

Identificación automática del nivel de daño de la cenicilla polvorienta a través de firmas espectrales en plantas de Cucurbita pepo L.

C. A. Rivera-Romero* E. R. Palacios-Hernández**
M. Trejo-Durán*** O. Vite-Chávez****
J. A. Morales-Saldaña†

* Centro de Investigación de Estudios de Posgrado (CIEP), Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México (e-mail: c.a.riveraromero@alumnos.uaslp.edu.mx).

** Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México (e-mail: epalacios@fciencias.uaslp.mx).

*** Departamento de Estudios Multidisciplinarios, División de Ingenierías Campus Irapuato-Salamanca, Universidad de Guanajuato, Guanajuato, México (e-mail: mtrejo@ugto.mx).

**** Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Jalpa, Zacatecas, México (e-mail: osvichz@uaz.edu.mx).

† Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México (e-mail: jmorales@uaslp.mx).

Resumen: La cenicilla polvorienta es un problema común en los cultivos de cucurbitáceas. Esta enfermedad minimiza el periodo de cosecha, el rendimiento de las plantas y reduce la fotosíntesis causando senescencia en las hojas. Se identificaron 4 niveles de daño (T1-hoja sana, T2-hoja en tiempo de germinación, T3-hoja con primeros síntomas y T4-hoja enferma) de la cenicilla polvorienta en hojas de Cucurbita pepo L (C. pepo L.). Se propone un algoritmo compuesto por 5 etapas para la identificación automática. A través de una base de datos compuesta de firmas espectrales e imágenes tomadas de hojas de C. pepo L., se obtuvieron datos para determinar los niveles de daño identificados. Para este trabajo, se consideraron las firmas espectrales para ser caracterizadas por medio del cálculo de 64 índices de vegetación. Se utilizaron herramientas estadísticas para descartar algunas de las características. Se trabajó con una etapa de clasificación binaria en base a los algoritmos del k-vecino más cercano (KNN) y las máquinas de soporte vectorial (SVM) para resolver casos en que las características resultantes fueran linealmente separables o no linealmente separables. En la etapa final, se planteó un problema de multclasificación y su evaluación de desempeño. Como resultado, se obtuvo la identificación automática de los datos de entrada a un vector resultante clasificado indicando el nivel de daño de la cenicilla polvorienta en las hojas de C. pepo L.

Palabras clave: Firma espectral, espectroscopia, índices de vegetación, Cucurbita pepo L., análisis de varianza, multclasificación.

1. INTRODUCCIÓN

La cenicilla polvorienta es una enfermedad común que se presenta en la mayoría de los cultivos de las cucurbitáceas. En el caso particular, las plantas de Cucurbita pepo L. (C. pepo L.) son muy susceptibles a los hongos *Podosphaera xanthii* y *Erysiphe cichoracearum*. El crecimiento fúngico blanco y polvoriento se desarrolla en ambas superficies de las hojas, pecíolos y tallos. Por lo general, las esporas aparecen primero en las hojas de la corona (hojas basales),

después, en las hojas inferiores sombreadas y en la superficie inferior de las plantas causando marchitez. En una planta enferma de cenicilla polvorienta, se minimiza el periodo de la cosecha y se reduce la fotosíntesis causando senescencia prematura en las hojas y en el fruto. Adicionalmente, en un cultivo, el daño de la cenicilla polvorienta hace a las demás plantas débiles a otras enfermedades (Barickman et al. (2017)). Después del día 16 del despliegue de la hoja, se pueden comenzar a observar los primeros síntomas. A través de la inspección

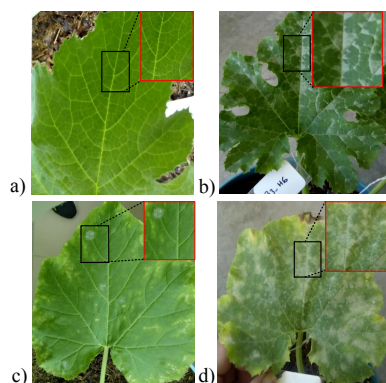


Fig. 1. Características visibles observadas en hojas de *C. pepo* L. con la enfermedad de la cenicilla polvorienta: a) Hoja sana, b) Hoja en germinación, c) Hoja con primeros síntomas y d) Hoja enferma.

visual, de los estadios fenológicos de crecimiento de la planta (Meier (2001)) y de la ocurrencia de los síntomas (Muhammed (2005)), se identificaron algunas características comunes (Figura 1 y Figura 2) en las hojas de *C. pepo* L.: hojas sanas, hojas en tiempo de germinación, hojas con primeros síntomas y hojas enfermas por el hongo.

