UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA



"APLICACIÓN DE ALGORITMOS INTELIGENTES PARA RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA"

TESIS

PRESENTADO POR:

GREGORIO TOMAS CASTRO ALVAREZ

Para optar el Título Profesional de:

INGENIERO DE SISTEMAS E INFORMÁTICA

ILO – PERÚ

2019

05

UNIVERSIDAD NACIONAL DE MOQUEGUA
CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

ACTA DE SUSTENTACIÓN

El Jurado suscribiente ha calificado el trabajo de Tesis:

TITULADO: Aplicación de Algoritmos Inteligentes para reconocimiento automático de enfermedades foliares de cultivo de Palta

PRESIDENTE: Hugo Euler Tito Chura

MIEMBROS: Aníbal Fernando Flores García

Alex Peter Zúñiga Incalla

ASESOR: Carlos Alberto Silva Delgado

PRESENTADO POR EL BACHILLER: Gregorio Tomas Castro Alvarez

DE LA PROMOCIÓN: 2014

CUYOS RESULTADOS HAN SIDO LOS SIGUIENTES: Quince

CALIFICATIVO DEL TRABAJO: Quince

CALIFICATIVO DE LA SUSTENTACIÓN: Quince

CALIFICATIVO FINAL: 15 (Quince)

POR LO EXPUESTO, EL BACHILLER: Gregorio Tomas Castro Alvarez

HA SIDO DECLARADO EXPEDITO PARA QUE SE LE CONFIERA EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS E INFORMÁTICA

EN FE DE ELLO QUEDA ASENTADA LA PRESENTE ACTA

ILO, A LOS 30 DÍAS DE mayo DE 2019

Presidente

Miembro

Miembro

DEDICATORIA Y/O AGRADECIMIENTO

El presente trabajo está dedicado a Dios por velar siempre por mi bienestar, a mis Padres, quienes supieron inculcar en mí valores y principios para cumplir mis metas, una de ellas ser Profesional, también a mis hermanos y hermanas que me apoyaron cuando más los necesitaba.

Seguidamente agradecer a la máxima casa de estudios, la Universidad Nacional de Moquegua, la cual me brindó la oportunidad de desarrollarme tanto académico, como culturalmente y permitiéndome conocer y vivir todo lo que rodea a esta gran Institución, en la que pasé los momentos más agradables y también los más complicados.

Finalmente agradecer a mi Asesor Mg. Carlos Alberto Silva Delgado, por apoyarme incondicionalmente durante el desarrollo de mi proyecto de tesis.

RESUMEN

Se desarrolló un trabajo de investigación de: “APLICACIÓN DE ALGORITMOS INTELIGENTES PARA RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA” en el que se aplicaron cuatro algoritmos inteligentes: Naive Bayes, Random forest, Redes Neuronales y Support Vector Machines. El proyecto de investigación tiene como objetivo general la determinación del algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta.

La presente investigación está dividida en cinco (5) capítulos tales como: datos generales, planteamiento del problema, marco teórico, marco metodológico, presentación de resultados, adicional a los capítulos mencionados también mencionar las conclusiones, recomendaciones, referencias bibliográficas y anexos.

Se trabajó mediante la metodología de etapas de procesamiento de imágenes tales como: a) adquisición de imagen, b) preprocesamiento de imágenes, c) extracción de características y d) reconocimiento; luego de la evaluación de su eficacia se obtuvo que el algoritmo Maquina de Soporte Vectorial tiene mayor asertividad de 96% en el reconocimiento de enfermedades foliares del cultivo de palta, esto después de ser evaluados con Matriz de Confusión.

ABSTRACT

We developed a research work: "IMPLEMENTATION OF INTELLIGENT ALGORITHMS FOR AUTOMATIC RECOGNITION OF FOLIAR DISEASES OF CULTIVATION OF AVOCADO" applied to four intelligent algorithms: Naive Bayes, Random forest, Neural Networks and Support Vector Machines. The research project has as its general objective the determination of the most effective, intelligent algorithm for the recognition of images of the foliar disease of avocado.

This research is divided into five (5) chapters such as: general data, problem approach, theoretical framework, the methodological framework, presentation of results, additional to those chapters also mention the findings, recommendations, bibliographical references and Annexs.

You work through the methodology of image processing stages such as: (a) image acquisition, image preprocessing (b), (c) Removal of features and d) Recognition; then evaluate its effectiveness was obtained that the algorithm Support Vector Machine has greater assertiveness of 96% in the recognition of foliar diseases of the cultivation of avocado, this after being evaluated with Confusion Matrix.

ÍNDICE

DEDICATORIA Y/O AGRADECIMIENTO i
RESUMEN ii
ABSTRACT iii
INTRODUCCIÓN 1
I. DATOS GENERALES 3
1.1 TÍTULO 3
1.2 Tipo de Investigación 3
1.3 Área de Investigación 3
1.4 Nombre de Autor 3
1.5 Localidad donde se realizó la Investigación 3
1.6 Asesor 3
II PLANTEAMIENTO DEL PROBLEMA 4
2.1 DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA 4
2.2 FORMULACIÓN DEL PROBLEMA 5
2.2.1 Interrogante General 5
2.2.2 Interrogantes Secundarias 5
2.3 JUSTIFICACIÓN 6
2.4 FORMULACIÓN DE OBJETIVOS 7
2.4.1 Objetivo General 7
2.4.2 Objetivos Específicos 7
2.5 FORMULACIÓN DE HIPÓTESIS 8
2.5.1 Hipótesis General 8
2.5.2 Hipótesis Específicas 8
III MARCO TEÓRICO 9
3.1 ANTECEDENTES DEL ESTUDIO 9
3.2 BASES TEÓRICAS 14
3.2.1 Naive Bayes 14
3.2.2 Random forest (Bosques Aleatorios) 15
3.2.3 Redes Neuronales 17
3.2.4 Las máquinas de soporte vectorial (SVMs) 20
3.2.5 Modo de color RGB 21

3.2.6 Matriz de Confusión 23
3.2.7 Estadística descriptiva univariante: variables cuantitativas ... 24
3.2.7.1 Medidas de centralización: 24
3.2.7.2 Medidas de orden o posición (localización): 25
3.2.8 Resumen de 5 números. 26
3.2.9 METODOLOGÍA 26
3.2.9.1 Adquisición de imagen. 27
3.2.9.2 Preprocesamiento de la imagen. 27
3.2.9.3 Extracción de característica. 28
3.2.9.4 Reconocimiento 28
3.2.10 R (lenguaje de programación). 28
3.3 DEFINICIÓN DE TÉRMINOS 29
IV MARCO METODOLÓGICO 30
4.1 LUGAR DE EJECUCIÓN 30
4.2 TIPO Y DISEÑO 30
4.3 NIVEL DE INVESTIGACIÓN 30
4.4 OPERACIONALIZACIÓN DE VARIABLES 30
4.5 POBLACIÓN Y MUESTRA 31
4.5.1 La Población 31
4.5.2 Muestra 31
4.6 TÉCNICAS E INTRUMENTOS PARA RECOLECCIÓN DE DATOS ... 32
4.6.1 Técnica 32
4.6.2 Instrumento para recolección de datos 32
4.7 DISEÑO EXPERIMENTAL Y EVALUACIÓN DE LA EFICACIA DE LOS ALGORITMOS DE CLASIFICACIÓN 32
4.7.1 DISEÑO EXPERIMENTAL 32
4.7.1.1 Metodología (Etapas para reconocimiento de patrones) .. 32
4.7.1.1.1 Adquisición de datos. 32
4.7.1.2 Preprocesamiento 33
4.7.1.3 Extracción de características. 38
4.7.1.4 Reconocimiento 43
4.7.2 EVALUACIÓN DE LA EFICACIA DE LOS ALGORITMOS DE CLASIFICACIÓN 43
4.7.2.1 Precisión (FP). 43

4.7.2.2 Valor de predicción Negativa (TN)..... 43
4.7.2.3 Error de Clasificación..... 44
4.7.2.4 Exactitud (Accuracy). 44
V PRESENTACIÓN DE RESULTADOS 45
5.1 Resultado de objetivo general 45
5.2 Resultados de Objetivos Específicos. 45
CONCLUSIONES..... 53
Conclusión General 53
Conclusiones Específicas 53
RECOMENDACIONES..... 54
REFERENCIAS BIBLIOGRÁFICAS..... 55
ANEXOS 57

ÍNDICE DE FIGURAS

Figura 1: Esquema de Random Forest 17
Figura 2: Esquema de una red Neuronal 19
Figura 3: Comparación entre una neurona biológica (izquierda) y una artificial (derecha). 19
Figura 4: La frontera de decisión debe estar tan lejos de los datos de ambas clases como sea. 21
Figura 5: Distribución de colores en el espacio RGB. 22
Figura 6: Se recolecto en total 330 imágenes sanas (se adjunta algunas imágenes) 33
Figura 7: Se recolecto en total 300 imágenes enfermas (Oligonychus sp.), se adjunta algunas imágenes. 33
Figura 8: Imagen enferma Oligonychus sp – Imagen original..... 34
Figura 9: Oligonychus sp – Procesado (se incrementa sumando brillo +0.3) 35
Figura 10: Oligonychus sp – se incrementa multiplicando contraste *0.3 a la imagen original. 36
Figura 11: Oligonychus sp – corrección gama (se incrementa brillo ^0.3.)..... 37
Figura 12: Imágenes sanas de 100x100 pixeles 38
Figura 13: Imágenes enfermas de 100x100 pixeles. 38
Figura 14: Imagen sana “1”..... 39
Figura 15: Imágenes enfermas de 100x100 pixeles de formato JPG. 46
Figura 16: Imágenes sanas de 100x100 pixeles de formato JPG. 46
Figura 17: Evaluación de desempeño de modelo de Naive Bayes..... 49
Figura 18: Evaluación de desempeño de modelo de Random Forest 50
Figura 19: Evaluación de desempeño de modelo de Redes Neuronales 51
Figura 20: Evaluación de desempeño de modelo de Support Vector Machine… 52

ÍNDICE DE TABLES

Tabla 1 : Pseudocódigo del algoritmo kDB – Naive Bayes (Sahami, 1996)..... 15
Tabla 2: Algoritmo de Random Forest (Lior, 2014) 17
Tabla 3: Algoritmo de Redes Neuronales(nnet) (Datahoek, miércoles, 30 de marzo de 2016) 20
Tabla 4: Algoritmo de entrenamiento de SVM(e1071) (Technology, 2015) 21
Tabla 5: Resumen de 5 números (estadísticas descriptivas). 26
Tabla 6: Operacionalización de variables 31
Tabla 7: Histograma de Oligonychus sp – imagen Original. 34
Tabla 8: Características estadísticas de Oligonychus sp – imagen Original. 34
Tabla 9: Histograma de Oligonychus sp - Procesado (se incrementa brillo +0.3 a la imagen original.) 35
Tabla 10: Características estadísticas de Oligonychus sp – Imagen procesado se incrementa brillo + 0.3 a la imagen original. 35
Tabla 11: Características estadísticas de Oligonychus sp – se multiplica para incrementar contraste *0.3. 36
Tabla 12: Histograma de Oligonychus sp – se multiplica para incrementar el contraste *0.3 36
Tabla 13: Histograma de Oligonychus sp - Corrección de gama ^0.3. 37
Tabla 14: Características estadísticas de Oligonychus sp – Corrección de gama ^0.3. 37
Tabla 15: Características estadísticas de la imagen 1 en R 39
Tabla 16: los datos de imagen “1” 100x100 = 10 000 datos. 40
Tabla 17: 1er cuartil de canal de color gris. 41
Tabla 18: 3er cuartil de canal de color gris. 42
Tabla 19: Evaluación de desempeño de modelos. 43
Tabla 20: Evaluación de desempeño de los modelos. 45
Tabla 21: Enfermedad de palta Oligonychus sp. 46
Tabla 22 : Promedio de las variables de imágenes enfermas 47
Tabla 23: Promedio de las variables de imágenes sanas 47
Tabla 24: Validación cruzada 48
Tabla 25: Matriz de Confusión general 48
Tabla 26: fórmulas para evaluación de desempeño de modelos 48
Tabla 27: Matriz de confusión de modelo de Naive Bayes 49
Tabla 28: Evaluación de desempeño de modelo de Naive Bayes 49
Tabla 29: Matriz de confusión de modelo de Ramdon forest 49
Tabla 30: Evaluación de desempeño de modelo de Random Forest 50
Tabla 31: Matriz de confusión de modelo de Redes Neuronales 50
Tabla 32: Evaluación de desempeño de modelo de Redes Neuronales 51
Tabla 33: Matriz de confusión de modelo de Support Vector Machine 51
Tabla 34: Evaluación de desempeño de modelo de Support Vector Machine 51
Tabla 35: Evaluación de desempeño de los modelos de clasificación. 52

ÍNDICE DE ECUACIONES

Ecuación 1: Naive Bayes 15
Ecuación 2: Sensibilidad 24
Ecuación 3: Especificidad 24
Ecuación 4: Precisión 24
Ecuación 5: Exactitud (Accuracy) 24
Ecuación 6: Valor de predicción negativa 24
Ecuación 7: Error de clasificación 24
Ecuación 8: Media 25
Ecuación 9: Percentil 25
Ecuación 10: posición Pos 26
Ecuación 11: Muestra 31
Ecuación 12: Cuartil Impar 40
Ecuación 13: Cuartil de Gris 40
Ecuación 14: 1er Cuartil de Gris 41
Ecuación 15: Mediana de Gris 41
Ecuación 16: Media de Gris 42
Ecuación 17: 3er Cuartil de Gris 42

ÍNDICE DE ANEXOS

ANEXOS 1: Preprocesamiento de imágenes 57
ANEXOS 2: Extracción de las características de las imágenes sanas y enfermas. 58
ANEXOS 3: Reconocimiento. 59
ANEXOS 4: modelos 60
ANEXOS 5: Matriz general de datos filas = 630 x columnas = 25 61
ANEXOS 6: Matriz de consistencia. 63

INTRODUCCIÓN

El cultivo de palta es un factor de gran importancia en la economía no solo en la Región de Moquegua, también Nacional e Intencional. Indicar que la producción se incrementa con el pasar de los años, con respecto a nuestro País indicar que es uno de principales productores de la palta nivel Mundial. Es evidente que el negocio de la palta está generando grandes beneficios al País. Es por ello que este fruto es muy importante para la Población, por ende, debemos implementar estrategias contra los factores negativos, una de ellas es usando herramientas tecnológicas, lo que ayudara a combatir a controlar a las enfermedades del dicho producto, una de las enfermedades más potenciales de la palta es Oligonychus sp o también es conocido como ácaros , que tienen gran potencial para dañar los cultivos de la palta, por lo mencionado se plantea el siguiente proyecto: APLICACIÓN DE ALGORITMOS INTELIGENTES PARA RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA. Brindando así la identificación de la enfermedad en su momento oportuno. Con diferentes algoritmos inteligentes de clasificación, los resultados que se pueden obtener pueden ser diferentes para un mismo problema; algunos pueden resolver muy bien un problema específico, pero tener resultados diferentes en otro. Estos resultados dependen de la complejidad del problema, del tipo de datos de entrenamiento y, sobre todo, de los fundamentos de funcionamiento de cada modelo. Por lo indicado se plantea entrenar 4 algoritmos de clasificación, como son: Naive Bayes, Random forest, Redes Neuronales y máquina de soporte vectorial; la manera en que trabajan estos algoritmos son similares a la vida cotidiana que tenemos los humanos usando la visión como una herramienta primordial para capturar un objeto. La visión nos permite identificar el objeto, su posición, su ubicación, su forma y el punto más idóneo para ser capturado, para poder realizar la evaluación de su desempeño en la clasificación de enfermedad foliar del cultivo de palta, se recolectó un total de 630 imágenes entre sanas e imágenes de hojas enfermas (Oligonychus sp). Todo este proceso se detallará en los respectivos puntos del proyecto.

Gracias al uso de estas herramientas tecnologías a nivel de la agricultura, se podrá mejorar la producción de la palta, de igual manera podremos dar a conocer que tan eficaz son estas herramientas tecnologías.

I. DATOS GENERALES

1.1 TÍTULO

“APLICACIÓN DE ALGORITMOS INTELIGENTES PARA RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA.

1.2 Tipo de Investigación

Tecnológico Aplicativo

1.3 Área de Investigación

Procesamiento de imágenes y Reconocimiento de Patrones

1.4 Nombre de Autor

G. Tomas Castro Alvarez

1.5 Localidad donde se realizó la Investigación

Provincia Mariscal Nieto - Moquegua

1.6 Asesor

Mg. Carlos Alberto Silva Delgado

II PLANTEAMIENTO DEL PROBLEMA

2.1 DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA

Descripción general de la enfermedad foliar de la palta. Las enfermedades en la palta es uno de los problemas más importantes debido al costo que significa la aplicación de métodos de control, teniendo como consecuencia pérdidas económicas en los Agricultores **(Flores).** Las enfermedades en la palta tienen como consecuencia directa en el aumento de los costos de producción, al mismo tiempo indicar que es uno de los problemas más importantes que afrontan los Agricultores **(Jaime Salvo Del P., 2017).**

El control de **Oligonychus sp(ácaros)** en el cultivo de palta ha sido un problema sanitario a nivel mundial, causa principalmente un bronceado en las hojas cuando las poblaciones del ácaro son altas, se confinan en las venas y ocasionan disminución en la producción por efecto secundario de su daño. Las poblaciones son abundantes durante periodos de altas temperatura que favorecen el aumento de las poblaciones. Es por esto que se ha generado una preocupación mundial. **(Vargas-S., 2017).** En cuanto al aspecto sanitario en el Perú existen ácaros que provocan daños económicos al cultivo. Entre ellos uno de los más resaltantes es (Oligonychus), pues puede provocar una caída importante de hojas, disminuyendo así la capacidad fotosintética y por ende el crecimiento de frutos y otros órganos de la planta **(Gamalier Lemus S., 2010).** El manejo de Oligonychus sp(ácaros) en sistemas orgánicos de producción de palta es uno de los principales desafíos encontrados por los agricultores. Es una de las principales plagas que ataca cultivos ocasionando pérdidas hasta del 80% de la producción, debido principalmente a su alto potencial reproductivo. **(G1, 2013).**

Aún existen problemas que impiden que la Región Moquegua se ubique entre las principales regiones productoras a nivel nacional, las limitantes fundamentales

que afectan a la palta son los siguientes: Poco compromiso de los productores en la implementación de nuevas tecnologías que puedan mejorar la producción del cultivo de palta, bajo nivel tecnológico en la producción de palta y la falta de investigación, desarrollo e innovación en el cultivo. **(Moquegua, Octubre - 2012) .** Existe escasa capacidad de generación y adopción de tecnologías. Por ello se hacen necesarias políticas de promoción e innovación tecnológica. Asimismo, se requiere una reforma de largo plazo en las instituciones del estado para que de ese modo se logre un cambio sostenido en el tiempo que provea de incentivos para la inversión en innovación que pueda fomentar la competitividad del sector **(Tapia2, 2011)**. Una de las debilidades es el limitado avance de investigación, innovación y tecnología en Agricultura **(Agrario, 2008)**

De acuerdo a las referencias citadas en punto de la descripción problemática como: **(Flores) , (Jaime Salvo Del P., 2017), (Vargas-S., 2017), (Moquegua, Octubre - 2012),** entre otros. Indicar lo siguiente:

Que las enfermedades foliares de la palta es un problema a nivel Local, Nacional e Internacional, tiene un gran impacto en la producción afectando directamente a la economía de los Agricultores, de igual manera mencionar que **Oligonychus sp(ácaros)** es uno de las principales enfermedades foliares de la palta. Uno de los problemas que existe especialmente en la Región de Moquegua es la falta de uso de herramientas tecnológicas, que hace que la identificación de la enfermedad no sea detectada en un momento apropiado.

