INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE CÓMPUTO

Trabajo TerminaI

“Prototipo de aplicación para la detección de deficiencia de nutrientes en cultivos de hidroponía”

*2017-A054*

Presenta:

Edgar Rodrigo Arredondo Basurto

Directores:

Ing. Eduardo Gutiérrez Aldana Dr. José Félix Serrano Talamantes



*Mayo de 2018*

INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE CÓMPUTO

SUBDIRECCIÓN ACADÉMICA

No de TT:2017-A054 de mayo de 2017

Documento técnico

“Prototipo de aplicación para la detección de deficiencia de nutrientes en cultivos de hidroponía”

*2017-A054*

Presenta:

Edgar Rodrigo Arredondo Basurto1

Directores:

Ing. Eduardo Gutiérrez Aldana Dr. José Félix Serrano Talamantes

RESUMEN

El objetivo de este proyecto es desarrollar un prototipo de aplicación de visión por computadora que tome como entrada imágenes de hojas de cultivos con una enfermedad visible y como resultado se obtiene un probable diagnóstico de la enfermedad de la planta, bajo ciertas restricciones de plantas y enfermedades. Es así que en este documento se describen el análisis, diseño y construcción del sistema que cumple con dicho objetivo.

Palabras clave: Clasificador, descriptor, machine learning, deep learning, red neuronal artificial.

I. Resumen

El objetivo de este proyecto es desarrollar un prototipo de aplicación de visión por computadora que analice imágenes de hojas de cultivos que presenten síntomas visibles de alguna enfermedad. Como resultado se obtiene un probable diagnóstico de cuál es la enfermedad que presenta la planta.

Es así como en este documento se definen y diseñan las diferentes etapas que requiere este prototipo para su implementación, así como detalles y evidencias del desarrollo y puesta a punto del mismo.

En el capítulo 1 se presenta una breve introducción en la que se explica el contexto de este proyecto y la motivación que dio origen a su desarrollo. Además, se definen los objetivos para la etapa de diseño del prototipo.

En el capítulo 2 se exponen algunos fundamentos teóricos útiles para entender de mejor forma el funcionamiento de una de las etapas fundamentales del prototipo, que es la detección de objetos durante la etapa de segmentación. En esta etapa se aíslan las hojas del resto de la imagen, ya que son los objetos que contienen la información útil para esta aplicación.

Inmediatamente después, en el capítulo 3 se abordan las etapas de captura de la imagen y segmentación. Aquí se definen el tipo de objetos a capturar y ciertas restricciones que debe tener la imagen que el usuario proporciona, como la especie de planta, el contenido de la imagen y la posición de las hojas dentro de la imagen. En la segmentación se presenta el método de detección de objetos seleccionado.

En el capítulo 4 se presenta lo relacionado a la clasificación de los objetos, incluyendo la selección del descriptor y el algoritmo de clasificación. En este capítulo se delimita cuáles son los nutrientes por identificar por este prototipo, siendo estos el potasio, nitrógeno, azufre y magnesio.

Finalmente, en el capítulo 5 se aborda el tema de las herramientas y lenguaje de programación para implementar el prototipo. Todas las etapas relacionadas a la visión por computadora se desarrollarán con el apoyo de la biblioteca de OpenCV la cual ofrece soporte para C++, java y python.

II. Abstract

The purpose of this project is the developing of a computer vision application that analyze plant leaves images with a visible disease. As the result, the application gives a probably diagnostic of the plant disease.

In this document are defined and designed the different phases that this prototype needs for its implementation.

In the chapter 1 is presented a brief introduction where is explained the context of this project and the motivation behind the its development. Also, the objectives for the design phase are presented.

In the chapter 2 is found the theoretical framework the object detection topic is discussed. The object detection is required in the segmentation step, where the leaves are separated from the background of the image, because the leaves contains the useful information for the next steps of the application.

In the chapter 3 are discussed the image capture and segmentation steps. In this chapter are defined the object type to capture and some restrictions about the captured image, such as the plant species, the image’s contents and the position of the leaves inside the image. In the segmentation section is presented the object detection method chosen.

In the chapter 4 is presented the feature descriptor and the image classifier. Also, in this chapter is defined the deficiencies to detect in the images, which are the potassium, nitrogen, sulfur and magnesium deficiencies, by reasons that are discussed in that section.

Finally, in the chapter 5 are discussed the developing tools and the programming language for the implementation. All the steps related to computer vision are going to be developed with the support of the OpenCV library, which support the C++, java and python languages.

III. Índice general

[I. Resumen 3](#_Toc497097033)

[II. Abstract 4](#_Toc497097034)

[III. Índice general 5](#_Toc497097035)

[Capítulo 1. Introducción 6](#_Toc497097036)

[1.1 Objetivos. 9](#_Toc497097037)

[1.1.2 Objetivo general. 9](#_Toc497097038)

[1.1.3 Objetivos particulares. 9](#_Toc497097039)

[1.2 Justificación. 9](#_Toc497097040)

[1.3 Trabajo propuesto. 10](#_Toc497097041)

[Capítulo 2. Marco teórico 12](#_Toc497097042)

[2.1 Detección de objetos mediante visión por computadora. 12](#_Toc497097043)

[2.1.1 Detección de objetos Viola-Jones 12](#_Toc497097044)

[2.1.2 Clasificación SVM con descriptores de histogramas de gradientes orientados (HOG) 14](#_Toc497097045)

[Capítulo 3. Captura y segmentación de la imagen. 15](#_Toc497097046)

[3.1 Imagen adquirida 16](#_Toc497097047)

[3.2 Segmentación 18](#_Toc497097048)

[Capítulo 4. Extracción de descriptores y clasificación. 21](#_Toc497097049)

[4.1 Selección de descriptores 22](#_Toc497097050)

[4.2 Clasificador 23](#_Toc497097051)

[Capítulo 5. Implementación. 25](#_Toc497097052)

[5.1 Lenguaje de programación 25](#_Toc497097053)

[5.2 Arquitectura del sistema 26](#_Toc497097054)

[5.3 Diseño de la aplicación. 28](#_Toc497097055)

[Referencias. 30](#_Toc497097056)

Capítulo 1. Introducción

Introducción.

