



Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), julio-agosto 2024,
Volumen 8, Número 4.

https://doi.org/10.37811/cl_rcm.v8i4

**CLASIFICACIÓN DE PLANTAS
MEDICINALES EN TIEMPO REAL MEDIANTE
LA UTILIZACIÓN DE INTELIGENCIA
ARTIFICIAL EN HUERTOS DEL CANTÓN
CAMILO
PONCE ENRÍQUEZ**

**REAL-TIME CLASSIFICATION OF MEDICINAL PLANTS
THROUGH THE USE OF ARTIFICIAL INTELLIGENCE IN THE
ORCHARDS OF CANTON CAMILO PONCE ENRÍQUEZ**

Galo Fabian Salinas Matamoros
Universidad Técnica de Machala - Ecuador

Jose Julian Coronel Reyes
Instituto Tecnológico Superior España - Ecuador

Leonor Margarita Rivera-Intriago
Universidad Técnica de Machala - Ecuador

DOI: https://doi.org/10.37811/cl_rcm.v8i4.12640

Clasificación de Plantas Medicinales en Tiempo Real Mediante la Utilización de Inteligencia Artificial en Huertos del Cantón Camilo Ponce Enríquez

Galo Fabian Salinas Matamoros¹gsalinas5@utmachala.edu.ec<https://orcid.org/0000-0001-7242-1152>Universidad Técnica de Machala
Ecuador**Jose Julian Coronel Reyes**jose.coronel@iste.edu.ec<https://orcid.org/0000-0002-7883-5388>Instituto Tecnológico Superior España
Ecuador**Leonor Margarita Rivera Intriago**lriviera@utmachala.edu.ec<https://orcid.org/0000-0002-9407-1525>Universidad Técnica de Machala
Ecuador

RESUMEN

Las plantas son esenciales para la vida humana. Ayudan a respirar, proporcionan alimentos, ropa, medicinas y combustible, y también protegen el medio ambiente. Este estudio, pretende clasificar plantas medicinales a partir de un banco de imágenes, haciendo uso del aprendizaje automático. Se dispuso de una base de datos de 1.513 plantas de 7 especies entre ellas Llantén (230), Toronjil (250), Stevia (222), Violeta (218), Dulcamara (245), Geranio (198) y Sábila (150) que contienen 7 variables morfológicas de forma. Para el análisis de estas variables, se crearon modelos de clasificación de Regresión Logística (RL), Máquina de Vectores de Apoyo (SVM), Árbol de decisión (AD) y k Vecino Más Cercano (k-NN) con validación cruzada de 5 veces y se compararon métricas de rendimiento. Se determinaron tasas generales de clasificación correcta en 95,55%; 93,12 %; 91,39% y 89,12% para RL, MSV, AD, y k-NN, respectivamente. El modelo de clasificación RL, que tiene los resultados de precisión más altos, ha clasificado de las plantas medicinales de llantén, toronjil, Stevia, Violeta, dulcamara, geranio y sábila con 96,63%; 93,85%; 91,14% y 88,36%, respectivamente. Con base en los valores de medición del rendimiento obtenido, es posible decir que el estudio logró el éxito en la clasificación de especies de plantas medicinales.

Palabras claves: imágenes de plantas medicinales, aprendizaje automático, matriz de confusión, métricas de rendimiento

¹ Autor principal

Correspondencia: gsalinas5@utmachala.edu.ec

Real-time Classification of Medicinal Plants Through the use of Artificial Intelligence in the Orchards of Canton Camilo Ponce Enríquez

ABSTRACT

Plants are essential to human life. They help to breathe, provide food, clothing, medicine and fuel, and also protect the environment. This study, aims to classify medicinal plants from an image bank, making use of machine learning. A database of 1,513 plants of 7 species including Llantén (230), Toronjil (250), Stevia (222), Violet (218), Dulcamara (245), Geranium (198) and Aloe (150) containing 7 morphological shape variables was available. For the analysis of these variables, Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT) and k Nearest Neighbor (k-NN) classification models were created with 5-fold cross-validation and performance metrics were compared. Overall correct classification rates were determined to be 95,55%; 93,12%; 91.39%, and 89,12% for RL, MSV, AD, and k-NN, respectively. The RL classification model, which has the highest accuracy results, classified of the medicinal plants of plantain, lemon balm, Stevia, Violet, dulcamara, geranium, and aloe with 96,63%; 93,85%; 91,14%; and 88,36%, respectively. Based on the performance measurement values obtained, it is possible to say that the study was successful in the classification of medicinal plant species.