2.2 FORMULACIÓN DEL PROBLEMA

2.2.1 Interrogante General

- ¿Cuál es el algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta?

2.2.2 Interrogantes Secundarias.

- ¿Cómo Construir una base de datos de imágenes de enfermedad foliar de palta?

- ¿Cuáles son los patrones más representativos al seleccionar en las imágenes?
- ¿Cómo entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento?
- ¿Qué algoritmo es más eficaz en la tarea de reconocimiento utilizando los patrones identificados?

2.3 JUSTIFICACIÓN

La Región Moquegua se ha constituido como uno de los principales productores de palta a nivel nacional, y cuenta con un enorme potencial agrícola por sus condiciones climáticas. Estas características le dan ventajas competitivas a la región para la siembra de este producto, pero que deben de ser complementadas con otras capacidades que potencien y/o desarrollen su oferta exportable siendo más competitiva y sostenible **(Moquegua, Octubre - 2012)**

Impulsando tareas de investigación y desarrollo (progreso tecnológico) en la Agricultura, las cuales a su vez dinamizan la economía Regional, Macro Regional y Nacional. **(Agrario, 2008)**. El Desarrollo de la biotecnología es otro aspecto relevante que da forma al actual panorama en el cual se desarrolla la actividad agraria mundial, está dado por el avance en materia de la biotecnología, la cual conjuga la tecnología para maximizar su impacto y aprovechamiento. Este representa un gran avance para nuestro País, por su gran biodiversidad, incremento de la productividad agropecuaria y control de enfermedades **(Elías, 2010).** La Corporación Colombiana de Investigación Agropecuaria, contribuye al bienestar de los Agricultores, mediante la generación y transferencia de tecnologías, para hacer más eficaz y rentable la producción agropecuaria con criterios de Competitividad, Equidad, Sostenibilidad y Desarrollo Científico y Tecnológico **(BERNAL E., 2008)**.

Por su parte, la tecnificación agraria mediante la provisión de servicios especializados (investigación y transferencia de tecnología), contribuye a mejorar la producción y articulación al mercado, logrando mayores tasas de rentabilidad y competitividad en la agricultura, además de propiciar un mejor uso de los recursos **(INIA, 2018)**. Mediante el uso de las herramientas tecnológicas, tiene efectos

positivos en la cadena de valor y comercialización de productos agrícolas **(Elin Baldárrago E., 2014).**

De acuerdo a las referencias citadas en punto de justificación como: **(Elin Baldárrago E., 2014), (Agrario, 2008), (Elías, 2010),** entre otros. Indicar que el **impacto** que tendrá la aplicación de herramientas tecnológicas en la Agricultura de Moquegua son los siguiente:

- Identificar en el momento apropiado las enfermedades foliares de palta.
- Incremento en la producción de palta.
- Uso necesario de recursos materiales y económicos
- Mejora la economía de los Agricultores.
- Mayores tasas de rentabilidad y competitividad en la agricultura
- Mediante el uso de las herramientas tecnológicas, tiene efectos positivos en la cadena de valor y comercialización.
- La idea general es aprovechar en su plenitud el enorme potencial agrícola por sus condiciones climáticas que posee la Región Moquegua.

2.4 FORMULACIÓN DE OBJETIVOS

2.4.1 Objetivo General

- Determinar el algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta.

2.4.2 Objetivos Específicos

- Construir una base de datos de enfermedad foliar de imágenes de palta.
- Seleccionar los patrones más representativos en las imágenes
- Entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento.
- Determinar cuál es el algoritmo más eficaz en la tarea de reconocimiento utilizando los patrones identificados

2.5 FORMULACIÓN DE HIPÓTESIS

2.5.1 Hipótesis General

- La eficacia de los algoritmos para el reconocimiento de imágenes de la enfermedad foliar de palta es estadísticamente significativos mayor al 90 % de eficacia propuesto.

2.5.2 Hipótesis Específicas

- Se construyó la base de datos de imágenes de enfermedad foliar de palta.
- Se realizó la selección de patrones más representativos en las imágenes.
- Entrenamiento de los algoritmos de clasificación son significativos en términos estadísticos.
- El algoritmo de mayor eficacia en la tarea de reconocimiento de enfermedad foliar de palta es Maquina de Soporte Vectorial.

III MARCO TEÓRICO

3.1 ANTECEDENTES DEL ESTUDIO

(Viera-Maza, 2017) El presente trabajo de investigación busca establecer los conceptos básicos para el desarrollo de la visión artificial aplicada a la agroindustria nacional. Siendo su principal objetivo el desarrollo de un sistema de visión artificial para la clasificación de granos de cacao según características externas como el tamaño en la fase final del secado.

La tesis está dividida en cuatro capítulos de desarrollo teórico y experimental, un capítulo de conclusiones y recomendaciones, las respectivas referencias, y finalmente los anexos. Los aspectos básicos de esta investigación se resumen en el primer capítulo de la tesis, con el estado del arte del sistema de visión artificial y la problemática de la producción de cacao.

En el segundo capítulo se presenta los conceptos básicos y los elementos de un sistema de visión artificial. En el tercer capítulo, se presenta una descripción del prototipo, del software y la interfaz gráfica para la visualización de los resultados.

En el cuarto capítulo se presentan los resultados del sistema implementado para la clasificación de grano de cacao orgánico según su tamaño, obteniéndose un 88.4% de acierto en la clasificación. Y finalmente en el quinto capítulo se presentan las conclusiones de la tesis.

(Fabian Edgardo Favret, 2018) El contenido de palo en la yerba mate elaborada con palo es un factor determinante de su calidad. Actualmente, la determinación del contenido del mismo está dada oficialmente por la zaranda; un método no muy preciso, dónde la molienda fina que pasa la malla de la zaranda es considerada hoja. Este trabajo propone un modelo para la determinación del nivel de contenido de palo en la yerba mate, mediante el algoritmo de clasificación Naive Bayes (NB)

y Redes Neuronales Artificiales (RNA) para el reconocimiento de patrones en imágenes digitalizadas. El prototipo incluye la preparación de las muestras, adquisición de las imágenes, extracción de las características mediante descriptores de imágenes, definición del conjunto de datos para los vectores de entrada, experimentos y análisis de los resultados obtenidos. El modelo RNA tuvo un correcto desempeño y se lo puede considerar un método confiable para la determinación del porcentaje de palos en la yerba mate.
Con respecto al clasificador Naive Bayes el mejor resultado está dado por el experimento 9, con un porcentaje de acierto del 85%. Además, se puede ver claramente en el experimento, con la combinación de la Media y la Correlación no se obtiene ningún acierto, dando el peor resultado.
Con respecto a Redes Neuronales el experimento 3 tuvo el mejor desempeño tomando como base la premisa de que el clasificador sin haber visto anteriormente el conjunto de 60 imágenes de prueba, pudo clasificar las mismas con una tasa de acierto del 96,67%.

(Cecilia Sullca, 2018) En el sector agrícola en Perú se utilizan distintos métodos para evitar que los sembríos de arándanos se vean afectados por distintas enfermedades y plagas, o detectarlas a tiempo, se ha utilizado desde drones que monitorean los campos hasta personas que se encargan de recorrer los cultivos de arándanos, pero aun así no es suficiente pues los drones solo se encargan del monitoreo de riego y no pueden acercarse demasiado a las plantas y el ojo humano solo es capaz de percibir hasta un cierto nivel, ocasionando que no detecten la enfermedad o plaga presente en la planta y estas se propaguen. La decisión de construir un sistema experto que detecte la enfermedad o plaga y tomar una acción correctiva a tiempo se tomó en base a la necesidad de este sector pues este problema afecta críticamente al rendimiento de las cosechas. Una de las áreas consideradas en este trabajo es el uso de visión artificial puesto que se utilizan técnicas para el procesamiento y análisis de imágenes, imágenes que tuvieron que ser recolectadas por los propios autores de este trabajo recorriendo sembríos de arándanos y tomando fotos a cada planta y con ello se armó una base de datos; además del uso de machine learning para la utilización de algoritmos de aprendizaje, uno de ellos fue la utilización del clasificador Support Vector Machine (SVM). Los resultados arrojaron un índice de reconocimiento

adecuado que fue capaz de clasificar si la planta de arándano estaba siendo afectada por una enfermedad o plaga.
Los resultados obtenidos luego de evaluar el modelo realizado con SVM, Random Forest y Redes neuronales y probado con los vectores característicos extraídos: con los algoritmos de extracción: entonces el mejor resultado fue un accuracy de 0.856 esto representa que nuestro modelo reconoce con un 85.6% de aceptación, lo cual demuestra que nuestro modelo entrenado con el algoritmo Redes neuronales y utilizando los vectores característicos extraídos es adecuado para predecir el estado de la planta de arándano, mientras que VSM y Random Forest tiene un acierto 82% y 62% respectivamente, sin embargo, puede mejorar si es que entrenamos nuestro modelo con más imágenes , es decir aumentemos nuestra base de datos, el proceso de este resultado fue luego de pasar las imágenes de hojas enfermas como sanas de nuestra clase prueba para entrenar, para luego pasar las imágenes de la hojas enfermas y sanas de la clase test para probar nuestro modelo, comparando así la cantidad de acertaciones que tuvo el modelo en cada hoja de arándano que entraba al algoritmo.

(**Fonseca**, **2016**) El presente review despliega una revisión bibliográfica sobre los diferentes temas que intervienen en la determinación del color rojo en imágenes digitales. Estos temas son: cómo se maneja el color en las imágenes digitales, cuáles son los modelos de color, qué son las imágenes digitales y lo referente a su formato; qué es el procesamiento de las imágenes digitales, las alteraciones que se presentan en las imágenes digitales y su corrección, así como la segmentación y la determinación de color.

(NÚÑEZ, 2018) Actualmente aún existen áreas de la Amazona donde se puede encontrar especies de plantas que necesitan ser identificadas y estudiadas para poder conocer sus propiedades médicas, nutricionales, industriales, etc. En muchos casos, los especialistas realizan un análisis manual, comando en sus habilidades sensoriales y experiencia, pero eso demanda tiempo y dinero. Por eso, es importante tener una herramienta efectiva que permita hacer un rápido y eficiente reconocimiento de las plantas.

Con este trabajo queremos dar un aporte al área de la investigación de reconocimiento y clasificación de plantas, presentando los resultados de la caracterización de plantas a través de la textura de la hoja.

El objetivo es evaluar el método Bouligand-Minkowski basado en dimensión fractal multiescalar, usando imágenes digitales para la caracterización de la textura de hojas de la Amazonia del Perú, con el fin de ayudar a mejorar su identificación y catalogación.

Para lograr el objetivo de la investigación se trabajó con dos bases de datos: ImageCLEF 2012, con 101 especies de plantas y PERALD de 27 especies. La primera es la base de datos de validación y la segunda es objeto de la investigación. El paso inicial de este trabajo fue la aplicación de un preprocesamiento de las imágenes de las plantas. Luego, las imágenes fueron divididas en cuadrados de 128 x 128 pixeles, seleccionando los 5 mejores por cada imagen.

Este paso era necesario para facilitar la caracterización de la textura. Después se aplicó el método Bouligand-Minkowski a cada muestra para obtener los descriptores de la textura de la planta. Estos descriptores fueron la entrada al clasificador Multilayer Perceptron generando así un modelo de clasificación de plantas de base de datos PERALD.

Se desarrolla un sistema que usa técnicas de visión artificial para clasificar frutos de café en ocho categorías según el estado de maduración en el que se encuentre. Las ocho categorías comprenden todo el proceso de maduración del café desde las etapas iniciales hasta sobremaduros y secos. A partir de un conjunto de 9 características que incluyen medidas de color, forma y textura obtenidas del análisis de una imagen del fruto, se logró implementar un clasificador Bayesiano cuyo desempeño, evaluado mediante el método de validación cruzada, corresponde al 96.88**%. (Zulma Sandoval, 2009)**

(Jair Cervantesa, 2017) El desarrollo de sistemas de identificación de hojas de plantas es un reto actual que comprende numerosas aplicaciones que van desde alimentación, medicina, industria y medio ambiente. En la literatura actual, se han propuesto varias técnicas con el objetivo de identificar plantas en diversos campos de aplicación. Sin embargo, las técnicas actuales están restringidas al

reconocimiento e identificación de tipos de plantas limitados, utilizando descriptores de características específicos. En este artículo, se realiza un análisis comparativo de diversos métodos de extracción de características (texturales, cromáticas y geométricas) y clasificación sobre conjuntos de plantas muy similares y disimiles entre sí. Doce conjuntos de hojas con características de forma similares son estudiados utilizando varios clasificadores. Se analiza el desempeño de diferentes combinaciones de características en cada conjunto. Los resultados obtenidos muestran que, para incrementar el desempeño de los clasificadores estudiados, es necesaria una combinación de las diferentes técnicas de extracción de características, esta necesidad es mayor cuando se trabaja con conjuntos de hojas con características muy similares. Además, se muestra el mejor desempeño de un clasificador con otro.
En los experimentos realizados, todos los conjuntos de datos fueron normalizados y se utilizó validación cruzada con k = 10 para validar resultados, esto es, se realizaron 10 pruebas con 90% y 10% de los datos para entrenamiento y prueba respectivamente (los datos de prueba son excluyentes para cada una de las pruebas), el promedio de las 10 pruebas el mejor resultado de aciertos fue de 98%.

En el presente artículo se plantea el desarrollo de dos algoritmos de procesamiento de imágenes para la identificación del café idóneo para producción, uno de acuerdo al color de este (estado de maduración) y otro para detectar la plaga de la broca. El clasificador de color consta de varias etapas: una base de conocimiento que consta de un banco de imágenes de frutos de café maduro y verde, una etapa de preprocesado para limpiar impurezas y filtrar ruido en la imagen; prosigue la segmentación para extraer el objeto de interés. Luego se extraen las características de color de la imagen y por último el proceso de reconocimiento e interpretación, el cual consta de una red neuronal artificial que clasifica los frutos en maduros o verdes. Por otra parte, el algoritmo de detección de broca fue desarrollado mediante un criterio de bancarización, esto para buscar las zonas negras en la imagen, como el orificio dejado por esta plaga sobre el fruto de café. El clasificador por redes neuronales propuesto tuvo una efectividad de 97% al detectar los estados de madurez de los frutos de café, demostrando así que las técnicas de visión artificial para el control de calidad en los frutos de café son un método viable y poco invasivo **(Jean Carlos Herrera Pérez, 2016).**

(Manuel Martínez Robles) Se propone una alternativa al proceso de segmentación tradicional en imágenes cromáticas, empleando separación de canales con umbrales obtenidos de manera experimental. Las operaciones morfológicas básicas (erosión, dilatación y cierre), también son empleadas, aunque para trabajar con la morfología matemática en imágenes cromáticas es necesario establecer un orden vectorial, el cual permite dar uniformidad a los píxeles que componen cada una de las zonas de la imagen. Realizando este proceso se logra eliminar el ruido en esas zonas y se logra una detección exacta de regiones de interés. Una vez terminado el proceso de segmentación, se analizan las regiones encontradas para establecer características que permiten construir vectores para entrenar, en una etapa posterior, una red neuronal tipo backpropagation para reconocer los patrones que determinan el grado de la patología en estudio. El mejor resultado de 4 pruebas es 99%.

3.2 BASES TEÓRICAS

3.2.1 Naive Bayes

Naive Bayes es una forma especial de red bayesiana que se usa ampliamente para la clasificación y la agrupación, pero su potencial para el modelado probabilístico general (es decir, para responder consultas conjuntas, condicionales y marginales, sobre distribuciones arbitrarias) permanece en gran medida sin explotar. Naive Bayes representa una distribución como una mezcla de componentes, donde dentro de cada componente todas las variables se suponen independientes (características) entre sí. se considera que todas las características son igualmente importantes y estadísticamente independientes. Es decir, saber el valor de una característica no da información sobre otra. Dados los componentes suficientes, puede aproximar una distribución arbitraria de forma arbitraria. debido a su simplicidad, su resistencia al ruido, poco tiempo de procesamiento y alto poder predictivo **(Daniel Lowd, 2005).**

Por ejemplo, ¿cuál es la probabilidad de que llueva si estás en Montreal? ¡¿Demasiado general?! ¡De acuerdo, pongamos algunos números en la mesa! Supongamos que la probabilidad de lluvia el sábado es de 0.05. Tuviste una maravillosa semana de trabajo soleada y hoy es sábado. ¿Cuál es la probabilidad

de lluvia hoy? Fácil, vamos a hacer los cálculos. Primero, ¿cuál es la probabilidad del sábado? Bueno, tenemos 7 días en una semana, por lo que la probabilidad del sábado es de 1/7 ≈ 0.14. Así que **(Daniel Lowd, 2005):**

Un clasificador es una función que clasifica o asigna a un objeto o individuo en uno de los grupos o clases predefinidas.