En las siguientes secciones se proporciona información general acerca del contexto de este proyecto; la situación que dio origen a su desarrollo, el alcance y objetivos del mismo, la metodología de desarrollo y el marco teórico relacionado con la propuesta solución planteada para el problema de la clasificación de enfermedades de plantas con base en el análisis de imágenes.

* 1. Planteamiento del problema.

La hidroponía es una rama de la agricultura. Es una técnica de producción intensiva de plantas, que se caracteriza por abastecer el agua y los nutrientes de manera controlada y de proporcionar a las plantas los elementos nutritivos en las concentraciones y proporciones más adecuadas, a través de una solución de nutrientes minerales. Para su aplicación se utilizan sustratos inertes diferentes al suelo agrícola a los que se les adiciona en forma constante una solución nutritiva, preparada a partir de fertilizantes comerciales; con esto se logra un medio que proporciona las condiciones físicas, químicas y sanitarias más adecuadas para el desarrollo de los cultivos.

La hidroponía ha sido utilizada en forma comercial desde hace 50 años y se ha adaptado a diferentes situaciones, tanto con cultivos al aire libre como bajo condiciones de invernadero. Este sistema de producción es importante porque permite cultivar especies para el consumo humano en regiones donde no existen tierras de cultivo o donde el clima no es favorable para el cultivo tradicional de ciertas especies [1].

Las ventajas anteriores han favorecido el mercado global de la hidroponía, el cual fue valuado en 411.88 millones de dólares (USD) en el año 2017 y se espera que crezca a una tasa anual de crecimiento compuesta (CARG) de 12.81% para alcanzar un mercado de 752.57 millones de dólares en el año 2022 [2].

Sin embargo, en México el 60% de los invernaderos de hidroponía que se han instalado han fracasado ante el desconocimiento de productores, la falta de capacitación de técnicos y de mercado [3]. Parte de este problema es el mantener saludables las plantas, ya que la deficiencia de algún nutriente en la solución puede causar que los frutos no sean aptos para consumo humano, o en casos extremos, que la planta muera. Además existen otro tipo de enfermedades derivadas de bacterias o insectos que una planta puede padecer. Identificar la enfermedad tomando como referencia los síntomas visibles que presenta la planta no es una tarea sencilla, ya que requiere de conocimientos y experiencia para realizar un diagnóstico preciso.

El prototipo propuesto tiene como finalidad ayudar al agricultor a diagnosticar posibles enfermedades de plantas con base en los síntomas visibles que se presenten. Como se verá más adelante, la solución propuesta permite clasificar diferentes tipos de enfermedades y plantas, siempre y cuando se tenga una cantidad suficiente de imágenes y una computadora con los recursos suficientes para procesarlas. En dicha solución se considera un conjunto de datos de 14,828 imágenes clasificadas en nueve enfermedades que presenta el tomate.

1.2 Objetivos.

Los objetivos de este proyecto están clasificados en un objetivo general y en diversos objetivos para metas particulares. Dichos objetivos se describen a continuación.

1.2.1 Objetivo general.

Diseñar y desarrollar el prototipo de una aplicación de visión por computadora que analiza imágenes de hojas de tomate con una anormalidad visible y realiza un diagnóstico de una posible enfermedad, bajo un subconjunto predefinido de enfermedades del tomate.

1.2.2 Objetivos particulares.

* Entrenar el modelo de clasificación de imágenes con el conjunto de datos predefinido.
* Realizar pruebas de eficiencia del clasificador, obteniendo un resultado superior al 90%.
* Implementar un sistema web en el que se aloje el clasificador y permita realizar identificaciones de enfermedades a los usuarios.

1.3 Justificación.

La arquitectura general de sistemas de clasificación de enfermedades de plantas a través del análisis imágenes consiste de tres etapas: pre-procesamiento, extracción de descriptores y clasificación [3].

En el pre-procesamiento las imágenes son preparadas usando algunas operaciones, tales como conversión de espacio de color de RGB a algún otro. También suele incluirse la eliminación el fondo de la imagen tratando de concentrar el análisis en el objeto de interés. Desafortunadamente, este tipo de operación es complicada y algunas veces requiere de la intervención del usuario, lo cual decrece la automatización del sistema.

Los descriptores propuestos por expertos son extraídos de la imagen para formar un vector descriptor. Ejemplos de descriptores son los momentos de color o de textura obtenidos a partir de la matriz de co-ocurrencia. El objetivo de un descriptor es, a partir de su valor, poder clasificar un objeto entre distintas clases.

La última etapa determina a que clase pertenece el objeto no identificado a través de un modelo o algoritmo de clasificación. Este modelo debe ser entrenado usando algoritmos de aprendizaje y muestras de imágenes pre-clasificadas (muestras etiquetadas). Ejemplos de este tipo de algoritmos son la Máquina de Vectores de Soporte (SVM), los k-vecinos más cercanos (KNN) y la Red Neuronal Artificial (ANN).

Las etapas de pre-procesamiento y extracción de descriptores implican operaciones que suelen ser complejas y consumir tiempo considerable. Además, debido a esto es común solicitar asistencia del usuario lo que hace que el sistema no esté completamente automatizado.

