



UNIVERSIDAD NACIONAL DE COLOMBIA

# Desarrollo de una metodología para identificación de características fisicoquímicas de productos agrícolas a partir de su correlación con técnicas de visión de máquina

Juan Pablo Bonilla Gonzalez

Universidad Nacional de Colombia  
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica  
Bogotá, Colombia  
Año 2016



# Desarrollo de una metodología para identificación de características fisicoquímicas de productos agrícolas a partir de su correlación con técnicas de visión de máquina

**Juan Pablo Bonilla Gonzalez**

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:  
**M.Sc. en Ingeniería - Automatización Industrial.**

Director:

Ph.D., M.Sc. Flavio Augusto Prieto Ortiz

Co-Directora:

Dr., M.Sc. Claudia Patricia Pérez Rodríguez

Línea de Investigación: Visión de Máquina

Grupo de Investigación: GAUNAL - Grupo de Automática de la Universidad Nacional

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica

Bogotá, Colombia

Año 2016





”La perfección de su ser no deja ningún lugar al hombre porque el hombre no podría trascenderse en Dios si Dios ya está todo entero dado. En tal caso el hombre no es más que un accidente indiferente a la realidad del ser; está en la tierra como un explorador perdido en el desierto; puede ir a la derecha o a la izquierda, puede ir a donde quiera; jamás irá a ningún lugar y la arena cubrirá sus huellas.”

Simone de Beauvoir

A mi madre y a mi hermano por ser el motor para seguir luchando.



## Resumen

La ausencia de prácticas automatizadas en la cosecha y poscosecha de los productos agrícolas, tiene como consecuencias un incremento en las pérdidas y una disminución en la competitividad de los mismos. La estimación de características fisicoquímicas en los productos mediante imágenes digitales puede permitir mejorar procesos tales como: selección y clasificación tanto para productores como consumidores. La metodología desarrollada con este fin permite correlacionar aspectos visibles en las imágenes tales como: color, textura, tamaño y forma con parámetros fisicoquímicos medidos. Se proponen cuatro fases fundamentales: *i.* identificación y medición de las características, *ii.* procesamiento de imágenes y extracción de características, *iii.* estimación de las características fisicoquímicas y, finalmente, *iv.* validación de la correlación. Ésto, permite generar sistemas de visión de máquina automáticos para estimar a futuro, dichas propiedades fisicoquímicas; mediante una técnica no destructiva y rápida. Los resultados obtenidos al aplicar la metodología para estimar algunas de las principales características fisicoquímicas en diferentes productos, son superiores al 80 % en términos del coeficiente de correlación, con una disminución significativa del porcentaje de error respecto a la desviación estándar de la muestra.

**Palabras clave:** Automatización, Cosecha, Inteligencia artificial, Poscosecha, Productos agrícolas, Visión de máquina.

## Abstract

The absence of automation technologies for harvest and postharvest practices on agricultural products, has as a consequence an increase in losses and a decline in the competitiveness of the same. The estimation of physicochemical characteristics in agricultural products using digital images can allow improving processes such as selection and classification for both: producers and consumers. The methodology developed for this purpose enables correlate aspects in visible images such as color, texture, size and shape with measured physicochemical parameters. Four stages was proposed: *i.* identification and measurement of the characteristics, *ii.* image processing and feature extraction, *iii.* physicochemical characteristics estimation and *iv.* validation of the correlation. By following these steps is possible to construct machine vision systems for future automatic estimations of these physicochemical properties; with a nondestructive and rapid technique. The results obtained by applying of these methodology to estimate some of the main physicochemical characteristics in different products, are over 80 % in terms of the correlation coefficient with a significant decrease of the rate error relative to the standard deviation of the sample set.

**Keywords:** Agricultural products, Artificial Intelligence, Automation, Harvest, Machine Vision, Postharvest.

# Contenido

<b>Resumen</b>	<b>vii</b>
Lista de figuras . . . . .	XI
Lista de tablas . . . . .	XIII
Lista de símbolos . . . . .	XV
<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico.</b>	<b>3</b>
2.1. Medición de propiedades físicoquímicas en productos agrícolas. . . . .	4
2.1.1. Sólidos solubles. . . . .	4
2.1.2. Firmeza. . . . .	5
2.1.3. Acidez titulable. . . . .	5
2.1.4. Masa. . . . .	6
2.1.5. Volumen. . . . .	6
2.1.6. Color. . . . .	7
2.2. Visión de máquina en productos agrícolas. . . . .	8
2.2.1. Espacio de color RGB. . . . .	9
2.2.2. Espacio de color CIELab. . . . .	10
2.2.3. Constancia de color. . . . .	11
2.2.4. Matriz de co-ocurrencia de niveles de gris. . . . .	13
2.2.5. Momentos de Hu. . . . .	13
2.3. Técnicas matemáticas y computacionales para correlación de parámetros . .	15
2.3.1. Análisis por componentes principales. . . . .	15
2.3.2. Análisis discriminante lineal. . . . .	15
2.3.3. Regresión Lineal Múltiple. . . . .	16
2.3.4. Redes Neuronales Artificiales . . . . .	18
<b>3. Metodología para la estimación de características físicoquímicas en productos agrícolas a partir del análisis de imágenes digitales.</b>	<b>20</b>
3.1. Identificación y medición de las características físicoquímicas de interés en el producto agrícola. . . . .	21

