

# AI-Driven Maturity Stage Identification of Amazonian Fruits

Willintong Marin, Ivan F. Mondragon, and Julian D. Colorado

**Abstract**—This paper presents a Machine Learning approach for the classification of Amazonian fruits (Moriche, Asai and Seje). Vegetative indices were used as features to drive the corresponding classification by processing RGB/VIS imagery. In this regard, we used four Machine Learning models to identify the stage of maturity for the fruits: Multi-variable regressions, Naives Bayes, Support Vector Machines and Artificial Neural Networks. These models were trained and tested with the features of each variety. Experimental results were validated by calculating ROC data, in which neural networks achieved an accuracy of 99% in the stage of maturity identification for the three amazonian varieties. These results allow us to conclude that the used vegetative indices accurately correlate with the physiological characteristics of the fruits, being relevant for the stage of maturity of the three varieties.

**Index Terms**—Image processing, Machine Learning, Vegetative Index, Amazonian Fruits Asai, Seje and Moriche).

## I. INTRODUCTION

El uso de técnicas de inteligencia artificial (IA) para identificar el estado de madurez de un fruto, conlleva dos desafíos: (i) el procesamiento de imágenes para la extracción de características relevantes, y (ii) el entrenamiento de modelos de Machine Learning (ML) que permitan asociar las características extraídas con la variable fisiológica respectiva; en este caso, el estado de madurez. El trabajo se enfoca principalmente en el primer punto, explorando qué tipo de características permiten un entrenamiento preciso de los modelos de ML. En ese sentido, se propone la extracción de la intensidad de los canales de color RGB de la imagen, incluyendo la escala de grises y el cálculo de índices de vegetación (IV).

En la literatura se encuentran avances en los métodos de clasificación de frutos con técnicas de IA en manzanas, piñas, bananas, entre otros, pero no en frutos amazónicos. Tanto los frutos como las plantas amazónicas han sido poco exploradas bajo estas técnicas de aprendizaje automático.

Este artículo propone un sistema integrado que combina las características obtenidas de los IV con las técnicas de aprendizaje automático para ser aplicados en la clasificación de las variedades de frutos Asai, Seje y Moriche. De igual manera, se estudia la descripción detallada de los frutos y las palmas en cuanto a la colorimetría que relaciona su estado de madurez con los tiempos de cosecha [1], y una clasificación del estado de madurez relacionando escala de color, descripción morfológica, tamaño en peso, diámetro y calidad [2], [3].

La Figura 1 muestra la arquitectura del sistema propuesto, la cual comprende: la captura y procesamiento de imágenes, la extracción de características, el análisis de variabilidad de los índices vegetativos, la clasificación del estado de madurez mediante diversos modelos de aprendizaje y su respectivo desempeño. Es importante resaltar, que los clasificadores (modelos de ML: Regresiones Lineales Multivariable -MLR, Naives Bayes -NB, Máquinas de Soporte Vectorial -SVM y Redes Neuronales Artificiales -ANN) fueron entrenados con bases de datos independientes, correspondientes a las imágenes de cada variedad, manteniendo como modelo los hiperparámetros en el procesamiento de las imágenes y de los modelos de ML, lo que propone la hipótesis de poder estimar el estado de madurez para las tres variedades a partir de los índices con mejor correlación.

Las palmas de Asai, Seje y Moriche, (*Euterpe precatoria*, *Oenocarpus bataua* y *Mauritia flexuosa*) [2] son variedades endémicas propias de la Amazonía. Generalmente, se ubican en las riveras de los ríos y sectores conocidos como Moricheras o cananguchales, los cuales son terrenos pantanosos y de difícil acceso. Estas palmas pueden medir hasta 40m de altura, de corona casi esférica, de hojas palmeadas de 2.5m de ancho y hasta de 4.5m de largo [1]. La fructificación para el Asai y Seje ocurre en abril y para Moriche en octubre; no obstante, estas fechas varían de acuerdo con la zona y cambios del clima [4].

El Instituto para la investigación científica de la Amazonía colombiana (SINCHI) realiza investigaciones con el ánimo de aportar con herramientas alternativas para el aprovechamiento sostenible de variedades no maderables del bosque por parte de los campesinos e indígenas de la región. Estos frutos son de gran interés internacional por su alto valor nutritivo y compuestos de bioactivos, lo que lo convierte en un potencial para el desarrollo del bosque y sus habitantes. La Figura 2 detalla algunas características morfológicas de las palmas y los frutos.

## II. REVISIÓN DE TRABAJOS RELACIONADOS

En la literatura se encuentran trabajos relacionados técnicamente con los procedimientos, herramientas y demás consideraciones técnicas con este trabajo, en cuanto a la extracción de características de las imágenes de los frutos en RGB/VIS para la identificación de su estado de madurez y entrenamiento de modelos de ML, etc., los cuales son abordados en esta sección.

