

# Identificación del Ácaro Oligonychus en Cultivos de Persea Americana Utilizando un Clasificador SVM

## Identification of the Oligonychus Mite in Persea Americana Crops Using an SVM Classifier

**Renzo Apaza Cutipa**

<https://orcid.org/0000-0002-7089-2812>, Escuela de posgrado de la Universidad Nacional del Altiplano - Perú

**Gina Fiorella Charaja Sánchez**

Escuela de posgrado de la Universidad Nacional del Altiplano - Perú

**Vladimiro Ibañez Quispe**

<https://orcid.org/0000-0002-0277-4945>, del Instituto de Investigación en Ciencias de la Computación UNA - Puno

**Ernesto Nayer Tumi Figueroa**

<https://orcid.org/0000-0003-0401-9027>, del Instituto de Investigación en Ciencias de la Computación UNA - Puno

### Resumen

El presente proyecto tuvo como objetivo realizar la identificación del ácaro de Palto. Para ello se planteó una aproximación basada en el aprendizaje de la máquina realizando extracción de características estadísticas y un clasificador basado en SVM. Se concluye que las características estadísticas utilizadas fueron suficientes para lograr que el clasificador basado en SVM logre una clasificación correcta del 100%, es decir para el caso particular de identificar una hoja sana de una hoja con presencia de ácaros el valor mínimo, primer cuartil, mediana, media aritmética, tercer cuartil y el valor máximo por cada canal de color del modelo aditivo RGB fueron suficientes para la correcta clasificación sobre la base de datos propuesta.

*Palabras clave:* identificación automática de enfermedades foliares, aprendizaje de la máquina, extracción de características, cultivo de palto, Oligonychus, clasificador

### Abstract

The objective of this project was to identify the Palto mite. For this, an approach based on machine learning was proposed, extracting statistical characteristics and a classifier based on SVM. It is concluded that the statistical characteristics used were sufficient to achieve that the SVM-based classifier achieves a correct classification of 100%, that is, for the particular case of identifying a healthy leaf of a leaf with the presence of mites, the minimum value, first quartile, median, arithmetic mean, third quartile and the maximum value for each color channel of the additive RGB model were sufficient for the correct classification on the proposed dataset.

*Keywords:* automatic identification of foliar diseases, machine learning, extraction of characteristics, avocado cultivation, Oligonychus, Support Vector Machine classifier.

### 1. Introducción

A medida que las plantaciones del palto van creciendo son atacadas por diferentes plagas y sus enemigos se hacen más numerosos. Teniendo como resultado (pérdida de productividad, ingresos e inversiones). Las enfermedades en los cultivos tienen efectos críticos sobre la calidad y cantidad de los productos agrícolas destruyendo el normal estado del cultivo y cambiando o interrumpiendo sus funciones vitales como el fotosíntesis, transpiración, polinización, fertilización y germinación (Wang, X.F., Wang, Z., Zhang, S.W. et al., 2015). El diagnóstico de enfermedades basado en la detección de síntomas tempranos es una tarea habitual que se tiene en cuenta para las estrategias integradas de control de plagas (Johannes et al., 2017). El reconocimiento consiste en la operación de adquirir una imagen nueva no incluida en la etapa de entrenamiento y posteriormente asignar a la imagen una clase almacenada en la base de datos (Apaza Cutipa et al., 2019), La precisión de reconocimiento y el diagnóstico de las

enfermedades en los cultivos en los estados iniciales es muy importante (Gavhale & Gawande, 2014). Dentro de las áreas de conocimiento de la informática debemos señalar que el procesamiento de imágenes es un área emergente donde las investigaciones y los avances están tomando un progreso geométrico en el campo agrícola. Varias investigaciones continúan en la detección de enfermedades de plantas. La identificación de enfermedades de las plantas no solo puede maximizar la producción del rendimiento, sino que también puede ser útil para diversos tipos de prácticas agrícolas (Pooja et al., 2017). La mayoría de los algoritmos existentes de reconocimiento de enfermedades de cultivos basados en imágenes se basan en la extracción de varios tipos de características a partir de imágenes de hojas de plantas enfermas (Zhang et al., 2017). Los síntomas de las enfermedades en cultivos se pueden apreciar de forma visual sobre las hojas de los plantíos así es posible detectar de automáticamente estas enfermedades mediante la aplicación de técnicas de aprendizaje de la máquina sobre las imágenes

de las hojas (Camargo & Smith, 2009). En este artículo se presenta una aproximación para la Identificación del ácaro de Palto (*Oligonychus*) basada en aprendizaje de la máquina, aprendizaje supervisado que utiliza un extractor de características estadístico sobre las representaciones de color de las imágenes y utilizando un clasificador basado en máquinas de soporte vectorial (SVM).

