

PONTIFICIA UNIVERSIDAD JAVERIANA

TRABAJO DE GRADO

ESTIMACIÓN DE NIVELES DE NITRÓGENO EN CULTIVOS DE ARROZ A PARTIR DE  
TÉCNICAS DE APRENDIZAJE DE MÁQUINA BASADOS EN IMÁGENES  
MULTIESPECTRALES

*Juan Sebastián Caldas Laverde*  
*Natalia Andrea Cera Bornacelli*

Director:  
Julián David Colorado Montaña, PhD

Co-Director:  
Carlos Andrés Devia Pinzón, M.Sc

Noviembre, 2018

# Contenido

1. INTRODUCCIÓN .....	7
2. MARCO TEÓRICO .....	8
2.1. Espectro electromagnético .....	8
2.2. Imágenes Multiespectrales .....	8
2.3. Índices Vegetativos .....	10
2.4. Regresiones Lineales.....	10
2.4.1. Regresiones Lineales Simples .....	10
2.4.2. Regresiones Lineales Multivariable .....	10
2.5. Inteligencia Artificial .....	11
2.6. Máquinas de Soporte Vectorial (SVM).....	12
2.6.1. Máquinas de soporte vectorial para regresión .....	13
2.7. Redes Neuronales Artificiales .....	14
2.8. Coeficiente de Correlación.....	17
2.9. Coeficiente de Determinación (R <sup>2</sup> ).....	17
2.10. <i>Root Mean Squared Error</i> (RMSE) .....	18
2.11. Curva ROC .....	18
3. OBJETIVOS .....	19
3.1. Objetivo General .....	19
3.2. Objetivos Específicos.....	19
3.2.1. Objetivo Específico 1 .....	19
3.2.2. Objetivo Específico 2 .....	19
3.2.3. Objetivo Específico 3 .....	19
3.2.4. Objetivo Específico 4 .....	20
3.2.5. Objetivo Específico 5 .....	20
4. DESARROLLO .....	21
4.1. Antecedentes del Macro Proyecto.....	21
4.2. Desarrollo del Trabajo de Grado .....	23
4.2.1. Identificación de Índices .....	23
4.2.2. Extracción de Índices .....	25
4.2.3. Estimación de Niveles de Nitrógeno .....	26
4.2.4. Correlación con Mediciones Realizadas .....	31
4.2.5. Comparación de Desempeño entre Técnicas.....	31
5. PROTOCOLO DE PRUEBAS.....	33
5.1. Regresiones Lineales Multivariable .....	33
5.2. Máquinas de Soporte Vectorial .....	34
5.3. Redes Neuronales Artificiales .....	37
6. ANÁLISIS DE RESULTADOS .....	41

6.1.	Implementación de Técnicas de Aprendizaje de Máquina.....	41
6.1.1.	Regresiones Lineales Multivariable .....	41
6.1.2.	Máquinas de Soporte Vectorial .....	42
6.1.3.	Redes Neuronales Artificiales .....	44
6.2.	Análisis ROC .....	46
7.	CONCLUSIONES .....	48
8.	REFERENCIAS .....	49

## ÍNDICE DE FIGURAS

Figura 1. Espectro Electromagnético. [14].....	8
Figura 2. Imagen RGB .....	8
Figura 3. Imagen Banda Rojo .....	9
Figura 4. Imagen Banda Infrarrojo Cercano .....	9
Figura 5. Banda Verde .....	9
Figura 6. Espectro de Absorción de Pigmentos en Plantas [37].....	9
Figura 7. Campo de la Inteligencia Artificial.....	11
Figura 8. Campo del Aprendizaje de Máquina.....	12
Figura 9. Híper-planos de Separación [15] .....	13
Figura 10. SVM Lineal [38].....	14
Figura 11 SVM aplicando la función Kernel [38].....	14
Figura 12. Modelamiento simplificado de una neurona biológica. [6] .....	15
Figura 13. Funcionamiento general de una neurona artificial. [8] .....	15
Figura 14. Arquitectura Perceptrón Multicapa.....	17
Figura 15. Ejemplo Curva ROC [17] .....	18
Figura 16. Drones instrumentados e integrados para el Proyecto: Asctec Pelican con cámara multiespectral ADC Lite de Tetracam y Asctec Hummingbird con cámara multiespectral Sequoia de Parrot.....	21
Figura 17. Interfaz de Control de Misión.....	22
Figura 18. Interfaz de usuario para la configuración y control de la cámara multiespectral abordo del dron. .....	22
Figura 19. Diagrama en Bloques.....	23
Figura 20. Pseudo-algoritmo para la obtención de los índices a partir de las imágenes. ....	25
Figura 21. Interfaz toolbox Regression Learner.....	27
Figura 22. Interfaz Regression Learner – Response Plot .....	28
Figura 23. Interfaz Regression Learner – Response Plot .....	28
Figura 24. Interfaz de diseño de red.....	30
Figura 25. Interfaz de entrenamiento de la red.....	31
Figura 26. Gráfica de desempeño de entrenamiento .....	31
Figura 27. Curva ROC a partir de clasificación .....	32
Figura 28. Evaluación Regresiones Lineales Etapa Vegetativa .....	33
Figura 29. Evaluación Regresiones Lineales Etapa Reproductiva.....	33
Figura 30. Evaluación Regresiones Lineales Etapa Cosecha.....	33
Figura 31. Comparación Etapa Vegetativa (1) – SVR.....	34
Figura 32. Comparación Etapa Vegetativa (2)- SVR.....	34
Figura 33. Comparación Etapa Vegetativa (3) - SVR.....	35
Figura 34. Comparación Etapa Reproductiva (1) – SVR.....	35
Figura 35. Comparación Etapa Reproductiva (2) - SVR.....	35
Figura 36. Comparación Etapa Reproductiva (3) - SVR.....	35
Figura 37. Comparación Etapa Cosecha (1) – SVR.....	35
Figura 38. Comparación Etapa Cosecha (2) - SVR.....	35
Figura 39. Comparación Etapa Cosecha (3) - SVR.....	36
Figura 40. Relación entre correlación y $\epsilon$ - Etapa Vegetativa .....	36
Figura 41. Comparación variación $\epsilon$ – Etapa Vegetativa .....	36
Figura 42. Relación entre correlación y $\epsilon$ - Etapa Reproductiva.....	36
Figura 43. Comparación variación $\epsilon$ – Etapa Reproductiva.....	36
Figura 44. Relación entre correlación y $\epsilon$ - Etapa Cosecha.....	37
Figura 45. Comparación variación $\epsilon$ – Etapa Cosecha.....	37
Figura 46. Red neuronal con 2 capas .....	37
Figura 47. Red neuronal con 5 capas .....	37
Figura 48. Comparación entre el número de capas de una red.....	38
Figura 49. Implementación Redes Neuronales – Etapa Vegetativa .....	38

Figura 50. Implementación Redes Neuronales – Etapa Reproducción .....	39
Figura 51. Implementación Redes Neuronales – Etapa Cosecha .....	39
Figura 52. Comparación No. Neuronas – Etapa Vegetativa .....	40
Figura 53. Comparación No. Neuronas – Etapa Reproductiva .....	40
Figura 54. Comparación No. Neuronas – Etapa Cosecha .....	40
Figura 55. Resultados Estimación total Regresiones Lineales.....	41
Figura 56. Resultados Entrenamiento Único Regresiones Lineales.....	41
Figura 57. Análisis $\epsilon=2$ .....	43
Figura 58. Análisis $\epsilon=0.5$ .....	43
Figura 59. Estimación Total - SVR.....	43
Figura 60. Estimación Total Entrenamiento Único - SVR.....	43
Figura 61. Estimación Total – Entrenamiento Único.....	46
Figura 62. Estimación Total – Entrenamiento Múltiple.....	46
Figura 63. Curva ROC .....	47

## ÍNDICE DE TABLAS

Tabla 1. Funciones de Transferencia para Redes Neuronales .....	16
Tabla 2. Índices Vegetativos Seleccionados .....	24
Tabla 3. Comparación Regresiones Lineales .....	41
Tabla 4. Comparación Regresiones Lineales – Tipo de Implementación .....	41
Tabla 5. Análisis de Resultados - Etapa Vegetativa.....	42
Tabla 6. Análisis de Resultados – Etapa Reproductiva.....	42
Tabla 7. Análisis de Resultados – Etapa Cosecha.....	42
Tabla 8. Análisis Variación $\epsilon$ – Etapa Vegetativa.....	42
Tabla 9. Análisis Variación $\epsilon$ – Etapa Reproductiva.....	42
Tabla 10. Análisis Variación $\epsilon$ – Etapa Cosecha.....	43
Tabla 11. Estimación Total - SVR .....	44
Tabla 12. Comparación SVR– Tipo de Implementación .....	44
Tabla 13. Comparación Número de Capas en Redes Neuronales .....	44
Tabla 14. Comparación funciones de aprendizaje – Etapa Vegetativa .....	44
Tabla 15. Comparación funciones de aprendizaje – Etapa Reproductiva .....	44
Tabla 16. Comparación funciones de aprendizaje – Etapa Cosecha.....	44
Tabla 17. Comparación No. Neuronas capa oculta – Etapa Vegetativa.....	45
Tabla 18. Comparación No. Neuronas capa oculta – Etapa Reproductiva .....	45
Tabla 19. Comparación No. Neuronas capa oculta – Etapa Cosecha .....	45
Tabla 20. Comparación Estimación Total – Redes Neuronales .....	46
Tabla 21. Comparación entre implementaciones de redes neuronales.....	46
Tabla 22. Tabla Resultante Total .....	46
Tabla 23. Accuracy .....	47

# 1. INTRODUCCIÓN

El nitrógeno es un nutriente esencial para los seres vivos, ya que es uno de los constituyentes principales de compuestos vitales tales como aminoácidos, proteínas, enzimas, nucleoproteínas y ácidos nucleicos, así como también de las paredes celulares y clorofila en los vegetales [1]. En las plantas este elemento se encuentra presente en grandes cantidades y su importancia es tal que puede representar diferencias importantes en los rendimientos de los cultivos.

A grandes rasgos la importancia del nitrógeno radica en crear masa vegetal. Siendo más específicos, este elemento estimula el crecimiento al favorecer la división celular. Además, está involucrado en una gran cantidad de procesos. Uno de los más importantes es la producción de clorofila, pues como forma parte de ella es necesaria para su síntesis.

El nitrógeno es, además, el nutriente que en general más influye en el rendimiento y calidad del producto a obtener en la actividad agropecuaria [1]. Éste es componente más limitante en la producción de cultivos y su rendimiento es importante para el sustento económico de los sistemas de cultivo [2].

En algunos países está emergiendo una nueva tecnología aplicada al manejo de los sistemas de producción agrícola. Esta tecnología, denominada Agricultura de Precisión, involucra el mapeo intensivo de las características del suelo y el seguimiento del cultivo con métodos de manejo computarizado de datos, Sistemas de Posicionamiento Global (GPS) y Sistemas de Información Geográfica (GIS), para la determinación de las densidades de siembra, así como de las cantidades y de la frecuencia de aplicación de pesticidas y de fertilizantes [4].

Teniendo en cuenta la necesidad de medir y controlar las variabilidades dentro del terreno del cultivo que influyen en el desarrollo adecuado de la planta y el rendimiento de la cosecha, las tecnologías utilizadas en cada etapa de la agricultura de precisión han evolucionado conjuntamente ofreciendo soluciones integradas que se han materializado en las maquinarias, equipos de medición, software especializado y métodos de manejo orientados al desarrollo eficiente de esta actividad, sobrellevando los retos técnicos planteados en el sector agropecuario [8]. El uso y desarrollo de tecnologías para agricultura de precisión es un aspecto clave para mejorar de manera acelerada la actividad agrícola en aspectos económicos y ecológicos en las regiones productivas de los países en vía de desarrollo.

Una de las tendencias tecnológicas de mayor utilidad es la teledetección a través de sensores multiespectrales aerotransportados por vehículos aéreos no tripulados (UAV). Estos vehículos permiten obtener información sobre el estado del cultivo de manera autónoma utilizando un amplio rango de sensores que lleva a bordo, tales como cámaras RGB, multiespectrales, térmicas, entre otras [13].

La captura de estas imágenes se realiza simultáneamente en varios rangos del espectro electromagnético llamados bandas, y la necesidad de su adquisición radica en las propiedades espectrales (reflectancia, transmitancia y absorbancia, entre otras) del suelo y de las plantas en sus diferentes etapas de desarrollo, las cuales van ligadas estrechamente a factores como actividad fotosintética, cantidad de agua, clorofila, biomasa, entre otros. [8]

Evidentemente, el sector agrícola está experimentando un gran cambio derivado directamente a los avances tecnológicos. La adopción de técnicas de percepción remota en este dominio ha llevado a la generación de grandes cantidades de datos. Este volumen se encuentra creciendo constantemente y ha llegado a un punto en el que analizar esta gran cantidad de datos va más allá de la capacidad humana, en particular cuando los datos no son homogéneos [9]. Es acá donde entra en escena el Aprendizaje de Máquina, la cual es una tecnología emergente que brinda herramientas para facilitar el descubrimiento de reglas y patrones en grandes conjuntos de datos.

Partiendo de este hecho, el presente trabajo de grado tiene como fin estimar la concentración de nitrógeno en las plantaciones de arroz ubicadas en las locaciones de Santa Rosa Meta y Palmira, Valle del Cauca a partir de la integración de un sistema UAV equipado con un conjunto de cámaras RGB-NIR, realizando el análisis de datos con distintas técnicas de aprendizaje de máquina tales como regresiones lineales multivariable, redes neuronales artificiales y máquinas de soporte vectorial.

## 2. MARCO TEÓRICO

### 2.1. Espectro electromagnético

Se denomina espectro electromagnético a la distribución energética del conjunto de las ondas electromagnéticas. El espectro electromagnético se extiende desde la radiación de menor longitud de onda, como los rayos gamma y los rayos X, pasando por la luz ultravioleta, la luz visible y los rayos infrarrojos, hasta las ondas electromagnéticas de mayor longitud de onda, como son las ondas de radio.

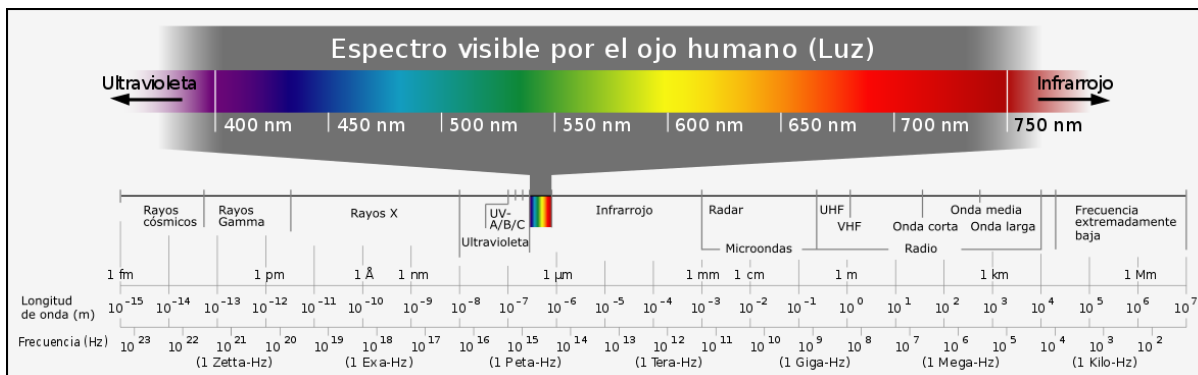


Figura 1. Espectro Electromagnético. [14]

La figura anterior (1) muestra el espectro electromagnético, relacionando los colores con su correspondiente longitud de onda, longitudes de onda menores a 400nm y mayores a 750nm, son imperceptibles al ojo humano.

### 2.2. Imágenes Multiespectrales

Las imágenes multiespectrales están definidas como la adquisición de la misma imagen en múltiples regiones del espectro electromagnético. Las longitudes de onda pueden estar separadas por filtros o mediante el uso de instrumentos sensibles a longitudes de onda particulares, frecuencias más allá del rango visible como infrarrojo y ultravioleta pueden brindar información adicional acerca del estudio que se esté realizando [20].

Para la presente aplicación, los instrumentos están diseñados para el estudio de parámetros relativos a la vegetación, por lo que las bandas están seleccionadas en los rangos del verde, rojo e infrarrojo cercano, donde la vegetación presenta su mayor respuesta de absorbancia y reflectancia [18]. A continuación, en las figuras 2, 3, 4 y 5 se muestra un ejemplo de las imágenes capturadas en distintas bandas del espectro electromagnético.



Figura 2. Imagen RGB



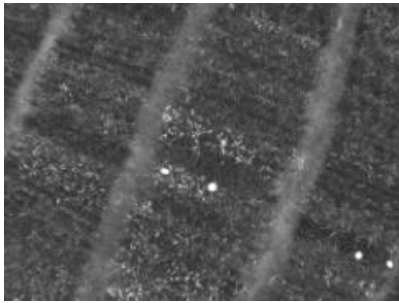


Figura 3. Imagen Banda Rojo

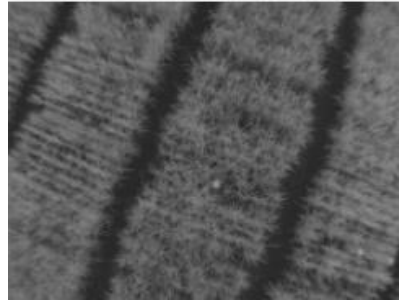


Figura 4. Imagen Banda Infrarrojo Cercano

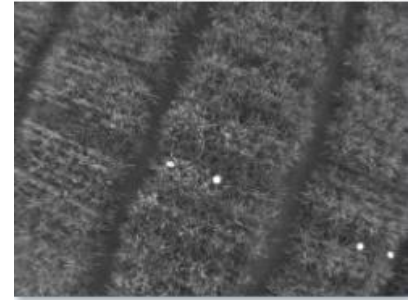


Figura 5. Banda Verde

Las plantas son expertas en capturar la energía de la luz y utilizarla para crear azúcares mediante un proceso llamado fotosíntesis. Este proceso comienza con la absorción de luz mediante moléculas orgánicas especializadas llamadas pigmentos, que se encuentran en los cloroplastos de las células de las plantas [19].

En la fotosíntesis, la energía solar se convierte en energía química mediante organismos fotosintéticos. Sin embargo, en la fotosíntesis no se usan de igual manera todas las distintas longitudes de onda en la luz del sol, ya que los pigmentos absorben solo longitudes de onda específicas de la luz visible, mientras que reflejan otras. Esta característica es denominada reflectancia, definida como la relación entre la potencia electromagnética incidente con respecto a la potencia que es reflejada en una interface [19].

El conjunto de longitudes de onda que absorbe un pigmento se conoce como su espectro de absorción siendo la absorbancia la relación de la potencia espectral radiante, incidente y reflejada a través de un material. En el siguiente diagrama, se pueden ver los espectros de absorción de tres pigmentos importantes en la fotosíntesis: clorofila a, clorofila b y  $\beta$ -caroteno. El conjunto de longitudes de onda que un pigmento no absorbe, se refleja, y la luz reflejada es lo que vemos como color. Por ejemplo, percibimos las plantas de color verde por su gran contenido de moléculas de clorofila a y b, que reflejan luz verde [19].

Espectro de absorción de los pigmentos

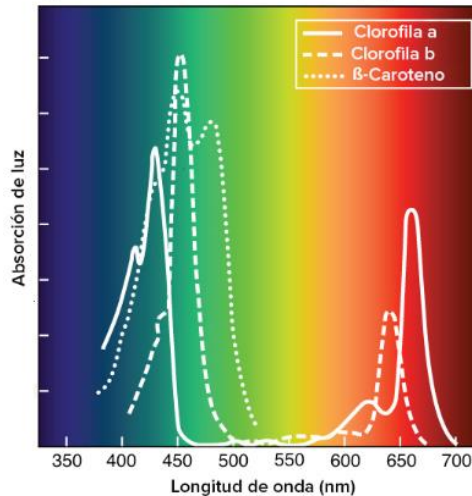


Figura 6. Espectro de Absorción de Pigmentos en Plantas [37]

La figura anterior (6) muestra el comportamiento de las moléculas de clorofila en cuanto a su absorbancia, a partir de esta gráfica se puede concluir que entre las longitudes de onda 450nm- 550nm existe un bajo índice de absorción de luz, por lo que la reflectancia aumenta y el color que muestran las plantas son de color entre verdes y amarillos.

