



UNIVERSIDAD NACIONAL DE MOQUEGUA
COMISIÓN ORGANIZADORA

RESOLUCIÓN DE COMISIÓN ORGANIZADORA N° 430-2017-UNAM

Moquegua, 06 de Setiembre de 2017

VISTOS, el Informe N° 204-2017-EPISI/UNAM-FILIAL ILO de 05 de Setiembre de 2017, Oficio N° 323-2017-VIPAC-CO/UNAM de 06 de Setiembre de 2017, Acuerdo de Sesión Extraordinaria de Comisión Organizadora de fecha 06 de Setiembre de 2017, y;

CONSIDERANDO:

Que, el párrafo cuarto del artículo 18° de la Constitución Política del Estado, concordante con el artículo 8° de la Ley N° 30220, Ley Universitaria, reconoce la autonomía universitaria, en el marco normativo, de gobierno, académico, administrativo y económico, que guarda concordancia con los artículos 6°, 7°, 8°, 9° y 10° del Estatuto Universitario;

Que, el Reglamento de Grados y Títulos de la Universidad Nacional de Moquegua, aprobado con Resolución de Comisión Organizadora N° 190-2016-UNAM de 05 de Agosto de 2016, establece en el Artículo 12°, que el proyecto de tesis es un trabajo de investigación individual que presentan los estudiantes del último año académico, egresados o bachilleres al Director de la Escuela Profesional, con la finalidad de resolver un problema objeto de estudio, asimismo, precisa en el Artículo 15° que todo proyecto de tesis debe tener un asesor, quien deberá ser docente ordinario de la Escuela Profesional o en forma facultativa un docente contratado en la especialidad en el área que se investiga. El jurado dictaminador del proyecto, será designado por el Comité Asesor y el Director de la Escuela Profesional, al mismo que estará compuesto por tres miembros elegidos entre los docentes ordinarios y/o contratados, conforme se indica en los artículos 18, 19° y 20° del precitado Reglamento;

Que, mediante Informe N° 204-2017-EFISI/UNAM-FILIAL ILO de 05 de Setiembre de 2017, el Mg. Ing. Carlos Alberto Silva Delgado, Director de la Escuela Profesional de Ingeniería de Sistemas e Informática, solicita a Vicepresidencia Académica la aprobación del proyecto de tesis denominado: "Aplicación de algoritmos inteligentes para: Reconocimiento automático de enfermedades foliares de cultivo de palta", presentado por el Bachiller Gregorio Tomás Castro Alvarez, el mismo que fue declarado apto según acta de aprobación de proyecto de tesis para optar el título profesional de Ingeniero de Sistemas e Informática de 15 de Agosto de 2017, solicitando se emita el acto resolutivo;

Que, con Oficio N° 323-2017-VIPAC-CO/UNAM de 06 de Setiembre de 2017, la Dra. María Elena Echevarría Jaime Vicepresidencia Académica de la Universidad Nacional de Moquegua, solicita al Dr. Washington Zeballos Gámez Presidente de la Comisión Organizadora – UNAM, la emisión de acto resolutivo de reconocimiento de aprobación de proyecto de tesis, así como la designación de asesor y miembros del jurado dictaminador, conforme se precisa en el Reglamento de Grados y Títulos de la Universidad Nacional de Moquegua;

Que, en Sesión Extraordinaria de Comisión Organizadora de fecha 06 de Setiembre de 2017, se acordó por UNANIMIDAD, Aprobar el Proyecto de Tesis en referencia presentado por el Bachiller Gregorio Tomás Castro Álvarez, asimismo se acordó designar como Asesor de Tesis al Mg. Ing. Carlos Alberto Silva Delgado y a los miembros del jurado dictaminador de la Escuela Profesional de Ingeniería de Sistemas e Informática de la UNAM, encargados de evaluar el trabajo de investigación, conforme a la propuesta remitida;

Por las consideraciones precedentes y en uso de las atribuciones que le concede la Ley Universitaria N° 30220, el Estatuto de la Universidad Nacional de Moquegua y lo acordado en Sesión Extraordinaria de Comisión Organizadora de fecha 06 de Setiembre de 2017;

SE RESUELVE:

ARTÍCULO PRIMERO.- APROBAR, el Proyecto de Tesis denominado: "APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA", presentado por el Bachiller GREGORIO TOMÁS CASTRO ÁLVAREZ, conforme a lo expuesto a la parte considerativa de la presente resolución.

ARTÍCULO SEGUNDO.- DESIGNAR, al MG. ING. CARLOS ALBERTO SILVA DELGADO como Asesor del proyecto de tesis aprobado en el artículo primero de la presente resolución.





UNIVERSIDAD NACIONAL DE MOQUEGUA
COMISIÓN ORGANIZADORA

RESOLUCIÓN DE COMISIÓN ORGANIZADORA N° 430-2017-UNAM

ARTÍCULO TERCERO.- DESIGNAR, al Jurado Dictaminador del Proyecto de Tesis: "APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA", presentado por el Bachiller GREGORIO TOMÁS CASTRO ÁLVAREZ, conforme al siguiente detalle:

- M.Sc. HUGO EULER TITO CHURA : PRESIDENTE
- Mg. ANÍBAL FERNANDO FLORES GARCÍA : PRIMER MIEMBRO
- Mgr. ALEX PETER ZUÑIGA INCALLA : SEGUNDO MIEMBRO

ARTÍCULO CUARTO.- ENCARGAR, a los profesionales designados el cumplimiento de lo establecido en el Reglamento de Grados y Títulos de la Universidad Nacional de Moquegua, asimismo, Vicepresidencia Académica deberá adoptar las acciones académicas necesarias, para el cumplimiento de la presente resolución.

Regístrese, Comuníquese, Publíquese y Archívese.




DR. WASHINGTON ZEBALLOS GÁMEZ
PRESIDENTE

Presidencia
VIPAC
VIPI
EPISI
Interesado
Arch. (2)




ABOG. GUILLERMO S. KUONG CORNEJO
SECRETARIO GENERAL

INFORME N°204-2017-EPISI/UNAM-FILIAL ILO



A : DRA. MARÍA ELENA ECHEVARRÍA JAIME
Vicepresidenta Académica - UNAM

DE : MG. CARLOS ALBERTO SILVA DELGADO
Director de la Escuela Profesional de Ingeniería de Sistemas e Informática
UNAM – Filial Ilo

ASUNTO : SOLICITO APROBACIÓN DE PROYECTO DE TESIS MEDIANTE ACTO RESOLUTIVO

REFERENCIA : INFORME N° 005-2017-HETCH/UNAM-FILIAL ILO
ACTA DE CONFORMIDAD DE PROYECTO DE TESIS

FECHA : Ilo, 05 de Setiembre del 2017.

Mediante el presente me dirijo a usted, para expresarle un cordial saludo y en virtud al documento de la referencia presentado por el Jurado Revisor de Tesis del candidato a Título Profesional Sr. **GREGORIO TOMÁS CASTRO ALVAREZ** Bachiller de la Escuela Profesional de Ingeniería de Sistemas e Informática, donde se aprueba por UNANIMIDAD el Proyecto de Tesis titulado **"APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA"**, proyecto que deberá ser ejecutado en un plazo máximo de dos años conforme indica el Reglamento de Grados y Títulos.

Los miembros del **JURADO REVISOR DE TESIS**, están integrados de acuerdo al siguiente detalle:

MSC. HUGO EULER TITO CHURA	PRESIDENTE
MG. ANIBAL FERNANDO FLORES GARCÍA	PRIMIER MIEMBRO
MSC. ALEX PETER ZUÑIGA INCALLA	SEGUNDO MIEMBRO
MGR. CARLOS ALBERTO SILVA DELGADO	ASESOR

En tal sentido solicito a través de su despacho la aprobación mediante **ACTO RESOLUTIVO**, del proyecto de Tesis titulado: **"APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA"**,

Sin otro particular me suscribo de usted no sin antes testimoniarle los sentimientos de mi especial consideración y estima personal.

Atentamente,

UNIVERSIDAD NACIONAL DE MOQUEGUA

Mg. Ing. Carlos Alberto Silva Delgado
Director de Escuela
Ingeniería de Sistemas e Informática



29



Universidad Nacional de Moquegua Vicepresidencia Académica

"Año del Buen Servicio al Ciudadano"

Moquegua 06 de Setiembre del 2017

UNIVERSIDAD NACIONAL MOQUEGUA
SECRETARIA GENERAL
RECIBIDO

06 SEP 2017 4005

Hora: 4:30 pm N° Reg.
Firma: gy Folio: 4 + ANUADO

OFICIO N° 323 -2017-VIPAC-CO/UNAM

SEÑOR:

**Dr. WASHINGTON ZEBALLOS GAMEZ
PRESIDENTE DE LA COMISIÓN ORGANIZADORA
UNIVERSIDAD NACIONAL DE MOQUEGUA**

Presente.-

ASUNTO : APROBACION DE PROYECTO DE TESIS, ASESOR, JURADO REVISOR DE TESIS

REFERENCIA : INFORME N° 204-2017-EPISI/UNAM-FILIAL ILO

Mediante el presente es grato dirigirme a usted, para saludarlo cordialmente y a la vez manifestarle que visto el documento de la referencia, presentado por el Mg. CARLOS ALBERTO SILVA DELGADO, Director de la Escuela Profesional de Ingeniería de Sistema e Informática, solicita la emisión de la respectiva resolución según el siguiente detalle:

1.- Aprobar el Proyecto de Tesis **"APLICACIÓN DE ALGORITMOS INTELIGENTES PARA : RECONOCIMIENTO AUTOMATICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA"**, del Bachiller Gregorio Tomás Castro Alvarez se adjunta el Acta de Aprobación del Proyecto de Tesis.