En este proyecto se plantea el uso de un modelo de aprendizaje profundo (en adelante llamado deep learning), específicamente de una Red Neuronal Convolucional (CNN) como una alternativa a la clasificación de enfermedades en plantas. Este modelo permite que los descriptores sean definidos y obtenidos de una forma completamente automática. Además permite realizar la etapa de entrenamiento y clasificación directamente sobre las imágenes, sin incluir operaciones de pre-procesamiento. Las características anteriores permitirán desarrollar un sistema completamente automatizado en el que la única acción que el usuario tendrá que realizar es proporcionar la imagen de la hoja enferma.

1.4 Metodología.

El objetivo de este proyecto y los métodos para alcanzarlo están bien definidos, tal como se describió en las secciones anteriores. Teniendo esto en cuenta y considerando que hay dos etapas fácilmente identificables en el desarrollo del proyecto (el clasificador y el desarrollo del sistema web) se plante un modelo incremental de desarrollo de software.

El primer incremento consiste en el análisis, diseño y construcción del clasificador haciendo uso de un algoritmo CNN. El segundo incremento consiste en el análisis, diseño y construcción del sistema web en el que será integrado el clasificador para que esté disponible a los usuarios. Lo anterior se encuentra representado en la figura 1.1.



Figura 1.1 Metodología de desarrollo propuesta.

En los capítulos siguientes se desarrolla cada una de las etapas del modelo incremental propuesto.

Capítulo 2. Marco teórico

Marco teórico.

El deep learning es una tendencia reciente en al aprendizaje automático (machine learning), el cual ha tenido éxito en áreas como la visión por computadora, el reconocimiento automático de voz y el procesamiento de lenguaje natural.

2.1 Comparación de clasificación de imágenes usando modelos tradicionales de Machine Leaning y Deep Learning.

La clasificación de imágenes usando algoritmos de machine learning está compuesta de dos fases:

* Fase de entrenamiento. En esta fase se entrena el algoritmo usando un conjunto de datos pre-clasificados (muestras etiquetadas).
* Fase de predicción. En esta fase se utiliza el algoritmo entrenado para predecir la etiqueta de imágenes fuera del conjunto de entrenamiento.

La fase de entrenamiento para el problema de clasificación de imágenes tiene dos etapas principales:

1. Extracción de descriptores. En esta etapa se hace uso del conocimiento general en el área para seleccionar y extraer descriptores que serán usados por el algoritmo de machine learning, de acuerdo con el tipo de imágenes a clasificar. HOG y SIFT (Histograma de Gradientes Orientados y Transformación de Características Invariante a la Escala respectivamente) son ejemplos de descriptores usados en la clasificación de imágenes.
2. Entrenamiento del modelo. En esta etapa se hace uso de un conjunto de entrenamiento compuesto por descriptores de imágenes y sus etiquetas correspondientes para entrenar el modelo de machine learning.

En la fase de predicción, se aplica el mismo proceso de extracción de descriptores a imágenes nuevas y los descriptores obtenidos se pasan al algoritmo de machine learning para predecir su etiqueta o clase a la que pertenece. Ambas etapas, entrenamiento y predicción se muestran de forma esquemática en la figura 2.1.

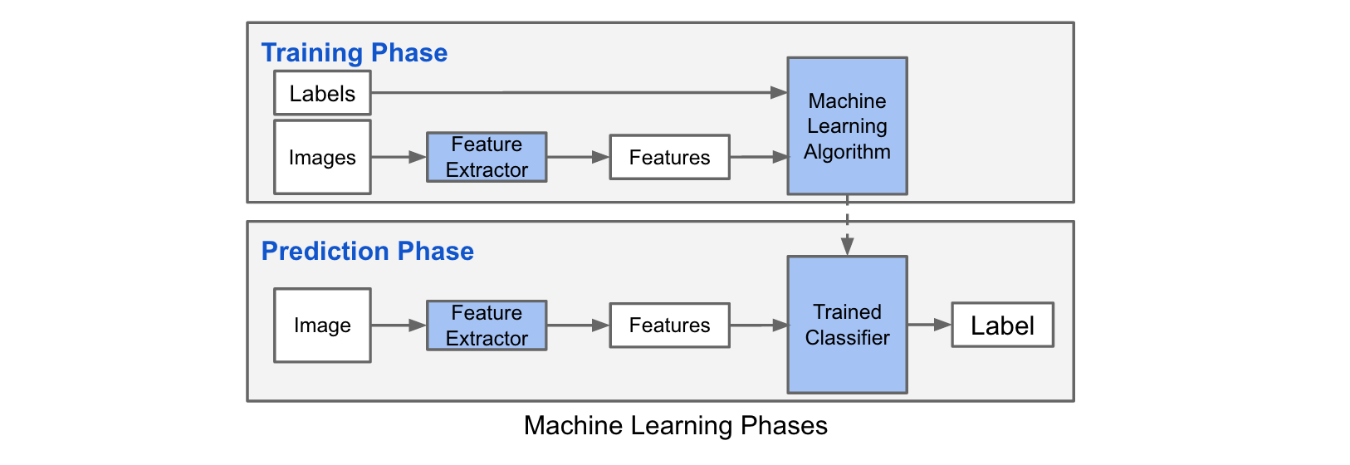


Figura 2.1 Clasificación con algoritmos de machine learning.