3.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	24
3.2.1. Constancia de color. . . . .	26
3.2.2. Segmentación de Regiones. . . . .	27
3.2.3. Medición de componentes de color. . . . .	30
3.2.4. Análisis de textura MCNG. . . . .	34
3.2.5. Características de tamaño y forma. . . . .	35
3.2.6. Composición del vector de características. . . . .	37
3.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	39
3.3.1. Análisis ACP. . . . .	39
3.3.2. Análisis ADL. . . . .	42
3.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	46
<b>4. Validación de la metodología en casos de estudio.</b>	<b>50</b>
4.1. Diseño experimental . . . . .	51
4.2. Caso de Estudio: Maracuyá - <i>P. edulis</i> . . . . .	52
4.2.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola. . . . .	52
4.2.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	54
4.2.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	57
4.2.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	61
4.3. Caso de Estudio: Granadilla - <i>P. Ligularis</i> . . . . .	63
4.3.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola. . . . .	63
4.3.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	64
4.3.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	65
4.3.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	66
4.4. Caso de Estudio: Mango Común - <i>M. Indica L.</i> . . . . .	68
4.4.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola. . . . .	68
4.4.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	69

---

4.4.3.	Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	70
4.4.4.	Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	70
4.5.	Caso de Estudio: Tomate - <i>S. lycopersicum</i> . . . . .	72
4.5.1.	Identificación y medición de las características fisicoquímicas de interés en el producto agrícola. . . . .	72
4.5.2.	Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	73
4.5.3.	Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	74
4.5.4.	Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	75
4.6.	Caso de Estudio: Mango - <i>Nam doc mai</i> . . . . .	77
4.6.1.	Identificación y medición de las características fisicoquímicas de interés en el producto agrícola. . . . .	77
4.6.2.	Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola. . . . .	78
4.6.3.	Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes. . . . .	80
4.6.4.	Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. . . . .	86
4.7.	Análisis y discusión de los resultados. . . . .	89
<b>5.</b>	<b>Conclusiones y trabajo futuro</b>	<b>93</b>
<b>A.</b>	<b>Datos físicos medidos en productos agrícolas.</b>	<b>95</b>
	<b>Referencias</b>	<b>98</b>

# Lista de Figuras

<b>2-1.</b>	Refractómetro de mano utilizado en la medición del contenido de sólidos solubles [50]. . . . .	4
<b>2-2.</b>	Penetrómetro utilizado para la medición de firmeza [76, 32]. . . . .	5
<b>2-3.</b>	Balanza utilizada para la medición de masa en gramos [32]. . . . .	6
<b>2-4.</b>	Métodos utilizados para la medición del volumen en productos agrícolas [32]. . . . .	7
<b>2-5.</b>	Equipo electrónico (Colorímetro) y ejemplo del proceso de medición de color . . . . .	7
<b>2-6.</b>	Elementos de un sistema de visión de máquina utilizando imágenes digitales [67]. . . . .	8
<b>2-7.</b>	Sistema de visión de máquina y los niveles del procesamiento de imágenes [33]. . . . .	9
<b>2-8.</b>	Modelo de color RGB [113]. . . . .	9
<b>2-9.</b>	Discretización de RGB reduciendo de (a) 16'777.216 niveles de color hacia (b) 64 niveles de color [124]. . . . .	10
<b>2-10.</b>	Modelo de color Lab. . . . .	11
<b>2-11.</b>	Tres adquisiciones de la misma escena bajo diferentes condiciones de iluminación [109]. . . . .	12
<b>2-12.</b>	Transición desde el proceso biológico a las redes neuronales artificiales a través de la simplificación del modelo de la neurobiología y la neurociencia computacional [5]. . . . .	18
<b>2-13.</b>	Regla de la detención temprana basada en la validación cruzada de la curva de aprendizaje [42]. . . . .	19
<b>3-1.</b>	Tipos de imágenes de productos agrícolas obtenidas. . . . .	24
<b>3-2.</b>	(a) Imagen original capturada por la cámara; (b) Imagen segmentada que muestra algunas regiones como la piel, el tallo y las regiones de daño; (c) Región que contiene todas las clases, excepto el vástago y el fondo, que se utiliza para calcular el tamaño; y (d) Imagen que muestra cómo se calcula el tamaño [13]. . . . .	25
<b>3-3.</b>	Elementos de un sistema de visión de máquina con imágenes a color y separación de planos RGB en la agricultura [24]. . . . .	30
<b>3-4.</b>	Los valores medios de coordenadas de color (centros de las esferas) y desviaciones estándar (límites de las superficies de las esferas) de diferentes lotes de hortalizas (nueve vegetales medidos por lote), representados en el espacio 3-D [72]. . . . .	31
<b>3-5.</b>	Variaciones del color durante el proceso de maduración en productos agrícolas [38]. . . . .	32
<b>3-6.</b>	Análisis de calidad y defectos en productos agrícolas [64]. . . . .	35
<b>3-7.</b>	Extracción de características de tamaño en productos agrícolas mediante imágenes [39]. . . . .	36
<b>3-8.</b>	Estimación de propiedades físicas mediante ACP. . . . .	41
<b>3-9.</b>	Estimación de propiedades químicas mediante ACP. . . . .	41
<b>3-10.</b>	Separación en hiperplanos mediante la técnica ADL [77]. . . . .	42
<b>3-11.</b>	Estimación de propiedades físicas mediante ADL. . . . .	45
<b>3-12.</b>	Estimación de propiedades químicas mediante ADL. . . . .	45
<b>3-13.</b>	Validación de la estimación de propiedades físicas mediante ACP y ADL. . . . .	47