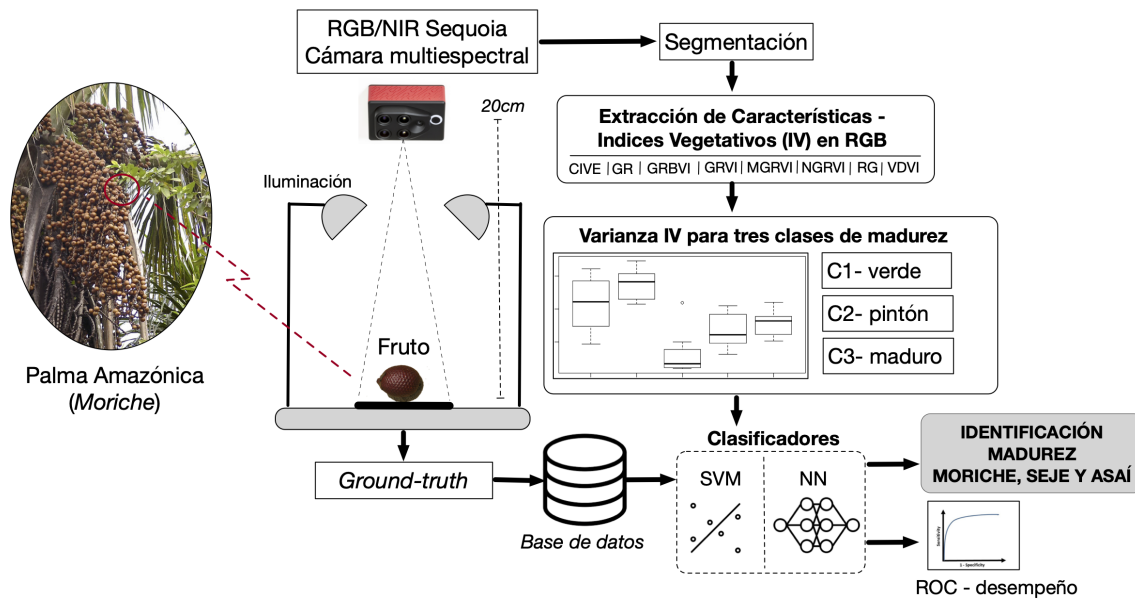


Fig. 1. Arquitectura del sistema.

#### A. Características por Intensidades de Color

En [5], se aplica un método para la extracción de características basado en el histograma de color y la distancia Kulback para identificar la variabilidad del color en la manzana, mientras que [6] propone un método basado en el espacio de color. Interesados en la selección de tomate de alta calidad, en [7] se desarrolla un sistema basado en modelos matemáticos tipo PCA, SVM y LDA, para la clasificación de un conjunto de 250 imágenes, alcanzando una precisión del 90%. En [8], se menciona que una clasificación basada sólo en características de color RGB no es suficiente para garantizar una alta precisión, por lo tanto, implementan una red de decisión logrando un proceso más riguroso. En [9], se desarrolla un sistema general de clasificación de frutas con atributos basados en el histograma 3D-RGB y modelo tipo Random Forest.



Fig. 2. Palma Amazónica Moriche (Mauritia flexuosa)

#### B. Características por Indices de Vegetación

Para estudiar el estado de las plantas, sus condiciones fisiológicas, su estado hídrico, condiciones nutricionales, establecer estimaciones de la biomasa, etc., esto a nivel del dosel

o la hoja, con técnicas de IA, los IV se han convertido en una muy buena opción para extraer características. [10] propone un método de clasificación de tierras mediante índices de vegetación basados en el espectro visible, el cual denominan VDVI (índice de vegetación de diferencia de banda visible). Los IV-RGB son sensibles al verdor de la planta y los emplea para extraer la vegetación verde en una imagen RGB [11]. La ampliación de los acrónimos aquí citados los encuentra en la Tabla I, la cual incluye la relación matemática que permite obtener los IV. Para [12] el GRVI es altamente preciso para extraer información de terrenos áridos y semiáridos, de igual manera es un indicador fenológico, eficiente para calcular la biomasa. El MGRVI ayuda ampliar las diferencias de reflectancia entre las bandas, producto de la cuadratura de estas. Adicionalmente, en [12] plantean un nuevo índice NGRVI para mejorar el contraste de los valores de brillos partiendo del MGRVI.

En [3] predicen el contenido de clorofila combinando varios índices, resaltando el gran aporte dado por RGVI. En [13] se redefine el uso de varios IV, tales como el índice de resistencia atmosférica visible (VARI) y el índice de diferencia verde-rojo normalizado (NGRDI), los cuales emplean para estimar la fracción de vegetación. Por otro lado, el NGRDI y el Índice de Vegetación Verde Roja Modificada (MGRVI) lo consideran un indicador fenológico y lo emplean para estimar la biomasa, y el índice de relación rojo-verde (RGRI) para analizar la sensibilidad angular de los índices de vegetación. Otros como el índice de exceso de verde (ExG) y el índice de color de la vegetación (CIVE) también han sido probados para identificar la vegetación verde.

Los IV normalmente son empleados para estudiar la reflectancia a nivel de la planta, del dosel y de las hojas, para así obtener información del estado de la planta, pero poco se han utilizado para estudiar las condiciones fisiológicas de los frutos directamente con técnicas de (IA) y no se encontró en la literatura aplicaciones de IA en frutos amazónicos.

TABLA I  
ÍNDICES DE VEGETACIÓN (IVS)

IV	Descripción	Fórmula
GRVI	Red Vegetation Index [14]	$\frac{G-R}{G+R}$
RGVI	Green Vegetation Index [3]	$\frac{R-G}{R+G}$
MGRVI	Modified Red Vegetation Index [14]	$\frac{G^2-R^2}{G^2+R^2}$
NGRVI	Reciprocal transformation based on MGRVI normalization [12]	$\frac{1}{MGRVI}$
GRBVI	Normalized green-red difference index [10]	$\frac{G^2-R^2}{G^2+R^2}$
VARI	Visible Atmospherically Resistant Index [14]	$\frac{G-R}{G+R-B}$
RG	Red Green Ratio [15]	$\frac{R}{G}$
GR	Red Green Ratio Index [11]	$\frac{G}{R}$
VDVI	Visible-band difference vegetation index [10]	$\frac{2G-R-B}{2G+R+B}$
EXG	Excess green [10]	$2G - R - B$
CIVE	Color Index of Vegetation [13]	$0.441R - 0.881G + 0.385B + 18.787$
NGBDI	Normalized green-blue difference index [16]	$\frac{G-B}{G+B}$
NRBDI	Normalized red-blue difference index [3]	$\frac{R-B}{R+B}$
RGBVI	Red, Green-Blue Vegetation Index [3]	$\frac{R-G}{R+G+B}$
RBGVI	Red, Blue-Green Vegetation Index [3]	$\frac{R-B}{R+G+B}$

### III. MÉTODOS Y MATERIALES

#### A. Protocolo de Captura de la Imagen

Esta fase comprende la captura de las imágenes, para lo cual se diseñó un sistema de iluminación controlada basado en luz led con 1.100 lúmenes. Se hizo uso de una estructura cerrada caja estudio, la cámara se ubica a una altura de 20 cm sobre el plano perpendicular al fruto, a una distancia paralela de 43 cm y una diagonal de 47 cm desde el fruto, tal como se puede observar en la Figura 1, el fondo en el que se ubica el fruto es de color blanco para facilitar el proceso de segmentación.