Keywords: medicinal plant imaging, machine learning, confusion matrix, performance metrics

Artículo recibido 20 julio 2024

Aceptado para publicación: 10 agosto 2024



INTRODUCCIÓN

Las plantas medicinales han sido utilizadas desde épocas remotas para el tratamiento de numerosas enfermedades. Desde los inicios de la civilización, han ayudado al hombre ofreciendo distintos tipos de medicinas capaces de curar ciertas dolencias gracias a sus compuestos naturales. El conocimiento sobre las plantas medicinales y sus propiedades se ha ido transmitiendo en las distintas culturas y a sus generaciones, a través del tiempo (Aguaiza Quizhpilema et al., 2021).

Ecuador es rico en una gran biodiversidad de especies vegetales medicinales, su identificación se determina por características físicas: como el tamaño, color, textura, así como por el sabor y aroma, de tal manera existen muchas plantas medicinales que se tornan dificultosas en el momento de ser reconocidas (Pisco Acebo, 2020).

Por lo tanto, es de gran utilidad una herramienta rápida y precisa que pueda clasificar las plantas medicinales a bajo costo y ampliamente estandarizado. En este sentido, la IA es una herramienta informática muy valiosa, sus técnicas son cada vez más utilizadas en diversos campos y además permiten crear sistemas que imiten un comportamiento inteligente (Ramón Fernández, 2020).

Machine Learning (ML) es una aplicación de inteligencia artificial (IA) que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia sin estar programados explícitamente (Jiménez, 2021). Un recurso sorprendente para Machine Learning es una herramienta GUI basada en la web para crear modelos de clasificación de aprendizaje automático personalizado sin experiencia técnica especializada como es, Teachable Machine (Carney et al., 2020).

A partir del análisis de un conjunto de datos de entrenamiento conocido, el algoritmo de aprendizaje automático supervisado produce una función inferida para hacer predicciones sobre los valores de salida. El sistema puede proporcionar objetivos para cualquier entrada nueva después de una capacitación suficiente (Rodríguez & Camacho, 2018).

Además, se han aplicado algoritmos para la evaluación de la calidad de los alimentos, incluida la carne de atún, cerdo y salmón (Chen et al., 2021), análisis y clasificación plantas acuáticas (Rodríguez Garlito, 2023). La combinación de técnicas estadísticas multivariantes se ha convertido en una poderosa herramienta para hacer frente a diversos problemas en el sector botánico, jardines, viveros, y los que se dedican a la compra-venta de estas especies vegetales.

Recientemente se han publicado trabajos que utilizan algoritmos para clasificar variedades de plantas medicinales, estudios comparativos con algoritmos de aprendizaje automático en base a la identificación de plantas ornamentales (Ortiz, 2023). Otra investigación logro desarrollar un sistema de visión artificial para la detección de hongos en plantas de cannabis medicinal mediante redes neuronales en ambientes controlados (Alvarez Acosta, 2021). Otra investigación de técnicas de aprendizaje de máquinas, usadas en el reconocimiento de especies vegetales (Molina Cajas & Salazar Segovia, 2022). (Pinto Armijos, 2023) desarrollo y diseñó un prototipo para la identificación de colores de semáforos vehiculares mediante la herramienta Teachable Machine de Google. (Juanazo Arriaga & Lamboggia Yar, 2024) Diseñaron e implementaron un sistema de reconocimiento facial mediante P5.JS y Teachable Machine para la apertura de una puerta.

Con este propósito el objetivo general de esta investigación es: clasificar plantas medicinales a partir de un banco de imágenes, haciendo uso de Inteligencia Artificial, y como objetivos específicos: utilizar una aplicación Web Teachable Machine para la clasificación de imágenes de plantas medicinales, y, evaluar el modelo construido y determinar si cumple con los objetivos de la investigación.

MATERIALES Y MÉTODOS

El presente estudio se desarrolló en la provincia del Azuay, ubicada en el litoral sur del Ecuador, centrando la investigación en Huertos del cantón Camilo Ponce Enríquez.

La metodología utilizada es CRIP – DM (Cross-Industry Standard Process for Data Mining) es una metodología estándar utilizada para la minería de datos, un proceso que implica descubrir patrones y conocimientos útiles a partir de grandes conjuntos de datos (Espinosa-Zúñiga, 2020).

Esta metodología es un enfoque estructurado en seis fases para la minería de datos, de las cuales utilizaremos solo cuatro que se describen a continuación: Adquisición de imágenes, extracción de características morfológicas de forma, entrenamiento del modelo e Interpretación de resultados, y, Evaluación del modelo.

El objetivo de este estudio es extraer características morfológicas, características de forma obteniendo imágenes de 7 especies de plantas medicinales. También se trata de realizar operaciones de clasificación de las características obtenidas utilizando diversas técnicas de inteligencia artificial.