Ahora, la principal diferencia de los algoritmos tradicionales de machine learning y los algoritmos de deep learning radica en la ingeniería de descriptores. En los algoritmos tradicionales de machine learning este proceso es elaborado manualmente, es decir, la selección y extracción debe ser diseñada e implementada por el programador. En cambio, en los algoritmos de deep learining la ingeniería de descriptores es realizada automáticamente por el algoritmo. La ingeniería de descriptores es costosa, consume tiempo importante y requiere de cierta experiencia. Un descriptor mal seleccionado condena al fracaso el resto del clasificador. Es así como los algoritmos de machine learning prometen resultados más precisos comparados con los algoritmos tradicionales de machine learning, con menos o incluso sin ingeniería de descriptores. Esta diferencia se muestra en la figura 2.2.

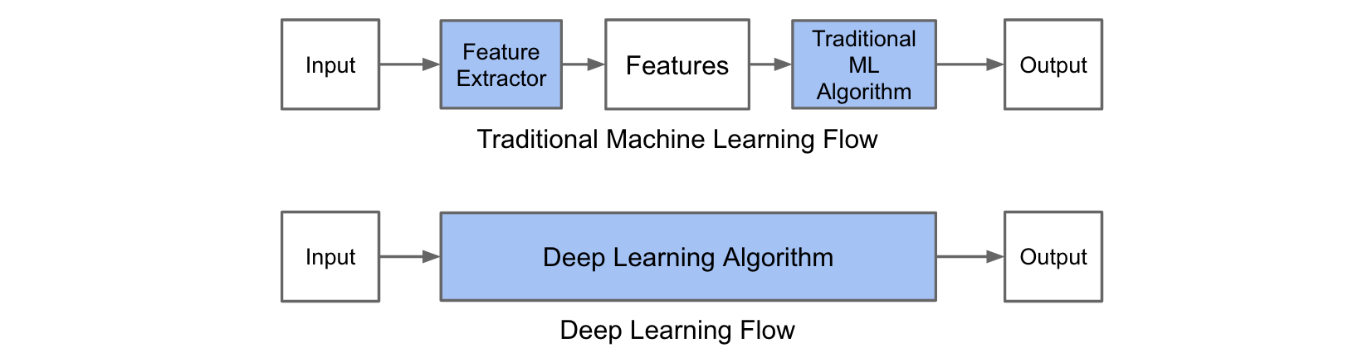


Figura 2.2 Diferencias entre la clasificación con modelos de machine learning y deep learning.

2.2 Redes Neuronales Artificiales (ANN).

Deep learning se refiere a una clase de redes neuronales artificiales compuestas de muchas capas de procesamiento. Las redes neuronales tienen décadas de existencia pero el reciente éxito y popularidad del deep learning puede rastrearse a la publicación “ImageNet Classification with Deep Convolutional Networks” realizada por Krizhevsky, Sutskever, and Hinton en el año 2012. En dicha publicación demostraron por primera vez como una CNN puede superar en desempeño, de forma contundente, a otros métodos tradicionales de clasificación de imágenes. Además, factores como el incremento del poder computacional de las computadoras, el uso de GPUs y la disponibilidad de grandes conjuntos de datos han favorecido el desarrollo de los modelos de deep learning.

Las neuronas artificiales están inspiradas en las neuronas biológicas. Una neurona artificial tiene un número finito de entradas con pesos asociadas a ellas, y una función de activación. La salida de la neurona es el resultado de la función de activación aplicada a la suma de las entradas con los pesos correspondientes (figura 2.3). Las neuronas artificiales se conectan entre sí para formar redes neuronales artificiales.

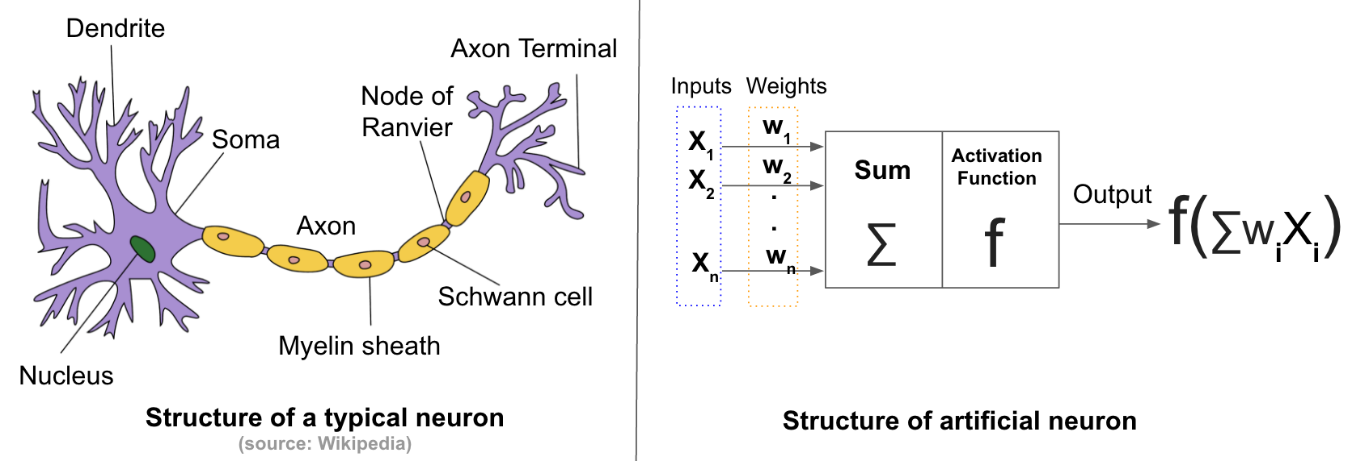


Figura 2.3 Estructura de una neurona artificial.

