VOLUMEN 37
XXX Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Agricultura de precisión vía espectroscopía e inteligencia artificial

Rodrigo Fabian Cervantes Martinez[1], Neil Otniel Moreno Rivera[2], Diana Paulina Moreno Miranda[3], Oleksiy Shulika[*4]

[1] Licenciatura en Datos e Inteligencia Artificial, División de Ingenierías del Campus Irapuato-Salamanca (DICIS), Universidad de Guanajuato.
[2] Licenciatura en Sistemas Computacionales, División de Ingenierías del Campus Irapuato-Salamanca (DICIS), Universidad de Guanajuato.
[3] Maestría en Ingeniería Eléctrica, División de Ingenierías del Campus Irapuato-Salamanca (DICIS), Universidad de Guanajuato.
[4] Departamento de Ingeniería Eléctrica, División de Ingenierías del Campus Irapuato-Salamanca (DICIS), Universidad de Guanajuato.

Resumen

En este trabajo se presenta el desarrollo de un sistema de clasificación de hojas de fresa utilizando mediciones espectrales y técnicas de inteligencia artificial. Se recolectaron hojas en tres estados fisiológicos (sanas, enfermas y deshidratadas) y se midió su espectro de reflectancia en el rango 400 - 750 nm. Los datos fueron preprocesados mediante filtros Savitzky - Golay y analizados con índices vegetativos (VARI, NDVI y Red Edge Index) y reducción de dimensionalidad (PCA). Finalmente, se entrenaron clasificadores LDA, SVM y KNN para automatizar la detección del estado de salud de la hoja. Los resultados muestran una alta precisión (alrededor de 90%) en la clasificación, lo cual refleja el potencial de estas herramientas para la agricultura de precisión. El sistema propuesto puede implementarse en campo para el monitoreo y detección temprana de enfermedades y estrés hídrico.

Palabras clave: espectroscopía; inteligencia artificial; clasificación; índices vegetativos; agricultura de precisión.

Introducción

La agricultura, como actividad esencial para garantizar la seguridad alimentaria, enfrenta actualmente la necesidad de incorporar herramientas tecnológicas que permitan una mayor eficiencia en el diagnóstico y resolución de problemas. En un mundo donde la demanda de alimentos sigue en aumento, la automatización y el análisis inteligente de datos representan una oportunidad crucial para optimizar los cultivos y asegurar su máximo rendimiento [1].

En este trabajo presentamos una introducción a cómo la inteligencia artificial puede contribuir a resolver una de las problemáticas más comunes en el sector agrícola: la detección oportuna de enfermedades en plantas productoras de frutos. Como caso de estudio, se analizaron hojas de fresa, clasificadas en tres estados fisiológicos: sana, enferma y totalmente deshidratada, que fueron previamente recolectadas y medidas en condiciones de laboratorio controladas.

Para diferenciar estos estados, se emplearon índices de vegetación, herramientas matemáticas derivadas del análisis espectral que permiten evaluar el estado fisiológico de una planta de forma no invasiva [2]. En este estudio se adecuaron dichos índices al rango espectral disponible (400–750 nm), el cual abarca la región visible y parte del infrarrojo cercano, incluyendo el llamado borde rojo, sensible a cambios en el contenido de clorofila y al estrés hídrico [3].

Entre los índices evaluados se encuentra el VARI (Visible Atmospherically Resistant Index), útil para detectar cambios en la coloración de la hoja asociados a procesos de deshidratación o enfermedad [4]. Los espectros de reflectancia fueron previamente filtrados mediante el algoritmo Savitzky-Golay, una técnica de suavizado digital ampliamente utilizada para preservar las características espectrales mientras se reduce el ruido [5].

[*] Toda correspondencia debe ser enviada a: oshulika@ugto.mx

Para el análisis de los datos, se aplicaron métodos de reducción de dimensionalidad como PCA (Análisis de Componentes Principales) y LDA (Análisis Discriminante Lineal), los cuales permiten visualizar la separación entre clases de hojas a partir de los datos espectrales [6]. Adicionalmente, se implementaron algoritmos de aprendizaje supervisado como SVM (Support vector machine) y KNN (K-Vecinos más Cercanos), con el objetivo de iniciar el entrenamiento de un clasificador que permita automatizar la identificación del estado de salud de una hoja [7].

Los resultados obtenidos muestran que, mediante estas técnicas, es posible clasificar eficazmente las hojas sanas, enfermas y deshidratadas, lo cual representa un primer paso hacia el desarrollo de una herramienta robusta de diagnóstico que pueda ser utilizada en el campo mexicano, particularmente en regiones como Guanajuato, donde la fresa representa uno de los cultivos de mayor relevancia económica.

Marco teórico

La espectroscopía es el estudio de la interacción entre la radiación electromagnética y la materia Fig. 1, en especial en la manera en cómo una sustancia absorbe, emite o dispersa la luz en diferentes longitudes de onda.

