

CONAMA 2020

CONGRESO NACIONAL DEL MEDIO AMBIENTE

Soluciones innovadoras para una enfermedad en bananos:

Diagnóstico de la marchitez del banano utilizando diferentes algoritmos



Autor Principal: Barlin Orlando Olivares (Universidad de Córdoba, España)

Otros autores: Andrés Vega (Consejo Nacional de Investigaciones Científicas y Tecnológicas, Argentina); M. Angélica Rueda Calderón (Consejo Nacional de Investigaciones Científicas y Tecnológicas, Argentina); Juan Carlos Rey (Instituto Nacional de Investigaciones Agrícolas, Venezuela); Miguel Araya-Alman (Universidad Católica del Maule, Chile); José A. Gómez (Instituto de Agricultura Sostenible, España); Blanca Landa (Instituto de Agricultura Sostenible, España); Juan Antonio Navas (Instituto de Agricultura Sostenible, España); Deyanira Lobo (Universidad Central de Venezuela).

Cita: Olivares BO, Vega A, Rueda MA, Araya-Alman M, Rey JC, Gómez JA, Landa B, Navas JA, Lobo D. 2021. Soluciones innovadoras para una enfermedad en bananos: diagnóstico de la marchitez del banano utilizando diferentes algoritmos. En: Fundación CONAMA (eds.). Congreso Nacional del Medio Ambiente. Madrid, España. p. 1-19.

SOLUCIONES INNOVADORAS PARA UNA ENFERMEDAD EN BANANOS: DIAGNÓSTICO DE LA MARCHITEZ DEL BANANO UTILIZANDO DIFERENTES ALGORITMOS

Resumen

La incidencia de la marchitez del banano 'Cavendish' se caracteriza por el marchitamiento progresivo, de rápida propagación, y acción violenta, afectando la superficie sembrada e incidiendo directamente en la producción de banano en la región central de Venezuela. El objetivo de este estudio fue diagnosticar la incidencia de marchitez del banano en una finca de Venezuela utilizando información de suelo mediante el desempeño de los algoritmos Random Forest (RF), máquinas de soporte vectorial (Support Vector Machines, SVM) y análisis discriminante lineal (ADL). Se realizó un muestreo sistemático en 39 lotes de una finca bananera, donde se determinó la granulometría (contenidos de arena, limo y arcilla), el contenido de nutrimentos (fósforo, potasio, calcio, sodio, magnesio, cobre, zinc, manganeso y hierro), pH, materia orgánica y conductividad eléctrica. También, se determinó la incidencia de la marchitez del banano durante el periodo 2016-2017. Se particionó el conjunto de datos de la siguiente forma: El 60% del conjunto de datos se utilizó para entrenar los modelos RF, ADL y SVM con Kernel lineal y radial, el 15% para testear y el 25% para validar los modelos. El área bajo la curva ROC (Característica Operativa del Receptor) se utilizó para evaluar el desempeño de los modelos ajustados, indicando que el valor mayor obtenido presenta el mejor ajuste. El área bajo la curva de RF, ADL, SVM con Kernel lineal y con Kernel radial fue de 0,93, 0,79, 0,76 y 0,86, respectivamente. La sensibilidad, especificidad y precisión para RF fue de 77.8%, 63.0% y 70.3%, respectivamente, siendo RF el algoritmo que se desempeñó mejor que el resto, obteniendo una mejor estimación de la incidencia de la marchitez en lotes de banano. De esta forma, se evidenció que el RF permitiría predecir la incidencia de marchitez en suelos lacustrinos de Venezuela con buena precisión y, aplicado adecuadamente, puede ser una herramienta eficaz para la toma de decisiones en campo. Además, la utilización de información de suelo en zonas bananeras de Venezuela permitió identificar lotes con alta y baja incidencia de la marchitez del banano a través de un algoritmo de aprendizaje automático como RF. Este estudio, permitiría evidenciar que, se podría anticipar la predisposición al desarrollo de la enfermedad identificando las características de suelo de una finca. Finalmente, la cuarentena de los sitios y desinfección de los implementos de manejo, entre otros, son medidas para controlar la marchitez del banano y evitar su propagación a áreas sanas de cultivo.

Palabras clave: aprendizaje automático, agricultura sostenible, banano, marchitez, suelo.

Introducción

Los bananos (*Musa spp*) representan un cultivo de suma importancia para la economía de Venezuela, basada predominantemente en el petróleo. Durante los últimos 20 años la producción bananera ha sufrido ciertos cambios atribuidos principalmente al desabastecimiento de agroinsumos (semillas, fertilizantes, agroquímicos), problemas de acceso a divisas para satisfacer la demanda interna, y a la inadecuada gestión de las políticas agrícolas, la afectación de la sequía, plagas y enfermedades (FAO, 2019).

Entre los principales factores limitantes de la producción de bananos en la región central de Venezuela se encuentran la ocurrencia de sequias meteorológicas (Olivares, 2018; Cortez et al. 2018), el ataque del gorgojo negro del banano (*Cosmopolites sordidus* (Germar)) (Rey et al. 2006; 2009); y la severa afectación de superficies de bananos a causa de las enfermedades como la sigatoka negra (*Mycosphaerella fijiensis* Morelet), sigatoka amarilla (*Mycosphaerella musicola* Leach et Mulder) y la marchitez del banano (MB) generada por un complejo hongo-bacteria y los cuales afectaron el potencial del cultivo como producto para la exportación.

En las zonas bananeras del estado Aragua de Venezuela, la MB está diezmando la producción de banano Cavendish desde el 2006, reduciendo sus consecuencias en el rendimiento del rubro desde el 2010 (Martínez et al. 2016; Rey et al. 2016). La MB es un desorden fisiológico y metabólico; en cuyo origen podría jugar un papel importante la unión de factores bióticos y abióticos, como las condiciones físicas, químicas y los microorganismos del suelo potencialmente patógenos (Domínguez et al. 2001). Pero lamentablemente hasta la fecha no se ha podido descubrir el agente causal de la enfermedad, por lo que su control y prevención es complicada.

Actualmente, la MB puede ser confundida con la marchitez por *Fusarium oxysporum* f.sp. cubense (Foc) considerada como una de las más destructivas de banano a nivel mundial (Ploetz, 2015; Dita et al. 2018). Específicamente la Raza 4 Tropical (TR4) ha causado graves pérdidas en el Sudeste Asiático, afectando gravemente la subsistencia de los pequeños y medianos productores, extendiéndose al continente africano y al Medio Oriente (Karangwa et al. 2018), causando preocupaciones de su propagación al subcontinente indio y América Latina (Pérez-Vicente & Porras, 2015) con consecuencias que pudieran ser devastadoras para los productores bananeros, y sobre todo para la cadena de valor del banano.