2.- Asesor del Proyecto de Tesis:

- Asesor : Mg. Carlos Alberto Silva Delgado

3.- Jurado Revisor:

- Presidente : MSc. Hugo Euler Tito Chura
- Primer Miembro : Mg. Anibal Fernando Flores García
- Segundo Miembro : Mgr. Alex Peter Zuñiga Incalla

UNIVERSIDAD NACIONAL DE MOQUEGUA
SECRETARIA GENERAL
RECIBIDO

06 SET. 2017

Hora: N° REG. 928
Firma: Folios:

Por lo expuesto, solicito a través de vuestro despacho la aprobación mediante acto resolutivo del Proyecto de Tesis, Asesor y Jurado Revisor.

Agradeciendo la atención al presente, hago propicia la ocasión para reiterarle los sentimientos de mi especial consideración y estima personal.

UNIVERSIDAD NACIONAL DE MOQUEGUA
SECRETARIA GENERAL

PROVEIDO: 0928

FECHA: 06/09/17

PASE A:

PARA:

Atentamente,

UNIVERSIDAD NACIONAL DE MOQUEGUA

[Firma]

Dra. MARIA ELENA ECHEVARRIA JAIME
VICEPRESIDENTA ACADÉMICA

PRESIDENCIA - UNAM Prov. 4005

Folios: 4 + 1 ANUADO Page a: 56

Fecha: 06 SET. 2017 Para: Sesión de COMISIÓN ORGANIZADORA

UNAM PRESIDENTE

Adjunto (03) folios + 01 Anillado

MEEJ/VIPAC
masm./sec
Cc.: Archivo.

Moquegua, Prolongación Calle Ancash S/N Telefax 053 – 461227 053 – 463514 Anexo (202) 053-461471

www.unam.edu.pe

Vice_presidencia@unam.edu.pe

INFORME N°-2017-HETCH/UNAM- FILIAL ILO

A : **MG. CARLOS ALBERTO SILVA DELGADO**
Director de Escuela Profesional Ingeniería de Sistemas e Informática
UNAM Filial Ilo

DE : **MG. HUGO EULER TITO CHURA**
Jurado Dictaminador – Presidente del Proyecto de Tesis

ASUNTO : **CONFORMIDAD DEL PROYECTO DE TESIS**

REFERENCIA : **ACTA DE CONFORMIDAD**

FECHA : Ilo, 15 Agosto del 2017



Por medio del presente me dirijo a usted, para saludarlo de manera muy cordial con la finalidad de elevar el documento de la referencia, el mismo que indica que es **APTO** el proyecto de tesis denominado: **"APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA"** presentado por la Bachiller GREGORIO TOMÁS CASTRO ALVAREZ.

En tal sentido, solicito se efectúe el trámite regular para la emisión de la Resolución para la ejecución del Proyecto.

Es todo cuanto informo a usted, para su conocimiento y demás acciones que estime conveniente.

Atentamente,



MG. HUGO EULER TITO CHURA
JURADO DICTAMINADOR
PRESIDENTE

ADJUNTO:
. Acta de conformidad del proyecto
. 01 File

CASD/Resp
lmll/Sec.
c.c. Archivo



"AÑO DEL BUEN SERVICIO AL CIUDADANO"

ACTA DE CONFORMIDAD DE PROYECTO DE TESIS

En la ciudad de Ilo, en el Auditorio de la UNAM Filial Ilo, siendo las 14:00 horas del día martes 15 de Agosto del 2017, estando designado como jurado dictaminador e integrado por los señores docentes: MSc. HUGO EULER TITO CHURA (Presidente), MG. ANIBAL FERNANDO FLORES GARCÍA (Primer Miembro), MSc. ALEX ZÚÑIGA INCALLA (Segundo Miembro), y candidato al Título Profesional de Ingeniero de Sistemas e Informática.

- BACHILLER TOMÁS CASTRO ALVAREZ

Seguidamente los jurados proceden a la revisión del perfil del Proyecto de Tesis de Titulado: "**APLICACIÓN DE ALGORITMOS INTELIGENTES PARA: RECONOCIMIENTO AUTOMÁTICO DE ENFERMEDADES FOLIARES DE CULTIVO DE PALTA**" el jurado revisor de Tesis emite observaciones del proyecto las cuales fueron levantadas por el candidato al Título Profesional de Ingeniería de Sistemas e Informática.

Terminando el acto de revisión los miembros del Jurado Dictaminador dictamina **APTO** para su emisión del acto Resolutivo aprobatorio del Proyecto de Tesis.

Siendo las 16:20 horas del mismo día, se dio por terminado la sesión y firmado los miembros del jurado dictaminador.


MG. ANIBAL FERNANDO FLORES GARCÍA
JURADO DICTAMINADOR
PRIMER MIEMBRO


MSC. EULER TITO CHURA
JURADO DICTAMINADOR
PRESIDENTE


MSC. ALEX ZÚÑIGA INCALLA
JURADO DICTAMINADOR
SEGUNDO MIEMBRO

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:

G. TOMAS CASTRO ALVAREZ

Para optar el Título de:

INGENIERO DE SISTEMAS E INFORMÁTICO

MOQUEGUA - PERÚ

2017

ÍNDICE

I	DATOS GENERALES	1
1.1	Título	1
1.2	Tipo de Investigación	1
1.3	Área de Investigación	1
1.4	Nombre de Autor	1
1.5	Localidad donde se realizara la Investigación	1
1.6	Asesor	1
II	PROBLEMA DE LA INVESTIGACIÓN	2
2.1	DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA	2
2.2	FORMULACIÓN DEL PROBLEMA	3
2.2.1	Interrogativa General	3
2.2.2	Interrogativa Específica	3
2.3	JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN	3
2.4	OBJETIVOS	4
2.4.1	Objetivo General	4
2.4.2	Objetivo Específico	4
2.5	HIPÓTESIS	5
2.5.1	Hipótesis General	5
2.5.2	Hipótesis Específico	5
III	MARCO TEÓRICO	5
3.1	ANTECEDENTES DEL ESTUDIO	5
3.2	BASES TEÓRICAS	10
3.3	DEFINICIÓN DE TÉRMINOS	17
IV	MARCO METODOLÓGICO	18
4.1	LUGAR DE EJECUCIÓN	18
4.2	TIPO Y DISEÑO	18
4.2.1	NIVEL DE INVESTIGACIÓN	18
4.3	OPERACIONALIZACIÓN DE VARIABLES	18
4.4	Variable Independiente	18
4.4.1	Indicadores	19
4.5	Variables Dependiente	19

4.5.1	Indicadores.....	19
4.6	POBLACIÓN Y MUESTRA.....	19
4.7	TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS DE VALIDACIÓN Y CONFIABILIDAD DE LOS INSTRUMENTOS.	20
4.7.1	Técnica de Muestreo	20
4.7.2	Metodología de reconocimiento de patrones.....	21
4.8	DISEÑO EXPERIMENTAL Y EVALUACIÓN DE LA EFICACIA DE LOS ALGORITMOS DE CLASIFICACIÓN.....	24
4.9	Matriz de consistencia	1
V	ASPECTOS ADMINISTRATIVOS.....	2
5.1	Cronogramas de actividad.....	2
5.1	Recursos humanos.....	3
5.2	Fuentes de financiamiento y presupuesto.....	4
VI	RESULTADOS ESPERADOS.....	5
VII	REFERENCIAS BIBLIOGRÁFICAS.....	6
VIII	ANEXOS	9
8.1	Arquitectura para la clasificación de enfermedades.....	9
8.2	Flujo de procesos para la clasificación de enfermedades	9
8.3	Objetivos Específicos.....	10
8.3.1	Base de datos de Enfermedades foliares.....	10
8.3.2	Selección de patrones representativos	11
8.3.3	Entrenamiento de los algoritmos.....	12
8.4	Fortalezas y Debilidades	12

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

Tipo Descriptivo, Explicativo y Experimental

1.3 Área de Investigación

Computación gráfica y procesamiento de imágenes.

1.4 Nombre de Autor

G. Tomas Castro Alvarez

1.5 Localidad donde se realizara la Investigación

Provincia Mariscal Nieto - Moquegua

1.6 Asesor

Mg. Carlos Alberto Silva Delgado

The logo of the National Autonomous University of Mexico (UNAM) is displayed in a large, semi-transparent, light gray font. It features a stylized sun or moon symbol above the letters 'UNAM'.

UNAM

II PROBLEMA DE LA INVESTIGACIÓN

2.1 DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA

En la actualidad las enfermedades foliares de Palta es un tema que preocupa al País y la Región de Moquegua no es ajeno a ello. Con relación a la Región Moquegua los Agricultores de Palta no cuentan con ayuda de herramientas tecnológicas para reconocimiento de enfermedades foliares de Palta, por ende, los Agricultores tienen desventajas, ya que al tener las hojas/foliares de palta una enfermedad generara al Agricultor lo siguiente:

- Disminuye la producción de la palta
- Disminuye las ganancias
- Aumenta los gastos económicos
- No se identifica la enfermedad en el momento oportuno

Dificultando al Agricultor el logro de sus objetivos planificados durante el inicio de cultivo de Palta.

La producción de frutos disminuye, tanto en cantidad como en tamaño, hasta desaparecer totalmente.

Las enfermedades foliares puede infectar hojas de cualquier edad, produciéndoles un decaimiento general de la parte aérea como una marchitez; de ello deriva su denominación de tristeza, que se caracteriza por tener las hojas más pequeñas de lo normal, de un color verde pálido y al mismo tiempo puede tener gran cantidad de frutos pequeños que no desarrollan. También presenta defoliación, ramas secas y finalmente el árbol muere.