En el estudio de la óptica particularmente en los sistemas biológicos como plantas, es fundamental comprender procesos de interacción entre la luz y las superficies. Estos fenómenos no solo son fascinantes, sino que también son esenciales para entender procesos de la hoja tales como contenido de clorofila, termorregulación, cantidad de agua, etc. Estos fenómenos son llamados reflectancia, transmitancia y absorbancia.

La reflectancia y la transmitancia son definidas como tasas de radiación reflejada o transmitida (atraviesa la hoja) de la radiación incidente. La radiación indecente que no es reflejada o transmitida por la hoja es absorbida, en otras palabras, la absorbancia es definida como la cantidad de radiación que la hoja absorbe. Visto de otra manera, la reflectancia es la fracción de luz que es reflejada por la superficie de la muestra, en este caso una hoja. Similar a un espejo, donde este refleja casi toda la luz que llega a él. La transmitancia también es conocida como la fracción de luz que atraviesa la muestra, sin ser absorbida ni dispersada, al igual que una ventana de vidrio, es transparente, porque la mayor parte de la luz atraviesa a la misma. La absorbancia es vista como la cantidad de luz que una superficie absorbe a una longitud específica.

La relación entre las propiedades físicas mencionadas con la salud vegetal, se utilizan como técnicas de percepción remota y análisis espectral, esto para estimar el contenido de clorofila y otros pigmentos clave en la fotosintética y salud de la planta. Las técnicas también sirven para detectar un estrés hídrico, nutricional o por contaminantes mediante la variación de índices como la reflectancia y la absorbancia, a su vez estás técnicas permiten evaluar el contenido de agua en las hojas y evaluar la estructura interna [9].

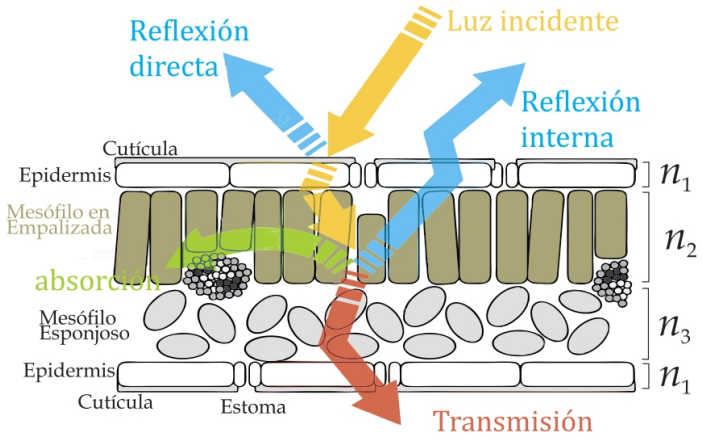

Figura 1. Procesos de interacción entre la luz y los tejidos de la hoja de una planta.

Importancia de los índices espectrales

Los índices espectrales son herramientas matemáticas derivadas de combinaciones específicas de longitudes de onda de la reflectancia, que permiten resaltar propiedades biofísicas de la vegetación y facilitar la discriminación de estados fisiológicos en cultivos. Su principal ventaja radica en su capacidad para sintetizar información compleja espectral en métricas fácilmente interpretables que se correlacionan con parámetros como contenido de clorofila, biomasa, o estrés hídrico.

- El VARI (Visible Atmospherically Resistant Index) utiliza únicamente bandas del espectro visible (rojo, verde y azul) y fue diseñado para estimar vegetación verde en condiciones atmosféricas variables. Es particularmente útil cuando no se dispone de sensores infrarrojos, como en imágenes obtenidas con cámaras RGB. Este índice detecta variaciones de color en la vegetación, asociadas comúnmente con cambios en el contenido de clorofila o estrés nutricional [8].
- El NDVI (Normalized Difference Vegetation Index) es uno de los índices más ampliamente utilizados en teledetección. Combina las bandas del rojo y el infrarrojo cercano para estimar la actividad fotosintética y la densidad de vegetación verde. Plantas sanas reflejan fuertemente en el NIR y absorben en el rojo, lo que produce valores altos de NDVI. Por lo tanto, este índice es esencial para monitorear el vigor de la vegetación y detectar áreas con estrés [12].
- El Red Edge Index explota la región del espectro conocida como “borde rojo” (~680–750 nm), donde ocurre un cambio abrupto en la reflectancia entre el rojo y el infrarrojo cercano. Esta zona es muy sensible a los cambios en el contenido de clorofila, permitiendo una detección temprana del estrés hídrico o senescencia. Este índice es especialmente valioso cuando se busca evaluar el estado fisiológico de las plantas antes de que se manifiesten síntomas visibles [3].

