INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

MODELO PARA LA DETECCIÓN DEL HUANGLONGBING MEDIANTE EL ANÁLISIS DE IMÁGENES DE HOJAS SINTOMÁTICAS DE ÁRBOLES DE LIMÓN PERSA

TESIS

PARA OBTENER EL GRADO DE MAESTRO EN
SISTEMAS COMPUTACIONALES

P R E S E N T A

GALDINO MARTÍNEZ FLORES

ASESOR:

M.I.A. ROBERTO ÁNGEL MELÉNDEZ ARMENTA

MISANTLA, VERACRUZ DICIEMBRE, 2016.

INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA
DIVISIÓN DE ESTUDIOS PROFESIONALES
AUTORIZACIÓN DE IMPRESIÓN DE TRABAJO DE TITULACIÓN MAESTRÍA

FECHA: 07 de Diciembre de 2016.

ASUNTO: **AUTORIZACIÓN DE IMPRESIÓN DE TESIS.**

A QUIEN CORRESPONDA:

Por medio de la presente se hace constar que el (la) C:

GALDINO MARTÍNEZ FLORES

estudiante de la maestría en SISTEMAS COMPUTACIONALES con No. de Control 142T0050 ha cumplido satisfactoriamente con lo estipulado por el **Lineamiento de Posgrado para la obtención del grado de Maestría** mediante **Tesis.**

Por tal motivo se **Autoriza** la impresión del **Tema** titulado:

MODELO PARA LA DETECCIÓN DEL HUANGLONGBING MEDIANTE EL ANÁLISIS DE IMÁGENES DE HOJAS SINTOMÁTICAS DE ÁRBOLES DE LIMÓN PERSA

Dándose un plazo no mayor de un mes de la expedición de la presente a la solicitud del examen para la obtención del grado de maestría.

ATENTAMENTE

M.I.A. Roberto Ángel Meléndez Armenta
Presidente

DIVISION DE ESTUDIOS PROFESIONALES

M.G.C. Eduardo Gutiérrez Almaraz
Secretario

M.C. Saúl Reyes Barajas
Vocal

Archivo.

VER. 01/03/09

F-SA--39

AGRADECIMIENTOS

Gracias a Dios por darme la vida, salud y sabiduría para terminar este proyecto profesional.

Gracias a CONACYT por brindarme una beca y hacer posible el desarrollo de este trabajo de investigación.

Gracias a mis maestros que contribuyeron en la realización de este trabajo de investigación, Dr. Luis Alberto Morales Rosales y al M.I.A. Roberto Ángel Meléndez Armenta, por compartir sus conocimientos y enseñanzas.

Gracias a mi asesor de tesis, a mis revisores y a los que contribuyeron de alguna forma en esta investigación.

Gracias al Tecnológico de Misantla, por las facilidades otorgadas para cursar este posgrado.

Gracias a mis hijos María Fernanda, Gonzalo Guillermo y Samuel, por la motivación que siempre me dieron para culminar esta tesis.

Gracias a mi esposa, I.B.Q Yadira Oliva Mendoza por apoyarme incondicionalmente en el desarrollo y terminación de este proyecto.

DEDICATORIA

A Dios porque para él todo es posible y porque me acompaña en todo momento.

A mis seres queridos.

A mis padres por su apoyo en cada etapa de mi formación académica y profesional.

A mis hijos María Fernanda, Gonzalo Guillermo y Samuel por ser buenos hijos y exitosos en lo que hacen.

A mi esposa I.B.Q. Yadira Oliva Mendoza por ser excelente madre, ser humano y esposa.

TABLA DE CONTENIDO

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

El propósito de esta investigación es diseñar un modelo para la identificación de la enfermedad de los cítricos llamada huanglongbing (HLB), específicamente en árboles de limón persa. Esta es la enfermedad más devastadora en los últimos años en el mundo. Ha ocasionado pérdidas millonarias a productores y a gran cantidad de personas que forman parte del proceso productivo de los cítricos, que va desde jornaleros, transportistas, empacadoras hasta las exportadoras.

El protocolo de actuación para la identificación de esta enfermedad del SENASICA (Servicio Nacional de Sanidad Inocuidad y Calidad Agroalimentaria), se debe implementar para productores pequeños, medianos, grandes. Sin embargo, los pequeños productores que incluyen a los huertos de traspatio, no lo aplican por diferentes factores. Esta situación podría ocasionar un problema de contaminación.

Este modelo sería de gran importancia para implementarse como medio de alerta en huertos de traspatio y para huertos de pequeños productores.

Existen diversas herramientas que ayudan a identificar enfermedades en plantas. Para esta enfermedad se han explorado diversos enfoques, como PCR, espectroscopía así como la identificación de compuestos orgánicos volátiles de plantas.

Por otro lado, la utilización de técnicas de tratamiento de imágenes así como la aplicación de técnicas de inteligencia artificial, es cada vez más aplicada para la identificación de diversas enfermedades o deficiencias nutrimentales en plantas bajo condiciones controladas o no controladas. Especialmente para el análisis y clasificación de imágenes se han aplicado con gran éxito las redes neuronales artificiales.

La propuesta es diseñar un modelo de identificación de HLB a través de los síntomas presentes en hojas de limón persa, usando tratamiento de imágenes, segmentación de imágenes, análisis de componentes principales y clasificación supervisada a través de una red neuronal backpropagation.

CAPÍTULO I GENERALIDADES.

1.1 Planteamiento del problema.

La citricultura es una de las actividades más importantes en el sector primario de nuestro país, posicionándose como el quinto productor a nivel mundial. La superficie sembrada con cítricos es de cerca de 550 mil hectáreas. Sin embargo, las diversas plagas y enfermedades que se presentan en los cultivos de cítricos, repercuten negativamente en la salud de las plantas, así como en la economía de las personas u organizaciones que dependen de ella directa o indirectamente. Entre ellas se encuentran el HLB o huanglongbing de los cítricos (Candidatus Liberibacter spp), la tristeza de los cítricos(VTC), leprosis de los cítricos (CiLV), el cancro (Xanthomonas axonopodis pv. citri) y la clorosis variegada de los cítricos (Xylella fastidiosa).

El Huanglongbing(HLB) o enverdecimiento de los cítricos, es considerada la enfermedad más destructiva en la actualidad por la severidad de los efectos en producción, por la rapidez en la que se dispersa, por afectar a todas las variantes de cítricos comerciales y por no tener cura aún. Cada árbol con HLB es eliminado para evitar la propagación de esta enfermedad. Esto puede causar la pérdida de gran parte de los cultivos sino se detecta y controla a tiempo, perdiendo paulatinamente la rentabilidad de los huertos. Ésta enfermedad es causada por la bacteria Candidatus Liberibacter asiaticus (presente en Asia y América); Candidatus Liberibacter africanus (presente en África) y Candidatus Liberibacter americanus (presente en América (Brasil). En México esta enfermedad es causada por la primera de ellas y el vector Diaphorina citri se encarga de trasportar la bacteria de una planta a otra. México, es el quinto productor de cítricos a nivel mundial y Veracruz es el principal productor de naranja, limón y toronja a nivel nacional, en esta entidad no está

presente la enfermedad, sin embargo, si se presenta y no se identifica a tiempo podría infectar a gran parte de huertos comerciales de cítricos, poniendo en riesgo la producción estatal y nacional, incrementando los costos e impactando negativamente en la economía de los que dependen directa o indirectamente de ella.

La secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA) aplica un protocolo de actuación para la detección del HLB de los cítricos, que se compone de un conjunto de actividades que tienen como objetivo detectar oportunamente la aparición de síntomas de esta enfermedad en las hojas de los cítricos, así como su control una vez detectada para evitar que más plantas se sigan contaminando. Sin embargo, sólo personal técnico capacitado de la dirección General de Sanidad Vegetal(DGSV) y de Organismos Auxiliares de Sanidad Vegetal (OASV) tienen el entrenamiento suficiente para explorar y reconocer los síntomas del HLB en las hojas de los árboles, quienes toman fotos de las hojas con características sintomáticas, envían las imágenes a expertos de la DGSV para incrementar el grado de certidumbre en la identificación de la enfermedad, estos expertos observan las imágenes solicitando a los técnicos muestras vegetales de los casos identificados por ellos como positivos, para finalmente realizar pruebas de laboratorio para confirmar la presencia de la enfermedad.

Este proceso de inspección se realiza de manera aleatoria a huertos comerciales. Para el caso de los pequeños y medianos productores de cítricos, el problema se incrementa debido a que se les recomienda contratar a Profesionales Fitosanitarios Autorizados en la materia, que laboren de manera independiente a los OASV para realizar esta exploración, incrementando los costos de producción.

Los expertos en identificación de síntomas del HLB observan la presencia de ellos en las hojas, los tallos y frutos de los árboles. Siendo las hojas donde se manifiestan en primer lugar, donde se presentan las siguientes características:

- En las hojas se manifiestan los primeros síntomas observándose una coloración amarilla pálida con áreas color verde, irregulares (moteado), manchas asimétricas, defoliación, engrosamiento y aclaración de las nervaduras, asimetría y difusión de colores en las nervaduras y folíolos, hojas pequeñas y rectas. Muchas veces puede confundirse con deficiencias de micronutrientes como: zinc, hierro, calcio, magnesio, manganeso y cobre.

- Los brotes vegetativos que crecen a partir de ramas con síntomas de HLB, desde que emergen hasta que completan su desarrollo y alcanzan su madurez, presentan aspecto vigoroso, sin síntomas del HLB y su color es verde claro, muy similar a lo observado en brotes vegetativos de la misma edad, desarrollados en árboles sanos. Sin embargo, cuando los brotes alcanzan entre 45 y 60 días de edad, las hojas empiezan a desarrollar pequeños puntos de color amarillento, distribuidos en toda la lámina foliar. De los 60 a 70 días, esas pequeñas manchas evolucionan hasta formar un moteado difuso con distribución irregular en la hoja. Las manchas que se forman generalmente no pasan al otro lado de la vena central, lo que le da la característica de distribución asimétrica, que caracteriza esta enfermedad y la distingue de síntomas de deficiencias nutricionales como lo señalan Gottwald, da Graça y Bassanezi (2007). A los 90 días de edad, las manchas amarillentas en la

mayoría de las hojas crecieron hasta que prácticamente desaparece el color verde característico de la especie.

Para automatizar el proceso de identificación de síntomas del HLB, es importante considerar los filtros que se deben aplicar a las imágenes para reducir o eliminar el ruido o información no deseada, causada por diversos factores en el proceso de adquisición de imágenes digitales, así como las diversas técnicas para la extracción de características principales, con el fin de reducir el número de variables empleadas o encontrar posibles relaciones entre ellas.

Los tipos de ruido básicos que pueden existir en las imágenes digitales son el impulsional o sal y pimienta, donde los pixeles de la imagen son muy diferentes en color e intensidad a los pixeles circundantes; el ruido gaussiano, donde todos y cada uno de los pixeles que componen la imagen cambian su valor a un pequeño valor de acuerdo con una distribución normal o gaussiana; y el ruido multiplicativo, generado por la falta de iluminación uniforme sobre la escena capturada. Las variables a considerar para la extracción de características son luz o iluminación, textura, tamaño, formas geométricas, contraste, color, posición, simetrías entre otras para el proceso de reconocimiento de patrones.

En investigaciones recientes, proponen soluciones a través de diversas técnicas de espectroscopia como la propuesta de Mishra, Karimi, Ehsani y Lee (2012), Mota et al. (2014), Mishra, Karimi, Ehsani y Albrigo (2011) y otras a través de visión por computadora como solución rápida y de bajo costo utilizando la densidad de la mezcla gaussiana (GMD) para extraer el objeto del primer plano de toda la imagen, seguido de la

extracción de características y el reconocimiento de la existencia de HLB en la hoja, propuesta de Deng, Li y Hong (2014). Sin embargo, para la implementación de estas propuestas se deben adquirir dispositivos y capacitación especializada que para los pequeños productores representan un alto costo de inversión.

Por lo tanto, en este trabajo de investigación se explora la identificación de la enfermedad conocida como huanglongbing, basado en la identificación de los síntomas que se manifiestan en las hojas de los cítricos infectados con el HLB, a través de reconocimiento de patrones en imágenes, usando modelos o técnicas de la inteligencia artificial para el análisis e interpretación.

1.2 Propuesta de solución

En esta investigación se propone desarrollar un modelo para identificación del HLB en imágenes adquiridas de las hojas del cítrico que presenta síntomas característicos de la enfermedad. Es importante mencionar que estos síntomas se muestran en las hojas de los brotes que tienen una edad de 45 a 70 días, por lo tanto, la selección de la muestra debe cumplir con esta condición, principalmente debe mostrar el moteado característico de esta enfermedad.

Para desarrollar este modelo se utilizarán técnicas de reconocimiento de patrones a través de redes neuronales usando clasificación supervisada. En esta propuesta se contemplan las siguientes etapas para su implementación: captura de datos, preprocesamiento, segmentación de imágenes, extracción de características, y aplicación de una red neuronal para su clasificación.

Los datos de entrada del modelo, serán las imágenes que se obtendrán en campo en un ambiente controlado usando un dispositivo capaz de adquirir y almacenar imágenes digitales. Estos datos, deberán ser preprocesados para disminuir el ruido adquirido durante el proceso de captura y segmentados para la identificación de bordes. Se aplicará el algoritmo de Canny para detección de bordes. Este algoritmo propuesto por Canny (1986) se fundamenta en los operadores de la primera derivada, su rendimiento es bueno porque extrae bordes del ancho de un pixel, cierra contornos y evita detectar falsos bordes. Valverde (2007) y Sánchez (2010) en sus respectivos trabajos confirman el gran rendimiento para la detección de bordes en imágenes de cualquier formato. Ellos catalogan a este algoritmo como uno de los mejores métodos para la detección de bordes.

Canny recibe como entrada la imagen de una hoja de limón persa, le aplica operaciones de filtrado gaussiano para reducir el ruido, obtiene la magnitud y dirección del vector gradiente en cada pixel, optimiza el adelgazamiento de los bordes hasta llegar al tamaño de un pixel de ancho. La salida de Canny es una imagen donde se marcan solo los bordes reales de la imagen que recibe como entrada este algoritmo.

Uno de los inconvenientes al trabajar con imágenes es la gran cantidad de datos que se deben procesar, esta cantidad depende de la resolución de la imagen, de tal manera que una imagen tiene de f filas y c columnas tiene una dimensionalidad de f X c coordenadas o pixeles para una imagen en escala de grises. Los sistemas de reconocimiento de patrones en imágenes no deberían implementarse directamente sobre estos espacios de datos de tan elevada dimensión ya que haría el clasificador muy lento. En esta propuesta se utilizará el algoritmo de Análisis de componentes principales (PCA) para reducir la dimensionalidad de los datos sin perder la información relevante. Esta técnica recibirá como entrada las imágenes de entrenamiento procesadas previamente con el algoritmo detector de bordes Canny. Como resultado de este proceso tendremos las mismas imágenes pero con menor dimensionalidad representado a través de sus componentes principales. Rodríguez, M. (2005) menciona que un clasificador necesita sólo los componentes principales más significativos, esto es, en lugar de usar todas las variables originales, utilizamos un número reducido de los componentes principales incorrelacionados generados, lo que en muchos casos aumenta la efectividad del clasificador al mismo tiempo que reduce sus necesidades de procesamiento y memoria.

Finalmente en la etapa de reconocimiento de patrones se utilizará una red neuronal artificial de tipo backpropagation para el análisis y clasificación de imágenes basándose

aprendizaje supervisado. Se tomarán como datos de entrada, los patrones que concentran la información más relevante de las hojas con síntomas del HLB; extraídos y proporcionados por el proceso del análisis de componentes principales. Esta técnica reducirá la dimensionalidad de los datos proporcionados inicialmente, pero, también obtendrá los patrones que mejor representen a esta enfermedad. Lo anterior ayudará a la red neuronal a trabajar con una cantidad menor de datos, siendo estos, los más representativos que beneficiarán la detección de la enfermedad a través de este modelo a desarrollar.

1.3 Justificación.

La problemática fitosanitaria en la agricultura impacta de diversas maneras a las entidades que confluyen en esta área. El mercado nacional e internacional, solicita a los productores frutas más sanas y de mayor calidad, libres de enfermedades ocasionadas por las diversas plagas, con el objetivo de garantizar la salud de quienes consumirán finalmente las frutas.

Para erradicar o controlar las enfermedades causadas por las plagas y enfermedades es necesario aplicar tratamientos fitosanitarios, trayendo consigo un incremento en los costos de producción, mermando la economía de los pequeños productores principalmente.

La Organización de la Naciones Unidas para la Alimentación y Agricultura menciona que El HLB (Huanglongbing) es la enfermedad más devastadora de los cítricos a nivel mundial, debido a los daños que causa, a la dificultad de su diagnóstico y a la velocidad de su dispersión. Esta enfermedad y los costos de su manejo traen consigo pérdidas directas en el rendimiento, volumen y valor de la producción con consecuencias económicas, sociales y ambientales negativas. El impacto económico del HLB está asociado a la importancia del producto, entre los que se encuentran todas las especies comerciales de cítricos.

Esta enfermedad en Sudáfrica ocasiona pérdidas anuales del 30% al 100% de la producción, siendo la enfermedad más importante desde hace algunas décadas. En la Isla Reunión y en Tailandia se han reportado plantaciones abandonadas por los estragos que causa el HLB. En Filipinas, la producción de cítricos disminuyó de 11 mil 700 toneladas a 100 toneladas de 1960 a 1970 por el ataque de este patógeno; lo anterior, debido a que

afectó a 7 millones de plantas en esa década; los registros muestran que en 1971 causó la muerte de un millón de árboles en una sola provincia de ese país. En Indonesia, más de 3 millones de plantas fueron afectadas entre 1960 y 1970. En Guandong, China, durante el período comprendido entre 1977 y 1981 fueron erradicadas 960 mil plantas de mandarinas y limones por causa del HLB, lo que disminuyó la producción de la región de 450 mil a 5 mil toneladas. Todas las plantaciones de mandarinas y naranja dulce de Arabia Saudita desaparecieron durante la década de 1975 a 1985.