<b>3-14.</b> Validación de la estimación de propiedades químicas mediante ACP y ADL. . . . .	47
<b>4-1.</b> Gráfico de barras comparativo de masa y volumen en términos de frecuencia absoluta para maracuyá. . .	53
<b>4-2.</b> Imagen de maracuyá del banco de imágenes obtenido para el análisis. . . . .	54
<b>4-3.</b> Procesamiento OHTA realizado a una muestra del banco de imágenes de maracuyá. . . . .	55
<b>4-4.</b> Procesamiento de imagen realizado a una muestra del banco de imágenes de maracuyá. . . . .	56
<b>4-5.</b> Curva de aprendizaje para la RNA luego del ACP; Iteraciones (Épocas) vs. E.C.M. (Error cuadrático medio)	60
<b>4-6.</b> Masa medida [g] vs. Masa estimada [g] utilizando una RNA luego de aplicar un ADL en maracuyá . . .	62
<b>4-7.</b> Volumen medido [ $cm^3$ ] vs. Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en maracuyá	62
<b>4-8.</b> Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para granadilla. . . . .	64
<b>4-9.</b> Procesamiento de imagen realizado a una muestra del banco de imágenes de granadilla. . . . .	65
<b>4-10.</b> Masa medida [g] vs. Masa estimada [g] utilizando una RNA luego de aplicar un ADL en granadilla. . . .	67
<b>4-11.</b> Volumen medido [ $cm^3$ ] vs Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en granadilla.	67
<b>4-12.</b> Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para mango común. . . . .	69
<b>4-13.</b> Procesamiento de imagen realizado a una muestra del banco de imágenes de mango común. . . . .	69
<b>4-14.</b> Masa medida [g] vs Masa estimada [g] utilizando una RNA luego de aplicar un ADL en mango común. .	71
<b>4-15.</b> Volumen medido [ $cm^3$ ] vs. Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en mango común . . . . .	71
<b>4-16.</b> Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para tomate. . . . .	73
<b>4-17.</b> Evolución de un fruto de tomate evidenciado en imágenes. . . . .	73
<b>4-18.</b> Procesamiento de imagen realizado a una muestra del banco de imágenes de tomate. . . . .	74
<b>4-19.</b> Masa medida [g] vs Masa estimada [g] utilizando una RNA luego de aplicar un ADL en tomate. . . . .	75
<b>4-20.</b> Volumen medido [ $cm^3$ ] vs Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en tomate.	76
<b>4-21.</b> Gráfico de barras para la comparación de masa, volumen, SS, acidez y pH en términos de su frecuencia absoluta para mango nam doc mai. . . . .	79
<b>4-22.</b> Procesamiento de imagen realizado a una muestra del banco de imágenes de Mango Nam doc mai. . . .	80
<b>4-23.</b> Estimación de propiedades físicas mediante ACP. . . . .	82
<b>4-24.</b> Estimación de propiedades químicas mediante ACP. . . . .	82
<b>4-25.</b> Estimación de propiedades físicas mediante ADL. . . . .	85
<b>4-26.</b> Estimación de propiedades químicas mediante ADL. . . . .	85
<b>4-27.</b> Validación de la estimación de propiedades físicas mediante ACP y ADL. . . . .	86
<b>4-28.</b> Validación de la estimación de propiedades químicas mediante ACP y ADL. . . . .	87
<b>4-29.</b> Comparación entre la media y desviación estándar medidos vs estimados para diferentes productos agrícolas.	92