Los diversos estudios sobre la epidemiología de la enfermedad establecen necesario considerar los efectos del tipo del suelo (Dita et al. 2018). Estas observaciones demuestran que la propagación de la marchitez en bananos en algunas zonas bananeras fue más rápida en unas regiones que en otras, lo que lleva a correlacionar la incidencia de ese tipo de enfermedad con las propiedades específicas del suelo, como textura (Deltour et al. 2017), pH (Li et al. 2018), capacidad de intercambio de cationes (Bosman, 2016), sales solubles totales, nutrientes disponibles (Segura et al. 2015), materia orgánica y drenaje (Lahav & Israeli, 2000).

El aprendizaje automático o Machine Learning permite tanto identificar patrones entre una cantidad considerable de datos que pueden ser de diferente naturaleza como predecir comportamientos a través de algoritmos capaces de aprender y evolucionar basándose en su propia experiencia (Chemura et al. 2017; Ma et al 2017; Ye et al., 2020). Con miras a anticipar la enfermedad de MB y conseguir un diagnóstico más preciso, surge el objetivo de este trabajo que es evaluar el desempeño de diferentes algoritmos, estos son, análisis discriminante lineal (LDA), máquinas vectoriales de soporte con Kernel Lineal (SVMkl), Random Forest (RF) y

máquinas vectoriales de soporte con Kernel Radial (SVMkr). Generando así, información valiosa capaz de detectar estas alteraciones en las plantas de banano y su correlación con las variables de suelo de manera objetiva y rigurosa.

Esta información resultaría muy valiosa para comprender en un sentido más amplio el problema fitosanitario que representa esta enfermedad en las plantaciones de bananos y sus relaciones con los demás componentes que conforman su manejo agronómico. La determinación de estas áreas permitiría ejecutar protocolos de prevención de la enfermedad enfocados en las áreas críticas para evitar su propagación en la finca considerando variables de suelo de fácil medición y económicas para el agricultor.

Metodología

Área de estudio

El área de estudio abarca una superficie de 180 ha de banano (*Musa paradisiaca* L) subgrupo Cavendish, ubicada en el municipio Libertador del estado Aragua, Venezuela. El clima es tropical de sabana con una precipitación promedio anual de 1100 mm y una evaporación promedio anual entre 1800-2200 mm (Hernández et al. 2017). Las lluvias son estacionales con 5 a 6 meses húmedos ubicados entre los meses de mayo-junio y octubre-noviembre (Olivares, 2018). Los suelos en su gran mayoría son lacustrinos, con texturas medias alta disponibilidad de nutrimentos, altos pH y con condiciones salinas en forma localizada (Viloria et al. 2003; Delgado et al. 2010).

Muestreo de suelos

Se realizó un muestreo sistemático siguiendo los lineamientos de (Lozano et al. 2004), con una distancia aproximada de 150 m entre sitios de muestreo, comprendiendo 90 puntos distribuidos en los 38 lotes de bananos de la finca. Se obtuvieron muestras compuestas en cada uno de los lotes de banano. Las muestras fueron sometidas a análisis de suelos con fines de fertilidad, determinando la proporción en porcentaje de arena (a), limo (L) y arcilla (A) (Gee & Or, 2002), la reacción del suelo (pH), conductividad eléctrica (CE) (dS/m) en suspensión 1:2 (suelo: agua) (Soil Survey Staff, 2014), los contenidos de MO: materia orgánica (%); K: potasio (mg/kg); Ca: calcio (mg/kg); Mn: manganeso (mg/kg); Fe: hierro (mg/kg); Zn: zinc (mg/kg), Cu: cobre (mg/kg); S: azufre (mg/kg) y fósforo (P) (mg/kg) (Mehlich, 1984).

Incidencia de marchitez en bananos (IMB)

Se evaluó la incidencia de la enfermedad, calculada como la proporción entre el número de plantas enfermas y el número total de plantas observadas (Akter et al. 2015) en cada lote de terreno cultivado durante los años 2016 y 2017 en los 39 lotes totales.

Agrupamiento de las variables del suelo

Se evaluaron dos algoritmos de agrupamiento, UPGMA y K-means, utilizando las variables de suelo estandarizadas como variables de entrada. UPGMA utilizó la distancia euclídea como índice de similitud. El número óptimo de conglomerados se determinó en base a 18 índices para agrupar las variables de suelo mediante el paquete NbClust (Charrad et al.2014), que aplica la regla de la mayoría (Kautzky et al.2015), estos índices son: 1-CH (Calinski & Harabasz 1974), 2-

Hartigan (Hartigan 1975), 3-C-Index (Hubert & Levin 1976), 4-DB (Davies & Bouldin 1979), 5-Silhouette (Rousseeuw 1987), 6-Ratkowsky (Ratkowsky & Lance 1978), 7-Ball (Ball & Hall 1965), 8-Ptbiserial (Milligan 1981), 9-McClain (McClain & Rao 1975), 10-Gamma (Baker & Hubert 1975), 11-Gplus (Milligan 1981), 12-Tau (Milligan 1981), 13-SDbw (Halkidi & Vazirgiannis 2001), 14-KL (Krzanowski & Lai 1988), 15-Duda (Duda & Hart 1973), 16-Beale (Beale 1969), 17-Gap (Tibshirani et al 2001), and 18-SD (Halkidi et al 2000). Por lo tanto, el número de conglomerados sugerido para el agrupamiento de las variables de suelo es el que presente una mayor frecuencia de los índices evaluados.

Validación de los agrupamientos

Para comparar los conglomerados realizados por estos algoritmos, se utilizaron los siguientes índices de validación interna: el índice de conectividad el cual está relacionado a la distancia entre objetos en un mismo conglomerado, mientras más bajo sea el valor de este índice es mejor (Kassambara 2017), el índice de ancho de silueta el cual mide la confianza con la que una observación es asignada a un conglomerado (Sekula et al. 2017) y el índice Dunn es el cociente entre la mínima distancia entre dos objetos que no pertenecen a un mismo conglomerado y la máxima distancia entre dos objetos de un mismo conglomerado. Combina la compactación (homogeneidad dentro del conglomerado) con el grado de separación entre conglomerados. (Dalton et al. 2009). Estos índices fueron calculados a través del paquete optCluster en R (Sekula et al. 2017).