La clorofila en la hoja está estrechamente relacionada con la concentración de N y, por lo tanto, refleja el estado nutricional con respecto a este importante nutriente. El N es necesario para la síntesis de la clorofila y como parte de esta molécula, está involucrado en el proceso de la fotosíntesis (Salisbury y Ross, 1992; Potash and Phosphate Institute, 1997). Cantidades adecuadas de N en la planta, producen hojas de color

verde oscuro debido a que estas tienen alta concentración de clorofila. El pigmento verde de la clorofila absorbe la energía de la luz necesaria para iniciar la fotosíntesis [18].

## 2.3. Índices Vegetativos

Un Índice de Vegetación, puede ser definido como un parámetro calculado a partir de los valores de la reflectancia a distintas longitudes de onda, y que es particularmente sensible a la cubierta vegetal (Gilbert et al, 1997). También, corresponde a un número generado por alguna combinación algebraica de bandas espectrales y que tienen cierta relación con la cantidad de la vegetación presente en una imagen dada. Estos índices, también son utilizados para mejorar la discriminación entre el suelo y la vegetación, reduciendo el efecto del relieve en la caracterización espectral de las diferentes cubiertas [21].

Un ejemplo de índice de vegetación muy simple puede ser debido a los cambios en alguno de los rangos del espectro electromagnético mencionados anteriormente, logrados dividiendo la medida de la reflectancia del infrarrojo por la del rojo. A mayor contraste entre las medidas del rojo e infrarrojo mayor vigor vegetal presentará la cubierta observada (más alta será la relación entre los sitios donde hay vegetación). Bajos valores de contraste significan una vegetación enferma hasta llegar a zonas sin vegetación (suelo y agua), cuyo resultado de la relación es muy pequeño [17].

## 2.4. Regresiones Lineales

Un modelo de regresión es un modelo matemático que busca determinar la relación entre una variable dependiente (Y) con respecto a otras variables llamadas explicativas o independientes (X) también llamadas variables predictoras o regresores [25]. Las regresiones pueden presentarse en diferentes tendencias dependiendo de la distribución de los datos.

### 2.4.1. Regresiones Lineales Simples

Regresión, supone que el propósito del experimento es la predicción. Significa que se intentará obtener una ecuación mediante la cual pueda predecirse el valor de una variable dependiente Y con la sola entrada de un regresor X. Si se utilizara más de una variable independiente para predecir el valor de Y entonces se emplearía el término regresión múltiple [26]. Y lineal alude a que la ecuación de la predicción tomará la forma de una línea recta. La ecuación 1 muestra el modelo de la regresión lineal simple.

$$Y = b_0 + b_1 * X$$

*Ecuación 1.*

Donde Y es la variable dependiente o explicada; X corresponde a la variable explicativa,  $b_1$  mide la influencia que la variable explicativa tiene sobre la variable dependiente, y  $b_0$  es la intersección o término constante.

### 2.4.2. Regresiones Lineales Multivariable

La regresión lineal múltiple contiene más de una variable independiente ( $X_1, X_2, X_3$ ) para ser relacionada a la variable dependiente Y; esto ofrece la ventaja de utilizar más información en la construcción del modelo y, consecuentemente, realizar estimaciones más precisas [27].

Sin embargo, al igual que en regresión lineal simple, se considerará que los valores de la variable dependiente Y han sido generados por una combinación lineal de los valores de una o más variables explicativas y un término aleatorio [28].

La fórmula que explica el comportamiento de las regresiones lineales multivariable se muestra en la ecuación 2.

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 + b_3 * X_3 + \dots + b_n * X_n$$

Ecuación 2.

Donde Y es la variable dependiente o explicada;  $X_n$  corresponde a cada una de las variables explicativas,  $b_1$  a  $b_n$  miden la influencia que cada variable explicativa correspondiente tiene sobre la variable dependiente, y  $b_0$  es la intersección o término constante.

## 2.5. Inteligencia Artificial

Inteligencia Artificial (IA) es la teoría y el desarrollo de sistemas de computadora capaces de desarrollar tareas que normalmente requieren de inteligencia humana, tales como la percepción visual, el reconocimiento de voz, la toma de decisión y la traducción entre dos idiomas [29]. Ampliamente usada en diagnóstico médico, comercio de acciones, control robótico, leyes, percepción remota, descubrimientos científicos incluso en el campo de la juguetería.

La inteligencia artificial engloba el concepto de aprendizaje automático o *Machine Learning*, cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento [30], la diferencia entre los 2 conceptos radica en que la IA es el concepto más amplio, en donde las máquinas son capaces de llevar a cabo las tareas de una manera considerada “inteligentes”, mientras que *Machine Learning* es una aplicación actual de la Inteligencia Artificial basada en la idea de que se debería dar a las máquinas acceso a los datos y dejarles aprender por sí mismos.

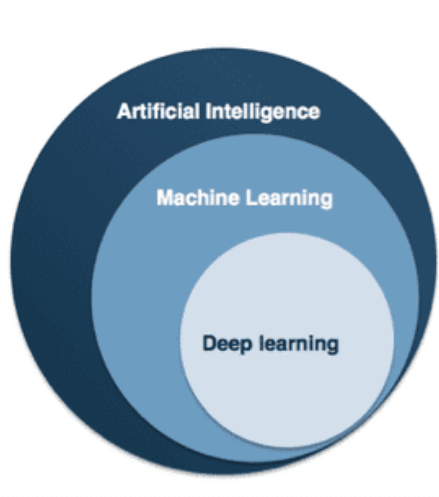


Figura 7. Campo de la Inteligencia Artificial

El aprendizaje automático emplea dos tipos de técnicas: el aprendizaje supervisado, que entrena un modelo con datos de entrada y salida conocidos para que pueda predecir salidas futuras, y el aprendizaje no supervisado, que encuentra patrones ocultos o estructuras intrínsecas en los datos de entrada [31].

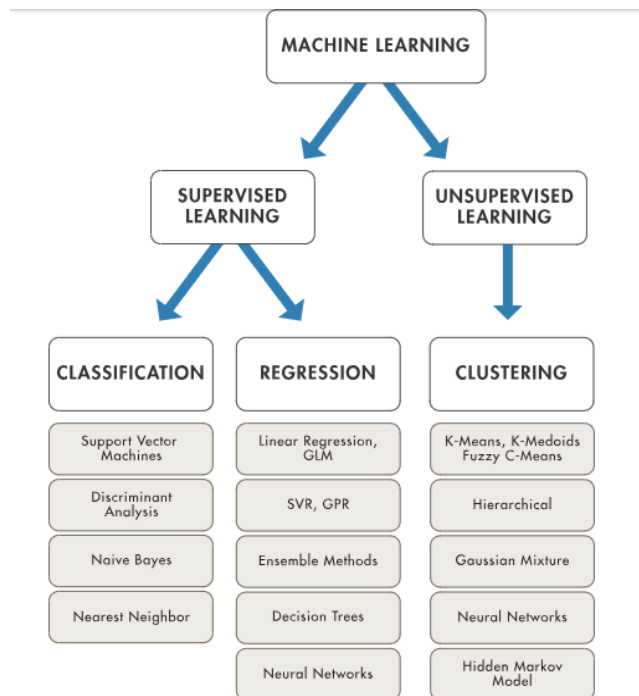


Figura 8. Campo del Aprendizaje de Máquina

Aterrizando los conceptos expuestos anteriormente el presente trabajo corresponde a un aprendizaje supervisado enfocado a regresión, se entrará en detalle en cuanto a las técnicas de *Machine Learning* aplicadas en el presente trabajo de grado.

## 2.6. Máquinas de Soporte Vectorial (SVM)

Las máquinas de soporte vectorial, o SVM (por sus siglas en inglés Support Vector Machine), corresponden a un nuevo sistema de aprendizaje que, en los últimos años, se ha convertido en el centro de atención, tanto en el desarrollo de nuevos algoritmos, como en las estrategias para su implementación. Las SVM fueron desarrolladas por Vapnik en el año 1995, para enfrentar los problemas de clasificación. Sin embargo, el actual método de Máquinas de soporte vectorial para regresión (SVMR por sus siglas en inglés) fue desarrollada en los laboratorios de AT&T por este mismo [32].

Este algoritmo trata de buscar el hiper-plano que separe con el mayor margen posible las clases u observaciones. Originalmente se usaba para resolver problemas binarios (2 clases), sin embargo, con el desarrollo de la técnica ahora es posible solucionar problemas de una mayor complejidad discriminando 3 o más clases.

Hay una gran cantidad de soluciones a hiper-planos de separación entre 2 clases (figura 9). No obstante, se escoge el plano que se acomoda a obtener un margen mayor tanto superior como inferior. Se crea el concepto de vectores de soporte que se definen como la observación de cada clase que están más cerca al hiper-plano de separación, generando 2 rectas o hiper-planos que pasan por los vectores de soporte siendo paralelos al hiper-plano de separación original.

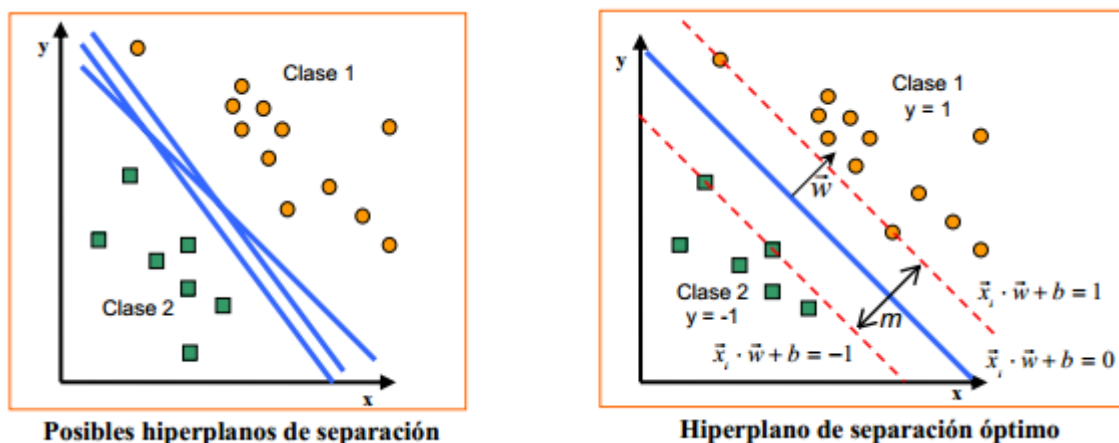


Figura 9. Híper-planos de Separación [15]

Para que el SVM pueda ser eficiente en conjuntos no linealmente separables se tienen dos opciones. Una en la que se permite un solapamiento entre clases, y otra en la cual se transforma el espacio de características original en otro linealmente separable (truco Kernel).

*Truco Kernel:* Un problema donde las entradas sean linealmente separables es muy utópico en el mundo real. Para esto existe una solución, mapear el problema a un nuevo espacio utilizando transformaciones no lineales,  $z = \Phi(x)$ , espacio donde el problema si es linealmente separable. Para esto se utiliza la función Kernel.

Por definición, una función Kernel es una función que asigna a cada par de elementos del espacio de entrada,  $X$ , un valor real correspondiente al producto escalar de las imágenes de dichos elementos en un nuevo espacio  $F$  (espacio de características). Los principales tipos de Kernel a utilizar en la implementación son:

- Kernel lineal:

$$K(x, x') = \langle x, x' \rangle$$

*Ecuación 3.*

- Kernel polinómico grado  $p$ :

$$K_p(x, x') = [\gamma \langle x, x' \rangle + \tau]^p$$

*Ecuación 4.*

- Kernel gaussiano

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right), \quad \gamma > 0$$

*Ecuación 5.*

El parámetro  $\gamma$ , conocido también como escala del Kernel, dependiendo de su valor expandirá las opciones en la función de Kernel gaussiano, dando origen los Kernel Gaussiano fino, Gaussiano medio y Gaussiano grueso, los cuales se diferencian en la flexibilidad en cuanto a la tendencia de datos. Un valor de escala de Kernel más bajo brindará un modelo más ajustado, mientras que un valor de escala de Kernel más elevado dará como resultado un modelo más rígido en cuanto a la tendencia de los datos.

### 2.6.1. Máquinas de soporte vectorial para regresión

En este caso, la idea es seleccionar el hiper-plano regresor que mejor se ajuste al conjunto de datos de entrenamiento. Ahora no se dispone de clases para separar. La idea se basa en considerar una distancia margen  $\epsilon$ , de modo que es de esperar que todos los ejemplos se encuentren en una banda o tubo entorno al hiper-plano, es decir, que disten una cantidad menor de  $\epsilon$  del hiper-plano. A la hora de definir el hiper-plano

sólo se consideran los ejemplos que disten más de  $\varepsilon$  del hiper-plano original. En este caso esos ejemplos serán los considerados como vectores soporte [34].

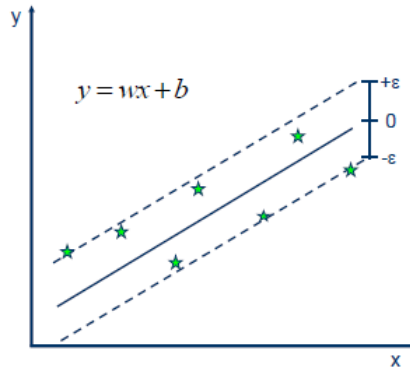


Figura 10. SVM Lineal [38]

El uso de Kernel en esta aplicación de regresión para máquinas de soporte vectorial tiene el mismo uso y principios que la aplicación de clasificación.

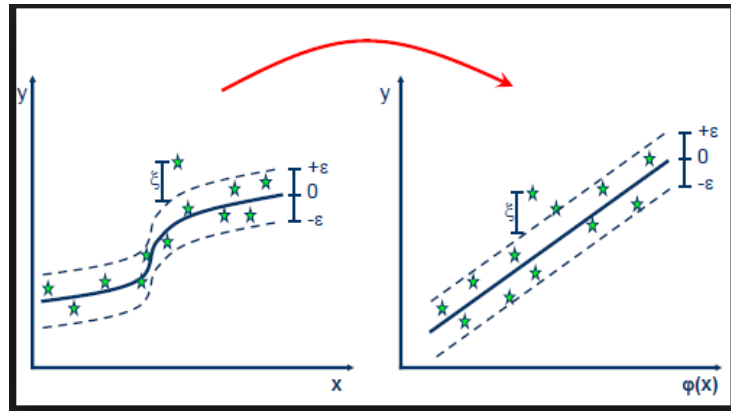


Figura 11 SVM aplicando la función Kernel [38]

## 2.7.Redes Neuronales Artificiales

Una red neuronal artificial (Artificial Neural Network ANN) es un modelo computacional que busca modelar el comportamiento biológico de las neuronas y cómo se organizan formando la estructura del cerebro [6]. Las ANN tienen una inclinación natural a adquirir el conocimiento a través de la experiencia, el cual es almacenado, al igual que el cerebro, en el peso relativo de las conexiones interneuronales.

Las redes neuronales artificiales no buscan modelar exactamente el comportamiento fisiológico de la neurona, sino que más bien solo sus características más relevantes, que entran en juego con su interacción con toda la red [6]. Aunque el tiempo de conmutación de las neuronas biológicas es casi un millón de veces mayor que las actuales componentes de las computadoras, las neuronas naturales tienen una conectividad miles de veces superior a la de las artificiales [9].

El objetivo principal de las redes neuronales de tipo biológico es desarrollar operaciones de síntesis y procesamiento de información, relacionadas con los sistemas biológicos [9]. Las neuronas y las conexiones entre ellas, llamadas sinapsis, constituyen la clave para el procesamiento de la información.

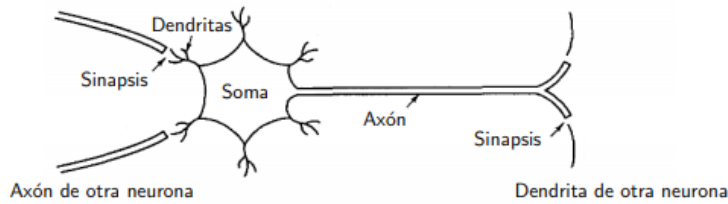


Figura 12. Modelamiento simplificado de una neurona biológica. [6]

Una neurona típica posee el aspecto y las partes que se muestran en la figura 12. Una neurona consta de tres partes:

1. El cuerpo de la neurona o soma
2. Las dendritas, que reciben las entradas
3. El axón, que lleva la salida de la neurona a las dendritas de otras neuronas.

La neurona envía su información de salida a otras por medio de su axón, codificando sus salidas como una serie de breves pulsos periódicos, llamados potenciales de acción. Luego, este pulso llega a las sinapsis y de ahí a las dendritas de la neurona siguiente [6].

Es importante resaltar que el impulso que llega a una sinapsis y el que sale de ella no son iguales en general. El tipo que saldrá depende muy sensiblemente de la cantidad de neurotransmisor. Esta cantidad de neurotransmisor cambia durante el proceso de aprendizaje, es aquí donde se almacena la información. Una sinapsis modifica el pulso, ya sea reforzándolo o debilitándolo [6].

En el caso de las ANN, las neuronas reciben señales de entrada de las neuronas vecinas ponderadas por los pesos de las conexiones. La suma de estas señales ponderadas proporciona la entrada total o neta de la neurona y, mediante la aplicación de una función matemática – denominada función de salida -, sobre la entrada neta, se calcula un valor de salida, el cual es enviado a otras neuronas. Tanto los valores de entrada a la neurona como su salida pueden ser señales excitatorias (cuando el valor es positivo) o inhibitoras (cuando el valor es negativo) [7].

La neurona o nodo de la red artificial es el elemento básico de computación. Éste recibe una entrada desde otras unidades o de una fuente externa de datos. Cada entrada tiene un peso asociado  $w$ , que se va modificando en el llamado proceso de aprendizaje [9]. El funcionamiento general de una neurona artificial es mostrado en la figura 13.

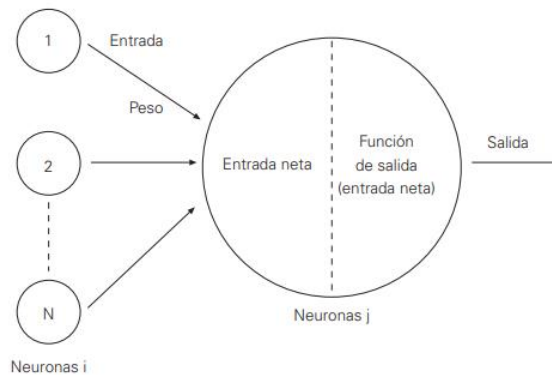


Figura 13. Funcionamiento general de una neurona artificial. [8]

Cada neurona transforma la señal recibida mediante una función de activación para determinar la señal de salida [10]. La función de activación, la cual determina el estado de activación actual de la neurona con base a la entrada neta de la neurona. Las principales funciones de activación utilizadas en el campo de redes neuronales artificiales se encuentran mostradas en la tabla 1.

	Función	Rango	Gráfica
--	---------	-------	---------

Lineal a tramos	$y = \begin{cases} 1, & \text{si } x > 1 \\ x, & \text{si } -1 < x < 1 \\ -1, & \text{si } x < -1 \end{cases}$	[-1,1]	
Sigmoidea logarítmica	$y = \frac{1}{1 + e^{-x}}$	[0,1]	
Sigmoidea tangencial	$y = \tanh(x)$	[-1,1]	

Tabla 1. Funciones de Transferencia para Redes Neuronales

Las redes neuronales que componen una ANN se organizan de forma jerárquica formando capas. Una capa o un nivel es un conjunto de neuronas cuyas entradas de información provienen de la misma fuente (que puede ser otra capa de neuronas) y cuyas salidas de información se dirigen al mismo destino (que puede ser otra capa de neuronas [7]). En este sentido, se distinguen tres tipos de capas: la capa de entrada, que recibe la información del exterior; la o las capas ocultas, que son aquellas cuyas entradas y salidas se encuentran dentro del sistema y, por tanto, no tienen contacto con el exterior; por último, la capa de salida, que envía la respuesta de la red al exterior.