Ecuación:

$$P(Raining \mid Saturday) = \frac{P(Raining\ AND\ Saturday)}{P(Saturday)} = \frac{0.05}{0.14} = 0.36 \quad = 36\%$$

Ecuación 1: Naive Bayes

Paso 1. Para cada variable predictora Xi i = 1,....,n calcular la cantidad de información mutua con respecto a la clase C, I (Xi ,C) Paso 2. Para cada par de variables predictoras calcular la cantidad de nformación mutua condicionada a la clase, I (Xi , Xj| C), cin i ≠j, i, j = 1,...., n Paso 3. Inicializar a vació la lista de variables usando N Paso 4. Inicializar la red Bayesiana a construir, BN, con un único nodo, el correspondiente a la variable C Paso 5. Repetir hasta que N incluya a todas las variables del dominio: Paso 5.1. Seleccionar de entre las variables que no están en N, aquella Xmax con mayor cantidad de información mutua respecto a C, I(Xmax, C) = max X∉NI(X,C) Paso 5.2. Añadir un nodo a BN, X∉N representando Xmax Paso 5.3. Añadir un arco C a Xmax en BN Paso 5.4. Añadir m = mín(|N, k) arcos de las m variables distintas Xj en que tengan los mayores valores I(Xmax, Xj|C) Paso 5.5. Añadir Xmax a Paso 6. Computar las probabilidades condicionadas necesarias para especificar la red Bayesiana BN

Tabla 1 : Pseudocódigo del algoritmo kDB – Naive Bayes **(Sahami, 1996)**

3.2.2 Random forest (Bosques Aleatorios)

El Bosque Aleatorio se compone de un conjunto de árboles de decisión (que también se llaman los árboles de clasificación o árboles de regresión "CART" y solucionan las tareas homónimas). Se utilizan en la estadística, análisis de los datos y en el aprendizaje automático. Cada uno de los árboles es un modelo bastante simple que tiene sus ramas, nodos y hojas. En los nodos se escriben los atributos de cuyos valores depende la función finalista. Luego, los valores de la

función finalista llegan en las hojas por las ramas. En el proceso de la clasificación de un nuevo caso, hay que bajar por el árbol a través de las ramas hasta la hoja, pasando por todos los valores del atributo aplicando el principio lógico «si..., entonces». Dependiendo de estas condiciones, a la variable finalista se le adjudicará un valor o clase (la variable finalista entra en una hoja determinada). El objetivo de la construcción del árbol de decisiones es la creación de un modelo que predice el valor de la variable finalista dependiendo de varias variables en la entrada **(Dmitrievsky, 2016).**

Se puede decir que Random forest es un caso especial del bagging, cuando se usan los árboles de decisión como una familia básica. Además, a diferencia del modo habitual de la construcción de árboles de decisión, no se usa la reducción del árbol (pruning). El método está orientado a la posibilidad de construir cuanto antes una composición usando las muestras grandes de los datos. Cada árbol se construye de una manera específica. La característica (atributo) para construir un nodo del árbol no se selecciona del número total de características, sino de su subconjunto aleatorio. Si estamos construyendo un modelo regresivo, el número de las características será igual a n/3. En caso de la clasificación, es $\sqrt{n}$. Todo eso representa las recomendaciones empíricas y se llama de correlación: diferentes árboles reciben diferentes conjuntos de características, y los árboles se entrenan en diferentes muestras **(Dmitrievsky, 2016).**

El algoritmo del bosque aleatorio ha resultado muy eficaz, permitiendo resolver problemas prácticos. Ofrece una alta precisión del entrenamiento, con una cantidad supuestamente alta de las contingencias que figuran en el proceso de la construcción del modelo. La ventaja ante otros modelos del aprendizaje automático es la evaluación out of bag para la primera parte del subconjunto que no forma parte de la muestra de entrenamiento. Por eso, no es obligatorio hacer la validación cruzada o la simulación en una muestra separada para los árboles de decisión. Será suficiente hacer la estimación out of bag para la siguiente mejora del modelo: selección del número de los árboles de decisión y el componente de regularización. **(Dmitrievsky, 2016).**

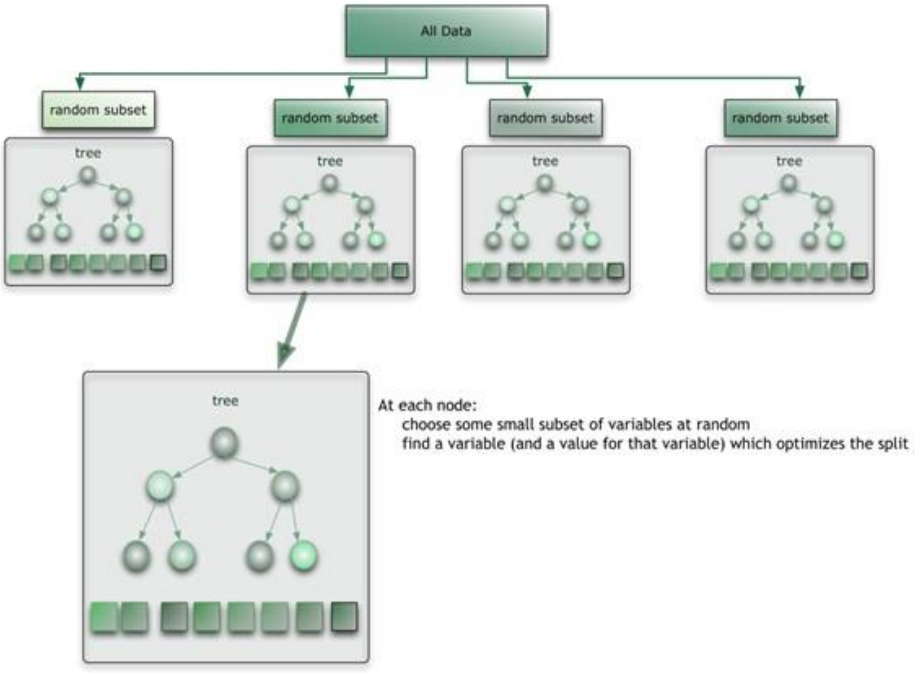


Figura 1: Esquema de Random Forest

Paso1. Muestra N casos al azar con el reemplazo para crear un subconjunto de los datos. El subconjunto debe ser aproximadamente 66% del conjunto total. Paso 2. En cada nodo: Paso 2.1. Para un número m, las variables predictoras m son seleccionados al azar entre todas las variables predictoras. Paso 2.2. La variable de predicción que proporciona la mejor división, de acuerdo con una función objetiva, se utiliza para hacer una división binaria en ese nodo. Paso 2.3. En el siguiente nodo, elije otras m variables al azar entre todas las variables predictoras y hace lo mismo. Paso 3. Encuentra la clase predicha más frecuente entre todos los nodos.

Tabla 2: Algoritmo de Random Forest **(Lior, 2014)**

3.2.3 Redes Neuronales

Definiciones de una red neuronal.

(Matich, 2001) Existen numerosas formas de definir a las redes neuronales; desde las definiciones cortas y genéricas hasta las que intentan explicar más detalladamente qué son las redes neuronales. Por ejemplo:

1) Una nueva forma de computación, inspirada en modelos biológicos.
2) Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.
3) Un sistema de computación compuesto por un gran número de elementos simples, elementos de procesos muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas.
4) Redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico **(Matich, 2001).**

Ventajas que ofrecen las redes neuronales.

Debido a su constitución y a sus fundamentos, las redes neuronales artificiales presentan un gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas. Entre las ventajas se incluyen:

- ✓ Aprendizaje Adaptativo. Capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.
- ✓ Auto-organización. Una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.
- ✓ Tolerancia a fallos. La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.
- ✓ Operación en tiempo real. Los cómputos neuronales pueden ser realizados en paralelo; para esto se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad.
- ✓ Fácil inserción dentro de la tecnología existente. Se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en

ciertas tareas. Ello facilitará la integración modular en los sistemas existentes **(Matich, 2001).**

Elementos básicos que componen una red neuronal.

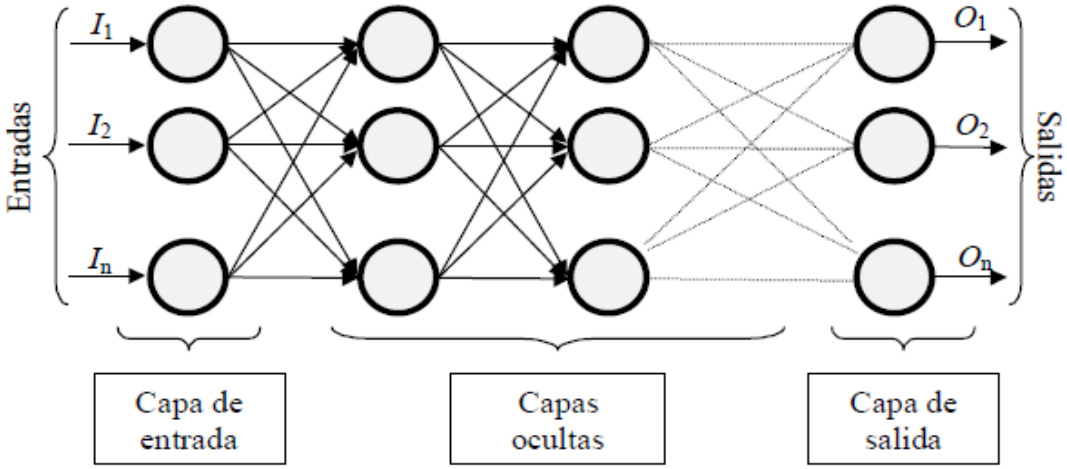


Figura 2: Esquema de una red Neuronal

La misma está constituida por neuronas interconectadas y arregladas en tres capas (esto último puede variar). Los datos ingresan por medio de la “capa de entrada”, pasan a través de la “capa oculta” y salen por la “capa de salida”. Cabe mencionar que la capa oculta puede estar constituida por varias capas **(Matich, 2001).**

se debe aprender algo sobre las neuronas y de cómo ellas son utilizadas por una red neuronal. En la **Figura N° 03** se compara una neurona biológica con una neurona artificial. En la misma se pueden observar las similitudes entre ambas (tienen entradas, utilizan pesos y generan salidas).

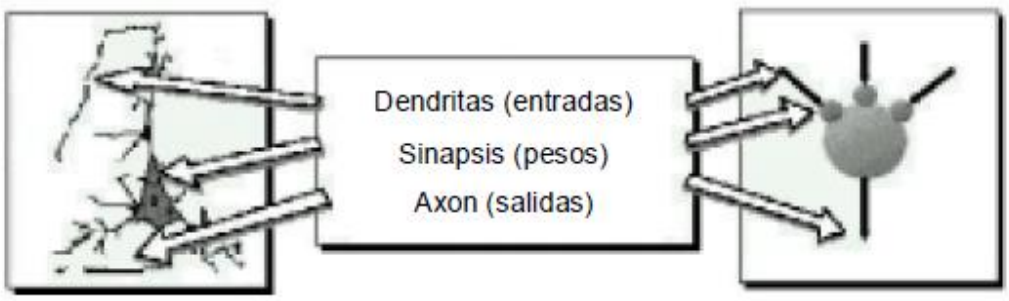


Figura 3: Comparación entre una neurona biológica (izquierda) y una artificial (derecha).

Mientras una neurona es muy pequeña en sí misma, cuando se combinan cientos, miles o millones de ellas pueden resolver problemas muy complejos. Por ejemplo, el cerebro humano se compone de billones de tales neuronas **(Matich, 2001).**

```
-Establecer los vectores de entrenamiento sm a la par que el vector
de objetivo (target) tm. m es el número de datos a ingresar.
-Establecer la velocidad de aprendizaje alfa dentro de un intervalo
(0,1).
-Establecer valor de umbral para la función de activación
-Iniciar los pesos w y la polarización (bias) b a cero, hacer wi = 0, bi
= 0, con i = 1,...,n.
-Hacer mientras contador = m:
 -Iniciar contador a cero
 -Para cada par de entrenamiento s:t realizar:
  -Colocar los valores xi = si
  -Calcular la respuesta de la unidad de salida
     y_in = b + suma(x * w)
  -Calcular la respuesta mediante la función de activación
  -Si y_in > umbral, entonces
    y = 1
  -Si y_in < umbral, entonces
    y = -1
  -En otro caso
    y = 0
  -Realizar la verificación si hubo cambio en los pesos
  -Si y es diferente de t, entonces
    w = w + alfa*t*x
    b = b + alta*t
  -En otro caso
   incrementar contador
Terminar
```

Tabla 3: Algoritmo de Redes Neuronales(nnet) **(Datahoek, miércoles, 30 de marzo de 2016)**

3.2.4 Las máquinas de soporte vectorial (SVMs)

Una Máquina de Soporte Vectorial (SVM) aprende la superficie decisión de dos clases distintas de los puntos de entrada. Como un clasificador de una sola clase, la descripción dada por los datos de los vectores de soporte es capaz de formar una frontera de decisión alrededor del dominio de los datos de aprendizaje con muy poco o ningún conocimiento de los datos fuera de esta frontera. Los datos son mapeados por medio de un kernel Gaussiano u otro tipo de kernel a un espacio de características en un espacio dimensional más alto, donde se busca la máxima separación entre clases. Esta función de frontera, cuando es traída de

regreso al espacio de entrada, puede separar los datos en todas las clases distintas, cada una formando un agrupamiento **(BETANCOURT, 2005).**

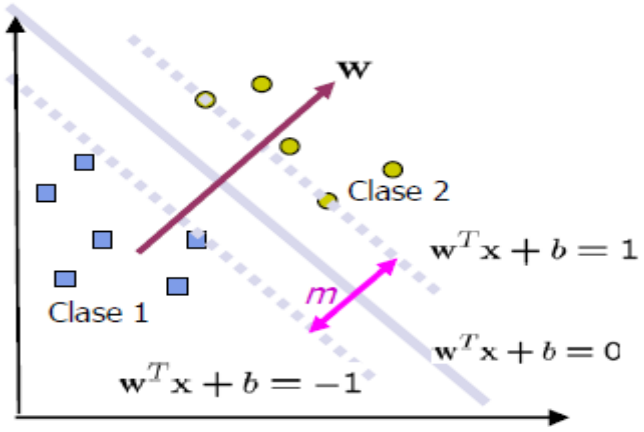


Figura 4: La frontera de decisión debe estar tan lejos de los datos de ambas clases como sea.

1. Normalize the data set
2. For Each C, γ:
   2.1. Cross Validate using leave-one-out.
      2.1.1. Train and test the SVM.
      2.1.2. Store the success rate.
   2.2. Compute the average success rate.
   2.3. Update the best C and γ if needed.
   2.4. Return to 2.1 with next C, γ.
3. Choose C, γ with best average success rate, and perform step (2) using fine scale around the selected parameters.

Tabla 4: Algoritmo de entrenamiento de SVM(e1071) **(Technology, 2015)**

3.2.5 Modo de color RGB

Las imágenes RGB utilizan tres colores para reproducir en pantalla hasta 16,7 millones de colores. RGB es el modo por defecto para las imágenes de digitales. Los monitores de ordenador muestran siempre los colores con el modelo RGB. La descripción RGB (del inglés Red, Green, Blue; "rojo, verde, azul") de un color hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores luz primarios. Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores

primarios, de manera, por ejemplo, que el valor 0 significa que no interviene en la mezcla y, a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera (valores reales entre 0 y 1, valores enteros entre 0 y 37, etc.), es frecuente que cada color primario se codifique con un byte (8 bits). Así, de manera usual, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255.
Por lo tanto, el rojo se obtiene con (255,0,0), el verde con (0,255,0) y el azul con (0,0,255), obteniendo, en cada caso un color resultante monocromático. La ausencia de color lo que nosotros conocemos como color negro se obtiene cuando las tres componentes son 0, (0,0,0) **(ERAÑA, 2009).**

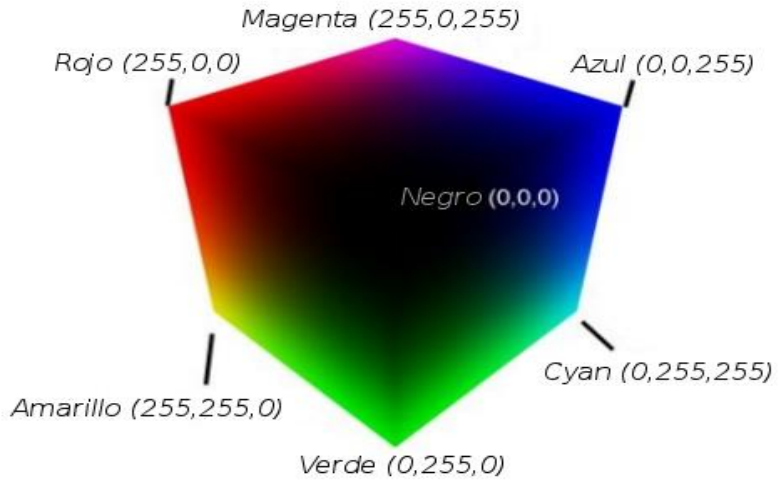


Figura 5: Distribución de colores en el espacio RGB.