En México, durante el primer año desde que el HLB fue detectado (2009) se estima que causó una reducción en el rendimiento de los árboles afectados de hasta un 50%, estimándose que en un plazo de cinco años, bajo un escenario de alto impacto de la enfermedad, las pérdidas potenciales de las zonas productoras serían de cerca de 3 millones de toneladas, equivalentes al 41 % de la producción total del país.

El HLB se encuentra presente en huertas y traspatios de 250 municipios de 16 estados de nuestro país, lo que representa el 6.1% de la citricultura nacional afectada por esta enfermedad, ya que no se encuentra ampliamente diseminada en dichos estados y aún no está presente en algunos de los estados que cuentan con la mayor superficie citrícola como Veracruz, San Luis Potosí, Tamaulipas y Nuevo León.

La reducción y pérdida en la producción de cítricos afecta directamente al empleo, tanto en campo como en la agroindustria, y en empresas relacionadas a la producción, procesamiento y distribución de cítricos.

Hoy en día, cada vez más se implementan propuestas de solución a diversas problemáticas del área agrícola a través de algoritmos y técnicas de inteligencia artificial.

La visión por computadora es un área muy utilizada para la identificación de diversas enfermedades o deficiencias nutricionales en las plantas.

El Servicio Nacional de Sanidad Inocuidad y Calidad Agroalimentaria (SENASICA), reconoce la reacción en cadena de la polimerasa (PCR) como única técnica para confirmar la presencia de HLB en un árbol de cítrico, ésta, es un proceso químico, que necesita de personal técnico capacitado y dispositivos especiales de un laboratorio químico. El presupuesto del gobierno para la identificación y tratamiento de esta enfermedad es para los grandes y medianos productores principalmente. Investigadores como Reyes y Cevallos (2009) en su investigación aseguran que PCR un proceso costoso y tardado para la identificación de esta enfermedad, por lo tanto, no es una opción para los pequeños productores ni para la aplicación de esta técnica de forma masiva; por otro lado, no permite el análisis en campo.

Lograr un eficiente manejo de HLB es indispensable para garantizar la supervivencia de la citricultura en nuestro país, principalmente en la región de Martínez de la torre donde Gil (2015), así como, el Servicio de Información Agroalimentaria y Pesquera (SIAP) de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), lo ubican como principal productor y exportador de limón persa. Por ello, identificar árboles enfermos de HLB es de vital importancia para un productor en esta región, para que la enfermedad no se disemine a otros árboles del mismo huerto o a huertos de otros productores. Por el contrario, no identificarlos, Mora (2011) menciona que generaría considerables pérdidas económicas a los productores, exportadoras, transportistas, y jornaleros relacionados con este producto, impactando económicamente a la región. Como ha sucedido en los estados de Michoacán, colima y Oaxaca importantes

productores de este cítrico sólo después de Veracruz. También menciona que, el HLB en México ha causado más daños en plantaciones de limón mexicano y persa; considerando que un panorama de infección de alto impacto ocasionaría una reducción de la producción anual de 146,954 toneladas, equivalente a 17.6% de una producción de 834,966 toneladas de limón persa.

El problema de identificación de la enfermedad del HLB ha sido atendido por varios investigadores a través de diversos enfoques y técnicas. Sin embargo, estas técnicas incluyen tecnología que incrementan su costo al implementarlos como método de solución a gran escala.

Con esta investigación, se propone identificar la enfermedad del huanglongbing usando visión por computadora, tratamiento de imágenes de las hojas de los cítricos con síntomas característicos de esta enfermedad, redes neuronales y algoritmos de clasificación supervisada. Esta propuesta beneficiará a principalmente a los pequeños productores, quienes son los menos tecnificados y poseen menor capital económico para la identificación y tratamiento de enfermedades en sus huertos.

1.4 Objetivo general.

Desarrollar un modelo de clasificación supervisado para la identificación del HLB en árboles de limón persa basado en redes neuronales tipo backpropagation.

1.5 Objetivos específicos.

- Implementar una base de datos de imágenes de hojas que manifiestan síntomas de esta enfermedad en árboles contaminados y hojas sanas para utilizarlos como datos de entrenamiento.

- Seleccionar las variables que se usarán en el modelo como discriminantes para identificar el HLB.

- Aplicar el algoritmo de Canny en el procesamiento de imágenes para identificar los objetos o segmentos que presentan las imágenes en las hojas.

- Implementar la técnica de Análisis de componentes Principales sobre las imágenes procesadas con el algoritmo Canny, para la extracción de características que sirvan como discriminantes de la enfermedad y para la reducción de la dimensionalidad de los datos.

- Desarrollar una red neuronal tipo backprogation, usando los componentes principales identificados como entrada a la red para clasificar las hojas sintomáticas con HLB y hojas sanas de limón persa.

- Realizar pruebas al modelo utilizando como entrada imágenes de hojas sintomáticas con HLB y hojas sin la enfermedad para validarlo.

1.6 Hipótesis.

Es posible identificar la enfermedad del HLB en imágenes de hojas sintomáticas de árboles de limón persa con una edad entre 45 y 70 días, recopiladas usando una cámara digital convencional, a través de un modelo basado en una red neuronal backpropagation con el fin de ayudar a los pequeños productores de la región de Martínez de la Torre, ubicados en Veracruz a prevenir brotes de infección.

1.7 Alcances y limitaciones

1.7.1 Alcances.

Proporcionar una herramienta a los pequeños y medianos productores de limón persa para la identificación de la enfermedad huanglongbing.

Identificación del HLB basado en imágenes de hojas de limón tipo persa con una edad de 45 a 70 días y que manifiestan características similares a las presentes en hojas de árboles contaminados con esta enfermedad.

Sólo se consideran hojas completas, no deben estar rotas ni dobladas, este modelo se diseña para aceptar imágenes en formato jpg capturadas desde cualquier dispositivo móvil que presta este servicio en un ambiente controlado.

1.7.2 Limitaciones.

Este modelo de clasificación para identificar la enfermedad del HLB, no es en tiempo real. Dependiendo de la ubicación de estos elementos y el método de la transferencia de

estas imágenes es el tiempo de retraso considerado para la entrega de resultados a cargo del modelo.

Existen varios escenarios para este proceso, cada uno con sus limitaciones; uno de ellos es tener una laptop en el huerto, otra, usar la red celular cuando está disponible y el más viable para un productor pequeño es tener una computadora en casa. En la primera y segunda se aceleran los resultados, pero se encarece, debido a que necesitan servicio y equipo de comunicación adicional y en la última, los resultados se retardan, pero es más viable económicamente.

CAPÍTULO II ESTADO DEL ARTE

Son varias las enfermedades que amenazan seriamente la producción de los cítricos en nuestro país. La que nos interesa estudiar es conocida como huanglongbing (HLB), esta es una de las principales preocupaciones de los productores de cítricos. Actualmente la inspección visual es el mecanismo utilizado para la identificación del HLB a gran escala, sin embargo, este mecanismo tiene varios inconvenientes, una de las más importantes es la subjetividad del diagnóstico ocasionando tasas bajas de detección.

La severidad de esta enfermedad, causada por la bacteria Candidatus Liberibacter en sus tres variantes, ha preocupado a diversos entes vinculados a la producción de cítricos. Uno de los sectores ocupados en proponer técnicas o modelos para la identificación de esta enfermedad, es el de investigación. Por ello, grupos y centros de investigación de todo el mundo, trabajan arduamente para mejorar las propuestas conocidas hasta ahora.

Para exponer las distintas investigaciones realizas en este ámbito, este trabajo se basa en la clasificación que realiza Sankaran, Mishra, Ehsani y Davis (2010), es una investigación sobre las técnicas avanzadas para la identificación de enfermedades en plantas. En este trabajo clasifican las investigaciones en técnicas moleculares, técnicas espectroscópicas y de imagen así como perfilado de compuestos orgánicos volátiles de plantas para la detección de enfermedades. Adicional e estos se incluye la técnica de tinción en yodo.

2.1 Técnicas utilizadas para la identificación del HLB

2.1.1 Técnicas moleculares

Las técnicas moleculares de detección de enfermedades en plantas, se han utilizado para confirmar la presencia o ausencia de una enfermedad. La cantidad mínima de microorganismo que puede ser detectado en la muestra se le llama técnica de sensibilidad molecular. López et al. (2003) informaron de que la sensibilidad de las técnicas moleculares para la detección de bacterias varió de 10 a 106 unidades formadoras de colonias/ml. Las técnicas moleculares comúnmente usadas para la detección de las enfermedades son el ensayo por inmunoabsorción ligado a enzimas (ELISA) y Reacción en Cadena de la Polimerasa (PCR), este último método es el único aprobado para el diagnóstico de HLB.

2.1.1.1 Reacción en cadena de la polimerasa PCR.

Es una técnica biológica molecular cuyo objetivo es obtener un gran número de copias de un fragmento de ADN particular. Sirve para amplificar un fragmento de ADN. La PCR es otra de las técnicas empleadas para la detección de patógenos. Se basa en el principio de la complementariedad de las bases de los ácidos nucleicos y la capacidad de síntesis del ADN por parte de la enzima polimerasa.

La reacción en cadena de la polimerasa (PCR) es uno de los métodos de diagnóstico más empleados en la actualidad debido a su elevada especificidad y sensibilidad (Collazo, Luis. & Llauger, 2009 y Collazo, Núñez, Luis & Llauger, 2011; Ramos-González, Hernández-Rodríguez & Banguela-Castillo, 2011). Para el diagnóstico de rutina de las bacterias asociadas a la enfermedad HLB se han implementado varios formatos de la PCR como son: PCR convencional, PCR dúplex, PCR anidada y la PCR cuantitativa en tiempo real (Collazo et al. 2009). Estos

métodos se basan en el uso de cebadores específicos diseñados a partir de varias regiones del genoma de estos microorganismos (Jagoueix, Bové & Garnier, 1997; Hocquellet, Toorawa, Bové & Garnier, 1999; Teixeira, et al., 2005).

Luis et al. (2014) en su investigación implementaron la PCR dúplex optimizada y PCR anidada optimizada, para ello, utilizaron diferentes cebadores o iniciadores. Sus resultados mostraron una alta eficacia de la PCR anidada para la detección del HLB a partir de hojas con moteado asimétrico difuso. Obteniendo alta especificidad (97,6%) y sensibilidad (75,4%) diagnóstica del método, así como la precisión (84,61%) y los valores de tasas de falsos positivos y negativos (2,6% y 24,5%, respectivamente), esta técnica la recomiendan aplicarla en viveros principalmente para la certificación de plantas sanas. La PCR dúplex permite la detección diferencial y simultánea de las tres especies de Ca. L. asociadas a la enfermedad, su aplicación se considera importante para la identificación de nuevas especies de esta bacteria en territorio cubano. Los costos de esta tecnología es de 10 y 20 USD por muestra, respectivamente, y requiere de laboratorios con equipamiento sofisticado y personal altamente calificado para realizar los análisis.

Li, Hartung, y Levy (2006) en su investigación PCR en tiempo real cuantitativo para la detección e identificación de la especie candidatus liberabacter asociado al HLB: en esta investigación desarrollaron PCR TaqMan cuantitative, usando un conjunto específico de sondas cebadoras o iniciadoras TaqMan 16S rDNA para diferentes tipos de candidatos Liberabacter (HLBas, HLBaf y HLBam,). Además, se utilizó un conjunto de sondas sebadoras basadas en citocromo oxidasa (COX) en planta como un control interno positivo, para evaluar la calidad de los extractos de ADN.

El ADN fue extraído de la nervadura central de tres hojas sintomáticas y asintomáticas de árboles sanos y enfermos de naranja dulce cultivados en invernadero. El diagnóstico puede ser ejecutado con extractos de ADN en campo, en menos de una hora, usando un equipo portable llamada SmartCycler para la detección de los patógenos en tiempo real. No hay discordancia entre los resultados mostrados con PCR en tiempo real y PCR convencional.

Sin embargo, sus autores comentan que el uso del TaqMan múltiple es muy útil para laboratorios en cuarentena, así como la gestión de las enfermedades y los programas de investigación. Por los equipos que manejan, los reactivos o enzimas necesarias se deben tener conocimientos técnicos sobre esta área, por lo que no puede aplicarse a mayor escala ni por cualquier persona.

Aksenov et al. (2014) afirma que la PCR es un proceso costoso y complicado que es especialmente difícil porque las cargas del patógeno C. Liberibacter se distribuyen de manera desigual en los tejidos de los vegetales, pueden estar presentes en niveles bajos y puede fluctuar con el tiempo. Debido a la posibilidad de falsos negativos de las muestras, muchos de los árboles que están detectados como sanos, son una fuente de infección importante durante mucho tiempo después de la prueba inicial.

Batista, Peña, López, Pérez y Llauger (2008) afirman que para la realización exitosa y repetible de la PCR, es necesario tener en cuenta las buenas prácticas de laboratorio y mantener un riguroso control de la calidad de los ensayos. Es de particular importancia, garantizar la calidad del agua que se emplee y evitar contaminaciones por problemas de manipulación (pipeteo, producción de aerosoles, uso de puntas o tubos no estériles, entre otros). El diagnóstico basado en la detección de ácidos nucleicos (PCR) presenta numerosas ventajas sobre otros

métodos. Estas técnicas pueden emplearse para la detección de cualquier tipo de patógeno, abarcan una mayor información relativa a éstos, son versátiles y su especificidad y sensibilidad son superiores. Sin embargo, tienen en contra los requerimientos de infraestructura, el costo del equipamiento, los gastos periódicos elevados por consumo de reactivos y la necesaria calificación del personal ejecutor, unido a las estrictas exigencias de bioseguridad si se emplea el marcaje radiactivo.

Las diferentes técnicas de reacción en cadena de la polimerasa (PCR) que se han implementado, para la identificación del Candidatus Liberibacter americanus, africanus y asiaticus, han sido de gran importancia y cumplen diferentes objetivos particulares.

Las técnicas moleculares son actualmente las más exactas para la detección de enfermedades en las plantas y para la identificación del HLB no es la excepción. Sin embargo, es una técnica cara, a menudo es un trabajo laborioso y requiere de instrumentos especializados. Por lo tanto, también, requiere de personal capacitado y con conocimiento técnicos en el área de química para el manejo de muestras y resultados. El tiempo de identificación depende de las muestras requeridas para ser analizadas, el número de personal y equipos y materiales. Debido a lo anteriormente mencionado, las técnicas moleculares no son recomendables para una inspección a gran escala, mucho menos como una herramienta para la detección del HLB en los cultivos de limón persa de pequeños productores.

Tinción en yodo es una reacción química usada para determinar la presencia o alteración de almidón u otros polisacáridos. Se han realizado diversos estudios de la tinción con yodo, como un método para la detección del HLB. Los resultados se han comparado con los de la PCR convencional.

Kawano, Tetsuya, Atsushi, Numazawa y Yasuda (2006) desarrollaron un kit para diagnóstico de HLB basado en la reacción del yodo con el almidón presente en las muestras infectadas. Taba, Nasu, Takaesu, Ooshiroy Moromizato (2006) demostraron que el método del almidón concuerda el 75% de las veces con PCR en hojas y el 95% en otras partes de los árboles.

Luis et al. (2014) en su trabajo de investigación en cuba, identificaron los porcentajes de coincidencia entre los resultados derivados del análisis por PCR y la tinción con yodo de las mismas muestras con diferentes sintomatologías fueron de un 96%, de 728 muestras analizadas. En el caso del síntoma más característico de la enfermedad (moteado asimétrico) se obtuvo un 100% de coincidencia, usando la misma técnica. Sin embargo, es necesario considerar que esta es una técnica de diagnóstico indirecto, que no diagnostica el patógeno como tal. Por lo que sería necesario el uso de la PCR para el diagnóstico de plantas, en donde sea necesaria una mayor certeza, como es el caso de los viveros propagadores y comerciales de las empresas citrícolas.

2.1.2 Técnicas de componentes orgánicos volátiles ((VOCs)).

Reyes y Cevallos (2009) mencionan que es posible desarrollar ensayos rápidos, sensores o biosensores portátiles y económicos para detectar cambios en concentraciones de metabolitos específicos del HLB. Utilizaron técnicas metabolómicas basadas en cromatografía líquida de alta presión, Cromatografía de Gas, y electroforesis de capilar para encontrar metabolitos marcadores del HLB en hojas de naranjos ‘valencia’ sanas, infectadas con HLB y con deficiencia de zinc.

Los resultados mostraron que naringenina, cuercitina, hesperidina y L-prolina son aparentemente los metabolitos más indicados para detectar HLB.

Aksenov et al. (2014) establecen un nuevo método de detección de la enfermedad basado en el análisis químico de los compuestos orgánicos volátiles (COV) liberados que emanan de los árboles de cítricos infectados. En esta investigación muestran como la aplicación de métodos analíticos para analizar perfiles de VOC emitidos por los árboles es posible identificar el HLB.

Encontraron que los biomarcadores de huella digital son específicos para el patógeno causal y podría ser interpretado utilizando métodos analíticos como cromatografía de gas/espectrometría de masas (GC/MS) y cromatografía de gas/espectrometría de movilidad diferencial (GC / DMS). Aplicando esta técnica obtuvieron una precisión de 90% durante todo el año. Ellos afirman que usando este método en muestras sintomáticas y con una técnica de PCR en tiempo real se puede detectar la enfermedad del enverdecimiento de los cítricos a una edad temprana. Antes de la clasificación se aplicó una inspección visual, así como el análisis de componentes principales (PCA) para identificar valores atípicos debido a diversos factores.

Por la tecnología especializada y utilizada para la identificación del HLB en esta investigación, éste método, no es viable aplicarlo en un programa a gran escala para pequeños productores, debido a que, es necesario un laboratorio de química o bioquímica y requiere de conocimientos técnicos de estas áreas necesariamente. Esta metodología está en etapa de desarrollo, por lo tanto, se deben realizar más investigaciones sobre esta técnica para comparar los resultados. El costo de la técnica depende de la precisión deseada.