# Lista de Tablas

<b>3-1.</b>	Propiedades fisicoquímicas en productos agrícolas. . . . .	22
<b>3-2.</b>	Constancia de color a partir de diferentes iluminaciones sobre un producto agrícola. . . . .	27
<b>3-3.</b>	Visibilidad de la Región de Interés(VRI) al aplicar la técnica de segmentación OHTA en diferentes productos agrícolas. . . . .	29
<b>3-4.</b>	Características de color extraídos en diferentes frutos. . . . .	34
<b>3-5.</b>	Características de textura MCNG extraídas en diferentes frutos. . . . .	34
<b>3-6.</b>	Características de forma (Momentos de Hu) extraídas en diferentes frutos. . . . .	36
<b>3-7.</b>	Media y desviación estándar de las 22 características extraídas para diferentes frutos. . . . .	38
<b>3-8.</b>	Vector propio del mayor valor propio ACP en términos de las 22 características en diferentes frutos. . . . .	40
<b>3-9.</b>	Vectores y valores propios del primer discriminante ADL en términos de las 22 características para la estimación de masa en diferentes frutos. . . . .	43
<b>3-10.</b>	Vectores y valores propios del primer discriminante ADL en términos de las 22 características para la estimación de volumen en diferentes frutos. . . . .	43
<b>3-11.</b>	Vectores y valores propios del ADL en términos de las 22 características para la separación por clases en Mango Nam doc mai. . . . .	44
<b>3-12.</b>	Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai. . . . .	48
<b>3-13.</b>	Análisis estadístico de correlación entre los datos de visión y el volumen de frutos de Mango Nam doc mai. . . . .	48
<b>3-14.</b>	Análisis estadístico de correlación entre los datos de visión y los sólidos solubles de frutos de Mango Nam doc mai. . . . .	48
<b>3-15.</b>	Análisis estadístico de correlación entre los datos de visión y la acidez de frutos de Mango Nam doc mai. . . . .	49
<b>3-16.</b>	Análisis estadístico de correlación entre los datos de visión y el pH de frutos de Mango Nam doc mai. . . . .	49
<b>4-1.</b>	Tabla de frecuencias para la masa [g] de frutos de maracuya. . . . .	53
<b>4-2.</b>	Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de maracuya. . . . .	53
<b>4-3.</b>	Análisis estadístico de correlación entre los datos de visión y la masa de frutos de maracuyá. . . . .	61
<b>4-4.</b>	Tabla de frecuencias para la masa [g] de frutos de granadilla. . . . .	63
<b>4-5.</b>	Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de granadilla. . . . .	63
<b>4-6.</b>	Tabla de frecuencias para la masa [g] de frutos de mango común. . . . .	68
<b>4-7.</b>	Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de mango común. . . . .	68
<b>4-8.</b>	Tabla de frecuencias para la masa [g] de frutos de tomate. . . . .	72
<b>4-9.</b>	Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de tomate. . . . .	72
<b>4-10.</b>	Tabla de frecuencias para la masa [g] de frutos de mango nam doc mai. . . . .	77
<b>4-11.</b>	Tabla de frecuencias para el volumen [ $cm^3$ ] de mango nam doc mai. . . . .	77

4-12. Tabla de frecuencias para sólidos solubles [ <i>Brix</i> ] de mango nam doc mai. . . . .	78
4-13. Tabla de frecuencias para acidez de mango nam doc mai. . . . .	78
4-14. Tabla de frecuencias para pH de mango nam doc mai. . . . .	78
4-15. Vectores y valores propios del ACP en términos de las 22 características para la separación por variabilidad en Mango Nam doc mai. . . . .	81
4-16. Vectores y valores propios del ADL en términos de las 22 características para la separación por clases en Mango Nam doc mai. . . . .	84
4-17. Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai. . . . .	88
4-18. Análisis estadístico de correlación entre los datos de visión y el volumen de frutos de Mango Nam doc mai. . . . .	88
4-19. Análisis estadístico de correlación entre los datos de visión y los sólidos solubles de frutos de Mango Nam doc mai. . . . .	88
4-20. Análisis estadístico de correlación entre los datos de visión y la acidez de frutos de Mango Nam doc mai. . . . .	88
4-21. Análisis estadístico de correlación entre los datos de visión y el pH de frutos de Mango Nam doc mai. . . . .	88
4-22. Estadísticos descriptivos de correlación entre la masa estimada mediante una RNA luego de ADL y la masa real medida en frutos de maracuyá, granadilla, mango y tomate. . . . .	89
4-23. Estadísticos descriptivos de correlación entre el volumen estimado mediante una RNA luego de ADL y el volumen real medido en frutos de maracuyá, granadilla, mango y tomate. . . . .	89
4-24. Análisis de resultados en proporción para la estimación de masa [g]. . . . .	90
4-25. Análisis de resultados en proporción para la estimación de volumen [ $cm^3$ ]. . . . .	91
4-26. Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai en diferentes características fisicoquímicas usando una RNA luego de ADL. . . . .	91
A-1. Datos medidos de masa [g] y volumen [ $cm^3$ ] en maracuyá. . . . .	95
A-2. Datos medidos de masa [g] y volumen [ $cm^3$ ] en mango común. . . . .	96
A-3. Datos medidos de masa [g] y volumen [ $cm^3$ ] en granadilla. . . . .	96
A-4. Datos medidos de masa [g] y volumen [ $cm^3$ ] en tomate. . . . .	97