Se capturaron 2 imágenes por fruto, una por cada cara, garantizando un buen estado de la superficie. Se utilizó una cámara fotográfica de 20 megapixels con  $5120 \times 3840$  de resolución. Por último, se garantiza que el fruto se ubique en el centro de la imagen. La figura 3 es una muestra de los frutos en condición natural, es decir, sin haber sido procesada. Teniendo en cuenta que el ciclo fenológico ocurre durante todo el año, y los tiempos de la fructificación para cada variedad es diferente, se diseñó un protocolo que permitiera la recolecta de los 100 frutos por cada estado de madurez en las fechas que correspondían. Se pudo observar que, dado los cambios climáticos, las palmas no generaron la fructificación en un mes exacto, sino en una ventana de tiempo de 3 meses para cada estado de madurez. Se inició con la captura de las imágenes en el mes de enero de 2019 y se terminó en el mes de abril de 2020 para la última de las 3 variedades.



Fig. 3. Estados de madurez de las tres variedades: Asaí, Seje y Moriche.

El instituto SINCHI [2] ha tipificado estos frutos en 5 estados de madurez. Como se mencionó anteriormente, para

este estudio se tuvieron en cuenta 3 estados; estado de madurez verde como clase 1; estado de madurez pintón como clase 2 y estado de madurez maduro como clase 3. En la clase 2 se agruparon los tres tipos de pintón que tiene cada una de las tres variedades.

#### B. Procesamiento y Segmentación

Se utiliza como método de segmentación la umbralización, estableciendo un umbral para la binarización y posterior identificación de la región de interés (ROI); de esta manera separar el fruto como la ROI del fondo de la imagen y así poder extraer las características únicamente de esta región. Para facilitar lo anterior, se extrae una ventana partiendo del centro de la imagen y estableciendo 512 filas/columnas, quedando una imagen resultante de  $1024 \times 1024$ . Posteriormente se pasa a escala de grises y se aplica un filtro Gaussiano para suavizar la imagen y eliminar posibles ruidos.

#### C. Extracción de Características

Como resultado de la revisión de trabajos relacionados, se han encontrado 15 índices de vegetación (IV) en el espectro visible. La Tabla I los resume. De igual manera se extrajeron 4 atributos correspondientes a la media del histograma de las intensidades de cada canal (Ch): rojo (R), verde (G), azul (B) y gris (Gr). Los vectores de características se organizaron en 3 grupos así: Grupo 1, corresponde a los 15 IV y los 4 Ch. El grupo 2 corresponde a los 15 IV solamente y el Grupo 3 a los 8 IV seleccionados mediante un análisis de variabilidad y un proceso de clasificación inicial con el modelo NB. El sentido que tiene la división en grupos de atributos, es encontrar los que mejor resultado de clasificación generan.

#### D. Análisis de la Variabilidad de los Índices de Vegetación

Se calcula la variabilidad de los datos en función de cada variable independiente, es decir, para cada estado de madurez. Para esto, se han definido 3 estados: Verde, Pintón y Maduro, etiquetados como C1, C2 y C3 respectivamente. Mediante la

varianza es posible identificar los IV de menor variabilidad para cada clase. Para el análisis se obtuvo una gráfica por cada conjunto de datos, es decir, para las 3 clases por separado, así como una gráfica con el cálculo de los IV para las 3 clases. Las gráficas obtenidas se encuentran en la figura 4 en la sección IV.

### E. Entrenamiento de los Modelos de Aprendizaje

Para la clasificación del estado de madurez de los frutos, se utilizaron 4 modelos de ML supervisados: NB, MLR, SVM y ANN. Para cada modelo se encontraron los hiperparámetros de mejor respuesta, los cuales se muestran en la Sección IV junto con las respectivas métricas. El conjunto de datos se compone de 900 imágenes en total, para cada variedad corresponden 300 (Asaí, Seje y Moriche), cada variedad la componen 3 clases (Verde, pintón y maduro), es decir, que para cada clase se contó con 100 imágenes. Para el entrenamiento y prueba se utilizó la validación cruzada con un número de folds de 10.

### F. Implementación del Modelo

Los algoritmos ?? y ?? detallan la etapa para el procesamiento de las imágenes, la extracción de las características, el ajuste de los hiperparámetros y configuración de los modelos de ML para la etapa de clasificación. El procesamiento de las imágenes se realizó en Python y la implementación de los modelos de aprendizaje para la clasificación en la herramienta Rapidminer<sup>1</sup>.

## IV. RESULTADOS

### A. Selección de Índices de Vegetación

La Figura 4 detalla la variabilidad de los IV para el Moriche. Este análisis permitió seleccionar un grupo inicial de 12 IV entre los 15 disponibles, descartando tres índices: EXG, VARI y NGRVI para los 3 estados de madurez. Para la variedad Asaí se descartaron dos índices: EXG y VARI, y finalmente para Seje: EXG. El motivo principal para descartar estos IV fueron los altos valores atípicos (*outliers*), la dispersión y alta tendencia a la mediana.

### B. Resultados de Clasificación

Se llevó a cabo un proceso preliminar de clasificación con el modelo Naives Bayes para validar los resultados de precisión del grupo de IV resultantes del ejercicio de variabilidad, obteniendo una exactitud del 94.33%.