Adquisición de imágenes

La cámara utilizada para el estudio tiene 12 megapíxeles, resolución de 2048×1088 y resolución completa a una frecuencia de imagen máxima de 53,7 fps. Dispone de funciones como el balance de blancos y la corrección de contraluz.

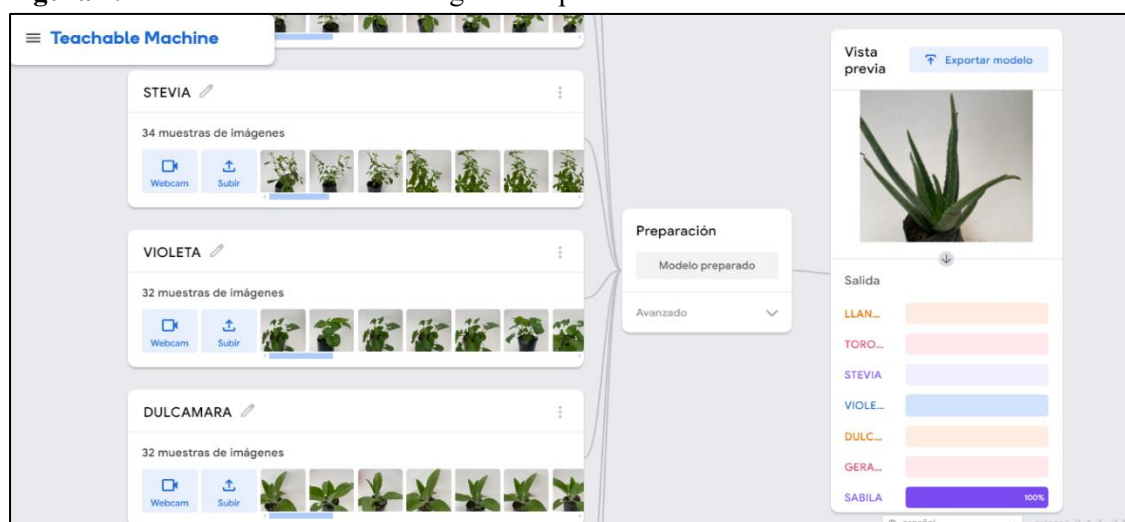
La cámara utilizada en el estudio se colocó en una caja cerrada con un dispositivo de iluminación en su interior y una estructura para evitar la recepción de luz del entorno exterior. El color de fondo de la caja se seleccionó como blanco para facilitar el procesamiento de la imagen. El tamaño de la caja se diseñó de forma que se pudieran capturar imágenes de un área de 14 cm de ancho y 18 cm de largo. La altura de la cámara se fijó a 15 cm.

Procesamiento de imágenes en Teachable Machine

Para realizar las operaciones de extracción de características y clasificación de la forma más precisa durante la fase de procesamiento de imágenes, se describieron las operaciones de preprocesamiento relacionadas con las imágenes. El procesamiento de imágenes se llevó a cabo con ayuda de la aplicación Teachable Machine.

Las imágenes tomadas de la cámara se convirtieron principalmente en imágenes en escala de grises. A continuación, se convirtieron en una imagen binaria utilizando el nivel de umbral global de la imagen en escala de grises con la ayuda del método otsu. Los objetos no deseados en las imágenes binarias resultantes se han eliminado y preparado para la etapa de extracción de características aplicando el proceso abierto.

Figura 1. Clasificación mediante imágenes de plantas medicinales



Extracción de características morfológicas de forma

Tras la etapa de procesamiento de imágenes, se obtuvieron imágenes que contenían plantas ornamentales separadas. Se extrajeron una serie de características que describían cada hoja. Las características de forma se determinaron mediante el análisis de características. Las características relacionadas con la geometría se obtuvieron con el software Matlab a partir de imágenes binarias. En total, se obtuvieron 7 características de forma para cada hoja (área, perímetro, longitud del eje mayor y menor, excentricidad, área convexa, y extensión). Los valores de todas las características se expresan en número de píxeles.

Conjunto de datos de características de plantas medicinales.

En esta investigación se utilizaron siete tipos especies de plantas medicinales (llantén, toronjil, stevia, violeta, dulcamara, geranio, y, sábila), teniendo en cuenta las características por su forma.

