intervalo para el color verde en OpenCV¹, ver (Changing Colorspaces 2018). Luego, todas las neuronas pertenecientes al clúster de este centroide son clasificadas como “vegetal”, y el resto como “no vegetal”.

A continuación, cada pixel de la imagen de entrenamiento empleada para obtener el SOM es clasificado como “vegetal” o “no vegetal”. Esto se realiza determinando para cada uno la neurona más cercana a ellos (o sea su BMU, pero ahora sobre el

¹ Librería para Visión por Computadora escrita en C++, empleada extensamente en el proyecto en cuestión.

SOM ya entrenado). Entonces, el pixel es clasificado con la misma etiqueta de la neurona seleccionada como BMU.

Con los pixeles de cada imagen RGB utilizada en el entrenamiento se crean dos histogramas, uno “vegetal” y otro “no vegetal” que contendrán respectivamente, la cantidad de veces que un pixel específico fue asociado con una neurona clasificada como de una u otra clase. Por ejemplo, del histograma vegetal podremos obtener la cantidad de veces que un pixel dado fue valorado como “vegetal” y, como veremos más adelante, esta información junto a la del histograma “no vegetal” será empleada para clasificar imágenes que no son consideradas durante el entrenamiento.

3.2. Segmentación

El proceso de segmentación para imágenes de prueba se realiza clasificando cada uno de sus pixeles como “vegetal” o “no vegetal”, comparando $P(vegetal | p_0)$ contra $P(noVegetal | p_0)$, donde p_0 es el pixel de interés (en el espacio de color *RGB*). Este par de probabilidades se determinan usando la definición de probabilidad condicional:

$$P(vegetal | p_0) = \frac{P(p_0 | vegetal) * P(vegetal)}{P(p_0)} \quad (5)$$

$$P(noVegetal | p_0) = \frac{P(p_0 | noVegetal) * P(noVegetal)}{P(p_0)} \quad (6)$$

Para resolver estas ecuaciones se necesitan obtener los valores de: $P(p_0 | vegetal)$, $P(p_0 | noVegetal)$, $P(vegetal)$, $P(noVegetal)$ y $P(p_0)$.

En la sección anterior se obtuvieron dos histogramas al final del entrenamiento, $H_{vegetal}$ y $H_{noVegetal}$, los cuales contienen para cada posible pixel p la cantidad de veces que este fue clasificado como “vegetal” y “no vegetal” (esto es $H_{vegetal}[p]$ y $H_{noVegetal}[p]$ respectivamente). Luego, si sumamos estas cantidades para todos los pixeles de las imágenes consideradas para el entrenamiento se obtiene que:

(8)

$$t_{vegetal} = \sum_{r=0}^{255} \sum_{g=0}^{255} \sum_{b=0}^{255} H_{vegetal}[(r, g, b)]$$

$$t_{noVegetal} = \sum_{r=0}^{255} \sum_{g=0}^{255} \sum_{b=0}^{255} H_{noVegetal}[(r, g, b)]$$

O sea $t_{vegetal}$ y $t_{noVegetal}$ son el total de veces que los pixeles fueron clasificados como “vegetal” y “no vegetal” respectivamente. Mientras que $t = t_{vegetal} + t_{noVegetal}$ es el número total de pixeles analizados.

Estas cantidades son importantes, ya que si un pixel p_0 fue clasificado como “vegetal” durante el entrenamiento una cantidad $H_{vegetal}[p_0]$ de veces, de un total de pixeles también clasificados de la misma forma ($t_{vegetal}$), entonces la probabilidad de que p_0 ocurra como pixel “vegetal” es

$$P(p_0 | vegetal) = \frac{H_{vegetal}[p_0]}{t_{vegetal}} \quad (9)$$

Mientras que para el caso de “no vegetal” sería

$$P(p_0 | noVegetal) = \frac{H_{noVegetal}[p_0]}{t_{noVegetal}} \quad (10)$$

Siguiendo una lógica similar. Si consideramos la cantidad total de veces que el pixel p_0 fue clasificado ($H_{vegetal}[p_0] + H_{noVegetal}[p_0]$), como “vegetal” o “no vegetal”, entre la cantidad total (t) de pixeles analizados, entonces la probabilidad de que este pixel aparezca en alguna de nuestras imágenes de entrenamiento es

$$P(p_0) = \frac{H_{vegetal}[p_0] + H_{noVegetal}[p_0]}{t} \quad (11)$$

En este punto observemos que para resolver (5) y (6), solo restan $P(vegetal)$ y $P(noVegetal)$, las cuales denotan la probabilidad de que un pixel cualquiera sea “vegetal” o “no vegetal” respectivamente. Sabiendo que $t_{vegetal}$ es el (13)

número de píxeles clasificados como “vegetal”, $t_{noVegetal}$ el número de aquellos clasificados como “no vegetal” y t el total analizado, entonces

$$P(vegetal) = \frac{t_{vegetal}}{t}$$

$$P(noVegetal) = \frac{t_{noVegetal}}{t}$$

Finalmente, si

$$P(vegetal | p_0) > P(noVegetal | p_0) \tag{14}$$

El píxel p_0 sería clasificado como “vegetal” y como “no vegetal” en el caso contrario.

3.2.1. Casos especiales

Aunque la fórmula denotada en (14) es, en esencia, el mecanismo mediante el cual se clasifican los píxeles para imágenes desconocidas. El sistema propuesto no la emplea explícitamente de esta manera, ya que varias consideraciones deben ser hechas con relación a p_0 .

Primero, cuando $P(vegetal | p_0)$ y $P(noVegetal | p_0)$ poseen valores similares (no necesariamente iguales) esto significa que p_0 es un píxel valorado como “vegetal” y “no vegetal” durante el entrenamiento un número similar de veces. Por lo que se podría decir que no existe un consenso abrumador sobre qué tipo de píxel es y podría resultar en un falso positivo o un falso negativo si es clasificado erróneamente.

En tal caso, para un píxel en esta situación se asume que es “no vegetal”, a menos que exista una diferencia mínima de 0.1 entre $P(vegetal | p_0)$ y $P(noVegetal | p_0)$. De donde, la condición especificada durante la sección anterior cambiaría a

$$P(vegetal | p_0) - P(noVegetal | p_0) > 0.1 \tag{15}$$

Obsérvese que esto aún implica que (14) es cierto, solo que ahora debe haber una mínima diferencia entre ellos.

El empleo de (15) como función comparativa en lugar de (14) implica que pueden presentarse situaciones donde $P(\text{vegetal} | p_0) > P(\text{noVegetal} | p_0)$, pero como $P(\text{vegetal} | p_0) - P(\text{noVegetal} | p_0) \leq 0.1$ entonces, p_0 es clasificado como “no vegetal”.

La elección del umbral (0.1) es experimental y la razón por la que se exige más de $P(\text{vegetal} | p_0)$ que de $P(\text{noVegetal} | p_0)$ al realizar la clasificación, es porque se considera que los falsos positivos resultaban más disruptivos al uso del sistema que un falso negativo, luego esta medida apunta a minimizar su ocurrencia. Sin embargo, es importante mencionar que la ocurrencia de pixeles en esta situación es muy baja, en especial si se emplean imágenes de entrenamiento tomadas en condiciones y horarios similares.

Un segundo caso a considerar es cuando un pixel p_0 , que se está intentando clasificar en una imagen desconocida nunca apareció en ninguna de las imágenes de entrenamiento. Luego, tanto para el histograma “vegetal” como el “no vegetal” se cumpliría que

$$H_{\text{vegetal}}[p_0] = H_{\text{noVegetal}}[p_0] = 0 \quad (16)$$

En tal caso la función (11) sería igual a cero, por lo que (5) y (6) no estarían bien definidas debido a la división por cero.

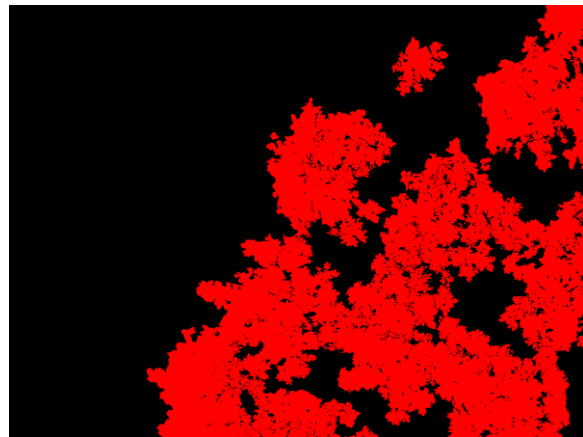
Para solucionar este problema el sistema propuesto asume que dichos pixeles son “no vegetales”, ya que de ser “vegetales” habrían aparecido al menos una vez en alguna imagen de entrenamiento y habrían sido clasificados como tal lo que se traduciría en que $H_{\text{vegetal}}[p_0] > 0$.

4. Resultados Experimentales

Con el objetivo de testear el funcionamiento del sistema se emplea un grupo de 10 imágenes aéreas de una resolución de 2000×1500 píxeles, de un sembradío de higo ubicado en el estado de Morelos de la República Mexicana. Cada una de ellas fueron tomadas empleando un dron en condiciones ambientales y lumínicas completamente naturales.