Persiguiendo mejorar el resultado de clasificación, se implementó un proceso de clasificación con un IV a la vez. La Tabla II detalla la exactitud obtenida cada vez que se agrega un IV al proceso de clasificación. Si la exactitud disminuye cuando se agrega un IV, este se excluye y se avanza con el siguiente, hasta lograr obtener los índices que mejor responden al proceso de clasificación.

<sup>1</sup><https://rapidminer.com>

TABLA II  
IDENTIFICACIÓN DE LOS IV QUE APORTAN AL PROCESO DE CLASIFICACIÓN (% EXACTITUD)

IV	%
RG	77
RG, CIVE	84.6
RG, CIVE, GR	90
RG, CIVE, GR, NGRBI	89
RG, CIVE, GR, VARI	87
RG, CIVE, GR, GRVI	91.3
RG, CIVE, GR, GRVI, RGVI	91.3
RG, CIVE, GR, GRVI, MGRVI	94
RG, CIVE, GR, GRVI, MGRVI, GRBVI	95
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI	96
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI, EXG	94.6
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI, NGBDI	96.6
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI, NGBDI, NRBDI	96
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI, NGBDI, RGBVI	95
RG, CIVE, GR, GRVI, MGRVI, GRBVI, VDVI, NGBDI, RBGVI	96

Como se puede observar en la Tabla II, se alcanza una exactitud de 96.67% con el grupo compuesto por 8 IV: CIVE, GR, GRBVI, GRVI, MGRVI, NGBDI, RG, VDVI.

Adicionalmente se realiza la clasificación con los índices de vegetación seleccionados y los atributos de las intensidades de cada canal, obteniéndose para el grupo de atributos No. 1 (IV & Ch) una exactitud de 92.33%; para el grupo 2 (15 IV) un 94.33% y para el grupo 3 (8 IV) un 96.67%.

La Tabla III muestra el vector de rendimiento del proceso de clasificación para el Grupo 3, se observa que el coeficiente de determinación  $R^2$  es confiable, acercándose a 1 y un error bajo.

TABLA III  
VECTOR DE RENDIMIENTO DE LA CLASIFICACIÓN POR NB

Exactitud	96.67% +/- 2.72% (promedio: 96.67%)
Error de clasificación	3.33% +/- 2.72% (promedio: 3.33%)
Error absoluto	0.044 +/- 0.027 (promedio: 0.044 +/- 0.177)
Error relativo	4.44% +/- 2.72% (promedio: 4.44% +/- 17.68%)
Error cuadrático medio	0.162 +/- 0.088 (promedio: 0.182 +/- 0.000)
Correlación cuadrada	0.872 +/- 0.095 (promedio: 0.866)

La Tabla IV, muestra la matriz de confusión del proceso de clasificación con los IV del Grupo 3, con una exactitud del 96,67%

TABLA IV  
MATRIZ DE CONFUSIÓN DE LA CLASIFICACIÓN NB, PARA EL GRUPO 3

	Verd. C1	Verd. C2	Verd. C3	Precisión de clase %
Pred. C1	98	2	5	93.33
Pred. C2	1	98	1	98
Pred. C3	1	0	94	98.95
Class recall	98	98	94	

Los siguientes resultados de clasificación con los modelos de ML restantes, se realizará con los atributos del Grupo 3, dado su mejor respuesta preliminar, al final se presentará el resultado comparativo basado en la exactitud para todos los grupos de atributos.

### C. Clasificación por Regresión Lineal Multivariable

A continuación, se presentan los resultados de aplicar la regresión lineal multivariable. Las siguientes ecuaciones son las regresiones para el estado de madurez de cada clase. (1) corresponde al estado de madurez de la clase 1 (EMC1); (2) corresponde al estado de madurez de la clase 2 (EMC2) y (3) corresponde al estado de madurez de la clase 3 (EMC3).

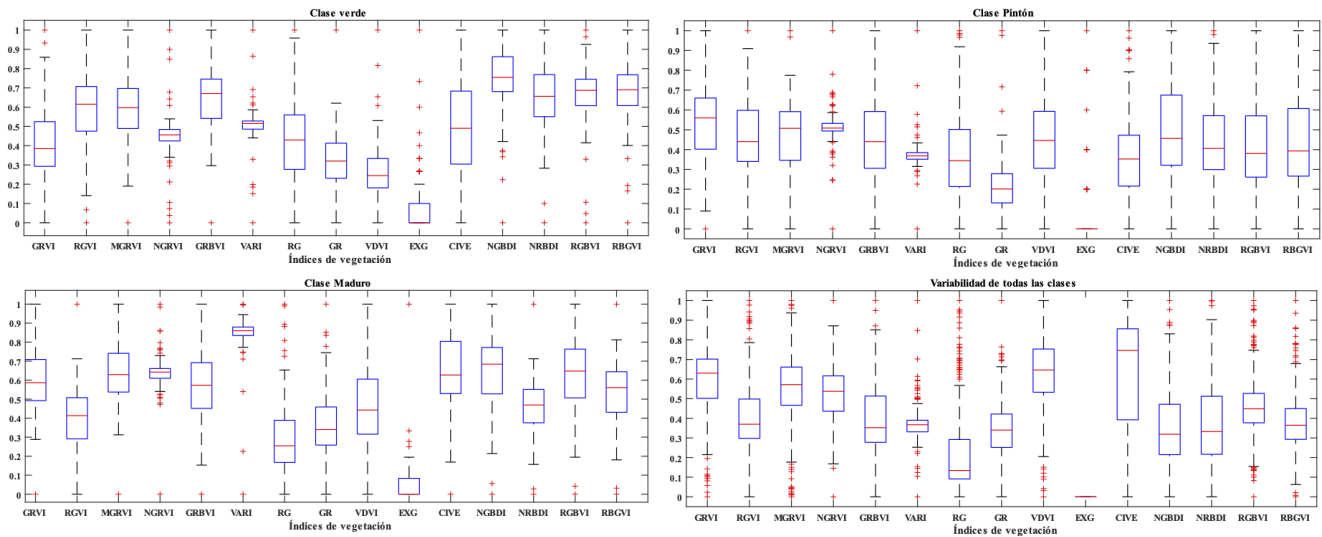


Fig. 4. Análisis de variabilidad de los índices vegetativos para el Moriche.