Tabla 1. Distribución de plantas de medicinales por especie

N°	Variedad	Total de hojas	Peso de hoja (promedio por hoja)
1	Llantén	230	0,90 g
2	Toronjil	250	0,13 g
3	Stevia	222	0,12 g
4	Violeta	218	0,45g
5	Dulcamara	245	32,55g
6	Geranio	198	2,07g
7	Sábila	150	367,50 g
	Total	1.513	57,67g

Partición de datos

Una vez que fueron procesadas las imágenes en el software Teachable Machine, se descargó el conjunto de datos y se dividió mediante una técnica de validación cruzada repetida de 5 veces para obtener subconjuntos de entrenamiento y prueba para la evaluación del modelo. Para este análisis se utilizó el lenguaje de programación Matlab y sus librerías.

Las medidas de rendimiento del modelo deben evaluarse en un conjunto de datos nuevos que no se hayan utilizado para entrenar el modelo. Un buen modelo debe ser capaz de realizar estimaciones precisas en estos datos de prueba.

La validación cruzada es una técnica común aplicada en el aprendizaje automático para maximizar el uso de los datos disponibles. En esta técnica, el conjunto de datos se divide aleatoriamente en múltiples

subconjuntos para entrenar y probar el modelo. La validación cruzada se utiliza para evitar el sobreajuste del modelo (Schaffer, 1993).

Para la obtención de los promedios al conjunto de datos se lo dividió en cinco particiones (k-folds). Se realizó un bucle a través de las 5 particiones. En cada iteración, una de las particiones se utilizó como conjunto de prueba, mientras que las otras cuatro se utilizaron como conjunto de entrenamiento (King et al., 2021). Para cada fold se entrenó los cuatro modelos utilizando el conjunto de entrenamiento asociado. Se evaluó el rendimiento de cada modelo en el conjunto de prueba, calculando las métricas de interés para cada fold.

Esto generó un conjunto de valores de métricas para cada modelo. Después de completar las cinco iteraciones, finalmente se calculó el promedio de cada métrica de rendimiento para cada modelo.

Los conjuntos de entrenamiento y prueba se cambiaron hasta que se probaron todos los pliegues (Coronel-Reyes et al., 2018). La partición de datos para cada pliegue dividió aleatoriamente el conjunto de datos, teniendo 1.210 muestras (80%) y 303 muestras (20%) para la prueba.

Métricas de Rendimientos

La creación de un nuevo modelo necesario para los problemas de clasificación o la utilización de modelos existentes y el éxito obtenido con este modelo se calculó en función del número de estimaciones precisas. Esto afecta más a la precisión de la clasificación que a la estimación de si el modelo es bueno o no. Por lo tanto, la matriz de confusión se utiliza para explicar las evaluaciones predictivas de la clasificación (de Castro et al., 2007).

Es la matriz de confusión la que proporciona información sobre las clases reales con las clases predichas realizadas por un modelo de clasificación sobre datos de prueba. En la Tabla 2 se da la matriz de confusión utilizada en la clasificación de plantas medicinales.

Tabla 2. Matriz de confusión y representación para clases múltiples

		Predicted				
		C ₁	C ₂	C ₃	...	C _n
Actual	C ₁	T₁	F12	F13	...	F1n
	C ₂	F21	T₂	F23	...	F2n
	C ₃	F31	F32	T₃	...	F3n

	C _n	Fn1	Fn2	Fn3	...	T_n

Este estudio empleó métricas de evaluación para valorar el rendimiento del método propuesto: accuracy, Error Rate, Precision, Recall y F1-Score. Los valores de estos parámetros se determinaron utilizando la matriz de confusión multiclase. Los datos de entrenamiento y de prueba se separaron mediante validación cruzada con un valor k-fold de 5. Los parámetros de evaluación se calcularon utilizando (1)-(5) (Puspitasari et al., 2023):

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn} * 100 \quad (1)$$

$$Error Rate = \frac{fp+fn}{tp+fp+tn+fn} * 100 \quad (2)$$

$$Precision = \frac{tp}{tp+fn} * 100 \quad (3)$$

$$Recall = \frac{tp}{tp+fn} * 100 \quad (4)$$

$$F1 - Score = \frac{2tp}{2tp+fp+fn} * 100 \quad (5)$$

Desarrollo de modelos

Regresión Logística (RL)

Es uno de los modelos estadísticos más utilizados. La variable dependiente en la RL se estima a partir de una o más variables. La RL aclara la relación entre las variables dependientes y las independientes (Cayambe Cajo, 2022). En LR no es necesario que las variables requieran una distribución normal. Dado que los valores predichos en LR son probabilidades, LR está acotado por 0 y 1. La razón de esto es que LR predice probabilidad en lugar de sí mismo en los resultados (Martínez et al., 2010).

Máquina Soporte Vectorial (MSV)

En los modelos MVS se utilizan diferentes funciones de núcleo. En este estudio, la clasificación se realizó utilizando la función de núcleo polinómico.