Entre las imágenes empleadas se observan un gran abanico de condiciones, por ejemplo, aquellas con una presencia predominante de hierbas malas o suelos no sembrados, otras con grandes extensiones de suelo, sombras y rocas, mientras que otras presentan evidencia de falta de agua en el cultivo.

De igual manera, para cada imagen del conjunto se realizó una segmentación manual consideradas como *Ground Truth* (GT) y que son empleadas durante el proceso de testeo para determinar la correctitud del sistema. Ver la figura 3.



a)

b)

c)

d)

Fig.3. Muestra de la imagen del conjunto: a) Imagen aérea original, b) GT correspondiente, c) Acercamiento de una región de la imagen original y d) GT correspondiente al acercamiento de la imagen original.

4.1. Detalles sobre la experimentación

Antes de continuar con el testeo del sistema es importante esclarecer los detalles relacionados con la obtención del conjunto de imágenes anterior, así como las características técnicas que poseen cada una de ellas. Estas especificaciones nos dan una visión clara sobre las condiciones en que se probará el sistema y, por tanto, de la relevancia de los resultados aquí presentados.

4.1.1. Especificaciones del algoritmo

- **Número de neuronas:** 25 (5x5)
- **Razón de aprendizaje:** 0.1
- **Razón de vecindad:** 2.5
- **Número de pesos en las neuronas:** 3
- **Interpretación del vector de entrenamiento:** *Componentes RGB del pixel*
- **Número de clústeres de neuronas:** 2
- **Mínimo umbral de probabilidad vegetal:** 0.1
- **Tiempo de entrenamiento típico:** ~127 minutos
- **Tiempo de segmentación típico:** ~2 segundos

4.1.2. Especificaciones del dron y el proceso de captura

- **Modelo:** *Phantom 4 de DJI*
- **Software:** *Drone Deploy*
- **Espacio cubierto:** *1 hectárea*
- **Altitud:** *20 metros*

- **Relación entre imagen y terreno:** 0.5 cm/px
- **Localización:** *Xalostoc, Morelos.*
- **Coordenadas GPS:** 18.72.1700, -98.907600
- **Resolución de la cámara:** *4K*

4.1.3. Especificaciones de las imágenes

Estas características se aplican tanto a las imágenes originales como al GT, a menos que se especifique lo contrario.

- **Resolución:** 2000×1500 *pixeles*
- **Espacio de color original:** *RGB*
- **Profundidad de color:** *8 bits*
- **Formato imágenes originales:** *JPEG*
- **Formato de imágenes GT:** *PNG*

Estas imágenes fueron tomadas a partir de cultivos de higos sin realizar ninguna modificación a su estado original por lo que las condiciones lumínicas para cada una de las imágenes pueden cambiar debido a la posición de la imagen cuando fueron tomadas.

4.2. Metodología de testeo

Para realizar la comprobación del sistema se ha empleado una validación cruzada exhaustiva sobre el conjunto de imágenes anteriormente mencionado. En esta validación cruzada se deja una muestra fuera para el testeo o LOOVC, por sus siglas en inglés *Leave-one-out cross-validation*.

Esto quiere decir que para un conjunto de n imágenes se emplean $n - 1$ para el entrenamiento y se prueba con la omitida. Este proceso se repite n veces para que cada imagen tome el lugar de aquella que es dejada fuera.

Una vez que se obtiene la segmentación del sistema para cada imagen del conjunto, entonces se clasifica cada pixel empleando el GT correspondiente como:

- **Verdadero positivo (VP):** Es aquel pixel clasificado como “vegetal” por el sistema e identificado como “vegetal” en el GT.
- **Verdadero negativo (VN):** Es aquel pixel clasificado como “no vegetal” por el sistema e identificado como “no vegetal” en el GT.
- **Falso positivo (FP):** Es aquel pixel clasificado como “vegetal” por el sistema e identificado como “no vegetal” por el GT.
- **Falso negativo (FN):** Es aquel pixel clasificado como “no vegetal” por el sistema e identificado como “vegetal” por el GT.

La siguiente matriz de confusión (Fawcett, 2005) muestra la relación entre los diferentes resultados presentados.

		<i>Condición Verdadera</i>	
		Vegetal	No Vegetal
<i>Predicción</i>	Vegetal	Verdadero Positivo (VP)	Falso Positivo (FP)
	No Vegetal	Verdadero Negativo (VN)	Falso Negativo (FN)

Tabla. 1. Matriz de confusión que muestra la relación entre VP, VN, FP y FN.

Como se puede observar aquellos valores que se intenta maximizar son los *verdadero positivos* y *falsos negativos*, mientras que se busca minimizar los *falsos positivos* y *verdaderos negativos*. Una vez se cuente con esta información, entonces se calculan varias medidas estadísticas con el objetivo de medir la correctitud del sistema (Alpaydin 2010):

- **OSR** (*Tasa de Éxito Global o Exactitud*)

$$OSR = \frac{VP + VN}{VP + VN + FP + FN}$$

- **TPR** (*Tasa de Verdaderos Positivos, Sensibilidad o Recuerdo*)

$$TPR = \frac{VP}{VP + FN}$$

- **TNR** (*Tasa de Verdaderos Negativos o Especificidad*)

$$TNR = \frac{VN}{VN + FP}$$

- **PPV** (*Valor Predictivo Positivo* o *Precisión*)

$$PPV = \frac{VP}{VP + FP}$$

- **NPV** (*Valor Predictivo Negativo*)

$$NPV = \frac{VN}{VN + FN}$$

- **F** (*Medida-F*)

$$F = \frac{2 * VP}{2 * VP + FN + FP}$$

De todas estas medidas **OSR** o Tasa de Éxito Global es la más importante ya que contabiliza el número de verdadero positivos y negativos contra el total de pixeles clasificados, dando una idea más completa de cómo lo realizó el algoritmo en vez de solo enfocarse en los verdaderos positivos o negativos de manera exclusiva.

Cada imagen del conjunto es segmentada empleando el mecanismo explicado en este trabajo, así como otros algoritmos de índice de color. Para cada segmentación se obtienen las medidas estadísticas anteriores, y luego se compara la efectividad del sistema propuesto contra estos algoritmos ya probados por la comunidad.

4.2.1. Curvas de Precisión y Recuerdo

Los algoritmos de índice de color considerados son: MExG, CIVE, ExG, COM2, ExR, ExGR, NDI, GB, VEG y COM1.

Las funciones que se emplean para obtenerlos son las siguientes:

- **MExG** (*Modified Excess Green Index*)

$$1.262 * G - 0.884 * R - 0.311 * B$$

- **CIVE** (*Colour Index of Vegetation Extraction*)

$$0.441 * R - 0.811 * G + 0.385 * B + 18.78745$$

- **ExG** (*Excess Green Index*)

$$2 * G - R - B$$

- **COM2** (*Combined Indices 2*)

$$0.36 * ExG + 0.47 * CIVE + 0.17 * VEG$$

- **ExR** (*Excess Red Index*)

$$1.3 * R - G$$

- **ExGR** (*Excess Green minus Excess Red Index*)

$$ExG - ExR$$

- **NDI** (*Normalised Difference Index*)

$$128 * (((G - R)/(G + R)) + 1)$$

- **GB** (*Green minus Blue*)

$$G - B$$

- **VEG** (*Vegetative Index*)

$$G/(R^a * B^{(1-a)}), a = 0.667$$

- **COM1** (*Combined Indices 1*)

$$ExG + CIVE + ExGR + VEG$$

Debido al alto número de ellos disponibles y, con el objetivo de simplificar las comparaciones con el sistema propuesto en este trabajo, vamos a elegir aquellos que mejor resultados ofrecen para el conjunto de imágenes analizadas.

Afortunadamente, todos estos algoritmos mencionados siguen un patrón de segmentación similar que consiste en convertir la imagen analizada a escala de

grises y luego encontrar un punto en dicha escala que sirva como umbral para separar pixeles “vegetales” de pixeles “no vegetales”.

Sin embargo, donde todos difieren, es en la fórmula empleada para convertir la imagen original a otra en escala de grises. Razón por la cual cada uno de ellos obtiene resultados muy diferentes en distintas condiciones lumínicas del cultivo.

Con el objetivo de contabilizar estas diferencias se empleó un gráfico de Precisión contra Recuerdo, donde para cada índice de color una curva fue representada. Dicha curva se obtiene a partir de todos los umbrales posibles en la escala de grises. O sea, para cada índice de color se grafican tantos puntos como umbrales, donde cada punto corresponde a la Precisión y Recuerdo obtenidas por el índice de color empleando ese umbral para la imagen analizada. Nótese que mientras más cerca de 1 se encuentren la Precisión y el Recuerdo, mejor es el umbral. Idealmente, buscamos un umbral, para cualquiera de los índices de color, cuyo punto en el gráfico corresponda a (1, 1). O sea, una Precisión y Recuerdo perfectos.

Las figuras 4, 5 y 6 corresponde a dichos gráficos para tres imágenes distintas de nuestro conjunto de clasificación.