$$EMC1 = -33.7384 + 10.7311GRVI + 13.2205MGRVI - 1.1264GRBVI + 7.4373RG - 1.068GR + 10.8747VDVI + 2.7143CIVE + 18.6374NGBDI \quad (1)$$

$$MSC2 = 9.0440 + 1.1196MGRVI - 19.733RG - 2.8046GR + 0.3795VDVI - 8.6318CIVE - 2.4574NGBDI \quad (2)$$

$$MSC3 = 11.7088 - 5.659GRVI - 130278MGRVI + 2.3543GRBVI + 6.5144RG + 2.1576GR - 8.4771VDVI + 6.7606CIVE - 13.5213NGBDI \quad (3)$$

La Tabla V muestra el vector de rendimiento del proceso de clasificación por regresiones. Se observa que el coeficiente de determinación  $R^2$  es bastante confiable.

TABLA V  
VECTOR DE RENDIMIENTO CLASIFICACIÓN POR MLR

Exactitud	98.67% +/- 2.33% (promedio: 98.67%)
Error de clasificación	1.33% +/- 2.33% (promedio: 1.33%)
Error absoluto	0.028 +/- 0.021 (promedio: 0.028 +/- 0.106)
Error relativo	2.79% +/- 2.14% (promedio: 2.79% +/- 10.62%)
Error cuadrático medio	0.086 +/- 0.072 (promedio: 0.110 +/- 0.000)
Correlación cuadrada	0.981 +/- 0.032 (promedio: 0.980)

Este método de clasificación por regresión multivariable arroja resultados con una exactitud del 98.67%. La Tabla VI, muestra la matriz de confusión, donde se puede observar que, para la clase 3, la precisión es del 100% y para la clase 2 y 1 es del 99% y 98% respectivamente.

TABLA VI  
MATRIZ DE CONFUSIÓN DE LA MLR, PARA EL GRUPO 3

	Verd. C1	Verd. C2	Verd. C3	Precisión de clase %
Pred. C1	98	1	0	98.99
Pred. C2	2	98	0	98
Pred. C3	0	1	100	99.01
Class recall	98	98	100	

D. Clasificación con Máquina de Soporte Vectorial (SVM)

Los resultados de clasificación con el modelo SVM alcanzan una exactitud de 97.33%. En la Tabla VII se dan a conocer los hiperparámetros del modelo SVM que mejor respuesta alcanzó.

TABLA VII  
HIPERPARÁMETROS SVM

Parámetro	Valor o tipo %
SVM type	un-SVC
Kernel type	rbf
Gamma	0.0
un	0.5
Cache size	80
Epsilon	0.001

La Tabla VIII muestra el vector de rendimiento, se observa que el coeficiente de determinación  $R^2$  disminuye levemente, pero mantiene un valor confiable cerca de 1.

TABLA VIII  
VECTOR DE RENDIMIENTO CLASIFICACIÓN POR SVM

Exactitud	97.33% +/- 4.10% (promedio: 97.33%)
Error de clasificación	2.67% +/- 4.10% (promedio: 2.67%)
Error absoluto	0.027 +/- 0.041 (promedio: 0.027 +/- 0.161)
Error relativo	2.67% +/- 4.10% (promedio: 2.67% +/- 16.11%)
Error cuadrático medio	0.110 +/- 0.128 (promedio: 0.163 +/- 0.000)
Correlación cuadrada	0.964 +/- 0.054 (promedio: 0.961)

La clasificación por SVM baja levemente la exactitud pasando al 97.33%. La Tabla IX, muestra la matriz de confusión, se observa que, la clase 3, la precisión es del 98% la clase 2 y 1 es del 97.92% y 96.12% respectivamente.

TABLA IX  
MATRIZ DE CONFUSIÓN CLASIFICACIÓN POR SVM

	Verd. C1	Verd. C2	Verd. C3	Precisión de clase %
Pred. C1	99	2	0	98.02
Pred. C2	1	94	1	97.92
Pred. C3	0	4	99	96.12
Class recall	99	94	99	

E. Clasificación con Redes Neuronales Artificiales (ANN)

Los resultados de clasificación de este modelo mejoran con respecto a las anteriores. Se logra una exactitud del 99% bajo los parámetros consignados en la Tabla X, aquí se dan a conocer los hiperparámetros del modelo ANN.

TABLA X  
HIPERPARÁMETROS ANN

Parámetro	Valor o tipo
Capas ocultas	1
Función de activación	Sigmoide
Número de Neuronas	8
Ciclos de entrenamiento	$243 \leq CE \leq 647$
Tasa de aprendizaje	0.01
Momentum	0.9
Epsilon	0.0001

Importante ajustar los ciclos de entrenamiento (CE) entre los valores indicados, de lo contrario el resultado de la exactitud se afecta.

La Tabla XI, muestra el vector de rendimiento, se observa que el coeficiente de determinación  $R^2$  aumenta y se convierte en el mejor alcanzado entre todos los modelos, acercándose a 1.