2.1.3 Técnicas de imágenes y Espectroscopia.

En los últimos años, investigadores de diversos centros de investigación, han desarrollado tecnología enfocada al área agrícola para la identificación de enfermedades en plantas. Se desatacan estos métodos por ser propuestas automatizadas no destructivas. Se busca detectar enfermedades a través de herramientas de una manera rápida, a una edad temprana y que se pueda aplicar de forma masiva en campo en tiempo real. Algunos de estos trabajos se mencionan a continuación.

Mishra, Karimi, Ehsani y Albrigo (2011) emplean un sensor óptico activo multibanda para la identificación de la enfermedad huanglongbing(HLB) en árboles de cítricos. Hace mediciones de la copa de los árboles. Este sensor se compone de cuatro iluminadores de banda estrecha con cuatro longitudes de onda multibanda diferentes; dos en la región visible y dos en la región infrarroja cercana. Se utilizó la técnica de Análisis de Componentes Principales (PCA) para la eliminación de datos atípicos en las muestras en un paso previo al análisis de datos.

Para la clasificación de árboles infectados con HLB se aplicaron las técnicas de árboles de decisión, regresión logística, k vecinos más cercanos, redes neuronales y máquinas de vectores de soporte.

Utilizando cinco mediciones de cada árbol, las máquinas de vectores de soporte obtuvieron un menor porcentaje de error de clasificación seguido por la técnica de árboles de decisión con un 2% y 3% correspondientemente.

Utilizando esta técnica, no fue posible discriminar árboles sanos e infectados en base a una sola medición de un árbol; sin embargo, el uso de múltiples mediciones de un árbol era posible lograr alta precisión en la clasificación de árboles enfermos.

Mishra, Karimi, Ehsani y Lee (2012) proponen un método para la identificación del HLB en árboles de cítricos usando una técnica de espectroscopia vis -nir (luz visible e infrarrojo cercano). Coleccionaron datos espectrales de árboles de naranja valencia de cuatro huertos de Florida. Adquirieron los datos de reflectancia espectral de la copa de los árboles utilizando dos espectroradiómetros. Utilizaron tres técnicas de clasificación para clasificar los datos: k vecinos más cercanos (KNN), regresión logística (RL) y máquinas de vectores soporte (SVM).

El inconveniente es, que utilizando sólo una observación de reflectancia de copa por árbol arroja resultados inadecuados en los tres enfoques estudiados, específicamente entre un 18% y 35%. Sin embargo, Cuando se utilizaron cinco espectros del mismo árbol para la clasificación, los métodos SVM y KNN espectros ponderados clasifican con un porcentaje de error del 3% y 6.5% respectivamente. Por lo tanto, esta propuesta requiere de múltiples mediciones de la copa de un solo árbol para una precisión mayor a 90%.

Además se utiliza equipo especializado que requiere personal capacitado con conocimientos técnicos para manejo y calibración. Peso de un dispositivo de este tipo es de 3 a 8 kg, esto no es tan manejable en campo.

Pourreza, Lee, Raveh, Hong y Kim (2013) en su investigación , el objetivo fue evaluar un sistema de detección basado en visión artificial para la identificación del HLB en hojas sintomáticas de esta enfermedad, identificación de hojas sanas y de hojas con deficiencias nutrimentales. Utilizaron filtros polarizados tanto en el sistema de iluminación como para la adquisición de imágenes para resaltar los síntomas de la enfermedad. Se basaron en que las imágenes de las hojas con síntomas de HLB tienen alta concentración de almidón y se pueden enfatizar a través de una banda de iluminación estrecha y un filtro polarizador.

Se basaron en los resultados de la técnica molecular PCR, y categorizaron las muestras en cinco grupos o clases de hojas: sanas, con síntomas de HLB, deficientes de zinc, deficientes de magnesio y una que cumplían con dos características deficientes de Zinc y sintomáticas de HLB. Una vez teniendo los datos, les aplicaron la técnica de Análisis de Componentes principales(PCA). Utilizaron los clasificadores lineal, naive bayes lineal, Mahalanobis, cuadrática, naive bayes cuadrática, Máquinas de vectores de soporte (SVM) y K-Vecinos más cercanos (KNN) para evaluar los datos.

Enfoques de diagnóstico como prueba de PCR , la medición de almidón y de exploración de cultivos, no son absolutamente precisas y no hay método de detección cien por ciento preciso que se ha informado todavía. Por lo tanto, la evaluación del método sugerido fue sin duda afectada por esta imprecisión. Aunado a ello se requiere de equipo especial, personal capacitado y de un sistema de iluminación controlado.

Pourreza, Lee y Ehsani (2014) desarrollan en este estudio, una nueva máquina de visión basada en sensor para la identificación del HLB a través de hojas sintomáticas bajo condiciones de campo controlado. Para ello, emplearon la combinación de una cámara en blanco y negro, con un sistema de iluminación de banda estrecha polarizada. A través de este sensor resaltan la acumulación de almidón en las hojas sintomáticas de los cítricos infectados por HLB además de diferenciarlas de las hojas amarillas causadas por otras situaciones como deficiencias nutrimentales. El sensor fue probado bajo condiciones simuladas y en campo directamente con tres clases de hojas (HLB sintomáticas, sanas y deficientes de zinc). De las imágenes extraen la media de escala de grises y la desviación estándar de cada imagen. Para la clasificación utilizaron características simples de histogramas estadísticos y la técnica de clasificación Support Vector Machine (SVM) con una precisión mayor a 95%.

Sin embargo el sensor de visión tiene un sistema de iluminación personalizado y fue diseñado para adquirir imágenes después de la puesta de sol para un verdadero experimento en el campo. Por lo tanto, para este experimento, las imágenes fueron adquiridas en un cuarto oscuro para que las muestras sólo reflejen la luz polarizada que recibieron del sistema de iluminación del sensor de visión.

El prototipo fue construido con materiales baratos, sin embargo, no incluye el costo de mano de obra, mercadotecnia, distribución entre otros. Un productor de cítricos pequeño necesita de conocimientos técnicos para producirlo en casa lo que lo hace imposible de usarlo a gran escala.

Deng et al. (2014) proponen en su investigación, desarrollar un método de respuesta rápida y de bajo costo, capaz de reconocer la enfermedad del HLB en las hojas de los cítricos usando visión por computadora. En esta propuesta, se extrae el primer plano del objeto hoja de la imagen utilizando el modelo de densidad de mezcla gaussiana (GMD), se extraen sus características y finalmente reconocimiento de la presencias del HLB en la hoja basada en árbol de vocabulario escalable (SVT), para generar el SVT provee un método de agrupamiento de descriptores locales.

Para evaluar la eficiencia del método propuesto, se creó un conjunto de datos (dataset) de 216 imágenes capturadas con cámaras móviles y cámaras DSLR, divididas en cinco categorías de hojas: Hojas sanas, hojas con HLB, hojas subsaludables de HLB por los tratamientos aplicados, hojas amarillas y hojas con deficiencia de zinc.

El método SVT, se compara con los métodos Baseline SVT (Nister & Stewenius, 2006) y BoW (Yap, Chen, Li y Wu, 2010). En la clasificación de las cinco categorías de hojas, se observa, que los métodos SVT producen alto reconocimiento, superando al método BoW(80%). De igual manera, el método propuesto de escala ponderada SVT(cercana al 100%) supera a Baseline SVT(88%). En general el método propuesto, muestra una mejora del 10% en promedio con respecto a los otros métodos.

En cuanto al tiempo de reconocimiento por imagen usando las técnicas manuales, K-means, baseline SVT y el método propuesto fue de más de 300, 4.663, 0.329 y 0.331 segundos correspondientemente. Hay un incremento muy pequeño del método propuesto con respecto al Baseline SVT.

Los resultados experimentales muestran que el método de reconocimiento propuesto con GMD para la extracción de objetos alcanza una precisión de 95-100% en 1 segundo.

Mota et al. (2014) desarrollan un método para la identificación del HLB usado una plataforma portable a través de espectroscopia fluorescente. Para ello, utilizaron una fuente de excitación LED azul, un espectrómetro comercial en miniatura así como software embebido. Los datos espectrales son adquiridos de hojas sanas; con HLB sintomáticas; con HLB asintomáticas. Las hojas fueron excitadas por un LED azul y sus espectros fluorescentes fueron coleccionados. Electrónica y SW embebido procesaron el espectro y la clasificación a través de regresión por mínimos cuadrados parciales. Siendo posteriormente analizadas por un software basado en redes neuronales. El objetivo principal es detectar si una nueva muestra está infectada de HLB o es una muestra de un árbol sano.

En pruebas de laboratorio tuvo una precisión del 82%, sin embargo, en condiciones de campo real, tuvo una precisión del 65.6%. Por otra parte, los dispositivos que utiliza no son de uso común y requieren una inversión y capacitación especial para su uso.

Kumar et al. (2012) En esta investigación hacen uso de imágenes aéreas hiperespectrales y espectrales para detectar rápidamente posibles árboles enfermos en una gran área mediante firmas espectrales. Dos conjuntos de imágenes hiperespectrales fueron adquiridos en 2007 y 2009, a partir de diferentes plantaciones de cítricos en Florida. Las imágenes multiespectrales fueron adquiridas en 2009.

Utilizaron imágenes hiperespectrales, observaciones de las mediciones en campo, imágenes de una librería espectral, MTMF (mixture tuned matched filtering), mapeo del ángulo espectral (SAM) y métodos LSU (linear spectral unmixing) a través de un software para la detección de áreas infectadas de HLB.

Para validar los resultados de las imágenes del 2007 utilizaron una verificación en campo basada en mediciones en campo y chequeos visuales en árboles de cítricos; los del 2009 se validaron con PCR. Utilizaron el SW para imágenes hiperespectrales (ENVI, ITT VIS) para el análisis.

Para imágenes del 2007, se utilizó el gráfico de dispersión n- D para el Análisis MTMF donde no todos los píxeles de vegetación fueron espectralmente puros, y los valores de píxel variaron desde el lado izquierdo a la parte derecha de una fila de árboles a través de un dosel. Debido a la similitud en el espectro de los pixeles de árboles sanos y enfermos de HLB en los bordes del dosel, los análisis de MTMF y SAM identificaron pixeles sanos como si estuvieran enfermos. Con SAM se observó una precisión global del 60%.

Existe una clara posibilidad de imprecisión en los datos al verificarlos en campo debido a errores de georeferenciación. Los autores sugieren utilizar Métodos de corrección atmosférica Better para el cuidado de la varianza de iluminación y la normalización de píxeles del borde del dosel para obtener mejores resultados.

Para el sitio seleccionado en 2009, se usó PCR para verificación de árboles sanos y enfermos de HLB en campo, adicionalmente se utilizó un espectrómetro manual para tomar muestras de árboles sanos y enfermos de HBL como una alternativa a los resultados PCR para validarlos. Registrándose una precisión de detección del 80% usando MTMF en imágenes hiperespectrales. SAM con imágenes multiespectrales registró una precisión del 87%. Observándose que Las imágenes multiespectrales dieron mejores resultados de detección que las hiperespectrales.

2.2 Conclusiones.

Se han conocido diversas propuestas de solución a esta problemática, a cargo de varios investigadores de diferentes países y centros de investigación. Estas propuestas se pueden clasificar en técnicas moleculares, técnicas de identificación de componentes orgánicos volátiles y a través de la espectroscopía y técnicas de imágenes. En la *Tabla 1* se mencionan las tecnologías utilizadas en estas investigaciones. La presente propuesta se encuentra dentro de las técnicas de reconocimiento a través de tratamiento y análisis de imágenes, la idea es usar equipos convencionales de bajo costo debido a que será una herramienta de identificación del HLB para los pequeños productores de limón persa.

Tabla 1

Comparación de las tecnologías utilizadas en investigaciones similares.

Investigación	**Tecnología utilizada**	**Preprocesamiento Extracción de características**	**Algoritmo de clasificación**
Mishra et al. (2011)	Sensor óptico multibanda Dos en región visible y dos en región infrarroja.	PCA	Árboles de decisión Regresión logística K vecinos más cercanos Redes neuronales Máquinas de vectores de soporte
Mishra et al. (2012)	Espectroscopia VIS-NIR Luz visible e infrarroja	PCA	KNN Regresión Logística Máquinas de vector de soporte.
Pourreza et al. (2013)	Cámara monocromática Filtros polarizados Banda de a	PCA	Lineal Naive bayes lineal Mahalanobis SVM Cuadrática Naive bayes cuadrática KNN
Pourreza et al. (2014)	Sensor prototipo (Cámara monocromática E iluminación de banda estrecha polarizada (95%)	Descriptores: promedio de escala de grises y desviación estándar	Histogramas estadísticos y SVM
Deng et al. (2014)	Cámara móvil Cámara DSRL (95%)	Densidad de mezcla gaussiana Descriptores locales	Árbol de vocabulario escalable histogramas
Mota et al. (2014)	Espectroscopia fluorescente Led azul Espectrómetro SW embebido (65.6%)	Regresión por mínimos cuadrados parciales	Red es neuronales
Kumar et al. (2012)	firmas espectrales	hyperspectral imaging software (ENVI, ITT VIS) (MNF)fracción mínima de ruido	(MTMF) Filtración de coincidencias de mezclas sintonizadas. 80% (SAM)Mapeo de ángulo espectral 87% Modelo lineal espectral sin mezcla

CAPÍTULO III PROPUESTA DEL MODELO

3.1 Introducción.

En esta investigación, se busca identificar la enfermedad huanglongbing en los árboles de limón persa, utilizando imágenes de hojas que presentan características sintomáticas a la enfermedad o descartar que este mal se encuentre presente. Para ello, se propone seguir la siguiente metodología.

1. Adquisición de las imágenes.
2. Procesamiento de imágenes.
3. Segmentación (detección de bordes).
4. Reducción de la dimensionalidad, mediante la técnica de análisis de componentes principales.
5. Diseño de la red neuronal multicapa backpropagation.
    a. Entrenamiento
    b. Validación
    c. prueba

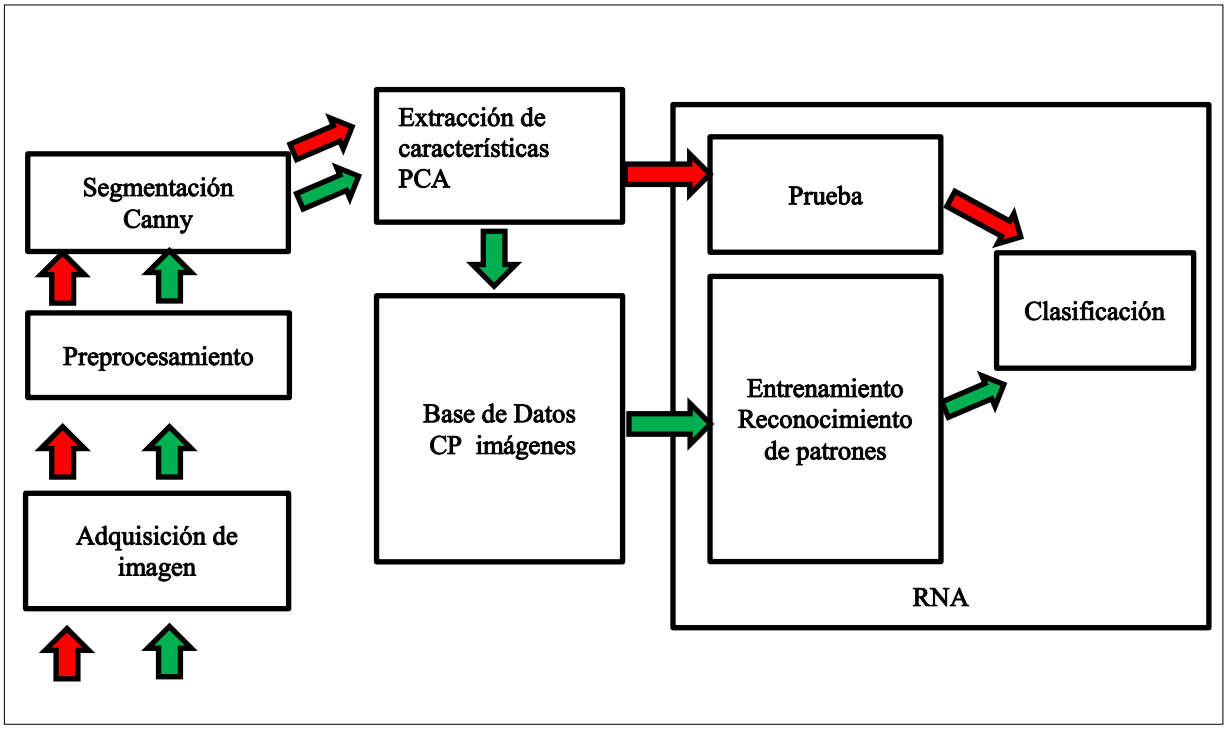

Figura 1. Modelo para la identificación del HLB.

3.2 Procesamiento de imágenes.

Una imagen es la representación visual de un objeto y está definida como una función bidimensional ***f(x,y)*** como se muestra en la *Figura 2*, donde ***x*** y ***y*** son coordenadas especiales y la amplitud f en un par de coordenadas **(x,y)** se llama la intensidad o nivel de gris de la imagen en ese punto. Cuando los valores de ***x*** y ***y***, así como su amplitud *f* son cantidades finitas y discretas de dice que es una imagen digital.

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

Figura 2. Representación de una imagen.

Las imágenes digitales se pueden clasificar en dos tipos principalmente, imágenes a color e imágenes en escala de grises.

Las primeras son representadas a través de una matiz de *m* X *n* X *p*; donde *n* representa el número de pixeles de ancho, *m* el número de pixeles de largo y *p* representa el nivel de rojo, verde y azul de un pixel de una imagen en formato RGB.