2.2.1 Redes neuronales unidireccionales.

Las redes neuronales unidireccionales (feedforward) son las más sencillas de las redes neuronales artificiales. Estas redes tienen tres tipos de capas, capa de entrada, capa oculta y capa de salida. Los datos se mueven de la capa de entrada hacia las neuronas ocultas y desde estas hacia las neuronas de salida. En la figura 2.4 se muestra un ejemplo de una red unidireccional en la que cada nodo o neurona está conectada a todas las de la siguiente capa. Este tipo de conexión es llamado “Completamente conectado” (fully-connected). El número de capas ocultas y su tamaño es variable. Mientras más grandes y de mayor profundidad sean las capas ocultas, pueden modelarse patrones más complejos.



Figura 2.4 Red neuronal unidireccional con dos capas ocultas.

2.2.2 Funciones de activación.

Las funciones de activación transforman la suma de las entradas con los pesos aplicados. Estas funciones deben ser no lineales para permitir codificar patrones de complejos de datos. Las funciones de activación más comunes son las funciones Sigmoidal, tangente hiperbólica y ReLU (figura 2.5). En las redes neuronales profundas (con muchas capas ocultas) la función más usada es la ReLU.



Figura 2.5 Funciones de activación Sigmoidal, tanh y ReLU.

2.2.3 Entrenamiento de una red neuronal artificial.

El objetivo de la fase de entrenamiento es determinar los valores de los pesos de la red neuronal. Para alcanzar este objetivo son necesarios dos elementos:

* El conjunto de datos de entrenamiento. En el caso de clasificación de imágenes el conjunto de datos está compuesto de imágenes y sus correspondientes etiquetas.
* Una función de pérdida. Está función mide la imprecisión de las predicciones.

Con estos dos elementos se entrena la red neuronal con un algoritmo llamado backpropagation [4].

2.3 Redes Neuronales Convolucionales (CNN).

Las redes neuronales convolucionales son un tipo especial de redes unidireccionales. Estos modelos están diseñados para emular el comportamiento de la corteza visual, por lo que asumen que la entrada de la red son imágenes. Las CNNs tienen un muy buen desempeño en tareas de reconocimiento visual. La arquitectura de una CNNs está compuesta por capas tres capas: Capa convolucional (CONV layer), capa de agrupamiento (pooling layer) y capa conectada completamente (Fully-connected layer). Estas capas permiten a la red codificar determinadas propiedades de la imagen. En la figura 2.6 se muestra la estructura de la CNN LeNet.



Figura 2.6 CNN LeNet.

2.3.1 Capa convolucional.

La capa convolucional es el bloque fundamental con que se construye una red convolucional y donde se realiza la mayor parte del trabajo computacional de la red. Los parámetros de la red convolucional consisten de un conjunto de filtros aprendibles. Cada filtro es pequeño espacialmente, pero es desplazado por toda la imagen calculando la convolución entre el filtro y la imagen de entrada. Conforme el filtro se desplaza por el alto y ancho de la imagen, se producen mapas de activación de dos dimensiones que indican la respuesta de dicho filtro en cada posición espacial. De forma intuitiva la red aprenderá los filtros que se activan cuando se encuentran algún tipo de característica visual, tales como bordes, orientaciones o regiones de algún color. Lo anterior en las primeras capas de la red. En capas superiores pueden detectarse características más complejas como ruedas, señales de tránsito, etc.

Cada neurona en la capa convolucional está conectada a una región local espacial de la imagen de entrada, pero considera todos los canales de color. En la figura 2.7 se tienen cinco neuronas de la capa convolucional, todas ellas conectadas a la misma región de la imagen de entrada.



Figura 2.7 Conectividad local de las neuronas de la capa convolucional.

2.3.2 Capa de agrupamiento.

Es común en una arquitectura CNN insertar periódicamente capas de agrupamiento entre capas convolucionales sucesivas, con la intención de progresivamente reducir el tamaño espacial de la representación para reducir la cantidad de parámetros y el cómputo en la red. De esta forma se controla el sobreajuste, para evitar entrenar de más la red.

La capa de agrupamiento ajusta el tamaño espacial empleando la operación MAX. La forma más común es tomando un filtro de 2x2 aplicado a toda la entrada espacialmente. De esta forma se reduce en un 75% las activaciones (figura 2.8).



Figura 2.8 Reducción del volumen en la capa de agrupamiento.

2.4 Arquitectura de las redes neuronales convolucionales.

La arquitectura más sencilla de una red neuronal convolucional comienza con una capa de entrada (imágenes) seguida de una secuencia de capas convolucionales y de agrupamiento, terminando con capas completamente conectadas. Las capas convolucionales usualmente están seguidas de una capa de funciones de activación ReLU (figura 2.9).

Las capas convolucionales, de agrupamiento y ReLU actúan como extractores de descriptores aprendibles, mientras que las capas completamente conectadas actúan como un clasificador de machine learning. Además, las primeras capas de la red codifican patrones genéricos de las imágenes, mientras que las capas posteriores codifican patrones detallados de las imágenes.

También debe mencionarse que solo las capas convolucionales y completamente conectadas tienen pesos, los cuales son aprendidos en la etapa de entrenamiento [5].



Figura 2.9 Ejemplo de arquitectura CNN.

Existen distintas arquitecturas con un nombre asignado. Las más comunes son las siguientes:

* LeNet. Desarrollada por Yann LeCun en la década de 1990. Con esta arquitectura se desarrollaron las primeras aplicaciones exitosas de redes convolucionales.
* AlexNet. Es el primer trabajo que popularizo las redes convolucionales en la visión por computadora, desarrollado por Alex Krizhevsky, Ilya Sutskever y Geoff Hinton. La AlexNet fue inscrita en el reto ImageNet ILSVRC challenge en 2012 y superó de forma significativa al segundo clasificado.
* ZF Net. Una mejora de la AlexNet que resultó ganadora del ILSVRC 2013, desarrollada por Matthew Zeiler y Rob Fergus.
* GoogLeNet. La ganadora del ILSVRC 2014, desarrollada por Szegedy et al [6] de Google.