TABLA XI  
VECTOR DE RENDIMIENTO CLASIFICACIÓN POR ANN

Exactitud	99.00% +/- 1.61% (micro promedio: 99.00%)
Error de clasificación	1.00% +/- 1.61% (promedio: 1.00%)
Error absoluto	0.044 +/- 0.021 (promedio: 0.044 +/- 0.105)
Error relativo	4.36% +/- 2.15% (promedio: 4.36% +/- 10.45%)
Error cuadrático medio	0.098 +/- 0.059 (promedio: 0.113 +/- 0.000)
Correlación cuadrada	0.986 +/- 0.023 (promedio: 0.985)

La clasificación por ANN incrementa la exactitud pasando al 99%, convirtiéndose así, entre el mejor modelo de clasificación en este trabajo. La Tabla XII, muestra la matriz de confusión, donde se puede observar que, para la clase 3, la precisión es del 98%, para la clase 2 y 1 es del 97.92% y 96.12% respectivamente.

TABLA XII  
MATRIZ DE CONFUSIÓN CLASIFICACIÓN POR ANN

	Verd. C1	Verd. C2	Verd. C3	Precisión de clase %
Pred. C1	99	1	0	99
Pred. C2	1	99	1	98.02
Pred. C3	0	0	99	100
Class recall	99	99	99	

### F. Resultados Consolidados

La Tabla XIII consolida los resultados de clasificación en función de los grupos de atributos, de acuerdo con la métrica de la exactitud. La primera columna corresponde a los modelos de ML aplicados, la segunda, tercera y cuarta columnas muestran los resultados de implementar los modelos en los Grupos 1, 2 y 3 de atributos.

TABLA XIII  
CONSOLIDADO DE LOS RESULTADOS DE CLASIFICACIÓN EN FUNCIÓN DE LOS GRUPOS DE ATRIBUTOS

ML	% Grupo 1	% Grupo 2	% Grupo 3
NB	92.33	94.33	96.67
MLR	96.67	99	98.67
SVM	79	97.33	97.33
ANN	99	99	99

La peor clasificación para el fruto Moriche la realiza el modelo SVM con el 79% para el Grupo 1 de atributos; no obstante, logra mejorar a un 97.33% con los grupos de atributos 2 y 3. La mejor clasificación se logra con el modelo

ANN para todos los grupos de atributos. La Figura 5, muestra la curva ROC resultante del proceso de clasificación de los 4 modelos. Se puede observar que el punto de corte con mayor exactitud y estabilidad se logra con ANN a una sensibilidad de 1 y una especificidad de 0.05, seguida de MLR con una diferencia muy pequeña, la curva ROC está basada en el grupo 3 de atributos.

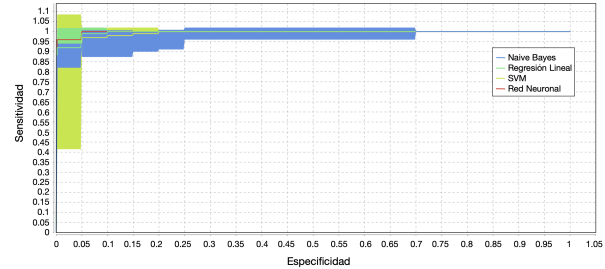


Fig. 5. Curva ROC. Proceso comparativo de los resultados de clasificación entre los 4 modelos de ML para la variedad Moriche.

### G. Resultados de la Clasificación en los Frutos Asaí y Seje

Se hizo una exploración para la clasificación de las variedades Asaí y Seje, partiendo del modelo obtenido en el entrenamiento de la variedad Moriche, pero el resultado obtenido no fue el esperado, el modelo termina clasificando todas las clases como una misma; así que, se determinó volver a entrenar los modelos con los conjuntos de datos obtenidos de las variedades Asaí y Seje, manteniendo los hiperparámetros en las etapas de preprocesamiento, procesamiento de las imágenes para la extracción de características y obtención de los atributos Grupos 1, 2 y 3 y posterior aplicación de la etapa de entrenamiento de los 4 modelos de ML, obteniendo resultados similares al entrenamiento con la variedad Moriche.

La Tabla XIV resume los resultados de clasificación de la variedad Asaí. Las flechas hacia arriba y hacia abajo son un indicador que el resultado de clasificación sube o baja respectivamente con respecto a los resultados del fruto Moriche. Se puede destacar que hubo una mejora considerable con el modelo SVM pasando de 79 a 88.33, todos los demás resultados tienen mejoras en menor proporción. Los grupos de atributos que mejor respondieron fueron los Grupos 1 y 2, mientras que, el Grupo 3 tuvo una caída en la exactitud, eso sí, muy leve.

TABLA XIV  
CONSOLIDADO DE LOS RESULTADOS DE CLASIFICACIÓN DEL FRUTO ASAÍ. LAS FLECHAS INDICAN UN AUMENTO (↑) O DECREMENTO (↓) EN LA EXACTITUD.

ML	% Atributos Grupo 1	% Atributos Grupo 2	% Atributos Grupo 3
NB	96.67 ↑	96 ↑	95.67 ↓
MLR	96.67	99	97.33 ↓
SVM	88.33 ↑	97 ↓	95 ↓
ANN	99	99	97.67 ↓

La Tabla XV resume los resultados de clasificación de la variedad Seje. Las flechas hacia arriba o hacia abajo son un indicador que el resultado de clasificación sube o baja respectivamente con respecto a los resultados del fruto Moriche. Se puede destacar que hubo una mejora considerable con el

modelo SVM pasando de 79 a 88.33 en Asaí y a 94.67 en Seje, todos los demás resultados tienen ascensos y descensos muy leves. El grupo de atributos que mejor respondió fue el Grupo 3.

TABLA XV

CONSOLIDADO DE LOS RESULTADOS DE CLASIFICACIÓN DEL FRUTO SEJE. LAS FLECHAS INDICAN UN AUMENTO (↑) O DECREMENTO (↓) EN LA EXACTITUD.