Escala de gris. A partir de una escala de claridad (escala de grises) será posible clasificar cualquier muestra de color de acuerdo a su parecido con uno u otro de los grises de la escala. La claridad en cuanto atributo de un color puede entonces definirse como aquella cualidad de un color percibido que permite clasificarlo como equivalente a alguno de los grises de una escala de intervalos uniformes que vaya del negro al blanco **(ERAÑA, 2009).**

Conversión a escala de gris. La manera más simple, de convertir una imagen en color a blanco y negro, es a lo que todo programa de retoque trae de "convertir a escala de grises. La forma en que el programa convierte a grises, aunque

parezca un misterio, no lo es tanto: en realidad hace un mezclado de los tres canales (RGB) para obtener uno solo en gris tomando los siguientes porcentajes: Rojo 30%, Verde 59%, Azul 11%, que es según los expertos lo más parecido a como nuestros ojos captan la intensidad de luz dependiendo del color de ésta **(ERAÑA, 2009).**

3.2.6 Matriz de Confusión

Una matriz de confusión contiene información sobre clasificaciones reales y pronosticadas realizadas por un sistema de clasificación. El rendimiento de tales sistemas se evalúa comúnmente utilizando los datos de la matriz **(Provost, 1998).**

Métricas de precisión

Podemos hacer más para analizar los resultados que solo medir la precisión. Si consideras los posibles resultados de las clases, entonces se pueden organizar en una tabla. (donde un verdadero positivo es una observación colocada correctamente en la clase 1, mientras que un falso positivo es una observación la observación se coloca incorrectamente en la clase 1, mientras que los ejemplos negativos (tanto verdaderos como falsos) son los puestos en clase 2) **(Marsland, 2015):**

True Positives	False Positives
False Negatives	True Negatives

La precisión se define entonces como la suma del número de verdaderos positivos y verdaderos negativos divido por verdaderos positivos, falso positivo, falso negativos y verdaderos negativos **(Marsland, 2015):**

Donde:

- ✓ **TP =** True Positivo
- ✓ **FN =** False Negativo
- ✓ **FP =** Falso Positivo

- ✓ **TN =**True Negativo

$$\boldsymbol{Sensibilidad(TP)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

Ecuación 2: Sensibilidad

$$\boldsymbol{Especificidad(TN)} = \frac{\boldsymbol{TN}}{\boldsymbol{TN + FP}}$$

Ecuación 3: Especificidad

$$\boldsymbol{Precisión(FP)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$$

Ecuación 4: Precisión

$$\boldsymbol{Exactitud(Accuracy)} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + FP + TN + FN}}$$

Ecuación 5: Exactitud (Accuracy)

$$\boldsymbol{Valor\ de\ predicción\ negativa} = \frac{\boldsymbol{TN}}{\boldsymbol{TN + FN}}$$

Ecuación 6: Valor de predicción negativa

$$\boldsymbol{Error\ de\ clasificación} = \frac{\boldsymbol{TP + FN}}{\boldsymbol{FP + FN + TP + TN}}$$

Ecuación 7: Error de clasificación

3.2.7 Estadística descriptiva univariante: variables cuantitativas

Para resumir variables de tipo cuantitativo tenemos un abanico de herramientas bastante más amplio que para el caso cualitativo. Podemos tabular los datos en tablas de frecuencias, o bien calcular medidas de resumen específicas de este tipo de variables, que se pueden clasificar a grandes rasgos de la siguiente forma **(Botella Rocamora, 2013-2014):**

3.2.7.1 Medidas de centralización:

Las medidas de centralización nos informan sobre la localización alrededor de la que se encuentran los valores de la variable en estudio. Hay diferentes estadísticos que nos informan sobre este valor, entre los que destacamos **(Botella Rocamora, 2013-2014):**

Media. Supongamos que tenemos una variable cuantitativa a la que llamamos X y tenemos recogidos n valores de esta variable que denotamos con x1, x2, ..., xn. La media de estos valores se representa x y se calcula mediante la fórmula **(Botella Rocamora, 2013-2014):**

$$\text{Media} = \frac{\sum_{i=1}^{n} X1}{n} = \frac{X1+X2....+Xn}{n}$$

Ecuación 8: Media

Mediana. es el valor que cumple que la mitad de los valores de la variable son inferiores a él y la otra mitad son superiores. Si el número de datos en la muestra es impar será el valor central de la muestra ordenada (muestra en la que las unidades experimentales aparecen ordenadas según el valor que toman). Si el número de datos es par la mediana se define como la media de los dos valores centrales de la muestra ordenada **(Botella Rocamora, 2013-2014).**

3.2.7.2 Medidas de orden o posición (localización):

Estas medidas indican, como refleja su nombre, el orden o posición de una observación entre los valores de una variable cuantitativa. Para el cálculo de estas medidas debemos ordenar de forma ascendente los valores de la muestra, al resultado de dicha reubicación de los valores se le conoce como muestra ordenada **(Botella Rocamora, 2013-2014)*:***

Mínimo: El mínimo es el valor menor.
Máximo: El máximo es el valor mayor.

Percentil al p %. El percentil al p % es el valor que cumple que el p % de las observaciones de la muestra son inferiores a él (y por tanto el resto son superiores a él). Para su cálculo deberíamos hallar la posición que ocupa dicho valor en la muestra ordenada. Dicha posición la podemos calcular mediante la siguiente expresión **(Botella Rocamora, 2013-2014):**

$$pos = (n+1).\frac{p}{100}$$

Ecuación 9: Percentil

Si la posición Pos resulta un número entero, indica que el valor en esa posición de la muestra ordenada es el percentil buscado. Si, en cambio, la posición resulta un número con decimales, el percentil se calculará utilizando por un lado su parte entera ([Pos]), y por otra su parte decimal (deci (Pos)) mediante la fórmula **(Botella Rocamora, 2013-2014):**

$$deci(pos).X[\text{pos}] + 1 + \big(1 - \text{dec}(\text{pos})\big).\text{X}[\text{pos}]$$

Ecuación 10: posición Pos

Los percentiles al 25 %, 50 % y 75 % reciben nombres concretos dada su importancia (y se denotan por P25 = Q1, P50 = Q2 y P75 = Q3). A estos percentiles se les dice cuartiles (primer, segundo y tercer cuartil respectivamente) ya que dividen la muestra en cuatro partes de igual tamaño. Si pensamos con calma nos podremos dar cuenta de que ya hemos definido anteriormente al segundo cuartil ya que este estadístico no es más que el valor que es superior al 50 % de las observaciones de la variable y esa propiedad era la condición que había de cumplir necesariamente la mediana. Por tanto, al hablar del percentil al 50 %, del segundo cuartil o de la mediana de una variable nos estamos refiriendo exactamente a la misma cantidad **(Botella Rocamora, 2013-2014).**

3.2.8 Resumen de 5 números.

El resumen de 5 números en **R** existe un resumen de seis números, es decir, la tabla incluye mínimo, Q1, mediana, media, Q3 y máximo (la media no es incluida en el resumen de cinco números) **(Saavedra, 2014). (Hoaglin D., 1983)**

Summary (Variable)



Tabla 5: Resumen de 5 números (estadísticas descriptivas).

3.2.9 METODOLOGÍA

Etapas de reconocimiento de patrones:

3.2.9.1 Adquisición de imagen.

La primera etapa, dentro de un proceso de visión computacional es la etapa de adquisición. En este primer paso, se trata de conseguir que la imagen sea lo más adecuada posible para que se pueda continuar con las siguientes etapas **(Marcos, 2006).**

Una correcta adquisición de la imagen supone un paso muy importante para que el proceso de reconocimiento tenga éxito. Dentro de esta etapa existen múltiples factores que atañen directamente al proceso de captura de la imagen, formados fundamentalmente por: el sistema hardware de visión artificial (cámara, óptica, tarjeta de adquisición de imagen, ordenador y software) y el entorno y posicionamiento de los elementos (la iluminación, el fondo, posición correcta de la cámara, ruido eléctrico-óptico externo, etc.)**. (Marcos, 2006)**

3.2.9.2 Preprocesamiento de la imagen.

Toda imagen que se adquiere por medios ópticos, electroópticos o electrónicos sufre en cierta medida los efectos de la degradación que se manifiestan en forma de ruido, pérdida de definición y de fidelidad de la imagen. La degradación viene provocada por el ruido de los sensores de captura, imprecisiones en el enfoque de la cámara, movimiento de la misma o perturbaciones aleatorias, entre las que tiene relevancia el efecto de la propagación de la radiación en el medio de transmisión. Los mecanismos que tratan de contrarrestar estos efectos se incluyen dentro de la etapa de preprocesado, recibiendo el nombre de operaciones de restauración **(Marcos, 2006).**

Generalmente el preprocesado pretende reparar en la imagen los desperfectos producidos o no eliminados por el hardware: deformación de ésta, ruido introducido, poco o mucho contraste o brillo, falta de ecualización apropiada, etc. Los algoritmos de preprocesado permiten modificar la imagen para eliminar ruido, transformarla geométricamente, mejorar la intensidad o el contraste, etc.; procesos que intentan mejorar el resultado final de la imagen capturada. Por supuesto, como estos algoritmos necesitan gran cantidad de tiempo de procesado, es lógico que lo mejor es utilizar un conveniente hardware que los evite **(Marcos, 2006).**

En un proceso de visión artificial estos algoritmos tienen que ser utilizados lo menos posible, un uso excesivo de ellos repercutirá en el tiempo de proceso total e indicará que la calibración, iluminación y selección de los elementos de la etapa de adquisición no ha sido la adecuada **(Marcos, 2006).**

A parte de la degradación de la imagen, existen características de la imagen que, en muchos casos, es conveniente mejorar. Así, muchas veces, se quiere mejorar el contraste, brillo, niveles de grises, eliminar brillos, aumentar los bordes, mejorar las texturas, etc. Todas estas técnicas vienen a denominarse operaciones de mejora de la imagen **(Marcos, 2006).**

3.2.9.3 Extracción de característica.

En general, extracción de características recibe un objeto como entrada y devuelve una representación matemática del mismo en un espacio vectorial donde se conserven las similitudes inherentes a dichos objetos. Este mismo enfoque también se aplica a las imágenes. En cuanto al video, la representación se basa principalmente en extraer descriptores a los fotogramas que lo componen **(Ruber Hernández García, 2014).**

3.2.9.4 Reconocimiento.

Una vez extraído las características más representativas de la imagen, se realiza el entrenamiento de los modelos para luego predecir si la imagen es sana o enferma, el desempeño de los modelos se evalúa mediante matriz de confusión.

3.2.10 R (lenguaje de programación).

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico de distribución libre, bajo Licencia GNU, y se mantiene en un ambiente para el cómputo estadístico y gráfico. Este software corre en distintas plataformas Linux, Windows, MacOS, e incluso en PlayStation 3. El término ambiente pretende caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación gradual de herramientas muy específicas y poco flexibles, como suele ser con otro software de análisis de datos. El hecho que R sea un lenguaje y un sistema, es porque forma parte de la filosofía de creación **(Julio Sergio Santana, 27 de noviembre de 2014).**

3.3 DEFINICIÓN DE TÉRMINOS

Adopción de Tecnología. Es el resultado de la decisión de los productos agrarios de incorporar o usar una tecnología determinada en sus sistemas de producción. Las tecnologías adoptadas por los productores corresponden a aquellas provenientes de fuentes externas (Centros de Investigación y otras fuentes); en consecuencia, constituyen innovaciones tecnológicas introducidas por los productores **(MINAGRI, 2014).**

Innovación Agraria o Innovación Tecnológica Agraria. Términos sinónimos, referidos a la generación de nuevos productos y/o procesos en el agro o a la mejora significativa de los mismos en un determinado espacio de tiempo. El proceso de la innovación tecnológica agraria implica la creación, desarrollo, uso y difusión de un nuevo producto, proceso o servicio en el agro y los cambios significativos de éstos **(MINAGRI, 2014).**

Palta. (Persea americana Mill) es un cultivo tropical que puede desarrollar muy bien en costa, valles interandinos y zonas tropicales. En el Perú, su manejo puede permitir cosechas de la misma variedad en diferentes estaciones del año, lo cual representa una ventaja para la comercialización y exportación de fruta a varios países **(SENASA, 2012).**

Oligonychus sp(ácaros). Conocido por su coloración rojiza como “araña roja” o “araña café”, es un ácaro que posee 8 pares de patas, es de las plagas más comunes en cualquier plantación palta; más aún, si se descuidan las plantaciones se presentan infestaciones tan severas que pueden llegar a secar el follaje o manchar los frutos **(MINAGRI, 2014).**

IV MARCO METODOLÓGICO

4.1 LUGAR DE EJECUCIÓN

El presente proyecto se realizó en los laboratorios de la Universidad Nacional de Moquegua, en la carrera Profesional de Ingeniería de Sistemas e Informática, Departamento de Moquegua.

4.2 TIPO Y DISEÑO

La presente investigación es de:

- Tipo tecnológico aplicativo y
- Diseño experimental.

4.3 NIVEL DE INVESTIGACIÓN

Nivel de investigación es descriptivo. Porque se describirán los principales patrones identificados en las imágenes de enfermedad foliar del cultivo de palta.

4.4 OPERACIONALIZACIÓN DE VARIABLES

Operacionalización de Variables		
Independiente	**Dimensión**	**Indicadores**
Características de palta.	Características de color	Canal Gris
		Canal Rojo
		Canal Verde
		Canal Azul
	Características de estadística	Valor Mínimo
		1er cuartil
		Mediana

Operacionalización de Variables		
Independiente	**Dimensión**	**Indicadores**
		Media
		3er Cuartil
		Valor Máximo
Dependiente	**Dimensión**	**Indicadores**
Eficacia de los algoritmos.	Precisión en el reconocimiento de imágenes.	Exactitud (Accuracy)

Tabla 6: Operacionalización de variables

4.5 POBLACIÓN Y MUESTRA

4.5.1 La Población

La población está conformada por la totalidad de los cultivos de palta, en la región Moquegua.

4.5.2 Muestra

El tamaño de la muestra está calculado al 95% de confianza y un margen de error de 0.05, con el siguiente estadístico para población y varianza desconocida **(CORO, JULIO DE 2012)**:

$$\mathbf{n} = \frac{\mathbf{Z^2} * \mathbf{P} * \mathbf{Q}}{\mathbf{d}^2}$$

Ecuación 11: Muestra

Donde:

n = Tamaño de muestra

Z = Certeza estadística deseada (95% de confianza, valor estándar de 1.96)

P = Proporción de cobertura (0.5)

q = 1 – Proporción de cobertura (1 – 0.5)

d = Margen de error (0.05)

Reemplazando valores, se tiene:

n= 384.16

n= 384; En consecuencia, la muestra es de 384.

4.6 TÉCNICAS E INTRUMENTOS PARA RECOLECCIÓN DE DATOS

4.6.1 Técnica

Se realizó un ***muestreo sistemático de grilla*** sobre la ubicación geográfica de los cultivos de palta hasta recolectar un total de 630 muestras enfermas y sanas.

4.6.2 Instrumento para recolección de datos

Cámara fotográfica profesional - CANÓN EOS REBEL T5I (Full HD + HDMI) – Para recolección de datos.

4.7 DISEÑO EXPERIMENTAL Y EVALUACIÓN DE LA EFICACIA DE LOS ALGORITMOS DE CLASIFICACIÓN

4.7.1 DISEÑO EXPERIMENTAL

4.7.1.1 Metodología (Etapas para reconocimiento de patrones).

4.7.1.1.1 Adquisición de datos.

Se hizo el uso de Cámara fotográfica profesional - CANÓN EOS REBEL T5I (Full HD + HDMI) para adquisición de imágenes sanas y enfermas (Oligonychus sp.).

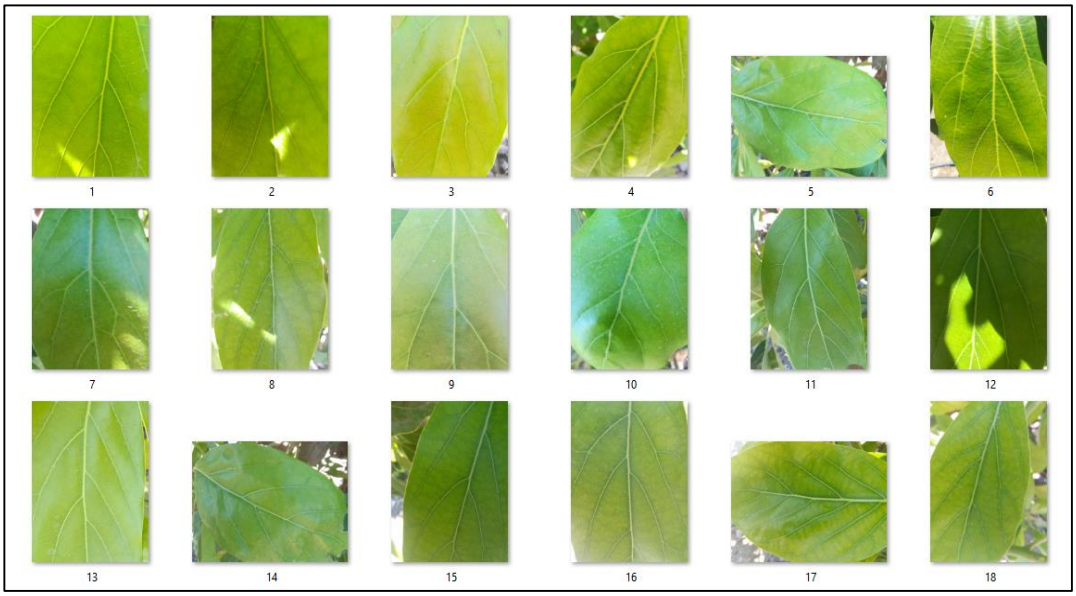


Figura 6: Se recolecto en total 330 imágenes sanas (se adjunta algunas imágenes)



Figura 7: Se recolecto en total 300 imágenes enfermas (Oligonychus sp.), se adjunta algunas imágenes.

4.7.1.2 Preprocesamiento

En este punto se consideró la siguiente imagen enferma (Oligonychus sp) como ejemplo para realizar el preprocesamiento.



Figura 8: Imagen enferma Oligonychus sp – Imagen original.