2.2 FORMULACIÓN DEL PROBLEMA

2.2.1 Interrogativa General

- ❖ ¿Cuál es la eficiencia del algoritmo para reconocimiento automático de enfermedades foliares mediante imágenes del cultivo de palta de la región Moquegua utilizando algoritmos?

2.2.2 Interrogativa Específica

- ❖ ¿Cómo Construir una base de datos de imágenes de enfermedades foliares del cultivo de palta?
- ❖ ¿Cuáles son los patrones representativos a seleccionar en las imágenes de enfermedades foliares en la palta?
- ❖ ¿Cómo entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento de enfermedades foliares?
- ❖ ¿Qué algoritmo es más eficaz en la tarea de reconocimiento de enfermedades foliares en imágenes del cultivo de palta utilizando los patrones identificados?

2.3 JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN

La importancia de presente Proyecto de Tesis, es la identificación de la enfermedad foliar de palta. El Agricultor a la hora de tomar una foto a foliares de palta, inmediatamente tendrá el resultado o el tipo de enfermedad que tiene las hojas/foliares.

De igual manera el aplicativo tendrá grandes ventajas tales como:

- Identificación de las enfermedades en el momento oportuno
- Reduce gastos económicos
- Aumento de ganancias
- Incremento de la producción de la palta

De igual manera indicar que la implementación del presente Proyecto, ayudará a los Agricultores de la Ciudad de Moquegua a lograr su objetivo planificado.

El control de las enfermedades es una de las actividades que tiene gran importancia, debido al costo que significa la aplicación de los métodos para combatir una enfermedad.

Cabe indicar que 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. Éstas características le dan ventajas comparativas a la Región para siembra de este producto, pero que deben de ser complementadas con otras capacidades que se potencien o desarrollen para que su oferta exportable sea competitiva y sostenible (**Plan Operativo de la Palta (POP) – Moquegua**).

2.4 OBJETIVOS

2.4.1 Objetivo General

Determinar la eficacia del algoritmo para reconocimiento automático de enfermedades en cultivos de palta mediante imágenes utilizando Algoritmos

2.4.2 Objetivo Específico

- ❖ Construir una base de datos de imágenes de enfermedades foliares del cultivo de palta.
- ❖ Seleccionar patrones representativos en imágenes de enfermedades foliares en la palta.
- ❖ Entrenar los algoritmos de clasificación.
- ❖ Determinar cuál es el algoritmo de mayor eficacia en la tarea de reconocimiento de enfermedades foliares en imágenes del cultivo de palta utilizando los patrones identificados.

2.5 HIPÓTESIS

2.5.1 Hipótesis General

El reconocimiento automático de enfermedades foliares en el cultivo de palta mediante imágenes se realiza de forma eficaz utilizando algoritmos.

2.5.2 Hipótesis Específico

- ❖ La construcción de una base de datos de imágenes de enfermedades foliares del cultivo de palta permitirá realizar el reconocimiento automático de enfermedades foliares del cultivo de palta.
- ❖ La selección de patrones representativos en imágenes de enfermedades foliares del cultivo de palta son significativos en términos estadísticos.
- ❖ Los entrenamientos de los algoritmos de clasificación son significativos en términos estadísticos.
- ❖ Los algoritmos seleccionados utilizando los patrones identificados son eficaces.

III MARCO TEÓRICO

3.1 ANTECEDENTES DEL ESTUDIO

(Hetzroni, Miles, Engel, Hammer, & Latin, 1994) su sistema trata de identificar deficiencias de hierro, Sio y Nitrógeno mediante el monitoreo de hojas de lechuga. La captura de las imágenes fue hecha por una cámara de video analógica, y solo después las imágenes fueron digitalizadas. El primer paso del algoritmo propuesto en este trabajo es la segmentación de las imágenes en hojas y el fondo. En el siguiente paso un número de características de tamaño u color son extraídas desde ambas representaciones RGB y HSI. Estos parámetros son finalmente alimentados a una red

neuronal y a un clasificador estadístico, el cual es utilizado para determinar la condición de la planta

(Pydipati, Burks, & Lee, 2005) realiza una comparación de dos diferentes aproximaciones para detectar y clasificar tres tipos de enfermedades de cítricos. El autor realiza una colección de 39 características de textura y crea cuatro diferentes subconjuntos de estas características para utilizarlas en dos diferentes aproximaciones de clasificación. La primera aproximación estuvo basada en un clasificador de mínima distancia de Mahalanobis, utilizando el principio del vecino más cercano. El segundo acercamiento utilizaba funciones de bases radiales (RBF) un clasificador basado en red neuronal entrenado con el algoritmo de backpropagation. De acuerdo a los autores, ambas aproximaciones de la clasificación tienen un desempeño similar y también utilizando lo mejor de los cuatro subconjuntos, los que contienen diez características de textura de color y saturación.

(Huang, 2007) propone un método para detectar y clasificar tres tipos diferentes de enfermedades que afectan a la variedad de planta de orquídea Phalaenopsis. El procedimiento de segmentación adoptado por el autor es significativamente más sofisticado que los hallados en otros artículos, y está compuesto por cuatro pasos: remoción del recipiente de la planta utilizando un clasificador bayesiano, ecualización de la imagen utilizando una transformación exponencial, una estimación tosca para la localización de la región afectada por la enfermedad y la ecualización de la sub imagen centrada en la localización tosca. Construyen características de color y textura extraídas de una matriz de co-ocurrencia de los niveles de gris de la imagen (Haralick, Shanmugan, & Dinstein, 1973). Finalmente estas características son alimentadas a una red neuronal artificial de

topológica Multilayer Perceptrón con una capa oculta, la cual realiza la clasificación final.

(Sanyal, Bhattacharya, Parui, Bandyopadhyay, & Patel, 2007), aborda el problema de detectar y clasificar seis tipos de deficiencias de mineral en cultivos de arroz. Primero, el algoritmo extrae un número determinado de características de textura y color. Cada una de las características (textura y color) es enviada a su propia específica red neuronal Multilayer perceptrón. Ambas redes neuronales tienen una capa oculta, pero el número de neuronas en la capa oculta es diferente (40 neuronas para textura y 70 neuronas para el color). Los resultados devueltos por ambas redes neuronales son combinados, flexiblemente para la clasificación final. Una aproximación muy similar es utilizada por el mismo autor en otro artículo (Sanyal & Patel, 2008), pero en este caso el objetivo es identificar dos tipos de enfermedades (explosión y mancha marrón) que afectan a los cultivos de arroz.

(Al Bashish, Braik, & Bani-Ahmad, 2010) el método que ellos presentan tratan de identificar cinco diferentes enfermedades de plantas. Los autores no especifican las especies de plantas utilizadas en las pruebas, y las imágenes fueron capturadas in situ. Después una etapa de pre-procesamiento para limpiar la imagen, es aplicada el algoritmo de agrupamiento de K-medias con el objetivo de dividir la imagen en cuadro grupos. De acuerdo a los autores, al menos uno de los grupos debería de corresponder a una de las enfermedades. Después que, para cada agrupamiento un numero determinador de características de color y textura mediante el llamado en el trabajo método de co-ocurrencias, el cual opera con imágenes en el formato HSI. Estas características son proporcionadas a una red neuronal multilayer perceptrón con diez capas ocultas, las que realizan la clasificación final.

(Meunkaewjinda, Kumsawat, Attakitmongcol, & Srikaew, 2008) proponen un método para identificar y clasificar enfermedades que afectan a la vid, El método utiliza muchas representaciones de color (HSI, L*a*b*, UVL y YCbCr) a lo largo de la ejecución de la investigación. La separación de las hojas y el fondo de las imágenes son realizadas por una red neuronal multilayer perceptron, el cual es ensamblado con una librería de color construida a priori por medio de un mapa no supervisado y auto organizado SOM. Los colores presentes sobre la hoja son luego agrupadas por medio de un no entrenado y no supervisado Selforganizing map. Un algoritmo genético determina el número de clúster a ser adoptados en cada caso. Las regiones de salud y de enfermedad son posteriormente separadas por un clasificador basado en Maquinas de Soporte Vectorial. Por ultimo algunas manipulaciones y la imagen segmentada es utilizada por una Maquina de Soporte Vectorial multiclase, la cual realiza la clasificación entre costra moho o no enferma.

(Youwen, Tianlai, & Yan, 2008) propone un método para identificar dos enfermedades que pueden manifestarse en las hojas de pepinos. La segmentación en regiones saludables y regiones enfermas es realizada utilizando una aproximación estadística de reconocimiento de patrones. A continuación algunas características de color, forma y textura son extraídas. Posteriormente estas características sirven de insumo del clasificador basado en la Maquina de Soporte Vectorial, la que realiza la clasificación final. Los autores indican que sus resultados obtenidos por la aplicación de la Máquina de Soporte Vectorial son mejores que los obtenidos utilizando redes neuronales artificiales.

(Yao et al., 2009) el sistema propuesto tiene como objetivo identificar y clasificar tres tipos de enfermedades que afectan los cultivos de

arroz. El algoritmo primero aplica una transformación de color particular a la imagen tradicional en RGB, resultando en dos canales (y_1 , y_2). Entonces la imagen es segmentada por el método de Otsu, después las regiones enfermas son aisladas. Las características de color forma y textura son extraídas, el ultimo del espacio de color HSV. Finalmente las características son enviadas a una Maquina de Soporte Vectorial que es la encargada de realizar la clasificación.