El objetivo del algoritmo MVS es encontrar un hiperplano que separe el conjunto de datos en un número predefinido y discreto de clases que sean las más consistentes con los ejemplos de entrenamiento (Arjona et al., 2018). El término separación óptima del hiperplano se refiere a la frontera de decisión que minimiza las clasificaciones erróneas durante el paso de entrenamiento.

Árbol de decisión (AD)

AD es un clasificador compuesto por múltiples AD. Para realizar una nueva clasificación, cada DT proporciona una clasificación para las entradas. Después, AD evalúa las clasificaciones y selecciona la estimación que tiene más votos.

La AD tiene la capacidad de gestionar un gran número de variables en un conjunto de datos. También tiene bastante éxito en la predicción de datos incompletos. El mayor inconveniente de la AD es su falta de repetibilidad. Además, el modelo final y los resultados posteriores son difíciles de interpretar (Borrero Becerra & Diaz Molano, 2016).

K-Nearest Neighbor (k-NN)

El método k-NN es un algoritmo de aprendizaje no paramétrico. k-NN utiliza la distancia euclidiana como parámetro en nombre de la clasificación del conjunto de datos, donde K representa el número de vecinos, para calcular la distancia entre los datos (Gonzalez-Cortés, 2019).

La k-NN está pensada para clasificar datos de muestra cuya clase es desconocida. Por esta razón, la distancia a los datos de muestra se calcula con el conjunto de datos preclasificados en el conjunto de entrenamiento. Dado que hay una cierta cantidad de datos a probar, los datos de prueba se procesan con todos los datos existentes individualmente (Durand Tarrillo & Abanto Robles, 2021).

RESULTADOS Y DISCUSIÓN

El método de clasificación de plantas se desarrolló para determinar siete tipos de variedades: Llantén, Toronjil, Stevia, Violeta, Dulcamara, Geranio, y, Sábila; por lo tanto, el número total de datos es de 1.513 hojas de plantas medicinales.

Se aplicaron los diferentes modelos de aprendizaje automático partiendo de la matriz de confusión para explicar las evaluaciones predictivas de la clasificación de la planta.

Tabla 3. Matriz de confusión de los algoritmos utilizados (a) RL, (b) MSV, (c) (AD) y (d) k-NN,

Actual	Predict						
	llantén	toronjil	Stevia	Violeta	dulcamara	geranio	sábila
Llantén	221	1	2	1	2	1	1
Toronjil	1	241	0	0	1	2	0
Stevia	0	2	216	1	2	1	2
Violeta	1	1	1	212	2	3	0
Dulcamara	3	0	1	1	239	3	0
Geranio	0	3	1	1	1	190	1
Sábila	2	1	0	2	2	1	143

(a)

Actual	Predict						
	llantén	toronjil	Stevia	Violeta	dulcamara	geranio	sábila
Llantén	218	3	2	1	2	1	4
Toronjil	1	235	4	3	1	2	3
Stevia	3	2	210	1	2	1	2
Violeta	1	1	1	209	2	3	3
Dulcamara	3	4	4	1	230	3	3
Geranio	5	3	1	3	1	181	1
Sábila	2	1	4	2	2	1	137

(b)

Actual	Predict						
	llantén	toronjil	Stevia	Violeta	dulcamara	geranio	sábila
Llantén	210	3	2	5	2	4	4
Toronjil	5	225	4	3	4	2	3
Stevia	3	2	208	5	2	4	2
Violeta	4	3	5	204	2	5	3
Dulcamara	3	4	4	4	226	3	3
Geranio	5	3	6	3	2	175	1
Sábila	2	1	4	2	2	1	131

(c)

Actual	Predict						
	llantén	toronjil	Stevia	Violeta	dulcamara	geranio	sábila
Llantén	209	4	3	5	2	4	4
Toronjil	5	220	6	3	4	2	5
Stevia	4	6	204	5	5	4	2
Violeta	4	3	5	198	6	5	5
Dulcamara	3	4	4	4	218	3	4
Geranio	5	3	6	3	2	164	6
Sábila	2	6	4	2	7	7	124

(d)

Esta matriz proporcionó información sobre las clases reales con las clases estimadas realizadas a través de un modelo de clasificación basado en los datos de la prueba. En función de los resultados, se clasificaron correctamente 1.462 hojas del algoritmo RL, 1.420 hojas del algoritmo MSV, 1.379 hojas del algoritmo AD y 1.337 hojas del algoritmo k-NN, el algoritmo RL lideró en esta categoría. El modelo

de clasificación RL, que tiene los resultados de precisión más altos, ha clasificado las especies de Llantén, Toronjil, Stevia, Violeta, Dulcamara, Geranio y Sábila con 96,63%, 93,85%, 91,14% y 88,37%, respectivamente.