Reducción de dimensionalidad y clasificación

En estudios espectrales, cada muestra puede contener cientos de longitudes de onda, lo que genera conjuntos de datos muy grandes y difíciles de procesar. Esta alta dimensionalidad puede provocar redundancia y ruido, dificultando el análisis y la clasificación de los datos. Para enfrentar este problema, se utilizan técnicas de reducción de dimensionalidad, como el Análisis de Componentes Principales (PCA).

El PCA es un método que transforma los datos originales en un nuevo conjunto de variables llamadas *componentes principales* (PC1, PC2 y PC3 en Fig. 2). Estas nuevas variables resumen la mayor parte de la información presente en los datos, pero usando menos dimensiones.

Una herramienta común para decidir cuántos componentes principales conservar es el análisis de la varianza explicada acumulada. Este análisis permite observar cuánto de la variabilidad total de los datos se preserva al agregar cada componente. Generalmente, se busca identificar un punto de saturación o de inflexión en la curva, a partir del cual agregar más componentes aporta mejoras marginales. Este criterio visual facilita reducir la dimensionalidad manteniendo la mayor parte de la información relevante, y se representa gráficamente para guiar decisiones objetivas sobre la selección de componentes [10].

Luego de reducir la cantidad de variables, se aplicaron tres modelos de clasificación supervisada. Estos modelos permiten predecir si una muestra corresponde a una hoja sana, enferma o seca, a partir de los datos previamente capturados. Los modelos antemencionados son:

- **Análisis Discriminante Lineal (LDA):** Es un método estadístico que busca proyectar los datos en un espacio donde las clases queden lo más separadas posible. Para lograrlo, LDA maximiza la distancia entre los grupos (por ejemplo, entre hojas sanas y enfermas en **Fig. 2**) y, al mismo tiempo, minimiza la dispersión dentro de cada grupo. A diferencia del PCA, que no utiliza etiquetas de clase, el LDA sí tiene en cuenta la categoría a la que pertenece cada muestra, lo que lo convierte en una herramienta eficaz para tareas de clasificación multiclase.
- **Máquinas de Vectores de Soporte (SVM):** Este modelo busca una frontera de decisión que separe las clases de la mejor manera posible. En los casos más simples, esta frontera puede ser una línea recta o un plano. Sin embargo, cuando los datos no pueden separarse fácilmente, SVM emplea funciones llamadas *kernels* que permiten transformar los datos a un espacio en el que la separación sea más clara. Esta capacidad de adaptación hace que SVM sea muy preciso incluso en contextos complejos y con muchas variables [3].
- **k-Vecinos más Cercanos (KNN):** Es un modelo basado en la similitud entre muestras. Para clasificar una nueva observación, KNN busca las *k* muestras más cercanas a ella (según una medida de distancia) y asigna la clase más común entre ellas. Es un enfoque intuitivo y fácil de implementar, útil cuando no se dispone de un modelo previamente entrenado. No obstante, su rendimiento puede verse afectado por la presencia de ruido o por diferencias en la escala de los datos [15].

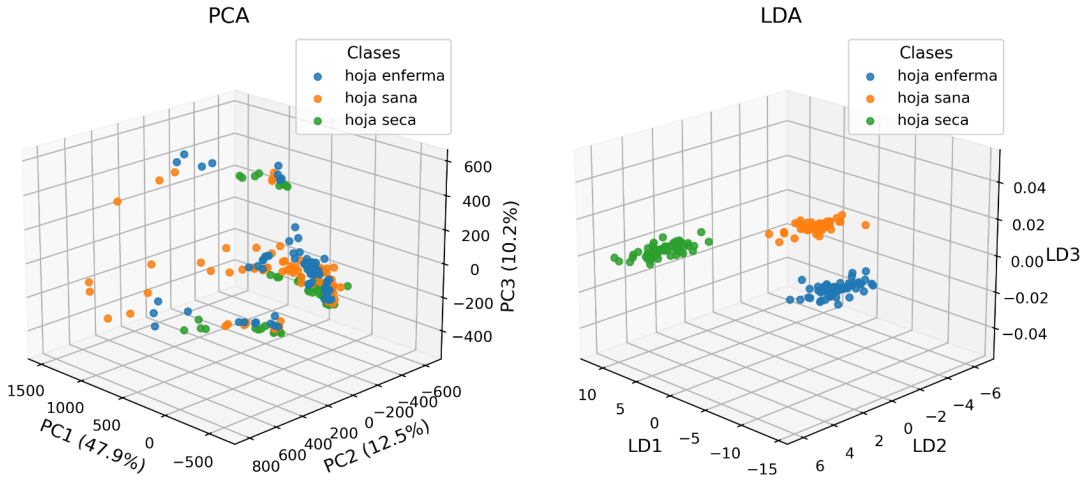

Figura 2. Visualización de separación entre clases con 3 componentes en PCA y LDA.