Una imagen en escala de grises, conocida también como escala de intensidades o escala monocromática se representa en una matriz de *m* X *n* valores en donde cada pixel es una sola muestra que contiene la información de la intensidad de la imagen.

Se utilizarán en esta investigación imágenes en escala de grises por ocupar menos espacio y principalmente por ser el formato apropiado para poder obtener mejor sus características.

En esta fase las imágenes se les pueden aplicar diversas operaciones, como entrada se proporciona una imagen y como salida obtenemos una nueva versión de esa imagen. El objetivo de aplicar estas operaciones, es mejorar las condiciones o calidad de estas imágenes, así como obtener de ellas ciertos datos de interés para su uso posterior.

Las imágenes se pueden adquirir con una cámara de celular, cámara fotográfica o una tableta electrónica. El siguiente paso es configurar la resolución de la imagen a 200 X 350 pixeles. Otra operación muy importante es la modificación del brillo para corregir algunas sombras y ruidos adquiridos durante la adquisición de las imágenes. Resultados de estas operaciones se muestran en las *Figura 3* y *Figura 4*.

Figura 3. Hoja a color con HLB.

Figura .4. Hoja a color sin HLB.

3.3 Segmentación

La segmentación es un proceso por el cual se extrae cierta información de la imagen para ser utilizada más adelante. La segmentación está basada en dos principios fundamentales: discontinuidad y similitud (Fu & Mui, 1981). Cabe pues enfocar la segmentación orientada a bordes (discontinuidad) y orientada a regiones (similitud).

3.3.1 Detección de Bordes

Los bordes son pixeles alrededor de los cuales la imagen presenta una brusca variación en los niveles de gris (González & Woods, 2002). El objetivo consiste en dada una imagen, localizar los bordes más probables generados por los elementos de la escena y no por el ruido.

La detección de los bordes de una imagen es de suma importancia y utilidad para esta investigación, debido a que facilita el reconocimiento de objetos, la segmentación de regiones, entre otras operaciones. Para ello, existen diversos algoritmos como Roberts, Prewitt, Sobel, Canny entre otros.

Se utilizará en esta fase el algoritmo de Canny para este fin. Este algoritmo está considerado como uno de los mejores métodos de detección de contornos mediante el empleo de máscaras de convolución y basado en la primera derivada. Los puntos de contorno son como zonas de píxeles en las que existe un cambio brusco de nivel de gris. En el tratamiento de imágenes, se trabaja con píxeles, y en un ambiente discreto, es así que en el algoritmo de Canny se utiliza máscaras, las cuales representan aproximaciones en diferencias finitas.

Canny (1986) propuso un método para la detección de bordes, este algoritmo consiste en los siguientes tres grandes pasos:

- Obtención del gradiente: en este paso se calcula la magnitud y orientación del vector gradiente en cada píxel.
- Supresión no máxima: en este paso se logra el adelgazamiento del ancho de los bordes, obtenidos con el gradiente, hasta lograr bordes de un píxel de ancho.
- Histéresis de umbral: en este paso se aplica una función de histéresis basada en dos umbrales; con este proceso se pretende reducir la posibilidad de aparición de contornos falsos.

3.3.2 Obtención del gradiente

Para la obtención del gradiente, lo primero que se realiza es la aplicación de un filtro gaussiano a la imagen original con el objetivo de suavizar la imagen y tratar de eliminar el posible ruido existente. Sin embargo, se debe de tener cuidado de no realizar un suavizado excesivo, pues se podrían perder detalles de la imagen y provocar un indeseable resultado final. Este suavizado se obtiene promediando los valores de intensidad de los píxeles en el entorno de vecindad con una máscara de convolución de media cero y desviación estándar *s*. En la *Figura 5* se muestran dos ejemplos de máscaras que se pueden usar para realizar el filtrado gaussiano. Una vez que se suaviza la imagen, para cada píxel se obtiene la magnitud y módulo (orientación) del gradiente, obteniendo así dos imágenes. El algoritmo para este primer paso se describe a continuación.

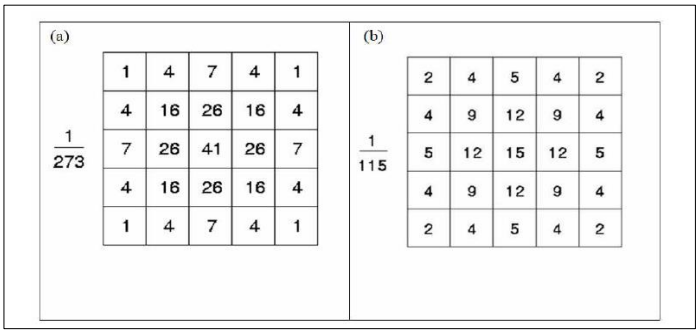

Figura 5. Máscaras de convolución recomendadas para obtener el filtro gaussiano.

3.3.3 Supresión no máxima al resultado del gradiente

Las dos imágenes generadas en el paso anterior sirven de entrada para generar una imagen con los bordes adelgazados. El procedimiento es el siguiente: se consideran cuatro direcciones identificadas por las orientaciones de 0º, 45º, 90º y 135º con respecto al eje horizontal. Para cada píxel se encuentra la dirección que mejor se aproxime a la dirección del ángulo de gradiente.

Algoritmo: Obtención de Gradiente

Entrada: imagen I

Máscara de convolución H, con media cero y desviación estándar σ .

Salida: imagen E_m de la magnitud del gradiente

Imagen E_0 de la orientación del gradiente

1. Suavizar la imagen I con H mediante un filtro gaussiano y obtener J como imagen de salida.

2. Para cada píxel (i, j) en J, obtener la magnitud y orientación del gradiente basándose en las siguientes expresiones:

El gradiente de una imagen f(x,y) en un punto (x,y) se define como un vector bidimensional dado por la ecuación:

$$G[f(x,y)] = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial_x} f(x,y) \\ \frac{\partial}{\partial_y} f(x,y) \end{bmatrix}$$

Siendo un vector perpendicular al borde, donde el vector G apunta en la dirección de variación máxima de *f* en el punto (x,y) por unidad de distancia, con la magnitud y dirección dadas por:

$$|G| = \sqrt{G_x^2 + G_y^2} = |G_x| + |G_y|,$$

$$\phi(x,y) = \tan^{-1}\frac{G_y}{G_x}$$

3. Obtener E_m a partir de la magnitud de gradiente y E_o a partir de la orientación, de acuerdo a las expresiones anteriores.

Posteriormente se observa si el valor de la magnitud de gradiente es más pequeño que al menos uno de sus dos vecinos en la dirección del ángulo obtenida en el paso anterior. De ser así se asigna el valor 0 a dicho píxel, en caso contrario se asigna el valor que tenga la magnitud del gradiente.

La salida de este segundo paso es la imagen I_n con los bordes adelgazados, es decir, $E_m(i\,j)$, después de la supresión no máxima de puntos de borde.

3.3.4 Histéresis de umbral a la supresión no máxima.

La imagen obtenida en el paso anterior suele contener máximos locales creados por el ruido. Una solución para eliminar dicho ruido es la histéresis del umbral.

El proceso consiste en tomar la imagen obtenida del paso anterior, tomar la orientación de los puntos de borde de la imagen y tomar dos umbrales, el primero más pequeño que el segundo. Para cada punto de la imagen se debe localizar el siguiente punto de borde no explorado que sea mayor al segundo umbral. A partir de dicho punto seguir las cadenas de máximos locales conectados en ambas direcciones perpendiculares a la normal del borde siempre que sean mayo res al primer umbral. Así se marcan todos los puntos explorados y se almacena la lista de todos los puntos en el contorno conectado. Es así como en este paso se logra eliminar las uniones en forma de Y de los segmentos que confluyan en un punto.

Algoritmo: Supresión no máxima

Entrada: imagen E_m de la magnitud del gradiente

imagen E_o de la orientación del gradiente

Salida: imagen I_n

Considerar: cuatro direcciones d_1 , d_2, d_3, d_4 identificadas por las direcciones de 0º, 45º, 90º y 135º con respecto al eje horizontal

- Para cada píxel (i, j) :
    - Encontrar la dirección d_k que mejor se aproxima a la dirección E_o(i, j), que viene a ser la perpendicular al borde.
    - Si E_m (i, j) es más pequeño que al menos uno de sus dos vecinos en la dirección d_k, al píxel (i, j) de I_n se le asigna el valor 0, $I_n(i, j) = 0$ (supresión), de otro modo $I_n(i,j)$ $E_m(i,j)$.
- Devolver I_n

3.3.5 Un cuarto paso

Frecuentemente, es común que un cuarto y último paso se realice en el algoritmo de Canny, este paso consiste en cerrar los contornos que pudiesen haber quedado abiertos por problemas de ruido. Un método muy utilizado es el algoritmo de Deriche y Cocquerez. Este algoritmo utiliza como entrada una imagen binarizada de contornos de un píxel de ancho. El algoritmo busca los extremos de los contornos abiertos y sigue la dirección del máximo gradiente hasta cerrarlos con otro extremo abierto.

El procedimiento consiste en buscar para cada píxel uno de los ocho patrones posibles que delimitan la continuación del contorno en tres direcciones posibles. Esto se logra con la convolución de cada píxel con una máscara específica. Cuando alguno de los tres puntos es ya un píxel de borde se entiende que el borde se ha cerrado, de lo contrario se elige el píxel con el valor máximo de gradiente y se marca como nuevo píxel de borde y se aplica nuevamente la convolución. Estos pasos se repiten para todo extremo abierto hasta encontrar su cierre o hasta llegar a cierto número de iteraciones determinado.

Algoritmo: Histéresis de umbral a la supresión no máxima

Entrada: imagen I_n obtenida del paso anterior

imagen E_o de la orientación del gradiente

umbral t_1

umbral t_2, donde $t_1 < t_2$

Salida: imagen *G* con los bordes conectados de contornos

- Para todos los puntos de I_n y explorando I_n en orden fijo:
    - Localizar el siguiente punto de borde no explorado previamente, *In*(*i*, *j*), tal que $I_n(i, j) > t_2$.
    - Comenzar a partir de I_n(i,j), seguir las cadenas de máximos locales conectados en ambas direcciones perpendiculares a la normal de borde, siempre que $I_n > t1$.
    - Marcar todos los puntos explorados y, salvar la lista de todos los puntos en el entorno conectado encontrado.
- Devolver G formada por el conjunto de bordes conectados de contornos de la imagen, así como la magnitud y orientación, describiendo las propiedades de los puntos de borde.

Al aplicar este algoritmo a las hojas en escala de grises mostradas en la *Figura 6* y *Figura 7*, da como salida la misma hoja pero donde se manifiestan solo los bordes identificados por el algoritmo de Canny tal como se muestra en la *Figura 8 y Figura 9* correspondientemente**.**

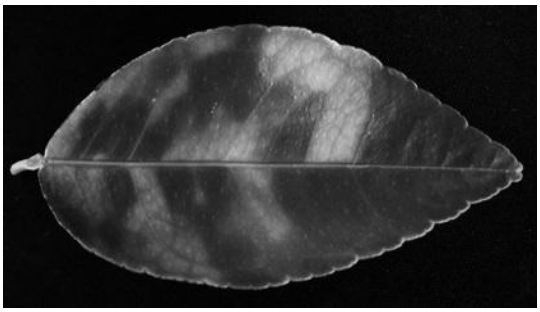

Figura 6. Hoja con HLB en escala de grises.

Figura 7. Hoja sin HLB en escala de grises.

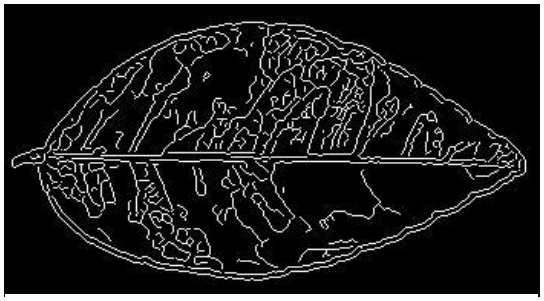

Figura 8. Hoja con HLB + Canny.

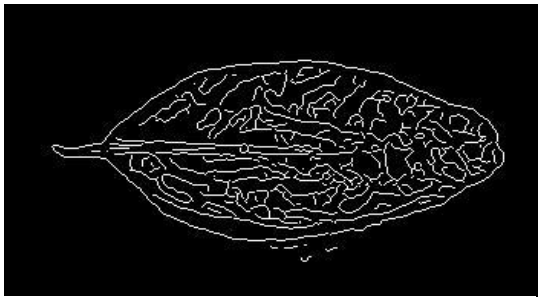

Figura 9. Hoja sin HLB + Canny.

3.4 Reducción de la Dimensionalidad.

Una de las desventajas del tratamiento de imágenes, es la gran cantidad de información que utilizan para su almacenamiento, representación o manipulación en formato digital. Las imágenes que se utilizarán en esta investigación son del tipo escala de grises, como se ha mencionado anteriormente, una imagen de este tipo se almacena en una matriz de *m* X *n* donde *m* son las filas y *n* las columnas de la matriz, donde cada posición de esta matriz se almacenará una intensidad de gris normalmente un valor entre 0 y 255, siendo 0 el gris más claro y 255 el gris más oscuro. Aunque las imágenes de este tipo son de tamaño menor que las de color, es importante destacar, que, siguen siendo de un tamaño muy elevado, siendo la resolución de la imagen la que define el tamaño de una imagen de esta clase. Por ejemplo, si se tiene una imagen de 198 X 355, se deben almacenar y manipular para cualquier operación con esta la cantidad de 70290 datos. Esto conlleva, a que los procesos futuros como el diseño y entrenamiento de la red neuronal así como la etapa de reconocimiento de patrones, sean muy lenta o inmanejables en algunos casos.

Para mejorar el rendimiento en el tratamiento y análisis de imágenes, existen diversas técnicas que reducen la dimensión o tamaño de las imágenes; en el presente trabajo, se aplicará la técnica de análisis de componentes principales (PCA), debido a que es una de las técnicas más utilizadas para este fin.

El análisis de componentes principales es una técnica de extracción de datos ampliamente utilizada en la actualidad. El objetivo principal que persigue dicha herramienta es reducir la dimensionalidad de un conjunto de observaciones con una gran cantidad de variables, ayudándose del estudio de la estructura de varianzas-covarianzas entre las variables que componen los datos de entrada. A partir de la proyección de los datos de entrada sobre las

direcciones de máxima varianza se obtendrá un nuevo espacio de representación de los datos en el que se puede eliminar fácilmente aquellos componentes con menor varianza, garantizando la mínima perdida de información (Sánchez, 2012).

El PCA, permite reducir un conjunto de variables altamente correlacionadas en un conjunto menor de nuevas variables, llamadas componentes principales, y estas, no se hallan correlacionadas entre sí.

Estadísticamente, el PCA tiene como objetivo reducir la dimensión original de un conjunto de *p* variables en un conjunto menor de *m* variables, a través de una combinación lineal de los datos originales que permitan lograr una mayor interpretación de la información disponible.

Cada imagen seleccionada para realizar este proyecto, además del tratamiento y segmentación, se le aplicará la técnica de análisis de componentes principales para reducir la dimensionalidad de estos datos. En esta etapa, se tomarán los cien primeros componentes principales que representarán más del noventa por ciento de la información original. Estas componentes, se almacenarán en una matriz, como se muestra en la *Figura 10*, donde cada columna representará los componentes principales de una imagen en particular.

Imágenes

Componentes

	H1	**H2**	**. . .**	**Hn**
Cp1	0.00125959809771935	0.000768769852153369	**. . .**	0.00120278679970731
. . .	. . .	. . .	. . .	. . .
Cp2	0.00950275265891108	0.00275670033908423	**. . .**	0.0200968838379384
• • •	• • •	• • •	. . .	• • •
Cp100	0.00415209667999859 . . .	0.0130510027093498 . . .		0.0117608123093509 . . .

Figura 10. Matriz de componentes principales de hojas de limón persa enfermas de HLB y sanas.

Es decir, si el grupo de imágenes de entrenamiento estuviese constituido por 50 elementos, entonces, resultará una matriz de tamaño FILAS x COLUMAS, donde:

FILAS = número de componentes X número de columnas de la imagen original.

COLUMNAS = número de imágenes;

FILAS representa los componentes principales por imagen y COLUMNAS representan cada una de las imágenes de este conjunto.

3.5 Diseño del clasificador.

Existen diversos algoritmos cuya función es la clasificar, entre ellos se encuentran el ID3, K-vecinos más cercanos, redes neuronales, entre otras. Para el reconocimiento de patrones se han utilizado principalmente las redes neuronales artificiales. Existen gran variedad de tipos de redes neuronales artificiales, normalmente se clasifican según su arquitectura (neuronas de entrada, capas ocultas, cantidad de neuronas en sus capas ocultas y número de neuronas en su capa de salida) o por su tipo de aprendizaje utilizado (supervisado y no supervisado).

En este trabajo de investigación se diseñará una red neuronal artificial tipo backpropagation con aprendizaje supervisado.

3.5.1 Red Neuronal Multicapa Backpropagation.

Rumelhart, Hinton y Williams (1986), formalizaron un método para que una red neuronal aprendiera la asociación que existe entre los patrones de entrada y las clases correspondientes, utilizando varios niveles de neuronas.

El método backpropagation (propagación del error hacia atrás), basado en la generalización de la regla delta, a pesar de sus limitaciones, ha ampliado de forma considerable el rango de aplicaciones de las redes neuronales. El funcionamiento de la red backpropagation (BPN) consiste en el aprendizaje de un conjunto predefinido de pares de entradas-salidas dados como ejemplo: primero se aplica un patrón de entrada como estímulo para la primera capa de las neuronas de la red, se va propagando a través de todas las capas superiores hasta generar una salida, se compara el resultado en las neuronas de salida con la salida que se desea obtener y se calcula un valor de error para cada neurona de salida. A continuación, estos errores se transmiten hacia atrás, partiendo de la capa de salida hacia todas las neuronas de la capa intermedia que

contribuyan directamente a la salida. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido un error que describa su aportación relativa al error total. Basándose en el valor del error recibido, se reajustan los pesos de conexión de cada neurona, de manera que en la siguiente vez que se presente el mismo patrón, la salida esté más cercana a la deseada (Valencia, Yáñez & Sánchez, 2006).