En la actualidad, los agricultores realizan prácticas como la aplicación de sustancias en los cultivos, las cuales pueden poner en riesgo la producción del campo a un elevado costo. Por otro lado, existen propuestas para atacar las enfermedades en los cultivos, pero consisten en pruebas destructivas de laboratorios a fin de identificar el daño. Por lo tanto, es importante contar con herramientas de detección automática considerando los diferentes niveles de daño de cualquier enfermedad. Bajwa et al. (2017), propuso el monitoreo de hojas de soya a través de la espectroscopia de reflectancia para identificar las longitudes de onda que mejor describen a los nemátodos. Cao et al. (2013) detectó la cenicilla polvorienta en cultivos de trigo a través de la reflectancia hiperespectral. Herrmann et al. (2017) realizó una evaluación con espectroscopia para identificar los niveles de daño por la araña roja en hojas de plantas de pimientos y frijol. Rumpf et al. (2010) implementó la identificación temprana y la clasificación de las plantas enfermas de cenicilla polvorienta, manchas y moho en hojas de betabel con máquinas de soporte vectorial en base a la reflectancia hiperespectral. Todos estos métodos obtienen los datos a través de las hojas de las plantas. En el caso de las cucurbitáceas, no existe propuesta para identificar los niveles de daño de la cenicilla polvorienta. Como anteriormente se han identificado algunas características particulares de las hojas enfermas de *C. pepo* L. se propone un algoritmo para la detección automática de los niveles de daño de la cenicilla polvorienta a través de firmas espectrales.

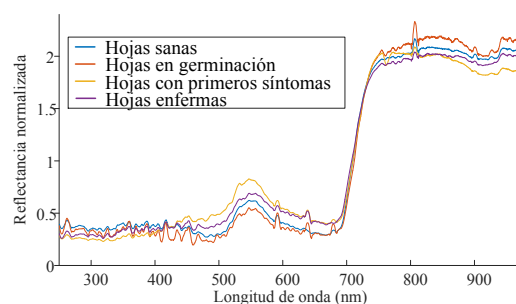


Fig. 2. Firmas espectrales promedio de las características observadas en las hojas de *C. pepo* L.

2. METODOLOGÍA

2.1 Protocolo

La base de datos contiene el registro de las firmas espectrales e imágenes respectivas de dónde se hizo la medición. Se realizó un análisis de los niveles de daño que puede presentar la cenicilla polvorienta a través de la inspección visual de las hojas, de los estadios fenológicos de crecimiento de la planta y de la ocurrencia de los síntomas en las hojas. Se identificaron 4 niveles de daño: T1-hojas sanas, T2-hojas en germinación, T3-hojas con primeros síntomas y T4-hojas enfermas por el hongo. Estos niveles fueron revisados por agricultores que cuentan con la experiencia en cultivos de cucurbitáceas. En la Tabla 1, se presenta a detalle los niveles de daño identificados en las hojas.

2.2 Definición de problema

Se considera que la planta muestra diferentes características en las hojas cuando un agente fúngico se presenta. Cuando las hojas son jóvenes, los primeros síntomas aún no son visibles. La cenicilla polvorienta aparece generalmente en las etapas de floración y del desarrollo del fruto. Una vez que la espora se establece en la hoja, existe un periodo de 3 a 7 días de tiempo de germinación. Después de estos días se hacen visibles los primeros síntomas en pequeñas manchas blancas circulares con esporas. A partir de este momento la enfermedad comienza a visualizarse en la mayoría de la planta, cubriendo tallos, flores y fruto. Por lo tanto, el problema a resolver parte de diseñar un algoritmo para detectar automáticamente en que nivel de daño se encuentra la planta para poder establecer un tratamiento.