Evaluación y selección de modelos

Para obtener el mejor rendimiento de cada modelo (SVM, LDA y KNN), se aplicó una técnica de ajuste llamada búsqueda bayesiana de hiperparámetros. Este método permite encontrar configuraciones óptimas (como el número de vecinos en KNN o el tipo de kernel en SVM) probando diferentes combinaciones de forma inteligente. A diferencia de métodos más básicos como la búsqueda en malla (grid search), la búsqueda bayesiana reduce el número de pruebas necesarias, lo que ahorra tiempo y mejora la precisión final de los modelos [14].

Para evaluar el desempeño de los modelos de clasificación, se utilizaron 4 métricas que permiten medir qué tan bien cada modelo identifica correctamente las clases. Estas métricas se basan en comparar las predicciones del modelo con las respuestas reales.

- **Accuracy (exactitud).** Representa el porcentaje total de predicciones correctas sobre el número total de casos. Es útil como primera referencia, aunque puede ser engañosa si las clases están desbalanceadas.

$$Accuracy = \frac{número\ de\ predicciones\ correctas}{total\ de\ muestras}$$

- **Precision (Precisión).** Mide la proporción de verdaderos positivos entre todas las veces que el modelo predijo una clase positiva. Es decir, de todas las veces que el modelo dijo "es clase A", cuántas veces acertó.

$$Precision = \frac{TP}{TP + FP}\ (TP: verdaderos\ positivos, FP: falsos\ positivos)$$

- **Recall (sensibilidad o exhaustividad).** Indica la capacidad del modelo para detectar correctamente todas las muestras de una clase positiva. Es decir, de todas las veces que realmente era "clase A", cuántas veces el modelo lo reconoció.

$$Recall = \frac{TP}{TP + FN}\ (FN: falsos\ negativos)$$

- **F1-score.** Es una media armónica entre precisión y recall. Se utiliza para balancear ambos aspectos, especialmente cuando hay un desbalance entre clases.

$$F1\ score\ = 2 \cdot \frac{Precision\ \cdot\ Recall}{Precision\ +\ Recall}$$

Estas métricas permiten una evaluación más completa que el uso exclusivo de la exactitud. Por ejemplo, en un caso donde una clase es mucho más frecuente que las otras, un modelo puede tener alta ***accuracy*** pero un bajo ***F1-score***, lo que indicaría un rendimiento deficiente en las clases menos representadas.

Aplicación local

El cultivo de fresa en Guanajuato es un sector agrícola de gran relevancia, aunque representa solo el 0.25% de la superficie del estado, ocupa el séptimo lugar de la producción de hortalizas cultivadas. Guanajuato, especialmente en la región del bajío, particularmente en municipios como Irapuato, Abasolo y Salamanca, han sido históricamente líderes en la producción de fresa, llegando a ser reconocidas como “La capital mundial de la fresa” [11]. Sin embargo, la producción ha enfrentado una tendencia a la baja en años recientes, con una reducción en superficie cultivada, debido a factores como precios bajos, altos costos de producción y problemas fitosanitarios.

Además, el sistema de producción tradicional predominante en Guanajuato depende del uso intensivo de fertilizantes y agua, con poca adopción de tecnologías modernas, lo que es un área de mejora, al usar las tecnologías de precisión en el cultivo de fresa. Al usar las tecnologías de precisión se podría incrementar la productividad y rentabilidad mediante el monitoreo para el uso eficiente de insumos, a su vez mejorando la calidad del producto al tener un control sobre la salud vegetal gracias a ese control de insumos podemos reducir el impacto ambiental evitando plagas y enfermedades, dando como resultado un aumento en la competitividad del sector [9].

Metodología

Primero se comenzó en una fase inicial de procesamiento de datos espectrales, lo que comprende una ingesta y un preprocesamiento de los archivos. Para la carga de archivos, se empleó el lenguaje Python, junto a librerías como Pandas para la gestión de estructuras de datos tabulares, a su vez se utilizó Numpy para la manipulación de arreglos de alta dimensionalidad. Este enfoque permitió la ingesta de grandes volúmenes de datos, para posteriormente realizar una estandarización y limpieza de los datos.

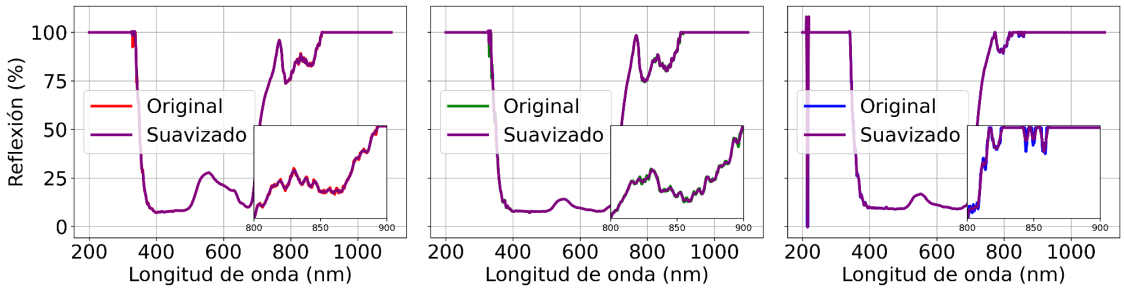

Figura 3. Comparación entre datos espectrales originales y suavizados (hoja enferma, sana y seca, respectivamente).