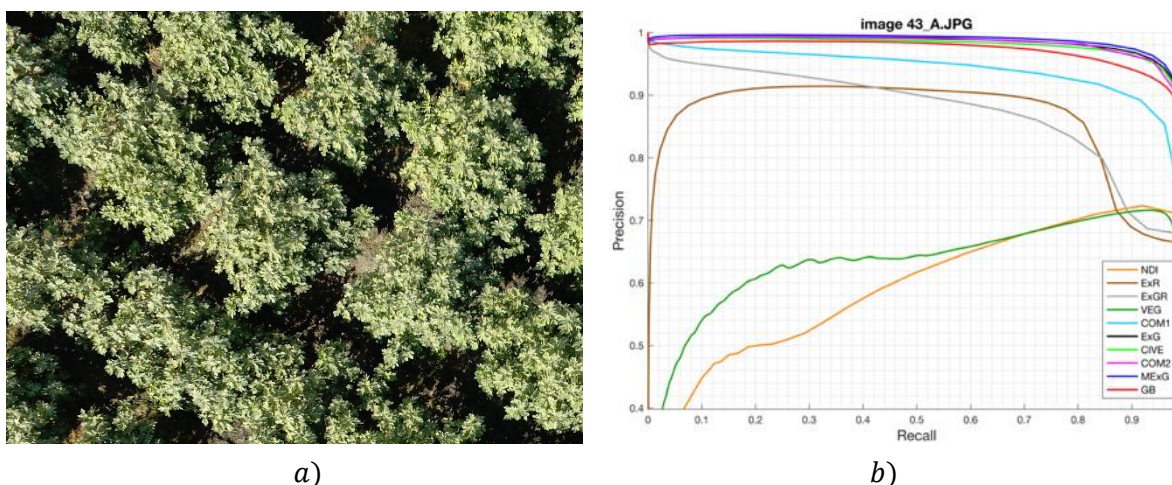
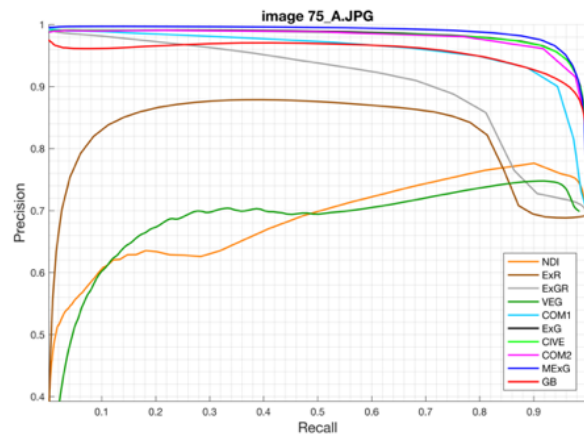


Fig.4. Imagen DJI_0043_A: a) Imagen original a segmentar, b) Gráfica de Precisión contra Recuerdo para cada uno de los índices de color.



a)

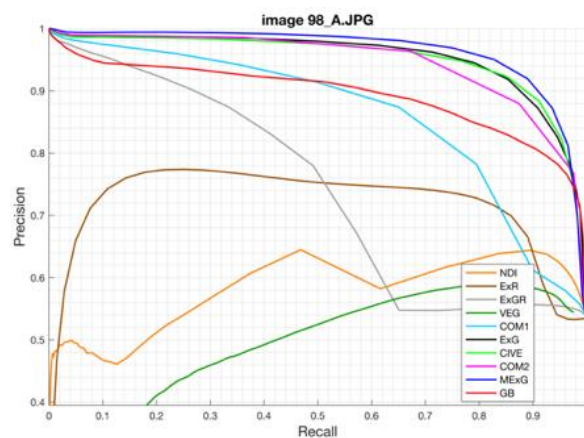


b)

Fig. 5. Imagen DJI_0075_A: a) Imagen original a segmentar, b) Gráfica de Precisión contra Recuerdo para cada uno de los índices de color.



a)



b)

Fig. 6. Imagen DJI_0098_A: a) Imagen original a segmentar, b) Gráfica de Precisión contra Recuerdo para cada uno de los índices de color.

Como se puede observar para los índices de color: MExG, CIVE, ExG, y COM2 presentan los resultados más precisos de todo el conjunto. Por lo que la comparación del sistema aquí propuesto se realizará contra estos índices de color, debido a que son los más efectivos.

4.2.2. Validaciones cruzadas

Para las imágenes del conjunto se emplearon aquellas que representaran la mayor variedad de condiciones contempladas durante el diseño del sistema. Por ejemplo,

las imágenes DJI_0010_A y DJI_0101_A se caracterizan por amplias extensiones de hierbas malas y suelos no cultivados. Los resultados obtenidos (ver tablas 1 y 10) sugieren que el algoritmo propuesto es notablemente más exacto que los índices de color. Esto se debe principalmente al menor número de falsos positivos generados a partir de las hierbas malas, ya que durante el aprendizaje el sistema no “aprende” que estos colores deberían ser clasificados como “vegetal”, a diferencia de la aproximación usada por los índices de color basada en rangos de color. Ver las figuras 7 y 17 para los resultados del algoritmo propuesto en las imágenes mencionadas.

Otras imágenes como la DJI_0043_A y la DJI_0075_A se caracterizan por presentar sombras muy pronunciadas, debido a la iluminación solar. En estos casos la sombra proyectada por las plantas de higo oculta porciones de plantas adyacentes. Como se observa en las tablas 5 y 7 el sistema es nuevamente más preciso que los índices de color. Lo cual se debe a la presencia de un menor número de falsos negativos provenientes de los píxeles de plantas de higo oscurecidas por las sombras. Ver las figuras 11 y 13.

En la imagen DJI_0098_A se observa una caída en la precisión del algoritmo en comparación a los índices de color (ver tabla 9 y figura 12). La razón para este decrecimiento se puede observar especialmente en la esquina inferior izquierda de la imagen donde varios píxeles de plantas de higo son clasificados erróneamente como “no vegetal”. Un análisis más detallado de esta imagen refleja que esta zona del cultivo se caracteriza por plantas secas, lo cual se refleja en el color marrón de las hojas, dándole a estos píxeles una baja probabilidad de ser clasificados como “vegetal” por el sistema, ya que en las imágenes de entrenamiento este color aparece pocas veces como parte de una planta.

El resto de las imágenes presentan una equilibrada proporción de cultivo, suelo y sombras. En cada una de ellas el sistema mantiene una precisión superior a los índices de color (ver tablas 2-4, 6 y 8). Ver las figuras 8-10, 12 y 14 para ver los resultados del algoritmo propuesto.

Imagen DJI_0010_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.810804	0.769541	0.734041	0.717012	0.695298
Recuerdo (TPR)	0.936377	0.93316	0.89384	0.84356	0.788428
Especificidad (TNR)	0.740639	0.678118	0.644751	0.646302	0.643261
Precisión (PPV)	0.668578	0.618304	0.584354	0.571299	0.552555
Predictivo Negativo (NPV)	0.954199	0.9478	0.91575	0.880864	0.844752
Medida-F (F)	0.780135	0.743783	0.706699	0.681234	0.649747

Tabla 1. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0010_A.

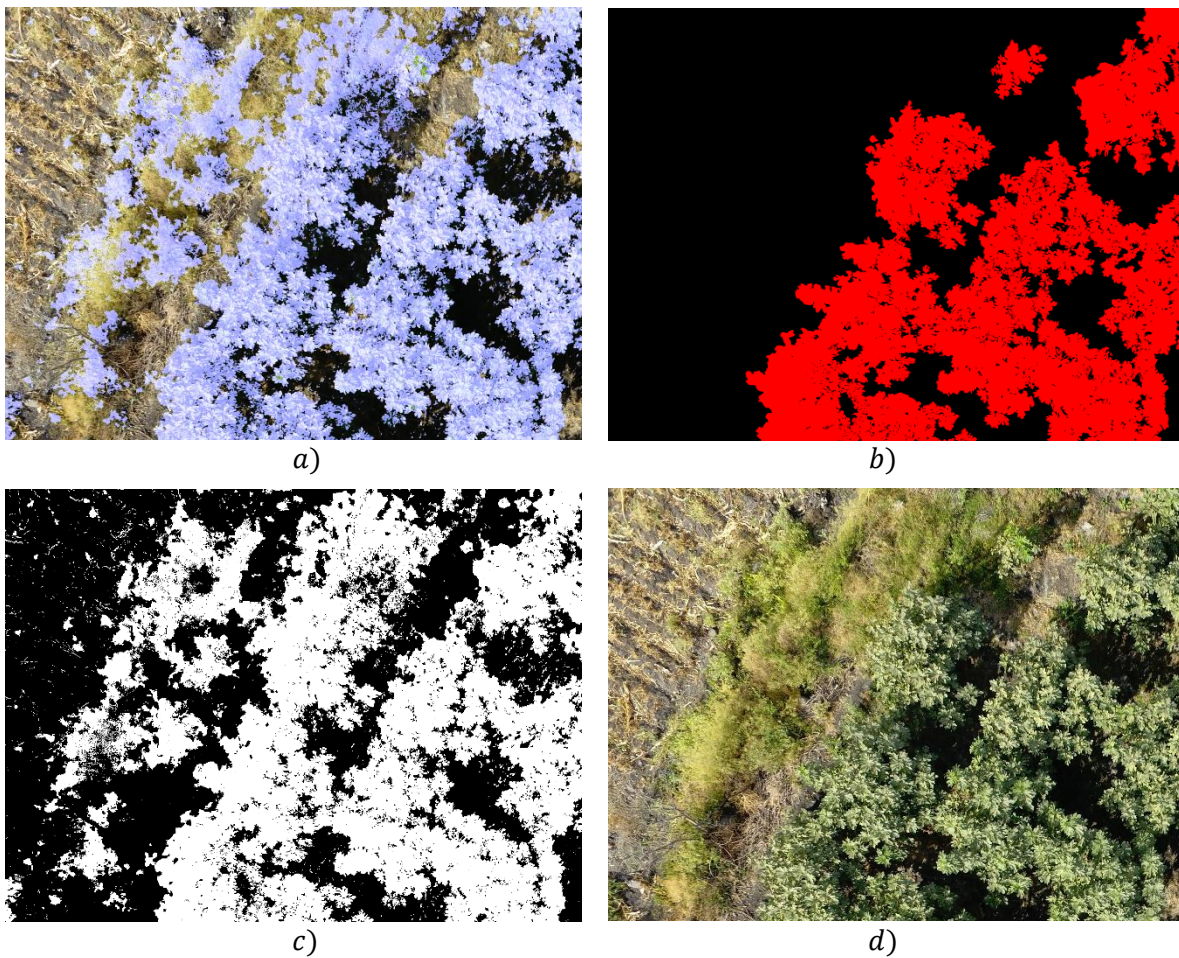


Fig.7. Imagen DJI_0010_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0010_B