3.5.2 Entrenamiento de la red.

En esta etapa se inicializarán los pesos de la red con valores aleatorios. Se debe proporcionar un patrón de entrada y especificar la salida deseada que debe generar la red. En este caso, como entrada se proporcionará a la red neuronal los componentes principales de una imagen cada vez. Como salida se le asignará un identificador de clase uno correspondiente a las hojas que presentan la enfermedad HLB o clase dos correspondiente a las hojas sin esa enfermedad.

3.5.3 Matriz de confusión.

La matriz de confusión, es una herramienta o instrumento utilizado para evaluar la sensibilidad y especificidad de los resultados generados por un algoritmo de clasificación, en el campo de la inteligencia artificial. Esta es una matriz cuadrada de n x n, donde n es el número de clases, como se muestra en la *Tabla 2*. En una matriz de confusión las columnas corresponden a los datos reales de referencia, mientras que las filas corresponden a las asignaciones del clasificador.

Tabla 2
Elementos de una Matriz de confusión de 2 X 2.

	Enfermedad		
Resultados prueba	**Enfermos HLB (+)**	**Sanos (-)**	**Totales**
Positivos HLB	Verdaderos positivos (a)	Falsos Positivos (c)	**Positivos a+c**
Negativos	Falsos Negativos (b)	Verdaderos negativos (d)	**Negativos b+d**
Total	**Enfermos a+b**	**Sanos c+d**	**N**

La sensibilidad de una prueba es la proporción de individuos u objetos enfermos clasificados como positivos; en otras palabras, es la probabilidad de clasificar correctamente a un objeto enfermo. Se utiliza la siguiente fórmula para calcularla:

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{total de casos positivos}} = \frac{a}{a+b} = \frac{VP}{VP+FN} X100$$

Donde:

a=verdaderos positivos

b=falsos negativos

a+b=total de casos positivos

VP=verdaderos positivos

FN=falsos negativos.

Por otra parte, la especificidad, es la proporción de individuos u objetos sanos clasificados como negativos; en otras palabras, es la probabilidad de clasificar correctamente a un objeto sano. Para obtenerla se utiliza la siguiente fórmula:

$$especificidad = \frac{verdaderosnegativos}{totaldecasosnegativos} = \frac{d}{d+c} = \frac{VN}{VN+FP} X100$$

Donde:

d=verdaderos negativos

c=falsos positivos

d+c=total de casos negativos

VN=verdaderos negativos

FP=falsos positivos.

La precisión de una prueba, es la proporción de predicciones correctas, tanto positivas como negativas. Está dada por la fórmula siguiente:

$$precisión = \frac{VP+VN}{VP+VN+FP+FN} X100$$

3.5.4 Validación de la red neuronal.

La validación del modelo es un una fase muy importante es este proyecto, se aplicará la validación cruzada usando K grupos (K-fold cross-validation) y validación cruzada dejando uno fuera (Leave-one-out cross-validation LOOCV) (James, Witten, Hastie & Tibshirani, 2013)

Pérez-Planells, Delegido, Rivera-Caicedo y Verrelst (2015) mencionan que en la investigación de Yang y Huang (2014) recomiendan usar el método k-fold cuando el conjunto de datos es pequeño. En este caso, el total de los datos se dividen en *k* subconjuntos, de manera que se aplicará el método *hold-out k* veces, utilizando cada vez, un subconjunto distinto para validar el modelo entrenado con los otros *k*-1 subconjuntos.

En la validación cruzada de K iteraciones o K-fold cross-validation los datos se dividen en K subconjuntos (folds). Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento.

El proceso de validación cruzada es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos de prueba.

La validación cruzada dejando uno fuera o Leave-one-out cross-validation (LOOCV) implica separar los datos de forma que para cada iteración tengamos un solo dato de prueba y todo el resto de los datos para entrenamiento. Este proceso se realiza durante n iteraciones. El dato de prueba puede ser aleatorio o en forma ordenada, desde el primer dato hasta él dato n. Se seleccionó la ordenada en este trabajo de investigación.

Por otra parte, debido a que, la muestra que se utilizará para el entrenamiento de la red es menor a treinta, se va a utilizar la t-student para validad los resultados.

IV IMPLEMENTACIÓN DEL MODELO

El modelo propuesto en esta investigación para la identificación del Huanglongbing, se ha descrito en la sección anterior. Ahora, se mostrarán los resultados obtenidos, así como la interpretación de estos. Para ello, se describirán los resultados en el orden de aplicación de cada parte de este modelo. Empezando con el los medios utilizados para la obtención de las imágenes, seguido del tratamiento de imágenes, la segmentación, el proceso de reducción de la dimensionalidad de los datos utilizados y finalmente las salidas mostradas por la red neuronal diseñada para la identificación del HLB.

4.1 Obtención de imágenes.

La base de datos de imágenes se obtuvo de dos fuentes diferentes. Las fotografías de hojas de limón persa con moteado asimétrico de HLB fueron proporcionadas por el comité estatal de sanidad vegetal del estado de Colima, mismas que fueron adquiridas en la campaña contra huanglongbing de los cítrico en ese estado. Las imágenes de hojas de limón persa sin la enfermedad del HLB fueron adquiridas en un huerto de la localidad de la espaldilla municipio de Misantla, Veracruz. En ambas situaciones se manejaron en un ambiente controlado. Para la adquisición de las fotografías de hojas sin la enfermedad se utilizó una cámara sony y un celular marca nokia con resolución de 640 X 480 pixeles.

4.2 Tratamiento de imágenes.

Las imágenes que se utilizaron en esta investigación, se le aplicaron operaciones básicas de tratamiento de imágenes. Estas operaciones se aplicaron con el objetivo de estandarizar el tamaño de las imágenes, cambiar las imágenes de color a escala de grises para percibir mejor los detalles de las hojas; para eliminar en la medida de lo posible las sombras que ocasionaban ruido se manipuló el brillo de las imágenes, con la intención de que se vieran más claras e identificar mejor los bordes.

Para estandarizar el tamaño se aplicó la operación de redimensión de resolución. El tamaño estándar seleccionado es de 200 X 350, también se manipuló el brillo necesario para poder difuminar sombras y bordes que no forman parte de la hoja. El resultado de ambas operaciones se muestra en la *Figura 11 (a) (b) (c) (d).*

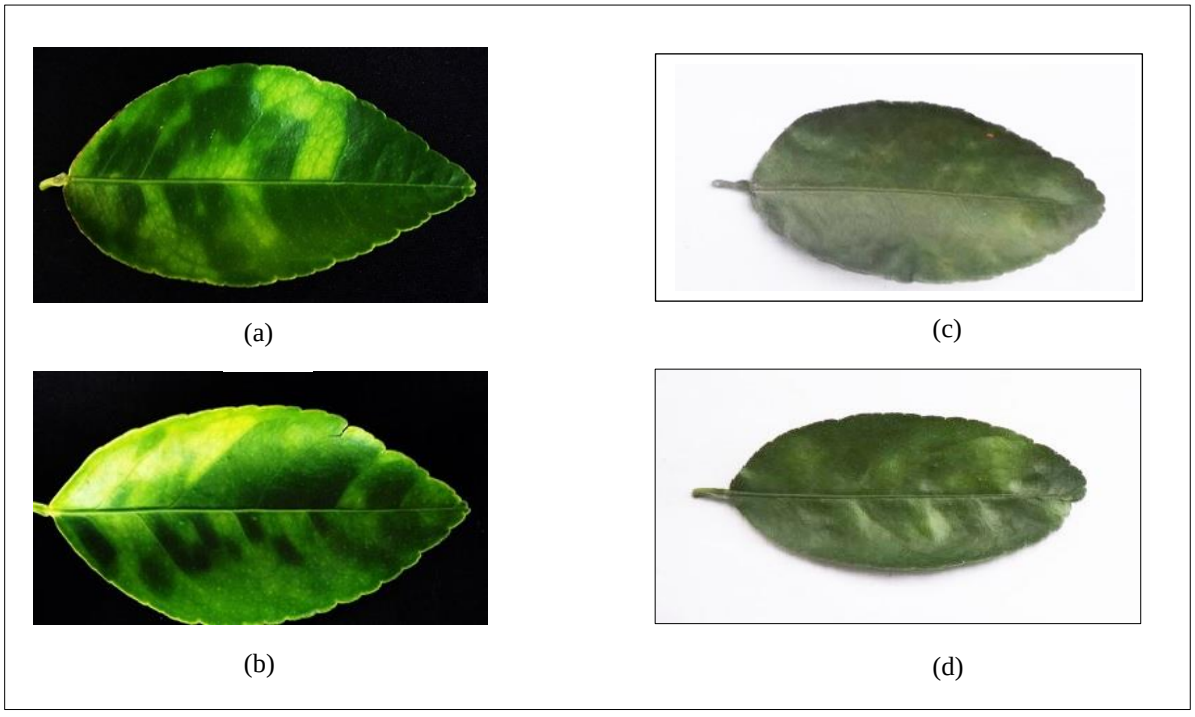

Figura 11. Hojas de limón persa estandarizadas con resolución de 200 X 350. (a) Hoja a color con HLB. (b) Hoja a color con HLB. (c) Hoja a color sin HLB. (d) Hoja a color sin HLB.

La salida a las operaciones anteriores, que son imágenes a color estandarizadas y con el brillo suficiente para su uso posterior, es la entrada a la siguiente operación, la conversión de imágenes a color a imágenes a escala de grises. Como muestra del resultado de este proceso se listan en la *Figura 12 (a) (b) (c) (b)* imágenes en escala de grises correspondientes a imágenes a color mostradas en la *Figura 11*.

Son imágenes con una intensidad de gris de 0 a 255 (256 niveles de gris). Este formato muestra detalles que a simple vista en una imagen a color no se perciben, esto mejora la identificación de bordes en el siguiente paso de este modelo.

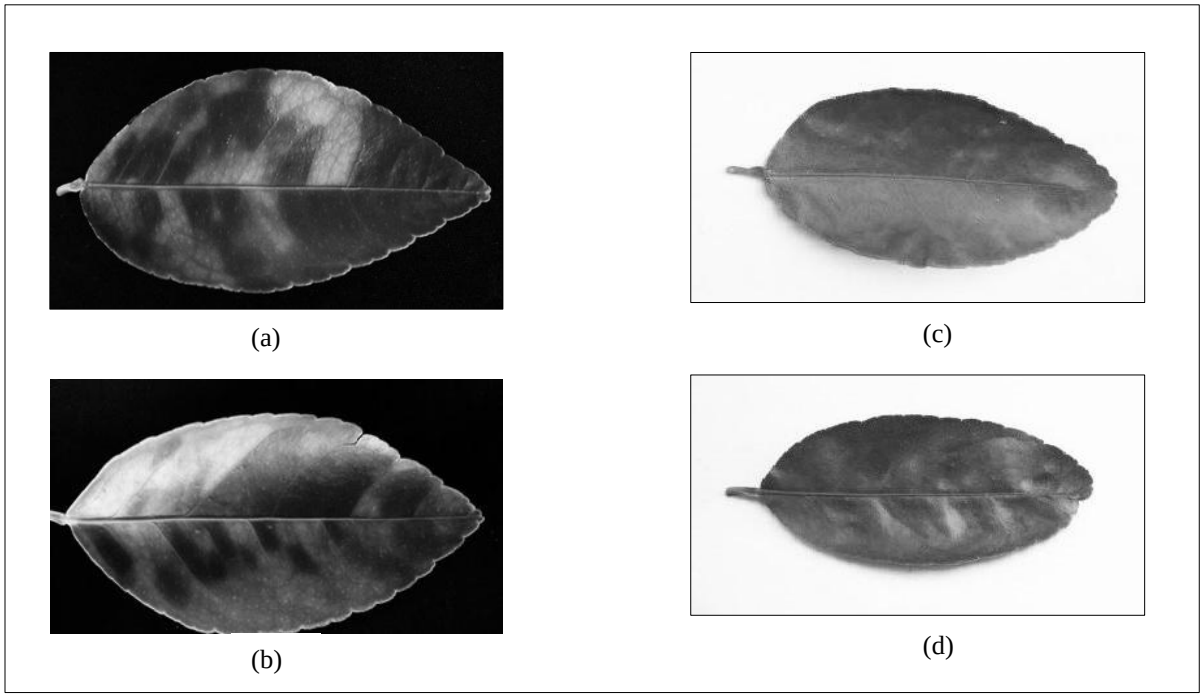

Figura 12. Hojas de limón persa en escala de grises estandarizadas con resolución de 200 X 350. (a) Hoja en escala de grises con HLB. (b) Hoja en escala de grises con HLB. (c) Hoja en escala de grises sin HLB. (d) Hoja en escala de grises sin HLB.

Las imágenes en escala de grises *(a) (b)* de las *Figura 12,* corresponden a hojas a color infectadas con HLB mostradas en la *Figura 11 (a) (b)* correspondientemente. Las hojas no

infectadas a color se muestran en las *Figura 11 (c) (d)* y su correspondiente en escala de grises en la *Figura 12 (c) (d).*

4.3 Segmentación.

La segmentación de imágenes divide la imagen en regiones u objetos de interés. Para realizarla se puede dividir la imagen basándose en cambios bruscos de nivel de gris (Detección de puntos aislados, detección de líneas, detección de bordes) o se puede dividir la imagen basándose en la búsqueda de zonas que tengan valores similares, conforme a unos criterios prefijados (Crecimiento de región o Umbralización). (Gonzalez & Woods, 1996).

Existen diversas técnicas para segmentar una imagen digital, dependiendo del área de interés, o los elementos importantes que deseamos resaltar o identificar es la operación u operaciones que debemos aplicar al objeto de estudio. En esta investigación, se deben identificar los bordes de una imagen, con el objetivo de visualizar zonas de la imagen que tengan patrones característicos de la enfermedad HLB. Existen varios algoritmos para la detección de bordes de una imagen, entre ellos Roberts, Prewitt, Sobel, Frei-Chen, laplaciano y el algoritmo de Canny.

El algoritmo seleccionado, para identificar los bordes en las imágenes en escala de grises de hojas de limón persa, es el algoritmo de Canny. En la *Figura 13 (a) (b)* se muestran los resultados obtenidos al aplicar este algoritmo a las dos hojas con HLB y las dos hojas sin HLB en la *Figura 13 (c) (d).*

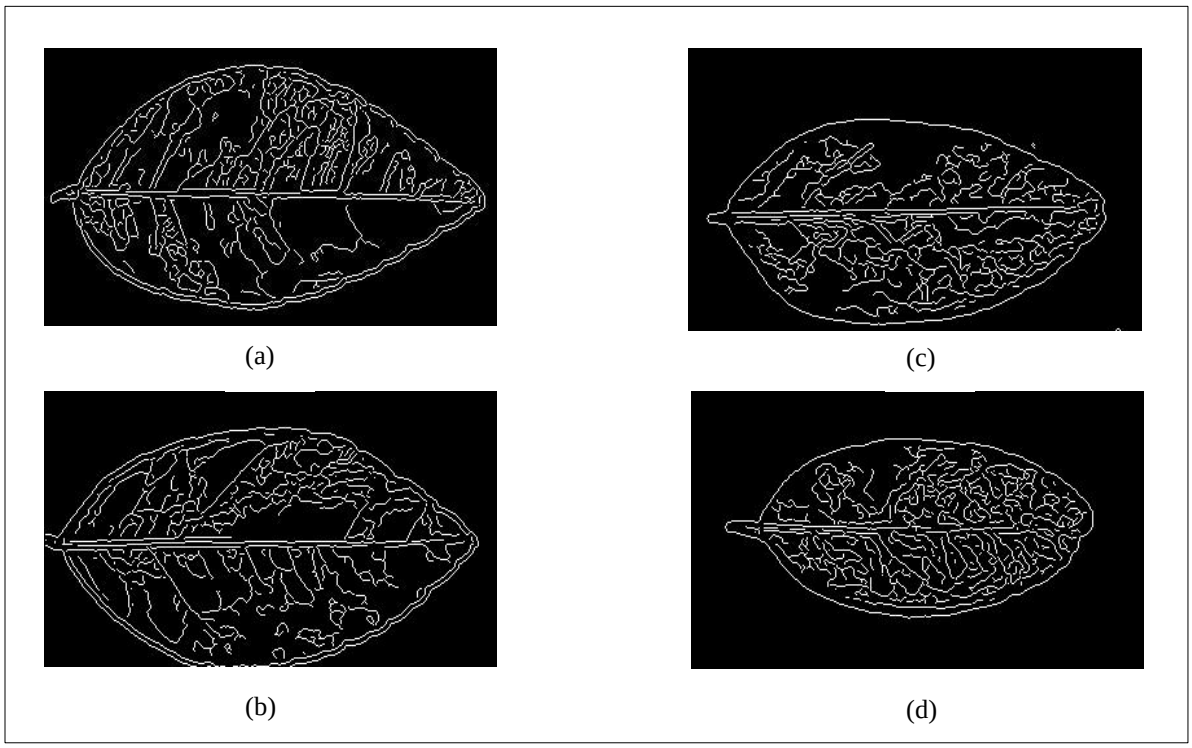

Figura 13. Hojas de limón persa segmentadas con el algoritmo de Canny. (a) Hoja con HLB+Canny. (b) Hoja con HLB+Canny. (c) Hoja sin HLB+Canny. (d) Hoja sin HLB+Canny.

En la salida del operador de detección de bordes Canny, se observan algunos patrones característicos de hojas enfermas y hojas libres de esta enfermedad. Se observa que en ambos tipos de hojas se manifiestan zonas o regiones sin bordes, mismas que representan una anomalía. Sin embargo, en las hojas infectadas con la bacteria que provoca la enfermedad HLB, las regiones que se observan son asimétricas con respecto a la nervadura central de la hoja.