El resto del artículo está organizado como sigue: en la sección materiales y métodos se describen la forma de la elaboración de la base de datos de imágenes, la extracción de características y el método de clasificación; en la sección de experimentos y resultados se expone los resultados obtenidos al aplicar el método propuesto; y la sección 3 está dedicada a la conclusión del trabajo y propuesta de futuros trabajos.

## 2. Métodos

Para poder realizar la identificación de una hoja de palta afectada por el acaro del palto y diferenciarla de una hoja sana de forma automática es necesario construir un banco de imágenes, extraer características representativas y posteriormente realizar la clasificación de las etiquetas de las imágenes. La descripción de estos métodos se amplía a continuación.

### 2.1. Elaboración del banco de Imágenes

Se realizó la captura fotográfica de ejemplares de hojas del cultivo de palta que contenían la presencia de ácaros y de hojas sanas algunos ejemplares se pueden apreciar a en la Figura 1.

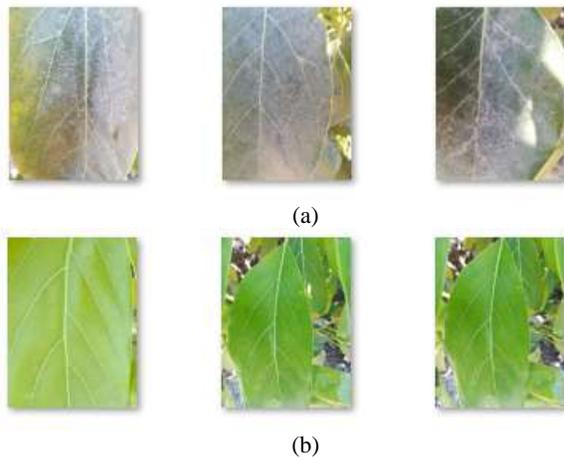


Figura 1. Ejemplares de hoja de palta (a) con presencia de ácaros, (b) sanas

Cada una de las imágenes fue segmentada y recortada en otras imágenes de 100 x 100 píxeles de dimensión en formato JPEG tanto para las sanas como para las enfermas, estas imágenes son las que conforman la base de datos de imágenes que hacen un total de 210 ejemplares sanos y 210 ejemplares con presencia de ácaros algunas de ellas se muestran en la Figura 2.

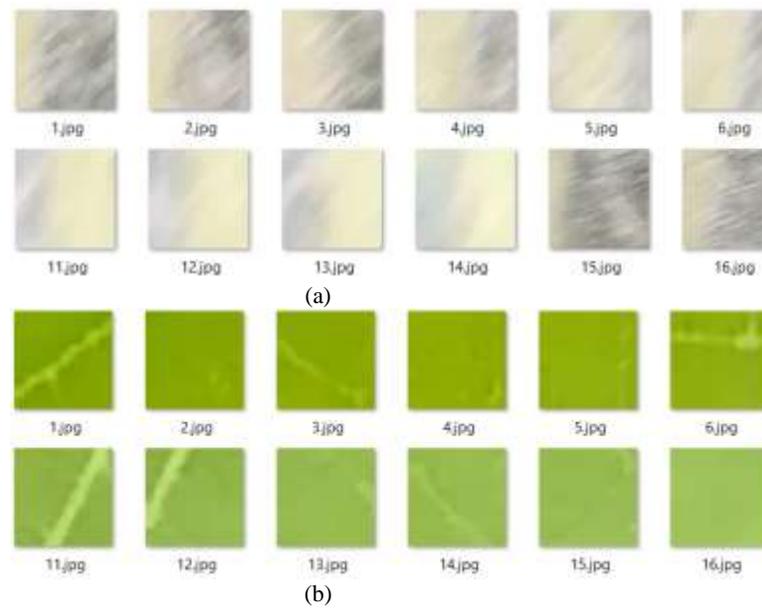


Figura 2. Imágenes que conforman la base de datos (a) con presencia de ácaros, (b) sanas.