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.913768	0.872614	0.829551	0.808593	0.853082
Recuerdo (TPR)	0.943442	0.875579	0.824132	0.791439	0.860304
Especificidad (TNR)	0.82456	0.8637	0.845842	0.860163	0.831371
Precisión (PPV)	0.941745	0.950767	0.941422	0.944488	0.938789
Predictivo Negativo (NPV)	0.829051	0.697807	0.615366	0.578402	0.664394
Medida-F (F)	0.942593	0.911625	0.878881	0.861217	0.897835

Tabla 2. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0010_B.

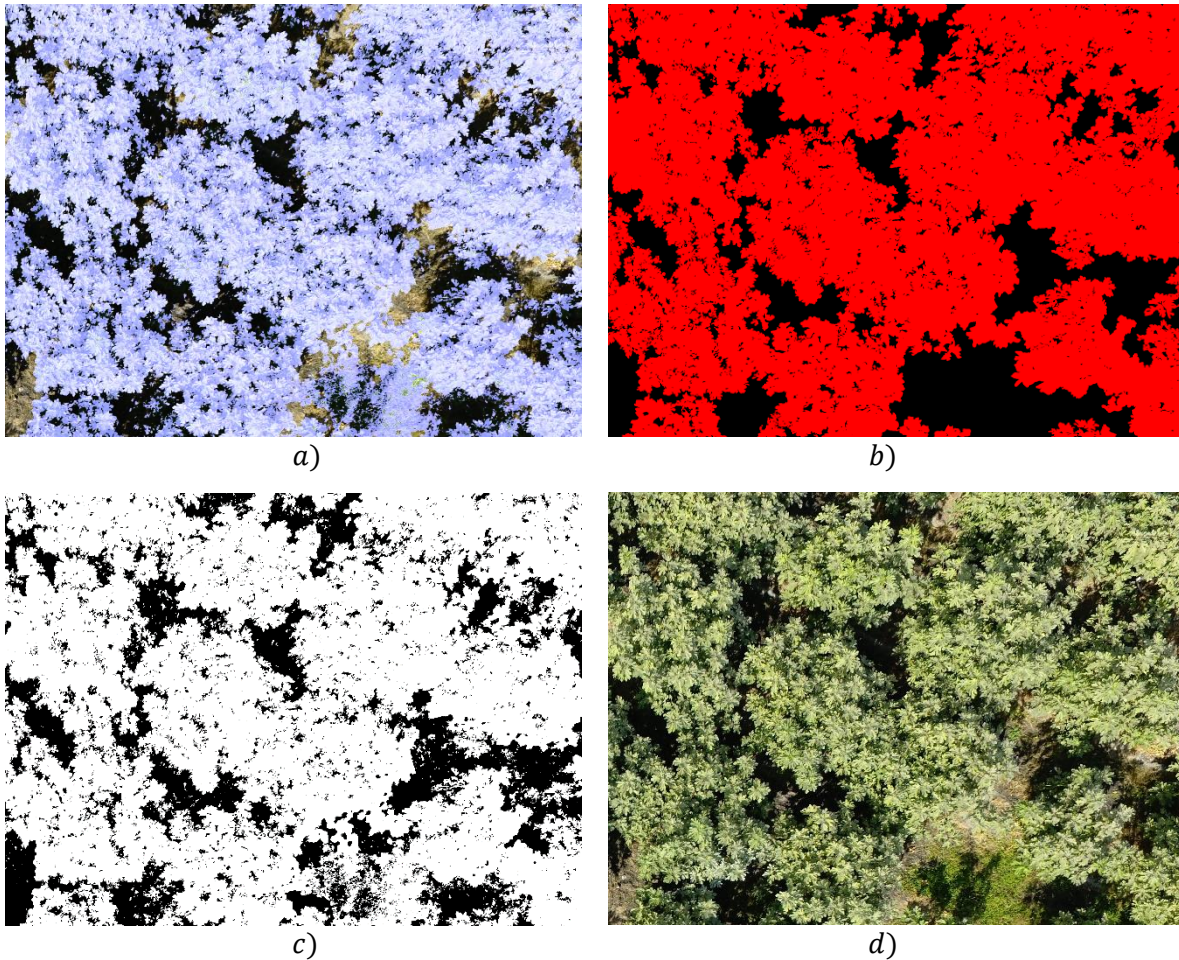


Fig.8. Imagen DJI_0010_B: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0018_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.943155	0.919375	0.880423	0.894743	0.910713
Recuerdo (TPR)	0.966118	0.918621	0.856626	0.877872	0.929
Especificidad (TNR)	0.911105	0.920426	0.913636	0.91829	0.885191
Precisión (PPV)	0.938151	0.941562	0.93263	0.93748	0.918656
Predictivo Negativo (NPV)	0.950659	0.890156	0.820331	0.843441	0.899324
Medida-F (F)	0.951929	0.92995	0.893014	0.906697	0.923799

Tabla 3. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0018_A.

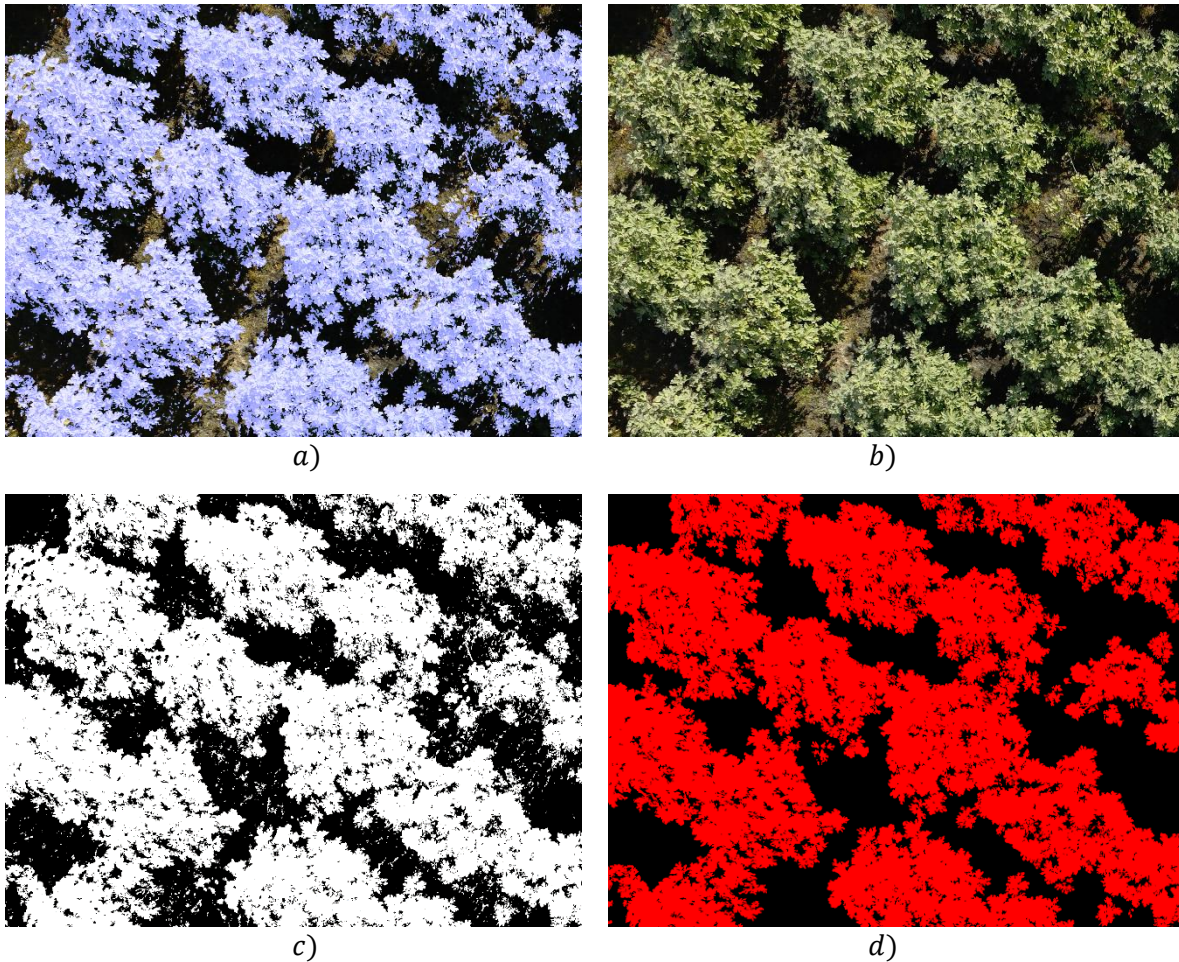


Fig.9. Imagen DJI_0018_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0036_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.930178	0.913085	0.891191	0.898783	0.880536
Recuerdo (TPR)	0.938522	0.901498	0.868444	0.87681	0.838622
Especificidad (TNR)	0.915054	0.934087	0.932421	0.938608	0.956506
Precisión (PPV)	0.952438	0.961225	0.958834	0.962806	0.972181
Predictivo Negativo (NPV)	0.891446	0.839538	0.796352	0.80783	0.765816
Medida-F (F)	0.945429	0.930404	0.911403	0.917798	0.900476

Tabla 4. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0036_A.