(Camargo & Smith, 2009b) El método propuesto trata de identificar tres diferentes tipos de enfermedades que afectan a las plantas de algodón. Los autores utilizaron no solo imágenes de hojas sino que incorporaron a su estudio imágenes de frutos y tallos. La segmentación de la imagen es realizada utilizando una técnica desarrollada por los autores (Camargo & Smith, 2009a), después que un número de características es extraídas de la regiones afectadas por la enfermedad, posteriormente estas características son utilizadas para alimentar una Maquina de Soporte Vectorial. El método uno contra uno fue utilizado para permitir a la Maquina de Soporte Vectorial manejar la clasificación de múltiples clases. El autor concluye que las características de la textura tienen potencialmente mejores características de discriminación.

(Zhang & Zhang, 2010) propusieron un método para reconocer tres tipos de enfermedades de las hojas de pepinos. Como muchas otras propuestas, la separación entre sanas o enfermas es realizada con un procedimiento simple de umbralización. En la siguiente una variedad de características de color, forma y textura son extraídas. Estas características son enviadas a una Maquina de Soporte Vectorial con funciones de Base Radial como Kernels (RBF), la que realiza la clasificación final.

3.2 BASES TEÓRICAS

Support Vector Machines (Maquina de Soporte Vectorial)

Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases, más cercanos al que se llama vector soporte. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o la otra clase.

SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta.

Las Máquinas de Soporte Vectorial (MSV) son una moderna y efectiva técnica de IA, que ha tenido un formidable desarrollo en los últimos años, a continuación se presentarán los fundamentos teóricos que definen estos sistemas de aprendizaje.

Uno de los conceptos fundamentales en esta técnica es el algoritmo Vector de Soporte (VS) es una generalización no-lineal del algoritmo Semblanza Generalizada, desarrollado en la Rusia en los años sesenta. El desarrollo de los VS trae consigo el surgimiento de las Máquinas de Soporte Vectorial. Estas son sistemas de aprendizaje que usan un espacio de hipótesis de funciones lineales en un espacio de rasgos de mayor dimensión, entrenadas por un algoritmo proveniente de la teoría de optimización.

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. El concepto de "Machine Learning" se refiere al aprendizaje de una estructura proporcionada por un conjunto de datos (Brian Sánchez ,2013).

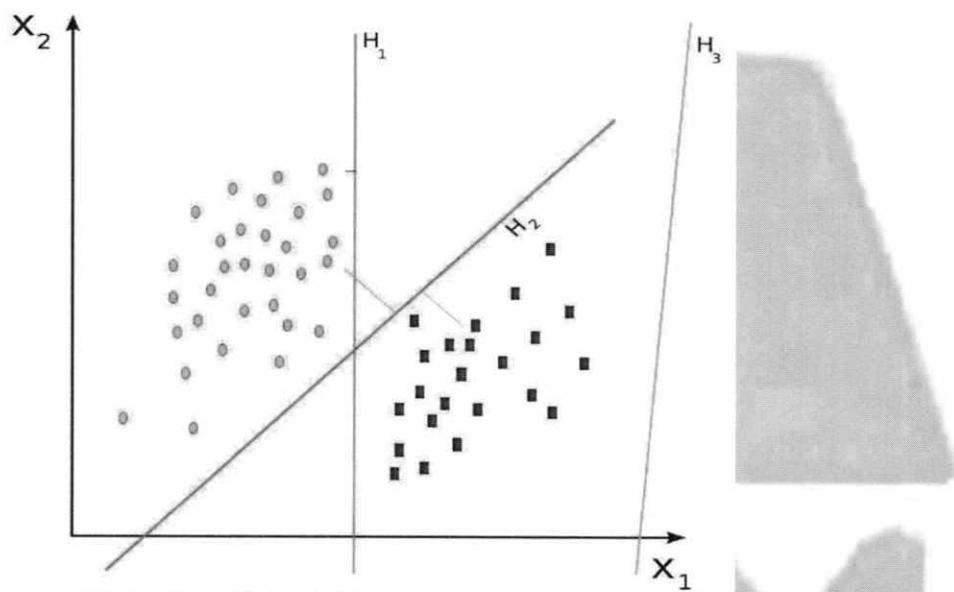


Figura 01. Máquina de soporte vectorial

Clasificación por redes neuronales

Se basan en el uso de redes neuronales artificiales que, se supone, imitan a las redes neuronales reales en el desarrollo de tareas de aprendizaje. Una neurona artificial es un objeto lógico (se trata de software no de hardware) que recibe diversas entradas, hace una suma ponderada de las mismas y produce una salida a partir de la aplicación de una función umbral a la media ponderada.

Si conectamos las salidas de unas neuronas como entradas de otras obtenemos una red neuronal. Uno de los ejemplos más típicos de red neuronal es el la Back Propagation Neural Network que aparece en el siguiente gráfico (Friedman, Geiger, & Goldszmit, 1997).

BPN Topology

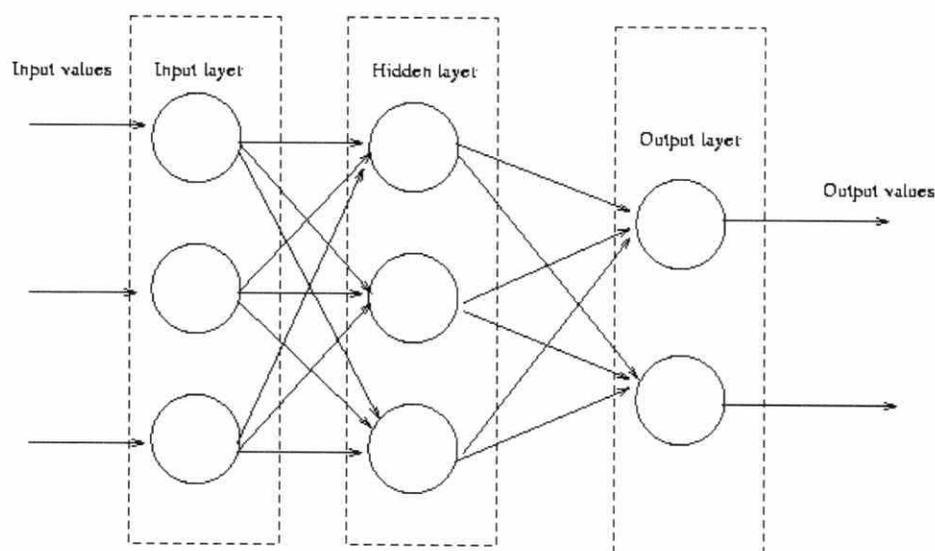


Figura 3: Red neuronal Artificial

Consta de una capa de entrada con tantas neuronas como variables de entrada se vayan a introducir en el modelo (en el caso de la teledetección sería una por cada banda utilizada para clasificar), una capa oculta que realiza la mayor parte del cálculo y una capa de salida con tantas neuronas como posibles clases existan. En teledetección esta salida suele consistir en un valor numérico entre 0 y 1 para cada clase, cuanto mayor sea este valor más verosímil resulta que el pixel pertenezca a la clase en cuestión

Para trabajar con una red neuronal existen varias fases:

- **Entrenamiento.** Se le introducen a la red la respuesta espectral de píxeles cuya clase se conoce y se compara la salida con la realidad. A partir de esta comparación se modifican los coeficientes de ponderación de todas las neuronas para que se obtenga la respuesta adecuada (se trata de un procedimiento automático) es decir un 1 en la clase correcta y ceros en las incorrectas
- **Estabilización.** Al principio del entrenamiento, los factores de ponderación cambian muy deprisa, pero conforme este se desarrolla (y si las áreas de entrenamiento se han seleccionado correctamente) se estabilizan (no se modifican, aunque se vuelvan a introducir los píxeles de entrenamiento). En este momento finaliza la fase de entrenamiento
- **Clasificación** Se introducen las respuestas espectrales de los píxeles cuya clase no se conoce y se adjudican a la clase que da una respuesta más alta (que no va a ser necesariamente 1).

Se trata en definitiva de un método de clasificación no paramétrico robusto que da buenos resultados cuando las respuestas espectrales de las clases no siguen una distribución normal. La clave está en el conjunto de coeficientes de ponderación que constituyen un conjunto de parámetros que deben ajustarse a unos datos de entrada y salida. Por tanto en cierto modo es equivalente a una regresión.

Aprendizaje profundo (deep learning).

El aprendizaje profundo es parte de un conjunto más amplio de métodos de aprendizaje automático basados en aprender representaciones de datos. Una observación (por ejemplo, una imagen) puede ser representada en muchas formas (por ejemplo, un vector de píxeles), pero algunas representaciones hacen más fácil aprender tareas de interés (por ejemplo, "¿es esta imagen una cara

humana?") sobre la base de ejemplos, y la investigación en esta área intenta definir qué representaciones son mejores y cómo crear modelos para aprender estas representaciones.

Varias arquitecturas de aprendizaje profundo, como redes neuronales profundas, redes neuronales profundas convolucionales, y redes de creencia profundas, han sido aplicadas a campos como visión por computador, reconocimiento automático del habla, y reconocimiento de señales de audio y música, y han mostrado producir resultados de vanguardia en varias tareas.

Random forest (Bosques Aleatorio)

Uno de los métodos más populares usados por los científicos de datos es el algoritmo Random Forest, uno de los mejores algoritmos de clasificación, capaz de organizar grandes cantidades de datos con exactitud.

Random forest es una combinación de árboles predictores en la que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. El algoritmo para inducir un random forest fue desarrollado por Leo Breiman y Adele Cutler, siendo Random forests su marca de fábrica.