# Lista de símbolos

<b>Abreviatura</b>	<b>Término</b>
<i>ACP</i>	Análisis por Componentes Principales
<i>ADL</i>	Análisis Discriminante Lineal
<i>ANN</i>	Artificial Neural Networks
<i>AT</i>	Acidez Titulable
<i>CSS</i>	Contenido de Sólidos Solubles
<i>FMA</i>	Factor Miliequivalente
<i>FFB</i>	Fresh Fruit Bunches
<i>GLCM</i>	Gray Level Co-occurrence Matrix
<i>IC</i>	Índice de Color
<i>LDA</i>	Linear Discriminant Analysis
<i>MCNG</i>	Matriz de Co-ocurrencia de Niveles de Gris
<i>MLR</i>	Multiple Linear Regression
<i>PCA</i>	Principal Component Analysis
<i>RFF</i>	Racimos de Fruta Fresca
<i>RGB</i>	Red, Green, Blue
<i>RLM</i>	Regresión Lineal Múltiple
<i>RNA</i>	Redes Neuronales Artificiales
<i>ROI</i>	Region of Interest
<i>SSC</i>	Soluble Solid Content
<i>TA</i>	Titratable Acidity
<i>cm<sup>3</sup></i>	Centímetros cúbicos
<i>m</i>	Masa
<i>g</i>	Gramos

# 1. Introducción

La ausencia de prácticas adecuadas durante las etapas de cosecha y poscosecha de los productos agrícolas que permitan conservar y prolongar la vida útil de los frutos; hace que se presenten altas pérdidas y bajos niveles de competitividad. Algunas de estas prácticas asociadas con la clasificación uniforme de los frutos según características referentes a su estado fisicoquímico y madurez como parámetros de calidad, son aún susceptibles a la automatización. Pese a que se han realizado avances en la estandarización de protocolos y métodos, no son aplicados de manera amplia. Éstos, se dejan en el plano de la observación y experiencia de quién manipula los productos agrícolas. Por lo tanto, aún es mucho lo que se puede mejorar y aparece la necesidad de automatizar como una alternativa útil a los métodos conocidos.

La identificación de dichas características en los productos agrícolas se realiza generalmente mediante inspección visual y al tacto, y es realizada por expertos productores, basándose en su experiencia empírica. Sin embargo, mediante equipos y métodos de análisis en laboratorio (algunos de ellos destructivos), es posible medir objetivamente dichos atributos físicos y químicos. La finalidad de esta tesis de investigación es el desarrollo de una metodología específica, que permita estimar algunas de estas características fisicoquímicas de manera consistente, con el fin de presentar una alternativa a procesos destructivos en laboratorio, y optimizar tiempos en la caracterización de los productos.

Para tal fin se utilizaron técnicas de visión de máquina y se correlacionaron variables fisicoquímicas en diferentes productos tomados como casos de estudio, con características extraídas por medio de imágenes digitales en el espectro visible (2D) del fruto. Éstas características extraídas, intentan emular el proceso realizado por expertos, basándose en información de color, textura, tamaño y forma; apreciables en el fruto, y que sirven como base para la estimación *a priori* de parámetros fisicoquímicos mediante la experiencia. La metodología determina un protocolo a seguir, para diseñar y desarrollar un sistema de visión de máquina capaz de ser aplicado no solo a los casos de estudio sino también extrapolable a otros productos agrícolas.

Para lograr este objetivo se desarrollaron cuatro fases principales: la primera basada en la medición tradicional en laboratorio de algunos parámetros fisicoquímicos asociados al desarrollo del fruto pero, adquiriendo paralelamente imágenes del producto (casos de estudio). En la segunda fase, se desarrollaron algoritmos de visión de máquina que procesaran las imágenes y extrajeran características de manera automática, análogamente como lo realizan los expertos. En la tercera etapa, se correlacionaron las características extraídas mediante imágenes, con los parámetros fisicoquímicos medidos experimentalmente y, finalmente, se evaluaron los resultados obtenidos por el sistema de visión de máquina contra la medición real realizada en laboratorio. La implementación de ésta metodología, servirá como herramienta para la estimación y/o predicción automática de las características fisicoquímicas, sin tener que destruir el fruto.

Para el desarrollo de cada una de las fases, este documento presenta la fundamentación teórica que incluye el estado del arte y el proceso de medición tradicional de los parámetros fisicoquímicos. También, se presenta en detalle la metodología propuesta utilizando técnicas de visión de máquina; pasando por su justificación hasta su aplicación. Posteriormente se presentan los resultados obtenidos a partir de la aplicación de dicha metodología en la estimación de características fisicoquímicas en: maracuyá, granadilla, mango y tomate. Se validó su funcionamiento, determinando una estimación de correlación mediante el coeficiente de correlación múltiple superior al 80 %, de manera consistente en todos los productos de prueba. Finalmente, se presentan las conclusiones y posibilidades de trabajos futuros.

## **Resumen del capítulo**

En este capítulo se presenta una introducción al por qué de identificar las características fisicoquímicas en productos agrícolas. Identificar las características fisicoquímicas en los productos agrícolas es útil en diferentes procesos. Para medir objetivamente estos atributos es necesario de diferentes sensores, instrumentos y equipos; dando lugar a la destrucción del fruto en algunas ocasiones. Debido a lo anterior, es común, que los encargados del manejo del producto desarrollen métodos basados en la experiencia para estimar subjetivamente aspectos como la madurez y/o calidad del producto, mediante inspección visual. La tesis de investigación, propone una metodología compuesta por cuatro fases fundamentales, para realizar estimaciones cuantificadas de manera automática, utilizando técnicas de visión de máquina. A continuación se presenta la fundamentación teórica para el desarrollo de la metodología.