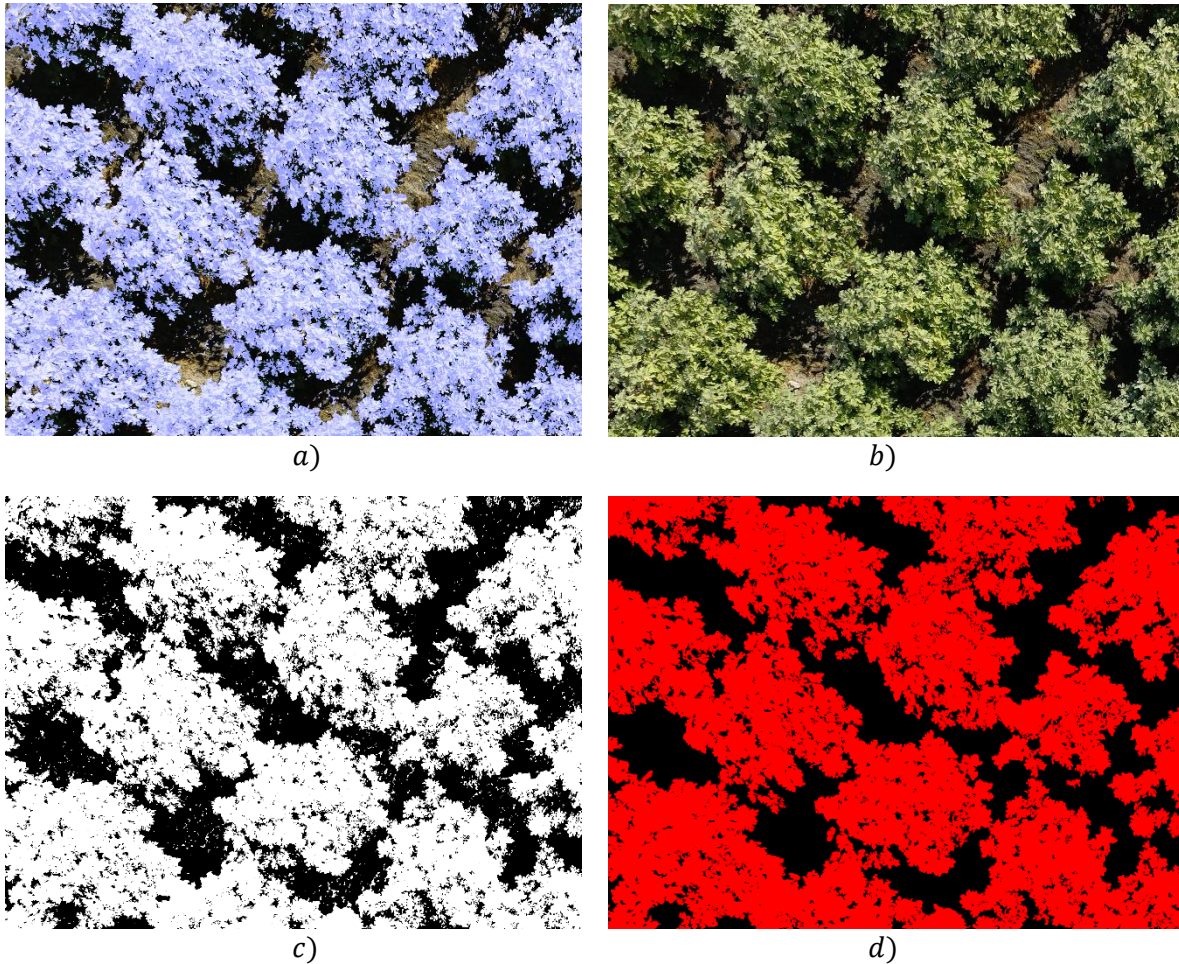


Fig.10. Imagen DJI_0036_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0043_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.941205	0.920966	0.911701	0.918407	0.931384
Recuerdo (TPR)	0.947544	0.906211	0.900193	0.907922	0.940245
Especificidad (TNR)	0.928557	0.950408	0.934663	0.939328	0.913704
Precisión (PPV)	0.963588	0.973305	0.964901	0.967594	0.956025
Predictivo Negativo (NPV)	0.898698	0.835489	0.824357	0.836406	0.884572
Medida-F (F)	0.955499	0.938561	0.931425	0.936809	0.948069

Tabla 5. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0043_A.

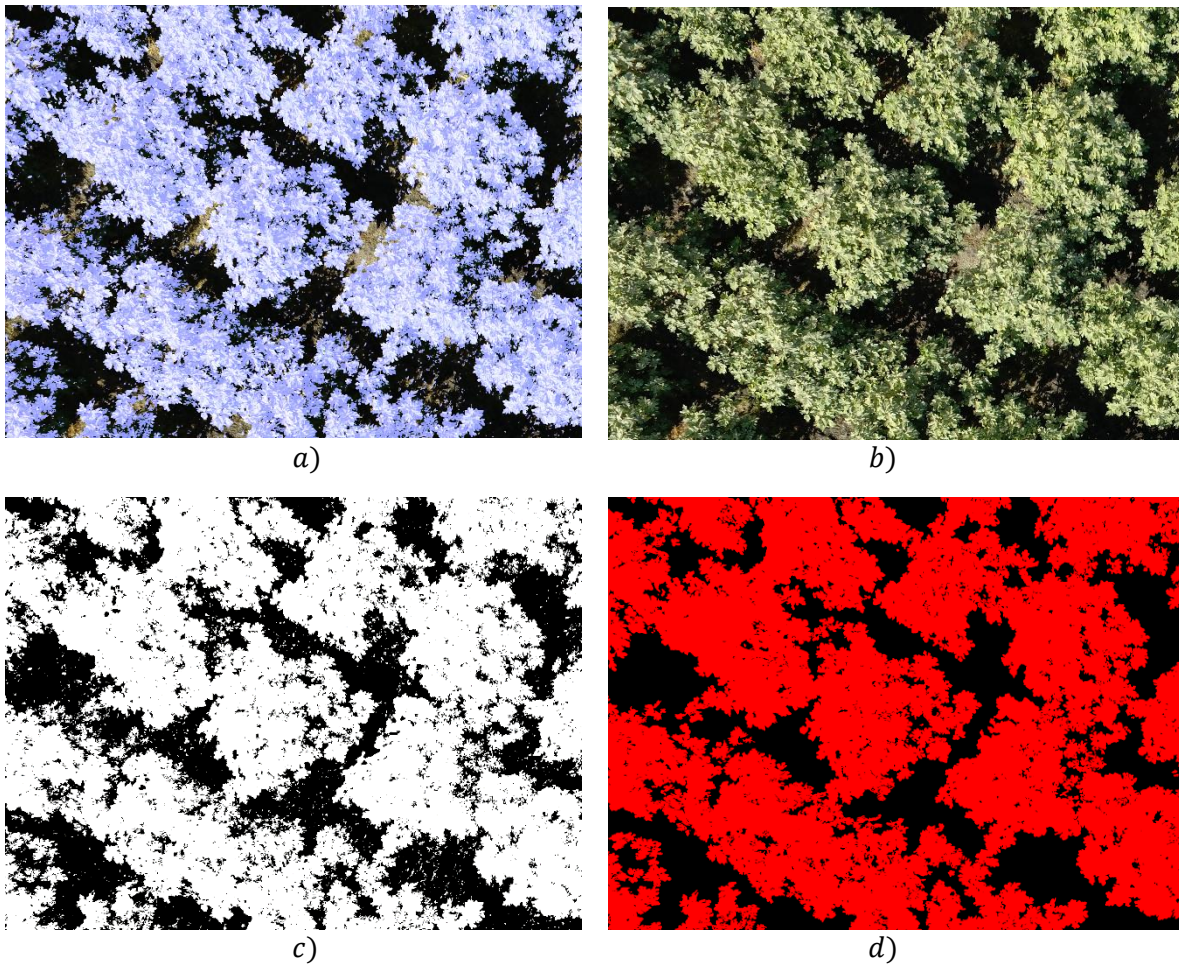


Fig.11. Imagen DJI_0043_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0051_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.899116	0.885467	0.844531	0.856487	0.874967
Recuerdo (TPR)	0.89639	0.844569	0.779673	0.805637	0.871135
Especificidad (TNR)	0.902581	0.937445	0.926959	0.921113	0.879838
Precisión (PPV)	0.921223	0.94493	0.931348	0.928465	0.902092
Predictivo Negativo (NPV)	0.872683	0.825955	0.768003	0.788536	0.843069
Medida-F (F)	0.908637	0.891935	0.848788	0.862701	0.886343

Tabla 6. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0051_A.