### 2.2. Extracción de Características

Las enfermedades de los cultivos pueden ser clasificadas o reconocidas analizando color, textura y forma de las imágenes de hojas afectadas con las enfermedades (Al-Hiary et al., 2011). Muchos métodos que utilizan la extracción de características han sido desarrollados, como ejemplo Das y Dey extraen características de la textura utilizando las características estadísticas de Co-Occurrence Matrix (GLCM) de nivel gris (Das & Dey, 2017). En la presente propuesta se elaboró un extractor de características estadísticas aplicado al modelo de color RGB cada canal de color de la imagen de hoja (Rojo, verde y azul) que incluyen al valor mínimo, primer cuartil, mediana, media aritmética, tercer cuartil y el valor máximo.

#### 2.2.1. Valor Mínimo.

Es el valor de una muestra mas pequeño sirve para identificar los valores atípicos o un error de entrada de datos.

#### 2.2.2. Cuartil 1 (Q1).

Es una medida de posición que divide al 25% de los datos su cálculo se realiza con:

$$Q1 = \frac{(N + 1)}{4} \quad (1)$$

#### 2.2.3. Mediana.

Es una medida de posición, indica la posición central de un conjunto de datos divide la muestra en el 50% de los datos.

$$M_e = \frac{(N + 1)}{2} \quad (2)$$

**2.2.4. Media aritmética.**

Describe a la muestra con un solo valor representa el centro de los datos.

$$\bar{X} = \sum_{i=1}^N \frac{X_i}{N} \quad (3)$$

**2.2.5. Cuartil 3 (Q3).**

Es una medida de posición que divide al 75% de los datos su cálculo se realiza con:

$$Q3 = 3 * \frac{(N + 1)}{4} \quad (4)$$

**2.2.6. Valor Máximo.**

Es el valor mas grande de una muestra sirve para identificar también los valores atípicos o errores de entrada de datos

**2.3. Clasificación**

La etapa de clasificación es el proceso final de un proceso automático de identificación de enfermedades foliares para ello existen diversas técnicas y algoritmos que pueden realizar este proceso. El presente artículo utiliza para la clasificación la Máquina de soporte vectorial (SVM) de base lineal. Dado un conjunto de muestras de entrenamiento con las clases etiquetadas la SVM busca el hiperplano que separe de forma óptima los puntos de una clase de la de otra (Bishop, 2013). La SVM busca el hiperplano que maximice la distancia (o margen) con los puntos que estén más cerca de él mismo, razón por la cual también se les conoce a las SVM como clasificadores de margen máximo (Bishop, 2013).

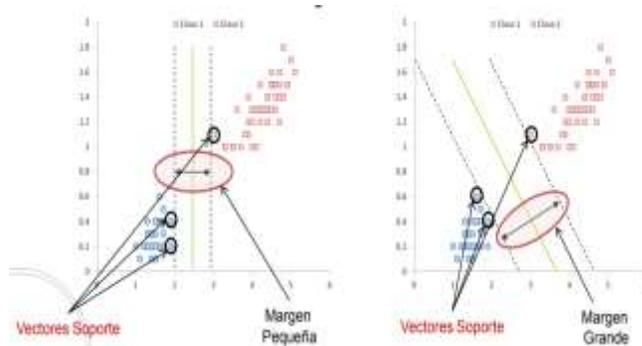


Figura 3. Vectores de soporte

El algoritmo de la Máquina de soporte Vectorial Utilizado se construyó basado en el algoritmo mostrado en al Tabla 1:

Tabla 1. Algoritmo de clasificación SVM

Algoritmo SVM
---------------

```

Input S, λ, T y k.
Elegimos w1 tal que ||w1|| ≤ 1/λ;
for n = 1 ... T do
(1) Elegimos At ⊆ S, donde |At| = k.
(2) Definimos At+ = {(x, y) ∈ At : ywtx < 1}
(3) ηt = 1/λt
(4) wt+1/2 = (1 - ηtλ)wt + ηt ∑(x,y) ∈ At+ yx
(5) wt+1 = min{1, 1/√λ} ||wt+1/2
end for
Output wT+1
    
```

El algoritmo que se ha utilizado fue el que esta implementado en el lenguaje R utilizando el paquete 'e1071'

**3. Resultados**

**3.1. Extracción de Características**

Se realizó el proceso de separación de las imágenes de las hojas del cultivo de paltas por canal de color basado en el modelo aditivo RGB tanto para las imágenes sanas como para las afectadas con el ácaro. Se aprecia en la Figura 4 los resultados de esta operación.