Este algoritmo mejora la precisión en la clasificación mediante la incorporación de aleatoriedad en la construcción de cada clasificador individual. Esta aleatorización puede introducirse en la partición del espacio (construcción del árbol), así como en la muestra de entrenamiento. El Random Forest comienza con una técnica de aprendizaje automático estándar llamada "árbol de decisiones", que, en cuanto al conjunto, corresponde a un aprendizaje. En un árbol de decisión, una entrada se introduce en la parte superior y hacia abajo

a medida que atraviesa el árbol de los datos, los cuales se acumulan en conjuntos más pequeños.

Principales beneficios:

- Precisión.
- Funciona de manera eficiente con bases de datos de gran tamaño.
- Maneja miles de variables de entrada sin necesidad de borrado.
- Da estimaciones de qué variables son importantes en la clasificación.
- Genera una estimación objetiva interna de la generalización de errores.
- Proporciona métodos eficaces para estimar datos que faltan.
- Los “bosques” generados se pueden guardar para uso futuro en otros datos.
- Los prototipos se calculan de manera que proporcionan información acerca de la relación entre las variables y la clasificación.
- Calcula proximidades entre pares de casos que se pueden utilizar en la agrupación, la localización de los valores extremos, o (en escala) dan interesantes vistas de los datos
- Ofrece un método experimental para la detección de interacciones de variables.

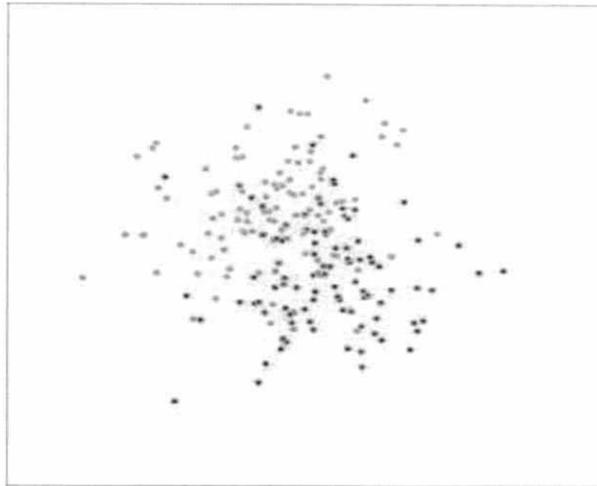


Figura 03. Datos de entrenamiento.

(Naive Bayes) Bayes ingenuos

En términos simples, un clasificador de Bayes ingenuo asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable. Por ejemplo, una fruta puede ser considerada como una manzana si es roja, redonda y de alrededor de 7 cm de diámetro. Un clasificador de Bayes ingenuo considera que cada una de estas características contribuye de manera independiente a la probabilidad de que esta fruta sea una manzana, independientemente de la presencia o ausencia de las otras características.

Estimación de parámetros y modelo de eventos

Todos los parámetros del modelo (por ejemplo, clases prioris y características de las distribuciones de probabilidad) se puede aproximar con frecuencias relativas del conjunto de entrenamiento. Estas son las estimaciones de máxima verosimilitud de las probabilidades. Una clase priori se puede calcular asumiendo clases equiprobables (es decir, $priori = 1 / (\text{número de clases})$), o mediante el cálculo de una estimación de la probabilidad de clase del conjunto de

entrenamiento (es decir, (el priori de una clase dada) = (número de muestras en la clase) / (número total de muestras)). Para la estimación de los parámetros de la distribución de una característica, se debe asumir una distribución o generar modelos de estadística no paramétrica de las características del conjunto de entrenamiento.

Las hipótesis sobre las distribuciones de características son llamadas el modelo de eventos del Clasificador Bayesiano Ingenuo. La distribución multinomial y la distribución de Bernoulli son populares para características discretas como las encontradas en la clasificación de documentos (incluyendo el filtrado de spam). Estas hipótesis conducen a dos modelos distintos, que a menudo se confunden. Cuando se trata con los datos continuos, una hipótesis típica es que los valores continuos asociados con cada clase se distribuyen según una Distribución normal.

Palta

Uno de los principales hitos de la industria de esta variedad durante los últimos 20 años ha sido su predominio sobre la variedad Fuerte en los principales países productores de palta de calidad. La amplia aceptación de palta en casi todos los mercados mundiales ha fortalecido la demanda. Una de las mayores virtudes de palta es su prolongada estación de cosecha, ha permitido aumentar el consumo mundial (AGROBANCO PERU ,2011).

3.3 DEFINICIÓN DE TÉRMINOS

Algoritmo: Conjunto ordenado de operaciones sistemáticas que permite hacer un cálculo y hallar la solución de un tipo de problemas.

Regresión: es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación

entre una variable dependiente y una o más variables independientes (o predictores).

Palta: La palta es el nombre que recibe un árbol del continente americano y su fruto. El árbol, de la familia Lauraceae, se denomina *Persea americana*.

SVM: Las máquinas de soporte vectorial, (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje.

IV MARCO METODOLÓGICO

4.1 LUGAR DE EJECUCIÓN

El presente proyecto se realizara 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 explicativo y experimental. Descriptivo porque se describirán los principales patrones identificados en las imágenes de enfermedades foliares del cultivo de palta, Explicativo porque se determinara como se relacionan los patrones entre ellos y el estado del cultivo. Experimental por que se manipula el uso de los patrones detectados en imágenes para determinar los mejores parámetros para el algoritmo de palta.

4.2.1 NIVEL DE INVESTIGACIÓN

La investigación es un trabajo; Descriptivo, Explicativo y Experimental

4.3 OPERACIONALIZACIÓN DE VARIABLES

4.4 Variable Independiente

- **Algoritmo de clasificación:** de imágenes de enfermedades del cultivo de palta.

- **Características:** cualitativas y cuantitativas que aportan información para el diagnóstico del cultivo de palta.

4.4.1 Indicadores

Características:

- ❖ Color
- ❖ Forma
- ❖ Presencia de cuerpos extraños
- ❖ Estado de las ramas

Algoritmos de clasificación:

- ❖ Naive Bayes
- ❖ Random forest
- ❖ Deep learning
- ❖ Clasificación por redes neuronales
- ❖ Support Vector Machines

4.5 Variables Dependiente

Eficacia de la clasificación automática de enfermedades foliares del cultivo de palta mediante imágenes utilizando ALGORITMOS INTELIGENTES, que será cuantificado por la tasa de acierto de tal forma que.

4.5.1 Indicadores

Tasa de acierto > 90% el algoritmo es eficaz.

4.6 POBLACIÓN Y MUESTRA

La Población

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

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:

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

Leyenda:

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.

4.7 TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS DE VALIDACIÓN Y CONFIABILIDAD DE LOS INSTRUMENTOS.

4.7.1 Técnica de Muestreo

Se realizara un muestreo sistemático de grilla sobre la ubicación geográfica de los cultivos de palta hasta recolectar un total de 384 muestras que contengan 192 ejemplares sanos y 192 con enfermedades.

4.7.2 Metodología de reconocimiento de patrones

4.7.2.1 Matemáticas

Este tipo de metodologías hacen uso de las propiedades comunes de los patrones y se basan en las reglas de clasificación formuladas en un marco matemático. Además, este enfoque se divide a su vez en otras dos categorías: las determinísticas y las estadísticas. Entre los métodos determinísticos podemos encontrar, por ejemplo, los algoritmos de aprendizaje iterativo tales como los algoritmos utilizados para el entrenamiento de redes neuronales. Por otra parte, dentro de los métodos estadísticos podemos encontrar a los clasificadores basados en la regla de clasificación de Bayes (**MEDINA, RECONOCIMIENTO DE CARACTERES MANUSCRITOS ON-LINE (EN LÍNEA) USANDO LA FUNCION SPLINE, OCTUBRE 2004**) (Correa).

4.7.2.1.1 Etapas para el reconocimiento de patrones

De manera general, cuando se intenta resolver algún problema que implique el reconocimiento de patrones se deben de tomar en cuenta las siguientes etapas (**MEDINA, OCTUBRE 2004**) (**TOBAR, Octubre 2007**):

- Adquisición de datos.
- Reprocesamiento.
- Extracción de características.
- Reconocimiento.

Adquisición de datos

Para poder realizar el reconocimiento es necesario, realizar o implementar alguna etapa de adquisición de los datos que describen al patrón que se desea clasificar. Esto implica que, el

método que se utilizará para realizar la adquisición de los datos, dependerá de la naturaleza del patrón. Por ejemplo, si se desea desarrollar un sistema para reconocer caracteres, esta etapa se podría realizar con la ayuda de una cámara digital o utilizando un escáner.

Preprocesamiento

Una vez que ya se cuenta con los datos que describen a cada uno de los patrones, es más conveniente realizar una etapa de preprocesamiento sobre cada uno de ellos en lugar de ser dados como entrada del sistema tal y como fueron obtenidos durante la etapa de adquisición de datos.

El realizar un preprocesamiento sobre los datos tiene varias ventajas; la principal de ellas es que puede reducir la dimensionalidad de los datos, lo cual mejora substancialmente la ejecución del sistema, sobre todo cuando se utiliza una metodología como la de redes neuronales. Dentro de esta etapa se puede preprocesar el patrón de entrada de tal forma que todos los patrones tengan el mismo tamaño (escala) consiguiendo con esto que el sistema sea invariante al escalamiento. Además de esto, también se busca lograr que el sistema sea invariante a la traslación. Cuando un sistema es invariante a la traslación y al escalamiento de los patrones, se dice que el sistema cuenta con un conocimiento previo.

Extracción de características

Uno de los principales problemas en el reconocimiento de patrones, es encontrar una manera óptima de representar la información original que describe a cada uno de los patrones. Este problema es conocido como extracción de características. Este proceso de extracción de características trata de reducir la

cantidad de información (reducción de dimensionalidad) que representa a cada uno de los patrones, obteniendo de esta forma, un vector de características que represente de la mejor manera posible al patrón original.