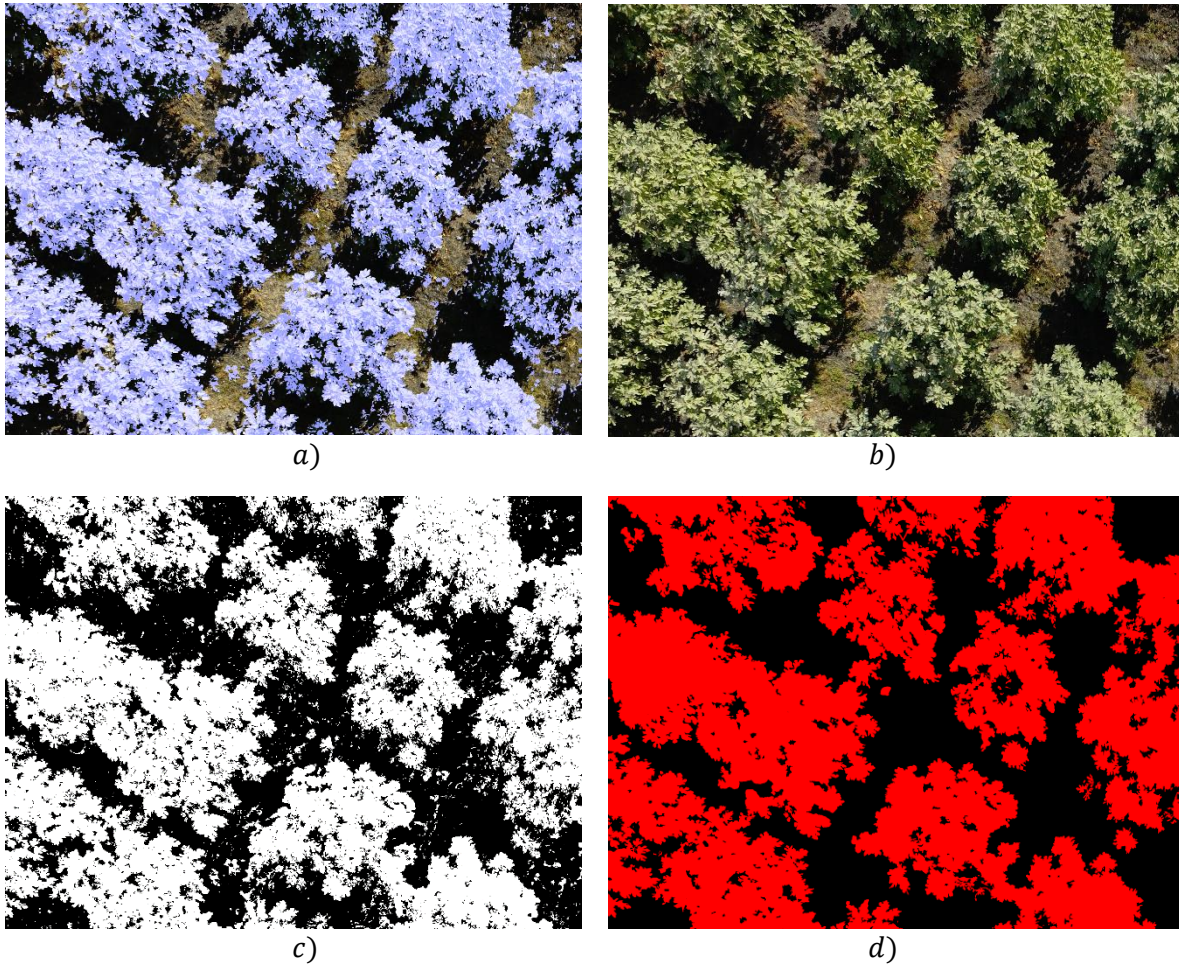


Fig.12. Imagen DJI_0051_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0075_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.921819	0.903561	0.890604	0.898002	0.917589
Recuerdo (TPR)	0.923304	0.876933	0.865002	0.876934	0.917641
Especificidad (TNR)	0.91848	0.963434	0.948168	0.945376	0.917471
Precisión (PPV)	0.962217	0.981793	0.974042	0.973044	0.96154
Predictivo Negativo (NPV)	0.841924	0.776869	0.757499	0.773573	0.832057
Medida-F (F)	0.942359	0.926405	0.91629	0.922492	0.939078

Tabla 7. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0075_A.

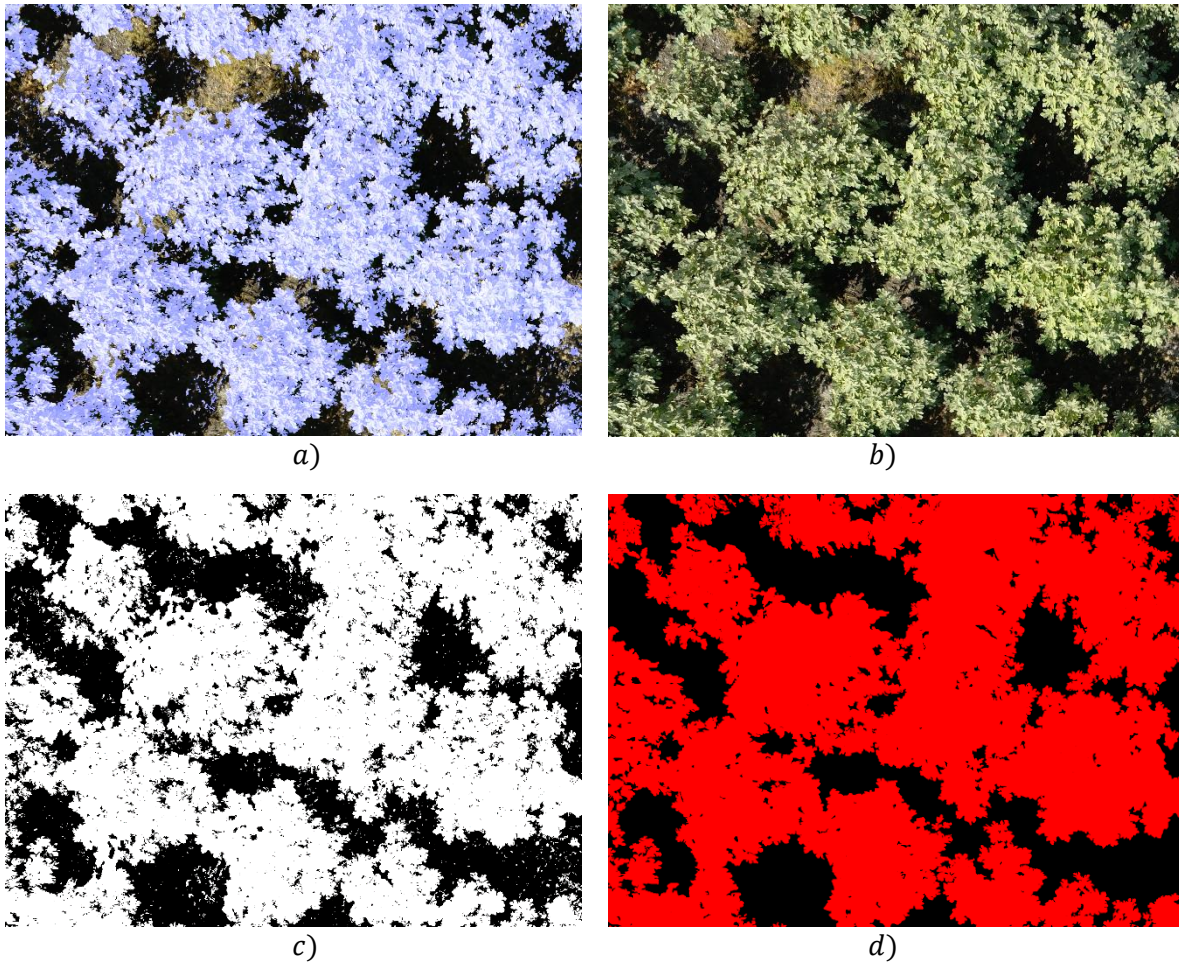


Fig.13. Imagen DJI_0075_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0083_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.922168	0.910821	0.862336	0.878482	0.883848
Recuerdo (TPR)	0.905451	0.883436	0.80525	0.834564	0.85489
Especificidad (TNR)	0.945774	0.949492	0.942948	0.940498	0.92474
Precisión (PPV)	0.959314	0.961088	0.952224	0.951936	0.941316
Predictivo Negativo (NPV)	0.876296	0.852256	0.774206	0.801031	0.818608
Medida-F (F)	0.931605	0.920628	0.872591	0.889395	0.896024

Tabla 8. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0083_A.