2.3 Algoritmo propuesto

El algoritmo propuesto se compone de 5 etapas de desarrollo (Figura 3). En la primera etapa, se considera para la generación de la base de datos de firmas espectrales e imágenes de las hojas de *C. pepo* L. La segunda etapa consiste en el pre-tratamiento y el análisis exploratorio de las firmas espectrales para determinar el

Tabla 1. Nivel de daño de la cenicilla polvorienta en hojas de cucurbitáceas.

Descripción	Nivel de daño	Etiqueta	Días de muestra	Estadios fenológicos (Meier (2001))
Hojas sanas	0	T1	19	Desarrollo de las hojas y floración
Hojas en tiempo de germinación	1	T2	5	Floración
Hojas con primeros síntomas	2	T3	8	Floración y desarrollo del fruto
Hojas enfermas	3	T4	6	Madurez del fruto y semilla

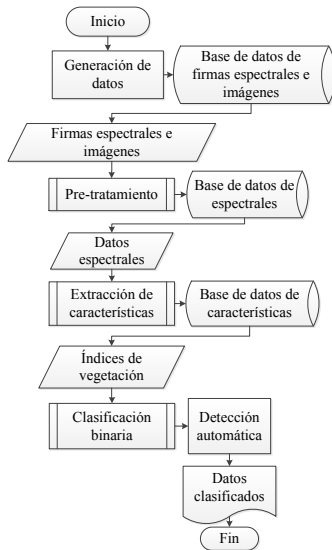


Fig. 3. Diagrama general del algoritmo propuesto para la identificación automática del nivel de daño de la cenicilla polvorienta en hojas de *C. pepo* L.

nivel de daño. La tercer etapa se encarga de la extracción de características, las cuales fueron definidas por el cálculo de 64 índices de vegetación. La cuarta etapa consistió en aplicar clasificadores binarios para identificar los datos de cada nivel de daño. La quinta etapa es la de la identificación automática en donde se muestra a que nivel de daño pertenecen los datos de entrada.

2.4 Etapa de la generación de la base de datos

Se obtuvieron de las firmas espectrales e imágenes de hojas de un cultivo de 12 plantas de *C. pepo* L. A este cultivo se le aplicó un tratamiento básico preventivo y de riego. Las plantas se sembraron en el invernadero del Departamento de Estudios Multidisciplinarios (DEM) de la División de Ingenierías Campus Irapuato-Salamanca de la Universidad de Guanajuato. Se trabajó con un espectrómetro UV-VIS-NIR HR4000CG-UV-NIR Ocean Optics. Las firmas espectrales son muy variables, por lo tanto no es posible definir una sola firma para un nivel de daño específico. Debido a esto, se tomaron una serie de firmas espectrales e imágenes de las hojas de *C. pepo* L. continuamente durante el crecimiento de las plantas sin cortar o destruir las hojas. Las firmas espectrales se tomaron con un rango de longitud de onda de 195 nm a 1122 nm en el espectro UV-VIS-NIR y en varios puntos de la hoja (lóbulos), con una herramienta tipo sargento mecánico diseñada especialmente para obtener los datos

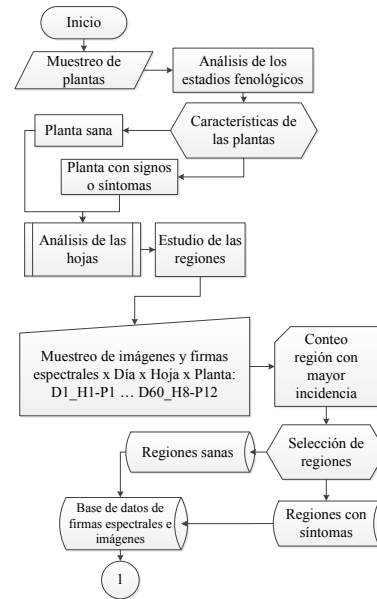


Fig. 4. Diagrama de la etapa de la generación de la base de datos.