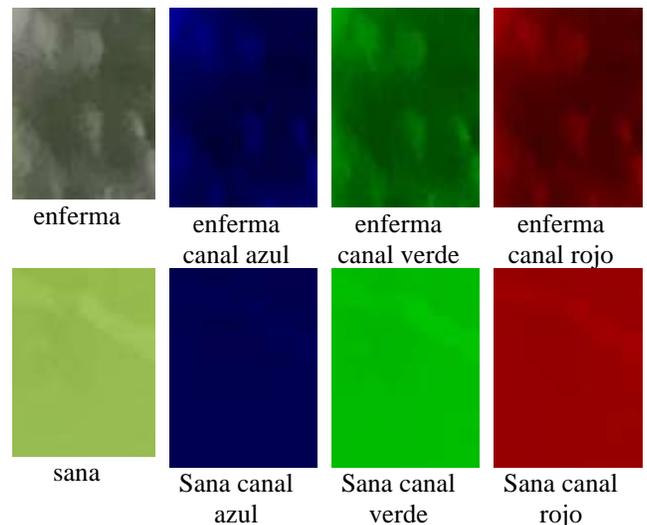


Figura 4. Imágenes de hojas de cultivo de palta por canal de color

Esto se realizó en el lenguaje R versión 3.5.0 utilizando la librería EBImage localizada en la URL <https://bioconductor.org/biocLite.R>. A continuación, se realizó la extracción de características de la totalidad de las imágenes de la base de datos. A continuación, en la Figura 5, se muestra fragmento del código utilizado para extraer data de las matrices de las imágenes y convertirlas en vectores de datos

```

imagen_gray <- channel(img, "gray")
img_gray <- imagen_gray@.Data #matriz
vector_gray <- as.vector(img_gray) #vector
imgMatGris <- rbind(imgMatGris, as.array(vector_gray))
imagen_red <- channel(img, "red")
img_red <- imagen_red@.Data
vector_red <- as.vector(img_red)
imgMatRed <- rbind(imgMatRed, as.array(vector_red))
imagen_green <- channel(img, "green")
img_green <- imagen_green@.Data
vector_green <- as.vector(img_green)
imgMatGreen <- rbind(imgMatGreen, as.array(vector_green))
imagen_blue <- channel(img, "blue")
img_blue <- imagen_blue@.Data
vector_blue <- as.vector(img_blue)
imgMatBlue <- rbind(imgMatBlue, as.array(vector_blue))
    
```

Figura 5. Extracción de datos

A continuación, se procedió a consolidar la matriz de características halladas el procedimiento utilizado se resume en la Figura 6.

```

resgray <- transpose(apply(imgMatGris, 1, summary))
resred <- transpose(apply(imgMatRed, 1, summary))
resgreen <- transpose(apply(imgMatGreen, 1, summary))
resblue <- transpose(apply(imgMatBlue, 1, summary))
caracteristicas <- as.data.frame(cbind(resgray, resred, resgreen, resblue, enfermedad))
    
```

Figura 6 consolidación de la matriz de características

Los valores iniciales del dataframe de las imágenes se muestran a continuación.

```

head(caracteristicas, n=1)
  gray_min gray_max gray_mean gray_std gray_min gray_max gray_mean gray_std gray_min gray_max gray_mean gray_std gray_min gray_max gray_mean gray_std
1  0.00  1.00  0.50  0.50  0.00  1.00  0.50  0.50  0.00  1.00  0.50  0.50  0.00  1.00  0.50  0.50
  enfermedad
1  SANA
    
```

Figura 7. Características

### 3.1. Resultados del entrenamiento del algoritmo de Clasificación

Para la etapa de entrenamiento se implementó el Algoritmo de SVM en el lenguaje de programación R utilizando el paquete CARET para entrenar los algoritmos bajo el siguiente criterio:

Tabla 2. Cantidad de utilizados para validación y entrenamiento del modelo

Grupo	Descripción	% de la base de datos
1	Grupo de Validación	25%
2	Grupo de Entrenamiento	75%

Se utilizó la siguiente rutina en R.

```

library(e1071)
modelo.svm <- svm(f, data=datos.train)
save.modelo.svm, file = "modelosvm.rda") #para guardar los modelos
pred <- predict(modelo.svm, datos.test, type="class")
# Matriz de confusión
mc <- table(pred, datos.test[,dep], dnn = c("Asignado", "Real"))
mc <- mc[order(rownames(mc)), order(colnames(mc))]
cat("** SVM\n")
print(mc)
    
```

Figura 8. Rutina para el clasificador basado en SVM

Se utilizó el Error de la validación cruzada de K iteraciones. El resultado final lo obtenemos a partir de realizar la media aritmética de los K valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{K} \sum_{i=1}^K E_i$$