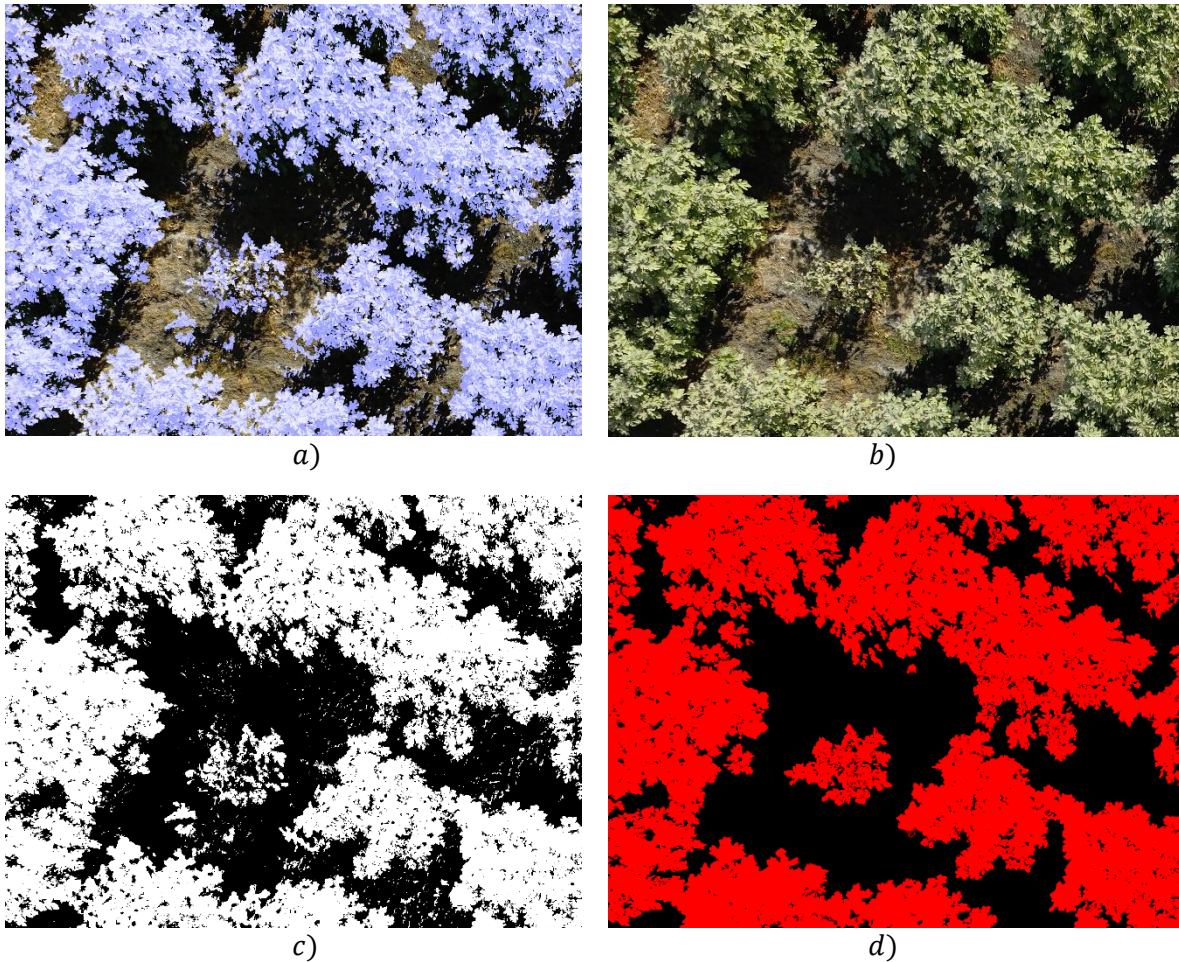


Fig.14. Imagen DJI_0083_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0098_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.823712	0.840506	0.844579	0.853802	0.797987
Recuerdo (TPR)	0.706344	0.730514	0.753709	0.774773	0.65575
Especificidad (TNR)	0.96755	0.975303	0.955943	0.950653	0.972303
Precisión (PPV)	0.963867	0.973155	0.954474	0.950596	0.966683
Predictivo Negativo (NPV)	0.728888	0.747036	0.760025	0.774985	0.697397
Medida-F (F)	0.815252	0.834556	0.842293	0.853726	0.781422

Tabla 9. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0098_A.

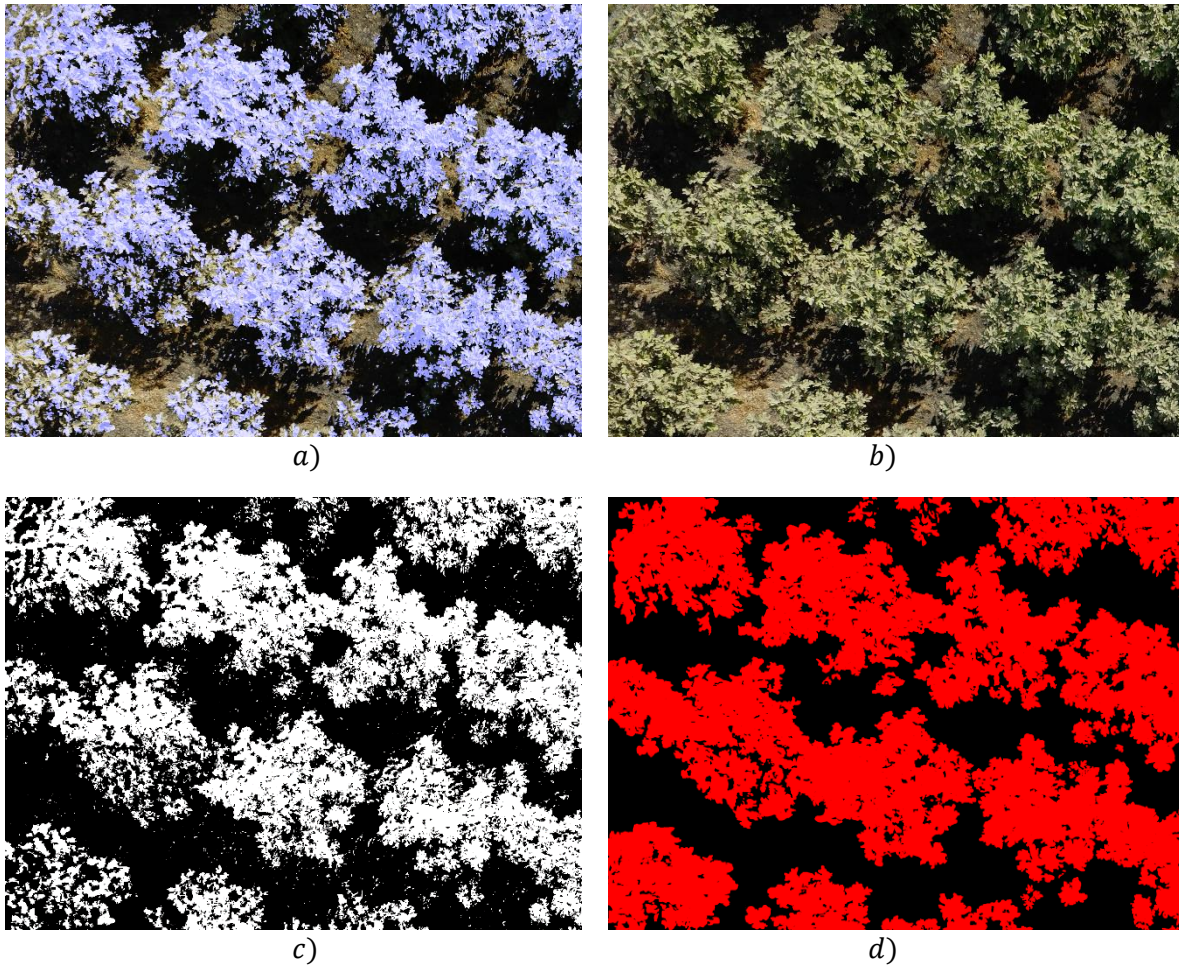


Fig.15. Imagen DJI_0098_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

Imagen DJI_0101_A

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.888749	0.877492	0.845855	0.83609	0.813151
Recuerdo (TPR)	0.88522	0.858496	0.865855	0.826591	0.711752
Especificidad (TNR)	0.890546	0.887163	0.835673	0.840926	0.864775
Precisión (PPV)	0.804596	0.794811	0.728454	0.725692	0.728243
Predictivo Negativo (NPV)	0.938421	0.924893	0.924448	0.904988	0.854919
Medida-F (F)	0.842984	0.825427	0.791234	0.772862	0.719903

Tabla 10. Comparación de los resultados obtenidos con el sistema propuesto y los índices color para la imagen DJI_0101_A.