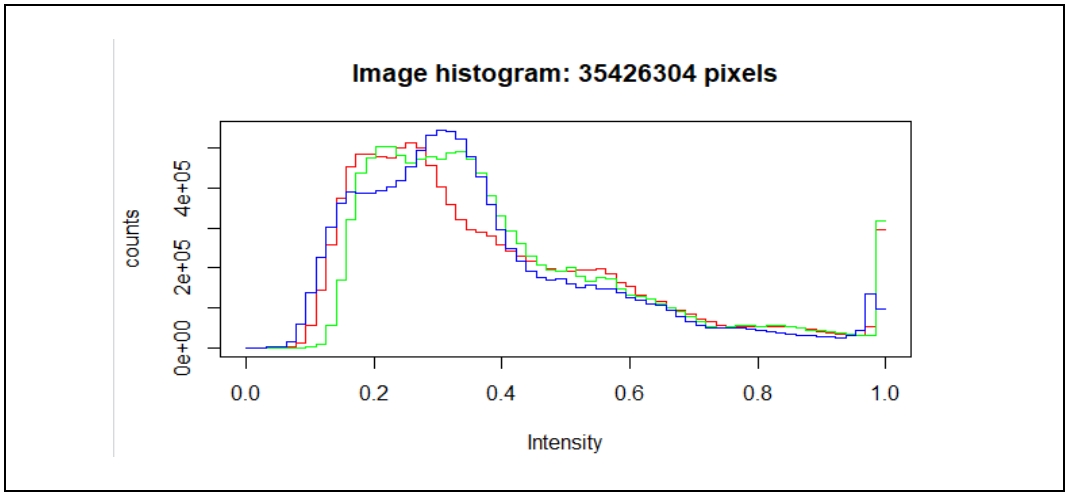


Tabla 7: Histograma de Oligonychus sp – imagen Original.

```
> summary(Image1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.2353  0.3333  0.3873  0.4941  1.0000
```

Tabla 8: Características estadísticas de Oligonychus sp – imagen Original.

Brillo: En este punto mayormente se incrementa o disminuye de acuerdo a la necesidad que tan oscuro u claro está la imagen. En reprocesamiento de imágenes para aumentar el brillo o disminuir se trabaja con la operación suma y resta respectivamente. En este caso específico se le suma un valor de 0.3 a la imagen origina.



Figura 9: Oligonychus sp – Procesado (se incrementa sumando brillo +0.3)

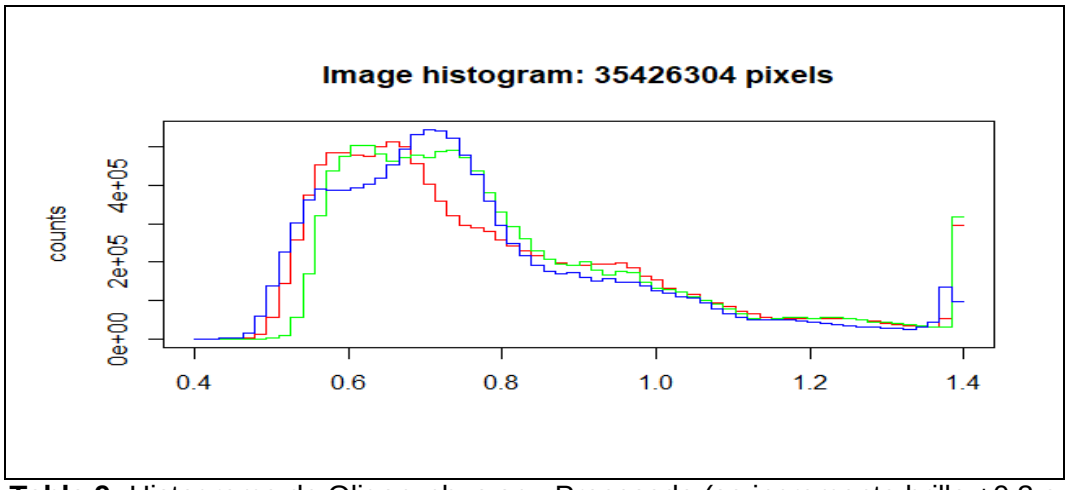


Tabla 9: Histograma de Oligonychus sp - Procesado (se incrementa brillo +0.3 a la imagen original.).

```
> summary(a)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4000  0.6353  0.7333  0.7873  0.8941  1.4000
```

Tabla 10: Características estadísticas de Oligonychus sp – Imagen procesado se incrementa brillo + 0.3 a la imagen original.

Contraste: En pre procesamiento de imágenes para aumentar el contraste o disminuir se trabaja con la operación multiplicación para incrementar y división para disminuir. En este caso específico se le multiplica un valor de 0.3 a la imagen origina. También indicar que se trabaja de acuerdo a la calidad de la imagen.



Figura 10: Oligonychus sp – se incrementa multiplicando contraste *0.3 a la imagen original.

```
> summary(e)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.1176  0.1667  0.1937  0.2471  0.5000
```

Tabla 11: Características estadísticas de Oligonychus sp – se multiplica para incrementar contraste *0.3.

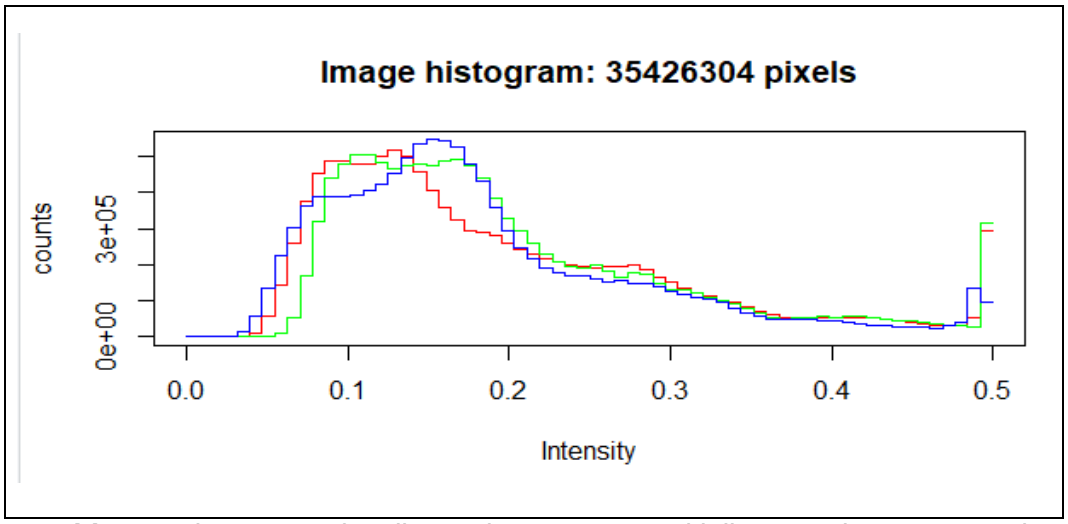


Tabla 12: Histograma de Oligonychus sp – se multiplica para incrementar el contraste *0.3

Gama: Es la combinación de brillo y contraste y por lo general se realiza para realizar el ajuste de la intensidad de imagen. Para realizar esta operación se realiza la operación raíz con el valor de ^0.3.

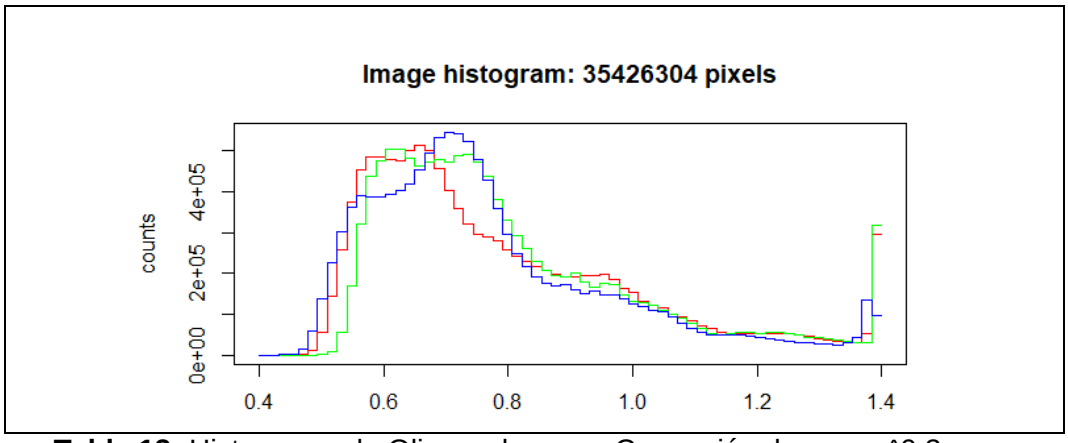


Tabla 13: Histograma de Oligonychus sp - Corrección de gama ^0.3.

```
> summary(g)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.4851  0.5774  0.6019  0.7029  1.0000
```

Tabla 14: Características estadísticas de Oligonychus sp – Corrección de gama ^0.3.



Figura 11: Oligonychus sp – corrección gama (se incrementa brillo ^0.3.)

Después de realizar el preprocesamiento se divide la imagen en subimagenes de 100x100 pixeles tanto sanas como enfermas.

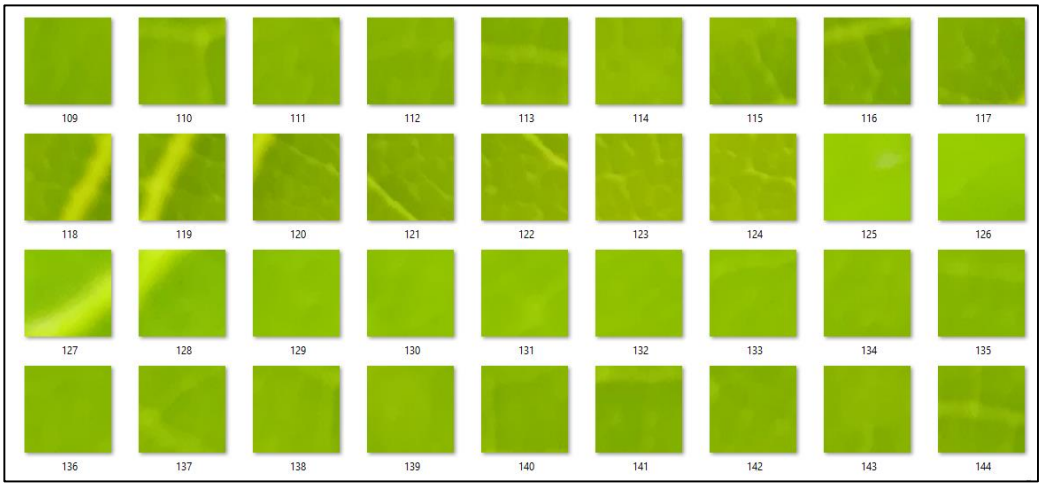


Figura 12: Imágenes sanas de 100x100 pixeles

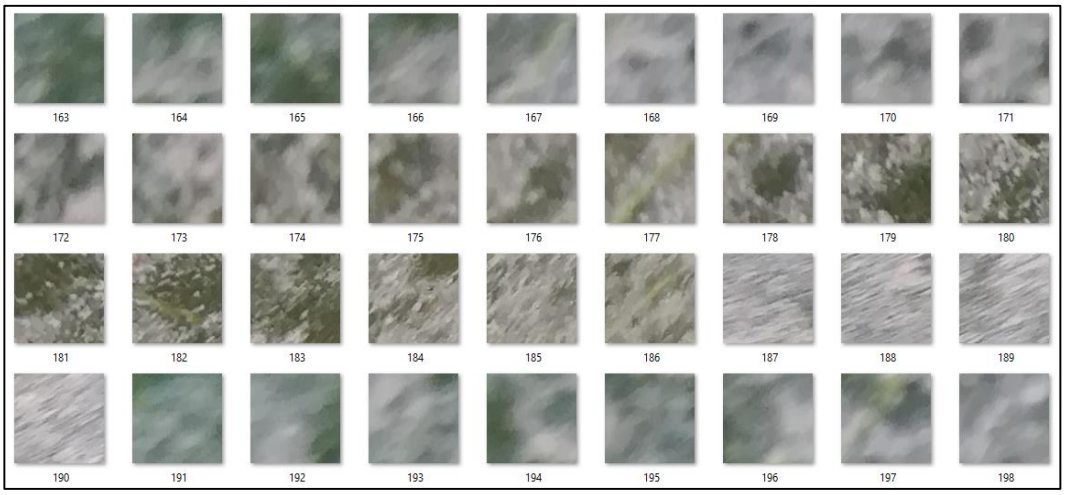


Figura 13: Imágenes enfermas de 100x100 pixeles.

4.7.1.3 Extracción de características.

De cada subimagen de 100 x 100 pixeles se extrajo los valores de los pixeles de su respectivo canal (R, G, B), rojo (R), verde (G) y azul (B). Además, se convirtió cada subimagen a una sola banda de tonos de gris y se extrajeron los respectivos valores de los pixeles una vez convertidos a gris. Con estos pasos se obtuvo entonces 10,000 valores x 4 canales (R, G, B, gris), para cada subimagen de 100 x 100 pixeles. Se calculó entonces el valor mínimo, primer cuartil, mediana, media, tercer cuartil y máximo de los valores de cada banda de cada subimagen de 100 x 100 pixeles. Por lo tanto, la información de cada subimagen se redujo a 24 estadísticos.

Este procedimiento se repitió para cada subimagen de 100 x 100 pixeles y se realizó tanto para las subimágenes generadas a partir de las hojas enfermas como para las de las hojas sanas. Los valores de los 24 estadísticos más la etiqueta según su estado, es decir 'enferma' o 'sana', para todas las subimágenes se almacenó en una matriz para el posterior entrenamiento de los modelos de clasificación.

Procedimiento para calcular las características estadísticas de los 4 canales de color RGB Y G(Gris):

Tenemos la matriz general que tomaremos de **630 x 25**.

Donde:

630 = la cantidad de imágenes sanas y enfermas (en este caso solo se visualiza sanas)

24 = las 6 características estadísticas de los canales de color (Rojo, Verde, Azul y Gris) más la etiqueta (sana o enferma) que son un total de 25.

Ver Anexo D pág. 61: *Matriz general de datos filas = 630 x columnas = 25.*

Tomaremos como ejemplo imagen sana denominada "1". En la matriz general la imagen "1" está en la columna "1".



Figura 14: Imagen sana "1"

```
> summary(vector_gray)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3464  0.3752  0.3830  0.3922  0.3908  0.5686
```

Tabla 15: Características estadísticas de la imagen 1 en R

Ahora la imagen "1" es de 100x100 pixeles para cada canal de color (Gris, Rojo Verde y Azul)

CANAL COLOR GRIS - IMAGEN (1)		
NUMERACIÓN	DATOS ORIGINALES	ORDENADO
1	0,389542484	0,346405229
2	0,389542484	0,346405229
3	0,389542484	0,346405229
4	0,389542484	0,346405229
5	0,389542484	0,346405229
9991	0,392156863	0,556862745
9992	0,392156863	0,556862745
9993	0,392156863	0,556862745
9994	0,392156863	0,558169935
9995	0,392156863	0,558169935
9996	0,392156863	0,560784314
9997	0,388235294	0,560784314
9998	0,388235294	0,564705882
9999	0,388235294	0,564705882
10000	0,388235294	0,568627451

Tabla 16: los datos de imagen "**1**" 100x100 = 10 000 datos.

Valor mínimo de Gris

Para calcular el valor mínimo lo único que se hace es ordenar de menor a mayor los datos.

Valor mínimo = 0,34640523

Cuartiles

Aclaración: para calcular los cuartiles se utilizan dos fórmulas para datos pares e impares:

Para impares:

$$Qk = \frac{k(n+1)}{4}$$

Ecuación 12: Cuartil Impar

Para pares:

$$Qk = \frac{k(n)}{4}$$

Ecuación 13: Cuartil de Gris

Donde:

K = Cuartil (1er cuartil, 2do cuartil o 3er cuartil)

N= Cantidad de números

1er Cuartil de Gris

En esta oportunidad la cantidad de datos con relación a primera imagen es 100x100 pixeles igual 10000, por lo tanto, es un numero par:

$$Qk = \frac{k(n)}{4}$$

Ecuación 14: 1er Cuartil de Gris

Donde:
K= 1er cuartil, n = 10 000, reemplazamos en la fórmula:

$Q1 = \frac{1(10000)}{4} = 2500$, primeramente, los datos se deben ordenar de menor a mayor, seguidamente ubicarnos en el dato 2500:

CANAL COLOR GRIS - IMAGEN (1)		
NUMERACIÓN	**DATOS ORIGINALES**	**DATOS ORDENADOS**
2500	0,375163399	0,375163399

Tabla 17: 1er cuartil de canal de color gris

Por lo tanto, el 1er cuartil que está en la numeración 2500 es = **0,375163399**(siempre se debe considerar los datos ordenados de menor a mayor)

Mediana de Gris

Para calcular la mediana se debe considerar lo siguiente:

Primeramente, los datos se ordenan de menor a mayor, seguidamente se considera el valor central (cuando los datos son impares el valor central siempre será único dato), mientras cuando los datos son pares (tendremos 2 datos centrales, los cuales se divide entre 2). Para este ejemplo tendremos 2 datos centrales:

$$Mediana = \frac{0{,}383006536 + 0{,}383006536}{2} = \mathbf{0,383006536}$$

Ecuación 15: Mediana de Gris

Media de Gris

La media o promedio es simplemente:

$$\boldsymbol{Media} = \frac{\sum_{i=1}^{n} X1}{n} = \frac{X1 + X2 \ldots . + Xn}{n}$$

Ecuación 16: Media de Gris

$$\boldsymbol{Media} = \frac{0{,}346405229 + \cdots + 0{,}568627451}{10000} = \mathbf{0,39223817}$$

3er Cuartil de Gris

Para calcular 3er cuartil con relación a primera imagen es 100x100 pixeles igual 10000, por lo tanto, es un numero par aplicaremos la siguiente fórmula:

$$Qk = \frac{k(n)}{4}$$

Ecuación 17: 3er Cuartil de Gris

Donde:
K= 3er cuartil, n = 10 000, reemplazamos en la fórmula:

$Q3 = \frac{3(10000)}{4} = 7500$, de igual manera los datos deben estar ordenados de menor a mayor, seguidamente ubicarnos en el dato 7500:

CANAL COLOR GRIS - IMAGEN (1)		
NUMERACIÓN	DATOS ORIGINALES	DATOS ORDENADOS
7500	0,392156863	0,390849673

Tabla 18: 3er cuartil de canal de color gris

Por lo tanto, el 3er cuartil que está en la numeración 7500 es = **0,390849673**(siempre se debe considerar los datos ordenados de menor a mayor).

Valor Máximo de Gris

Para calcular el valor máximo lo único que se hace es ordenar de menor a mayor los datos.