Desbalanceo en las categorías de IMB

Para abordar el tema de desbalanceo presente en la variable categórica IMB (diferente cantidad de clases de incidencia alta o baja), se aplicó una técnica de pre-procesamiento de datos conocida como re-muestreo o muestreo. Esta técnica consiste en realizar modificaciones directas sobre los elementos de la base de datos con el fin de generar un equilibrio entre las distintas clases que la componen (Galindo Huerta, 2018; He & Garcia, 2009). Se usaron en este trabajo tres variantes de remuestros que son: submuestreo (undersampling), sobremuestreo (oversampling) y sobremuestreo aleatorio (ROSE). La técnica undersampling trabaja sobre la clase mayoritaria del conjunto de datos y consiste en la eliminación de muestras dentro de esta clase de tal forma que se genere un equilibrio entre clases. Por otro lado, tenemos la técnica de remuestreo oversampling, que en contra posición a la previamente nombrada, realiza su trabajo sobre la clase minoritaria del conjunto de datos, añadiendo muestras a esta clase de tal forma que se equilibre la diferencia entre clases. El sobremuestreo aleatorio (ROSE) produce una muestra sintética, posiblemente equilibrada, de datos simulados según el enfoque de remuestreo suavizado Bootstrap (Longadge & Donge, 2013; Lunardon et al., 2014; Guo et al. 2017).

Modelación estadística

Para el conjunto de datos se realizaron 30 submuestras aleatorias de la siguiente forma: 60% para entrenamiento, 15% para validación, 25% para testeo. Esto se realizó con el fin de identificar el mejor tipo remuestreo acorde a nuestro conjunto de datos. Para cada una de las submuestras se aplicaron los siguientes tres tipo de remuestros: undersampling, oversampling y ROSE. La librería de R que se usó en esta instancia fue ROSE (Menardi & Torelli, 2014; Lunardon et al. 2014). Esta librería permitió abordar problemas de clasificación binaria cuando hay presencia de clases desbalanceadas. Para medir el desempeño de los algoritmos evaluados se consideró: el área bajo la curva ROC, la sensibilidad, la especificidad y la precisión.

Resultados

Características del suelo bananero

Los resultados del análisis descriptivo (figura 1), indicaron importantes diferencias entre las características de los suelos de los lotes, ocurriendo suelos de textura franca a franco limosa, suelos afectados ligeramente por sales (Lotes 1-5), así como la variabilidad en los contenidos de nutrientes (fósforo y potasio) y las amplias diferencias en contenidos de calcio y magnesio; y en las relaciones de las bases cambiables (Calcio/Magnesio /Potasio) reflejadas en la figura 2.

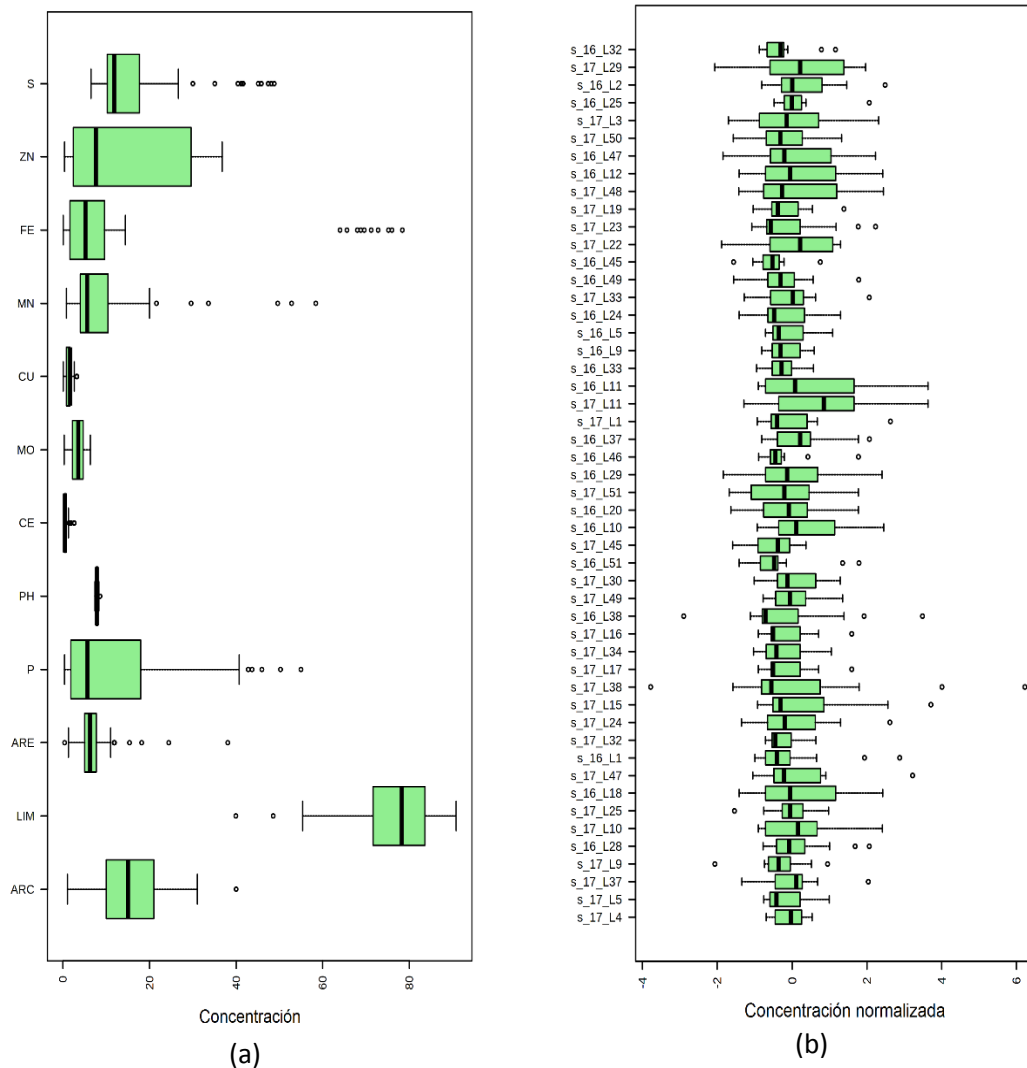


Figura 1. Diagramas de caja y antes y después de la normalización. (a) Concentración de algunas variables bajo estudio: ARC: arcilla (%); LIM: Limo (%); ARE: Arena (%); P: fósforo (mg/kg); pH: reacción del suelo; CE: conductividad eléctrica (ds/m); MO: materia orgánica (%); Mn: manganeso (mg/kg); Fe: hierro (mg/kg); Zn: zinc (mg/kg); Cu: cobre (mg/kg); S: azufre (mg/kg). (b) Concentración de lotes de bananos muestreados normalizados. En la figura se presentan 50 muestras (o sitios de muestreo) como referencia.