En función de la organización de las neuronas de la red formando capas o agrupaciones, se encuentran dos tipos de arquitecturas básicas: redes multicapa y redes monocapa [7]. Las redes multicapa disponen de conjuntos de neuronas agrupadas en dos o más capas, formadas por una capa de entrada, una capa de salida y una o más capas intermedias u ocultas [7].

Por su parte, en las redes monocapa, se establecen conexiones entre neuronas que pertenecen a la única capa que constituye la red. Las redes monocapa se utilizan generalmente en tareas relacionadas con la autoasociación, que consiste en regenerar información de entrada que se presenta a la red de forma incompleta o distorsionada [10].

La conectividad entre los nodos de una red neuronal está relacionada con la forma en la que las salidas de las neuronas están canalizadas para convertirse en entradas de otras neuronas. La señal de salida de un nodo puede ser una entrada de otro elemento de proceso, o incluso ser una entrada de sí mismo (conexión autorrecurrente) [10].

En cuanto al número de capas, las redes neuronales monocapa tienen conexiones de tipo lateral, es decir, con neuronas de su mismo nivel [7]. Estas conexiones pueden ser hacia un elemento distinto de la red o el elemento mismo.

Por otro lado, para las redes multicapa, las conexiones entre las distintas capas pueden ser hacia adelante, donde todas las neuronas de una capa reciben señales de entrada desde otra capa anterior (la más cercana a la entrada de la red), y envían señales de salida a una capa posterior (que está más cerca a la salida de la red) o, hacia atrás, conectando la salida de las neuronas de capas posteriores a la entrada de capas anteriores [10].

La implementación de distintas arquitecturas varía según el uso final al que es destinado la red neuronal. Las ANN son un método para resolver problemas para aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción [26]. Al diseñar una red neuronal, el objetivo es encontrar la combinación que mejor se ajusta entrenando a la red neuronal.

Una de las arquitecturas más utilizadas en la resolución de problemas es el perceptrón multicapa (red multicapa con conexiones hacia adelante), el cual es utilizado como aproximador universal, en el sentido de que cualquier función continua puede aproximarse con un perceptrón multicapa, con al menos una capa de neuronas ocultas [27]. El perceptrón multicapa es un modelo matemático útil a la hora de aproximar o interpolar relaciones no lineales entre datos de entrada y salida. Se puede observar la arquitectura del perceptrón multicapa en la figura 14.



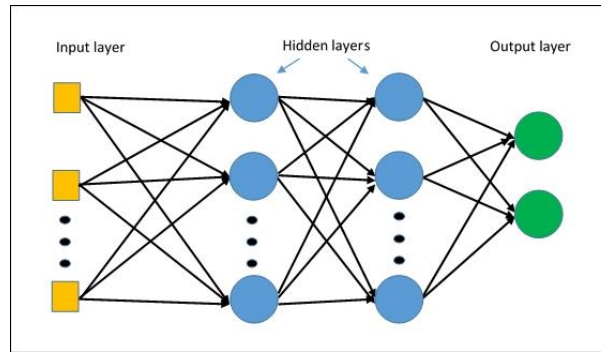


Figura 14. Arquitectura Perceptrón Multicapa

El proceso mediante el cual una red neuronal modifica sus pesos, es decir, destruye, modifica o crea conexiones en respuesta a una información de entrada es llamado mecanismo de aprendizaje [7]. El proceso de aprendizaje finaliza, cuando los valores de los pesos dejan de cambiar, entonces se dice que la red ha aprendido.

Un aspecto importante respecto al aprendizaje de las redes neuronales es el conocer bajo qué criterios se modifican los valores de los pesos. Dependiendo de la arquitectura de la red empleada, se puede considerar un proceso que fuerce a una red a entregar una respuesta dada ante una entrada específica [7]. Una regla de aprendizaje determina la forma de actualización de los pesos de la red.

## 2.8. Coeficiente de Correlación

El coeficiente de correlación mide el grado de relación o asociación existente generalmente entre las variables aleatorias [13]. Éste busca estimar el nivel de correspondencia entre los valores de ambas variables. El coeficiente de correlación de dos variables aleatorias es una medida de su dependencia lineal. Si cada variable tiene N observaciones escalares, el coeficiente de correlación entre las variables A y B se define como se muestra en la ecuación 6.

$$\rho(A, B) = \frac{1}{N - 1} \sum_{i=1}^N \left( \frac{A_i - \mu_A}{\sigma_A} \right) \left( \frac{B_i - \mu_B}{\sigma_B} \right)$$

Ecuación 6.

Donde  $\mu_A$  y  $\sigma_A$  son la media y la desviación estándar de A respectivamente y  $\mu_B$  y  $\sigma_B$  son la media y la desviación estándar de B.

## 2.9. Coeficiente de Determinación (R<sup>2</sup>)

El coeficiente de determinación o R<sup>2</sup> se define como la proporción de la varianza total de variable explicada por la regresión. El R<sup>2</sup> refleja la bondad del ajuste de un modelo a la variable que se pretende explicar [16].

El resultado del coeficiente de determinación oscila entre 0 y 1. Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que se desea explicar. De forma inversa, cuanto más cerca de cero, menos ajustado estará el modelo y, por tanto, menos fiable será. La ecuación 7 muestra cómo se define el valor de R<sup>2</sup>.

$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Ecuación 7.

Donde  $\hat{Y}$  corresponde a la estimación del modelo a evaluar y Y es el vector de valores reales sobre los que se hace la comparación.

## 2.10. Root Mean Squared Error (RMSE)

La Raíz del Error Cuadrático Medio o RMSE (*Root Mean Squared Error*) es una medida de desempeño entre 2 funciones. En este contexto, RMSE consiste en la raíz cuadrada de la sumatoria de los errores cuadráticos entre dos conjuntos de datos. En comparación con la Desviación Media Absoluta o MAD, RMSE amplifica y penaliza con mayor fuerza aquellos errores de mayor magnitud. La fórmula de cálculo del RMSE se muestra en la fórmula a continuación (8): [35]

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - F_i)^2}$$

Ecuación 8.

Donde  $A$  es el vector de  $N$  predicciones y  $F$  es el vector de los verdaderos valores.

## 2.11. Curva ROC

Una curva ROC es una representación gráfica de la tasa de éxito (probabilidad de detectar correctamente una señal cuando dicha señal está efectivamente presente) frente a la tasa de falsa alarma (probabilidad de detectar una señal cuando efectivamente NO está presente) para tareas de detección con sólo dos resultados posibles (sí / no, presente / ausente). La curva se crea según se varía el umbral o criterio para detectar la señal a lo largo de la escala de valores, a partir de los cuales se hace la detección [36]. Los ejes de esta representación gráfica pueden ser representados en términos de sensibilidad vs 1-especificidad, evaluando los mismos parámetros. Un ejemplo de una curva ROC se puede ver a continuación en la figura 15.

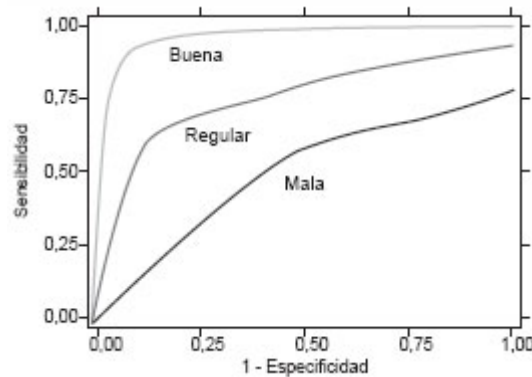


Figura 15. Ejemplo Curva ROC [17]

Se define la exactitud, acuracidad o *Accuracy* de una variable de decisión como la probabilidad de discriminar correctamente. Cuanto más se aproxime la cantidad de resultados acertados al total de individuos de la muestra, más cercano a 1 será la exactitud y se obtendrá, por tanto, una prueba con alta capacidad discriminante [17]. La exactitud o *accuracy* está dado por la forma (Ecuación 9):

$$AC = \frac{\text{resultados acertados}}{\text{total de la muestra}}$$

Ecuación 9.

## 3. OBJETIVOS

### 3.1. Objetivo General

El objetivo general propuesto para este trabajo de grado corresponde a estimar la concentración de nitrógeno en plantas de arroz a partir del análisis de imágenes multiespectrales mediante técnicas de *Machine-Learning*.

### 3.2. Objetivos Específicos

Los objetivos específicos que permiten el cumplimiento del objetivo general son:

1. Definir un conjunto de Índices Vegetativos (IV) necesarios para la medición indirecta de nitrógeno a partir de imágenes multiespectrales.
2. Desarrollar un algoritmo para la extracción de los IV a partir de imágenes multiespectrales tomadas en tres fases de crecimiento del cultivo: vegetativo, reproducción y cosecha.
3. Desarrollar un algoritmo para la estimación de nitrógeno por medio de la implementación de tres (3) técnicas de *Machine-Learning*: regresiones multivariable, SMV y Redes Neuronales.
4. Validar las tres (3) técnicas implementadas a partir de una correlación superior al 65% entre la estimación de nitrógeno y un *Ground-truth* establecido (medición directa de nitrógeno - SPAD).
5. Realizar un comparativo de desempeño entre las tres (3) técnicas desarrolladas a partir de un análisis estadístico ROC.

#### 3.2.1. Objetivo Específico 1

Se realizó una evaluación del estado del arte respecto la relación de los índices vegetativos y la concentración de nitrógeno en cultivos de cereales, específicamente arroz. Los índices seleccionados fueron aquellos que, según la literatura brindan una mayor información sobre la acumulación de nitrógeno en las hojas del cultivo.

#### 3.2.2. Objetivo Específico 2

Una vez conocidos los índices vegetativos se realizó en *Matlab* un código el cuál, calcula cada uno de estos índices para los distintos conjuntos de imágenes multiespectrales para una sección de cultivo específica.

Basados en el estado del arte de agricultura de precisión se tomaron las ecuaciones que describen cada uno de estos índices vegetativos y se generó una matriz que relaciona los índices vegetativos calculados para cada sección de cultivo fotografiada. A cada conjunto de índices vegetativos por sección se asocia la medida de nitrógeno correspondiente obtenida directamente en las mediciones del cultivo con el sistema de medición SPAD.

El producto final obtenido es un conjunto de tres matrices como la descrita previamente, cada una correspondiente a una etapa de cultivo a evaluar.

#### 3.2.3. Objetivo Específico 3

Ahora, con cada una de las matrices que relacionan los índices vegetativos con el nivel de nitrógeno presente en distintas secciones de cultivo, se procedió a realizar la estimación de los niveles de nitrógeno con tres técnicas diferentes de aprendizaje de máquina (Regresiones lineales multivariable, máquinas de soporte vectoriales y redes neuronales artificiales). Para esto se realizó el entrenamiento de cada una de las máquinas con el 70% de cada una de las matrices por etapa y se evaluó el desempeño de las distintas máquinas con el 30% restante.

Para cada una de las máquinas se realizó un procedimiento evaluativo entre distintos tipos y arquitecturas de las mismas para definir cuál de todas las configuraciones posibles brinda una mejor aproximación a los datos obtenidos en las mediciones.

#### **3.2.4. Objetivo Específico 4**

Para la evaluación de las distintas técnicas de aprendizaje de máquina se calculó el coeficiente de correlación entre los datos obtenidos en el 30% de evaluación y las mediciones de nitrógeno realizadas previamente en las plantaciones de arroz. Como manera de asegurar que la correlación obtenida es la deseada se realizaron además pruebas de desempeño por medio de la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación.

#### **3.2.5. Objetivo Específico 5**

Finalmente, con el fin de definir cuál de las tres técnicas de aprendizaje implementadas brinda un mayor acercamiento a los datos reales, se realizó un cuadro comparativo por medio del análisis estadístico ROC. En ella se evaluó la habilidad de cada una de las técnicas para clasificar los niveles de nitrógeno estimados según la etapa de cultivo a la que corresponden, ya sea etapa de reproducción o de cosecha.

## 4. DESARROLLO

### 4.1. Antecedentes del Macro Proyecto

El presente trabajo de grado se realiza con base en el proyecto del equipo conformado por: Julián Colorado (PI), Camila Rebolledo, Eliel Preto, Iván Mondragón, Carol Martínez, Diego Patiño, Carlos Devia, Juan P. Rodríguez, llamado “*Desarrollo de una Herramienta para la Agricultura de Precisión en los Cultivos de Arroz: Sensado del Estado de Crecimiento y de Nutrición de las Plantas usando un Drone autónomo*”. Dicho proyecto tiene como fin desarrollar e implementar un sistema de alta precisión para el seguimiento del estado de desarrollo y nutrición nitrogenada e hídrica de los cultivos de arroz mediante el uso de un drone autónomo. Dada la similitud de los objetivos, parte de las labores realizadas por el mencionado proyecto son de gran utilidad para el presente trabajo de grado.

La captura de imágenes fue realizada a través de 2 cuadricópteros con referencias, *Astec Pelican* y *Astec Hummingbird*, ambos fabricados por la empresa Alemana: *Ascending Technologies* de Intel. Los cuadricópteros tienen una capacidad de carga de 1.2Kg y 500g respectivamente e incorporan un sistema de comunicación *datalink Xbee* a 2.4 GHz para la transmisión de datos sensoriales e imágenes de video a la estación base. También incorpora comunicación vía RF para posible tele-operación en vuelo. Los cuadricópteros cuentan con dos sistemas de procesamiento a bordo: i) Un procesador Intel Core i7 de bajo nivel que contiene el autopiloto, lo cual garantiza una alta robustez y precisión del robot en vuelo, y ii) Un procesador *Atomboard* de alto nivel a 1.6 GHz, 1GB de RAM y sistema operativo Linux-Ubuntu preinstalado en una memoria SD. El *Atomboard* contiene (7) puertos USB 2.0, UART, puerto LVDS para video, puerto mini *PCI-express* y entrada de tarjeta SD. La plataforma también cuenta con sensores inerciales (IMU: *Inertial Measurement Unit*) y posicionamiento (GPS: *Global Positioning System*) [18].



Figura 16. Drones instrumentados e integrados para el Proyecto: Astec Pelican con cámara multispectral ADC Lite de Tetracam y Astec Hummingbird con cámara multispectral Sequoia de Parrot.

La figura 16 detalla los dos cuadricópteros antes mencionados. Las fotografías corresponden a pruebas de campo realizadas durante la ejecución del proyecto en los cultivos de arroz del CIAT en Santa Rosa, Departamento del Meta, Colombia. Adicionalmente, se detallan las dos cámaras multispectrales integradas en los drones. Ambas cámaras fueron seleccionadas en función de las especificaciones técnicas otorgadas por los fabricantes. El objetivo de usar dos cámaras radica en probar los algoritmos y métodos desarrollados para la extracción de biomasa, nitrógeno y estado hídrico a partir de imágenes multispectrales capturadas por diferentes ópticas [18].

El *ADC-lite Tetracam* contiene un sensor de 3.2 mega-píxeles optimizado para la captura de luz infrarroja cercana, para longitudes de onda hasta 1150nm (NIR: *Near Infrared*). Por otro lado, la cámara multispectral *sequoia de Parrot* está conformado por 4 arreglos (lentes) que permiten capturar imágenes a diferentes longitudes de onda: R-G-B y NIR (rojo, verde, azul y cercano infrarrojo-NIR) [18].

La figura 17 detalla la interfaz de usuario para la definición y control de la misión de vuelo del drone. Esta interfaz se ejecuta en cualquier computador portátil con sistema operativo Windows. La interfaz permite al usuario: [18]

- Cargar mapa satelital del cultivo.

- Determinar la trayectoria de vuelo del dron para cobertura de cultivo de gran extensión. Se puede definir la altura de vuelo, velocidad del vuelo, resolución de vuelo a través de los puntos de control (coordenada GPS de latitud y longitud), tiempo de vuelo estacionario en cada punto de control y orientación del dron con respecto al punto de control.
- Supervisar la misión: batería del dron, variables de vuelo, comando de “go home” para abortar misión y regrese al punto de despegue entre otros.

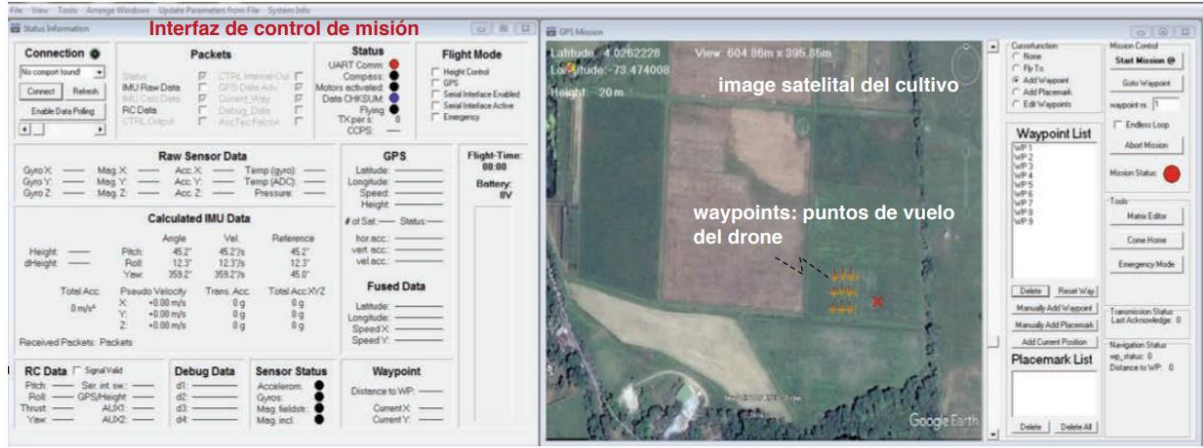


Figura 17. Interfaz de Control de Misión

Adicionalmente, se desarrolló una segunda interfaz de usuario para la configuración y control de la cámara multispectral abordo del dron. La Figura 18 detalla las funcionalidades del software/interfaz desarrollado [18].

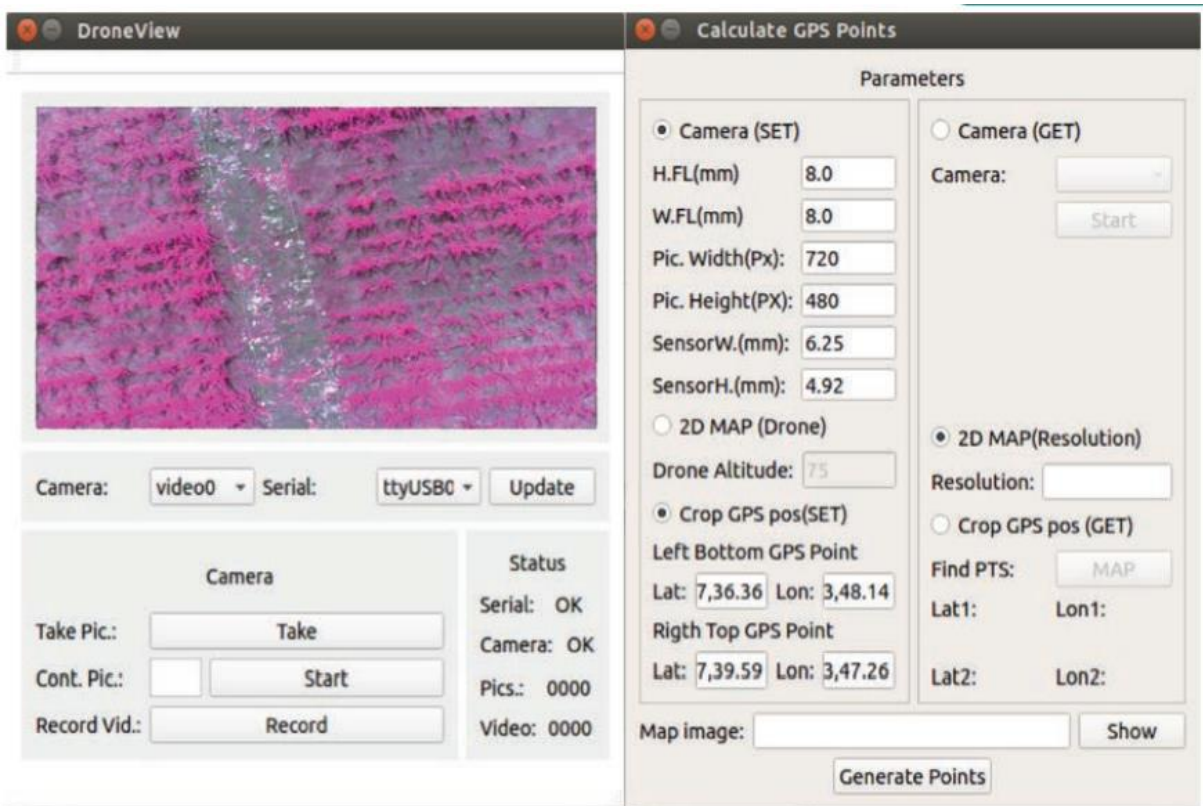


Figura 18. Interfaz de usuario para la configuración y control de la cámara multispectral abordo del dron.