ML	% Atributos Grupo 1	% Atributos Grupo 2	% Atributos Grupo 3
NB	95.33 ↑	96.33 ↑	96.67
MLR	95.67 ↓	98 ↓	98.67
SVM	94.67 ↑	95.67 ↓	96.67 ↓
ANN	98 ↓	97.33 ↓	98 ↓

La tabla XVI consolida los resultados de clasificación en términos del promedio de cada variedad y los grupos de atributos. La columna 2 se obtiene de hallar el promedio de la exactitud por la variedad Asaí, en el grupo 1 de atributos. Con esta tabla se puede determinar que, el grupo que en promedio mejor clasifica es el grupo 2 con el 97.33% y la variedad mejor clasificada fue Seje, con el 96.75%, obteniendo una mínima diferencia.

TABLA XVI

PROMEDIOS DE CLASIFICACIÓN, GRUPOS DE ATRIBUTOS Y VARIEDADES EN %.

variedad	Grupo 1	Grupo 2	Grupo 3	Promedio
Asaí	95.42%	97.75%	96.42%	96.53%
Seje	95.92%	96.83%	97.50%	96.75%
Moriche	91.75%	97.42%	97.92%	95.69%
Promedio	94.36%	97.33%	97.28%	

La tabla XVII consolida los resultados de clasificación en términos del promedio de cada modelo de ML y las variedades. La columna 2 se obtiene de hallar el promedio de la exactitud por el modelo NB para la variedad Asaí. Con esta tabla, se puede determinar, que el modelo de ML que en promedio mejor clasifica es ANN con el 98.44% y la clasificación más baja fue con SVM con el 93.44%.

TABLA XVII

PROMEDIOS DE LA CLASIFICACIÓN, VARIEDADES VS MODELOS DE ML EN %.

ML	Asaí	Seje	Moriche	Promedio
NB	96.11%	96.11%	94.44%	95.56%
MLR	98%	97.45%	98.11%	97.85%
SVM	93.44%	95.67%	91.22%	93.44%
ANN	98.56%	97.78%	99%	98.44%
Promedio	96.53%	96.75%	95.69%	

## V. CONCLUSIONES

No se encuentran aportes suficientes en la literatura orientados a la clasificación de la madurez de estas variedades amazónicas con técnicas de IA, lo que determina un campo poco explorado y una oportunidad de contribución. Por el contrario, sí se encuentra en la literatura una alta exploración de los índices vegetativos en el espectro visible para estimar cambios en las plantas, más no enfocados al fruto directamente.

Los resultados de este trabajo indican que no todos los índices vegetativos (IV) hacen un aporte significativo para la

clasificación del estado de la madurez de los frutos de estudio. Este trabajo logra identificar 8 IV que tienen un comportamiento confiable. Basados en la exactitud, el modelo de ML que mejor clasificación generó para la variedad Moriche fue ANN con un 99%. El grupo de atributos que mejor respondió en promedio entre todos los modelos de ML fue el grupo 3, correspondiente a los 8 IV seleccionados, con un promedio del 97.91%, seguido del grupo 2 con 97.33% y el grupo 1 con 91.75%.

En términos de comportamiento entre las 3 variedades, el modelo de ML que en promedio logró una mejor clasificación fue ANN con el 98.44%, seguida de MLR con 97.85%, NB 95.55% y SVM 93.44%. El grupo de atributos que aportó una mejor identificación de la madurez, fue el grupo 2 con un 97.33%, el más bajo el grupo 3 con 94.36%.

La variedad mejor clasificada fue Seje con 96.75%, con una diferencia marginal respecto a las otras. Estos resultados demuestran que la clasificación del estado de madurez haciendo uso de IV en el espectro visible, es una técnica viable. El aporte de las características obtenidas mediante la media de las intensidades por cada canal de color es insignificante en el momento de trabajar con características obtenidas mediante los IV.

Finalmente, se comprobó que los índices extraídos y los modelos entrenados en la variedad Moriche no se pudieron aplicar a las variedades Asaí y Seje, pero sí se lograron calibrar hiperparámetros en el procesamiento de las imágenes y en el entrenamiento de los modelos de ML, para ser aplicados a las 3 variedades. El trabajo futuro estará enfocado en el uso de índices vegetativos en diferentes longitudes de onda, calculados mediante el procesamiento de imágenes hiperespectrales. Adicionalmente, se propone encontrar relaciones morfológicas y fisiológicas entre el dosel y los frutos.

## AGRADECIMIENTOS

Este trabajo es financiado por el Departamento del Guaviare y Ministerio de Ciencia, Tecnología e Innovación (MinCiencias) en Colombia. Convocatoria No. 752 de 2016. *Formación de capital humano de alto nivel para el Departamento del Guaviare, 2016 - II corte, por la financiación de los estudios de doctorado*. Los autores agradecen al Instituto Amazónico SINCHI por su apoyo a la investigación. Asimismo, se agradece al Programa OMICAS por su financiación parcial de los equipos para esta investigación: Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles, Convocatoria Colciencias Ecosistema Científico Colombiano financiado por el Banco Mundial, MinCiencias, ICETEX, Ministerio de Educación Nacional y Ministerio de Industria y Comercio Cont. ID: FP44842-217-2018, OMICAS Award ID: 792-61187.

## REFERENCIAS

- [1] J. Hernández, M.S., Castro, S.Y., Giraldo, B., & Barrera, *Seje, moriche, asaí: Palmas amazónicas con potencial*, primera ju ed., Bogotá D.C., 2018.
- [2] SINCHI, "Fichas Palmas amazónicas con potencial Seje, Moriche y Asaí," 2018.
- [3] X. Ma, J. Feng, H. Guan, and G. Liu, "Prediction of chlorophyll content in different light areas of apple tree canopies based on the color characteristics of 3d reconstruction," *Remote Sensing*, vol. 10, no. 3, 2018.