En algunos lotes el pH fue ligeramente más alto que en otros, lo cual se relacionó con altos niveles de Na y deficiencia de Fe y Mn. Por otra parte, las características particulares del material parental de los suelos generan niveles muy altos de Ca, lo cual se asocia con muy altas relaciones

Ca/Mg; Ca/K e incluso existen altas relaciones Mg/K, que provocan desbalance nutricional y dificultad en la absorción de estos elementos. Por otra parte, los contenidos de Materia Orgánica y Zinc fueron las únicas variables químicas que presentaron valores normales en todos los suelos

En la mayoría de los lotes predominan las texturas francas (FL/F – FL), con predominancia de partículas con diámetro equivalente entre 2 y 50 μm , con valores bajos de densidad aparente (0,45 a 0.89 Mg.m^{-3}), lo cual es de esperarse, ya que se trata de suelos de origen lacustrino, cuyo material parental le proporciona tal característica. Por esta razón, los valores de Porosidad Total son bastante altos (63 a 78 %). Estas características permiten que estos suelos tengan una tasa de infiltración moderada a moderadamente alta.

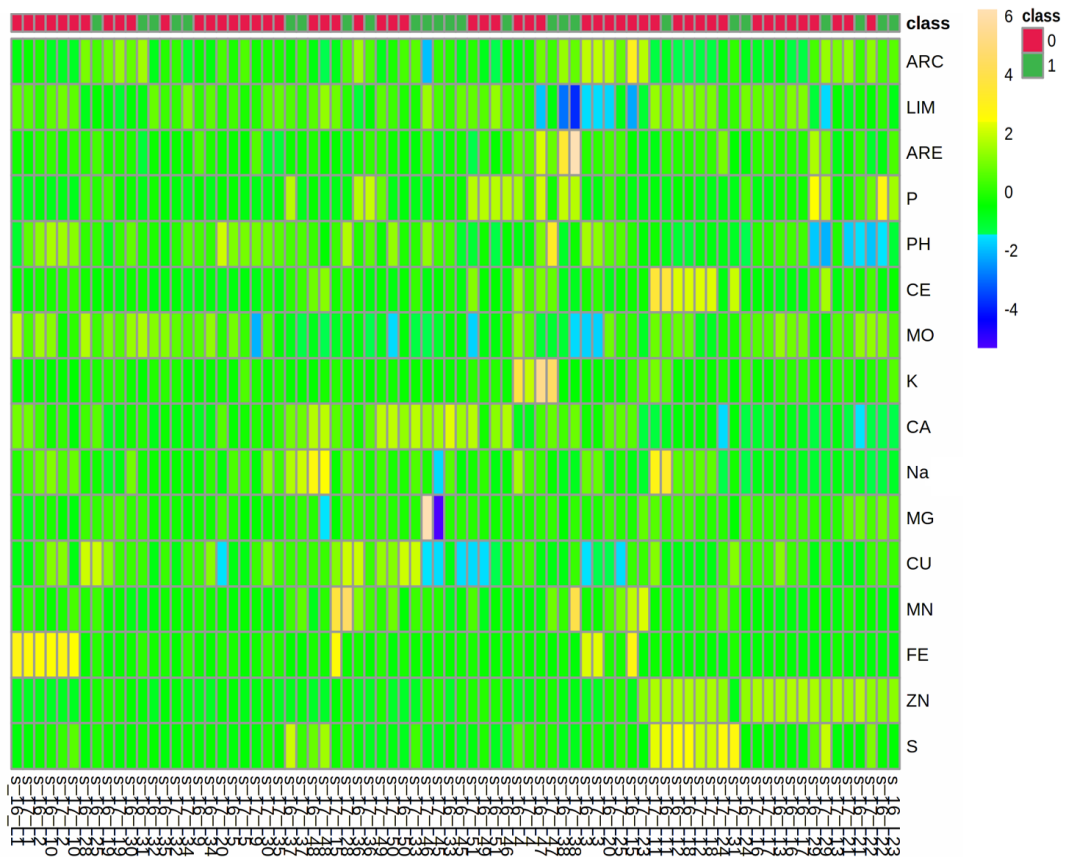


Figura 2. Mapa de calor generado a partir de datos de suelo de los lotes bananeros evaluados en el 2016 y 2017, el cual representa los valores de concentración de las variables de suelo (color azul a amarillo) para el periodo de estudio.

Incidencia de la marchitez del banano (IMB)

La figura 3 muestra el transcurso de la incidencia de marchitez del banano en el sitio evaluado. Para el periodo 2016 la mayor incidencia ocurrió en el lote 38 con 5.57% y el menos afectado fue el lote 3 con 0.54%. Con relación al periodo 2017, la mayor proporción de incidencia se registró en el lote 36 con 8.47% mientras que el lote 17 presentó una baja incidencia (0.11%).

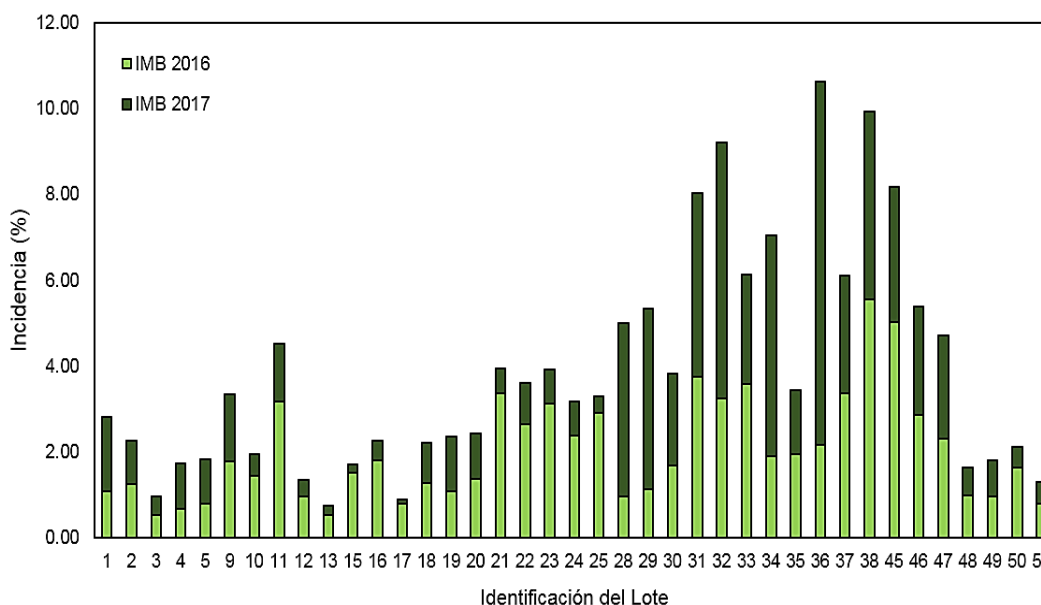


Figura 3. Transcurso de la incidencia de la marchitez del banano (IMB) durante el periodo 2016-2017 en la zona bananera.