## 2. Marco Teórico.

La diversidad del conjunto de frutas, hortalizas y granos comercializados internacionalmente establece una gran riqueza, que nos ofrece productos a lo largo de todo el año, con amplio origen botánico, de características muy diversas (cultivo, estructura, morfología, color, valor nutritivo, coste, manipulación, conservación, consumo, etc.) que, en muchos casos se encuentran ligadas a la cultura de los grupos étnicos [6, 15, 10, 40].

Diferentes operaciones entre ellas: selección y clasificación en los productos agrícolas suelen estar estrechamente relacionados con la comercialización, realizándose generalmente de forma manual pero con un potencial uso de medios automatizados, en algunos de ellos [25, 55, 56, 100]. La creciente exigencia de calidad por parte de los consumidores de los productos, junto a factores económicos, ha inducido a empresas y cultivadores a introducir innovaciones tecnológicas en sus instalaciones, en particular para realizar procedimientos no destructivos de evaluación. Entre éstos se destacan aquellos que utilizan las propiedades ópticas, tales como la visión de máquina o visión por computador [24, 116].

Los procesos de clasificación y estimación de parámetros en los frutos, se convierten en tareas repetitivas para los agricultores y/u operarios, a través de inspección física y visual [26, 28, 47, 76]. Para la cosecha, por ejemplo, en casos extremos los cultivadores esperan a que el fruto sufra desprendimiento natural, o según su experiencia, determinan el momento adecuado para su recolección. Para solucionar problemas como éste y reducir tiempos, es posible diseñar e implementar algoritmos de visión por computador. Dichos algoritmos, permiten correlacionar los procesos naturales asociados a la maduración de los productos agrícolas, con los aspectos visibles en imágenes digitales [70, 112]. Es decir, traducen lo que normalmente se obtendría mediante métodos destructivos a nivel de laboratorio.

En adelante entonces, se presentarán: las herramientas, los métodos, los equipos, la tecnología y la teoría para realizar la medición de las características fisicoquímicas de ciertos productos agrícolas. Posteriormente, se presentan las técnicas de visión de máquina utilizadas en la agricultura y algunas herramientas algorítmicas para cuantificar los parámetros usados por los expertos de manera cualitativa. Por último, las herramientas matemáticas y computacionales comúnmente utilizadas para la correlación estadística de los parámetros.

## 2.1. Medición de propiedades físicoquímicas en productos agrícolas.

La apariencia visual de frutos frescos y vegetales es uno de los primeros elementos determinantes para el consumidor. Ésta permite hacerse una idea de las características físicoquímicas asociadas a la calidad del producto. Frecuentemente la apariencia, suele ser uno de los factores más críticos en todos los eslabones de la cadena de producción (productores, intermediarios y compradores/consumidores). Por esta razón, es necesario poder medir y/o cuantificar a partir de métodos y prácticas, cada una de éstas características que permitan clasificar los productos, acorde a los requerimientos del consumidor.

### 2.1.1. Sólidos solubles.

El contenido de sólidos solubles (SS), es uno de los principales parámetros para evaluar el estado de madurez de los frutos. Éste depende en gran medida de una serie de factores genéticos, ambientales y de cultivo [18, 75, 95]. Los azúcares son los principales sólidos solubles presentes en el zumo de las frutas frescas y, por lo tanto, la estimación de estos puede referirse a la cantidad de azúcar presente en el fruto (una medida de “dulzura”). Los ácidos orgánicos, aminoácidos y diferentes compuestos, contribuyen también a los sólidos solubles, pero en una significativa menor proporción.

Para medir la cantidad de sólidos solubles, es necesario destruir el fruto y utilizar una pequeña cantidad del zumo de la fruta para realizar la medición utilizando un refractómetro. El refractómetro mide el índice de refracción, lo que indica que tan lento pasa la cantidad de un haz de luz a través del zumo de la fruta [27, 76]. El refractómetro posee una escala de medición llamada índice de refracción y otra equivalente al porcentaje de SS, expresado en °Brix, que es posible leer por un operario directamente en el equipo, Figura 2-1. Hoy en día existen algunos dispositivos digitales, que pretenden eliminar los posibles errores de lectura asociados al operador.

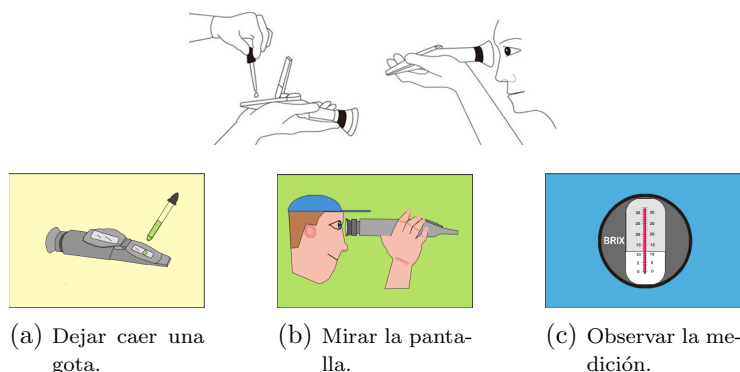


Figura 2-1.: Refractómetro de mano utilizado en la medición del contenido de sólidos solubles [50].