La extracción de características debe de cumplir con las siguientes condiciones:

- ❖ La dimensionalidad del vector de características debe de ser menor que la del patrón original.
- ❖ Las características deben representar una codificación óptima de la entrada, perdiendo la información que no sea muy importante.

Especialmente, cuando se realiza el reconocimiento de patrones utilizando redes neuronales, la extracción de características debe tratar de obtener un vector de características con una dimensionalidad mucho menor a la del patrón original, puesto que, un vector con una dimensionalidad más pequeña que sea dado como entrada a la red neuronal tiene varios beneficios. En primer lugar, la cantidad de pesos que deben de ser aprendidos es menor, y en segundo lugar, al tener menos pesos, el tiempo de entrenamiento puede ser reducido considerablemente.

Reconocimiento

Finalmente, una vez que se han realizado las etapas de preprocesamiento y extracción de características, se procede a realizar el objetivo principal del sistema: el reconocimiento.

De manera general, la etapa de reconocimiento consiste en obtener una salida del sistema como respuesta a un conjunto de estímulos dados como entrada (patrones).

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

Sensibilidad

Indica la capacidad reconocer de forma efectiva el estado de los cultivos de palta que si presentan una enfermedad mediante el análisis de imágenes.

Especificidad

Indica la capacidad reconocer de forma efectiva el estado de los cultivos de palta que no presentan una enfermedad mediante el análisis de imágenes.

Precisión

Indica la capacidad de dar el mismo resultado en mediciones diferentes realizadas en las mismas condiciones.

Tasa de falsa aceptación (rate of false acceptance - FAR)

$$FAR = \frac{\text{número de falsas aceptaciones}}{\text{numero de impostores}}$$

Tasa de falso rechazo (rate of false rejection - FRR)

$$FRR = \frac{\text{número de falsos rechazos}}{\text{numero de rostros evaluados}}$$

Tasa de igual error (equal error rate -ERR)

$$EER = \frac{\text{número de falsas aceptaciones} + \text{número falsos rechazos}}{\text{número total de entradas}}$$

Prueba T para la diferencia de proporciones

El estadístico de prueba que permite contrastar a partir de dos muestras aleatorias e independientes es p la estimación de obtenida del total de observaciones.

$$z = \frac{p1 - p2}{\sqrt{\frac{p(1-p)}{n1} + \frac{p(1-p)}{n2}}}$$

Si se consideran las proporciones como medias y se aplica la prueba t utilizada para comparar medias poblacionales los resultados no son fiables ya que la estimación del error típico que realiza el programa no coincide con la del estadístico de prueba. Contingencia) y realizar el contraste de independencia Chi-cuadrado.



4.9 Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES
<p>PROBLEMA GENERAL:</p> <p>¿Cuál es la eficiencia del reconocimiento automático de enfermedades foliares mediante imágenes del cultivo de palta de la región Moquegua utilizando algoritmos inteligentes?</p> <p>PROBLEMAS ESPECÍFICOS:</p> <ol style="list-style-type: none"> 1) ¿Cómo construir una base de datos de imágenes de enfermedades foliares del cultivo de palta? 2) ¿Cuáles son los patrones representativos a seleccionar en las imágenes de enfermedades foliares en la palta? 3) ¿Cómo entrenar los algoritmos de clasificación para realizar la tarea de reconocimiento de enfermedades foliares? 4) ¿Qué algoritmo es más eficaz en la tarea de reconocimiento de enfermedades foliares en imágenes del cultivo de palta utilizando los patrones identificados? 	<p>OBJETIVO GENERAL:</p> <p>Determinar la eficacia del reconocimiento automático de enfermedades en cultivos de palta mediante imágenes utilizando algoritmos.</p> <p>OBJETIVOS ESPECÍFICOS:</p> <ol style="list-style-type: none"> 1) Construir una base de datos de imágenes de enfermedades foliares del cultivo de palta. 2) Seleccionar patrones representativos en imágenes de enfermedades foliares en la palta. 3) Entrenar los algoritmos de clasificación. 4) Determinar cuál es el algoritmo de mayor eficacia en la tarea de reconocimiento de enfermedades foliares en imágenes del cultivo de palta utilizando los patrones identificados. 	<p>HIPÓTESIS GENERAL:</p> <p>El reconocimiento automático de enfermedades foliares en el cultivo de palta mediante imágenes se realiza de forma eficaz utilizando algoritmos inteligentes.</p> <p>HIPÓTESIS ESPECÍFICAS:</p> <ol style="list-style-type: none"> 1) La construcción de una base de datos de imágenes de enfermedades foliares del cultivo de palta permitirá realizar el reconocimiento automático de enfermedades foliares del cultivo de palta. 2) La selección de patrones representativos en imágenes de enfermedades foliares del cultivo de palta son significativos en términos estadísticos. 3) Los entrenamientos de los algoritmos de clasificación son significativos en términos estadísticos. 4) Los algoritmos seleccionados utilizando los patrones identificados son eficaces. 	<p>VARIABLE INDEPENDIENTE</p> <ul style="list-style-type: none"> + Características + algoritmos de clasificación <p>INDICADORES característicos:</p> <ul style="list-style-type: none"> o color o forma o presencia de cuerpos extraños o estado de las ramas <p>algoritmos de clasificación:</p> <ul style="list-style-type: none"> o Naive Bayes o Random forest o Deep learning o Clasificación por redes neuronales o Support Vector Machines <p>VARIABLE DEPENDIENTE</p> <p>Eficacia de la clasificación automática de enfermedades foliares del cultivo de palta mediante imágenes utilizando algoritmos inteligentes, que será cuantificado por la tasa de acierto de tal forma que.</p> <p>INDICADORES</p> <p>tasa de acierto > 90% el algoritmo es eficaz</p>
<p>MÉTODO Y DISEÑO</p> <p>MÉTODO: comparativo</p> <p>NIVEL: comparativa</p>	<p>POBLACIÓN Y MUESTRA</p> <p>POBLACIÓN: La población está conformada por la totalidad de los cultivos de palta, en la región Moquegua.</p> <p>MUESTRA: el tamaño de la muestra está calculada al 95% de confianza y un margen de error de 0.05</p>	<p>TÉCNICAS E INSTRUMENTOS</p> <p>TÉCNICAS: determinar el mejor algoritmo</p> <p>INSTRUMENTO: prueba t para la diferencia de proporciones</p>	

V ASPECTOS ADMINISTRATIVOS.

5.1 Cronogramas de actividad.

Nro.	ACTIVIDAD	AÑOS											
		2017						2018					
		Jun.	Jul.	Ago.	Set.	Oct.	Nov.	Dic.	Ene.	Febr.	Mar.		
1	Elaboración del Perfil de Proyecto de Tesis	X											
2	Evaluación del Perfil de Proyecto de Tesis		X										
3	Levantamiento de Observación		X										
4	Compra de Equipos		X										
5	Servicio de consultoría con Expertos			X									
6	Recolección de datos y Muestreo de los Ejemplares			X									
7	Obtención de las características			X									
8	Determinación de las características				X								
9	Entrenamiento y Construcción del algoritmos					X							
10	Implementación del algoritmo en una plataforma computacional						X						
11	Evaluación y Optimización de la eficacia del algoritmo							X					
12	Elaboración del informe final								X				
13	Levantamiento de Observaciones									X			
14	Sustentación de Proyecto final										X		
15	Levantamiento de observaciones finales											X	
16	Presentación del Proyecto final												X

5.1 Recursos humanos.

NOMBRES	CARGO
MG. CARLOS ALBERTO SILVA DELGADO	ASESOR
ING.ARMANDO RUEDAS ALFARO	CO-ASESOR
G. TOMAS CASTRO ALVARES	TESISTA