Capítulo 3. Análisis del clasificador.

Análisis del clasificador.

En esta sección se describe el análisis del clasificador, correspondiente al primer incremento propuesto en la metodología de la sección 1.4.

Esta sección se realizó teniendo en consideración el conjunto de datos que se logró recopilar. Este conjunto de datos es un subconjunto del repositorio de acceso abierto publicado en [www.PlantVillage.org](http://www.PlantVillage.org). En el momento del desarrollo de este proyecto, dicho repositorio se encontraba temporalmente inhabilitado, por lo que las imágenes usadas en este proyecto fueron facilitadas por Brahimi M. [7].

El conjunto de datos usados en este sistema está formado por 14,828 imágenes divididas en nueve enfermedades del tomate, tal como se describe en la tabla 3.1.

|  |  |
| --- | --- |
| Enfermedad | Número de imágenes |
| Virus del rizado amarillo del tomate (Tomato yellow leaf curl virus) | 4032 |
| Virus del mosaico del tomate (Tomato mosaic virus) | 325 |
| Corynespora cassiicola. Mancha en forma de blanco. (Target spot) | 1,356 |
| Araña roja (Spider mites) | 1,628 |
| Septoriosis (Septoria spot) | 1,723 |
| Passalora fulva. Moho en la hoja (Leaf mold) | 904 |
| Tizón tardío (Lateblight) | 1,781 |
| Tizón temprano (Earlyblight) | 952 |
| Mancha bacteriana (Bacterial spot) | 2,127 |
| Total | 14,828 |

Tabla 3.1 Conjunto de datos del proyecto.

3.1 Requisitos del sistema.

Este proyecto nació de la necesidad de poder identificar de forma automática enfermedades en plantas, a partir de síntomas visuales. Y además, teniendo en consideración los objetivos y el alcance de este proyecto, los requisitos de software de este incremento se describen a continuación.

3.1.1 Requisitos de software.

Requisitos funcionales.

**RF01.** Clasificación.

**Nivel de madurez:** Alta.

**Prioridad:** Alta.

**Descripción:** El sistema clasificará la imagen en alguna de las nueve enfermedades descritas en la tabla 3.1, indicando la probabilidad de la predicción.

Requisitos no funcionales.

**RNF01**. La eficiencia de clasificación deberá ser superior al 90%.

**RNF02**. El clasificador identificará la clase de la imagen de entrada en un tiempo no mayor a cinco segundos.

3.1.2 Requisitos de hardware.

El mayor trabajo computacional se realiza durante la etapa de entrenamiento del clasificador, donde dependiendo del cpu, gpu y de la cantidad de memoria, el proceso puede extenderse por horas, o incluso días. Sin embargo, una vez entrenado el clasificador, clasificar una imagen es una tarea muy sencilla que se lleva a cabo en cuestión de algunos segundos. Por lo anterior, el dispositivo en el que se ejecuten las predicciones, ya con el clasificador entrenado, no se ven restringido en este sentido.

La etapa de entrenamiento será realizada con el framework Caffe. Caffe es un framework para deep learning desarrollado en C++ por Berkeley Al Research (BAIR). Los requisitos del sistema para este framework realmente dependen de las acciones que se vayan a realizar. Para este proyecto se realizará un fine-tuning (red pre-entrenada), lo que reduce el cómputo. Esto se detalla en el capítulo de diseño de este incremento. Por la razón anterior, una computadora con los requisitos siguientes es suficiente:

* CPU: Intel core i3.
* RAM: 4 GB.
* GPU\*: Se require CUDA para el modo GPU, versión 6 o superior.

La GPU es opcional pero su uso puede reducir el tiempo de cómputo a unas pocas horas. De forma similar, un mejor CPU o más RAM ayudan en el mismo sentido.

3.2 Modelo de casos de uso.

El clasificador no tiene interacción directa con el usuario, por lo que en este incremento un modelo de casos de uso no resulta aplicable. El clasificador se encontrará en un entorno en el que a través de una interfaz gráfica el usuario pueda clasificar sus imágenes. Esto se aborda en el análisis del segundo incremento.

3.3 Interfaz gráfica.

Por la misma razón expuesta en la sección 3.2, una interfaz gráfica no resulta aplicable en este incremento.

Capítulo 4. Diseño del clasificador.

Diseño del clasificador.

Como ya se ha mencionado anteriormente, el algoritmo clasificador será un algoritmo de deep leaning, en particular una red neuronal convolucional (CNN). Lo anterior debido a las ventajas que presenta al ser comparado con los algoritmos tradicionales de machine learning, especialmente en la etapa de la ingeniería de descriptores.

Algunas de las arquitecturas de CNNs fueron mencionadas en la sección 2.4. Entre ellas la AlexNet y GoogLeNet son las más populares debido a la calidad de los resultados obtenidos en distintas aplicaciones, tales como el “ImageNet Large Scale Visual Recognition Competition” (ILSVRC). De estas dos arquietecturas se ocupará la GoogLeNet debido a que fue la que gano dicha competición en una fecha posterior.

4.1 Fine-Tuning.

Para implementar una red neuronal convolucional se tiene básicamente dos opciones: entrenar toda la red neuronal desde cero (con valores iniciales aleartorios), o usar una red pre-entrenada con un conjunto muy grande de datos y adaptarla a las necesidades particulares.