Una vez son tomadas las fotografías, éstas son filtradas puesto que algunas presentan defectos en el momento de la captura. Considerando que el dron se suspende a una altura considerable, una ráfaga de viento puede generar defectos en la fotografía en el momento de la toma.

Es preciso señalar que las imágenes tienen asociadas una medida de nitrógeno aproximada tomada a través del dispositivo SPAD, que de igual forma tiene ligada una ubicación GPS a ellas, sin embargo, la toma de datos de nitrógeno no corresponde exactamente a una fotografía en especial. Se realizaron varias mediciones de nitrógeno a lo largo del campo y después se buscó la posición más cercana de las fotografías para vincular ambos parámetros, esto con el propósito de realizar la conexión entre las fotografías y medidas reales de la variable a estimar.

Considerando que las imágenes ya han sido capturadas y esta información está unida a las medidas de nitrógeno el proceso a seguir el cual comprende el presente trabajo de grado se muestra a continuación en la figura 19.

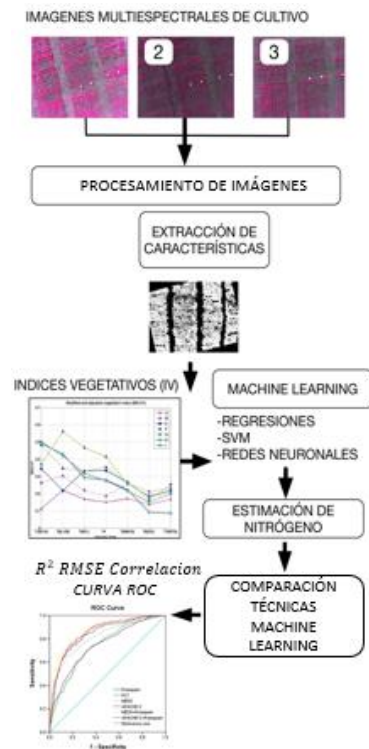


Figura 19. Diagrama en Bloques

Para el desarrollo del Trabajo de Grado se utilizaron un total de 801 fotografías procesadas; 363 correspondientes a la etapa vegetativa, 239 a la reproductiva y 199 a la de cosecha.

## 4.2.Desarrollo del Trabajo de Grado

### 4.2.1. Identificación de Índices

Basados en el estado del arte revisado y analizado, el cual recopila un amplio estudio en el campo de agricultura de precisión enfocada en cultivos de arroz fueron escogidos unos índices vegetativos, que dado su alto desempeño al momento de brindar información en el estado del cultivo fueron seleccionados para el desarrollo de este trabajo de grado. Los índices vegetativos implementados son mostrados en la tabla 2.

Nombre	Fórmula
--------	---------

Normalized Difference Vegetation Index	$NDVI = \frac{\rho_{780} - \rho_{670}}{\rho_{780} + \rho_{670}}$
Simple Ratio	$RVI = \frac{\rho_{780}}{\rho_{670}}$
Green Normalized Difference Vegetation Index	$GNDVI = \frac{\rho_{780} - \rho_{500}}{\rho_{780} + \rho_{500}}$
Corrected Transformed Vegetation Index	$CTVI = \frac{NDVI + 0.5}{ NDVI + 0.5 } \sqrt{ NDVI + 0.5 }$
Soil-Adjusted Vegetation Index	$SAVI = 1.5 \cdot \frac{(\rho_{800} - \rho_{670})}{(\rho_{800} - \rho_{670} + 0.5)}$
Modified-SAVI	$MSAVI = \frac{1}{2} (2\rho_{800} + 1 - \sqrt{((2\rho_{800} + 1)^2 - 8(\rho_{800} - \rho_{670}))})$
Triangular Vegetation Index	$TVI = \frac{1}{2} (120(\rho_{780} - \rho_{500}) - 200(\rho_{670} - \rho_{500}))$

Tabla 2. Índices Vegetativos Seleccionados

$\rho_{780-800}$  = Reflectancia correspondiente al infrarrojo cercano

$\rho_{670}$  = Reflectancia correspondiente al rojo

$\rho_{500}$  = Reflectancia correspondiente al verde

#### 4.2.1.1. NDVI

Este es el índice más comúnmente usado porque retiene la habilidad para minimizar los efectos topográficos y a su vez producir una escala de medida lineal. Además, los errores de división por cero son reducidos en gran medida. Más aún, la escala de medida posee la apreciada propiedad de oscilar de -1 a 1 con el 0 representando el valor aproximado de la falta de vegetación. Entonces, los valores negativos representan las superficies sin vegetación [22].

#### 4.2.1.2. RVI

Utilizado para altos niveles de biomasa. RVI es un índice muy sensible a la vegetación y tiene una alta correlación con los niveles de biomasa de las plantas. Cuando el nivel de vegetación es escaso, menor al 50%, el índice es sensible a los efectos atmosféricos y su representación de los niveles de biomasa es débil. Toma valores altos para vegetación y bajos para suelo y agua. Reduce los efectos atmosféricos y topográficos [23].

#### 4.2.1.3. GNDVI

El índice GNDVI (Índice de Vegetación de Diferencia Normalizada Verde) es un indicador del “verdor” de la planta o actividad fotosintética. Es uno de los índices de vegetación más utilizados para determinar la captación de agua y nitrógeno en el dosel del cultivo.

Los valores arrojados por dicho índice también oscilan entre -1 y 1. Valores entre -1 y 0 están asociados a la presencia de agua o suelo desnudo. Este índice se emplea fundamentalmente en la etapa intermedia y final del ciclo de cultivo [24].

#### 4.2.1.4. CTVI

Apunta a corregir el TVI, Agrega una constante de 0,5 a todos los valores del NDVI con el fin de eliminar los valores negativos. Sin embargo, no lo logra del todo ya los valores del NDVI pueden tener el rango -1 a +1. Los valores menores que -0,5 dejan pequeños valores negativos luego de la operación de adición [22].



#### 4.2.1.5. SAVI

Propuesta por Huete (1988). Está preparado para minimizar los efectos del fondo de suelo sobre la señal de vegetación al incorporar un factor 0.5 de ajuste del suelo constante en el denominador de la ecuación NDVI [22].

#### 4.2.1.6. M-SAVI

El Índice de vegetación ajustada de suelo modificado (MSAVI2) intenta minimizar el efecto del terreno desnudo en el SAVI [33].

#### 4.2.1.7. TVI

El índice de vegetación transformado incorpora un valor de 0.5 con el fin de evitar resultados negativos, la raíz cuadrada intenta corregir valores que se aproximan a una distribución de *Poisson* e introduce una distribución normal. Como deficiencia es posible decir que no elimina del todo los valores negativos, pues necesita como dato de entrada valores positivos, de no ser así la operación continúa con valores negativos [23].

### 4.2.2. Extracción de Índices

Se utilizó la herramienta *Matlab* para la extracción de índices. El procedimiento fue realizado de la siguiente manera:

Por cada *set* de fotografías se calculó el valor promedio entre los valores de píxeles en cada una de las imágenes de banda específica. Teniendo un valor correspondiente a cada banda, éstos se reemplazaron en las distintas ecuaciones de los índices seleccionados, encontrando de esta forma, el valor correspondiente a cada uno de los índices en la sección del cultivo analizada. En la figura 20 se muestra el pseudo-algoritmo de la función implementada para la extracción de los índices vegetativos.

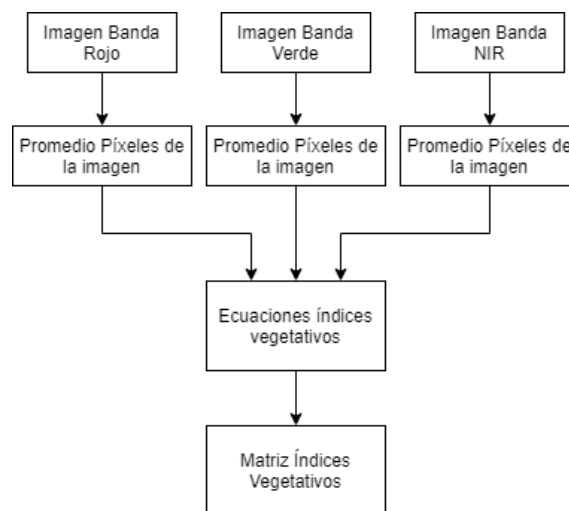


Figura 20. Pseudo-algoritmo para la obtención de los índices a partir de las imágenes.

A medida que se van hallando los valores de cada uno de los índices correspondientes al *set* de fotos analizado, éstos se fueron almacenando en una matriz de ocho columnas donde las primeras siete columnas corresponden a los valores de cada uno de los índices vegetativos correspondientes al grupo de imágenes y la octava columna corresponde al valor de concentración de nitrógeno correspondiente a la sección del cultivo.

Para tener una mayor claridad en el desarrollo del trabajo de grado, se decidió tener como producto final tres distintas matrices, una por cada una de las etapas de cultivo tenidas en cuenta para el desarrollo del trabajo de grado (siembra, reproducción y cosecha).

La formación de las matrices sigue el procedimiento expuesto anteriormente, sin embargo, dada la afinidad con el proyecto previamente nombrado, el desarrollo siguiente fue ejecutado con las matrices generadas por éste. A partir de este punto los proyectos toman objetivos diferentes y los procesos son ajenos uno del otro.

### 4.2.3. Estimación de Niveles de Nitrógeno

En esa sección, una vez obtenidas las distintas matrices de índices vegetativos de cada etapa de cultivo junto con sus correspondientes valores de concentración de nitrógeno se procede a implementar las técnicas de aprendizaje de máquina propuestas para la estimación de los niveles de nitrógeno (Regresiones lineales, redes neuronales artificiales y SVM). Debido a que el objetivo de las máquinas en el presente trabajo de grado consiste en realizar una aproximación a modelos reales, los algoritmos de aprendizaje son de tipo supervisado y por esto, las matrices de bases de datos están divididas en un 70% asignadas a entrenamiento de la máquina y el 30% restante para la evaluación de las mismas.

#### 4.2.3.1. Regresiones Lineales Multivariable

Para llevar a cabo una regresión lineal multivariable, como se explicó en el marco teórico, es necesario contar con un grupo de variables independientes, cuya combinación lineal da resultado a un valor que obedece a los primeros. Se conoce que existen siete variables independientes, cada una asignada a un índice vegetativo calculado por set de fotografías, y la variable dependiente hace referencia a los valores de SPAD obtenidos en cada una de las imágenes.

La combinación lineal propuesta es mostrada a continuación en la ecuación 10. Los términos de la ecuación están constituidos por un término independiente  $A_0$  y por la suma de las variables independientes del problema (los índices vegetativos), cada uno acompañado de su respectivo coeficiente de modelamiento ( $A_1$  a  $A_7$ ), los cuales buscan aproximar de manera más acertada la combinación de las variables independientes al resultado objetivo.

$$\text{Nitrógeno} = A_0 + A_1IV_1 + A_2IV_2 + A_3IV_3 + A_4IV_4 + A_5IV_5 + A_6IV_6 + A_7IV_7$$

*Ecuación 10.*

La implementación de regresiones lineales multivariable desarrollada en *Matlab* busca hallar tanto el término independiente,  $A_0$ , como los demás coeficientes de modelamiento para entregar una aproximación más acertada del valor de concentración de nitrógeno. Siguiendo con lo mencionado anteriormente, el 70% de la base de datos de cada etapa de cultivo es utilizada para calcular dichos coeficientes y el 30% restante es utilizado para la evaluación de la técnica.

El código Matlab implementado es encontrado en los anexos. Los resultados obtenidos se muestran en la sección de pruebas.

#### 4.2.3.2. Máquinas de Soporte Vectorial

Una vez los datos han sido separados aleatoriamente para realizar el entrenamiento con un 70% de los mismos en cada una de las matrices y un 30% para realizar las pruebas. El *toolbox* de *Regression Learner* de *Matlab* es la herramienta que será utilizada para esta parte del proceso, dado que ésta cuenta con los algoritmos de SVM en la aplicación de regresiones con diversos Kernel.

El entrenamiento se realiza para cada set de datos por etapa de cultivo. La información es importada desde el *toolbox*, estos datos constan de una matriz de 8 columnas por N filas, dependiendo del set de datos

que se esté entrenando. Las primeras 7 columnas contienen las características o regresores (IV) de la regresión mientras que la octava columna comprende el resultado o target (Concentración de nitrógeno).

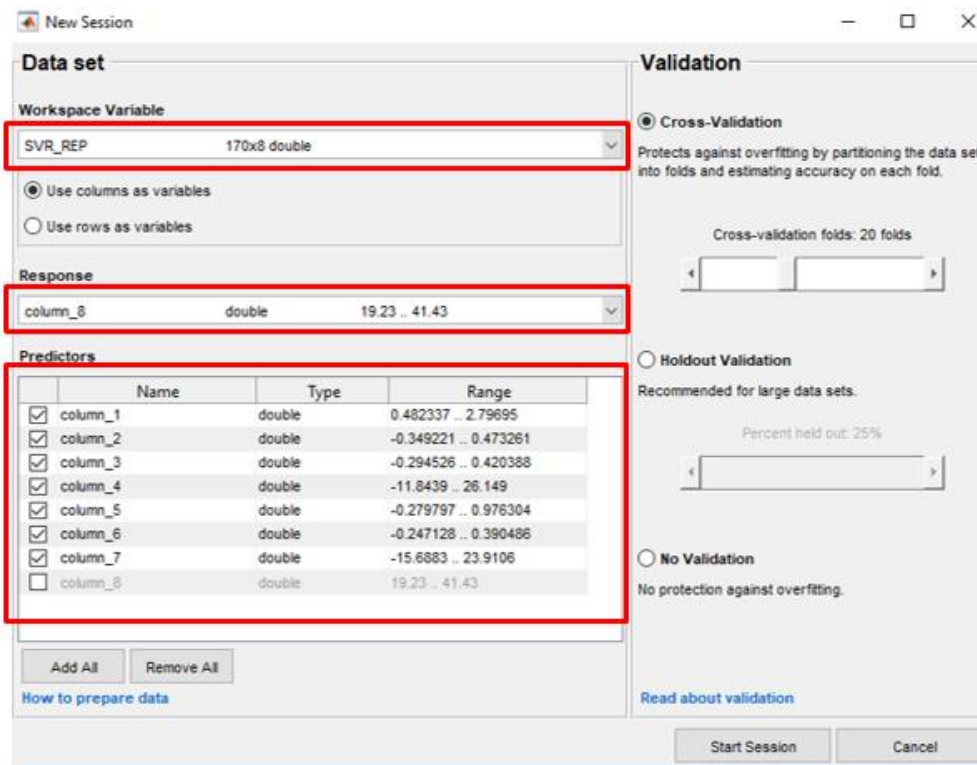


Figura 21. Interfaz toolbox Regression Learner

En la figura 21 se muestra la interfaz del *toolbox Regression Learner*; se tomará como ejemplo el set de datos correspondientes a la etapa de Reproducción, por lo que en el cuadro *Workspace Variable* aparece esta matriz de 170 filas por 8 columnas, donde la última columna corresponde a la a estimar, como se especifica en el cuadro *Response*. En la tabla de *Predictors* han sido seleccionado todos los IV con el objetivo de obtener más información acerca de la medida de nitrógeno respectiva por cada fila.

Una vez los datos son importados al *toolbox*, se selecciona la técnica de regresión que se utilizará, aterrizado a este trabajo de grado la opción *All SVMs* es la escogida, como se muestra en la figura 22. Ésta presenta la opción de entrenar diferentes máquinas de soporte vectorial simultáneamente, variando el parámetro Kernel.

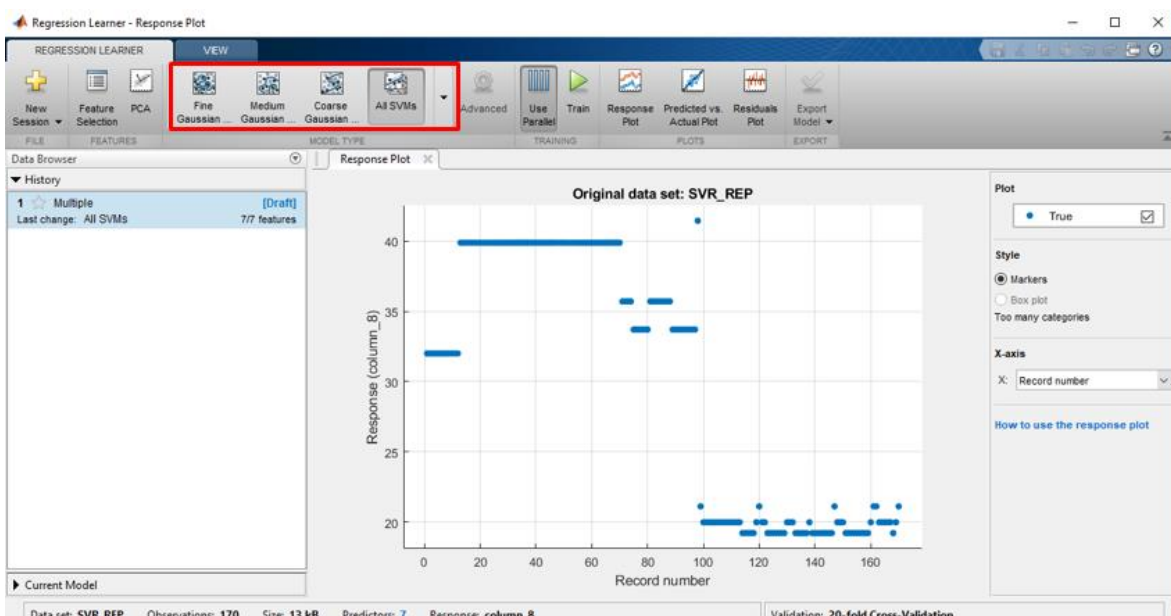


Figura 22. Interfaz Regression Learner – Response Plot

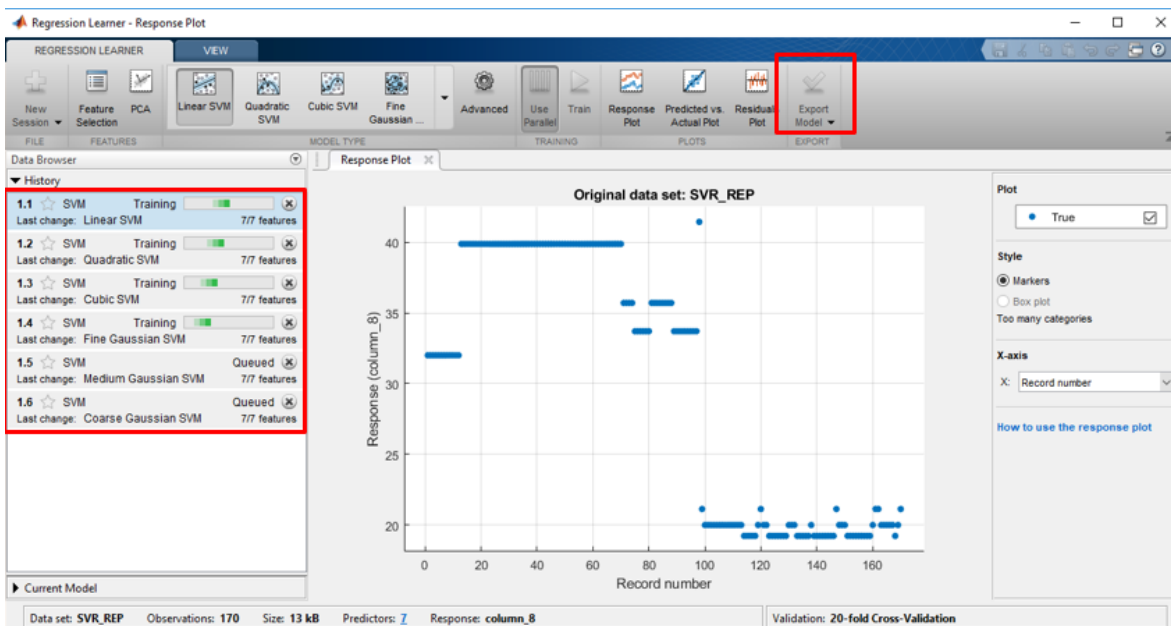


Figura 23. Interfaz Regression Learner – Response Plot

Las máquinas son entrenadas variando el parámetro Kernel, los demás parámetros permanecen temporalmente por defecto del software, la figura 23 muestra los diferentes tipos de máquinas siendo entrenadas.