de reflectancia con un ángulo de luz incidente de 45°. Las imágenes se tomaron del mismo punto que las firmas con una estructura plástica a una distancia de 8 cm de distancia entre la hoja y el lente de una cámara digital resultando imágenes en espacio de color RGB con una dimensión de 4160x3120 a 13 megapíxeles. El periodo de muestreo fue del 23 de Septiembre al 23 de Noviembre del 2016, durante los días Lunes, Miércoles y Viernes. La base de datos se conformó por muestras de 12 hojas por cada planta, por 19 días de muestreo por hoja, por 2 regiones de mayor incidencia a síntomas, resultando en un total de 5,472 firmas espectrales UV-VIS-NIR e imágenes. Esta etapa y los elementos que la forman se describen en la Figura 4.

2.5 Etapa de la extracción de características

2.6 Etapa de pre-tratamiento y análisis exploratorio

Al tener la base de datos de firmas espectrales, se procedió a aplicar un filtro de suavizado Savitzky-Golay basado en el cálculo de una regresión polinomial con al menos $k + 1$ puntos espaciados para determinar el nuevo valor de cada punto. Se consideró un polinomio de grado 3 con un espaciado de 10 puntos. Este filtro permitió desaparecer ruidos ajenos al sistema provocados por el espectrómetro. Después de obtener las firmas filtradas se normalizaron los datos de acuerdo a la media. A

Tabla 2. Ecuaciones de índices de vegetación (Green Normalized Difference Vegetation (GNDVI), Near-Infrared Red-Edge NDVI (NRENDVI), Triangular greenness index (TGI), Simple Difference (SD4), Anthocyanin Reflectance Index (ARI)) que miden variables internas de la hoja. Los límites del centro y rango de la banda espectral ρ : azul (Blue(B)) 490nm/470 510nm, verde (Green(G)) 56nm/540 580nm, rojo (Red(R)) 66nm/656 67nm, borde-rojo (Red-edge(RE)) 715nm/710 720nm, cercano infrarrojo (Near infrared(NIR)) 790nm/770 810nm para la relación del valor de la reflectancia relativa Herrmann et al. (2017).

Índices de vegetación	Ecuación
Vegetación de Diferencia Normalizada Verde	$GNDVI = \frac{\rho_{NIR} - \rho_G}{\rho_{NIR} + \rho_G}$
Gitelson	$GIT1 = \frac{1}{R_{700}}$
NDVI de rojo al infrarrojo cercano	$NRENDVI = \frac{\rho_{NIR} - \rho_{RE}}{\rho_{NIR} + \rho_{RE}}$
Verdor triangular	$TGI = 0.5((R_B)(\rho_R - \rho_G) + (R_G)(\rho_R - \rho_B))$
Diferencia simple	$SD4 = \frac{1}{R_{550}} - \frac{1}{R_{750}}$
Reflectancia de antocianinas	$ARI = \frac{1}{R_{550}} - \frac{1}{R_{700}}$

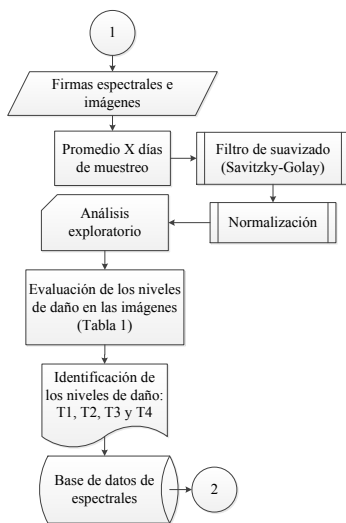


Fig. 5. Diagrama de la etapa del pre-tratamiento y análisis exploratorio.

continuación, se tomaron las imágenes relacionadas con cada firma espectral y se agruparon por días de muestra conforme las características de la enfermedad. Se realizó un análisis exploratorio de los datos para encontrar características por rangos de longitudes de onda para evaluar el nivel de daño con las firmas. En esta etapa se identificaron los diferentes niveles de daño: T1, T2, T3, T4 entre las firmas espectrales promedio (definidos en la Tabla 1) y como resultado se obtuvo la base de datos espectrales. La Figura 5 describe los elementos de la estructura interna de esta etapa.