- [4] A. Montero, I.M., Barrera, J.A., Giraldo, B., Lucena, "Fichas Técnicas de Especies de uso Forestal y Agroforestal de la Amazonia Colombiana," Bogotá D.C., 2016.
- [5] A. Taparugssanagorn, S. Siwamogsatham, and C. Pomalaza-Ráez, "A non-destructive oil palm ripeness recognition system using relative entropy," *Computers and Electronics in Agriculture*, vol. 118, pp. 340–349, oct 2015. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0168169915002938>
- [6] S. Cárdenas-Pérez, J. Chanona-Pérez, J. V. Méndez-Méndez, G. Calderón-Domínguez, R. López-Santiago, M. J. Perea-Flores, and I. Arzate-Vázquez, "Evaluation of the ripening stages of apple (Golden Delicious) by means of computer vision system," *Biosystems Engineering*, vol. 159, pp. 46–58, 2017.
- [7] N. El-Bendary, E. El Hariri, A. E. Hassanien, and A. Badr, "Using machine learning techniques for evaluating tomato ripeness," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1892–1905, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2014.09.057>
- [8] A. Marín-cano, "Sistema de clasificación por visión artificial de mangos tipo Tommy Classification system for artificial vision type Tommy mango," no. June, 2015.
- [9] D. Heras, "Fruit image classifier based on artificial intelligence," *Revista Killkana Técnica.*, vol. 1, no. 2, pp. 21–30, 2017. [Online]. Available: [https://doi.org/10.26871/killkana\[\\_\]tecnica.v1i2.79](https://doi.org/10.26871/killkana[_]tecnica.v1i2.79)
- [10] W. Xu, Y. Lan, Y. Li, Y. Luo, and Z. He, "Classification method of cultivated land based on uav visible light remote sensing," *International Journal of Agricultural and Biological Engineering*, vol. 12, no. 3, pp. 103–109, 2019.
- [11] H. Cen, L. Wan, J. Zhu, Y. Li, X. Li, Y. Zhu, H. Weng, W. Wu, W. Yin, C. Xu, Y. Bao, L. Feng, J. Shou, and Y. He, "Dynamic monitoring of biomass of rice under different nitrogen treatments using a lightweight UAV with dual image-frame snapshot cameras," *Plant Methods*, vol. 15, no. 1, pp. 1–17, 2019.
- [12] X. Zhang, F. Zhang, Y. Qi, L. Deng, X. Wang, and S. Yang, "New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV)," *International Journal of Applied Earth Observation and Geoinformation*, vol. 78, no. December 2018, pp. 215–226, 2019. [Online]. Available: <https://doi.org/10.1016/j.jag.2019.01.001>
- [13] L. Wan, Y. Li, H. Cen, J. Zhu, W. Yin, W. Wu, H. Zhu, D. Sun, W. Zhou, and Y. He, "Combining UAV-based vegetation indices and image classification to estimate flower number in oilseed rape," *Remote Sensing*, vol. 10, no. 9, 2018.
- [14] C. A. Devia, J. P. Rojas, E. Petro, C. Martinez, I. F. Mondragon, D. Patino, M. C. Rebolledo, and J. Colorado, "High-Throughput Biomass Estimation in Rice Crops Using UAV Multispectral Imagery," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 96, no. 3-4, pp. 573–589, 2019.
- [15] D. García-cervigón and J. José, "Estudio de Índices de vegetación a partir de imágenes aéreas tomadas desde UAS / RPAS y aplicaciones de estos a la agricultura de precisión ." p. 78, 2015.
- [16] M. Du and N. Noguchi, "Monitoring of wheat growth status and mapping of wheat yield's within-field spatial variations using color images acquired from UAV-camera System," *Remote Sensing*, vol. 9, no. 3, 2017.



**Iván F. Mondragón B.** (Colombia, 1980) se tituló como Ingeniero Eléctrico (BSEE), Universidad Nacional de Colombia, 2002; Magister en Ingeniería Electrónica y de Computadores, Universidad de los Andes (Colombia), 2005; doctorado en Automática y Robótica, Universidad Politécnica de Madrid (España), 2011. Desde 2013, es profesor a tiempo completo y director del Centro Tecnológico de Automatización Industrial (CTAI), Departamento de Ingeniería Industrial de la Pontificia Universidad Javeriana. Actualmente está trabajando en visión artificial aplicada a vehículos aéreos no tripulados, robótica industrial así como en sistemas de fabricación flexible FMS, inspección de calidad, realidad virtual (sistema inmersivos tipo CAVE) e Industria 4.0.



**Julián D. Colorado** Profesor Asociado del Departamento de Electrónica de la Pontificia Universidad Javeriana en Bogotá, Colombia. Completó su Ph.D (2012) y M.Sc (2010) en Robótica y Automática en la Universidad Politécnica de Madrid, España, donde estudió el desarrollo de nuevos controladores de vuelo (GNC) para vehículos aéreos no tripulados, incluyendo cuadricópteros y drones biológicamente inspirados. Fue investigador visitante en la Universidad de Brown USA (2010-2011), donde estudió cómo integrar actuadores inteligentes basados en aleaciones con memoria (SMA) para controlar la modulación en robots aéreos de alas mórficas. Ha dirigido y participado en proyectos Europeos FP7, y fue miembro de la comisión de la Unión Europea, para definir legislación y normativa para el uso de UAVs en el espacio aéreo Europeo (Bruselas, Bélgica). Autor de más de 50 publicaciones científicas en revistas indexadas ISI/SCOPUS, conferencias internacionales y capítulos de libro. Sus intereses de investigación incluyen robótica de campo, agricultura de precisión, robótica aérea, robótica bio-inspirada y sistemas de control.



**Willington Marín Rodríguez** Ingeniero electrónico, 2006, Universidad de Cundinamarca, Magister en Administración de Empresas, 2013, Universidad Externado de Colombia, actualmente adelanta estudios de doctorado en Ingeniería en la Pontificia Universidad Javeriana. Con interés en el campo de la investigación orientada a la agricultura de precisión para explorar alternativas de aprovechamiento sostenible en la Amazonía Colombiana. Con de 14 años de experiencia profesional, desempeñándose en sectores y actividades de la administración y la ingeniería.