En la práctica, muy pocas personas entrenan completamente una red convolucional desde cero, porque es muy raro disponer de un conjunto de datos del tamaño suficiente. En cambio es muy común pre-entrenar una red convolucional sobre un conjuntos de datos muy grande (por ejemplo ImageNet, que contiene 1.2 millones de imágenes en mil categorías distintas), y entonces usar la red convolucional, ya sea como inicialización o como un extractor de descriptores fijo para la tarea de interés modificando solo las últimas capas. Los tres principales escenarios de transferencia de aprendizaje son los siguientes:

* **CNN como un extractor de descriptores fijo.** Se toma una red convolucional pre-entrenada sobre ImageNet y se remueve la última capa completamente conectada (las salidas de esta red son las mil clases distintas de ImageNet). Entonces el resto de la red convolucional se trata como un extractor de descriptores fijo para un nuevo conjunto de datos. Este extractor de descriptores puede usarse para entrenar un clasificador lineal (SVM por ejemplo) para el nuevo conjunto de datos.
* **Calibración fina de la red convolucional (Fine-tuning).** Esta estrategia consiste en reemplazar y reentrenar el clasificador al final de la red convolucional sobre el nuevo conjunto de datos. También se contempla volver a ajustar los pesos de la red pre-entrenada continuando con el backpropagation. Es posible volver a ajustar todas las capas de la red o también conservar algunas de las primeras capas intactas y las demás ajustarlas. Esto último considerando la observación de que las primeras capas de la red contienen descriptores genéricos (bordes o colores por ejemplo) que deberían ser útiles en muchas aplicaciones, pero en las capas siguientes la red convolucional se vuelve progresivamente más específica de acuerdo a los detalles contenidos en las imágenes del conjunto de datos original.
* **Modelos pre-entrenados.** Debido a que el proceso de entrenamiento sobre ImageNet de las redes convolucionales modernas toma entre dos y tres semanas, considerando el uso de múltiples GPUs, es muy común que las personas compartan sus modelos entrenados para el beneficio de otros, quienes pueden usar esos modelos para realizar fine-tuning.

Ahora, teniendo en consideración que el conjunto de datos de este proyecto es relativamente grande (14,828 imágenes) es factible la realización de un fine-tuning sobre una red pre-entrenada en ImageNet [8].

4.2 Caffe. Deep Learning Framework.

Caffe es un framework para deep learning desarrollado en C++ por Berkeley Al Research (BAIR). Algunas de los beneficios de este framework son los siguientes [9]:

* Los modelos y la optimización son definidos mediante archivos de configuración sin la necesidad de hard-coding. Puede seleccionarse el modo CPU + GPU o únicamente CPU modificando una sola bandera.
* Código extensible. Durante su primer año, el repositorio de caffe fue bifurcado (forked) por más de mil desarrolladores que contribuyeron a cambios significativos mediante su retroalimentación. Gracias a estos colaboradores el framework se mantiene alineado con el estado del arte del deep learning.
* Caffe puede procesar alrededor de sesenta millones de imágenes por día con una sola GPU NVIDIA K40 (1 ms/imagen).
* Caffe cuenta con una comunidad formada por académicos, investigadores, ingenieros, etc. Incluso tiene un “Model Zoo” en el que los usuarios comparten sus modelos de Caffe para el beneficio mutuo de la comunidad.

Para realizar el entrenamiento de la red se tomará un modelo de caffe pre-entrenado de la GoogLeNet sobre ImageNet y se realizará un proceso de fine-tuning sobre el conjunto de datos de este proyecto.

Para lo anterior se instalará Caffe en una computadora con las siguientes características.

* Sistema operativo: Ubuntu 16.04
* CPU: Intel Core i3
* RAM: 4 GB

Dicho equipo no cuenta con una GPU (CUDA) compatible con Caffe, por lo que la instalación se realizara en modo solo CPU (CPU only). Esto genera el inconveniente de un mayor tiempo de cómputo, sin embargo para fines de este prototipo, no representa un problema mayor.

Capítulo 5. Construcción del clasificador.

Construcción del clasificador.

Capítulo 6. Análisis del sistema Web.

Análisis del sistema Web.

En esta sección se describen los requisitos de hardware y software, se presenta un modelos de caso de usos y la especificación de la interfaz gráfica, correspondientes al segundo incremento de la metodología propuesta en la sección 1.4.

6.1 Requisitos del sistema.

A continuación se presentan los requisitos funcionales y no funcionales de este incremento. Considerando que de cierta forma el prototipo de este incremento incluye al del primer incremento, aquí se incluyen los requisitos enunciados en la sección 3.1 más los propios de este incremento.

6.1.1 Requisitos de software.

Requisitos funcionales.

**RF01.** Selección de imagen.

**Nivel de madurez:** Alta.

**Prioridad:** Media.

**Descripción:** El sistema permitirá seleccionar una imagen del sistema de archivos local del dispositivo del usuario.

**RF02.** Clasificación.

**Nivel de madurez:** Alta.

**Prioridad:** Alta.

**Descripción:** El sistema clasificará la imagen en alguna de las nueve enfermedades descritas en la tabla 3.1, indicando la probabilidad de que esa sea la enfermedad correcta.

Requisitos no funcionales.

**RNF01**. La eficiencia de clasificación deberá ser superior al 90%.

**RNF02**. El clasificador identificará la clase de la imagen de entrada en un tiempo no mayor a cinco segundos.

**RNF05**. El sistema será desarrollado para un ambiente Web.

**RNF06**. El sistema podrá usarse a través del navegador Chrome v.64.

6.1.2 Requisitos de hardware.

En este incremento ya se tiene el clasificador entrenado. Con el modelo entrenado solo se requiere realizar forward propagation en la red para realizar las predicciones, tarea sencilla que se ejecuta en algunos segundos. Por lo que se tienen los siguientes requisitos de hardware, tomando como referencia las especificaciones de las instancias gratuitas de Amazon Web Services.