En esta etapa, se plantea la extracción de características que consistió en el cálculo de 64 índices de vegetación (VI). Estos índices se relacionan con diferentes variables internas de una hoja como la biomasa, el área foliar, contenido de clorofila, nitrógeno, antocianinas, carotenos y otros compuestos. Los resultados de los VI fueron normalizados para facilitar su análisis. En la Tabla 2 se presentan algunos índices de vegetación y sus especificaciones. Para verificar que las características

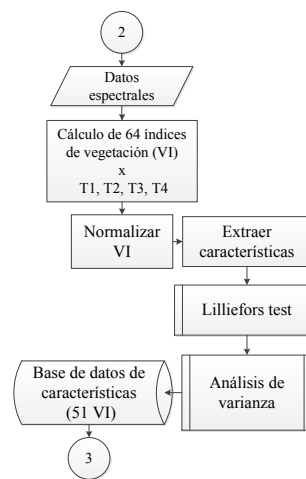


Fig. 6. Diagrama de la etapa para la extracción de características.

describieran los 4 niveles de daño de forma separada (T1, T2, T3, y T4) se realizó una prueba de normalidad Lilliefors de los datos para someterlos después en un análisis de varianza (ANOVA) y reducir el número de características. El ANOVA permitió descartar algunos índices de vegetación en los cuales no existe diferencia entre niveles de daño, lo cual resultaron 51 características útiles. En la Tabla 3 se muestran resultados de las características que tiene diferencia significativa entre niveles de daños. En la Figura 6 se presentan los elementos de esta etapa.

2.7 Etapa de clasificación binaria

En esta etapa se propuso utilizar clasificadores binarios para la detección automática. Se construyeron gráficas de dispersión (mapas de características) para visualizar las características linealmente separables (LS) y las que no resultarían ser linealmente separables (NLS). Primero, se consideró un clasificador lineal: el algoritmo del *k-vecino más cercano* (KNN, k-nearest neighbors) para someter características LS. Para el caso de las características NLS, se consideró un segundo clasificador: el algoritmo

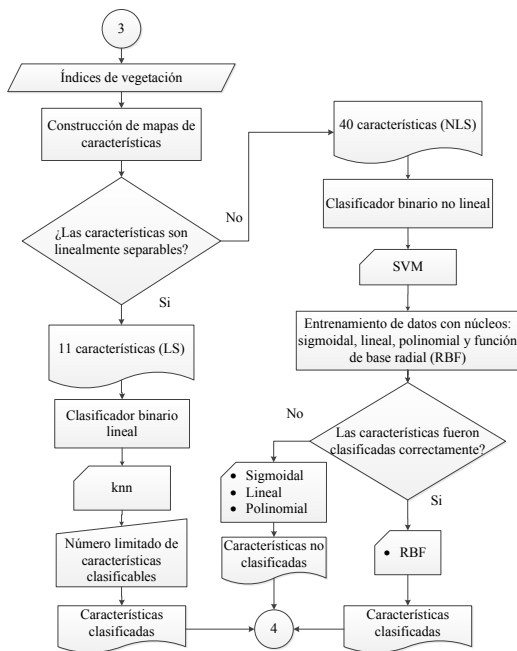


Fig. 7. Diagrama de la etapa de la clasificación binaria.

de las *máquinas de soporte vectorial* (SVM, support vector machine). KNN logró clasificar un número limitado de características y diferenciar entre las clases T1, T3 y T4. SVM separó todas las clases de T1 hasta T4 utilizando como núcleo la función de base radial. Al final se obtuvieron los clasificadores binarios entrenados con los datos de diferente nivel de daño. En la Figura 7 se muestran los elementos y las funciones que se utilizaron para esta etapa.