4.4 Reducción de la dimensionalidad.

Existen diversas técnicas para reducir la dimensión de las variables. En esta investigación, se aplica el análisis de componentes principales (PCA) para transformar un conjunto de variables *originales*, en un nuevo conjunto de variables, llamadas componentes principales (CP). Los CP obtenidos, tienen como principal característica, el que no existe una correlación entre ellas. También el número de CP es menor al número de variables iniciales u originales.

Como se ha mencionado anteriormente, cada imagen que se utiliza en esta investigación, tiene una resolución de 200 X 350 pixeles. Por lo tanto, la cantidad de datos que se están tratando con estas especificaciones es de 70000. Cada dato o pixel representa una intensidad de nivel de gris del rango de 0 a 255.

Al aplicar el algoritmo de análisis de componentes principales a una imagen de 200 X 350, genera como salida una matriz de 350 X 350. Cada columna, representa un componente principal, es decir, genera 350 componentes y cada una consta de 350 datos (filas).

Hay algo importante que se debe mencionar, elegir pocos componentes se traduce en mayor reducción de la dimensionalidad de los datos, sin embargo, puede suceder que elegir pocos CP signifique que se está perdiendo información valiosa de la imagen.

Para demostrar gráficamente este comportamiento, se mostrarán los resultados al generar los componentes principales de las hojas que se han venido ilustrando en las fases anteriores.

La hoja infectada con HLB de la *Figura 13 (a)*, al calcular sus componentes principales con veinte, cincuenta, cien y doscientos componentes; y reconstruirla con el número de componentes antes mencionados, se observa, que mientras menos CP seleccionemos la pérdida de información es mayor, como se muestra claramente en la *Figura 14 (a) (b) (c) (d)* con veinte, cincuenta, cien y doscientos componentes correspondientemente.

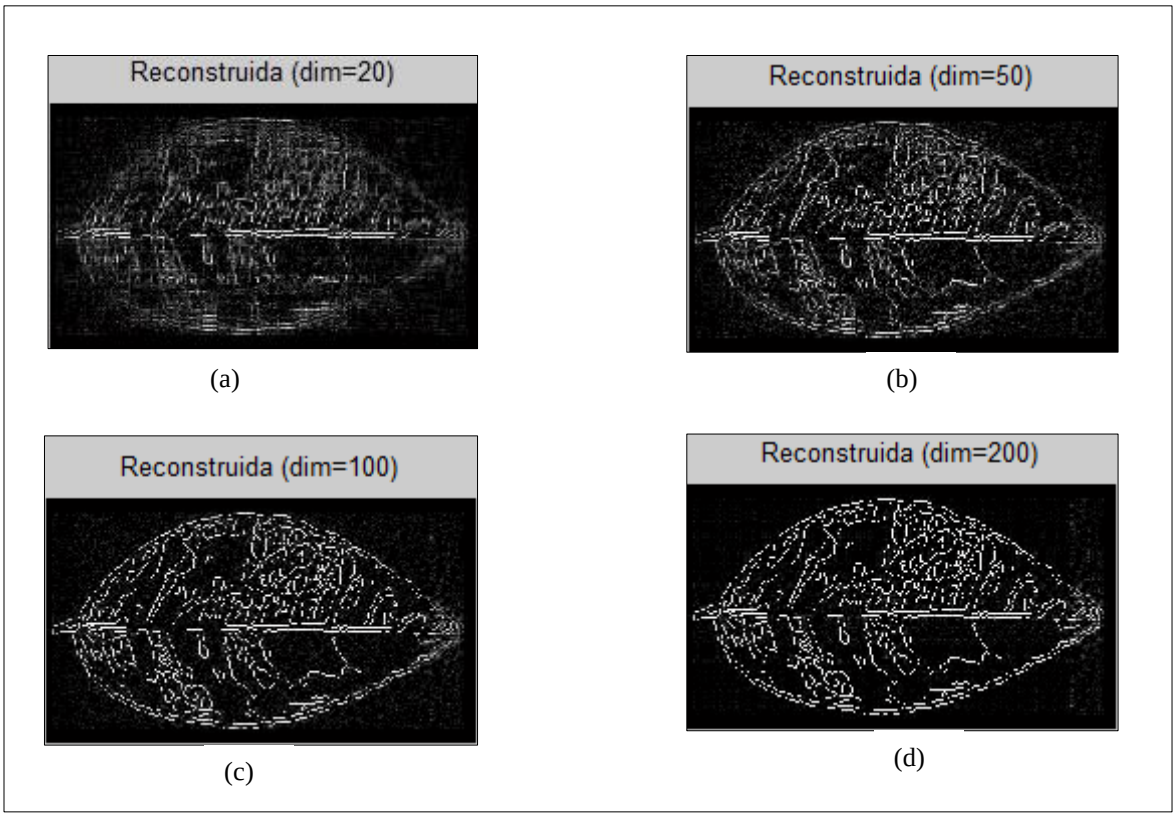

Figura 14. Hoja de limón persa con HLB reconstruida a partir de diferentes cantidades de PCA's. (a) Hoja con HLB reconstruida con 20 PCA´s. (b) Hoja con HLB reconstruida con 50 PCA´s.. (c) Hoja con HLB reconstruida con 100 PCA´s.. (d) Hoja con HLB reconstruida con 200 PCA´s.

La imagen de la *Figura 13 (b)* reconstruida a partir de veinte, cincuenta, cien y doscientos componentes principales se muestra gráficamente en la *Figura 15 (a) (b) (c) y (d)* correspondientemente.

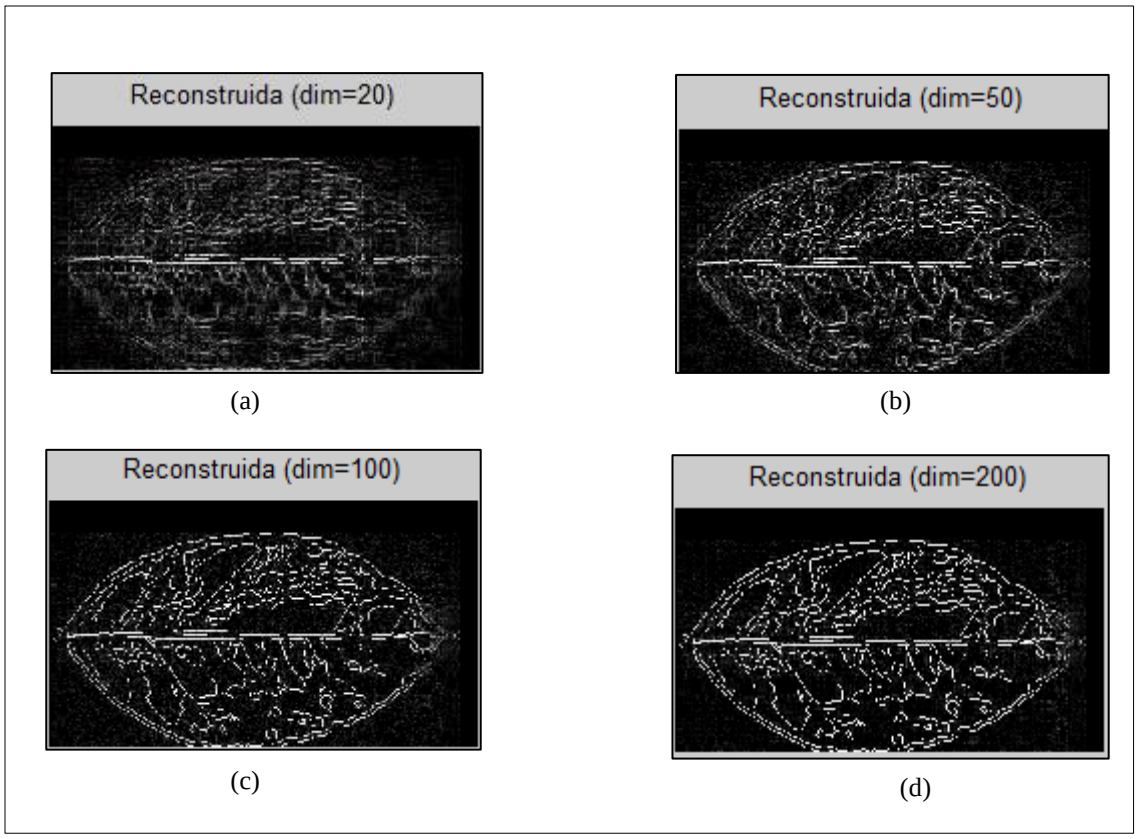

Figura 15. Hoja de limón persa con HLB reconstruida a partir de diferentes cantidades de PCA's. (a) Hoja con HLB reconstruida con 20 PCA´s. (b) Hoja con HLB reconstruida con 50 PCA´s.. (c) Hoja con HLB reconstruida con 100 PCA´s.. (d) Hoja con HLB reconstruida con 200 PCA´s.

La imagen reconstruida con los primeros veinte CP mostrada a través de la *Figura 15 (a)*, hace notar la pérdida del sesenta por ciento de los datos, debido a que los primeros veinte CP representan sólo el cuarenta por ciento de la varianza de los datos en promedio del total de componentes. Es importante notar también, que entre las imágenes reconstruidas con cien y doscientos CP concentran más o menos la misma información, representando del 90 al 100 de la información contenida en la imagen, ver la *Figura 15 (c) y (d).*

El comportamiento en las imágenes de hojas de limón persa sin HLB, tanto de la *Figura 13 (c)* como la *Figura 13 (d)*, con respecto al número de CP para su reconstrucción, es muy similar al comportamiento de la *Figura 14* y *Figura 15*. Podemos observar los resultados en la *Figura 16 (a) (b) (c) (d)* y *Figura 17 (a) (b) (c) (d).*

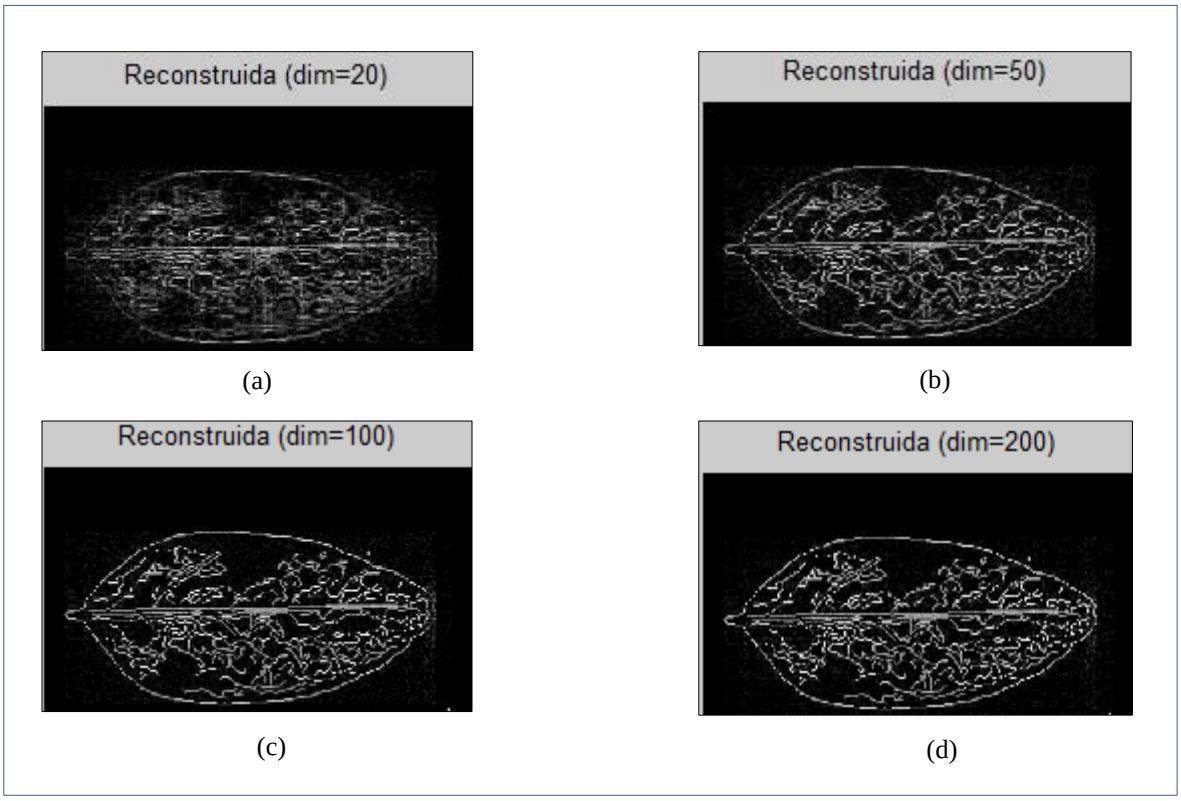

Figura 16. Hoja de limón persa sin HLB reconstruida a partir de diferentes cantidades de PCA's. (a) Hoja sin HLB reconstruida con 20 PCA´s. (b) Hoja sin HLB reconstruida con 50 PCA´s.. (c) Hoja sin HLB reconstruida con 100 PCA´s.. (d) Hoja sin HLB reconstruida con 200 PCA´s.

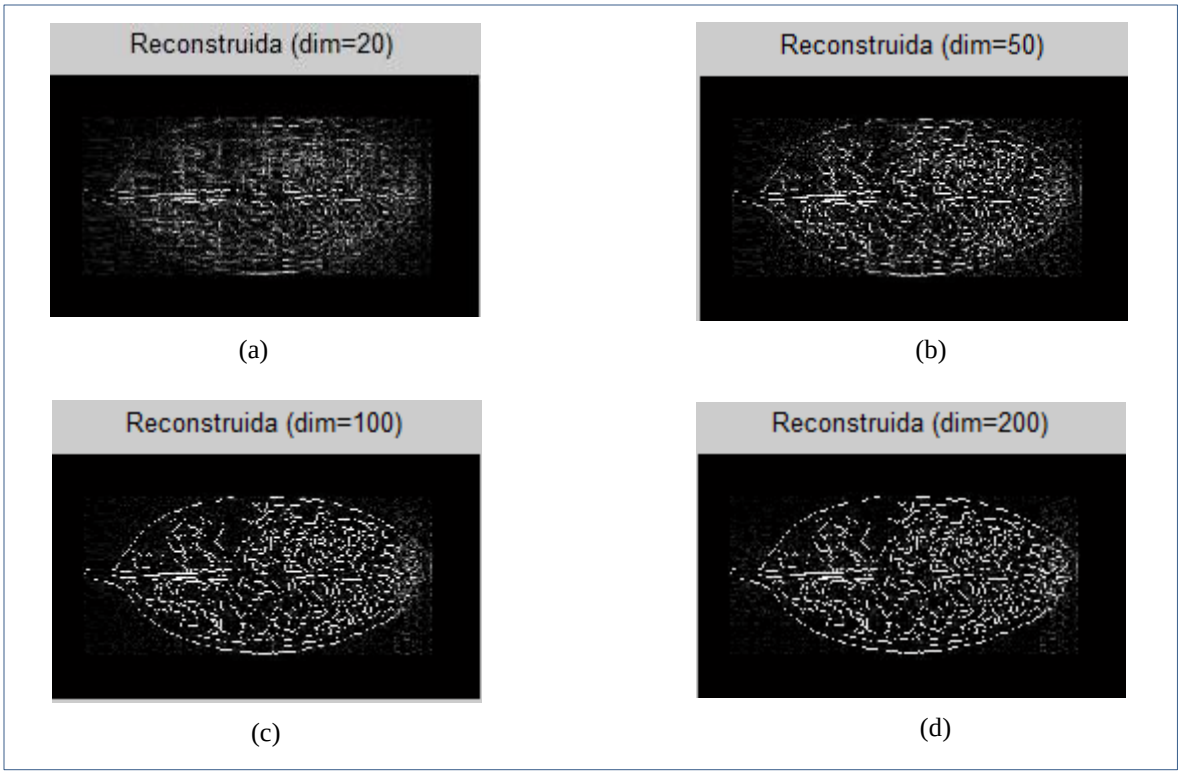

Figura 17. Hoja de limón persa sin HLB reconstruida a partir de diferentes cantidades de PCA's. (a) Hoja sin HLB reconstruida con 20 PCA´s. (b) Hoja sin HLB reconstruida con 50 PCA´s.. (c) Hoja sin HLB reconstruida con 100 PCA´s.. (d) Hoja sin HLB reconstruida con 200 PCA´s.

Al analizar los resultados para tomar una descisión de cuántos CP se deben seleccionar para reducir la dimención de los datos, se debe considerar el porcentaje de varianza que representan esos componentes principales. Mientras más componentes principales se seleccionen la pérdida de datos disminuye.

En esta investigación, se deben utilizar los componentes principales suficientes para no perder detalles de las imágenes, mismas, que podrían ser determinantes para su correcta clasificación como enferma o no enferma de HLB.

Al aplicar el algoritmo de análisis de componentes principales a una imagen de una hoja de limón persa, con tamaño de 200 X 350 pixeles, se generan 350 componentes principales. Ahora, la tarea más importante es, tomar la decisión de cuántos CP se deben seleccionar para reducir la dimensionalidad de las variables.

Para calcular o conocer, la cantidad de información incorporada a un componente se utiliza la varianza. Mientras mayor sea la varianza de un CP, significa que es más importante. Debido a esto, se debe calcular la varianza de cada componente y ordenarlos de mayor a menor para seleccionar los primeros. Pero, la pregunta es ¿Cuántos CP seleccionar de esos primeros? La respuesta podría ser diez, veinte, cien, ciento cincuenta o más.