En el contexto de la clasificación de plantas medicinales mediante métodos de aprendizaje automático, las métricas de rendimiento son herramientas esenciales para evaluar la eficacia y la precisión de los modelos. Para las métricas de rendimiento de la clasificación, se calcularon criterios de éxito como Accuracy, Error Rate, Recall, Precision, F1-Score. Los resultados de la evaluación del método desarrollado a partir de las características morfológicas de forma se resumen en la Tablas 4, respectivamente.

Tabla 4. Medidas de rendimiento promedio de los modelos utilizados

Measure	RL	MSV	AD	k-NN
Accuracy (%)	95,55	93,12	91,39	89,12
Error Rate (%)	4,45	6,88	8.61	10,88
Recall (%)	95,89	93,26	92,32	90,37
Precision (%)	95,74	93,34	92,15	90,26
F1-Score (%)	95,12	93,10	92,70	90,30

Cuando se examina la Tabla 4, se observa que todos los modelos excepto el k-NN tienen éxito en la clasificación con una tasa superior al 90%. El modelo RL presenta el mejor valor con un 95,55% de accuracy. Además, el modelo de clasificación RL obtenido tiene los mejores valores para todas las métricas de rendimiento calculadas. Se observa que el valor de Precisión, que es realmente positivo en todos los clasificadores, determina la relación entre el número de los clasificados positivos y el de todos los positivos, es proporcional a la accuracy.

Los hallazgos destacan que el modelo de LR demostró la mejor precisión y sensibilidad en comparación con otros algoritmos evaluados. Esto sugiere que LR es eficaz en identificar correctamente instancias de plantas medicinales, superando a otros modelos en esta categoría específica. Al examinar estudios previos en la clasificación de plantas medicinales como el de (Kan et al., 2017), también identificó al LR como un modelo eficaz para la clasificación de especies de plantas medicinales, con resultados comparables en términos de precisión y sensibilidad. Sin embargo, es importante señalar que las

métricas numéricas específicas pueden variar en comparación con otros estudios debido a diferencias en conjuntos de datos, preprocesamiento y configuración experimental. Por ejemplo, (Dileep & Pournami, 2019) informó resultados similares con los Vp pero obtuvo variaciones en los resultados al priorizar Vn, lo cual destaca la sensibilidad de los resultados a diferentes condiciones experimentales. En (Kaur & Singh, 2021), utilizando un algoritmo de Support Vector Machine para lograr el 95,8% de exactitud. El tamaño de cada imagen es de 256 * 256 píxeles. En (Naeem et al., 2021), se introduce una técnica de medición fractal para reconocer y ordenar las hojas de las plantas en función de la forma de la hoja y el diseño de las venas. Utilizando 20 clasificadores de función k de vecino más cercano, pueden lograr un alto ritmo de reconocimiento del 87,1%. La estrategia de medición fractal del volumen se utiliza para producir algoritmos para la hoja y de investigación discriminante directa (LDA), marcas de textura listas para vencer a la técnica convencional dependiente del filtro de Gabor y el examen de Fourier.

Finalmente, el rendimiento de k-NN puede depender significativamente de la elección del parámetro k. Valores de k demasiado pequeños pueden hacer que el modelo sea sensible a ruido, mientras que valores demasiado grandes pueden reducir la capacidad del modelo para capturar patrones locales (Tiwari et al., 2022). Los resultados del presente artículo respaldan la eficacia de LR en la clasificación de plantas medicinales. La comparación con estudios anteriores destaca consistencias y variabilidades, subrayando la importancia de considerar el contexto específico de la aplicación al seleccionar el modelo óptimo. Para futuras investigaciones, se sugiere explorar estrategias adicionales de preprocesamiento de datos o la aplicación de técnicas avanzadas de aprendizaje automático para mejorar aún más el rendimiento del modelo en este desafiante problema de clasificación.

CONCLUSION

En el presente artículo, en lo que respecta a la clasificación de plantas medicinales, las características de forma de las especies tienen rasgos externos distintivos, lo que hace que este proceso de clasificación sea complejo. En este estudio, los resultados obtenidos a lo largo de la validación cruzada de 5 veces muestran que la precisión de los modelos y las puntuaciones de rendimiento F1-Score se superponen. Esto demuestra que la distribución de los datos es regular. Las mayores tasas de éxito en todas las métricas indican que los modelos tienen éxito en la clasificación.