2.8 Etapa de la identificación automática

Una vez que se determinaron los clasificadores binarios, se optó por construir la etapa de la identificación automática con una multclasificación. Se utilizó el método uno contra uno (OVO, one vs one, T1 vs T2, T1 vs T3, T1 vs T4, T2 vs T3, T2 vs T4, y T3 vs T4) que es un esquema de votación para identificar sobre un conjunto de datos de prueba (formado por vectores de características) a que nivel de daño pertenece cada dato. Para establecer cada nivel se usó un subsistema de votación. Se evaluó el desempeño del sistema utilizando matrices de confusión y coeficientes *kappa de Cohen*. En la Figura 8 se presentan los elementos que describen la etapa de la identificación automática del nivel de daño de la cenicienta polvorienta en las hojas de *C. pepo L.*

3. RESULTADOS

La base de datos de firmas espectrales y de imágenes de las hojas de plantas de *C. pepo L.* quedó integrada en la primera etapa. En la segunda etapa se definió la base de datos espectrales a partir de filtrado, normalización y

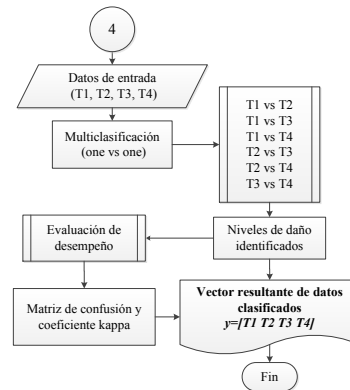


Fig. 8. Diagrama de la etapa de la identificación automática

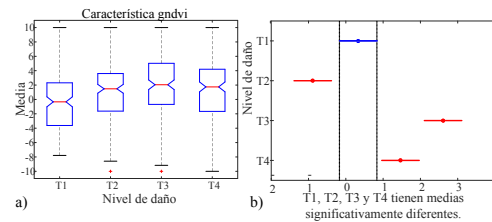


Fig. 9. ANOVA de un sólo factor a) ANOVA de la característica *gndvi* con medias significativamente diferentes, b) comparación múltiple del nivel de daño con la característica *gndvi*.

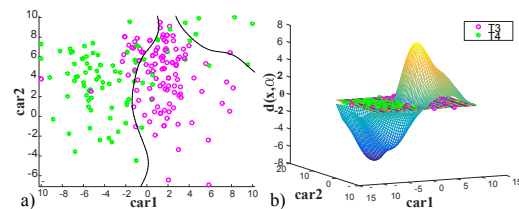


Fig. 10. Gráficas de los hiperplanos óptimos con SVM a) hiperplano óptimo en 2D con características NLS con núcleo de base radial, b) hiperplano óptimo en 3D.

el análisis exploratorio. En la tercer etapa se obtuvieron las características a través del cálculo de los índices de vegetación y se realizó una reducción del total. Se observó que la mayoría de las características de dos o mas conjuntos de datos tuvieron medias significativamente diferentes entre sí y se pueden comparar entre cada posible par de niveles de daño (T1 T2, T1 T3, T1 T4, T2 T3, T2 T4, T3 T4). La Figura 9a y la Figura 9b muestra el resultado con el ANOVA en datos con medias significativamente diferentes y separables entre todas las clases para la característica *gndvi*.

En la etapa de los clasificadores binarios, se obtuvieron las gráficas de los hiperplanos con las SVM y con algunas de las características NLS con diferencias significativas. En la Figura 10 se observa un hiperplano óptimo para los niveles T2 y T4 de estas características.

Tabla 3. Capacidad de algunos índices de vegetación para ser clasificados según el nivel de daño como resultado del ANOVA de un sólo factor.

Referencia	VI	T1	T2	T3	T4	F/significancia
Bajwa et al. (2017); Herrmann et al. (2017)	*gndvi	a	b	c	d	29.37 ($p < 0.0001$)
Cammarano et al. (2014)	git1	a	b	a, c	a, d	24.11 ($p < 0.0001$)
Herrmann et al. (2017)	nrendvi	a	b	c	b, c, d	12.6 ($p < 0.0001$)
Herrmann et al. (2017)	*tgi	a	b	c	d	930.4 ($p < 0.0001$)
Lu et al. (2015)	*sd4	a	b	c	d	549.02 ($p < 0.0001$)
Rumpf et al. (2010)	*ari	a	b	c	d	658.33 ($p < 0.0001$)

*VI discrimina significativamente el nivel de daño.

Los resultados de la etapa de identificación automática, se pueden visualizar en la Figura 11. La detección automática funciona con características LS y NLS. Estos son los resultados finales del algoritmo propuesto en donde se puede observar que el dato de entrada al sistema termina en un vector de salida que indica el nivel de daño de la hoja.