Para determinar el número de componentes a utilizar en este modelo, se debe calcular que porcentaje de la varianza representan los primeros ***n*** CP. La *Figura 18* muestra la varianza de los primero veinte CP, se aprecia que representan en promedio el cuarenta y cuatro por ciento de la varianza total. Al considerar solo los primeros cincuenta CP se consigue representar en promedio el setenta y dos por ciento de la varianza total, tal como se observa en la *Figura 19.*

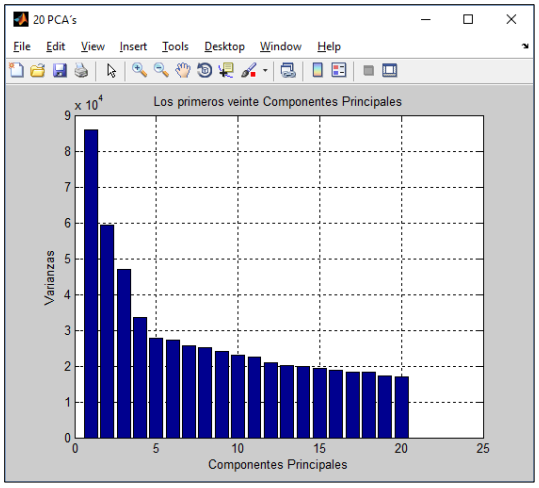

Figura 18. Varianza de los primeros 20 componentes principales. Representan el 44% de la información importante.

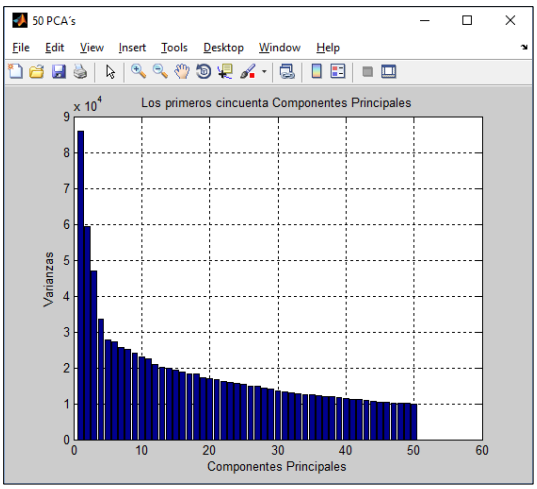

Figura 19. Varianza de los primeros 50 componentes principales. Representan el 72% de la información importante.

De igual manera, en la *Figura 20* se observa que se alcanzó en promedio el noventa y tres por ciento de la varianza total tomando en consideración cien CP. Con ciento cincuenta CP se alcanza el noventa y nueve por ciento de la varianza total, como se observar en la *Figura 21*.

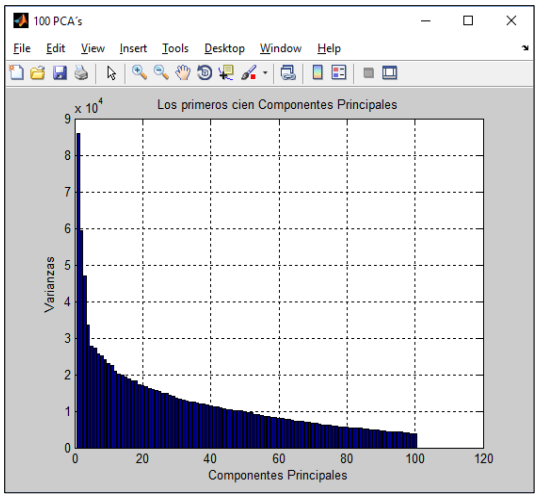

Figura 20. Varianza de los primeros 100 componentes principales. Representan el 93% de la información importante.

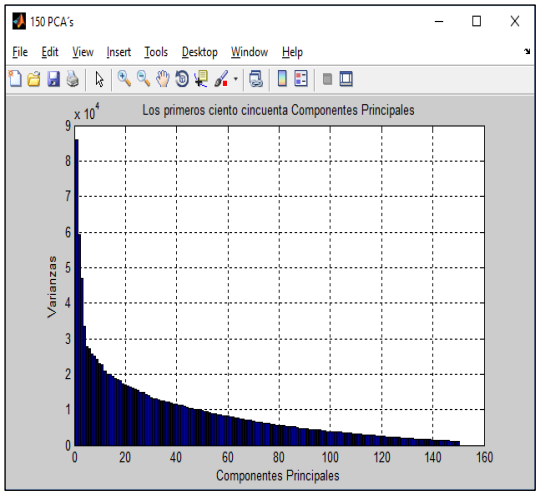

Figura 21. Varianza de los primeros 150 componentes principales. Representan el 99% de la información importante.

Se observa, que con los primeros veinte y cincuenta CP se pierde información importante, principalmente los detalles de la imagen con bordes. Los cien primeros componentes si cumplen con un alto porcentaje de variabilidad de los datos, la pérdida de información realmente es muy poca, por lo que en esta investigación se toma la decisión de elegir los primeros cien componentes principales, esto debido a representan noventa y tres por ciento de la variabilidad de los datos, tal como se manifiesta en la *Figura 20.*

Se puede observar también en la *Figura 22*, que los primeros doscientos CP contabilizan el cien por ciento de la varianza total, incluso alcanza ese porcentaje antes de los primeros doscientos.

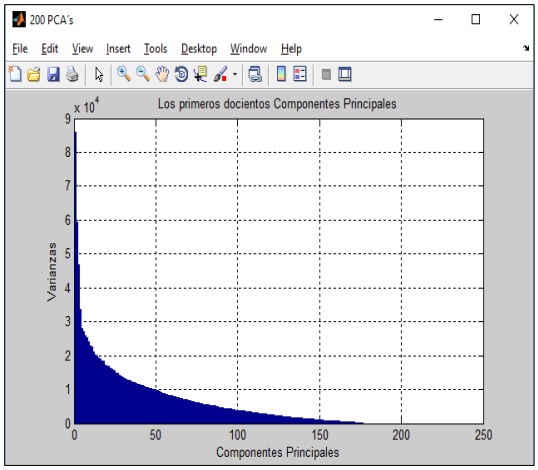

Figura 22. Varianza de los primeros 200 componentes principales. Representan el 100% de la información importante.

4.5 Clasificación.

En esta etapa, se diseñó una red neuronal multicapa backpropagation. La arquitectura de la red neuronal que mejores resultados proporcionó durante la fase pruebas es, una red con siete capas como se muestra en la *Figura 23*.

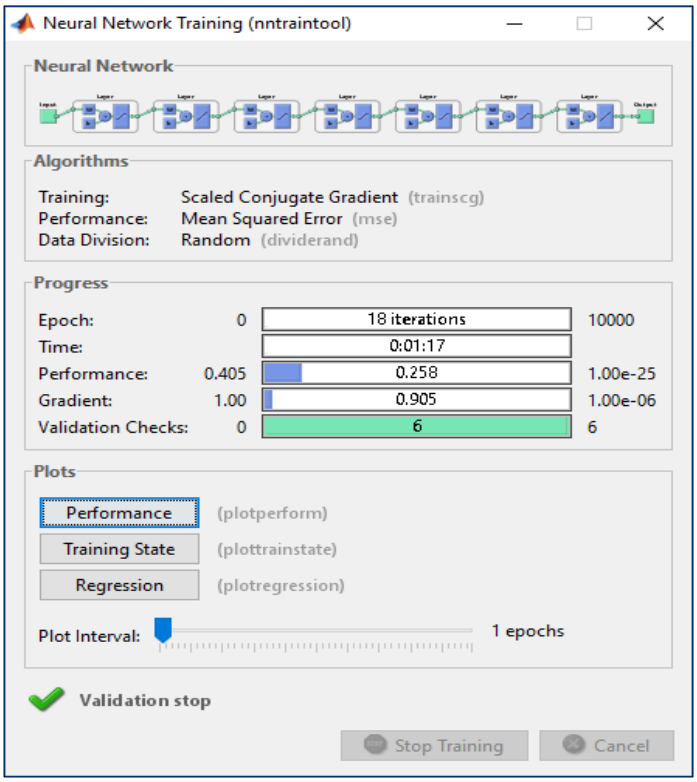

Figura 23. Interface de la red neuronal multicapa.

La primea capa o capa de entrada está constituida por 35000 neuronas, que corresponden a los valores correspondientes a los componentes principales de cada hoja; la capa dos está formada 80 neuronas, la capa tres de 40, la capa cuatro de 20, la capa cinco de 10, la capa seis de 5 y la capa siete de una neurona correspondiente a la capa de salida, donde genera dos valores posibles, estos son cercanos a $\begin{bmatrix}1\\0\end{bmatrix}$ ó $\begin{bmatrix}0\\1\end{bmatrix}$; si el clasificador lo relaciona con la primera clase

(hojas enfermas con HLB) genera la primera salida, en caso contrario, la salida se acerca a la segunda mostrada anteriormente, que corresponde a la clase de hojas sin la enfermedad HLB.

Finalmente, los pesos correspondientes a las neuronas de entrada se asignan de manera aleatoria, y la función de transferencia utilizada es la sigmoidea ("logsig"). Es importante mencionar que el aprendizaje de la red neuronal es supervisado.

4.6 Validación de los resultados.

Se aplicó la validación cruzada k-fold a los veinte datos, diez de ellos pertenecientes a hojas enfermas y con síntomas característicos del HLB, los otros diez datos corresponden a hojas sin la enfermedad HLB. El conjunto de datos se debe dividir en grupos (folds) iguales preferentemente.

Se dividió el conjunto de datos en 5 k-fodls, cada uno de ellos de cuatro datos. El primer fold estuvo formado del dato uno al dato cuatro, el segundo del dato cinco al dato ocho; el tercero del dato nueve al dato doce; el cuarto del trece al dieciséis y el quinto del diecisiete al veinte.

Se realizaron cinco iteraciones debido a que son cinco k-folds. Dicho de otra forma, se entrena cinco veces con cuatro k-folds diferentes cada vez con su correspondiente prueba aplicado sobre el k-fold restante.

En la implementación de la validación cruzada dejando uno fuera, se utilizaron los mismos veinte datos. Se realizaron veinte iteraciones de entrenamiento y prueba. En la primera iteración, el dato de prueba fue el primer dato y el resto de entrenamiento, en la segunda el segundo dato y

así sucesivamente hasta la iteración veinte, donde los primeros diecinueve datos se utilizaron de entrenamiento y el dato veinte de prueba.

Finalmente, la aplicación de la t-student, se realizó con datos generados por ejecución de la validación cruzada dejando uno fuera durante veinte ocasiones. Y su fórmula es la siguiente:

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n-1}}}$$

V ANÁLISIS DE RESULTADOS.

5.1 Validación cruzada k-fold.

Los resultados obtenidos con k-folds se muestran en la *Tabla 3*. La base de datos está constituida por veinte objetos. Los objetos del 1 al 10 corresponden a hojas de limón persa con síntomas y enfermas de HLB; los objetos del 11 al 20 corresponden a hojas de limón persa sanas.

En la *Tabla 3* se muestran los resultados al hacerles la prueba a los 5 k-folds. En el primer fold los cuatro resultados corresponden a las hojas 1,2,3 y 4 identificadas como enfermas. La primera hoja la evalúa como enferma, porque la salida es $\begin{bmatrix} 0.4223 \\ 0.0636 \end{bmatrix}$, como el dato superior es mayor que el, se interpreta como $\begin{matrix} 1 \\ 0 \end{matrix}$, indicando que la clasifica como enferma. La hoja 2, 3 y 4 también las identifica como enfermas. De tal manera, que tiene 4 aciertos de 4 en ese fold, teniendo una precisión de 100%.

En el fold 2 son hojas enfermas y tiene una precisión del 85%. En el fold 3, las dos primeras hojas son enfermas y las otras dos son hojas sanas; su precisión es del 100%. El fold 4 está formado por 4 hojas sanas y su precisión es del 25%. El fold 5 está constituido por 4 hojas sanas y su precisión fue de 100%.

Una hoja es identificada como sana, cuando el valor inferior es mayor que el valor superior. Por ejemplo, la cuarta hoja del fold 5. Donde la salida es $\begin{bmatrix} 0.3092 \\ 0.5496 \end{bmatrix}$ se interpreta como $\begin{matrix} 0 \\ 1 \end{matrix}$ que corresponde a la clase de hojas sanas.

Tabla 3
Resultados de validación cruzada K-fold

	RESULTADOS					
K-FOLD	**HLB**	**HLB**	**HLB**	**HLB**	**ACIERTOS**	**PRECISIÓN**
	1	1	1	1	4	1
1	0	0	0	0		
	HLB	**HLB**	**HLB**	**HLB**		
	1	1	0	1	3	0.75
2	0	0	1	0		
	HLB	**HLB**	**SANA**	**SANA**		
	1	1	0	0	4	1
3	0	0	1	1		
	SANA	**SANA**	**SANA**	**SANA**		
	1	1	0	1	1	0.25
4	0	0	1	0		
	SANA	**SANA**	**SANA**	**SANA**		
	0	0	0	0	4	1
5	1	1	1	1		
					PRECISIÓN GLOBAL	0.8

De acuerdo a los resultados anteriores, la matriz de confusión resultante se muestra en la *Tabla 4*. De acuerdo a los resultados de esta red neuronal, la probabilidad de clasificar correctamente a las hojas enfermas es del 90% (sensibilidad), la probabilidad de clasificar correctamente a las hojas sanas es del 70% (especificidad), obteniendo una precisión del 80%.

Tabla 4
Matriz de confusión de validación cruzada K-fold.

	DATOS REALES ENFERMEDAD		
RESULTADOS PRUEBA	**HLB**	**SANOS**	**TOTALES**
POSITIVOS	9	3	12
NEGATIVOS	1	7	8
TOTAL	10	10	20

Nueve hojas enfermas son correctamente clasificadas como positivas (verdaderos positivos). Una hoja enferma es incorrectamente clasificada como sana o negativa (falso negativo). Siete hojas sanas son clasificadas correctamente como negativas (verdaderos

negativos) y tres hojas sanas son incorrectamente clasificadas como enfermas o positivas (falsos positivos).

5.2 Validación cruzada dejando uno fuera.

En la *Tabla 5*, se muestran los resultados de la red neuronal utilizando validación cruzada dejando uno fuera. En el proceso de entrenamiento y prueba, se utilizó la misma base de datos usada en la validación cruzada k-folds. Son veinte iteraciones, por lo tanto son veinte resultados. Los diez resultados de la izquierda, corresponden a la prueba de las diez hojas con síntomas y que pertenecen a árboles con la enfermedad HLB. Los diez resultados de la derecha, pertenecen a hojas de árboles sin la enfermedad.

En la iteración 1, el resultado de la red neuronal al probar la hoja enferma número 1 es $\begin{bmatrix} 0.5606 \\ 0.4952 \end{bmatrix}$, como el valor superior es mayor, se interpreta como salida $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, esto significa que la clasifica correctamente en la clase de hojas con la enfermedad HLB. También clasifica correctamente las hojas 2, 3, 5, 6,7 y 9 en las iteraciones correspondientes. Las hojas 4,8 y 10 enfermas de HLB son incorrectamente clasificadas como sanas.

Tabla 5
Resultados de validación cruzada leave-one-out.

ITERACIÓN	HLB	RESULTADO	ITERACIÓN	SANA	RESULTADO
	0.5606	1		0.6171	0
1	0.4952	0	**11**	0.6876	1
	0.7489	1		0.4374	0
2	0.4882	0	**12**	0.6304	1
	0.5311	1		0.8086	1
3	0.2364	0	**13**	0.3611	0
	0.2853	0		0.4546	0
4	0.3925	1	**14**	0.5354	1
	0.8429	1		0.2641	0
5	0.7868	0	**15**	0.774	1
	0.7983	1		0.3495	0
6	-0.0134	0	**16**	0.4135	1
	0.686	1		0.5857	1
7	0.4326	0	**17**	0.4325	0
	0.8546	0		0.3892	0
8	0.8867	1	**18**	0.676	1
	0.5523	1		0.5128	0
9	0.3562	0	**19**	1.039	1
	0.475	0		0.7618	1
10	0.5644	1	**20**	0.404	0

Las hojas sanas, ubicadas desde la hoja 11 hasta la hoja 20, son clasificadas correctamente como negativas o sanas sólo las hojas 11, 12, 14, 15, 16,18 y 19 en sus correspondientes iteraciones. Pero las hojas 13, 17 y 20 son clasificadas incorrectamente como enfermas, cuando realmente son sanas.

Derivado de lo anterior, se calcula la matriz de confusión de los resultados anteriores de la red neuronal, misma que se muestra en la *Tabla 6.* Se concluye que la sensibilidad de la prueba es del 70%, la especificidad también es del 70% y de igual forma la precisión es del 70%.

Tabla 6
Matriz de confusión de validación cruzada leave-one-out.

	DATOS REALES ENFERMEDAD		
RESULTADOS PRUEBA	**HLB**	**SANOS**	**TOTALES**
POSITIVOS	7	3	10
NEGATIVOS	3	7	10
TOTAL	10	10	20

Siete hojas enfermas son clasificadas como positivas o con la enfermedad (verdaderos positivos). Tres hojas enferma son clasificadas como sanas o negativas (falsos negativos). Siete hojas sanas son clasificadas correctamente como negativas (verdaderos negativos) y tres hojas sanas son incorrectamente clasificadas como enfermas o positivas (falsos positivos).

5.3 Validación mediante t-student.

La t-student, se utilizó para comprobar la efectividad del modelo, especialmente porque el tamaño de la muestra es pequeño. En *la Tabla 7*, se muestran los datos utilizados para la t-student.

Tabla 7
Resultados de la validación cruzada leave-one-out en 20 corridas.