La RL presentó el mejor rendimiento global en términos de precisión en la clasificación de plantas medicinales. Esto le permite convertirse en una opción sólida y eficaz para esta investigación específica. No obstante, la RL demostró ser particularmente eficaz en la identificación precisa de plantas medicinales, liderando en la categoría TP. Esto es crucial para la aplicación práctica de la clasificación correcta de las plantas medicinales: llantén, toronjil, stevia, violeta, dulcamara, geranio, y, sábila. La RL es conocida por su capacidad para manejar conjuntos de datos complejos y no lineales. En el contexto de la clasificación de las de plantas medicinales, donde las relaciones pueden ser no lineales, la capacidad de la RL para encontrar límites de decisión óptimos es una ventaja práctica.

Dada su gran precisión y sensibilidad, la RL se perfila como la opción principal cuando es importante una clasificación precisa de las plantas medicinales. Esto puede ser crucial en aplicaciones agrícolas o industriales en las que se requiere una distinción clara entre distintos tipos de plantas medicinales. Al aplicar la RL en un entorno práctico, es esencial tener en cuenta el contexto específico de la aplicación. Evaluar factores como la interpretabilidad del modelo, la disponibilidad de recursos informáticos y el afán de aplicación es fundamental para tomar decisiones informadas. Dado que el rendimiento del modelo puede depender de las características específicas del conjunto de datos, se recomienda seguir evaluando el rendimiento de la RL en entornos de producción, con ajustes o la consideración de otras técnicas de aprendizaje automático.

REFERENCIAS BIBLIOGRAFICAS

- Aguaiza Quizhpilema, J., Simbaina Solano, J. C., Aguaiza Quizhpilema, J., & Simbaina Solano, J. C. (2021). Uso de plantas medicinales y conocimientos ancestrales en las comunidades rurales de la provincia de Cañar, Ecuador. *Revista CENIC Ciencias Biológicas*, 52(3), 223-236. http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2221-24502021000300223&lng=es&nrm=iso&tlng=es
- Alvarez Acosta, O. E. (2021). *Desarrollo de un sistema de visión artificial para la detección de hongos en plantas de cannabis medicinal mediante redes neuronales en ambientes controlados*. <http://repositoriodspace.unipamplona.edu.co/jspui/handle/20.500.12744/5316>
- Arjona, M. A. C., Canul, G. P. P., & Canché, M. J. U. (2018). U le' ts'aak: Una propuesta de prototipo de aplicación móvil para el reconocimiento y ubicación de plantas medicinales en el Estado de

- Yucatán: U le' ts'aak: a prototype proposal of a mobile application for the recognition and location of medicinal plants in the State of Yucatán. *Tecnología Educativa Revista CONAIC*, 5(2), Article 2. <https://doi.org/10.32671/terc.v5i2.72>
- Borrero Becerra, J. C., & Diaz Molano, C. A. (2016). *Elaboración de Base de Datos de Fotografías de Granos de Café Seco con Diferentes Defectos Físicos, Caracterizados con Métodos Estándar de PDI y Clasificación*. <http://repository.udistrital.edu.co/handle/11349/3824>
- Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., & Chen, A. (2020). Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-8. <https://doi.org/10.1145/3334480.3382839>
- Cayambe Cajo, F. R. (2022). *Clasificación automática de plantas monocotiledóneas y dicotiledóneas usando minería de datos* [bachelorThesis, Ecuador: Latacunga: Universidad Técnica de Cotopaxi (UTC)]. <http://localhost/handle/27000/9176>
- Chen, Z., Goh, H. S., Sin, K. L., Lim, K., Chung, N. K. H., & Liew, X. Y. (2021). *Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques* (arXiv:2106.12747). arXiv. <https://doi.org/10.48550/arXiv.2106.12747>
- Coronel-Reyes, J., Ramirez-Morales, I., Fernandez-Blanco, E., Rivero, D., & Pazos, A. (2018). Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. *Computers and Electronics in Agriculture*, 145, 1-10. <https://doi.org/10.1016/j.compag.2017.12.030>
- de Castro, A. I., Jurado-Expósito, M., & Peña-Barragán, J. M. (2007). *DISCRIMINACIÓN DE MALAS HIERBAS CRUCÍFERAS EN CULTIVOS DE INVIERNO PARA SU APLICACIÓN EN AGRICULTURA DE PRECISIÓN*.
- Dileep, M. R., & Pournami, P. N. (2019). AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 321-325. <https://doi.org/10.1109/TENCON.2019.8929394>