Agrupamiento de las variables del suelo

Para los algoritmos de agrupamiento evaluados, el número óptimo de conglomerados fue dos. En el cuadro 1 se muestran los ocho índices que señalaron la presencia de dos grupos para el método UPGMA y los seis índices que señalan la presencia de dos grupos para el método K-means.

Cuadro 1. Número de índices que clasificaron las variables del suelo en dos conglomerados (k=2), para los métodos UPGMA y k-means.

Método	Numero de índices	Descripción de los índices
UPGMA	8	CH, C-Index, Silhouette, Duda, Beale, Gap, McClain, Gplus
K-means	6	CH, Duda, Gap, McClain, SD, and SDbw

En general, para los métodos UPGMA y K-means, los índices de validación identificaron 2 conglomerados para el agrupamiento de las variables de suelo (cuadro 2). Todos los métodos de conglomerado evaluados sugieren que la conectividad más baja y los valores de silueta más altos se obtuvieron para k = 2. Sin embargo, el índice de Dunn, no fue sensible en identificar el número óptimo de conglomerados, dando valores diferentes para los dos métodos de agrupamiento (Cuadro 2).

Cuadro 2. Índices de validación de conglomerados de variables de suelo formados por dos métodos de conglomerado

Índices de validación [§]	Métodos [¥]	
	UPGMA	K-means
Conectividad	4.77 (2)	8.80 (2)
Ancho de silueta	0.37 (2)	0.35 (2)
Dunn	0.47 (5)	0.44 (3)

¥UPGMA: Método de agrupamiento jerárquico aglomerativo; K-means: método de agrupamiento no jerárquico.

§Conectividad: un valor bajo indica que los objetos dentro de un conglomerado son más similares. Ancho de Silueta: un valor alto indica una alta confianza con la que se asigna una observación a un grupo. Dunn: un valor alto implica una alta varianza entre grupos y una baja varianza dentro de grupo. La cantidad óptima de conglomerados entre paréntesis.

La regla de la mayoría se aplicó con el fin de identificar el número de grupos relacionados con las categorías de IMB. Luego, con la clasificación obtenida para cada algoritmo de agrupamiento, se contrastó la clasificación obtenida con la categorización de la variable IMB para ver si existe o no independencia entre esas dos variables. Los dos grupos formados por el método UPGMA son dependientes de la categorización realizada para la variable IMB (que considera dos categorías bajo: <2.5% y alto: ≥2.5%); ya que el p-valor del estadístico Chi cuadrado de Pearson ($p=0,0267$) es estadísticamente significativo ($p<\alpha$) a un nivel de significancia $\alpha=0,05$ (Cuadro 3). Por lo tanto, se rechaza la hipótesis nula de independencia, concluyendo que la categoría IMB es dependiente de los grupos obtenidos por el método de conglomeración UPGMA. Este resultado indicaría que la cantidad de categorías para la variable IMB es dos. UPGMA fue más sensible que K-means a la hora de identificar la cantidad de grupos asociados a las variables de suelo en lotes de banano con alta y baja incidencia.

Cuadro 3. Tabla de contingencia de la categorización de la variable IMB vs la clasificación obtenida por el método UPGMA.

Estadístico	Valor	gl	p
Chi Cuadrado Pearson	4,91	1	0,0267*
Chi Cuadrado MV-G2	5,01	1	0,0252
Irwin-Fisher bilateral	-0,09		0,0842
Coef.Conting.Cramer	0,18		
Kappa(Cohen)	-0,05		
Coef.Conting.Pearson	0,24		
Coeficiente Phi	-0,25		

*significativo ($p<\alpha$) a un nivel de significancia $\alpha=0,05$.

Modelación estadística

Se observó que cuando no se utilizó algún tipo de remuestreo a los datos originales (Raw data set) se obtuvieron en general valores bajos para cada uno de los criterios usados para medir el desempeño de los algoritmos aplicados en este estudio. El uso de las técnicas de remuestreo permitió obtener mejores resultados que al considerar los datos originales. Los diferentes algoritmos y tipos de remuestreos utilizados en este estudio, se pueden ver en el cuadro 4. En general, se observó un comportamiento similar en cuanto a que el remuestreo Oversampling

fue el que presentó una mayor área bajo la curva ROC para todos los algoritmos evaluados. Además, se observó que el tipo de remuestreo ROSE mostró un menor desempeño para el conjunto de datos y para los algoritmos evaluados, destacándose por presentar en general valores bajos del área bajo la curva ROC tanto para el conjunto de validación como el de testeo, baja precisión, baja sensibilidad y baja especificidad (Cuadro 4). El algoritmo RF presentó en promedio valores altos para cada uno de los criterios usados AUC: área bajo la curva ROC, SN: sensibilidad, SP: especificidad y AC: precisión; siendo este, el algoritmo con el mejor desempeño respecto a los otros algoritmos usados en este trabajo. Por otra parte, los algoritmos LDA y SVMkl fueron los que presentaron en promedio valores bajos en la mayoría de los criterios (Cuadro 4).

Cuadro 4. Desempeño promedio de diferentes algoritmos, análisis discriminante lineal (LDA), máquinas de soporte vectorial con Kernel Lineal (SVMkl), Random Forest (RF) y máquinas de soporte vectorial con Kernel Radial (SVMkr) para las categorías de IMB.

Algoritmos	Método de remuestreo [‡]	AUC [¥]		SN	SP	AC
		Conjunto de Validación	Conjunto de Testeo			
LDA	Raw dataset	0.65	0.57	54.87	52.05	53.46
	Undersampling	0.75	0.74	65.60	57.59	61.60
	Oversampling	0.79	0.69	62.24	56.07	59.15
	ROSE	0.66	0.59	56.31	52.62	54.47
SVMkl	Raw dataset	0.63	0.56	53.85	51.72	52.79
	Undersampling	0.76	0.73	65.00	57.11	61.06
	Oversampling	0.76	0.75	66.07	57.53	61.80
	ROSE	0.66	0.60	56.86	52.93	54.89
RF	Raw dataset	0.69	0.63	58.10	54.15	56.13
	Undersampling	0.92	0.93	77.80	62.78	70.29
	Oversampling	0.93	0.92	76.97	63.03	70.00
	ROSE	0.65	0.60	56.96	52.6	54.78
SVMkr	Raw dataset	0.67	0.62	57.91	53.26	55.58
	Undersampling	0.83	0.82	70.59	59.99	65.29
	Oversampling	0.86	0.85	72.98	60.88	66.93
	ROSE	0.66	0.85	56.28	53.49	54.88

AUC: área bajo la curva ROC, SN: sensibilidad, SP: especificidad, AC: precisión.