### 2.1.2. Firmeza.

La firmeza es frecuentemente medida utilizando diferentes instrumentos. Generalmente, mediante palpación pueden realizarse algunas estimaciones rápidas que permiten evidenciar grandes diferencias en los productos [75, 32]. Esta característica es un parámetro físico asociado a la calidad, estructura de la pared celular y estado de madurez de los productos agrícolas. Su determinación es fundamental en los procesos de aceptación y/o rechazo de un producto, así como también en los procesos de empaque y transporte [51, 60, 76, 84]. Conforme el tiempo va pasando, la firmeza de un producto agrícola disminuye, causando deshidratación y por tanto el ablandamiento de la pulpa. Este proceso conlleva pérdida de calidad por sobremaduración del producto, debido a tener una mayor susceptibilidad a sufrir daño mecánico.

La forma más común para medir la firmeza en laboratorio es mediante la resistencia a la compresión, utilizando un dispositivo conocido como: penetrómetro, Figura 2-2. Este instrumento utiliza un sistema de perforación para determinar la cantidad de fuerza necesaria para vencer la resistencia del producto, utilizando generalmente como unidad de medida las unidades de fuerza (N).

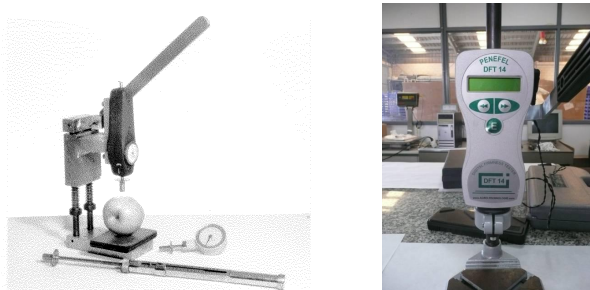


Figura 2-2.: Penetrómetro utilizado para la medición de firmeza [76, 32].

### 2.1.3. Acidez titulable.

Por otro lado, los frutos pueden ser estudiados en términos de sabor y acidez, ya que estos elementos suelen ser importantes en el procesamiento de las frutas. Uno de los métodos más conocidos es la denominada acidez titulable (AT). Ésta puede ser determinada mediante la técnica de titulación de un volumen conocido del zumo de la fruta, con 0.1N de NaOH (Hidróxido de Sodio), hasta llegar a un punto final de  $\text{pH} = 8.2$ , como es mostrado al usar fenolftaleína o sencillamente mediante un medidor de pH [29, 76, 51, 79]. El NaOH es adicionado al zumo hasta alcanzar el cambio esperado en el pH. Los mililitros de NaOH necesarios son usados para calcular la AT, expresado como porcentaje málico, cítrico o ácido tartárico, calculado como la Ecuación 2-1.



$$AT = ml(NaOH) \times N(NaOH) \times FMA \times 100ml(zumo). \quad (2-1)$$

Donde  $ml(NaOH)$  son los mililitros de NaOH necesarios para alcanzar el pH y FMA es el factor miliequivalente dependiente del ácido orgánico predominante en el producto.

#### 2.1.4. Masa.

Tener una cantidad de producto con una masa uniforme puede ser útil en términos de manipulación y transporte, ayudando a satisfacer las preferencias del consumidor. Sin embargo, algunos productos que presentan una forma irregular, no pueden ser fácilmente estimados y/o clasificados en términos de su masa [84, 94, 104].

Debido a ésta problemática, es necesario realizar mediciones de masa en términos de gramos [g] ya sea para un producto o un lote [40, 93]. Este proceso ha presentado desde siempre una serie de limitaciones en términos de tiempo, ejecución y disponibilidad de los elementos para su medición, Figura 2-3. Pero esta medición es de vital importancia, ya que normalmente los productos agrícolas son comercializados en función de su masa pero sin ser clasificados por parte de los productores.



Figura 2-3.: Balanza utilizada para la medición de masa en gramos [32].

#### 2.1.5. Volumen.

Otra de las características físicas difícil de identificar, pero que está ligada al desarrollo del producto, es el volumen [5, 20, 81]. La medición de este producto agrícola posee un grado de dificultad mayor, debido a la necesidad de utilizar algunos principios matemáticos para estimarlo. Si bien es ampliamente conocido el principio de Arquímedes a partir de la cantidad de desplazamiento de agua que genera un cuerpo sobre una superficie líquida; la estimación del volumen por este método no suele ser tan sencilla, Figura 2-4. Es posible calcular el volumen de cierto producto, modelando el mismo en función de una superficie conocida (esfera, elipsoide, cilindro, etc.), y realizar las mediciones necesarias para estimar dicha superficie sobre el producto.