Teniendo las máquinas de soporte vectorial entrenadas, cada una con un Kernel diferente, éstas son exportadas con la herramienta *Export Model*. Ahora es posible realizar las predicciones correspondientes; para esto se usa la función *Predict*, que con la máquina de soporte vectorial entrenada y los predictores (IV) de test, estima los niveles de nitrógeno a dicho set de características. Estos datos predichos serán posteriormente contrastados contra los datos de *groundtruth* correspondientes a estos predictores.

#### 4.2.3.3. Redes Neuronales Artificiales

Como se definió previamente en el marco teórico, una red neuronal artificial tiene los siguientes elementos:

Un conjunto de entradas a la neurona, los cuales tienen unos pesos asociados. Un número de capas y a su vez, un número de neuronas que conforman cada una de las capas, ya sean de entrada, ocultas o de salida. Una función de activación o de transferencia, la cual representa la salida de la neurona en función a su conjunto de entradas. Y, por último, una función de aprendizaje, que determina de qué manera se van ajustando los pesos asociados a cada rama para obtener la salida deseada.

En el proceso de aprendizaje de una red, la salida computada por la red se compara con la salida deseada, y los pesos de la red se modifican en el sentido de reducir el error cometido. Se repite iterativamente este proceso hasta que la diferencia entre salida computada y deseada sea aceptablemente pequeña [12]. Cada iteración realizada en el proceso de aprendizaje es toma el nombre de época.

Para conocer qué máquina se ajusta más al modelo deseado se debe hacer una elección adecuada de características de la red neuronal, tanto de su arquitectura como de cada una de las partes que lo componen. No existe un procedimiento exacto y contundente para concretar cuál es la red neuronal que presenta un mejor desempeño respecto al objetivo planteado. Por esta razón, en el trabajo de grado se muestran distintas configuraciones de redes neuronales artificiales variando algunos de los elementos que la componen en espera de obtener una mejor aproximación a los datos deseados.

Inicialmente se desea conocer el número de capas requeridas para la implementación del trabajo de grado. Conociendo el número de capas va cercanamente ligado al número de datos que se desean procesar en la máquina se hace una comparación entre una red neuronal con solo una capa oculta y otra red que tiene un número de capas significativamente mayor a la primera red planteada.

Por otro lado, el desempeño de la red neuronal también depende de la función de aprendizaje, el cual es uno de los parámetros más relevantes en cuanto a la práctica de redes neuronales. Por esta razón, se diseñan varias redes con distintas funciones de aprendizaje y transferencia para, de esta manera observar cuál de todas se ajusta de mejor manera al conjunto de datos a predecir.

Las redes neuronales utilizadas para solución son:

- Función de aprendizaje ‘Regresión Bayesiana’(BR) con función de activación sigmoidea tangencial en la capa oculta.
- Función de aprendizaje ‘Levenberg-Marquardt’(LM) con función de activación sigmoidea logarítmica en la capa oculta.
- Función de aprendizaje ‘Gradiente Escalado Conjugado’(SCG) con función de activación sigmoidea logarítmica en la capa oculta.
- Función de aprendizaje ‘Quasi-Newton BFGS’(BFG) con función de activación sigmoidea tangencial en la capa oculta.

En este sentido, el número de neuronas que poseen las capas ocultas de la red también interviene en el funcionamiento de la misma y, por esta razón, en el trabajo de grado también se hace un cuadro comparativo variando el número de neuronas de la red oculta en cada red diseñada.

Para el desarrollo de las redes neuronales artificiales fue utilizada la interfaz de la herramienta *Matlab*, llamada *nntool*, la cual permite crear, diseñar e implementar redes neuronales, permitiendo modificar cada uno de los parámetros de la red con el fin de alcanzar el objetivo deseado.

En la figura 24 se muestra la interfaz donde la red es creada y diseñada. En ella destacan entradas como ‘*Input data*’ y ‘*Target data*’, donde se especifica el número de neuronas en las capas de entrada y salida respectivamente, la función de entrenamiento ‘*Training function*’, el número de neuronas por capa y la función de transferencia en cada una de estas capas.

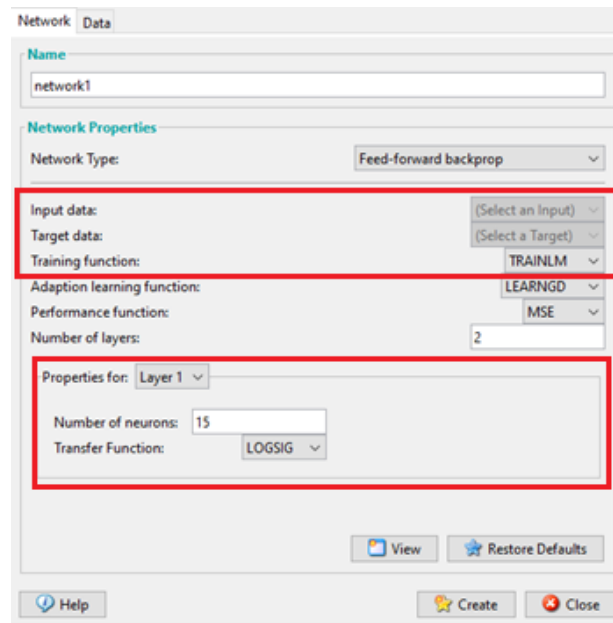
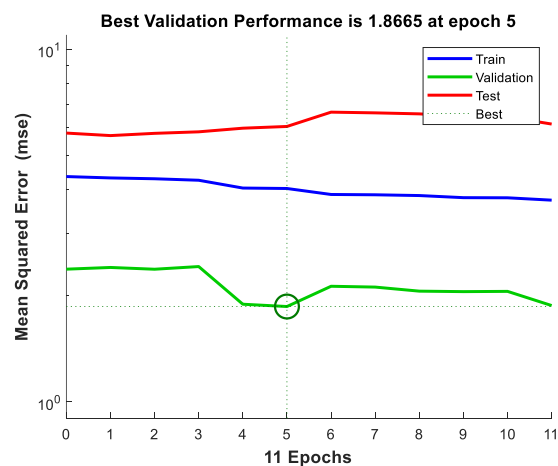
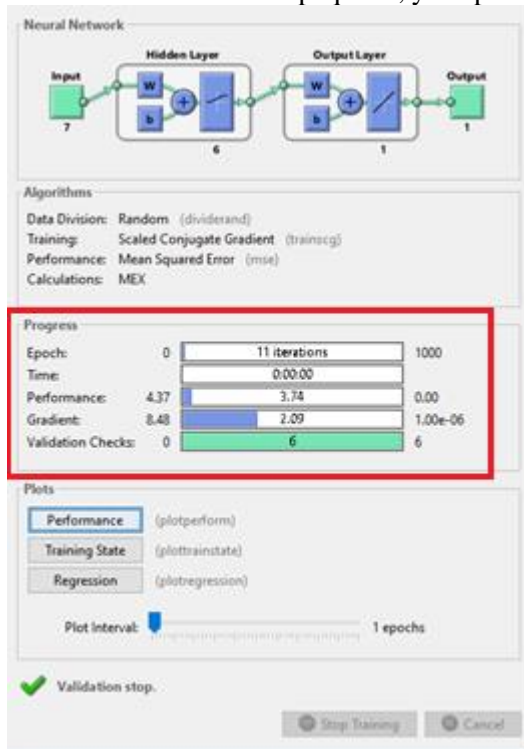


Figura 24. Interfaz de diseño de red

Para el proceso de entrenamiento, la maquina divide el set de datos de entrenamiento en tres subgrupos de datos, entrenamiento, validación y prueba. El set de entrenamiento es el conjunto de datos a partir de los cuales los pesos son ajustados, el set de validación busca certificar que efectivamente el desempeño de la red está disminuyendo en el entrenamiento y el set de prueba evalúa los pesos calculados en cada época. La figura 25 muestra el momento de entrenamiento de la red neuronal.

En la interfaz se observa la arquitectura de la red neuronal y las especificaciones asignadas a la misma. En la sección de 'Progress' se ve el proceso de las distintas variables de entrenamiento, las cuales son constantemente actualizadas en esta ventana. Entre las de mayor interés están el desempeño de la red (performance), medido en este caso en función del error cuadrático medio, el tiempo tomado en el entrenamiento, la magnitud del gradiente de desempeño, el cual tiende a disminuir en lo que el desempeño de la red se va haciendo más pequeño, y los puntos de validación utilizados para finalizar en entrenamiento.



Todas las variables del progreso presentan un límite el cual, si la red alcanza, el entrenamiento automáticamente se detiene.

La interfaz de entrenamiento permite también, observar una gráfica del desempeño de entrenamiento, observado en la figura 26. En ella se observan tres distintas líneas, correspondiente al comportamiento de cada uno de los subgrupos de datos de entrenamiento, validación y prueba durante el proceso de entrenamiento de la red.

En la sección de pruebas se muestra la evaluación de las distintas redes neuronales desarrolladas durante el trabajo de grado en cada una de las etapas de cultivo analizadas en éste.

#### 4.2.4. Correlación con Mediciones Realizadas

A manera de conocer qué tanto se asemejan los valores obtenidos por las diferentes técnicas de aprendizaje de máquina a los valores medidos en los diferentes cultivos es utilizado el coeficiente de correlación. El cual, como fue definido previamente en el marco teórico, es un índice el cual es utilizado para medir el grado de relación de dos variables cuantitativas.

En el desarrollo del trabajo de grado, para la obtención del coeficiente de correlación entre las mediciones realizadas en los cultivos y los valores estimados por cada una de las técnicas implementadas se utiliza la función de Matlab *corrcoef*, la cual tiene como entrada los vectores de datos a comparar. Uno es el vector de valores obtenidos en las mediciones de nitrógeno y el otro es el vector de valores estimados por cada una de las técnicas a relacionar.

Esta función calcula los valores de correlación entre cada uno de los vectores y la salida obtenida es una matriz de la forma mostrada en la fórmula en la que A corresponde al vector con las mediciones realizadas y B el vector resultante de la implementación de la técnica, tal como se muestra en la ecuación 11.

$$R = \begin{pmatrix} \rho(A, A) & \rho(A, B) \\ \rho(B, A) & \rho(B, B) \end{pmatrix}$$

Ecuación 11.

Por otro lado, una manera de conocer la semejanza de un conjunto de datos respecto a otros es a través de la raíz de la desviación cuadrática media, RMSE, el cual mide la cantidad de error que existe entre dos conjuntos de datos. En la implementación del código de Matlab encontrado en los anexos del documento se encuentra el cálculo de este valor para cada una de las técnicas implementadas.

#### 4.2.5. Comparación de Desempeño entre Técnicas

Se desea conocer cuál de las diferentes técnicas de aprendizaje de máquina implementadas brinda una mejor aproximación a las medidas obtenidas en los cultivos. Para esto es utilizada la técnica de análisis estadístico ROC.

Una vez conocidos los valores de clorofila correspondientes a cada una de las etapas del cultivo calculados en las distintas técnicas de aprendizaje de máquina, se valorará, por medio de un análisis estadístico ROC, la efectividad de las distintas técnicas implementadas evaluando su nivel de clasificación de valores de clorofila en las etapas de cultivo de reproducción y cosecha. Así, de esta manera se puede determinar qué técnica brinda valores más acertados a una etapa de cultivo específica.

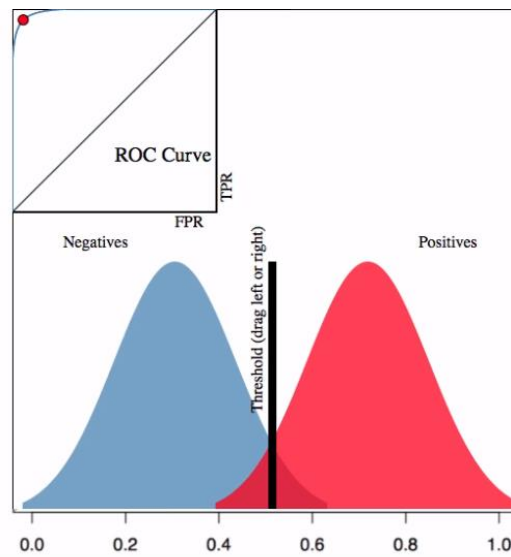


Figura 27. Curva ROC a partir de clasificación

En el desarrollo del trabajo de grado fue utilizada la función de *Matlab*, *perfcurve*, la cual calcula los valores de la curva de desempeño ROC para la salida del clasificador. En la sección de pruebas se evidencia el desempeño de las técnicas de aprendizaje de máquina.



## 5. PROTOCOLO DE PRUEBAS

### 5.1. Regresiones Lineales Multivariable

Para el método de regresiones lineales, se reiteró el procedimiento de entrenar y evaluar cada set de datos correspondiente a una etapa de cultivo por separado. Dado que se obtiene solo una estimación basada en los coeficientes de modelamiento que acompañan a las variables independientes, se realizará una comparación entre dicha estimación y la misma con la variante de que ésta ha sido modificada por un filtro promedio con el fin de eliminar el ruido de alta frecuencia y suavizar la función. Los resultados se muestran en las figuras a continuación.

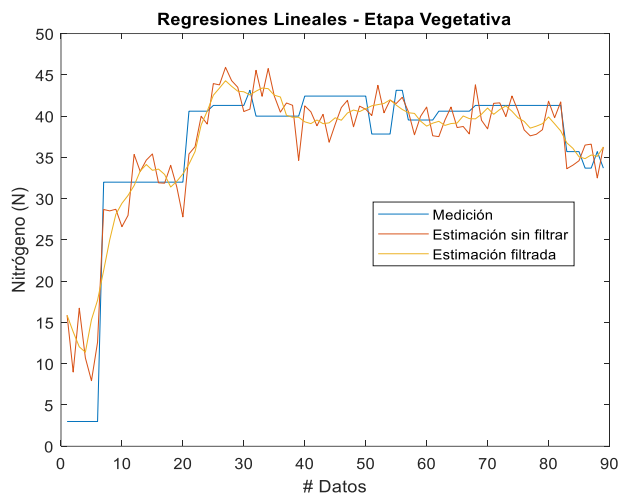


Figura 28. Evaluación Regresiones Lineales Etapa Vegetativa

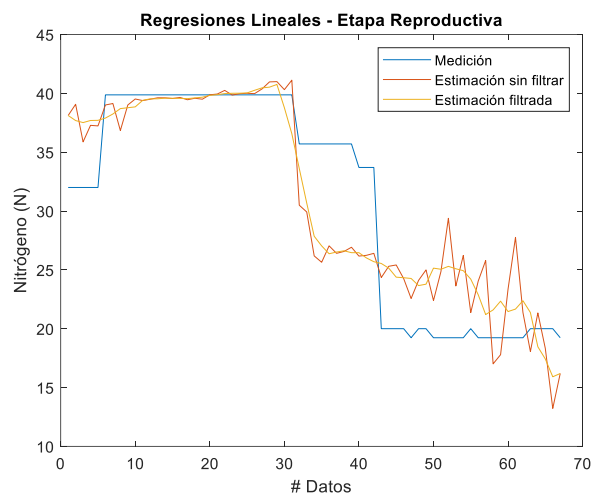


Figura 29. Evaluación Regresiones Lineales Etapa Reproductiva

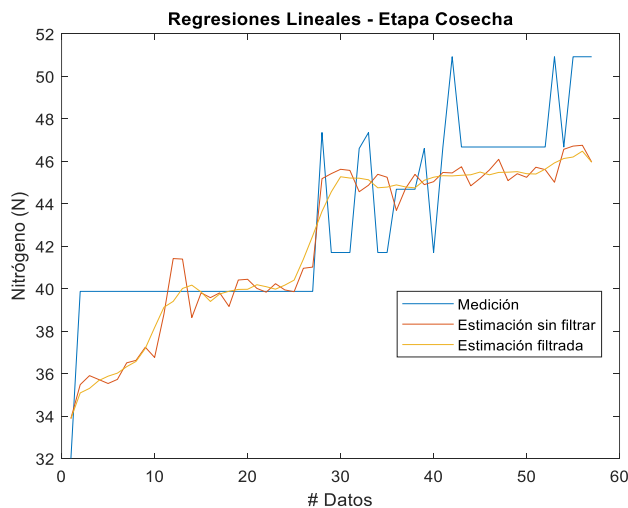


Figura 30. Evaluación Regresiones Lineales Etapa Cosecha

Las figuras 28, 29 y 30 muestran 3 funciones superpuestas: *groundtruth*, la estimación sin filtrar y la estimación filtrada, cada una correspondiente a las etapas vegetativa, reproductiva y cosecha respectivamente. Se muestra que la función de color dorado correspondiente a la estimación filtrada no presenta cambios tan bruscos como lo presenta la función de color rojo, los picos han sido disminuidos y las esquinas no son tan pronunciadas. La evaluación de RMSE y coeficiente de correlación, las cuales evalúan la estimación de manera cuantitativa, se presentarán en la sección de análisis de resultados.

## 5.2. Máquinas de Soporte Vectorial

En esta etapa del documento se mostrarán las pruebas realizadas a las diferentes SVMs planteadas en el capítulo anterior.

Dado que la distribución de datos para cada etapa (vegetativa, reproducción, cosecha) presenta una tendencia diferente, se realizarán pruebas separadas para cada fase de cultivo, tal como en el entrenamiento de las SVMs. Las máquinas fueron entrenadas con los conjuntos de datos previamente seleccionados, este proceso se reiteró para cada uno de los sets de datos correspondientes a cada etapa. Para las 3 fases se realizó el entrenamiento y predicción utilizando los 6 Kernel mencionados en capítulos anteriores.

Se realizará así mismo una prueba en el cuál sea involucrado el parámetro  $\epsilon$ , esta prueba se llevará a cabo una vez se tengan los resultado de correlación y se concluya ¿Cuál fue la máquina que por etapa, presentó una mayor correlación? ¿Cuál Kernel fue el utilizado en dicha máquina? Esto con el fin de analizar el comportamiento de la correlación frente a alteraciones en el valor del parámetro  $\epsilon$

Se presentan en las figuras, los resultados de las predicciones generadas por parte de cada una de las 6 diferentes máquinas entrenadas con un Kernel en particular para todas las etapas. Los resultados por etapa se presentan en 3 gráficas separadas para una mejor visualización de la tendencia de los datos. En cada gráfica está presente el valor de *groundtruth* como referencia, esto con el fin de contrastar la distribución de datos predicha por las máquinas de soporte vectorial contra los datos reales.