- Durand Tarrillo, A. G., & Abanto Robles, F. (2021). Comparación de técnicas de visión artificial para detección del defecto del fruto del café. *Repositorio Institucional - USS*. <http://repositorio.uss.edu.pe/handle/20.500.12802/9070>
- Gonzalez-Cortés, A. (2019). Morfología de semillas de siete especies del género *Opuntia* (Cactaceae) del sureste de Coahuila (México). *Boletín de la Sociedad Argentina de Botánica*, 54(4), 1-10. http://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1851-23722019000400003&lng=es&nrm=iso&tlng=es
- Jiménez, E. C. (2021). *Introducción al Machine Learning con MATLAB*. Marcombo.
- Juanazo Arriaga, B. A., & Lamboggia Yar, M. X. (2024). *Diseño e implementación de un sistema de reconocimiento facial mediante P5.JS y Teachable Machine para la apertura de una puerta* [bachelorThesis]. <http://dspace.ups.edu.ec/handle/123456789/27738>
- Kan, H. X., Jin, L., & Zhou, F. L. (2017). Classification of medicinal plant leaf image based on multi-feature extraction. *Pattern Recognition and Image Analysis*, 27(3), 581-587. <https://doi.org/10.1134/S105466181703018X>
- Kaur, P. P., & Singh, S. (2021). Classification of Herbal Plant and Comparative Analysis of SVM and KNN Classifier Models on the Leaf Features Using Machine Learning. En N. Marriwala, C. C. Tripathi, S. Jain, & S. Mathapathi (Eds.), *Soft Computing for Intelligent Systems* (pp. 227-239). Springer. https://doi.org/10.1007/978-981-16-1048-6_17
- King, R. D., Orhobor, O. I., & Taylor, C. C. (2021). Cross-validation is safe to use. *Nature Machine Intelligence*, 3(4), 276-276. <https://doi.org/10.1038/s42256-021-00332-z>
- Martínez, M., Moschini, R., Barreto, D., Bodega, J., Forjan, H., Piatti, F., Presello, D., & Valentinuz, O. (2010). Factores ambientales que afectan el contenido de fumonisina en granos de maíz. *Tropical Plant Pathology*, 35, 277-284. <https://doi.org/10.1590/S1982-56762010000500002>
- Molina Cajas, J. E., & Salazar Segovia, T. M. (2022). *Revisión de técnicas de aprendizaje de máquinas, usadas en el reconocimiento de especies vegetales* [bachelorThesis, Ecuador: Latacunga: Universidad Técnica de Cotopaxi (UTC)]. <http://localhost/handle/27000/9183>

- Naeem, S., Ali, A., Chesneau, C., Tahir, M. H., Jamal, F., Sherwani, R. A. K., & Ul Hassan, M. (2021). The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach. *Agronomy*, 11(2), Article 2.
<https://doi.org/10.3390/agronomy11020263>
- Ortiz, D. J. B. (2023). *Universidad Agraria Del Ecuador*.
- Pinto Armijos, E. R. (2023). *Diseño e implementación de un prototipo para la identificación de colores de las luces de los semáforos vehiculares mediante la herramienta Teachable Machine de Google* [bachelorThesis]. <http://dspace.ups.edu.ec/handle/123456789/25682>
- Pisco Acebo, V. H. (2020). *Inventario De Biodiversidad De Plantas Medicinales En La Parroquia La Unión Del Cantón Jipijapa* [bachelorThesis, Jipijapa.UNESUM].
<http://repositorio.unesum.edu.ec/handle/53000/2391>
- Puspitasari, N., Septiarini, A., Hairah, U., Tejawati, A., & Sulastri, H. (2023). Betel leaf classification using color-texture features and machine learning approach. *Bulletin of Electrical Engineering and Informatics*, 12(5), Article 5. <https://doi.org/10.11591/eei.v12i5.5101>
- Ramón Fernández, F. (2020). Inteligencia artificial y agricultura: Nuevos retos en el sector agrario. *Campo Jurídico (Online)*, 8(2), 123-139. <https://doi.org/10.3749/revistacampjur.v8i2.662>
- Rodríguez Garlito, E. C. (2023). *Desarrollo de nuevos métodos y algoritmos de análisis y clasificación de imágenes multiespectrales orientadas a la gestión y monitorización de plantas acuáticas invasoras* [doctoralThesis].
<https://dehesa.unex.es:8443/handle/10662/18325>
- Rodríguez Tapia, S., & Camacho-Cañamón, J. (2018). *Los métodos de aprendizaje automático supervisado en la clasificación textual según el grado de especialización*.
<http://helvia.uco.es/xmlui/handle/10396/17614>
- Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning*, 13(1), 135-143. <https://doi.org/10.1007/BF00993106>
- Tiwari, V., Joshi, R. C., & Dutta, M. K. (2022). Deep neural network for multi-class classification of medicinal plant leaves. *Expert Systems*, 39(8), e13041. <https://doi.org/10.1111/exsy.13041>