Valor Máximo = **0,56862745**

4.7.1.4 Reconocimiento.

En esta fase se realizó el entrenamiento de los modelos y la evaluación de la exactitud. Se utilizaron 4 algoritmos: Redes Neuronales, Support Vector Machines y Random Forest, Naive Bayes.

La matriz con los valores extraídos de las subimágenes se particionó en dos partes de manera aleatoria: 80% de las observaciones (filas de la matriz) para entrenamiento de los modelos (dataset de entrenamiento), y 20% de las observaciones para validación de los modelos (dataset de validación). El dataset de entrenamiento se usó entonces para entrenar los 4 modelos.

Una vez que se entrenó el modelo para cada clasificador, se usó el dataset de validación para evaluar la exactitud del modelo respectivo, comparando la clase de referencia de cada observación (‘enferma’ o ‘sana’) con la clase predicha por el clasificador.

Evaluación de desempeño de los modelos		
Algoritmos	**Correctos**	**Margen de Error**
Support Vector Machine	95,23810%	4,76190%
Redes Neuronales	94,44444%	5,55556%
Random Forest	91,26984%	8,73016%
Naive Bayes	86,50794%	13,49206%

Tabla 19: Evaluación de desempeño de modelos.

4.7.2 EVALUACIÓN DE LA EFICACIA DE LOS ALGORITMOS DE CLASIFICACIÓN.

4.7.2.1 Precisión (FP).

Indica la capacidad de reconocer de forma efectiva el estado de los cultivos de palta es enferma.

4.7.2.2 Valor de predicción Negativa (TN).

Indica la capacidad de reconocer de forma efectiva el estado de los cultivos de palta es sana.

4.7.2.3 Error de Clasificación.

Es el margen del error del acierto a la hora de identificar entre sana y enferma de los algoritmos.

4.7.2.4 Exactitud (Accuracy).

Es el margen de la exactitud del acierto a la hora de identificar entre sana y enferma de los algoritmos.

V PRESENTACIÓN DE RESULTADOS

A continuación, se exponen los resultados de la ejecución de la presente investigación por cada objetivo propuesto.

5.1 Resultado de objetivo general

Según el ***objetivo general*** planteado: Determinamos que el algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta.

De acuerdo a evaluación de desempeño de los modelos mediante Matriz de Confusión detallamos los resultados en la tabla **20**.

MATRIZ DE CONFUSIÓN						
Algoritmo	**Sensibilidad (TP)**	**Especificidad (TN)**	**Precisión**	**Valor de predicción Negativa**	**Error de Clasificacion**	**Exactitud (Accuracy)**
SVM	95,08%	96,92%	96,67%	95,45%	3,97%	96,03%
Random Forest	90,63%	96,77%	96,67%	90,91%	6,35%	93,65%
Naive Bayes	82,19%	100,00%	100,00%	80,30%	10,32%	89,68%
Redes Neuronales	95,00%	95,45%	95,00%	95,45%	4,76%	95,24%

Tabla 20: Evaluación de desempeño de los modelos.

El algoritmo Support Vector Machine es más eficaz con exactitud (Accuracy) de **96,03**.

5.2 Resultados de Objetivos Específicos.

Según el ***objetivo específico 1*** planteado: **Construimos una base de datos de enfermedad foliar de imágenes de palta,** para realizar este objetivo se procedió a seleccionar imágenes representativas tanto para las hojas con enfermedad y sanas:

Nombre científico	Nombre Común
Oligonychus sp.	Ácaro de palta

Tabla 21: Enfermedad de palta Oligonychus sp.

Se construyó banco de imágenes enfermas de cultivo de palta, que tiene la dimensión de 100x100 pixeles, así mismo indicar que se trabajó con un total de 300 imágenes de 100x100 pixeles.

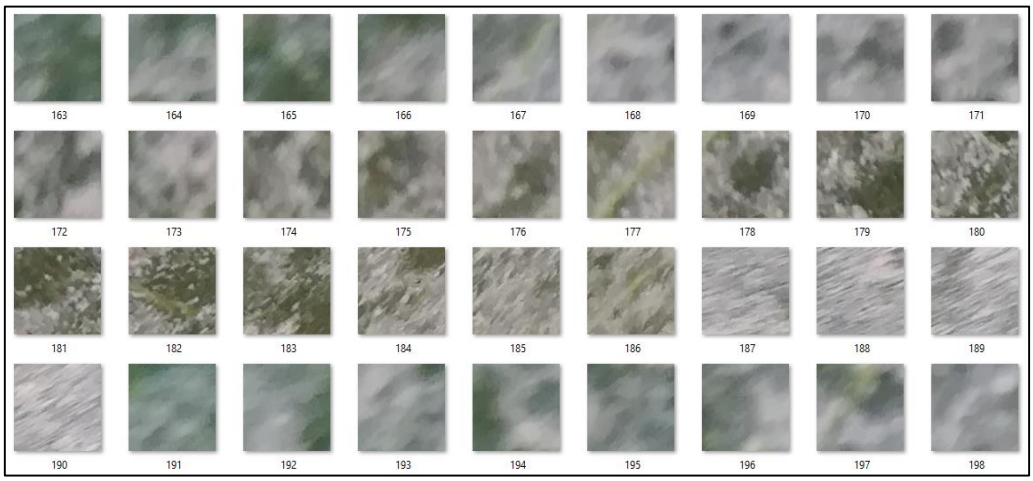


Figura 15: Imágenes enfermas de 100x100 pixeles de formato JPG.

Se construyó banco de imágenes sanas de cultivo de palta, que tiene la dimensión de 100x100 pixeles, así mismo indicar que se trabajó con un total de 330 imágenes de 100x100 pixeles.

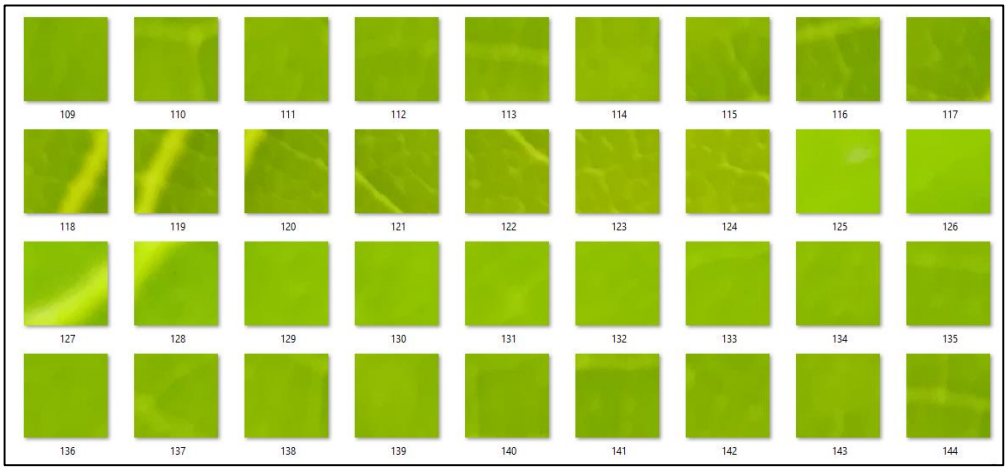


Figura 16: Imágenes sanas de 100x100 pixeles de formato JPG.

Según el ***objetivo específico 2*** planteado: **Seleccionamos los patrones más representativos en las imágenes.** Los patrones más representativos son los promedios de cada variable, tanto para sanas como enfermas.

Promedio de las variables de imágenes enfermas para los 4 canales:

PROMEDIO - ENFERMA (Oligonychus sp)					
gmin	g1q	gmedian	gmean	g3q	gmax
0,516057	0,590915	0,619318	0,622001	0,651208	0,743473
rmin	r1q	rmedian	rmean	r3q	rmax
0,563647	0,642556	0,672072	0,673480	0,703804	0,787686
grmin	gr1q	grmedian	grmean	gr3q	grmax
0,596732	0,669425	0,695026	0,696786	0,723235	0,803634
bmin	b1q	bmedian	bmean	b3q	bmax
0,367059	0,454641	0,491092	0,495736	0,532618	0,657935

Tabla 22 : Promedio de las variables de imágenes enfermas

Promedio de las variables de imágenes sanas para los 4 canales:

PROMEDIO - SANA					
gmin	g1q	gmedian	gmean	g3q	gmax
0,417643	0,449970	0,459972	0,465673	0,475659	0,558713
rmin	r1q	rmedian	rmean	r3q	rmax
0,496043	0,536910	0,550220	0,557144	0,571420	0,664753
grmin	gr1q	grmedian	grmean	gr3q	grmax
0,645514	0,685012	0,695989	0,701714	0,713039	0,797647
bmin	b1q	bmedian	bmean	b3q	bmax
0,106239	0,125930	0,133250	0,138161	0,145348	0,222721

Tabla 23: Promedio de las variables de imágenes sanas

Según el ***objetivo específico 3*** planteado: **Entrenamos los algoritmos de clasificación para realizar la tarea de reconocimiento, p**ara la etapa de

entrenamiento de los modelos se usaron los algoritmos de clasificación bajo el siguiente criterio:

Grupo	Descripción	% de la base de datos
1	Grupo de Entrenamiento	80%
2	Grupo de Evaluación	20%

Tabla 24: Validación cruzada

Se realiza la evaluación de desempeño de los modelos mediante matriz de confusión:

Donde en total se trabajaron con 126 imágenes que corresponde al 20% del total que 630 imágenes (sanas y enfermas) que corresponde al 80%.

N = 126	POSITIVO(ENFERMA)	NEGATIVO(SANA)
POSITIVO(ENFERMA)	TP	FN
NEGATIVO(SANA)	FP	TN

Tabla 25: Matriz de Confusión general

Donde:

TP = True Positivo

FP = False Negativo

FN = False Negativo

TN = True Negativo

$Sensibilidad(TP) = \frac{TP}{TP+FN}$
$Especifidad(TN) = \frac{TN}{TN+FP}$
$Precisión = \frac{TP}{TP+FP}$
$Valor\ de\ predicción\ Negativa = \frac{TN}{TN+FN}$
$Error\ de\ Clasificacion = \frac{FP+FN}{FP+FN+TP+TN}$
$Exactitud(Accuracy) = \frac{TP+TN}{TP+TN+FP+FN}$

Tabla 26: fórmulas para evaluación de desempeño de modelos

Se muestran a continuación los resultados de la evaluación de desempeños de los modelos.

Naive Bayes.

N =126	POSITIVO(ENFERMA)	NEGATIVO(SANA)
POSITIVO(ENFERMA)	60	13
NEGATIVO(SANA)	0	53

Tabla 27: Matriz de confusión de modelo de Naive Bayes

Sensibilidad (TP)	82,19%
Especificidad (TN)	100,00%
Precisión	100,00%
Valor de predicción Negativa	80,30%
Error de Clasificacion	10,32%
Exactitud (Accuracy)	89,68%

Tabla 28: Evaluación de desempeño de modelo de Naive Bayes

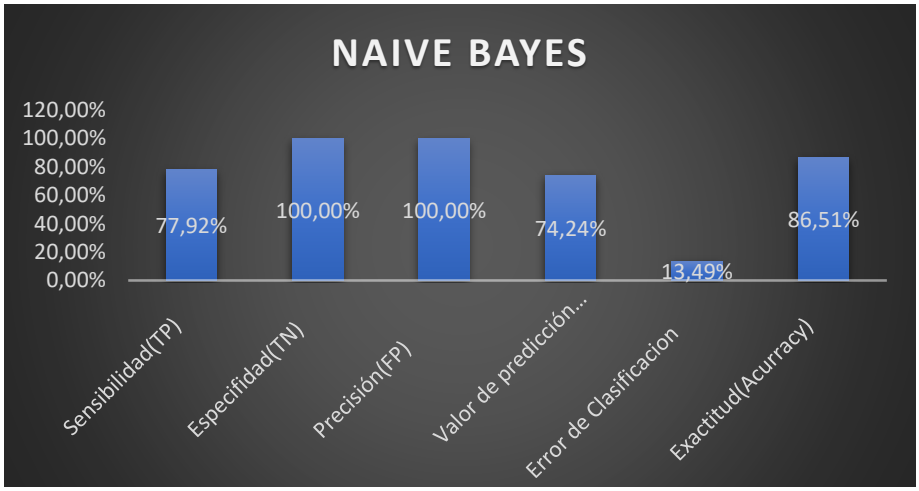


Figura 17: Evaluación de desempeño de modelo de Naive Bayes

Random forest

N=126	POSITIVO(ENFERMA)	NEGATIVO(SANA)
POSITIVO(ENFERMA)	58	6
NEGATIVO(SANA)	2	60

Tabla 29: Matriz de confusión de modelo de Ramdon forest

Sensibilidad (TP)	90,63%
Especificidad (TN)	96,77%
Precisión	96,67%
Valor de predicción Negativa	90,91%
Error de Clasificacion	6,35%
Exactitud (Accuracy)	93,65%

Tabla 30: Evaluación de desempeño de modelo de Random Forest

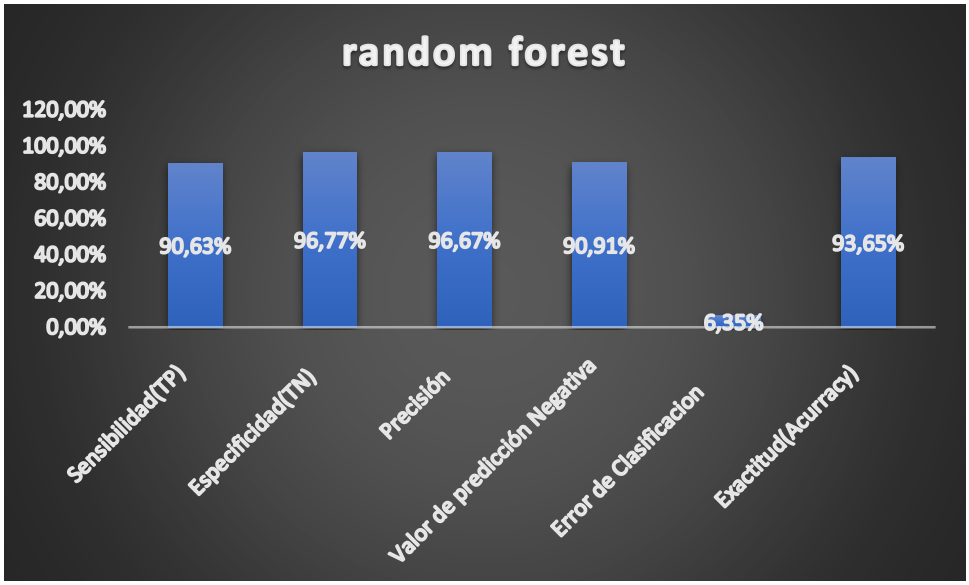


Figura 18: Evaluación de desempeño de modelo de Random Forest

Redes Neuronales

N = 126	POSITIVO(ENFERMA)	NEGATIVO(SANA)
POSITIVO(ENFERMA)	57	3
NEGATIVO(SANA)	3	63

Tabla 31: Matriz de confusión de modelo de Redes Neuronales

Sensibilidad (TP)	95,00%
Especificidad (TN)	95,45%
Precisión	95,00%
Valor de predicción Negativa	95,45%
Error de Clasificación	4,76%

Exactitud (Accuracy)	95,24%

Tabla 32: Evaluación de desempeño de modelo de Redes Neuronales

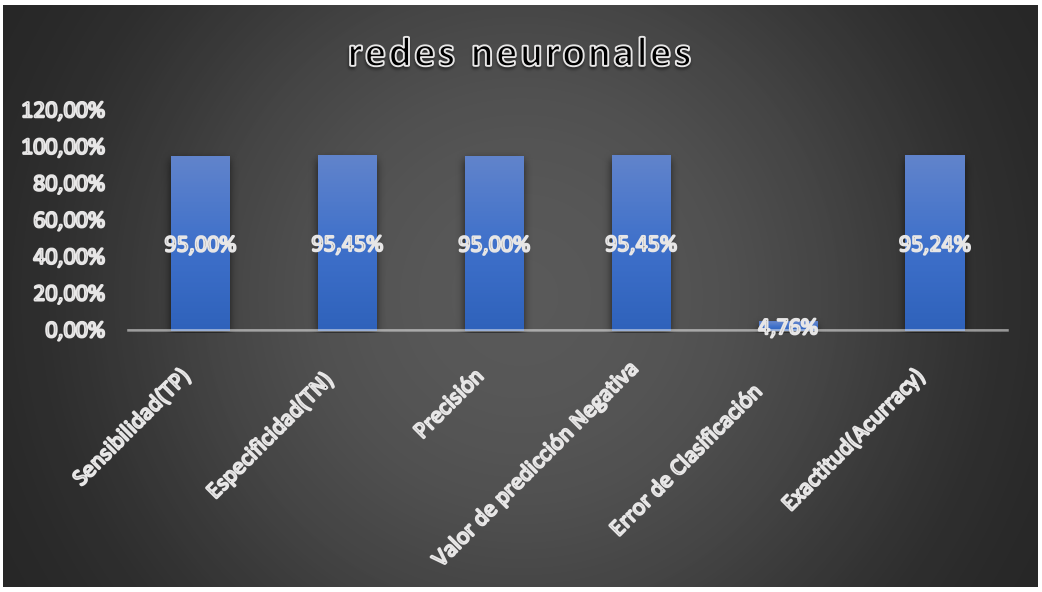


Figura 19: Evaluación de desempeño de modelo de Redes Neuronales

Support Vector Machines

N =126	POSITIVO(ENFERMA)	NEGATIVO(SANA)
POSITIVO(ENFERMA)	58	3
NEGATIVO(SANA)	2	63

Tabla 33: Matriz de confusión de modelo de Support Vector Machine

Sensibilidad (TP)	95,08%
Especificidad (TN)	96,92%
Precisión	96,67%
Valor de predicción Negativa	95,45%
Error de Clasificación	3,97%
Exactitud (Accuracy)	96,03%

Tabla 34: Evaluación de desempeño de modelo de Support Vector Machine

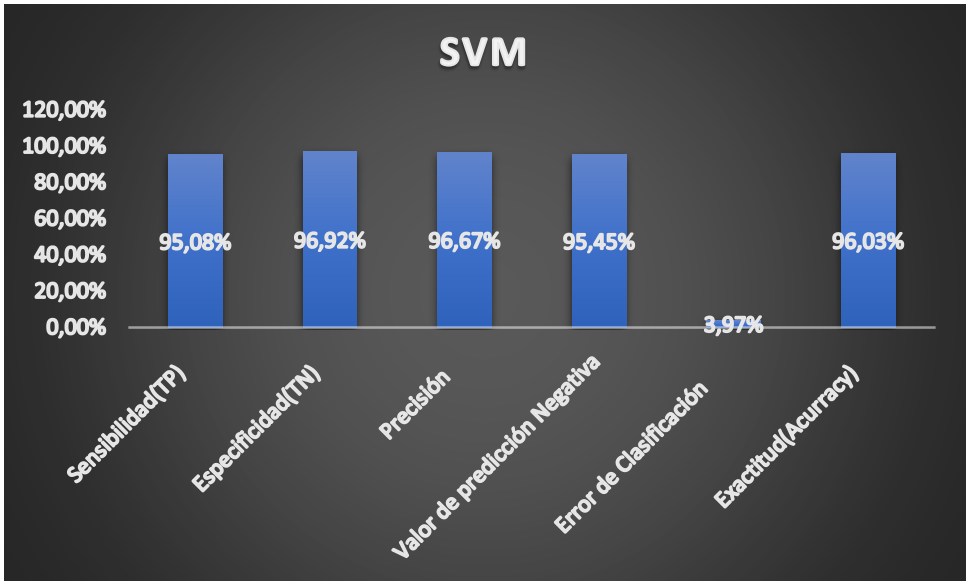


Figura 20: Evaluación de desempeño de modelo de Support Vector Machine.