Corrida	Porcentaje eficiencia	$(x-\overline{X})$	$(x-\overline{X})^2$
1	60	1.25	1.5625
2	50	-8.75	76.5625
3	70	11.25	126.5625
4	60	1.25	1.5625
5	55	-3.75	14.0625
6	50	-8.75	76.5625
7	70	11.25	126.5625
8	55	-3.75	14.0625
9	60	1.25	1.5625
10	65	6.25	39.0625
11	55	-3.75	14.0625
12	55	-3.75	14.0625
13	55	-3.75	14.0625
14	60	1.25	1.5625
15	55	-3.75	14.0625
16	50	-8.75	76.5625
17	65	6.25	39.0625
18	55	-3.75	14.0625
19	65	6.25	39.0625
20	65	6.25	39.0625

Hipótesis de trabajo: La hipótesis nula afirma que la media es igual a 60%, mientras que la hipótesis alternativa afirma que la media es diferente a 60%. Representándose de la siguiente forma:

H0: $\mu=0.6$

H1: $\mu\neq0.6$

Con una significancia de $\alpha=0.05$, por lo tanto, la confianza es del 95%. En la *Figura 24*, se muestran los valores críticos -2.093 y 2.093 para la cola izquierda y derecha correspondientemente en la distribución t de student.

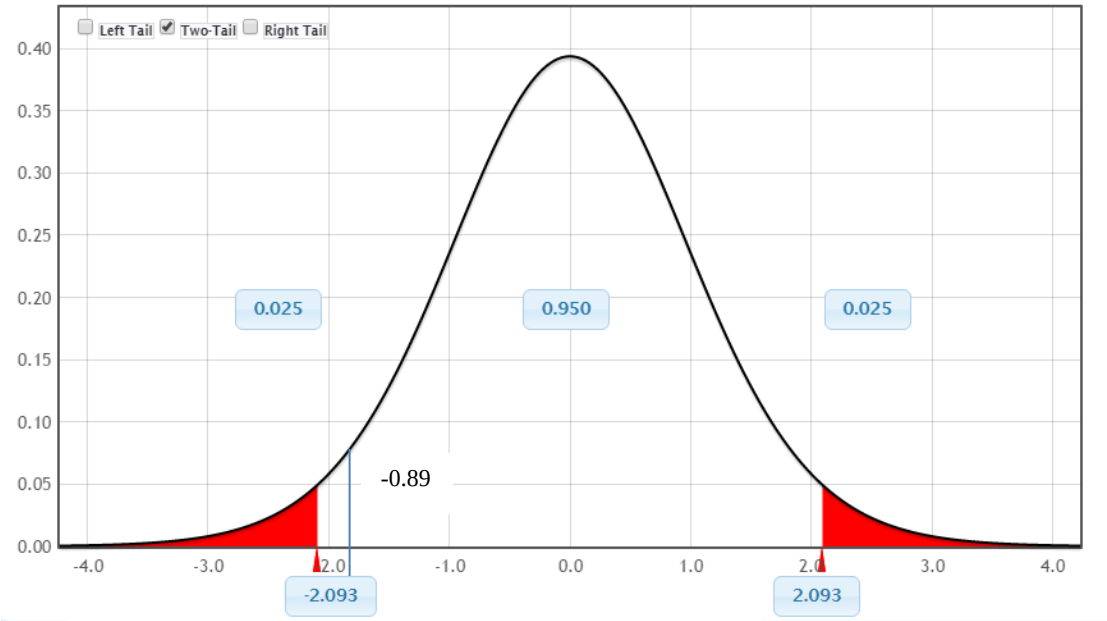

Figura 24. Distribución t-student.

t-student: $$t=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}=\frac{58.75-60}{\frac{6.25657549}{\sqrt{20}}}=\frac{-1.25}{1.39901281}=-0.8934872$$

Tamaño de la muestra:	20
Grados de libertad:	19
Media:	58.75
Desviación Estándar :	6.25657549

Se acepta la hipótesis nula Ho: $\mu = 0.6$
El intervalo de confianza de 95 % está entre 0.5582183 y 0.6167817

Se acepta la hipótesis nula y se concluye que la media de los aciertos obtenidos mediante el modelo de identificación de HLB en hojas de limón persa es de $\mu=0.6$ con un 95% de confianza.

CONCLUSIONES

Debido a los avances tecnológicos y desarrollo de herramientas informáticas, actividades que realizaban sólo los humanos ahora también las pueden realizar sistemas que integran software, hardware y técnicas de inteligencia artificial. La identificación de enfermedades, deficiencias nutrimentales u otros elementos en plantas está tomando gran importancia. En este sentido, se puede mencionar el reconocimiento de patrones, procesado y clasificación de imágenes, como herramientas de gran importancia que se han utilizado en este tipo de investigaciones.

En este trabajo de investigación se logró identificar la enfermedad del HLB en imágenes de hojas sintomáticas de árboles de limón persa, mismas que se adquirieron usando una cámara digital convencional. Se utilizó clasificación con aprendizaje supervisado, basado en una red neuronal tipo backpropagation.

Los resultados de las pruebas que proporciona la red neuronal, se probaron con el algoritmo de validación cruzada k-folds (K-fold cross-validation) y validación cruzada dejando uno fuera (Leave-one-out cross-validation). Con la primera se obtiene una precisión máxima del ochenta por ciento y con la segunda se tiene una precisión máxima del setenta por ciento. La red se diseñó con siete capas. Una capa e entrada, cinco capas ocultas y una capa de salida. Observándose que sus resultados eran aceptables.

Se validó también con la t de student debido a que la muestra de esta investigación es menor a treinta elementos. Se utilizó una muestra de veinte imágenes de hojas de limón persa. Diez eran de imágenes de hojas de limón persa con la enfermedad y con síntomas característicos

de ésta. Y diez hojas sanas. Esta prueba afirma que la media de la precisión de la red neuronal es del sesenta por ciento.

Como parte de esta investigación, se obtienen varias imágenes de hojas de limón persa enfermas de HLB, de las cuales, sólo diez son aptas para esta investigación. Las imágenes fueron proporcionadas por el comité estatal de sanidad vegetal del estado de colima. Sin embargo, queda pendiente trabajar para tener una base de datos de imágenes de hojas de limón con HLB para seguir desarrollando soluciones sobre esta línea de investigación.

No se seleccionaron variables directamente en el proceso de clasificación, realmente, la red neuronal se encargó de reconocer patrones a partir de los componentes principales que toma como entrada.

En el proceso de segmentación, se aplicó el algoritmo de Canny para la obtención de bordes de las imágenes de hojas de limón persa, tanto de las hojas sanas como enfermas. Y se obtuvieron resultados excelentes.

Se logró reducir la dimensionalidad de los datos, aplicando la técnica de análisis de componentes principales. Se generan 350 componentes al aplicar PCA, sin embargo, se determina utilizar solamente los primeros 100 componentes principales. Debido a que representan el noventa y tres por ciento de la variabilidad de los datos.

La aportación de esta tesis, es el modelo propuesto para la identificación de la enfermedad huanglongbing de los cítricos a través del tratamiento y segmentación de imágenes digitales, aplicando el análisis de componentes principales para reducir la dimensión de las variables utilizadas para el reconocimiento de patrones, basado en una red neuronal del tipo backpropagatión con aprendizaje supervisado.

Este trabajo, asienta las bases para desarrollar más investigaciones que resuelvan problemáticas del campo, especialmente en área citrícola, debido a que es la actividad económica más importante de la región de Martínez de la Torre, Veracruz.

TRABAJO FUTURO

La línea de continuación de este trabajo de investigación, es la aplicación de otras técnicas de tratamiento de imágenes digitales; así como, la aplicación de otras técnicas de inteligencia artificial o una combinación de ellas para mejorar la precisión del clasificador y principalmente la sensibilidad para identificar a la enfermedad huanglongbing en las hojas de limón persa conocidos como verdaderos positivos.

Otra línea de investigación es el desarrollo herramientas con interfaces amigables para la presentación de resultados e información en cualquier dispositivo electrónico, principalmente para dispositivos móviles.

También es prioridad, la generación de una base de datos de imágenes de hojas de limón persa con síntomas de la enfermedad. Esta debe ser lo suficientemente grande con la intención de generar mejores resultados, en especial, cuando se utilizan técnicas de minería de datos.

LISTA DE REFERENCIAS

Aksenov, A. A., Pasamontes, A., Peirano, D. J., Zhao W., Dandekar, A. M., Fiehn, O., Ehsani, R., y Davis C. E. (2014). Detection of Huanglongbing Disease Using Differential Mobility Spectrometry. *Analytical Chemistry*, 2014, 86, 2481−2488. dx.doi.org/10.1021/ac403469y.

Batista, L., Peña, I., López, D., Pérez, J. P. y Llauger, R. (2008). Técnicas de diagnóstico de enfermedades que afectan a los cítricos. *Manual de saneamiento y diagnóstico para la producción de material de propagación certificado de cítricos*. Recuperado el 14 de octubre de 2015, de http://www.concitver.com/publicaciones.html

Canny, J. (1986). A Computational Approach to Edge Detection. *Transactions on pattern analysis and machine intelligence,* Vol. PAMI-8, No. 6,679-698.

Collazo, C., Luis, M. y Llauger, R. (2009). Técnicas empleadas para el diagnóstico del Huanglonging de los cítricos. *Citrifrut,* 25 (2): 24-31.

Collazo, C., Núñez, J., Luis, M. y Llauger, R. (2011). Optimización de una reacción en cadenade la polimerasa anidada para el diagnóstico de la enfermedad "huanglongbing" de loscítricos. *CitriFrut*, 28(2):19-30. ISSN 1607-5072.

Deng, X. L., Li, Z. y Hong, T. S. (2014). Citrus disease recognition based on weighted scalable vocabulary tree. *Precision Agriculture*, *15*, 321–330. doi: 10.1007/s11119-013-9329-2.

Fui y Mui (19981). A survey on image segmentation. Pattern Recognition, 13:3-16.

Gil, X. (2015, 21 de mayo). Limón persa mexicano en el mundo (II). *El economista.* Recuperado el 16 de noviembre de 2015 de http://eleconomista.com.mx/columnas/agro-negocios/2015/05/21/limon-persa-mexicano-mundo-ii.

Gonzalez RC y Woods RE (1996). *Tratamiento digital de imágenes*, Addison-Wesley Publishing Co, Reading, Washington.

Gonzalez, R. y Woods, R. (2002). *Digital Image Processing*. 2nd Edition. Prentice Hall.

Gottwald, T. R., da Graça, J. V., y Bassanezi, R. B. (2007). Citrus Huanglongbing: The pathogen and its impact. Online. *Plant Health Progress* doi:10.1094/PHP-2007-0906-01-RV.

Hocquellet, A., Toorawa, P., Bové, J.M. y Garnier, M. (1999). Detection and identification of the two ´Candidatus Liberibacter´ species associated with the citrus Huanglongbing by PCR amplification of ribosomal protein genes of the β operon. Mol. Cell. *Probes* 13: 373-379.

Jagoueix, S., Bové, J. y Garnier, M. (1997). Comparison of the ribosomal intergenic regions of "Candidatus Liberobacter asiaticum" and "Candidatus Liberobacter africanum", the two species associated with the citrus Huanglongbing (Greening) disease. Int. J. Syst. Bacteriol. 47(1):224-227.

James, G., Witten, D., Hastie, T. y Tibshirani, R. (2013). An introduction to statistical learning. 1st ed. New York: *Springer*.

Kawano, S., Tetsuya, T., Atsushi, O., Numazawa, M., y Yasuda, K. (2006). The simple and rapid diagnosis of citrus huanglongbing (Citrus Huanglongbing) by Scratch method. Okinawa Prefectural Agricultural Research Center, *Makabe* 820, Itoman, Okinawa, 901-0336.

Kumar, A., Lee, W. S., Ehsani, R. J., Albrigo, L. G., Yang, C. y Mangan, R. L. (2012). Citrus greening disease detection using aerial hyperspectral and multispectral imaging techniques. *Society of Photo-Optical Instrumentation Engineers (SPIE), 6*(2012), 063542-1-063542-22.

Li, W., Hartung, J. S. y Levy, L. (2005). Quantitative real-time PCR for detection and identification of Candidatus Liberibacter species associated with citrus huanglongbing. *Journal of Microbiological Methods, 66* (2006), 104–115.

López, M.M., Bertolini, E., Olmos, A., Caruso, P., Gorris, M.T., Llop, P., Penyalver, R. y Cambra, M. (2003). Innovative tools for detection of plant pathogenic viruses and bacteria. *International Microbiology* 6, 233–243.

Luis, P. M., Hernández, R. L., Collazo, C. C., Peña, B. I., Zamora, R. V., López, H. D., Llauger, R. R. y Batista, L. R. L. (2014). Diagnóstico y caracterización de la enfermedad huanglongbing de los cítricos para el establecimiento de su manejo en Cuba. *Propuesta a Premio de la Academia de Ciencias de Cuba, 2014.* Instituto de Investigaciones en Fruticultura Tropical. La Habana, Cuba.

Mishra, A. R., Karimi, D., Ehsani, R. y Lee, W. S. (2012). Identification of citrus greening (hlb) using a VIS-NIR spectroscopy technique. *American Society of Agricultural and Biological Engineers, 55*(2), 711-720.

Mishra, A., Karimi, D., Ehsani, R. y Albrigo, L. G. (2011). Evaluation of an active optical sensor for detection of Huanglongbing (HLB) disease. *Biosystems Engineering*, *110*(2011), 302-309.

Mora, G. (2011). Ficha técnica HLB HuangLongBing. Servicio Nacional de Sanidad Inocuidad y Calidad AgroAlimentaria (SENASICA). Recuperado el 13 de octubre de 2015 de http://langif.uaslp.mx/plagasdevastadoras/documentos/fichas/Huanglongbing.pdf.

Mota, A. D., Rossi, G., De Castro, G. C., Ortega T. A. y De Castro, J. C. (2014). Portable fluorescence spectroscopy platform for Huanglongbing(HLB) citrus disease in situ detection. Light-Emitting

Diodes: Materials, Devices, and Applications for Solid State Lighting XVIII. *Society of Photo-Optical Instrumentation Engineers (SPIE), 9003* (90031U), 1–9.

Nister, D. y Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In Computer Vision and Pattern Recognition, *IEEE*, 2161–2168.

Pérez-Planells, Ll., Delegido, J., Rivera-Caicedo, J.P. y Verrelst, J. (2015) Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista de teledetección*, 44, 55-65. Recuperado el 25 de junio de 2016 de http://polipapers.upv.es/index.php/raet/article/view/4153/4616.

Pourreza, A., Lee, W. S. y Ehsani, R. (2014). A Vision Based Sensor for Huanglongbing Disease Detection under a Simulated Field Condition. *ASABE and CSBE/SCGAB Annual International Meeting*. Paper Number :141900251.

Pourreza, A., Lee, W. S., Raveh, E., Hong, Y. y Kim, H. J. (2013). Identification of citrus greening disease using a visible band image analysis. . *ASABE Annual International Meeting*. Paper Number: 131591910.

Ramos-González, P. L., Hernández-Rodríguez, L. y Banguela-Castillo, A. (2011). Plataformas genéricas para el diagnóstico a gran escala en fitopatología. *CitriFrut*, 28(1): 25-37.

Reyes, J. I. y Cevallos, J. M. (Agosto 2009) Perfilamiento Metabólico Para La Detección De Hlb En Cítricos. *1er. encuentro Internacional de Investigación en Cítricos***.** Encuentro llevado a cabo en CONCITVER, Martínez de la Torre, Veracruz, México.

Rodríguez, M. (2005), Optimización de la genotipificación de ADN como un problema de Selección de Características. *Tesis de Maestría en Ciencias con Especialidad en Ingeniería en Sistemas Computacionales*. Escuela de Ingeniería Departamento de Ingeniería en Sistemas Computacionales, Universidad de las Américas Puebla.

Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533--536.

Sánchez, A. (2012). Análisis de componentes principales: versiones dispersas y extensiones con diferentes costes. *Universidad Carlos II de Madrid*. *Departamento de Teoría de la*

Señaal y Comunicaciones. . Recuperado el 09 de febrero de 2016, de http://e-archivo.uc3m.es/bitstream/handle/10016/15618/PFC_Andres%20Sanchez%20Mangas.pdf?sequence=1

Sánchez, f. J. (2010). Medición y análisis de las variaciones en el nivel de un modelo físico empleando imágenes. *Maestría en ciencias de la computación*, universidad autónoma metropolitana , unidad azcapotzalco, división de ciencias básicas.

Sankaran, S., Mishra, A., Ehsani, R. y Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(2010), 1–13.

Sivanandam, S. N., Sumathi, S., y Deepa, S. N. (2006). *Introduction to Neural Networks Using Matlab 6.0*. McGraw-Hill.

Taba, S., Nasu, K., Takaesu, K., Ooshiro, A. y Moromizato, Z. (2006). Detection of citrus Huanglongbing using an iodo-starch reaction. *Science bulletin* - Faculty of Agriculture University of the Ryukyus (53):19-24.

Teixeira, D. C., Danet, J. L., Eveillard, S., Martins, Cintra de Jesús, V. J., Yamamoto, P., Lopes, S. A., Bassanezi, A. B., Ayres, A. J. , Saillard, C. y Bové, J. M. (2005). Citrus Huanglongbing in Sao Paulo State, Brazil: PCR detection of the Candidatus Liberibacter species associated with the disease. Mol. Cell. *Probes* 19: 173-179.

Valencia, M. A., Yánez, C. y Sánchez, L.P. (2006). Algoritmo Backpropagation para Redes Neuronales: conceptos y aplicaciones. *Instituto Politécnico Nacional Centro de Investigación En Computación*. No. 125, Serie Verde. Recuperado el 14 de agosto de 2015, de http://www.repositoriodigital.ipn.mx/handle/123456789/8628.

Valverde, R. J. (2007). Detección de bordes mediante el algoritmo de Canny. *Escuela Académico Profesional de Informática*, Universidad Nacional de Trujillo. Recuperado el 10 de noviembre de 2016 de la base de datos ResearchGate.

Yang, Y. y Huang, S. (2014). Suitability of five cross validation methods for performance evaluation of nonlinear mixed-effecs forest models – a case study. *Forestry*, 87, 654-662. http://dx.doi.org/10.1093/ forestry/cpu025.

Yap, K. H., Chen, T., Li, Z. y Wu, K. (2010). A comparative study of mobile-based landmark recognition techniques. Intelligent Systems, *IEEE*, 25(1), 48–57.