Discusión

Los resultados de IMB coinciden con los reportaos por González (2003), que indica que la enfermedad se desarrolla en clima tropical y subtropical, con presencia de alta humedad y suelos con mal drenaje, con fuertes desequilibrios nutricionales. Cuando no hay aireación, la infección se produce en las raíces sanas por encontrar un exceso perjudicial de anhídrido

carbónico originado por la respiración, y aunque la raíz principal es poco afectada, las raicillas laterales enferman y quedan destruidas.

Por otra parte, Martínez et al. (2016) establecieron que las variables más relacionadas con la incidencia de la marchitez del banano en la zona de estudio son las relacionadas con la granulometría, conductividad eléctrica, carbono orgánico, nitrógeno total, contenidos de fósforo, calcio y magnesio intercambiables y la relación calcio – magnesio.

De acuerdo a la distribución espacial de los focos de incidencia de marchitez de banano en la zona de estudio, se deduce que el manejo inadecuado en los inicios de la enfermedad ha causado su mayor incidencia entre los lotes 36 al 38 de la unidad de producción bananera; sin embargo, Rey et al. (2016) indicaron que el área de estos lotes presenta suelos con características específicas de altos contenidos de arcilla, potasio, zinc, cobre y sales en profundidad y los más bajos contenidos de hierro y calcio de la finca que se repiten en otras áreas de la finca con banano y pasto, constituyéndose en zonas susceptibles a ser afectadas por la enfermedad, abarcando alrededor de 108 has.

Se pudo observar que, aunque la mayoría de los lotes (66.7%) presentaban valores de hasta 2.5% de incidencia, existía un 33.3 % de los lotes con valores superiores al 2.5% de incidencia de la enfermedad. Es importante resaltar que los lotes con altos niveles de incidencia (> 5%) de la enfermedad son considerados del alto riesgo para la zona. Este riesgo se viene dando no solo por comprometer la producción rentable de la finca, sino por el potencial de actuar como fuente de inóculo para otros lotes dentro de la plantación y para las plantaciones vecinas.

Sin embargo, sólo unos pocos países estudian y examinan la incidencia de la marchitez de bananos causada en esta ocasión por Foc, como es el caso de zonas bananeras de Costa Rica con 7.3% de incidencia (Lichtemberg et al. 2010); en Colombia se reportaron zonas con incidencia de 0,3% con una prevalencia de 4.3% (Merchán, 2002); la investigación de Hossain & Rashid (2001) arrojó una incidencia entre 4,27-24% en Bangladesh donde se consideró alta la incidencia de Foc; mientras que en fincas de pequeños productores (<2 ha) de Malasia la enfermedad llegó hasta el 50% (Jamaluddin et al. 2001). Entre tanto, investigaciones de Tushemereirwe & Ploetz (1993), en Uganda reportaron la Marchitez por Fusarium en bananos sobre los 1400 msnm, en estado continuo de cultivo mayores de 30 años, donde indica que la incidencia de la enfermedad es generalmente menos del 5%, coincidiendo con los reportes de la incidencia de esta investigación.

La incidencia de esta enfermedad es usualmente baja, aunque hasta el 60% de una plantación puede ser afectada. Ya que la marchitez puede ser confundida fácilmente con el marchitamiento por Fusarium, a menudo no es posible para el productor bananero distinguir entre las dos enfermedades. En las regiones afectadas por el marchitamiento por Fusarium, por lo tanto, la ocurrencia de la marchitez del banano puede ser subestimada.

En este estudio, se observó que el método de remuestreo ROSE presentó valores bajos de los criterios usados tales como: área bajo la curva ROC, sensibilidad, especificidad y precisión, esto puede deberse a que este tipo de remuestreo genera datos sintéticos que distan a la realidad de los datos tomados en campo (Lunardon et al. 2014). Los algoritmos como el Random Forest (RF) y las máquinas de soporte vectorial con kernel radial (SVMkr) se usan generalmente cuando las clases de una variable no pueden ser separables linealmente (Bauckhage, 2019). En nuestro trabajo, se observó que los algoritmos que presentaron un mejor desempeño fueron los que permitían una separación no lineal de las clases de IMB (RF y SVMkr). Esto indicaría que las clases de IMB deberían ser abordadas desde un enfoque no separable linealmente. Por otra parte, en

nuestro trabajo se observó que el desempeño más bajo lo presentaron los algoritmos LDA y SVMk; siendo estos algoritmos comúnmente usados cuando las clases son separables linealmente.

Estos resultados son similares a los reportados por Gomez Selvaraj et al. (2020) los cuales establecen que modelos de Machine Learning basados en imágenes aéreas tienen un gran potencial para proporcionar un sistema de apoyo a la toma de decisiones para las principales enfermedades del banano en África. También, el estudio desarrollado por Ye et al. (2020) concluyó que el clasificador de RF fue el más adecuado para la identificación y mapeo de la enfermedad del marchitamiento por *Fusarium* del banano a partir de imágenes de detección remota basadas en el uso de Vehículos Aéreos No Tripulados. La máquina de soporte vectorial (SVM), el Random Forest (RF) y el algoritmo basado en redes neuronales artificiales (ANN) han sido usados por Ye et al. (2020) y Aruraj et al. (2019) para identificar ubicaciones que estaban infestadas o no infestadas con marchitez por *Fusarium*.

Conclusión

Obtener información relevante a partir de los datos, no solo podría ayudar a mejorar la experiencia del agricultor, sino también a tomar decisiones pertinentes desde las primeras etapas de la aparición de los síntomas a campo. Asimismo, los datos de suelo pueden ser una importante fuente de información con respecto a la predisposición de algunas enfermedades en banano. En este sentido, los algoritmos aquí evaluados permitirían analizar, interpretar y predecir futuros escenarios que ayudan a prevenir y reducir la carga de enfermedades en lotes de banano de Venezuela y su impacto en la producción.

Bibliografía

Akter H., Hassan, Md Kamrul, Rabbani, Md, Al Mahmud, Abdullah. (2015). Effects of Variety and Postharvest Treatments on Shelf Life and Quality of Banana. *Journal of Environmental Science and Natural Resources*. 6. doi:10.3329/jesnr.v6i2.22113.

Aruraj, A. Alex, M. S. P. Subathra, N. J. Sairamya, S. T. George and S. E. V. Ewards. (2019). Detection and Classification of Diseases of Banana Plant Using Local Binary Pattern and Support Vector Machine, 2nd International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 2019, pp. 231-235, doi: 10.1109/ICSPC46172.2019.8976582.

Baker, F. B., and L. J. Hubert. (1975). Measuring the Power of Hierarchical Cluster Analysis. *Journal of the American Statistical Association* 70 (349): 31–38. doi:10.1080/01621459.1975.10480256.