* 1 GB de memoriA RAM.
* 1 CPU Intel Xeon de alta frecuencia.

6.2 Modelo de casos de uso.

Desde el punto de vista del usuario, la aplicación le permitirá hacer las siguientes acciones. Seleccionar una imagen desde su dispositivo y obtener el diagnóstico de enfermedad de la planta capturada en la imagen. Esto se ilustra en el diagrama de casos de uso de la figura 6.1.



Figura 6.1 Diagrama de casos de uso del prototipo.

6.2.1 CU1. Seleccionar imagen.

Resumen.

El usuario accederá a la interfaz del sistema y seleccionará la imagen de la hoja con la enfermedad que quiere identificar.

Descripción.

|  |  |
| --- | --- |
| Caso de uso: CU1. Seleccionar imagen | |
| Versión | 1.0 |
| Actor | Usuario. |
| Propósito | Seleccionar la imagen con la hoja enferma. |
| Entradas | Ninguna. |
| Origen | IU1.0 Pantalla de inicio. |
| Salidas | Imagen seleccionada |
| Destino | IU1.0 Pantalla de inicio. |
| Precondiciones | Dar click en el botón “Seleccionar imagen”. |
| Postcondiciones | Imagen seleccionada y lista para enviar. |
| Errores | Extensión incorrecta del archivo seleccionado. |
| Reglas de negocio | Solo son admitidas las extensiones jpg, png. |
| Tipo | Caso de uso primario. |
| Observaciones | Solo se selecciona la imagen, aún no se envía para su clasificación. |
| Autor | Edgar Rodrigo Arredondo Basurto. |
| Estatus | Versión 1.0 Revisada. |

Trayectorias del caso de uso.

Trayectoria principal.

1 Da click en el botón “Seleccionar imagen”.

2 Selecciona un archivo en la ventana de dialogo. [Trayectoria A].

- - - - - - - *Fin del caso de uso.*

Trayectoria alternativa A.

Condición: El usuario selecciono un archivo con una extensión distinta a jpg o png.

1  Muestra el mensaje “Seleccione un archivo válido (\*.jpg, \*.png)”.

2  Termina el caso de uso.

- - - - - - - *Fin de la trayectoria.*

6.2.1 CU2. Obtener diagnóstico.

Resumen.

El usuario enviará la imagen preseleccionada y obtendrá como resultado la clasificación de la misma en alguna de las nueve categorías mostradas en la tabla 3.1.

Descripción.

|  |  |
| --- | --- |
| Caso de uso: CU2. Obtener diagnostico | |
| Versión | 1.0 |
| Actor | Usuario. |
| Propósito | Obtener la clasificación de la imagen con la hoja enferma. |
| Entradas | La imagen a clasificar. |
| Origen | IU1.0 Pantalla de inicio. |
| Salidas | Diagnóstico y su probabilidad. |
| Destino | IU1.0 Pantalla de inicio. |
| Precondiciones | Dar click en el botón “Enviar imagen”. |
| Postcondiciones | Ninguna. |
| Errores | Ninguno. |
| Reglas de negocio | Solo se puede clasificar en alguna de las nueve categorías listadas en la tabla 3.1. |
| Tipo | Caso de uso primario. |
| Observaciones | Ninguna. |
| Autor | Edgar Rodrigo Arredondo Basurto. |
| Estatus | Versión 1.0 Revisada. |

Trayectorias del caso de uso.

Trayectoria principal.

1 Da click en el botón “Enviar imagen”.

2 Muestra el resultado de la clasificación y su probabilidad.

- - - - - - - *Fin del caso de uso.*

6.3 Interfaz gráfica.

En el sistema solo es necesaria una pantalla, desde la cual el usuario selecciona la imagen la envía y recibe el diagnóstico.

6.3.1 IU1.0 Pantalla de inicio.

Objetivo.

Desde esta pantalla el usuario puede seleccionar la imagen desde su dispositivo y enviarla para obtener el diagnóstico. El resultado le será informado con un mensaje en esta misma pantalla.

Diseño.



Figura 6.2 IU1.0 Pantalla de inicio.

Referencias.

1. SAGARPA. Hidroponia rústica. [En línea] Disponible en: http://www.sagarpa.gob.mx.
2. Research and Markets. Global Hydroponics Market - Forecasts from 2017 to 2022. [En línea] Disponible en: https://www.researchandmarkets.com.
3. Mohanty S., Hughes D., Salathé M. Using Deep Learning for Image-Based Plant Disease Detection. 2016.
4. Moujahid A. A practical Introduction to Deep Learning with Caffe and Python. [En línea] Disponible en http://adilmoujahid.com
5. CS231n. Convolutional Neural Networks for Visual Recognition. [En linea] Disponible en http://cs231n.github.io/convolutional-networks/
6. Szegedy et al. Going Deeper with Convolutions. [En línea] Disponible en: https://arxiv.org/abs/1409.4842
7. Mohamed Brahimi. Computer Science Department. Mohamed El Bachir El Ibrahimi University.
8. CS231n. Transfer Learning. [En línea] Disponible en http://cs231n.github.io/transfer-learning/
9. Caffe. Deep Learning Framework. [En línea] Disponible en http://caffe.berkeleyvision.org/
10. Haralick, R. M., Shanmugan, K., and Dinstein, I. Textural features for image classification. 1973.