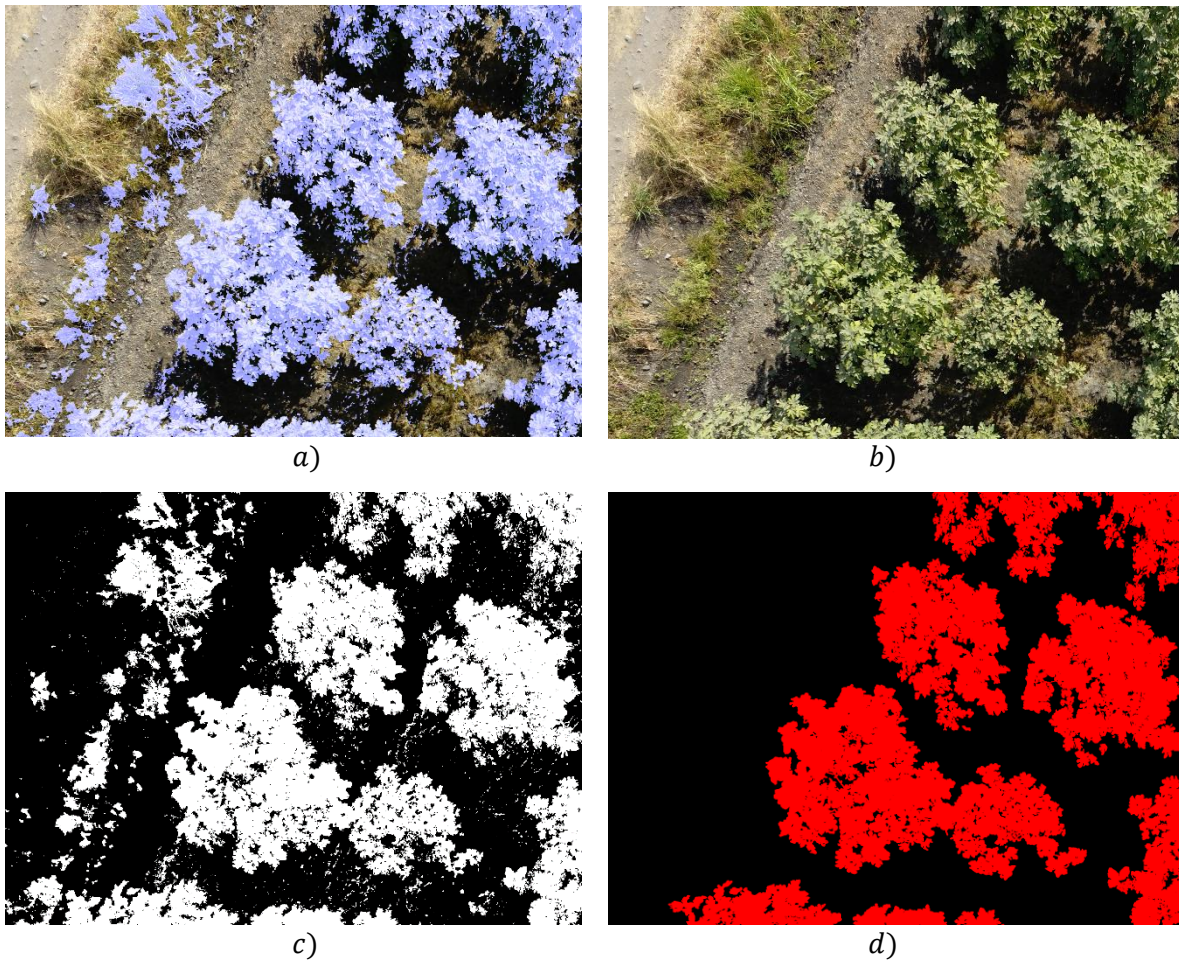


Fig.16. Imagen DJI_0101_A: a) Imagen segmentada por el sistema propuesto conservando colores originales, b) Imagen del GT, c) Imagen binaria generada por el sistema propuesto indicando los pixeles pertenecientes a la vegetación y d) Imagen original.

	Sistema propuesto	MExG	CIVE	ExG	COM2
Exactitud (OSR)	0.899116	0.885467	0.844531	0.856487	0.874967
Recuerdo (TPR)	0.89639	0.844569	0.779673	0.805637	0.871135
Especificidad (TNR)	0.902581	0.937445	0.926959	0.921113	0.879838
Precisión (PPV)	0.921223	0.94493	0.931348	0.928465	0.902092
Predictivo Negativo (NPV)	0.872683	0.825955	0.768003	0.788536	0.843069
Medida-F (F)	0.908637	0.891935	0.848788	0.862701	0.886343

Tabla 12. Media de los resultados anteriores para todos los algoritmos

5. Conclusiones y trabajos futuros

Los resultados obtenidos en este trabajo demuestran la viabilidad de los sistemas de aprendizaje de máquina no supervisados para la segmentación de cultivos en condiciones naturales de iluminación y/o sembrado. Esto constituye un importante avance en hacer a estas herramientas más accesibles para potenciales consumidores, como son dueños de cultivos, agricultores y otros, ya que elimina la necesidad de etiquetar las capturas para realizar el entrenamiento, lo cual muchas veces no resulta factible y puede afectar el rendimiento del sistema basado en la calidad del etiquetado.

Por otra parte, como el sistema propuesto fue ideado para funcionar en escenarios lumínicos no controlados, esto facilita el proceso de obtención de las imágenes del cultivo al no requerir condiciones específicas para asegurar su buen funcionamiento. No obstante, es importante mencionar que el proceso de capturas panorámicas mediante drones o dispositivos similares podría constituir un obstáculo para algunos usuarios.

Los resultados, aunque positivos, nos plantean nuevos retos, ya que es importante mejorar el reconocimiento del sembrado para evitar falsos positivos relacionados con otros tipos de vegetación también presentes en el cultivo. Así como realizar pruebas en plantaciones diferentes al higo.

El contar con un mecanismo robusto y eficiente de clasificación de plantas nos permite realizar otros procesos más complejos y de vital importancia económica y social. Ejemplo de ello es la detección automática de enfermedades y/o deficiencias de nutrientes en los cultivos, razón por la que emplearemos este sistema en trabajos futuros relacionados con esta problemática.

6. Referencias

- Aksoy, Selim, Ismet Zeki Yalniz, y Kadim Taşdemir. «Automatic Detection and Segmentation of Orhards Using Very High Resolution Imagery.» *IEEE Transactions on Geoscience and Remote Sensing*, 2012: 3117-3131.
- Alpaydin, Ethem. *Introduction to Machine Learning*. The MIT Press, 2010.
- Billingsley, J., y M. Schoenfisch. «The successful development of a vision guidance system for agriculture.» *Computers and Electronics in Agriculture*, 1997: 147-163.
- Campos, Yerania, Erik Rodner, Joachim Denzler, Humberto Sossa, y Gonzalo Pajares. «Vegetation segmentation in cornfield images using Bag of Words.» 2016.
- Changing Colorspaces*. 2018. https://docs.opencv.org/3.2.0/df/d9d/tutorial_py_colorspaces.html.
- Cuautla, El Sol de. «Morelos encabeza producción de higo en el país.» *El Sol de Cuautla*, 4 de Enero de 2017.
- Elangovan, K., y S. Nalini. «Plant Disease Classification Using Image Segmentation and SVM Techniques.» *International Journal of Computational Intelligence Research*, 2017: 1821-1828.
- Guo, Jingwei, y Xu Lihong. «Automatic Segmentation for Plant Leaves via Multiview Stereo Reconstruction.» *Hindawaj*, 2016.
- Hamuda, Esmael, Brian Mc Ginley, Martin Glavin, y Edward Jones. «Automatic crop detection under field conditions using the HSV colour space and morphological operations.» *Computers and Electronics in Agriculture*, 2017: 97-107.
- Haug, Sebastian , Andreas Michaels, Peter Biber, y Jorn Ostermann . «Plant classification system for crop/weed discrimination without segmentation.» *Applications of Computer Vision (WACV)*, 2014.
- Heraldo, Alianza Uninorte con El. «La agricultura consume el 70% del agua en el mundo.» *El Herald*, 22 de 3 de 2015.
- Kam Ho, Tim. «Random Decision Forests.» *AT&T Bell Laboratories*, 1995: 278-282.
- Katherine León, Domingo Mery, Franco Pedreschi, JorgeLeón. «Color measurement in L*a*b* units from RGB digital images.» *Food Research International*, 12 de 2006: 1084-1091.
- Kohonen, Teuvo. «Self-Organized Formation of Topologically Correct Feature Maps.» *Biological Cybernetics*, 1982: 59-69.
- Nube de Puntos*, Wikipedia. 2017. https://es.wikipedia.org/wiki/Nube_de_puntos.
- Ronghua, Ji, Fu Zetian, y Qi Lijun. «Real-time plant image segmentation algorithm under natural outdoor light conditions.» *New Zealand Journal of Agricultural Research*, 2007: 847-854.

- Roser, Max, y Hannah Ritchie. *Our World in Data - Fertilizers and Pesticides*. 2017.
<https://ourworldindata.org/fertilizer-and-pesticides> (último acceso: 27 de 3 de 2018).
- Sathyamurthy, Ravishankar , Harris Samuel, Pk Nagarajan, y S.A. El-Agouz. «A Review of Different Solar Still for Augmenting Fresh Water Yield.» *Journal of Environmental Science and Technology*, 2015.
- School, The USGS Water Science. *The World's Water*. 2 de 12 de 2016.
<https://water.usgs.gov/edu/earthwherewater.html>.
- Singh, Vijai, y A.K. Misra. «Detection of plant leaf disease using image segmentation and soft computing techniques.» *Information Processing in Agriculture*, 2017: 41-49.
- Torok, Leonardo. «Método de Otsu.» s.f.
- Whitley, Darrell. «A Genetic Algorithm Tutorial.» *Statistics and Computing*, 1994: 65-85.