Las pruebas se harán con los siguientes Kernel:

- Kernel lineal
- Kernel polinomial cuadrático
- Kernel polinomial cubico
- Kernel gaussiano fino
- Kernel gaussiano medio
- Kernel gaussiano grueso

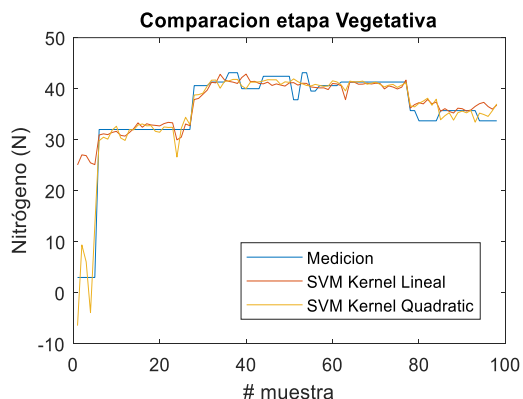


Figura 31. Comparación Etapa Vegetativa (1) – SVR

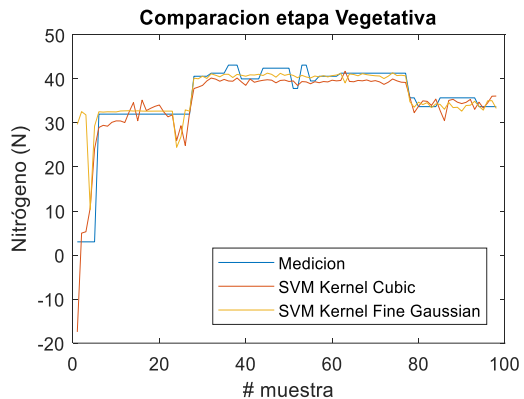


Figura 32. Comparación Etapa Vegetativa (2)- SVR

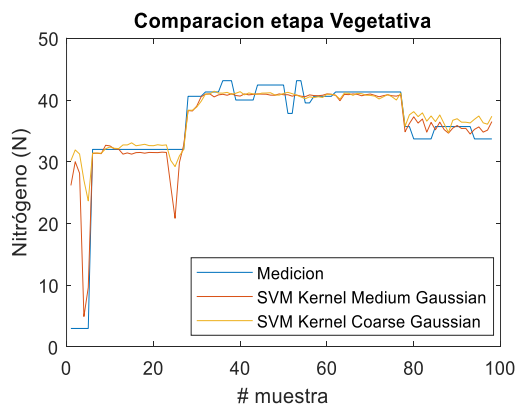


Figura 33. Comparación Etapa Vegetativa (3) - SVR

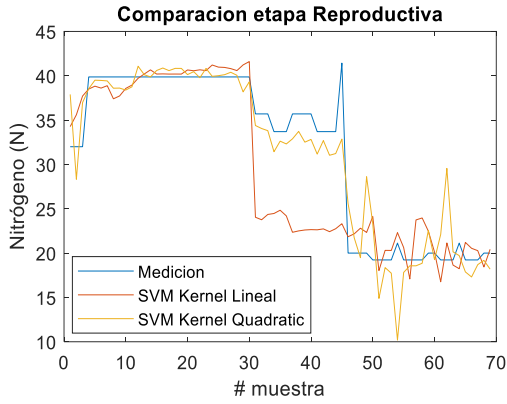


Figura 34. Comparación Etapa Reproductiva (1) – SVR

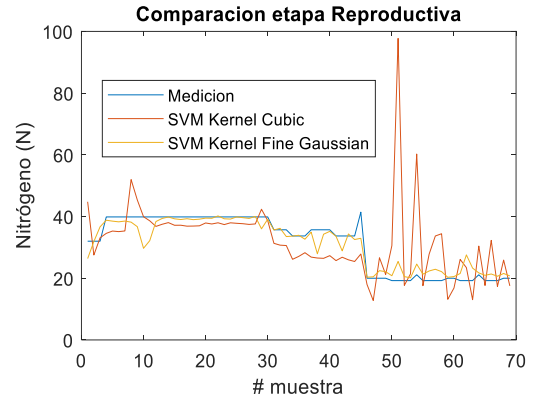


Figura 35. Comparación Etapa Reproductiva (2) - SVR

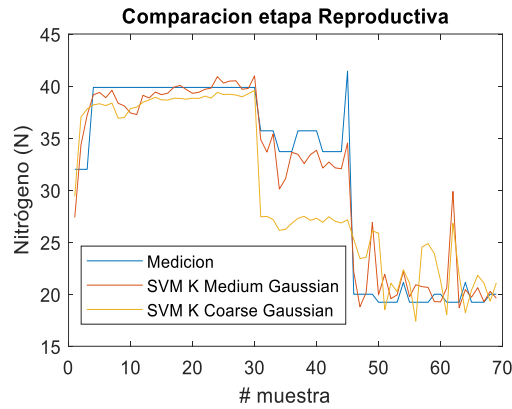


Figura 36. Comparación Etapa Reproductiva (3) - SVR

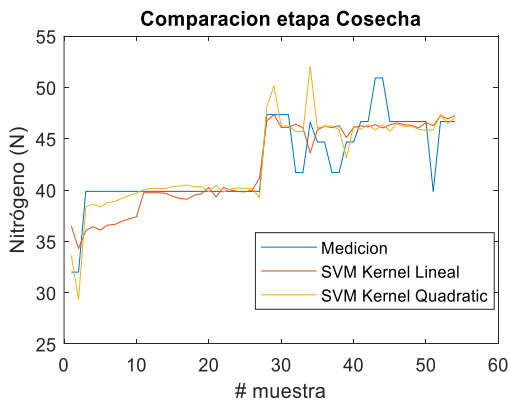


Figura 37. Comparación Etapa Cosecha (1) – SVR

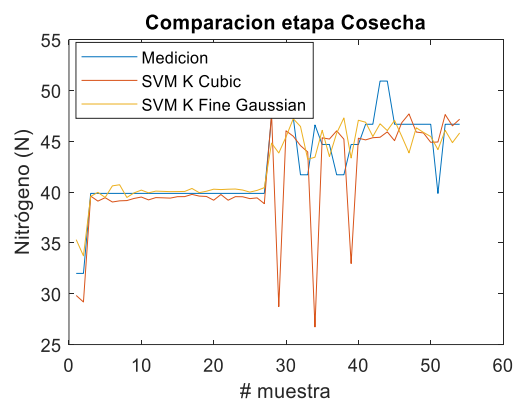


Figura 38. Comparación Etapa Cosecha (2) - SVR

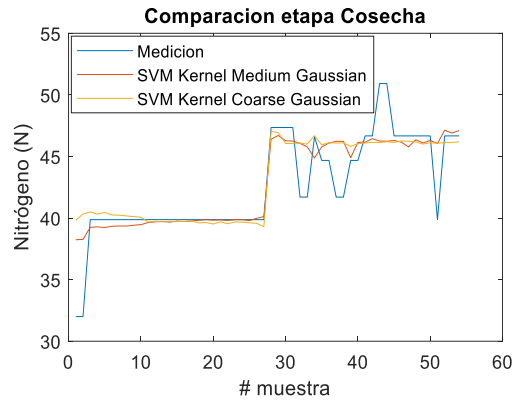


Figura 39. Comparación Etapa Cosecha (3) - SVR

Una vez se obtienen las máquinas que manifiestan una mayor correlación y sus correspondientes Kernel con base a las tablas, se procede a modificar el parámetro, con el fin de analizar el comportamiento de la máquina con diferentes valores de  $\epsilon$ . Se presentarán los resultados por etapa variando el parámetro  $\epsilon$  en la sección de análisis de resultados.

- 10
  - 7.5
  - 5
  - 2.5
- 2
  - 1
  - 0.75
  - 0.5

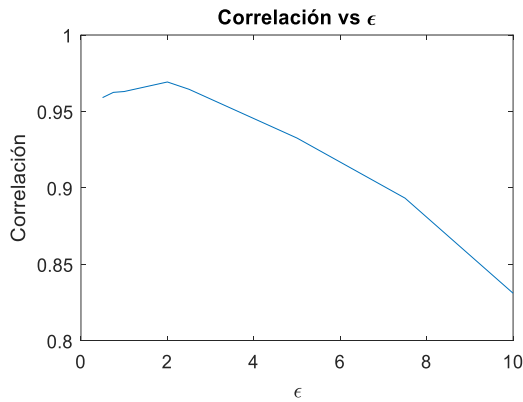


Figura 40. Relación entre correlación y  $\epsilon$  - Etapa Vegetativa

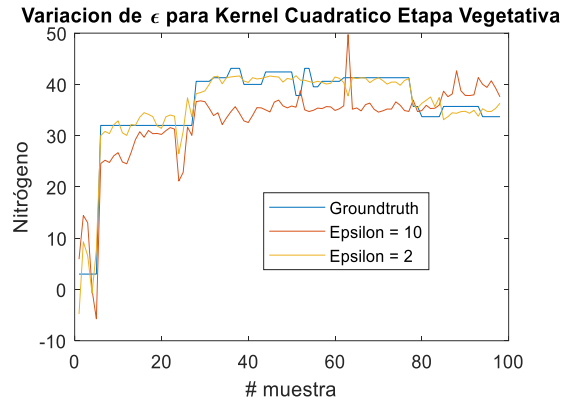


Figura 41. Comparación variación  $\epsilon$  - Etapa Vegetativa

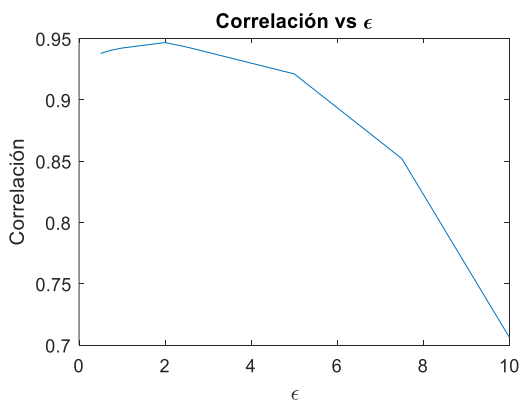


Figura 42. Relación entre correlación y  $\epsilon$  - Etapa Reproductiva

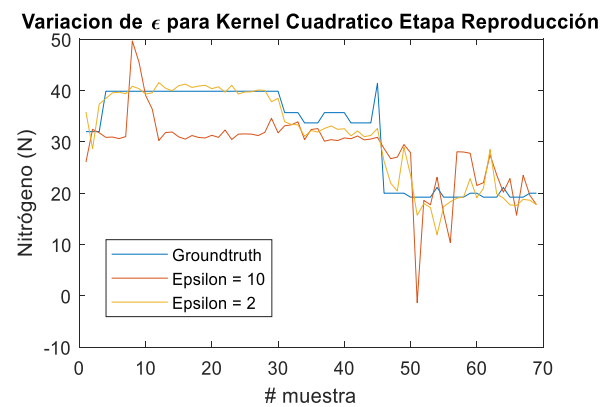


Figura 43. Comparación variación  $\epsilon$  - Etapa Reproductiva

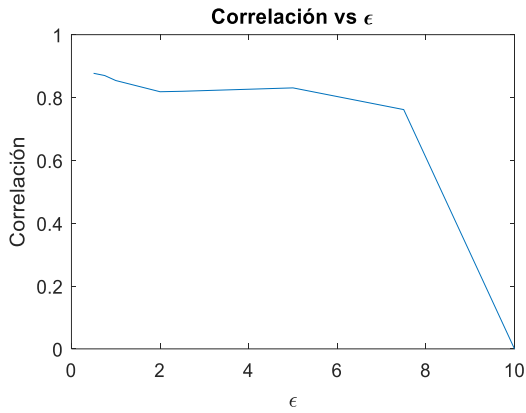


Figura 44. Relación entre correlación y  $\epsilon$  - Etapa Cosecha

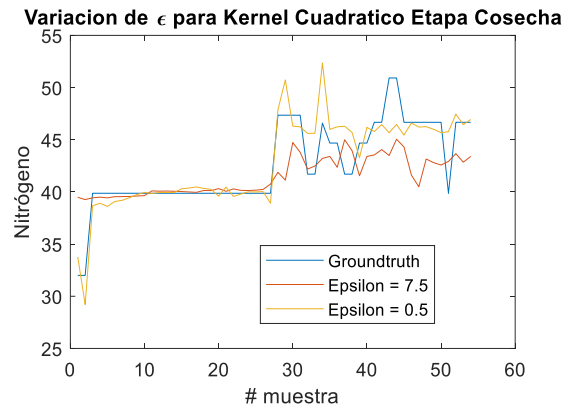


Figura 45. Comparación variación  $\epsilon$  - Etapa Cosecha

### 5.3. Redes Neuronales Artificiales

Como se mencionó previamente en el documento, el set de datos total de imágenes se encuentra dividido por cada una de las etapas de cultivo tenidas en cuenta para el desarrollo del trabajo de grado. A continuación, se muestran cada uno de los resultados de la evaluación de las diferentes redes neuronales implementadas para la solución de la problemática.

Inicialmente se puede apreciar la comparación entre redes neuronales con distinto número de capas. En las figuras 46 y 47 se muestran las dos redes neuronales contrastadas en esta parte. Consisten en una red de dos capas (una oculta y una externa) y una red de cinco capas (cuatro ocultas y una de salida) y en la figura 48 se observa el resultado obtenido a partir de estas dos arquitecturas.

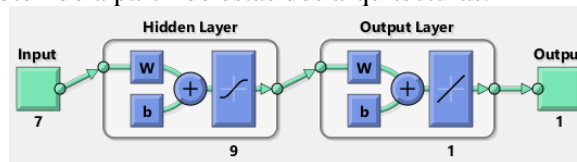


Figura 46. Red neuronal con 2 capas

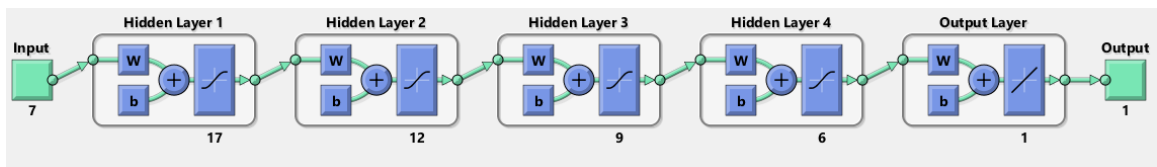


Figura 47. Red neuronal con 5 capas

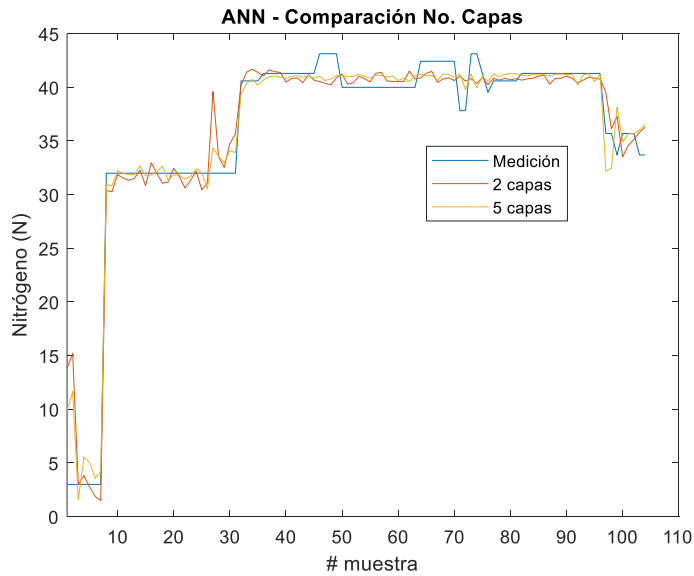


Figura 48. Comparación entre el número de capas de una red

El siguiente paso consiste en determinar qué función de aprendizaje modifica de manera más acercada los pesos de la red. Para esto, como fue mencionado previamente en la sección de desarrollo, en cada una de las etapas de cultivo se modelan redes con distintas funciones de aprendizaje y se evalúa su desempeño respecto a los datos esperados obtenidos en la medición de los cultivos. Vale mencionar nuevamente que cada una de estas redes fueron entrenadas con el 70% de los datos totales obtenidos y evaluados con el 30% restante.

En las figuras mostradas a continuación (49, 50, 51), se muestran los datos obtenidos por cada una de las redes neuronales utilizadas en comparación con las mediciones realizadas en la etapa vegetativa, reproductiva y cosecha respectivamente.

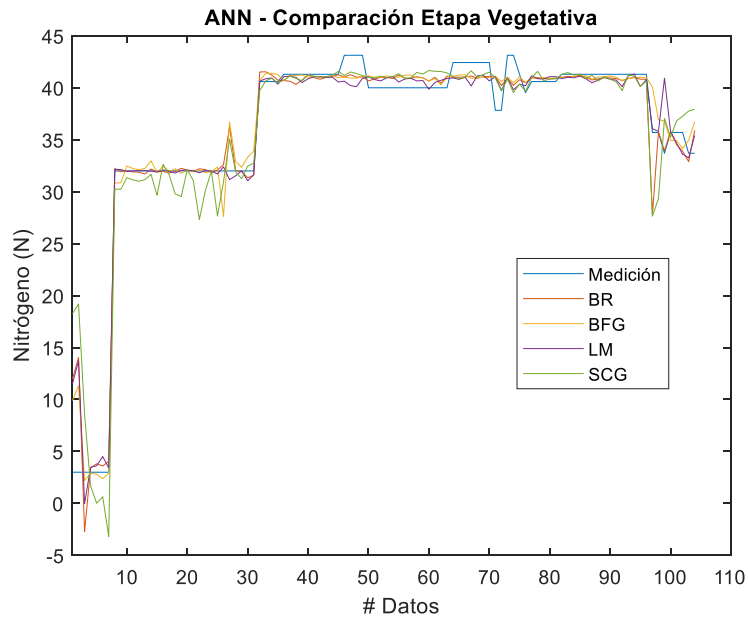


Figura 49. Implementación Redes Neuronales – Etapa Vegetativa

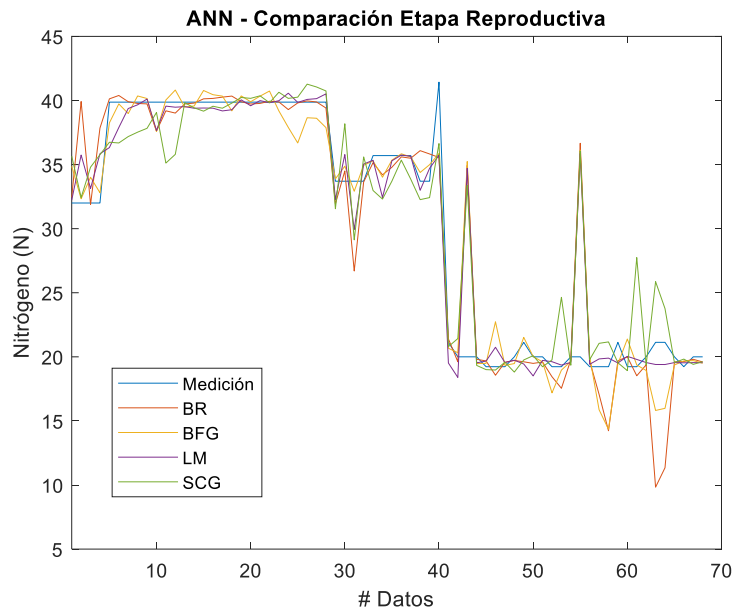


Figura 50. Implementación Redes Neuronales – Etapa Reproducción

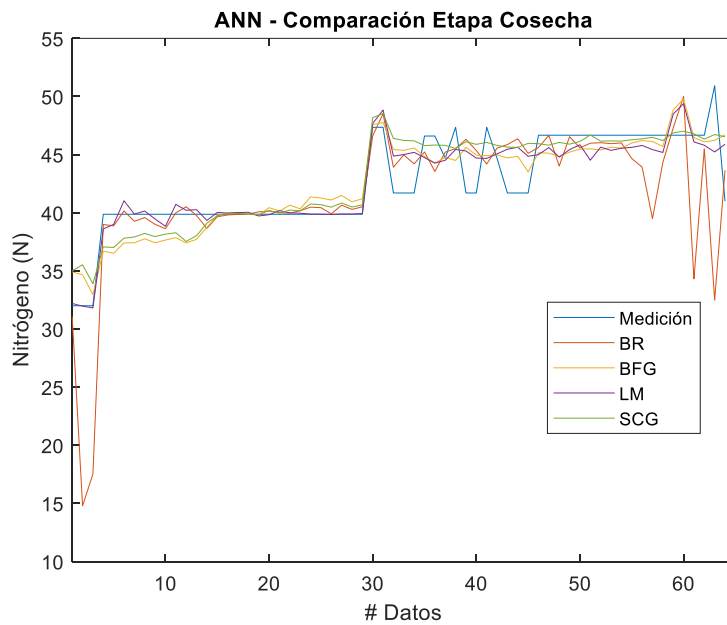


Figura 51. Implementación Redes Neuronales – Etapa Cosecha

Realizando un análisis cuantitativo entre las diferentes funciones de aprendizaje, evidenciado en la tabla X en la sección de pruebas, se escoge aquella red con mejor desempeño para cada una de las etapas y se varía el número de neuronas en la capa oculta, esto con el fin de visualizar de qué manera afecta esta variable en el desempeño de la red. En las figuras 52, 53 y 54 se observan los resultados obtenidos por cada máquina en cada una de las etapas de cultivo.