Dentro de la limpieza se incluyó una identificación y clasificación de los archivos dependiendo el estado de la hoja, el cual podría ser sana, enferma o seca, para posteriormente realizar un compilado de datos, antes de este compilado se realizó un suavizado y filtrado de las series espectrales, utilizando la media móvil para atenuar el ruido de las lecturas de alta frecuencia y el filtro Savitzky-Golay tomando 21 muestras y utilizando un polinomio de segundo orden para preservar las características morfológicas de las señales, al mismo tiempo que elimina el ruido.

Para mostrar los datos suavizados y el análisis de componentes principales (PCA), se decidió usar la librería de Python Matplotlib.pyplot, mostrando gráficas de reflectancia comparativa entre las distintas clases (Sana, Enferma o Seca), esto para mostrar una diferencia visual dentro de los perfiles espectrales, a su vez mostrar una correlación entre los perfiles espectrales de una misma clase.

Con las señales ya suavizadas Fig. 3, se analizó la varianza explicada acumulada por cada componente principal para determinar cuántas dimensiones conservar sin perder información relevante. A través del gráfico correspondiente Fig. 4, se identificó que a partir de 27 componentes (se utilizaron 30), la ganancia adicional de varianza explicada se vuelve mínima. Este punto de inflexión se marcó en la gráfica con una anotación, lo que permitió justificar de forma objetiva el número de componentes retenidas para el entrenamiento de los modelos. De este modo, se evita el uso de dimensiones redundantes y se mejora la eficiencia computacional sin comprometer el desempeño predictivo.

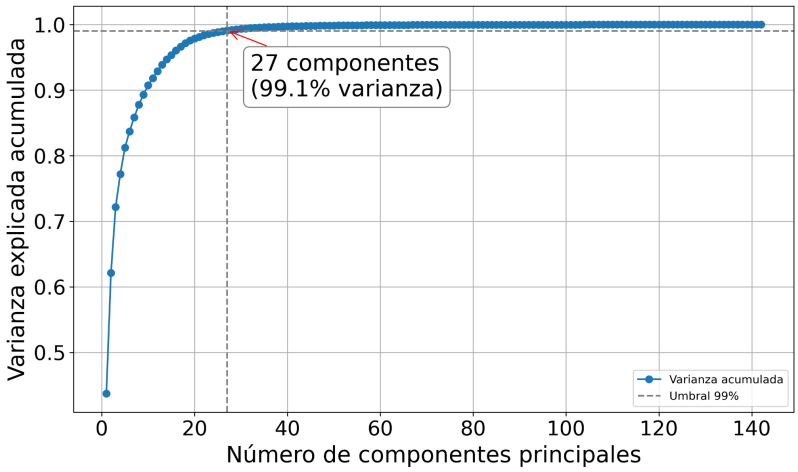

Figura 4. Análisis de varianza explicada acumulada.

Posteriormente, se construyó un conjunto de datos balanceado y listo para la fase de modelado. En esta etapa, se implementaron tres clasificadores supervisados (todos con la misma partición de datos, 70% para entrenamiento y 30% para prueba) ampliamente utilizados en problemas de teledetección y clasificación de casos multiples:

- Análisis Discriminante Lineal (LDA)
- Máquinas de Vectores de Soporte (SVM)
- k-Vecinos más Cercanos (KNN)

Cada uno de estos algoritmos fue entrenado y evaluado utilizando tres conjuntos de datos: los datos espectrales originales, los datos espectrales transformados mediante PCA y los índices vegetativos calculados (tres en total). Cabe destacar que no se aplicó PCA a los índices vegetativos, ya que al contar únicamente con tres características no requerían reducción de dimensionalidad. Esto permitió analizar el impacto de la reducción de dimensionalidad en el desempeño de los modelos.

Adicionalmente, para optimizar el rendimiento de los clasificadores, se implementó un proceso de optimización bayesiana de hiperparámetros, utilizando la técnica de Bayesian Search sobre los espacios definidos para cada modelo. Esta estrategia permitió encontrar de forma eficiente la mejor combinación de parámetros para maximizar la precisión en la predicción de clases.

Como última parte del análisis de los datos, se mostraron los índices vegetativos Fig. 5, en este caso se usaron a VARI, NDVI Y Red Edge Index a manera de comparación, utilizando histogramas para un mejor entendimiento. El índice vegetativo VARI, dentro del histograma una mayor medida de este índice representa una mayor pérdida de clorofila. Otro índice revisado es NDVI; este índice muestra una cuantificación de la salud y vigor vegetal, de esta forma una mayor medida representa un mejor estado. El último índice revisado fue Red Edge Index, que evalúa transiciones espectrales sensibles a la clorofila y por lo que una mayor medida de este representa una mejor salud vegetal.