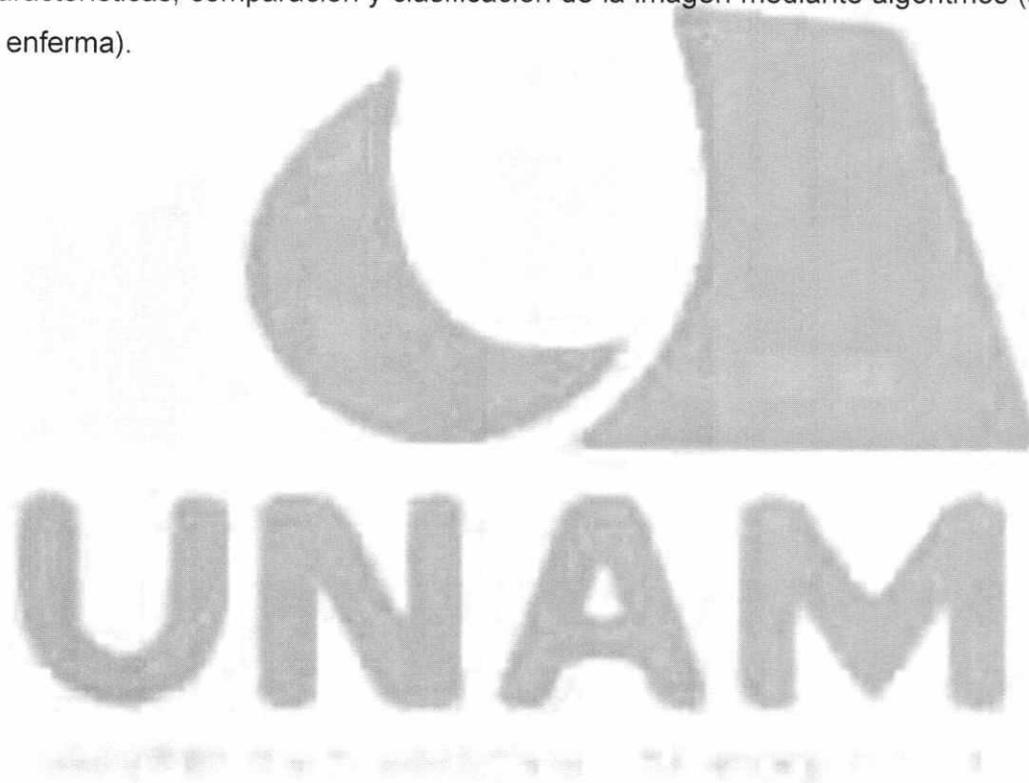


5.2 Fuentes de financiamiento y presupuesto

EQUIPOS											
COMPUTADORA			S/. 5.000,00								S/. 5.000,00
IMPRESORA			S/. 1.000,00								S/. 1.000,00
CAMARA FOTOGRAFICA			S/. 4.000,00								S/. 4.000,00
MATERIAL											
alojamiento WEB						S/. 500,00	S/. 500,00				S/. 1.000,00
Material Fungible 2 (aún no determinado)						S/. 500,00	S/. 500,00				S/. 1.000,00
PROGRAMAS INFORMÁTICOS											
Servidor										S/. 1.500,00	S/. 1.500,00
Colección de Biblioteca Digital						S/. 500,00					S/. 500,00
GASTOS											
Pasajes y viáticos			S/. 400,00			S/. 400,00	S/. 400,00			S/. 400,00	S/. 2.000,00
Consultoría			S/. 1.000,00			S/. 1.000,00	S/. 1.000,00			S/. 1.000,00	S/. 4.000,00
Total, de costos del proyecto			S/. 11.400,00			S/. 2.900,00	S/. 1.400,00			S/. 1.400,00	S/. 20.000,00

VI RESULTADOS ESPERADOS

Se espera obtener como resultado una tasa de acierto mayor al 90% en el reconocimiento automático de enfermedades foliares del cultivo de palta. Este objetivo involucra el desarrollo de un sistema de clasificación, con sub módulos para la adquisición de imágenes, pre procesamiento de las imágenes, extracción de características, comparación y clasificación de la imagen mediante algoritmos (sana o enferma).



VII REFERENCIAS BIBLIOGRÁFICAS

- Al Bashish, D., Braik, M., & Bani-Ahmad, S. (2010). A framework for detection and classification of plant leaf and stem diseases. *Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010*, 113–118. <https://doi.org/10.1109/ICSIP.2010.5697452>
- Camargo, A., & Smith, J. S. (2009a). An image-processing based algorithm to automatically identify plant disease visual symptoms. *Biosystems Engineering*, 102(1), 9–21. <https://doi.org/10.1016/j.biosystemseng.2008.09.030>
- Camargo, A., & Smith, J. S. (2009b). Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture*, 66(2), 121–125. <https://doi.org/10.1016/j.compag.2009.01.003>
- Friedman, N., Geiger, D., & Goldszmit, M. (1997). Bayesian Network Classifiers. *Machine Learning*, 29, 131–163. <https://doi.org/10.1023/a:1007465528199>
- Haralick, R., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. <https://doi.org/10.1109/TSMC.1973.4309314>
- Hetzroni, A., Miles, G. E., Engel, B. A., Hammer, P. A., & Latin, R. X. (1994). Machine vision monitoring of plant health. *Advances in Space Research*, 14(11), 203–212. [https://doi.org/10.1016/0273-1177\(94\)90298-4](https://doi.org/10.1016/0273-1177(94)90298-4)
- Huang, K.-Y. (2007). Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, 57(1), 3–11. <https://doi.org/10.1016/j.compag.2007.01.015>
- Marín, J. M. (2009). Análisis de Cluster y Árboles de Clasificación. *Springer*, 1–21.
- Meunkaewjinda, a, Kumsawat, P., Attakitmongcol, K., & Srikaew, A. (2008). Grape leaf disease detection from color imagery using hybrid intelligent system. *IEEE - Institute of Electrical and Electronics Engineers*, 1, 513–516.

<https://doi.org/10.1109/ECTICON.2008.4600483>

Pydipati, R., Burks, T., & Lee, W. (2005). Statistical and neural network classifiers for citrus disease detection using machine vision. *Transactions of the ASAE*, 48(5), 2007–2014. Retrieved from <http://asae.frymulti.com/abstract.asp?aid=19994&t=1>

Sanyal, P., Bhattacharya, U., Parui, S. K., Bandyopadhyay, S. K., & Patel, S. (2007). Color Texture Analysis of Rice Leaves Diagnosing Deficiency in the Balance of Mineral Levels towards Improvement of Crop Productivity. *10th International Conference on Information Technology (ICIT 2007)*, 85–90. <https://doi.org/10.1109/ICIT.2007.40>

Sanyal, P., & Patel, S. C. (2008). Pattern recognition method to detect two diseases in rice plants. *The Imaging Science Journal*, 56(6), 319–325. <https://doi.org/10.1179/174313108X319397>

Yao, Q., Guan, Z., Zhou, Y., Tang, J., Hu, Y., & Yang, B. (2009). Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Color Texture Features. *2009 International Conference on Engineering Computation*, 79–83. <https://doi.org/10.1109/ICEC.2009.73>

Yin, S., Qian, Y., & Gong, M. (2017). Unsupervised hierarchical image segmentation through fuzzy entropy maximization. *Pattern Recognition*, 68, 245–259. <https://doi.org/10.1016/j.patcog.2017.03.012>

Youwen, T., Tianlai, L., & Yan, N. (2008). The recognition of cucumber disease based on image processing and support vector machine. *Image and Signal Processing, 2008. CISP'08. Congress on*, 2, 262–267. <https://doi.org/10.1109/CISP.2008.29>

Zhang, J., & Zhang, W. (2010). Support vector machine for recognition of cucumber leaf diseases. *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, 5(1), 264–266. <https://doi.org/10.1109/ICACC.2010.5487242>

- Andres Morales, G., & Gomes Ruis, A. (2006). ESTUDIO E IMPLEMENTACION DE UNA HERRAMIENTA BASADA EN MAQUINAS DE SOPORTE VECTORIAL APLICADA A LA LOCALIACION DE FALLAS EN SISTEMAS DE DISTRIBUCION. Bucaramanga: Universidad Industrial de Santander.
- Debandi , N. (s.f.). RECONOCIMIENTO Y CLASIFICACION DE HONGOS DERMATOFITOS USANDO MAQUINA DE SOORTE VECTORIAL. Buenos Aires.
- Esther Gutierrez , A. (Junio , 2007). APLICACION DE LAS MAQUINAS DE SOORTE VECTORIAL PARA EL RECONOCIMIENTO DE MATRICULAS. Madrid.
- Gavidia Bobadilla, G. E. (s.f.). CLASIFICADORES BASADOS EN MAQUINAS DE SOPORTE VECTORIAL PARA EL DIAGNOSTICO Y PREDICCIÓN DE LA ENFERMEDAD DE ALSHEIMER. Barcelona , España .
- Gil Gonsales , W. (2013). CLASIFICADOR ROBUSTO BASADO EN MÁQUINAS DE SOPORTE VECTORIAL PARA LA LOCALIZACIÓN DE FALLAS EN SISTEMAS DE DISTRIBUCIÓN. Colombia: Universidad Tecnologia de Pereira .
- Resendis Trejo, J. (2006). LA MAQUINA DE VECTORES DE SOPORTE PARA IDENTIFICACION DE LINEA. MEXICO.
- Riobo Otero, V. (s.f.). RECONOCIMIENTO DE LOCALIACION MEDIANTE MAQUINAS DE SOPORTE VECTORIAL.
- Saldaña Alatamirano, K., & Rodrigues Urquiaga, R. (s.f.). DETECCION DE SOMNOLENCIA USANDO MAQUINA DE SOPORTE VECTORIAL A ARTIR DE DATOS DE UN ELECTROENFALOGRAMO. Trujillo.
- Solera Ureña, R. (2011). MAQUINA DE VECTORES SOPORTE PARA RECONOCIMIENTO ROBUSTO DE HABLA. Madrid.

VIII ANEXOS

8.1 Arquitectura para la clasificación de enfermedades

La arquitectura propuesta para la implementación de la aplicación “Reconocimiento de enfermedades foliares del cultivo de palta” consiste de un servidor para la aplicación y base de datos. Además se podrá acceder a la aplicación desde diferentes dispositivos. La FIGURA 1 muestra el diagrama de Arquitectura Propuesto

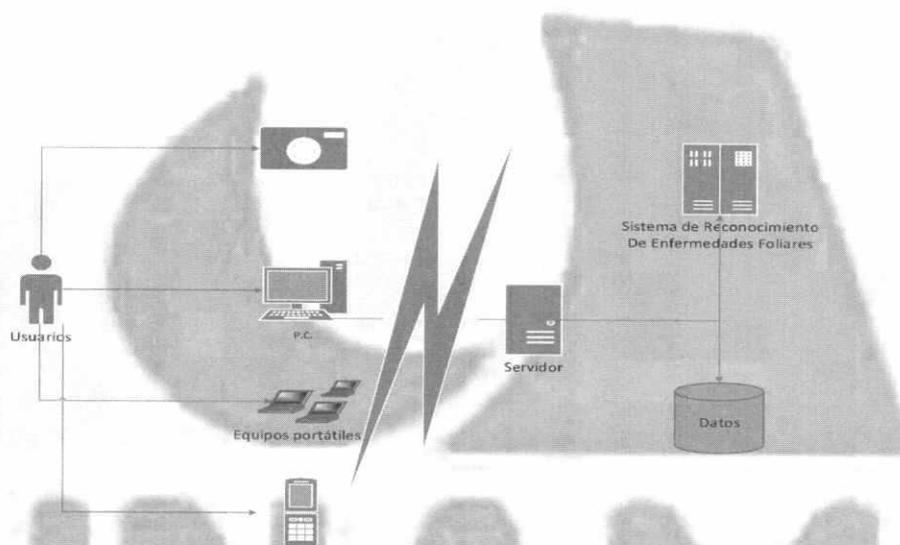


FIGURA 1
Propuesta de Arquitectura

8.2 Flujo de procesos para la clasificación de enfermedades

La FIGURA 2 resume el flujo del proceso para la construcción del sistemas de reconocimiento de enfermedades foliares del cultivo de palta



FIGURA 2
Flujo de procesos para la clasificación de enfermedades Foliares del cultivo de palta

8.3 Objetivos Específicos

8.3.1 Base de datos de Enfermedades foliares

La base de datos de imágenes contempla la recolección de imágenes de hojas con enfermedades y hojas sanas, La base de datos debe contemplar una interfaz de usuarios, además de contemplar la indexación de imágenes mediante.