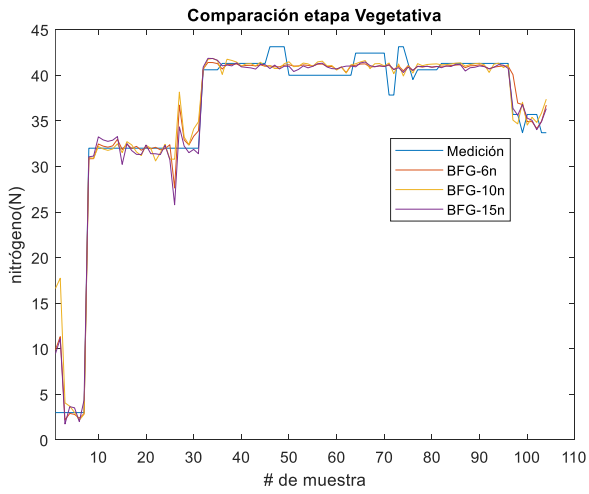


Figura 52. Comparación No. Neuronas – Etapa Vegetativa

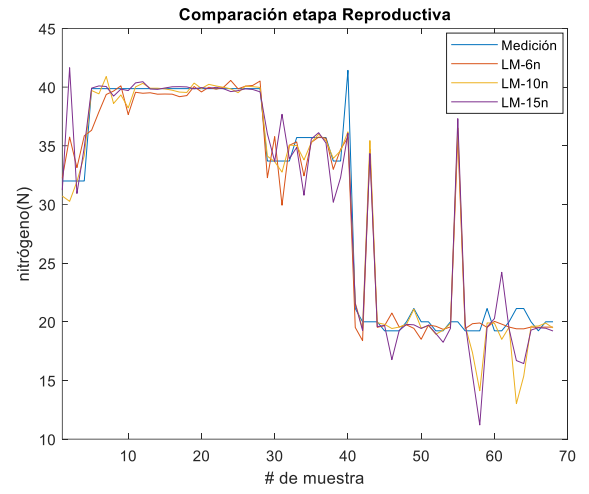


Figura 53. Comparación No. Neuronas – Etapa Reproductiva

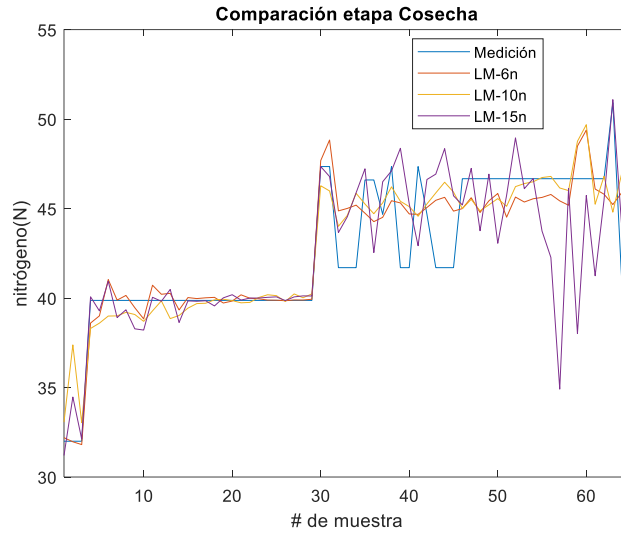


Figura 54. Comparación No. Neuronas – Etapa Cosecha



## 6. ANÁLISIS DE RESULTADOS

### 6.1. Implementación de Técnicas de Aprendizaje de Máquina

#### 6.1.1. Regresiones Lineales Multivariable

Se mostrarán los indicadores numéricos correspondientes al método de regresiones lineales en la tabla a continuación (3):

REGRESIONES LINEALES	Estimación sin filtrar		Estimación filtrada	
	CORRELACIÓN	RMSE	CORRELACIÓN	RMSE
Vegetativa (VEG)	0.935	3.688	0.929	3.914
Reproducción (REP)	0.851	4.80	0.890	4.356
Cosecha (RIP)	0.819	2.355	0.822	2.318

Tabla 3. Comparación Regresiones Lineales

Realizando una comparación entre los datos numéricos y las gráficas (28, 29, 30), se analiza que; aunque los datos filtrados suavizan como tal la tendencia de los datos, no siguen la referencia de cambios bruscos de los datos *groundtruth*. Si se fija la atención en la primera pendiente de subida de la figura 28, cabe notar que la estimación sin filtrar sigue esta recta de forma más aproximada que la función filtrada. Esto se remite al hecho de evitar vértices muy marcados por parte del filtro. Sin embargo, si se realiza una estimación poco acertada, el filtro minimizará este error, caso de los sets de datos correspondientes a las etapas de reproducción y cosecha (Figuras 29 y 30).

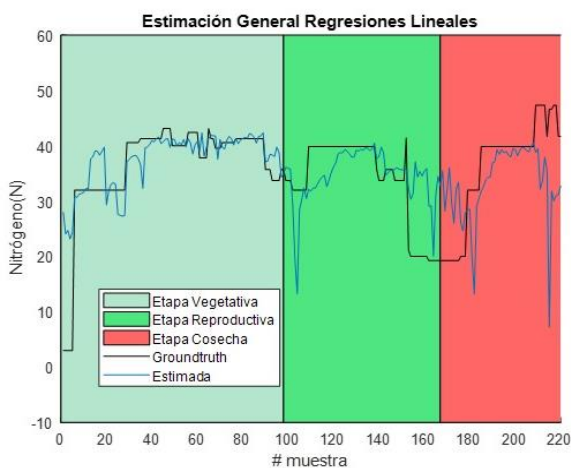


Figura 55. Resultados Estimación total Regresiones Lineales

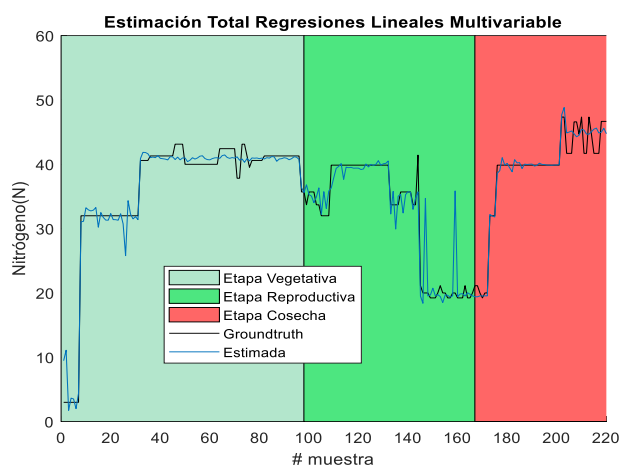


Figura 56. Resultados Entrenamiento Único Regresiones Lineales

	RMSE	CORRELACIÓN	R2
Regresiones – Múltiples Entrenamientos	3.660	0.921	0.849
Regresiones – Único Entrenamiento	7.974	0.467	0.218

Tabla 4. Comparación Regresiones Lineales – Tipo de Implementación

La figura 56 muestra las tendencias de datos superpuestas, tanto la estimada como los datos *groundtruth*. Esta gráfica muestra la unión de los datos estimados a partir de entrenamientos separados. A partir de esta gráfica se muestra una linealidad en los datos de entrada que pueden ser representados a través del método de regresiones lineales.

En la figura 55 se muestra la unión de los datos, buscando los coeficientes de modelamiento todos a la misma vez, por lo que al intentar modelar una función más compleja éstos no presentan una buena correlación, la linealidad entre todo el conjunto de datos no se encuentra tan presente y los datos estimados se presentan muy distantes de los datos de referencia *groundtruth*, tal como se ve en la tabla 4.

## 6.1.2. Máquinas de Soporte Vectorial

El análisis cuantitativo de cada una de las fases se presenta a continuación, el cuál involucra el *Root Mean Squared Error* o RMSE, el coeficiente de correlación y coeficiente de determinación.

Kernel	RMSE	CORRELACIÓN	R2
Cuadrático	2.34	0.96	0.927
Cúbico	3.75	0.90	0.827
Lineal	4.80	0.86	0.744
Gaussiano Medio	5.42	0.84	0.708
Gaussiano Grueso	5.87	0.79	0.634
Gaussiano Fino	6.07	0.76	0.579

Tabla 5. Análisis de Resultados - Etapa Vegetativa

Kernel	RMSE	CORRELACIÓN	R2
Cuadrático	2.33	0.94	0.892
Gaussiano Medio	3.03	0.92	0.887
Gaussiano Fino	4.02	0.91	0.882
Gaussiano Grueso	4.60	0.87	0.726
Lineal	5.88	0.81	0.656
Cúbico	12.28	0.34	0.119

Tabla 6. Análisis de Resultados - Etapa Reproductiva

Kernel	RMSE	CORRELACIÓN	R2
Cuadrático	2.04	0.87	0.770
Gaussiano Fino	2.16	0.87	0.761
Gaussiano Medio	2.35	0.84	0.713
Lineal	2.38	0.83	0.689
Gaussiano Grueso	2.97	0.81	0.660
Cúbico	4.39	0.57	0.326

Tabla 7. Análisis de Resultados - Etapa Cosecha

Dados los resultados presentados por las tablas 5, 6 y 7, basadas en las figuras desde la 31 a la 39, se concluye que el Kernel que presenta el mejor desempeño es el Kernel cuadrático teniendo como criterio el coeficiente de correlación.

En las figuras 40, 42 y 44 se muestra la correlación de los datos estimados en función del  $\epsilon$ , en las gráficas 41, 43 y 45 se evidencia una mayor correlación cuando el parámetro  $\epsilon$  disminuye su valor. Los valores presentados en las tablas 8, 9 y 10 corresponden a variaciones del parámetro épsilon en el Kernel cuadrático.

$\epsilon$	RMSE	CORRELACIÓN	R2
10	5,682	0,83	0,6378967
7,5	4,046	0,89	0,7979753
5	3,275	0,93	0,8141485
2,5	2,286	0,9645	0,9302674
2	2,119	0,9693	0,9310137
1	2,344	0,963	0,9274257
0,75	2,365	0,9625	0,9263674
0,5	2,469	0,9691	0,9197893

Tabla 8. Análisis Variación  $\epsilon$  - Etapa Vegetativa

$\epsilon$	RMSE	CORRELACIÓN	R2
10	6,992	0,7063	0,498895
7,5	6,063	0,8521	0,7260538
5	4,628	0,9211	0,8484772
2,5	3,080	0,9431	0,8893557
2	2,956	0,9467	0,8962258
1	3,056	0,9422	0,8878
0,75	3,158	0,9405	0,8845495
0,5	3,212	0,9378	0,8795224

Tabla 9. Análisis Variación  $\epsilon$  - Etapa Reproductiva

$\epsilon$	RMSE	CORRELACIÓN	R2
10	4,030	0.0000	0
7,5	3,073	0.7615	0,5798482
5	2,324	0.8306	0,6898657
2,5	2,401	0.8198	0,6720121
2	2,478	0.8184	0,6697144
1	2,284	0.8540	0,7292768
0,75	2,110	0.8699	0,7567429
0,5	2,060	0.8770	0,7691721

Tabla 10. Análisis Variación  $\epsilon$  – Etapa Cosecha

Partiendo del hecho que la región tubular que comprende los datos está en función del parámetro  $\epsilon$ , el hiper-plano generado debe aproximarse más a los vectores soporte generando una función más aproximada a la tendencia de datos *groundtruth*.

Las figuras 57 y 58 muestran la diferencia en cuanto a tendencia de datos con 2 valores diferentes del parámetro  $\epsilon$ .

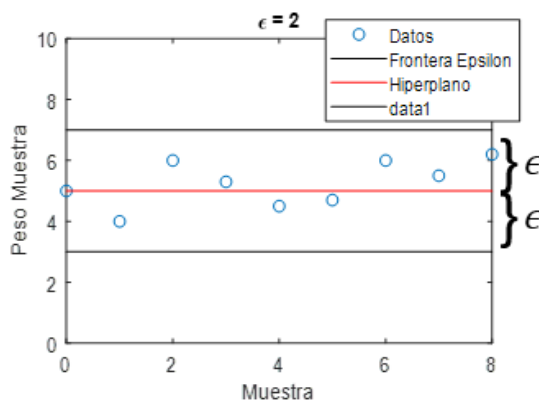


Figura 57. Análisis  $\epsilon=2$

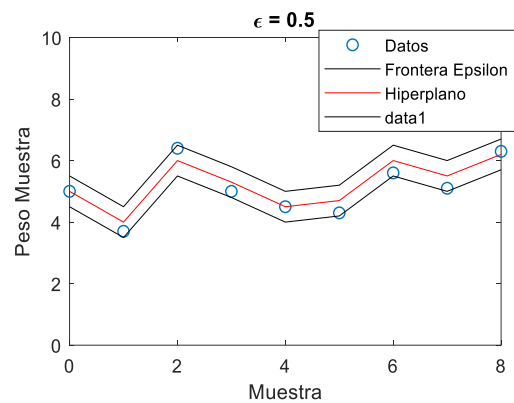


Figura 58. Análisis  $\epsilon=0.5$

La figura 58 presenta un modelado más acertado si es contrastado contra el modelado de la figura 57, dado que, al tener el parámetro  $\epsilon$  menor, el hiper-plano debe tener una distancia menor a los datos a estudiar, siguiendo la tendencia de una manera más aproximada.

Las siguientes gráficas muestran la unión de todos los datos diferenciando la forma de entrenamiento. La figura 59 muestra el resultado de la estimación con los datos entrenados por separado con las máquinas que mejor correlación y  $R^2$  presentaron, mientras que la figura 60 muestra el resultado de la unión de todos los datos y un solo entrenamiento para todos. Dado el resultado encontrado en esta sección, este set de datos fue entrenado con Kernel cuadrático con un valor  $\epsilon = 2$ , parámetros que mostraron el mejor desempeño en el entrenamiento por separado.

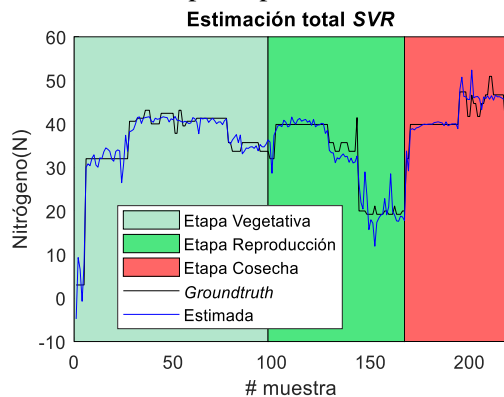


Figura 59. Estimación Total - SVR

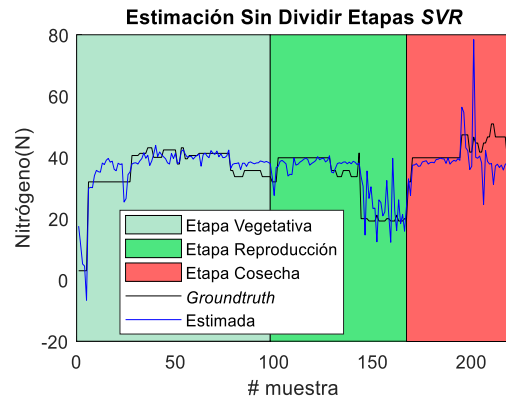


Figura 60. Estimación Total Entrenamiento Único - SVR

Las tablas 11 y 12 presentan los indicadores de la evaluación final de los datos entrenados por separado y con un solo entrenamiento.

ETAPA	MAQUINA SVR	RMSE	CORRELACION	$R^2$
VEGETATIVA	Kernel cuadrático $\epsilon=2$	2,1190	0,9693	0,931
REPRODUCCIÓN	Kernel cuadrático $\epsilon=2$	2,9560	0,9467	0,896
COSECHA	Kernel cuadrático $\epsilon=0.5$	2,0600	0,8700	0,769
TOTAL	-	2,3991	0,9638	0,928

Tabla 11. Estimación Total - SVR

	RMSE	CORRELACION	R2
Entrenamiento Único	5,442	0,799	0,638
Múltiples Entrenamientos	2.399	0.963	0.928

Tabla 12. Comparación SVR– Tipo de Implementación

### 6.1.3. Redes Neuronales Artificiales

En la tabla 13 se evidencia la comparación cuantitativa entre las redes neuronales de dos y cinco capas. En esta se nota que efectivamente a mayor número de capas, mayor será la correlación respecto a los valores reales. Sin embargo, la tabla muestra, además que, a mayor número de capas, mayor será el tiempo de procesamiento requerido para el entrenamiento de la red.

COMPARACIÓN NÚMERO DE CAPAS				
RED NEURONAL	CORRELACIÓN	RMSE	TIEMPO PROCESAMIENTO	R2
2 capas (9:1)	0.974	2.215	0 segundos	0.95
5 capas (17:12:9:6:1)	0.986	1.703	4 segundos	0.97

Tabla 13. Comparación Número de Capas en Redes Neuronales

Buscando no solo eficiencia en desempeño de la red sino también, en aras de reducir los tiempos de implementación y observando, asimismo, que la diferencia entre las correlaciones de las dos redes neuronales no es significativamente notoria, se decide diseñar las redes a utilizar con solo una capa oculta ya que ésta es suficiente para alcanzar el objetivo del proyecto.

Ahora, como fue evidenciado en la sección anterior, cada una de las redes neuronales presenta un desempeño distinto respecto al *groundtruth* obtenido en las mediciones de nitrógeno según la función de aprendizaje utilizada para modificar los pesos de la red. El análisis cuantitativo y comparación de cada una de las fases de cultivo se presenta enseguida en las tablas a continuación, evaluados a partir del RMSE, coeficiente de determinación ( $R^2$ ) y el coeficiente de correlación.

RED NEURONAL	CORRELACIÓN	RMSE	R2
Regresión Bayesiana (BR)	0.979	1.986	0.959
BFGS Quasi-Newton (BFG)	0.985	1.687	0.971
Levenberg-Marquardt (LM)	0.983	1.855	0.966
Gradiente escalado conjugado (SCG)	0.952	2.952	0.907

Tabla 14. Comparación funciones de aprendizaje – Etapa Vegetativa

RED NEURONAL	CORRELACIÓN	RMSE	R2
Regresión Bayesiana (BR)	0.921	3.779	0.848
BFGS Quasi-Newton (BFG)	0.938	3.203	0.880
Levenberg-Marquardt (LM)	0.9442	3.007	0.891
Gradiente escalado conjugado (SCG)	0.925	3.446	0.856

Tabla 15. Comparación funciones de aprendizaje – Etapa Reproductiva

RED NEURONAL	CORRELACIÓN	RMSE	R2
Regresión Bayesiana (BR)	0.67	4.405	0.462
BFGS Quasi-Newton (BFG)	0.851	2.149	0.724
Levenberg-Marquardt (LM)	0.890	1.835	0.792
Gradiente escalado conjugado (SCG)	0.862	2.115	0.742

Tabla 16. Comparación funciones de aprendizaje – Etapa Cosecha

Se observa en la tabla 14, correspondiente a la etapa vegetativa que, la red neuronal que presenta una mayor correlación es la red cuya función de entrenamiento es el método de optimización quasi-newton BFGS (Broyden – Fletcher - Goldbarb - Shanno) y por ende una tasa de error menor respecto a los datos adquiridos en las mediciones.

En ese mismo sentido, para la etapa de reproducción, mostrado en la tabla 15, la red que más asemeja sus resultados al *groundtruth* es la red con función de aprendizaje Levenberg-Marquardt, es decir, es la red que presenta una mayor correlación entre las demás redes implementadas. En la tabla 16, correspondiente a la etapa de cosecha, se evidencia que, de igual manera que para la etapa reproductiva, la red con mejor desempeño es aquella con función de aprendizaje Levenberg-Marquardt.