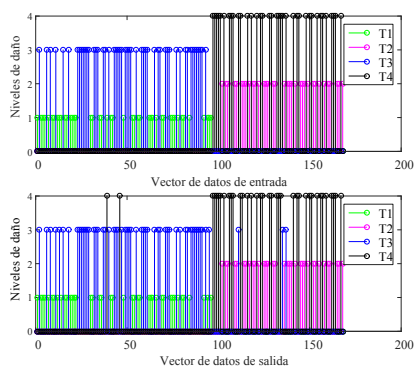


Fig. 11. Resultado de la identificación automática con características LS y NLS.

Estos resultados fueron evaluados con la exactitud de la matriz de confusión y valores altos del coeficiente $kappa$ (exactitud de entre 0.9776 y 0.9866, $kappa$ de entre 0.9360 y 0.9601) que indican un número de datos clasificados correctamente.

4. CONCLUSIONES

La identificación automática es útil para el manejo integrado de enfermedades en cultivos y para mejorar el rendimiento de la productividad. Este algoritmo identifica el nivel del daño en las hojas de las plantas de *C. pepo* L. y es significativo porque en algunos niveles de la enfermedad, el daño no es visible como en el caso de T2. Este sistema da solución a problemas de detección temprana de enfermedades de hongos, debido a que se definen los niveles de daño comenzando por un nivel 0 o de planta sana. Los resultados de T1, se consideran para tratamientos preventivos. En el caso de la cenicilla polvorienta, se asigna un pre-diagnóstico en la identificación de T2, debido a que es el tiempo de germinación del hongo y es una etapa previa a los primeros síntomas de la aparición de esporas blancas. Por lo tanto, al obtener datos de salida del sistema que

se ubiquen en T2, la planta se encuentra en tiempo oportuno de tratamientos de control. Los niveles T3 y T4, son niveles en los que la planta se encuentra enferma y se deben buscar otras alternativas apropiadas. Entre las ventajas que ofrece este algoritmo es que las hojas de las plantas no son cortadas, sólo monitoreadas continuamente. Además se considera la implementación en cualquier plataforma y para plantas con hojas similares a la planta de *C. pepo* L u otras cucurbitáceas.

REFERENCIAS

- Bajwa, S., Rupe, J., and Mason, J. (2017). Soybean disease monitoring with leaf reflectance. *Remote Sensing*, 9, 127.
- Barickman, T.C., Horgan, T.E., and Wilson, J.C. (2017). Efficacy of fungicide applications and powdery mildew resistance in three pumpkin cultivars. *Crop Protection*, 101(Supplement C), 90 – 94.
- Cammarano, D., Fitzgerald, G.J., Casa, R., and Basso, B. (2014). Assessing the robustness of vegetation indices to estimate wheat n in mediterranean environments. *Remote Sensing*, 6(4), 2827–2844.
- Cao, X., Luo, Y., Zhou, Y., Duan, X., and Cheng, D. (2013). Detection of powdery mildew in two winter wheat cultivars using canopy hyperspectral reflectance. *Crop Protection*, 45, 124 – 131.
- Herrmann, I., Berenstein, M., Paz-Kagan, T., Sade, A., and Karnieli, A. (2017). Spectral assessment of two-spotted spider mite damage levels in the leaves of greenhouse-grown pepper and bean. *Biosystems Engineering*, 157, 72 – 85.
- Lu, S., Lu, X., Zhao, W., Liu, Y., Wang, Z., and Omasa, K. (2015). Comparing vegetation indices for remote chlorophyll measurement of white poplar and chinese elm leaves with different adaxial and abaxial surfaces. *Journal of Experimental Botany*, 66(18), 5625–5637.
- Meier, U. (2001). Growth stages of mono- and dicotyledonous plants, bbch monograph.
- Muhammed, H.H. (2005). Hyperspectral crop reflectance data for characterising and estimating fungal disease severity in wheat. *Biosystems Engineering*, 91(1), 9 – 20.
- Rumpf, T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., and Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74(1), 91 – 99.