Ball, G. H., and D. J. Hall. (1965). *ISODATA, a Novel Method of Data Analysis and Pattern Classification*. Menlo Park, CA: Stanford Research Inst.

Beale, E. M. L. (1969). *Cluster Analysis*. London: Scientific Control Systems.

Bosman, M. (2016). *Role of the environment on the incidence of Panama disease in bananas*. Wageningen University. Netherlands

Bauckhage, C. (2019). *Lecture Notes on Machine Learning: Binary Linear Classifiers*. B-IT,

University of Bonn.

Caliński, T., and J. Harabasz. (1974). A Dendrite Method for Cluster Analysis. *Communications in Statistics* 3 (1): 1–27.

Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set.” *Journal of Statistical Software* 61 (6): 1–36. doi:10.18637/jss.v061.i06.

Chemura A, Mutanga O, Dube T. (2017). Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precis Agric*, 18(5): 859–881.

Cortez, A., Olivares, B., Parra, R., Lobo, D., Rodríguez, M.F. y Rey, J.C. (2018). Descripción de los eventos de sequía meteorológica en localidades de la cordillera central, Venezuela. *Ciencia, Ingenierías y Aplicaciones*. 1 (1):22-44. <http://dx.doi.org/10.22206/cyap.2018.v1i1.pp23-45>

Dalton, L., V. Ballarin, and M. Brun. (2009). Clustering Algorithms: On Learning, Validation, Performance, and Applications to Genomics. *Current Genomics* 10 (6): 430–445. doi:10.2174/138920209789177601.

Davies, D. L., and D. W. Bouldin. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2: 224–227. doi:10.1109/TPAMI.1979.4766909.

Delgado E, Trejos J, Villalobos M, Martínez G, Lobo D, Carlos Rey J, Rodríguez G, Rosales FE, Pocasangre LE. (2010) Determination of a soil quality and health index for banana plantations in Venezuela. *Interciencia* 35:927-933

Deltour, P., Franca, S. C., Pereira, O. L., Cardoso, I., De Neve, S., Debode, J., and Hofte, M. (2017). Disease suppressiveness to Fusarium wilt of banana in an agroforestry system: Influence of soil characteristics and plant community. *Agriculture Ecosystems & Environment*, 239, 173-181. doi:10.1016/j.agee.2017.01.018

Dita, M., Barquero, M., Heck, D., Mizubuti, ESG. and Staver, CP. (2018). Fusarium Wilt of Banana: Current Knowledge on Epidemiology and Research Needs Toward Sustainable Disease Management. *Front. Plant Sci.* 9 (1468), 1-21. doi: 10.3389/fpls.2018.01468

Domínguez, J., Negrin, M. A., and Rodriguez, C. M. (2001). Aggregate water-stability, particle-size and soil solution properties in conducive and suppressive soils to Fusarium wilt of banana from Canary Islands (Spain). *Soil Biology & Biochemistry*, 33(4-5), 449-455. doi:10.1016/s0038-0717(00)00184-x

Duda, R. O., and P. E. Hart. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.

FAO (2019) *Banana Market Review and Banana Statistics 2018*. Rome. <http://www.fao.org/economic/est/est-commodities/bananas/en/> Accessed jan122020

Galindo Huerta, A. (2018). Algoritmos de clasificación para datasets desequilibrados: análisis y comparativa. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S.I. de Sistemas Informáticos (UPM), Madrid.

Gee, G.W and D. Or. (2002). Particle-size analysis. En: J.H. Dane and G.C. Topp (Ed.) Methods of soil analysis. Part 4. SSSA Book series Nº 5, SSSA, Madison, WI. p 255-293.

Gomez Selvaraj, M; A Vergara, F Montenegro, HA Ruiz, N Safari, D Raymaekers, W Ocimati, J Ntamwira, L Tits, A B Omondi, G Blomme. (2020). Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin, ISPRS Journal of Photogrammetry and Remote Sensing. 169: 110-124. <https://doi.org/10.1016/j.isprsjprs.2020.08.025>

González, S. (2003). Etiología y Epidemiología del “Falso Mal de Panamá” de La Platanera en Canarias. Tesis Doctoral. Instituto Canario de Investigaciones Agrarias. Tenerife, España. 286p.

Guo H, Li Y, Shang J, Mingyun G, Yuanyue H, Bing G. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications. 73: 220-239. <https://doi.org/10.1016/j.eswa.2016.12.035>.

Halkidi, M., M. Vazirgiannis, and Y. Batistakis. (2000). Quality Scheme Assessment in the Clustering Process”. In European Conference on Principles of Data Mining and Knowledge Discovery, Lyon, France. 265–276. Springer-Verlag Berlin Heidelberg.

Halkidi, M., Y. Batistakis, and M. Vazirgiannis. (2001). On Clustering Validation Techniques.” Journal of Intelligent Information Systems 17 (2): 107–145. doi:10.1023/A:1012801612483.

Hartigan, J. A. (1975). Clustering Algorithms. New York: John Wiley & Sons.

He, H. and E.A. Garcia. (2009). Learning from imbalanced data. Knowledge and Data Engineering, IEEE Transactions on, 21(9): p. 1263-1284

Hernández, R; Pereira, Y; Molina, JC; Coelho, R; Olivares, B and Rodríguez, K. (2017). Calendario de siembra para las zonas agrícolas del estado Carabobo en la República Bolivariana de Venezuela. Sevilla, Spain, Editorial Universidad Internacional de Andalucía. 247 p.

Hossain, MM; Rashid, MH. (2001). Status of fusarium wilt of banana in Bangladesh. In: A.B. Molina, A.B.; Nik Masdek, N.H.; Liew, K.W. (Eds) Banana Fusarium wilt management: towards sustainable cultivation. 306 p. ISBN: 971-91751-14-1.

Hubert, L. J., and J. R. Levin. (1976). A General Statistical Framework for Assessing Categorical Clustering in Free Recall.” Psychological Bulletin 83 (6): 1072–1080. doi:10.1037/0033-2909.83.6.1072.

Jamaluddin, SH; Nik Madesk, NH; Ibrahim, O. (2001). Preliminary survey of banana fusarium wilt disease in Malaysia, implications on production and the industry. In: A.B. Molina, A.B.; Nik Masdek, N.H.; Liew, K.W. (Eds) Banana Fusarium wilt management: towards sustainable cultivation. 306 p. ISBN: 971-91751-14-1.