Ahora, conociendo qué función de aprendizaje brinda una unos resultados más aproximados a los obtenidos en las mediciones, se procede a variar la arquitectura de la red neuronal variando el número de neuronas de la capa oculta. A continuación, se encuentran las tablas comparativas de dicho procedimiento.

La tabla 17 indica que, para la etapa vegetativa, la red que brinda un menor error respecto al *groundtruth* contiene 15 neuronas en la capa oculta. Por otro lado, tanto para la etapa reproductiva y cosecha la red que brinda resultados más acercados a la realidad es aquella con 6 neuronas en la capa oculta, tal cual se evidencia en las tablas 18 y 19.

RED NEURONAL	CORRELACIÓN	RMSE	R2
BFGS Quasi-Newton (BFG)– 6 N	0.985	1.687	0.971
BFGS Quasi-Newton (BFG) – 10 N	0.971	2.38	0.944
BFGS Quasi-Newton (BFG) – 15 N	0.986	1.643	0.972

Tabla 17. Comparación No. Neuronas capa oculta – Etapa Vegetativa

RED NEURONAL	CORRELACIÓN	RMSE	R2
Levenberg-Marquardt (LM) – 6 N	0.9442	3.007	0.891
Levenberg-Marquardt (LM) – 10 N	0.939	3.201	0.883
Levenberg-Marquardt (LM) – 15 N	0.928	3.558	0.861

Tabla 18. Comparación No. Neuronas capa oculta – Etapa Reproductiva

RED NEURONAL	CORRELACIÓN	RMSE	R2
Levenberg-Marquardt (LM) – 6 N	0.890	1.835	0.792
Levenberg-Marquardt (LM) – 10 N	0.857	2.081	0.735
Levenberg-Marquardt (LM) – 15 N	0.739	2.926	0.546

Tabla 19. Comparación No. Neuronas capa oculta – Etapa Cosecha

Hecho el análisis previo, se concluye ahora cuáles, entre las redes neuronales propuestas, se escogen para representar de manera más acertada los niveles de nitrógeno a estimar. Las redes escogidas por etapa son:

- Etapa vegetativa: Red neuronal de 3 capas, 7 neuronas de en la capa entrada, 15 neuronas en la capa oculta, con función de transferencia sigmoidea tangencial y una neurona en la capa de salida, con función de transferencia lineal.  
Función de aprendizaje de la red: Método de optimización quasi-newton BFGS (Broyden – Fletcher - Goldbarb - Shanno)
- Etapa reproductiva: Red neuronal de 3 capas, 7 neuronas de en la capa entrada, 6 neuronas en la capa oculta, con función de transferencia sigmoidea logarítmica y una neurona en la capa de salida, con función de transferencia lineal.  
Función de aprendizaje de la red: Algoritmo Levenberg-Marquardt.
- Etapa de cosecha: Red neuronal de 3 capas, 7 neuronas de en la capa entrada, 6 neuronas en la capa oculta, con función de transferencia sigmoidea logarítmica y una neurona en la capa de salida, con función de transferencia lineal.  
Función de aprendizaje de la red: Algoritmo Levenberg-Marquardt.

En la tabla 20 se observa la red definitiva escogida para cada etapa de cultivo y la evaluación total de las redes implementadas.

ETAPA DE CULTIVO	RED NEURONAL	CORRELACIÓN	RMSE	R2
VEGETATIVA	BFGS Quasi-Newton (BFG) – 15 N	0.986	1.643	0.972
REPRODUCTIVA	Levenberg-Marquardt (LM) – 6 N	0.9442	3.007	0.891
COSECHA	Levenberg-Marquardt (LM) – 6 N	0.890	1.835	0.792
TOTAL	-	0.973	2.170	0.948

Tabla 20. Comparación Estimación Total – Redes Neuronales

En la figura 62 se muestra el comportamiento total de las redes neuronales definidas previamente. A su vez en la figura 61 se observan los resultados obtenidos si en vez de realizar una red neuronal por etapa se implementa solo una red para el conjunto de datos total. Esta red neuronal tiene seis neuronas en su capa oculta y su función de aprendizaje es el algoritmo Levenberg-Marquardt, ya que esta fue la red que predominó en el mayor desempeño entre las demás redes implementadas.

En la tabla 21 se logra observar la comparación cuantitativa entre las dos implementaciones de redes neuronales.

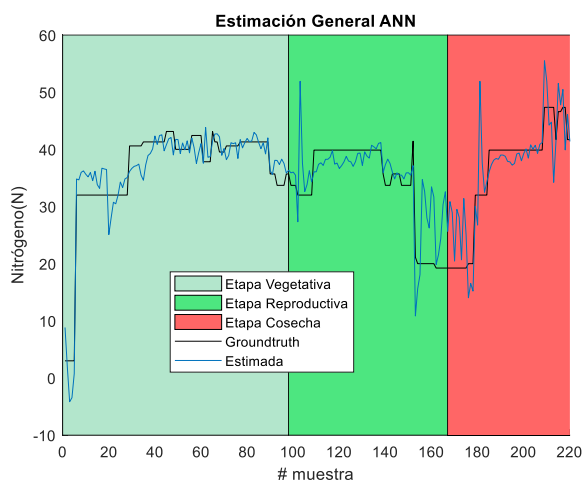


Figura 61. Estimación Total – Entrenamiento Único

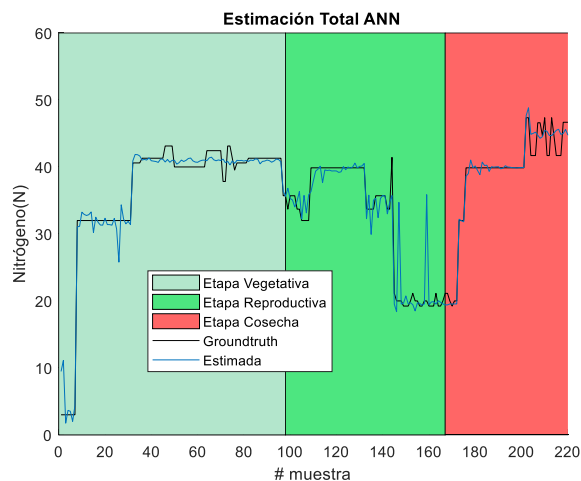


Figura 62. Estimación Total – Entrenamiento Múltiple

RED IMPLEMENTADA	CORRELACIÓN	RMSE	R2
Múltiple Entrenamiento	0.973	2.170	0.948
Entrenamiento Único	0.85	4.85	0.706

Tabla 21. Comparación entre implementaciones de redes neuronales

La tabla 22 que se presenta a continuación, se realiza la comparación general de las máquinas con mejor desempeño en cuanto a su correlación en las diferentes etapas de cultivo, comparando los valores promedio de los datos estimados contra los datos *groundtruth*, y finalmente su correlación.

Método	Valor Nitrógeno Promedio Estimado			Valor Nitrógeno Promedio Real			Correlación		
	Fases de cultivo			Fases de cultivo			Fases de cultivo		
	Vegetativa	Reproductiva	Cosecha	Vegetativa	Reproductiva	Cosecha	Vegetativa	Reproductiva	Cosecha
RL	36.923	30.766	42.669	35.7906	29.6338	42.4207	0.935	0.890	0.82
SVR	35.7472	29.2902	42.5226				0.9699	0.9467	0.8770
ANN	36.173	30.495	432.648				0.986	0.9442	0.890

Tabla 22. Tabla Resultante Total

## 6.2. Análisis ROC

Una vez conocidas cada una de las máquinas que mayor acercamiento tienen a la realidad para cada una de las técnicas de Aprendizaje de Máquina, se procede a valorar el desempeño de clasificación de las distintas máquinas respecto a las etapas de cultivo. Para esto, se evalúa la capacidad que tiene cada una de las máquinas para discriminar los niveles de nitrógeno estimados según la etapa de cultivo a la que

corresponden, ya sea etapa de reproducción o de cosecha. En la figura 63 se puede observar las curvas de desempeño ROC entregadas por cada una de las máquinas.

Cabe destacar que esta evaluación se hace únicamente entre las etapas de cultivo de reproducción y cosecha debido a que es en estas fases donde es más relevante conocer los niveles de concentración de nitrógeno para usos agrícolas.

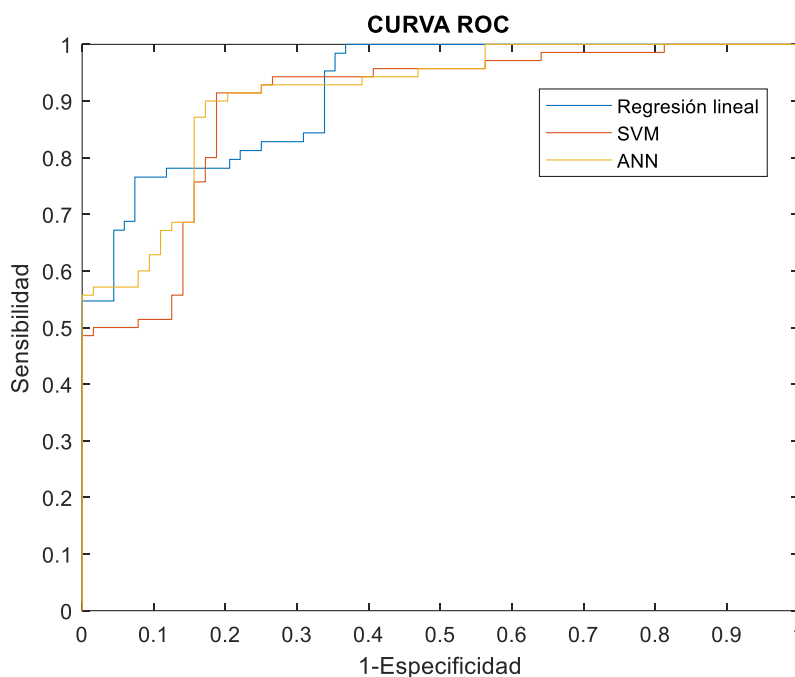


Figura 63. Curva ROC

Además, en la tabla 23 se observa la tabla de *Accuracy*, que entrega el porcentaje de exactitud de cada una de las máquinas, indicando cuál de todas posee una mayor probabilidad de clasificar los datos correctamente. La máquina con un mayor nivel de exactitud según esta medición son las redes neuronales.

	Exactitud
Regresiones Lineales	0.82
Máquinas de Soporte Vectorial	0.78
Redes Neuronales	0.85

Tabla 23. Accuracy

## 7. CONCLUSIONES

A lo largo del proyecto se pueden concluir ciertos aspectos relevantes. En primera instancia cabe resaltar que el entrenamiento separado de los datos brinda una correlación mucho mayor a un entrenamiento unificado de todos los datos. Esto debido a que, al analizar cada etapa por separado, se hace un análisis más detallado del comportamiento de los datos.

Cabe destacar que, la técnica que más aproxima los valores estimados a los niveles de nitrógeno obtenidos en las mediciones es el modelo computacional de redes neuronales artificiales; ya que éstas atacan de manera más contundente las no-linealidades de los datos en comparación con las regresiones lineales multivariable y, tienen un mayor grado de libertad a diferencia de las máquinas de soporte vectorial, las cuales dependen de una función matemática base para brindar una tendencia a los datos.

Por otro lado, se evidencia que, a mayor número de datos a la hora de entrenar, más se aproximarán las estimaciones a los datos esperados. Se ve reflejado que, la etapa que más datos tenía para ser entrenado obtuvo una mayor correlación en comparación con la que menos tuvo. Por lo que para trabajos futuros se recomienda finalizar con *sets* de datos de igual dimensión para todas las etapas, así el entrenamiento y las máquinas pueden ser evaluadas con las mismas condiciones.

Dado que la medida de nitrógeno fue generalizada para varias zonas de cultivo, es posible que los datos presenten cierta desviación de la realidad. El uso de un protocolo más especializado y estricto en cuanto a la ubicación de las medidas y su correspondiente fotografía aumentaría el nivel de confiabilidad de los datos.

El uso de Inteligencia Artificial (IA) aumenta de manera significativa la fiabilidad de los datos en comparación con técnicas más elementales, como fue el caso de las regresiones lineales, puesto que estas técnicas cubren la no linealidad de las tendencias de los datos. Aunque el proceso de las técnicas de IA presente un poco más de complejidad, estas técnicas brindan una alta certeza en cuanto a toma de decisiones si se refiere a temas de estimación o predicción.



## 8. REFERENCIAS

- [1] C. Perdomo, M. Barbazán, NITRÓGENO, Facultad de Agronomía, Universidad de la República
- [2] N. K. Fageria, V. C. Baligar, ENHANCING NITROGEN USE EFFICIENCY IN CROP PLANTS, Advances in Agronomy, 2005
- [3] MCL502, MC502DL, The SPAD.  
Ir a <http://www.gisiberica.com/Medidor%20de%20clorofila/mcl502.htm>
- [4] R. Cárdenas-Navarro, J.M. Sánchez, R. Farías, J.J. Peña, LOS APORTES DEL NITRÓGENO EN LA AGRICULTURA, Revista Chapon Serie Horticultura, 2004.
- [5] Nele Verhulst, Isabelle Francois, Rachael Cox, Bram Govaerts, EFICIENCIA DEL USO DE NITRÓGENO Y OPTIMIZACIÓN DE LA FERTILIZACIÓN NITRÓGENADA EN LA AGRICULTURA DE CONSERVACIÓN, Centro Internacional de Mejoramiento de Maíz y Trigo, 2015.
- [6] F. Izaurieta, C. Saavedra, REDES NEURONALES ARTIFICIALES, Departamento de Física, Universidad de Concepción, 2010
- [7] A. Palmer, J.J. Montaña, ¿QUÉ SON LAS REDES NEURONALES? APLICACIONES REALIZADAS EN EL ÁMBITO DE LAS ADICCIONES, Departamento de Psicología, Universidad de las Islas Baleares, 2012.
- [8] C. Tablada - G. Torres, REDES NEURONALES ARTIFICIALES, Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba. 2013.
- [9] INTRODUCCIÓN A LAS REDES NEURONALES APLICADAS. Ir a <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema3dm.pdf>
- [10] E. Soria, A. Blanco, REDES NEURONALES ARTIFICIALES.
- [11] D. J. Matich, REDES NEURONALES: CONCEPTOS BÁSICOS Y APLICACIONES, Facultad Regional Rosario, Universidad Tecnológica Nacional, 2001
- [12] F. Gómez, M. A. Fernández, M. T. López, M. Díaz, APRENDIZAJE CON REDES NEURONALES, Departamento de Informática, Escuela Universitaria Politécnica de Albacete.
- [13] L. F. Restrepo, J. González, DE PEARSON A SPEARMAN, Revista Colombiana de Ciencias Pecuarias, 2007.
- [14] Espectro Visible, Wikipedia Ir a [https://es.wikipedia.org/wiki/Espectro\\_visible](https://es.wikipedia.org/wiki/Espectro_visible)
- [15] S. Abad, DETECCIÓN DE ANOMALÍAS CARDÍACAS CON APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)
- [16] J. F. López, R CUADRADO (COEFICIENTE DE DETERMINACIÓN), Economipedia.
- [17] M. E. Burgos, C. Manterola, ASSESSMENT OF DIAGNOSTIC TEST STUDIES, Departamento de Cirugía y Traumatología, Universidad de la Frontera, 2010.

- [18] J. P. Rojas, C. A. Devia, E. Petro, C. Martínez, I. Mondragón, D. Patino, MC. Rebolledo, J. Colorado, AERIAL MAPPING OF RICE CROPS USING MOSAICING TECHNIQUES FOR VEGETATIVE INDEX MONITORING, International Conference on Unmanned Aircraft Systems, 2018.
- [17] M. K. Gómez, ÍNDICE DE VEGETACIÓN EN ÁREAS DEL BOSQUE SECO DEL NOROESTE DEL PERÚ A PARTIR DE IMÁGENES SATELITALES, Universidad de Piura, 2005
- [18] J. J. Díaz García-Cervigón, ESTUDIO DE ÍNDICES DE VEGETACIÓN A PARTIR DE IMÁGENES AÉREAS TOMADAS DESDE UAS/RPAS Y APLICACIONES DE ESTOS A LA AGRICULTURA DE PRECISIÓN, Universidad complutense de Madrid, 2015
- [19] REACCIONES DE LA FOTOSÍNTESIS DEPENDIENTES DE LA LUZ, de OpenStax College (CC BY 3.0).
- [20] A. Consentino, IDENTIFICATION OF PIGMENTS BY MULTISPECTRAL IMAGING; A FLOWCHART METHD, Heritage Science, 2014
- [21] P. Muñoz Aguayo, ÍNDICES DE VEGETACIÓN, Centro de información de recursos naturales, 2013
- [22] M. Abaurrea Pereda, COMPARACION DE INDICES DE VEGETACION EN ZONA SEMIÁRIDA DE NAVARRA, Universidad Pública de Navarra, 2013
- [23] M. A. Juzga Solanilla, COMPARACIÓN DE ÍNDICES DE VEGETACIÓN EN EL CERRO DE LA CONEJERA DE LA CIUDAD DE BOGOTÁ, Universidad Militar Nueva Granada, 2016
- [24] ¿QUÉ ES EL GNDVI? Ir a : <https://www.auravant.com/knowledge-base/que-es-gndvi/>
- [25] MODELO DE REGRESIÓN Ir a: <https://economipedia.com/definiciones/modelo-de-regresion.html>
- [26] J. Susan Milton, ESTADÍSTICA PARA BIOLOGÍA Y SALUD, S.A. Mcgraw-Hill/ Interamericana De España, 2007
- [27] M. J. Rodríguez R Mora Catalá, ANÁLISIS DE REGRESIÓN MÚLTIPLE, Universidad de Alicante, 2001
- [28] J. M. Rojo Abuín, REGRESIÓN LINEAL MÚLTIPLE, Instituto de Economía y Geografía, 2007
- [29] R. Norvig, INTELIGENCIA ARTIFICIAL: UN ENFOQUE MODERNO, Pearson 2004
- [30] A. Moreno E. Armengol J Béjar, APRENDIZAJE AUTOMÁTICA, Edición de la Universidad Politécnica de Catalunya, 1994
- [31] MACHINE LEARNING, Ir a: [https://la.mathworks.com/discovery/machine-learning.html?s\\_tid=srchtitle](https://la.mathworks.com/discovery/machine-learning.html?s_tid=srchtitle)
- [32] E. Cuevas Alfaro, MÁQUINAS DE SOPORTE VECTORIAL CON ALGORITMOS BASADOS EN POBLACIONES PARA EL PRONÓSTICO DEL PRECIO DE ACCIONES LAN CHILE, Pontificia Universidad Católica De Valparaíso, 2010
- [33] FUNCIÓN ARITMÉTICA DE LA BANDA Ir a: <http://desktop.arcgis.com/es/arcmap/10.3/manage-data/raster-and-images/band-arithmetic-function.htm>

[34] J.J. Martín, SUPPORT VECTOR REGRESSION: PROPIEDADES Y APLICACIONES, Universidad de Sevilla, 2012.

[35] CÁLCULO DE LA RAÍZ DEL ERROR CUADRÁTICO MEDIO O RMSE (ROOT MEAN SQUARED ERROR), Ir: <https://www.gestiondeoperaciones.net/proyeccion-de-demanda/calculo-de-la-raiz-del-error-cuadratico-medio-o-rmse-root-mean-squared-error/>

[36] P. Concejero, COMPARACIÓN DE MODELOS DE CURVAS ROC PARA LA EVALUACIÓN DE PROCEDIMIENTOS ESTADÍSTICOS DE PREDICCIÓN EN INVESTIGACIÓN DE MERCADOS, Universidad Complutense de Madrid, 2004

[37] Khan Academy, LUZ Y PIGMENTOS FOTOSINTETICOS Ir a: <https://es.khanacademy.org/science/biology/photosynthesis-in-plants/the-light-dependent-reactions-of-photosynthesis/a/light-and-photosynthetic-pigments>

[38] Support Vector Machine Ir a: [http://www.saedsayad.com/support\\_vector\\_machine\\_reg.htm](http://www.saedsayad.com/support_vector_machine_reg.htm)