Se muestran a continuación los resultados de la evaluación  
Tabla 3

Nro	Iteración K	$E_i$
1	10	0,2
2	50	0,18
3	100	0,0888
4	300	0,008999

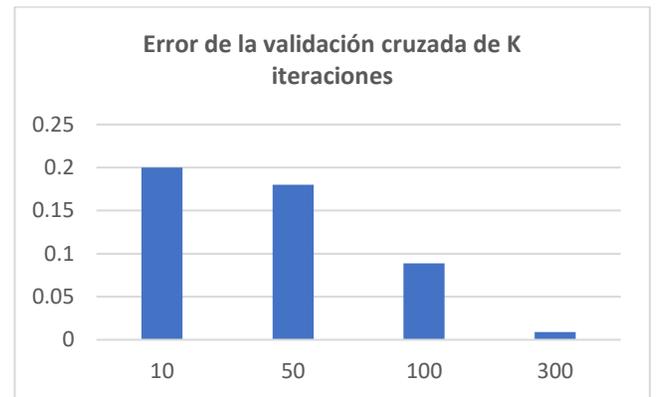


Figura 9. Gráfico de barras del Error de validación cruzada de K iteraciones

También se obtuvieron como resultados de la validación cruzada la matriz de confusión de la evaluación de las características para determinar si una imagen de hoja de palto está afectada por la presencia de ácaros.

```

** SVM
      Real
Asignado ENFERMA SANA
ENFERMA   52     0
SANA      0     52
Correctamente clasificados: 100 %
    
```

Figura 10. Matriz de confusión de la clasificación basado en SVM

### 4. Conclusiones

La identificación de plagas en etapas tempranas es necesario para un adecuado manejo de los cultivos de palta. En el caso particular de la identificación del ácaro de Palto el presente proyecto planteó una aproximación para su identificación basado en características estadísticas y un clasificador basado en SVM. Se concluye que las características estadísticas utilizadas fueron suficientes para lograr que el clasificador basado en SVM logre una clasificación correcta del 100%, es decir para el caso particular de identificar una hoja sana de una hoja con presencia de ácaros el valor mínimo, primer cuartil, mediana, media aritmética, tercer cuartil y el valor máximo por cada canal de color fueron suficientes para la

correcta clasificación. Dados los resultados y en busca de generalizar la detección de enfermedades a otras plagas se plantea como futuros proyectos la extracción de características geométricas y de puntos característicos.

## Referencias

- Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., & Al-Rahamneh, Z. (2011). Fast and Accurate Detection and Classification of Plant Diseases. *International Journal of Computer Applications*, 17(1), 31-38. <https://doi.org/10.5120/2183-2754>
- Apaza Cutipa, R., Charaja Flores, G., Tumi Figueroa, N. E., & Ibañez Quispe, V. (2019). Facial detection and recognition system using hybrid techniques in images and video sequences. *Revista de Investigación C&T Riqchary 1 (1)*, 41-46.
- Bishop, C. M. (2013). Pattern Recognition and Machine Learning. En *Journal of Chemical Information and Modeling* (Vol. 53, Número 9). <https://doi.org/10.1117/1.2819119>
- Camargo, A., & Smith, J. S. (2009). Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture*, 66(2), 121-125. <https://doi.org/10.1016/j.compag.2009.01.003>
- Das, A., & Dey, A. K. (2017). Leaf Disease Detection , Quantification and Classification Using Digital Image Processing. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS) Detection*, 5(11), 45-50. <https://doi.org/10.17148/IJIREEICE.2017.51110>
- Gavhale, M. K. R., & Gawande, P. U. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering*, 16(1), 10-16. <https://doi.org/10.9790/0661-16151016>
- Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., & Ortiz-Barredo, A. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Computers and Electronics in Agriculture*, 138, 200-209. <https://doi.org/10.1016/j.compag.2017.04.013>
- Pooja, V., Rahul, D., & Kanchana, V. (2017). Identification of Plant Leaf Diseases using Image Processing Techniques. *2017 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2017)*, Tiar, 3-6.
- Wang, X.F., Wang, Z., Zhang, S.W. et al., 2015. Monitoring and discrimination of plant disease and insect pests based on agricultural IOT. In: International. (2015). *Monitoring and discrimination of plant disease and insect pests based on agricultural IOT. In: International Conference on Information Technology and Management Innovation (ICITMI 2015)*, pp. 112–115. *Icitmi*, 112-115.
- Zhang, S., Wu, X., You, Z., & Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Computers and Electronics in Agriculture*, 134, 135-141. <https://doi.org/10.1016/j.compag.2017.01.014>