Según el ***objetivo específico 4*** planteado: **Determinamos el algoritmo más eficaz en la tarea de reconocimiento utilizando los patrones identificados, r**ealizado la evaluación de desempeño de los modelos a través de matriz de confusión tenemos los siguientes resultados:

RESULTADOS DE EVALUACION DE DESEMPEÑO DE MODELOS		
Algoritmos	**Exactitud**	**Margen de Error**
Support Vector Machine	96,03175%	3,96825%
Redes Neuronales	95,23810%	4,76190%
Random Forest	93,65079%	6,34921%
Naive Bayes	89,68254%	10,31746%

Tabla 35: Evaluación de desempeño de los modelos de clasificación.

Se observa que la menor tasa de margen de error es igual a 4 % correspondiente al algoritmo de Support Vector Machine, también podemos indicar que el mencionado algoritmo proporciona resultados exactos en un 96%.

Se concluye que el algoritmo que presenta mayor eficacia en la tarea de clasificación de enfermedad del cultivo de palta mediante imágenes es el **algoritmo Support Vector Machine**.

CONCLUSIONES

Conclusión General

Concluimos que el algoritmo inteligente **Support Vector Machine** es el más eficaz en reconocimiento de imágenes de la enfermedad foliar más potencial de palta.

Conclusiones Específicas

Concluimos los siguientes:

- Se construyó un banco de imágenes compuesta con las imágenes seleccionadas para la enfermedad Oligonychus sp (ácaro de palta) conformada por 300 elementos cada una con una resolución de 100x100 pixeles en formato JPG, y compuesta por imágenes de palta sanos que está conformado por 330 elementos.
- La tarea de selección de patrones representativos más representativos son los promedios de cada variable, tanto para sanas como enfermas.
- Que para entrenamiento de los modelos se usaron los algoritmos de clasificación bajo el siguiente criterio:

    - Grupo de Entrenamiento 80%(data)
    - Grupo de Evaluación 20%(data)

- Finalmente se realizó la evaluación de desempeño de los modelos mediante matriz de confusión, teniendo como resultado Maquina Soporte Vectorial es más eficaz en reconocimiento de imágenes de la enfermedad foliar de palta.

RECOMENDACIONES

- ✓ Se recomienda realizar investigaciones que involucren la aplicación de métodos computacionales de las áreas inmersas en inteligencia artificial, tales como: machine learning, data mining, entre otros, para aplicarlos a áreas como la agricultura.
- ✓ Se recomienda trabajar con modelos de Maquina Soporte Vectorial en proyectos de procesamiento de imágenes.
- ✓ Se recomienda también abordar el tema a través de datos colectados en tiempo real y/o en registro digital de videos.
- ✓ Finalmente, se recomienda abordar nuevas alternativas para la implementación de los algoritmos aprovechando el multiprocesamiento paralelo de las computadoras modernas.

REFERENCIAS BIBLIOGRÁFICAS

Agrario, D. R. (2008). *Plan estrategico de desarollo Agrario de la Region Moquegua.* Moquegua.

BERNAL E., D. D. (2008). *Tecnología para el Cultivo del Palto.* Colombia.

BETANCOURT, G. A. (2005). *LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs).* Pereira, Colombia.

Botella Rocamora, A. G. (2013-2014). *Estadística en Ciencias de la Salud.* Valencia - España.

Cecilia Sullca, C. M. (2018). *Detección de enfermedades y plagas en las hojas de arándanos utilizando técnicas de visión artificial.* Lima – Perú.

CORO, S. A. (JULIO DE 2012). *Estadística aplicada a la Investigación (Electiva).* Venezuela.

Daniel Lowd, P. D. (2005). *Naive Bayes modelos para la estimación de probabilidad.* Washington.

Datahoek. (miércoles, 30 de marzo de 2016). *Pseudocódigo Red Neuronal Perceptrón.*

Dmitrievsky, M. (2016). *Descripción del algoritmo Random Forest.*

Elías, I. (2010). *LA ESTRATEGIA COMPETITIVA DEL SECTOR AGRARIO A TRAVÉS DE LA INNOVACIÓN Y DESARROLLO.* UPC - Lima.

Elin Baldárrago E., Y. C. (2014). *El uso de las tecnologías de información y comunicación para la mejora de la producción agrícola .* Universidad Católica San Pablo - Arequipa.

ERAÑA, I. J. (2009). *SINTESIS DIGITAL DE COLOR UTILIZANDO TONOS DE GRIS.* SAN LUIS POTOSI.

Fabian Edgardo Favret, K. E. (2018). *DETERMINACIÓN DE LOS PORCENTAJES DE PALO EN LA YERBA MATE MEDIANTE TÉCNICAS DE IA.* Brasil.

Flores, D. F. (s.f.). *Cultivo de palto - Manual practico para Productores.* Lima.

Fonseca, J. P. (2016). *Revisión sobre la detección del color rojo en imágenes digitales independiente de su luminosidad y tonalidad.* Bogota.

G1, A. S. (2013). *MANEJO ALTERNATIVO DE ÁCAROS PLAGAS.* Caldas, Colombia.

Gamalier Lemus S., R. F. (2010). *El Cultivo del Palto - INIA.* Santiago, Chile.

Hoaglin D., M. F. (1983). *Entendimiento del análisis de datos robusto y exploratorio.*

INIA, I. N. (2018). *AGENDA REGIONAL DE INNOVACIÓN AGRARIA AREQUIPA.* Arequipa.

Jaime Salvo Del P., L. L. (2017). *Manual del cultivo del palto.* Santiago, Chile,.

Jair Cervantesa, J. T.-L. (2017). *Analisis Comparativo de las tecnicas utilizadas en un Sistema de Reconocimiento de Hojas de Planta.* Mexico.

Jean Carlos Herrera Pérez, S. M. (2016). *Clasificación de los frutos de café según su estado de maduración y detección de la broca mediante técnicas de procesamiento de imágenes.* Colombia.

Julio Sergio Santana, E. M. (27 de noviembre de 2014). *El arte de programar en R: Un lenguaje para la estadística.* Mexico.

Lior, R. &. (2014). *Data mining with decision trees: theory and applications.* Singapur.

Manuel Martínez Robles, T. M. (s.f.). *SEGMENTACIÓN DE IMÁGENES, MORFOLOGÍA MATEMÁTICA Y REDES NEURONALES.* Mexico.

Marcos, A. G. (2006). *TÉCNICAS Y ALGORITMOS BÁSICOS DE VISIÓN ARTIFICIAL.* España.

Marsland, S. (2015). *MACHINE LEARNING.* New York.

Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones.* Rosario - Argentina.

MINAGRI, P. . (2014). *ACRONIMOS Y GLOSARIO DE TERMINOS.* LIma.

Moquegua, G. R. (Octubre - 2012). *PLAN OPERATIVO DE LA PALTA - PALTA.* Moquegua.

NÚÑEZ, S. K. (2018). *CARACTERIZACIÓN DE TEXTURA EN IMÁGENES DIGITALES DE HOJAS DE PLANTAS MEDIANTE DIMENSIÓN FRACTAL MULTIESCALA.* Lima - Peru.

Provost, K. y. (1998). *Machine Learning.* Boston.

Ruber Hernández García, E. G. (2014). *Modelos de representación de características para la clasificación de acciones humanas en video: estado del arte.* Malaga, España.

Saavedra, P. P. (2014). *Estadistica Aplicada I (ESMA3101).* Esma - Argentina.

Sahami, M. (1996). *Learning Limited Dependence Bayesian Classifiers.* United States.

SENASA. (2012). *Procedimiento de Prospección y Monitoreo de Plagas de Palto.* La Molina, Lima .

Tapia2, E. L. (2011). *COMPETITIVIDAD DEL SECTOR AGRARIO PERUANO, PROBLEMÁTICA Y PROPUESTAS DE SOLUCION.* Lambayeque - Peru.

Technology, I. I. (2015). *Automatic Assessment of Parkinson's Disease From Natural Hands Movements Using 3D Depth Sensor.* Convention of Electrical and Electronics Engineers in Israel.

Torres, R. R. (s.f.). *Reconocimiento de Caracteres en Imagen Digital, una Aplicación al Reconocimiento Automático de Placas de Vehículos.* Paraguay.

Vargas-S., C. A. (2017). *CONTROL BIOLÓGICO DEL ÁCARO Oligonychus SOBRE AGUACATE.* Mexico.

Viera-Maza, G. (2017). *PROCESAMIENTO DE IMÁGENES USANDO OPENCV APLICADO EN RASPBERRY PI PARA LA CLASIFICACIÓN DEL CACAO.* Piura.

Zulma Sandoval, F. P. (2009). *Procesamiento de imágenes para la clasificación de café cereza.* Bogota - Colombia.

ANEXOS

.

Primeramente, se realiza la mejora de las imágenes. El código correspondiente para esta fase se encuentra en el archivo.

Preprocesamiento.r

```
library(EBImage)
file <- dir(path = "C:/Users/TOMAS/Documents/SCRIPT/eliminaruido/", full.names = T)
Image1 <- readImage(file)

hist(Image1)

a <- Image1 + 0.3
display(a)
hist(a)

e <- Image1*0.3
display(e)
hist(e)

g <- Image1^0.3
display(g)
hist(g)

par(mfrow = c(2, 2))
plot(a)
plot(e)
plot(g)

a <- Image1 + 0.3
e <- a*0.3
g <- e^0.3
display(g)

par(mfrow = c(1,2))
plot(Image1)
plot(g)

par(mfrow = c(2,2))
plot(Image1)
plot(g)
hist(Image1)
hist(g)
```

ANEXOS 1: Preprocesamiento de imágenes

Seguidamente se hizo el corte de las imágenes para generar subimágenes con un tamaño de 100 x 100 pixeles. Esto se realizó tanto para las imágenes de plantas sanas como para las imágenes de las plantas enfermas. El código correspondiente para esta fase se encuentra en el archivo.

Dividir.r

```r
#install.packages("magick")
library(magick)


numero <- 0
for (l in 1:7) {
  numero <- numero +1

  #ruta <- paste('C:/Users/TOMAS/Documents/FOTOS ULTIMO/ENFERMAS/',l ,'.jpg', sep = "")
  #C:\Users\TOMAS\Documents\FOTOS ULTIMO
  ruta <- paste('C:/Users/TOMAS/Documents/FOTOS ULTIMO/SANAS/',l ,'.jpg', sep = "")

  palta <- image_read(ruta)
  for(i in 0:20){
    for(j in 0:30){
      x <- i*100
      y <- j*100
      corte <- paste("100x100",x, y, sep = "+")
      img1 <- image_crop(palta, corte)
      #nombre <- paste(i,j, sep = "_")
      nombre <- paste(numero,"jpg", sep = ".")
      print(nombre)
      #pat <- paste("C:/Users/TOMAS/Documents/FOTOS ULTIMO/ENFERMAS/recortadas",nombre, sep = "/")
      pat <- paste("C:/Users/TOMAS/Documents/FOTOS ULTIMO/SANAS/recortadas",nombre, sep = "/")
      image_write(img1, path = pat, format = "jpeg")
      numero <- numero +1
    }
  }
}
```

ANEXOS 2: Extracción de las características de las imágenes sanas y enfermas.

De cada subimagen de 100 x 100 pixeles se extrajeron los valores de los pixeles en cada uno de los canales de color RGB, rojo (R), verde (G) y azul (B). Además, se convirtió cada subimagen a una sola banda de tonos de gris y se extrajeron los respectivos valores de los pixeles una vez convertidos a gris. Con estos pasos se obtuvo entonces 10,000 valores x 4 canales (R, G, B, gris), para cada subimagen de 100 x 100 pixeles. Se calculó entonces el valor mínimo, primer cuartil, mediana, media, tercer cuartil y máximo de los valores de cada banda de cada subimagen de 100 x 100 pixeles. Por lo tanto, la información de cada subimagen se redujo a 24 estadísticos. El código correspondiente para esta fase se encuentra en el archivo.

datospalta_1.r

```
    imagen_gray <- channel(img, "gray")
      img_gray <- imagen_gray@.Data
      vector_gray <- as.vector(img_gray)
      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))
  }
}
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))

nombres <- c("gmin","g1q", "gmedian", "gmean", "g3q", "gmax", "rmin","r1q", "rmedian", "rmean",
             "r3q", "rmax", "grmin","gr1q", "grmedian", "grmean", "gr3q", "grmax","bmin","b1q", "bmedian", "bmean", "b3q", "bmax")
names(caracteristicas)<-nombres
caracteristicas$enfermedad <- as.factor(enfermedad)
```

ANEXOS 3: Reconocimiento.

En esta fase se realizó el entrenamiento de los modelos y la evaluación de la exactitud. Se utilizaron 4 clasificadores:

- Redes neuronales
- Support Vector Machines
- Random Forest
- Naive Bayes

El código correspondiente para esta fase se encuentra en el archivo.

modelos.r

```
library(nnet)
parametros <- train(f, data=datos.train, method="nnet", trace=F)
size <- parametros$bestTune$size
decay <- parametros$bestTune$decay

modelo.nnet <- nnet(f,trace=F, data=datos.train, size=size, decay=decay)
save(modelo.nnet, file = "modelonnet.rda")
pred.nnet <- predict(modelo.nnet, datos.test, type="class")
mc <- table(pred.nnet,datos.test[,dep], dnn = c("Asignado","Real"))
mc <- mc[order(rownames(mc)),order(colnames(mc))]
cat("** Red neuronal: nnet\n")
print(mc)
# Aciertos en %
aciertos1 <- sum(diag(mc)) / sum(mc) * 100
cat("\nCorrectamente clasificados:",round(aciertos1,2),"%\n\n\n")



library(e1071)
modelo.svm <- svm(f, data=datos.train)
save(modelo.svm, file = "modelosvm.rda")
pred <- predict(modelo.svm, datos.test, type="class")
mc <- table(pred,datos.test[,dep], dnn = c("Asignado","Real"))
mc <- mc[order(rownames(mc)),order(colnames(mc))]
cat("** SVM\n")
print(mc)
aciertos3 <- sum(diag(mc)) / sum(mc) * 100
cat("\nCorrectamente clasificados:",round(aciertos3,2),"%\n\n\n")
```

```
library(randomForest)
modelo.rf <- randomForest(f, data=datos.train)
save(modelo.rf, file = "modelorf.rda")
pred <- predict(modelo.rf, datos.test, type="class")
mc <- table(pred, datos.test[,dep], dnn = c("Asignado","Real"))
mc <- mc[order(rownames(mc)),order(colnames(mc))]
cat("Random Forest\n")
print(mc)
aciertos5 <- sum(diag(mc)) / sum(mc) * 100
cat("\nCorrectamente clasificados:",round(aciertos5,2),"%\n\n\n")

modelo.nb <- naiveBayes(f, data=datos.train)
save(modelo.nb, file = "modelonb.rda")
pred <- predict(modelo.nb, datos.test, type="class")
mc <- table(pred, datos.test[,dep], dnn = c("Asignado","Real"))
mc <- mc[order(rownames(mc)),order(colnames(mc))]
cat("** Clasificador bayesiano ingenuo: naiveBayes\n")
print(mc)
aciertos10 <- sum(diag(mc)) / sum(mc) * 100
cat("\nCorrectamente clasificados:",round(aciertos10,2),"%\n\n\n")
```