Karangwa, P., Mostert, D., Ndayihanzamaso, P., Dubois, T., Niere, B., zum Felde, A., . . . Viljoen, A. (2018). Genetic Diversity of Fusarium oxysporum f. sp cubense in East and Central Africa. Plant Disease, 102(3), 552-560. doi:10.1094/pdis-02-17-0282-re

Kassambara, A. (2017). Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. STHDA. www.sthda.com

Kautzky, A., P. Baldinger, D. Souery, S. Montgomery, J. Mendlewicz, J. Zohar, A. Serretti, R. Lanzenberger, and S. Kasper. (2015). The Combined Effect of Genetic Polymorphisms and Clinical Parameters on Treatment Outcome in Treatment-resistant Depression. *European Neuropsychopharmacology* 25 (4): 441–453. doi:10.1016/j.euroneuro.2015.01.001.

Lavah, E. & Israeli, Y. (2000). Mineral deficiencies of Banana. In *Diseases of banana*, Abacá and enset. (Ed. D Jones). CABI Publishing: Wallingford, Oxon, UK. pp. 339-350.

Li, Z., Deng, Z., Chen, S., Yang, H., Zheng, Y., Dai, L., Zhang, F., Wang, S. & Hu, S. (2018). Contrasting physical and biochemical properties of orchard soils suppressive and conducive to Fusarium wilt of banana. *Soil Use and Management*, 34(1), 154-162. doi:10.1111/sum.12390

Lichtemberg, PS; Pocasangre, LE; Sikora, RA; Staver, C. (2010). Current status of Fusarium Wilt (*Fusarium oxysporum* f. sp. *cubense*) in bananas Gros Michel (AAA) at smallholder level of Costa Rica. XIX Reunión Internacional ACORBAT 2010. Medellín. COL. 32-44p.

Lozano P, Z, Bravo C, Ovalles F, Hernández RM, Moreno B, Piñango, L, Villanueva, JG. (2004). Selección de un diseño de muestreo en parcelas experimentales a partir del estudio de la variabilidad espacial de los suelos. *Bioagro*, 16(1),61-72. <https://n9.cl/nhvf>

Longadge, R., & Dongre, S. (2013). Class Imbalance Problem in Data Mining Review. ArXiv, abs/1305.1707.

Lunardon, N; Menardi, G; Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. *R Journal*, 6(1), 82-92.

Ma L, Fu T, Blaschke T, Li M, Tiede D, Zhou Z, et al. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *Isprs Int J Geo-Inf*, 6(2). doi: 10.3390/ijgi6020051

Martínez, G.; J. C. Rey; L. Castro; E. Micale; O. Lopez; R. Pargas; y E. Manzanilla. (2016). Marchitez en banano Cavendish en la Región Central de Venezuela, asociado a un complejo hongo – bacteria. Reunión Acorbat, 2016. Miami USA. Memorias.

McClain, J. O., and V. R. Rao. (1975). CLUSTISZ: A Program to Test for the Quality of Clustering of a Set of Objects." *Journal of Marketing Research* 12 (4): 456–460.

Mehlich A. (1984). Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Comm Soil Sci Plant Anal*15:1409-1416

Merchán, VM. (2002). Manejo Integrado de plagas del plátano y banano. In: Acorbat. Memorias XV reunión. Asociación de Bananeros de Colombia AUGURA. Medellín COL. 354 -361p.

Milligan, G. W. (1981). A Monte Carlo Study of Thirty Internal Criterion Measures for Cluster Analysis. *Psychometrika* 46 (2): 187–199. doi:10.1007/BF02293899.

Menardi, G. and Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28(1):92–122,

Olivares BO. (2018). Tropical rainfall conditions in rainfed agriculture in Carabobo, Venezuela. *Lgr Lif Sci* J27:86-102. <https://doi.org/10.17163/lgr.n27.2018.07>

Pérez-Vicente, L. and Porras, Á. (2015). Impacto potencial del cambio climático sobre las plagas de bananos y plátanos en Cuba. *Fitosanidad*, 19 (3), 201-211.

Ploetz, R. C. (2015). Management of Fusarium wilt of banana: A review with special reference to tropical race 4. *Crop Protection*, 73, 7-15. doi:10.1016/j.cropro.2015.01.007

Ratkowsky, D. A., and G. N. Lance. (1978). Criterion for Determining the Number of Groups in a Classification. *Australian Computer Journal* 10 (3): 115–117.

Rey JC, Chacín M, Sapuky M, Núñez M, Martínez G, Rodríguez G, Espinoza J, Arturo M, Pocasangre L, Delgado E, Rosales F. (2006). Aptitud de las tierras para banano en suelos de Venezuela y su relación con la productividad. XVII Reunión Internacional ACORBAT: Banano un negocio sustentable. Joinville. Santa Catarina, Brasil. Nov 15-20. p. 362.

Rey JC, Martínez G, Lobo D, Trejos J, Pocasangre L, Rosales F. (2009). Aspectos sobre calidad y salud de suelos bananeros en Venezuela. *Producción Agropecuaria* 2: 52-55

Rey, J. C.; G. Martínez; N. Pizzo; E. Micale; and N. Fernandez. (2016). Áreas susceptibles a la enfermedad Falso Mal de Panamá en Banano Cavendish, en la Región Central de Venezuela. Reunión Acorbat, 2016. Miami USA. Memorias.

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20: 53–65. doi:10.1016/0377-0427(87)90125-7.

Segura, R. A., Serrano, E., Pocasangre, L., Acuna, O., Bertsch, F., Stoorvogel, J. J., & Sandoval, J. A. (2015a). Chemical and microbiological interactions between soils and roots in commercial banana plantations (Musa AAA, cv. Cavendish). *Scientia Horticulturae*, 197, 66-71. doi:10.1016/j.scienta.2015.10.028

Sekula, M., S. Datta, and S. Datta. (2017). OptCluster: An R Package for Determining the Optimal Clustering Algorithm. *Bioinformatics* 13 (3): 101–103. doi:10.6026/bioinformatics.

Soil Survey Staff. (2014). Kellogg Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report No. 42, Version 5.0. R. Burt and Soil Survey Staff (ed.). U.S. Department of Agriculture, Natural Resources Conservation Service.

Tibshirani, R., G. Walther, and T. Hastie. (2001). Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society: Series B* 63 (2): 411–423. doi:10.1111/1467-9868.00293.

Tushemereirwe, WK; Ploetz, RC. (1993). First Report of Fusarium Wilt on East African Highland Cultivars of Banana. *Plant Disease* 77 (10): 1063 -1064

Viloria JA, Pineda C, López D, Gabaldón O. (2003). Aptitud de la tierra para caña de azúcar y banano en el sistema de riego Taguaguay (Aragua-Venezuela). *Agron Trop* 53:189-208

Ye H C, Huang W J, Huang S Y, Cui B, Dong Y Y, Guo A T, et al. (2020). Identification of banana fusarium wilt using supervised classification algorithms with UAV-based multi-spectral imagery. *Int J Agric & Biol Eng*, 13(3): 136–142. DOI: 10.25165/j.ijabe.20201303.5524