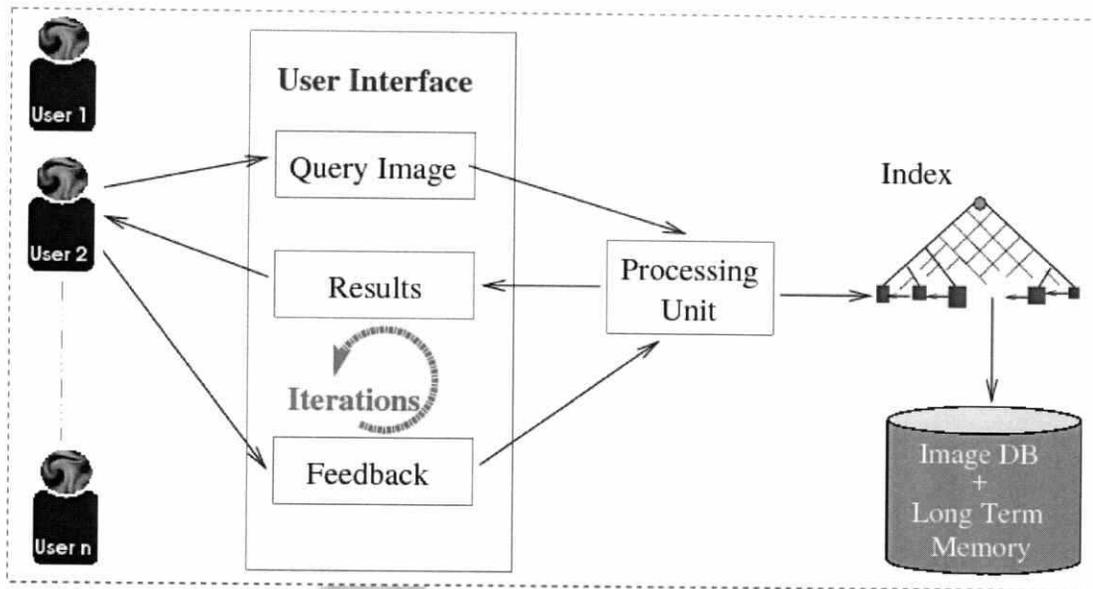


FIGURA 3
Base de Datos para la recuperación de imágenes
Basado en contenidos

8.3.2 Selección de patrones representativos

La FIGURA 4 resume el proceso de extracción de características de un sistema de reconocimiento de patrones.

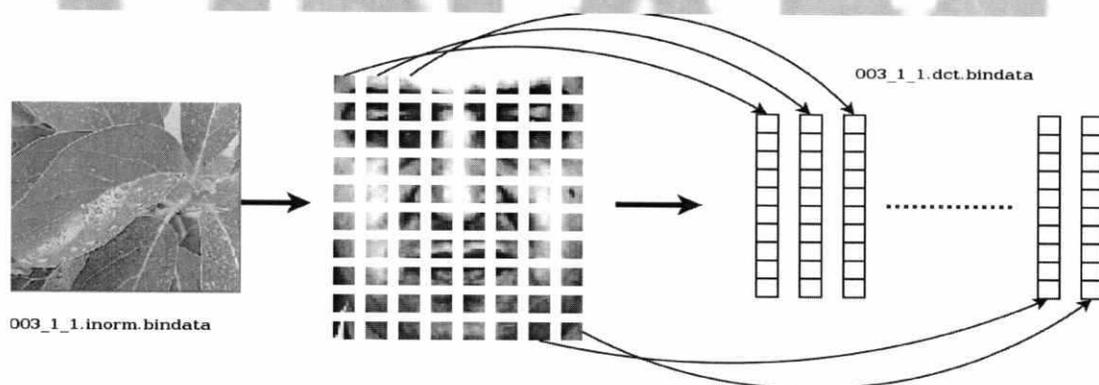


FIGURA 4
Proceso de extracción de características

8.3.3 Entrenamiento de los algoritmos

Diagrama

La FIGURA 5 muestra el proceso de entrenamiento para el modelo de clasificación.

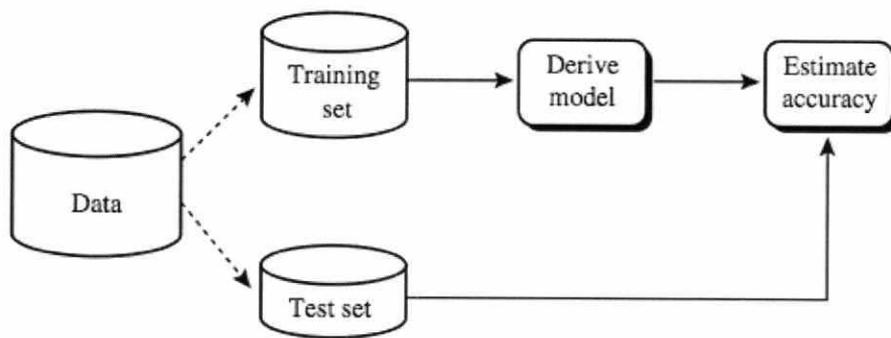


FIGURA 5
Proceso de entrenamiento para el modelo de clasificación

8.4 Fortalezas y Debilidades

Fortalezas

- ❖ Conocimiento de servidores Linux.
- ❖ Conocimiento de Lenguajes de programación C++ y R.
- ❖ Conocimiento de algoritmos

Debilidades

- ❖ Escasa información, respecto a estudios similares.
- ❖ Desconocimiento de las enfermedades de la palta, puesto que no está relacionado con la carrera profesional de ingeniería de sistemas e Informática.

PALTA



Figura 01

PRODUCCIÓN NACIONAL DE PALTA 1994 - 2012 (TM)

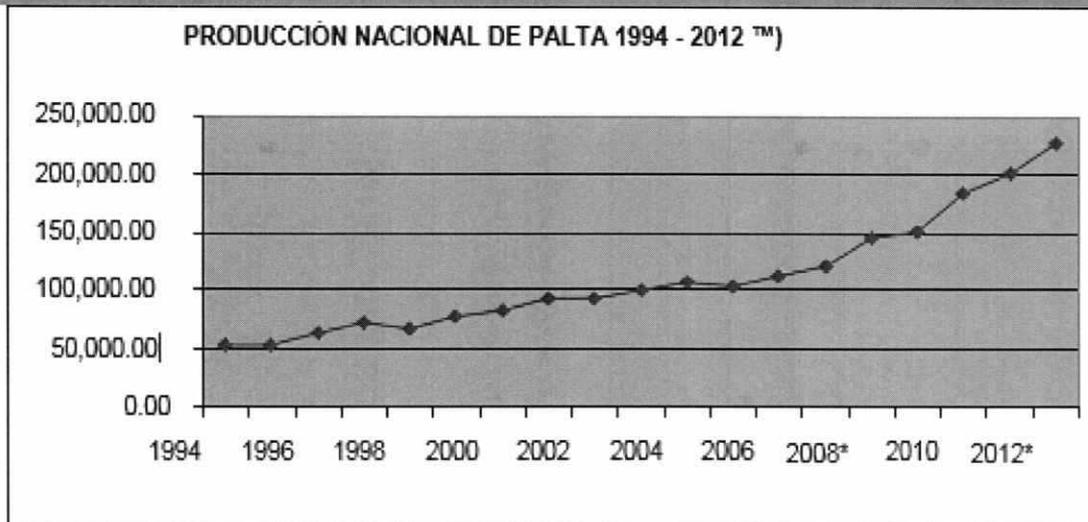


Figura 02

PRODUCCIÓN REGIONAL DE PALTA MOQUEGUA 2007 – 2011 (Tm)

UBICACIÓN	PRODUCCION (Tm)				
	2007	2008	2009	2010	2011
REG. MOQUEGUA	2,005	2,686	3,124	3,114	3,144
Prov. Mcal. Nieto	1,123	1,407	1,608	1,667	1,455
Moquegua	517	612	618	611	518
Samegua	379	510	634	577	540
Torata	209	268	261	375	295
Carumas	0	0	0	10	9
Cuchumbaya	9	9	32	34	33
San Cristóbal	9	9	62	60	61
Prov. Gral. Sánchez Cerro	881	1,279	1,516	1,447	1,689
Omate	731	989	1,136	1,065	1,281
Coalaque	55	192	231	221	202
Quinistaquillas	76	77	111	104	109
Puquina	5	0	8	8	8
La Capilla	5	0	12	12	12
Matalaque	9	21	19	37	76

Figura 03

UNAM