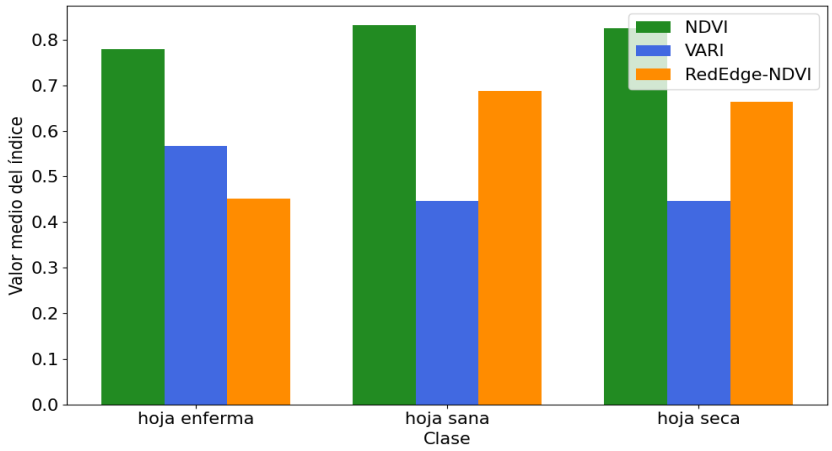

Figura 5. Comparación de índices vegetativos por clase.

Resultados

Se evaluaron tres modelos de clasificación (LDA, KNN y SVM) bajo tres enfoques: uso de todas las características espectrales, aplicación de PCA con 30 componentes principales y uso exclusivo de tres índices vegetativos. Los mejores resultados globales se obtuvieron sin reducción de dimensionalidad, donde SVM alcanzó un **accuracy del 91%** Tab. 1, seguido de LDA (87%) y KNN (85%), con métricas F1 superiores a 0.90 en varias clases. Al aplicar PCA, el rendimiento disminuyó ligeramente, siendo LDA el mejor con un **accuracy del 79%** Tab. 2, mientras que KNN y SVM bajaron a 72% y 64%, respectivamente. Finalmente, el entrenamiento con sólo los índices vegetativos arrojó el menor desempeño: SVM logró un **accuracy de 74%** Tab. 3, pero tanto LDA como KNN se quedaron por debajo del 71%, con métricas inconsistentes en la mayoría de las clases. Estos resultados evidencian que la reducción de dimensionalidad mediante PCA puede afectar negativamente el rendimiento, especialmente cuando se dispone de información espectral completa, mientras que el uso de pocos índices vegetativos limita considerablemente la capacidad predictiva de los modelos.

Tabla 1. *Comparativa entre modelos utilizando los índices vegetativos.*

Modelo de clasificación	Accuracy	*precision* - hoja enferma	*recall* - hoja enferma	*f1 score* - hoja enferma	*precision* - hoja sana	*recall* - hoja sana	*f1 score* - hoja sana	*precision* - hoja seca	*recall* - hoja seca	*f1 score* - hoja seca
LDA	0.70	1.00	0.85	0.91	0.60	0.80	0.69	0.58	0.47	0.52
K-NN	0.67	1.00	0.90	0.94	0.56	0.66	0.60	0.52	0.47	0.50
SVM	**0.74**	**1.00**	**0.90**	**0.94**	**0.62**	**0.85**	**0.72**	**0.66**	**0.47**	**0.55**

Tabla 2. *Comparativa entre modelos utilizando PCA en reducción de dimensionalidad.*

Modelo de clasificación	Accuracy	precision - hoja enferma	recall - hoja enferma	f1 score - hoja enferma	precision - hoja sana	recall - hoja sana	f1 score - hoja sana	precision - hoja seca	recall - hoja seca	f1 score - hoja seca
LDA	0.79	0.93	0.75	0.83	0.83	0.71	0.76	0.67	0.90	0.77
K-NN	0.72	0.73	0.70	0.71	0.63	0.66	0.65	0.80	0.80	0.80
SVM	0.64	0.85	0.60	0.70	0.66	0.57	0.61	0.53	0.76	0.62

Tabla 3. *Comparativa entre modelos utilizando todas las características.*

Modelo de clasificación	Accuracy	precision - hoja enferma	recall - hoja enferma	f1 score - hoja enferma	precision - hoja sana	recall - hoja sana	f1 score - hoja sana	precision - hoja seca	recall - hoja seca	f1 score - hoja seca
LDA	0.87	0.85	0.90	0.87	0.88	0.71	0.78	0.87	1.00	0.93
K-NN	0.85	0.88	0.75	0.81	0.81	0.85	0.83	0.86	0.95	0.90
SVM	0.91	1.00	0.95	0.97	0.94	0.80	0.87	0.84	1.00	0.91

Análisis de resultados

Con base en la métrica de **accuracy** y el balance entre precisión y **recall** por clase, el mejor modelo fue SVM sin reducción de dimensionalidad (todas las características).