ANEXOS 4: modelos

	gmin	g1q	gmedian	gmean	g3q	gmax	rmin	r1q	rmedian	rmean	r3q	rmax	grmin	gr1q	grmedian	grmean	gr3q	grmax	bmin	b1q	bmedian	bmean	b3q	bmax	enfermedad
1	0.3464052	0.3751634	0.3830065	0.3922382	0.3908497	0.5686275	0.4470588	0.4901961	0.5058824	0.5130212	0.5176471	0.6980392	0.5882353	0.6313725	0.6392157	0.6501906	0.6549020	0.8274510	0.000000000	0.000000000	0.000000000	0.0135027451	0.007843137	0.18039216	SANA
2	0.3660131	0.3843137	0.3882353	0.3915591	0.3960784	0.4588235	0.4784314	0.5137255	0.5215686	0.5240141	0.5333333	0.6000000	0.6196078	0.6392157	0.6431373	0.6474510	0.6509804	0.7137255	0.000000000	0.000000000	0.000000000	0.0032121569	0.003921569	0.06274510	SANA
3	0.3725490	0.3960784	0.4000000	0.4057222	0.4091503	0.4745098	0.5058824	0.5372549	0.5411765	0.5465314	0.5529412	0.6156863	0.6117647	0.6509804	0.6549020	0.6608941	0.6666667	0.7294118	0.000000000	0.000000000	0.003921569	0.0097411765	0.011764706	0.12549020	SANA
4	0.3738562	0.4078431	0.4130719	0.4134146	0.4169935	0.5098039	0.5019608	0.5529412	0.5607843	0.5601941	0.5647059	0.6509804	0.6156863	0.6627451	0.6705882	0.6704890	0.6745098	0.7647059	0.000000000	0.003921569	0.007843137	0.0095607843	0.011764706	0.11372549	SANA
5	0.3830065	0.4222222	0.4287582	0.4297516	0.4366013	0.4993464	0.5137255	0.5607843	0.5686275	0.5667996	0.5725490	0.6352941	0.6352941	0.6745098	0.6823529	0.6807655	0.6862745	0.7450980	0.000000000	0.023529412	0.039215686	0.0416898039	0.058823529	0.12156863	SANA
6	0.3738562	0.4287582	0.4339869	0.4437991	0.4496732	0.6104575	0.5019608	0.5568627	0.5647059	0.5700400	0.5725490	0.7254902	0.6196078	0.6745098	0.6823529	0.6872165	0.6901961	0.8392157	0.000000000	0.047058824	0.062745098	0.0741407843	0.090196078	0.30196078	SANA
7	0.4209150	0.4522876	0.4601307	0.4604846	0.4705882	0.5111111	0.5490196	0.5725490	0.5803922	0.5788286	0.5843137	0.6156863	0.6666667	0.6862745	0.6901961	0.6915251	0.6941176	0.7254902	0.043137255	0.090196078	0.113725490	0.1111000000	0.129411765	0.19215686	SANA
8	0.4522876	0.4797386	0.4875817	0.4942833	0.5006536	0.6104575	0.5529412	0.5882353	0.5960784	0.5990922	0.6039216	0.7019608	0.6666667	0.6980392	0.7019608	0.7067180	0.7098039	0.8000000	0.121568627	0.156862745	0.168627451	0.1770396078	0.188235294	0.32941176	SANA
9	0.4418301	0.4823529	0.4954248	0.4996175	0.5137255	0.6169935	0.5176471	0.5686275	0.5843137	0.5876847	0.6039216	0.7058824	0.6352941	0.6862745	0.6980392	0.7021863	0.7137255	0.8039216	0.141176471	0.180392157	0.207843137	0.2089815686	0.235294118	0.34117647	SANA
10	0.4091503	0.4810458	0.4915033	0.5104235	0.5137255	0.7281046	0.4941176	0.5529412	0.5607843	0.5800020	0.5803922	0.7843137	0.6235294	0.6784314	0.6901961	0.7068953	0.7058824	0.9098039	0.109803922	0.207843137	0.235294118	0.2443733333	0.266666667	0.49411765	SANA
11	0.4601307	0.5032680	0.5137255	0.5345566	0.5267974	0.7215686	0.5254902	0.5607843	0.5686275	0.5906831	0.5843137	0.7764706	0.6549020	0.6941176	0.7019608	0.7208898	0.7137255	0.8980392	0.180392157	0.258823529	0.270588235	0.2920968627	0.290196078	0.49019608	SANA
12	0.4575163	0.5176471	0.5320261	0.5418163	0.5437908	0.7176471	0.5137255	0.5647059	0.5764706	0.5908851	0.5882353	0.7725490	0.6392157	0.7019608	0.7176471	0.7262353	0.7294118	0.8941176	0.219607843	0.282352941	0.298039216	0.3083286275	0.309803922	0.48627451	SANA
13	0.5150327	0.5398693	0.5464052	0.5471541	0.5516340	0.6130719	0.5647059	0.5882353	0.5960784	0.5952784	0.6000000	0.6666667	0.6980392	0.7254902	0.7333333	0.7317867	0.7372549	0.7960784	0.278431373	0.309803922	0.313725490	0.3143972549	0.317647059	0.38039216	SANA
14	0.5058824	0.5411765	0.5477124	0.5506974	0.5555556	0.6156863	0.5607843	0.5921569	0.5960784	0.6001863	0.6078431	0.6627451	0.6941176	0.7294118	0.7333333	0.7369718	0.7411765	0.8039216	0.262745098	0.301960784	0.313725490	0.3149341176	0.321568627	0.38039216	SANA
15	0.4967320	0.5437908	0.5490196	0.5519966	0.5594771	0.6169935	0.5411765	0.5882353	0.5960784	0.5973169	0.6039216	0.6588235	0.6784314	0.7294118	0.7333333	0.7361247	0.7411765	0.7960784	0.270588235	0.309803922	0.321568627	0.3225482353	0.333333333	0.39607843	SANA
16	0.5581699	0.5777778	0.5803922	0.5819678	0.5843137	0.6169935	0.6000000	0.6196078	0.6274510	0.6272298	0.6313725	0.6627451	0.7372549	0.7568627	0.7607843	0.7603125	0.7647059	0.7882353	0.337254902	0.349019608	0.356862745	0.3583611765	0.360784314	0.40000000	SANA
17	0.5660131	0.5895425	0.6000000	0.5982532	0.6052288	0.6549020	0.6000000	0.6313725	0.6392157	0.6400541	0.6470588	0.6901961	0.7333333	0.7647059	0.7686275	0.7699294	0.7764706	0.8235294	0.345098039	0.372549020	0.384313725	0.3847760784	0.396078431	0.45098039	SANA
18	0.5738562	0.5934641	0.5973856	0.5975405	0.6013072	0.6496732	0.6235294	0.6431373	0.6431373	0.6455200	0.6470588	0.7019608	0.7450980	0.7647059	0.7686275	0.7681173	0.7725490	0.8196078	0.345098039	0.372549020	0.376470588	0.3789843137	0.388235294	0.43137255	SANA
19	0.5555556	0.5908497	0.5947712	0.5941170	0.6013072	0.6248366	0.6000000	0.6431373	0.6470588	0.6451761	0.6549020	0.6784314	0.7372549	0.7686275	0.7725490	0.7712741	0.7764706	0.8000000	0.329411765	0.360784314	0.368627451	0.3659007843	0.372549020	0.39607843	SANA
20	0.5464052	0.5790850	0.5856209	0.5869115	0.5908497	0.6522876	0.5882353	0.6313725	0.6392157	0.6395212	0.6470588	0.7019608	0.7254902	0.7607843	0.7686275	0.7675937	0.7725490	0.8313725	0.317647059	0.341176471	0.352941176	0.3536196078	0.360784314	0.42352941	SANA
610	0.4836601	0.6232026	0.6653595	0.6653590	0.7124183	0.8209150	0.3764706	0.5607843	0.6117647	0.6037851	0.6549020	0.7686275	0.5176471	0.6274510	0.6627451	0.6675482	0.7098039	0.8156863	0.5529412	0.6823529	0.7215686	0.7247435	0.7725490	0.8823529	ENFERMA(Oligonychus sp)
611	0.5647059	0.5816993	0.5869281	0.5907259	0.6000000	0.6379085	0.6235294	0.6431373	0.6470588	0.6486110	0.6549020	0.6980392	0.7529412	0.7686275	0.7725490	0.7761831	0.7843137	0.8274510	0.3098039	0.3333333	0.3450980	0.3473835	0.3647059	0.4000000	ENFERMA(Oligonychus sp)
612	0.5490196	0.5843137	0.5895425	0.5923301	0.5947712	0.7215686	0.5764706	0.6196078	0.6313725	0.6330373	0.6431373	0.7529412	0.7176471	0.7529412	0.7647059	0.7648745	0.7725490	0.8901961	0.3411765	0.3647059	0.3725490	0.3790784	0.3843137	0.5215686	ENFERMA(Oligonychus sp)
613	0.5490196	0.5882353	0.5934641	0.6026263	0.6026144	0.7241830	0.5803922	0.6156863	0.6235294	0.6314655	0.6313725	0.7568627	0.7176471	0.7568627	0.7607843	0.7694341	0.7686275	0.8862745	0.3490196	0.3882353	0.4000000	0.4069792	0.4117647	0.5333333	ENFERMA(Oligonychus sp)
614	0.5477124	0.6039216	0.6143791	0.6168170	0.6339869	0.6549020	0.5686275	0.6352941	0.6431373	0.6461890	0.6588235	0.6823529	0.7019608	0.7686275	0.7764706	0.7809533	0.7921569	0.8196078	0.3607843	0.4039216	0.4235294	0.4233086	0.4431373	0.4745098	ENFERMA(Oligonychus sp)
615	0.5647059	0.6143791	0.6366013	0.6396292	0.6549020	0.8117647	0.5882353	0.6352941	0.6549020	0.6599369	0.6705882	0.8392157	0.7215686	0.7725490	0.7921569	0.7967353	0.8078431	0.9686275	0.3843137	0.4313725	0.4588235	0.4622153	0.4823529	0.6274510	ENFERMA(Oligonychus sp)
616	0.5686275	0.6196078	0.6339869	0.6443416	0.6457516	0.8026144	0.5960784	0.6431373	0.6549020	0.6662961	0.6666667	0.8235294	0.7215686	0.7764706	0.7882353	0.8003275	0.8039216	0.9490196	0.3843137	0.4392157	0.4549020	0.4664012	0.4745098	0.6392157	ENFERMA(Oligonychus sp)
617	0.5712418	0.6235294	0.6339869	0.6523882	0.6588235	0.8235294	0.6000000	0.6431373	0.6509804	0.6727451	0.6745098	0.8509804	0.7294118	0.7803922	0.7882353	0.8075835	0.8117647	0.9803922	0.3764706	0.4392157	0.4627451	0.4768361	0.4980392	0.6509804	ENFERMA(Oligonychus sp)
618	0.5607843	0.6143791	0.6235294	0.6379650	0.6457516	0.8143791	0.5921569	0.6392157	0.6470588	0.6635702	0.6705882	0.8470588	0.7294118	0.7764706	0.7882353	0.8010906	0.8078431	0.9764706	0.3607843	0.4235294	0.4392157	0.4492341	0.4627451	0.6235294	ENFERMA(Oligonychus sp)
619	0.5607843	0.6326797	0.6418301	0.6465505	0.6627451	0.7647059	0.5843137	0.6470588	0.6549020	0.6629443	0.6705882	0.7882353	0.7254902	0.7882353	0.7960784	0.8047192	0.8156863	0.9215686	0.3607843	0.4549020	0.4745098	0.4719878	0.4980392	0.5843137	ENFERMA(Oligonychus sp)
620	0.5542484	0.5830065	0.5947712	0.6304731	0.6496732	0.7973856	0.5843137	0.6078431	0.6274510	0.6637031	0.6862745	0.8392157	0.7137255	0.7411765	0.7568627	0.7880259	0.8039216	0.9529412	0.3647059	0.3921569	0.4078431	0.4396902	0.4666667	0.6039216	ENFERMA(Oligonychus sp)
621	0.4666667	0.5176471	0.6078431	0.6137135	0.7150327	0.7385621	0.4980392	0.5490196	0.6549020	0.6585000	0.7725490	0.7960784	0.6588235	0.7098039	0.7921569	0.7937588	0.8823529	0.9058824	0.2431373	0.2941176	0.3803922	0.3888816	0.4901961	0.5137255	ENFERMA(Oligonychus sp)
622	0.4901961	0.5490196	0.7006536	0.6463094	0.7124183	0.7333333	0.5215686	0.5960784	0.7647059	0.7024675	0.7764706	0.7960784	0.6823529	0.7372549	0.8784314	0.8268039	0.8862745	0.9098039	0.2549020	0.3137255	0.4588235	0.4096569	0.4784314	0.4941176	ENFERMA(Oligonychus sp)
623	0.4993464	0.6013072	0.6862745	0.6598541	0.7176471	0.7490196	0.5372549	0.6352941	0.7450980	0.7092357	0.7843137	0.8039216	0.6705882	0.7764706	0.8666667	0.8316965	0.8901961	0.9137255	0.2901961	0.3843137	0.4470588	0.4386302	0.4823529	0.5450980	ENFERMA(Oligonychus sp)
624	0.4888889	0.5594771	0.6470588	0.6458839	0.7346405	0.7712418	0.5098039	0.5803922	0.6823529	0.6790647	0.7843137	0.8156863	0.6666667	0.7333333	0.8196078	0.8128580	0.8901961	0.9294118	0.2823529	0.3725490	0.4470588	0.4457290	0.5294118	0.5686275	ENFERMA(Oligonychus sp)
625	0.5202614	0.5738562	0.6431373	0.6541558	0.7411765	0.8065359	0.5411765	0.5960784	0.6745098	0.6867835	0.7882353	0.8509804	0.7019608	0.7529412	0.8117647	0.8217282	0.8980392	0.9647059	0.3137255	0.3686275	0.4666667	0.4539557	0.5333333	0.6039216	ENFERMA(Oligonychus sp)
626	0.5267974	0.6287582	0.6901961	0.6960878	0.7712418	0.7973856	0.5529412	0.6392157	0.7294118	0.7270071	0.8196078	0.8431373	0.6901961	0.7843137	0.8509804	0.8522447	0.9254902	0.9529412	0.3254902	0.4745098	0.5019608	0.5090118	0.5686275	0.6000000	ENFERMA(Oligonychus sp)
627	0.5529412	0.6470588	0.7267974	0.7070094	0.7725490	0.8326797	0.5921569	0.6745098	0.7686275	0.7479380	0.8235294	0.8823529	0.7294118	0.8117647	0.9019608	0.8705294	0.9294118	0.9921569	0.3372549	0.4509804	0.5098039	0.5025608	0.5686275	0.6274510	ENFERMA(Oligonychus sp)
628	0.5241830	0.5934641	0.6575163	0.6730085	0.7686275	0.8287582	0.5333333	0.6117647	0.6862745	0.7029027	0.8156863	0.8627451	0.6823529	0.7529412	0.8196078	0.8296404	0.9215686	0.9764706	0.3254902	0.4235294	0.4705882	0.4864824	0.5686275	0.6470588	ENFERMA(Oligonychus sp)
629	0.5568627	0.5986928	0.6313725	0.6718292	0.7712418	0.8039216	0.5529412	0.6000000	0.6627451	0.6971176	0.8156863	0.8509804	0.7137255	0.7568627	0.8000000	0.8308153	0.9254902	0.9607843	0.3411765	0.4313725	0.4588235	0.4875545	0.5686275	0.6078431	ENFERMA(Oligonychus sp)
630	0.5673203	0.6274510	0.6470588	0.6895302	0.7830065	0.8326797	0.5686275	0.6196078	0.6588235	0.7081824	0.8313725	0.8784314	0.7176471	0.7725490	0.8000000	0.8406286	0.9333333	0.9803922	0.3529412	0.4823529	0.4980392	0.5197796	0.5843137	0.6392157	ENFERMA(Oligonychus sp)

ANEXOS 5: Matriz general de datos filas = 630 x columnas = 25

<table>
<tr><th>PROBLEMAS</th><th>OBJETIVOS</th><th>HIPÓTESIS</th><th>OPERACIONALIZACIÓN DE VARIABLES</th></tr>
<tr>
<td>PROBLEMA GENERAL:

¿Cuál es el algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta?

PROBLEMAS ESPECÍFICOS:

1) ¿Cómo Construir una base de datos de enfermedad foliar mediante imágenes?
2) ¿Cuáles son los patrones más representativos al seleccionar en las imágenes?
3) ¿Cómo entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento?
4) ¿Qué algoritmo es más eficaz en la tarea de reconocimiento utilizando los patrones identificados?</td>
<td>OBJETIVO GENERAL:

Determinar el algoritmo inteligente más eficaz para el reconocimiento de imágenes de la enfermedad foliar de palta.
OBJETIVOS ESPECÍFICOS:

1) Construir una base de datos de imágenes de enfermedad foliar de palta.
2) Seleccionar los patrones más representativos en las imágenes
3) Entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento.
4) Determinar cuál es el algoritmo más eficaz en la tarea de reconocimiento utilizando los patrones identificados</td>
<td>HIPÓTESIS GENERAL:

La eficacia de los algoritmos inteligente para el reconocimiento de imágenes de la enfermedad foliar más potencial de palta es estadísticamente significativos mayor al 90 % de eficacia propuesto.

HIPÓTESIS ESPECÍFICAS:

1) Se construyó la base de datos de imágenes de enfermedad foliar de palta.
2) Se realizó la selección de patrones más representativos en las imágenes son significativos en términos estadísticos.
3) Entrenamiento de los algoritmos de clasificación son significativos en términos estadísticos.
4) El algoritmo de mayor eficacia en la tarea de reconocimiento de enfermedad foliar de palta es Support Vector Machine.</td>
<td><u>VARIABLE INDEPENDIENTE</u>
○ <u>Características de palta.</u>

▪ DIMENSIONES
○ Características de color
❖ INDICADOR
Canal Gris
Canal Rojo
Canal Verde
Canal Azul
○ Características de estadística
❖ INDICADOR
Min (Mínimo)
1st Qu (primer cuartil)
Mediana
Media
3er Qu (Tercer Cuartil)
Max(Máximo)

<u>VARIABLE DEPENDIENTE</u>
○ <u>Eficacia de los algoritmos.</u>
DIMENSIÓN
Precisión en el reconocimiento de imágenes.
Indicador
Exactitud (Accuracy)</td>
</tr>
<tr><th>TIPO Y DISEÑO</th><th>POBLACIÓN Y MUESTRA</th><th colspan="2">TÉCNICAS E INSTRUMENTOS</th></tr>
<tr>
<td>TIPO:
Tecnológico aplicativo

Diseño:
Experimental</td>
<td>POBLACIÓN:
La población está conformada por la totalidad de los cultivos de palta, en la región Moquegua.</td>
<td colspan="2">TÉCNICAS:
muestreo sistemático de grilla

INSTRUMENTO:
Instrumento para recolección de datos:
○ Cámara fotográfica profesional CANÓN EOS REBEL T5I (Full HD + HDMI)</td>
</tr>
</table>

	MUESTRA: el tamaño de la muestra está calculado al 95% de confianza y un margen de error de 0.5	

ANEXOS 6: Matriz de consistencia.