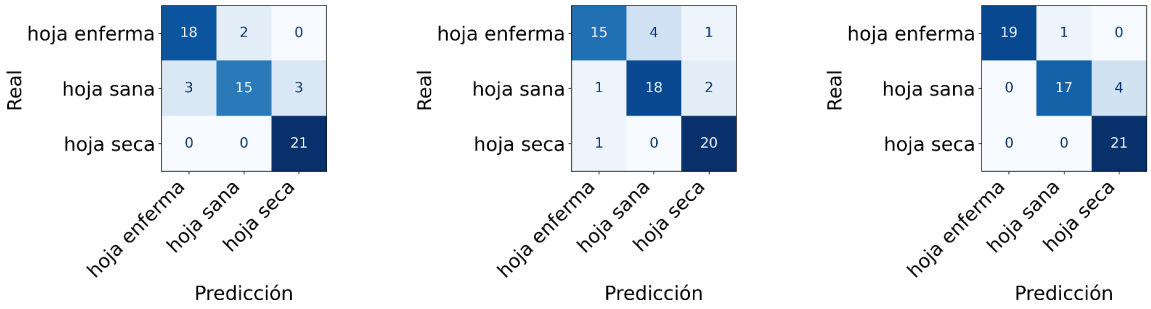

Figura 6. Matrices de confusión LDA, K-NN, SVM, respectivamente.

Este modelo logró capturar de manera eficaz los patrones espectrales asociados al estado fisiológico de las hojas, demostrando un alto rendimiento incluso al trabajar con un espacio de características de elevada dimensionalidad. La matriz de confusión Fig. 6 revela una clasificación precisa, especialmente en las categorías de hojas secas y enfermas, evidenciando su capacidad discriminativa. Asimismo, la curva de aprendizaje Fig. 7 de este modelo (líneas verdes) está por encima de las otras, además muestra una tendencia clara a la convergencia entre los errores de entrenamiento y validación, lo que sugiere una buena capacidad de generalización del modelo y permite descartar la presencia de sobreajuste o “memorización” de los datos.

Conclusión

En este trabajo se demostró la viabilidad de combinar técnicas de espectroscopía óptica e inteligencia artificial para el diagnóstico del estado fisiológico de hojas de fresa en tres categorías (sanas, enfermas y deshidratadas). A través de un riguroso proceso de preprocesamiento (suavizado de señales con media móvil y filtro de Savitzky–Golay), cálculo de índices vegetativos (VARI, NDVI y RedEdge-NDVI) y reducción de dimensionalidad (PCA y LDA), se logró sintetizar la información espectral en características altamente discriminatorias.

La fase de modelado incluyó la implementación de tres clasificadores supervisados —LDA, KNN y SVM— entrenados tanto sobre el espacio original de características como sobre el conjunto transformado por PCA (retención de las 30 componentes más significativas tras un análisis de varianza explicada). Para cada algoritmo se llevó a cabo una optimización bayesiana de hiperparámetros, garantizando el mejor rendimiento en cada caso.

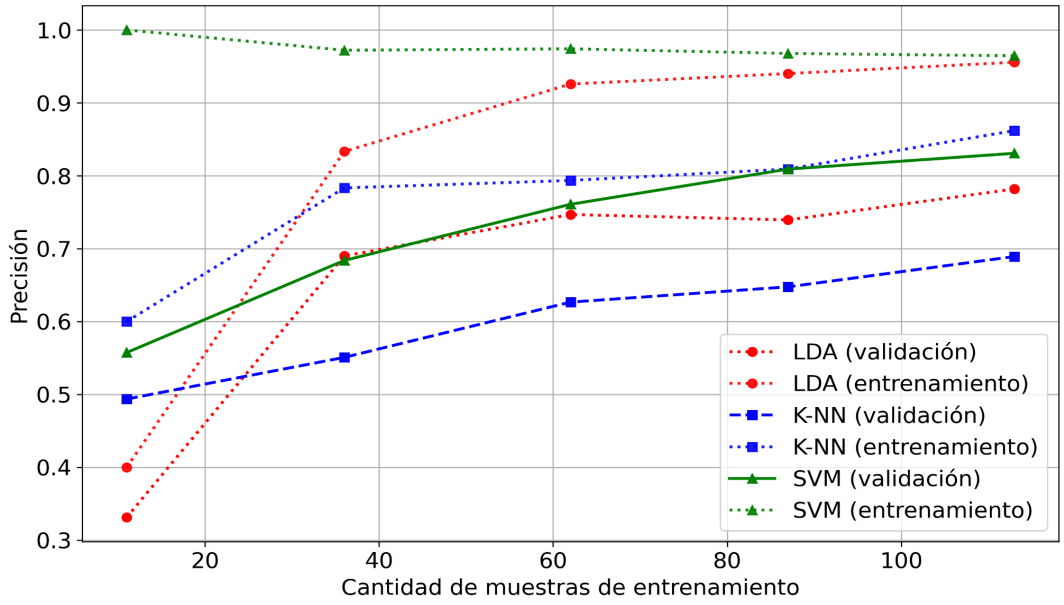

Figura 7. Curvas de aprendizaje de los modelos entrenados sin reducción de dimensionalidad.

Los resultados revelaron que, si bien la reducción de dimensionalidad con PCA ofreció ventajas en eficiencia computacional, el modelo SVM sin aplicar PCA alcanzó el mejor desempeño global (accuracy 91%, precisión y F_1 superiores a 0.90 en la mayoría de las clases). La matriz de confusión confirmó su capacidad para distinguir con alta exactitud en hojas sanas, enfermas y secas, y la curva de aprendizaje evidenció convergencia entre entrenamiento y validación, lo que descarta la presencia de sobreajuste.

Estos hallazgos validan el potencial de la propuesta para su implementación en sistemas de monitoreo de campo, especialmente en regiones productoras de fresa cómo Guanajuato. Su adopción permitiría un diagnóstico temprano de enfermedades y estrés hídrico extremo, optimizando el uso de insumos y reduciendo pérdidas económicas y ambientales.

Trabajo futuro

1. **Validación en campo:** Evaluar el sistema con datos espectrales obtenidos en condiciones reales de cultivo, ajustando la toma de datos a diferentes horarios y condiciones climáticas.
2. **Ampliación del catálogo de especies:** Adaptar y recalibrar el modelo para otras hortalizas de interés regional, como tomate o chile.
3. **Implementación en dispositivos móviles:** Desarrollar una aplicación ligera que, mediante cámaras o espectrometros portátiles, permita a productores obtener diagnósticos en tiempo real.

Con estas líneas de trabajo, se busca consolidar una herramienta práctica de agricultura de precisión que contribuya a la sostenibilidad y competitividad del sector agrícola.

Bibliografía

1. Ángeles-Núñez, J. G., & Cruz-Acosta, T. (2015). Aislamiento, caracterización molecular y evaluación de cepas fijadoras de nitrógeno en la promoción del crecimiento de frijol. *Revista Mexicana de Ciencias Agrícolas, 6*(5), 929–942.

2. Zeng, Y., Hao, D., Huete, A. *et al.* (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally, *Nat. Rev. Earth Environ.,* 3, 477– 493.

3. Chandra, M. A., & Bedi, S. S. (2021) Survey on SVM and their application in image classification**.** *Int. Journ. Inform. Tecnol., 13*, 1–11.

4. Cruz Durán, J. A., Sánchez García, P., Galvis Spínola, A., & Carrillo Salazar, J. A. (2011). Índices espectrales en pimiento para el diagnóstico nutrimental de nitrógeno. *Terra Latinoamericana, 29*(3), 259–265. Colegio de Postgraduados, Campus Montecillo.

5. Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2.ª ed.). Wiley.

6. Zhao, S., Zhang, B., Yang, J. *et al.* (2024) Linear discriminant analysis. *Nat. Rev. Methods Primers* 4, 70.

7. Cervantes, C., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends, *Neurocomputing*, 408, 189-215.

8. Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (2002). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment, 58*(3), 289–298.

9. González Álvarez, A. D., Cruz Jiménez, D., Núñez Díaz, D. A., Rodríguez Lunar, J. M., Armendáriz Flores, J. U., Canchola Martínez, M. I., & Abraham Juárez, M. R. (2025). En busca de una producción inocua de fresa con el control ecoamigable de Neopestalotiopsis rosae. *Jóvenes en la Ciencia, 28*, 1–15.

10. Greenacre, M., Groenen, P.J.F., Hastie, T. *et al.* (2022) Principal component analysis. *Nat. Rev. Methods Primers* 2, 100.

11. León López, L. L., Guzmán-Ortíz, D. L. A., García Berumen, J. A., Chávez Marmolejo, C. G., & Peña-Cabriales, J. J. (2014). Consideraciones para mejorar la competitividad de la región “El Bajío” en la producción nacional de fresa. *Revista Mexicana de Ciencias Agrícolas, 5*(4), 673–686.

12. Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS (NASA SP-351, p. 309).

13. Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639.

14. Garnett R. (2023). *Bayesian optimization. Cambridge Univ. Press, 358*

15. Ewees, A.A., Alshahrani, M.M., Alharthi, A.M. et al. (2025). Optimizing feature selection and remote sensing classification with an enhanced machine learning method, *Supercomput*, 81, 370.