**Figura 2-4.:** Métodos utilizados para la medición del volumen en productos agrícolas [32].

### 2.1.6. Color.

El color es un importante factor que tiene un especial interés en los productos agrícolas. A lo largo de muchos años, han sido utilizadas cartas de color para la clasificación de frutas, verduras y en general una gran variedad de productos agrícolas [8, 28, 71]. Sin embargo, éstas presentan un componente subjetivo y no facilitan la separación visual de algunos detalles presentes por ejemplo, entre variedades de una misma fruta. Para superar esta dificultad, se han desarrollado equipos electrónicos de medición conocidos como colorímetros, que se basan generalmente en la adquisición de las coordenadas acromáticas en el sistema de color CIELab o Lab.

El colorímetro genera un sistema compuesto por tres parámetros denominados  $L^*$ ,  $a^*$  y  $b^*$ , Figura 2-5. Este método ha sido uno de los más utilizados para medir y separar los objetos en términos de sus coordenadas de color. Industrias como la textil, tintas, plásticos, papel y diversos materiales, se basan en él de manera rutinaria en sus procesos [103, 47].



(a) Colorímetro utilizado en la medición de las coordenadas de color [103].



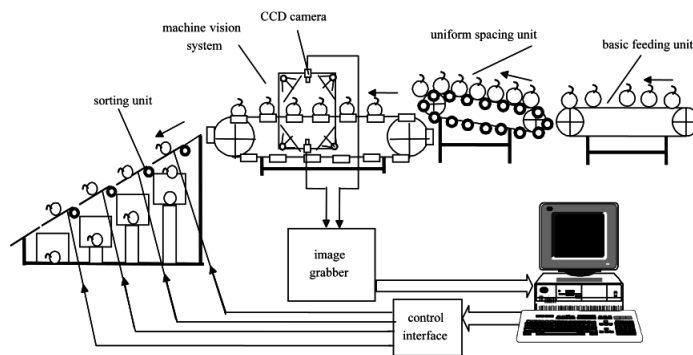
(b) Proceso de medición de color sobre frutos [32].

**Figura 2-5.:** Equipo electrónico (Colorímetro) y ejemplo del proceso de medición de color

En los productos agrícolas, diversos autores han utilizado diferentes combinaciones y operaciones entre éstos elementos, para definir un parámetro aplicable a un conjunto o subconjunto de frutas o vegetales. Uno de los más ampliamente usado se refiere a la relación entre las distintas componentes como  $IC = (a \times 1000)/(L \times b)$  [32], donde  $L^*$ ,  $a^*$  y  $b^*$  son cada una de las coordenadas de color, e  $IC$  es el parámetro ampliamente conocido como índice de color.

## 2.2. Visión de máquina en productos agrícolas.

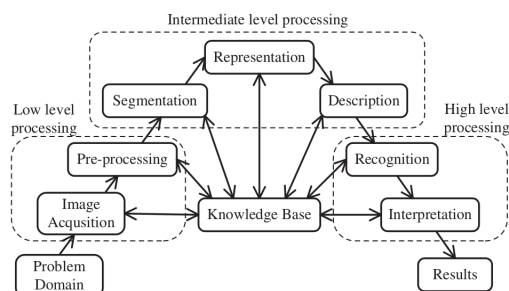
Los sistemas empleados en la medición de las propiedades ópticas, proporcionan una amplia información de las características del objeto de interés, siendo esto útil para la valoración de los productos hortofrutícolas. Esto se realiza por medio de sensores o cámaras, a partir de un sistema de transmisión de datos y centralización de información en computadores, ver Figura 2-6. Sin embargo, para poder realizar una correcta discriminación es necesario un complejo tratamiento de los datos, que presenta diferentes métodos para cada caso; siendo necesario también una periódica calibración de los equipos, que garantice su adecuado mantenimiento y uso.



**Figura 2-6.:** Elementos de un sistema de visión de máquina utilizando imágenes digitales [67].

Los sistemas de visión de máquina constan, generalmente, de tres subsistemas: procesamiento de imágenes, identificación y actuación, Figura 2-7. El procesamiento de nivel bajo, que es el procesamiento básico de la imagen, implica la adquisición de las imágenes y el pre-procesamiento. El procesamiento de nivel intermedio, implica la segmentación del área de interés en las imágenes, la extracción de características, y la representación o descripción de las escenas. Finalmente, el procesamiento de alto nivel, siendo el paso clave del análisis de imagen, donde se pasa de la representación al reconocimiento, interpretación y clasificación [124]. Los equipos ópticos utilizados comúnmente, suelen estar centrados en las zonas del espectro visible (Longitud de onda: 400-800 nm). Este rango de longitudes de onda es fácilmente percibido por el ojo humano de manera directa.

Muchas de las aplicaciones desarrolladas en el sector agrícola, que utilizan la tecnología de visión de máquina, basan su análisis en este tipo de imágenes para tareas como: tele-detección aérea para evaluaciones de los recursos naturales, calidad poscosecha de productos, clasificación y/o automatización de otros procesos. Esto se debe a que los sistemas de visión de máquina no sólo reconocen tamaño, forma, color y textura en los objetos de manera no destructiva, sino que también proporcionan atributos numéricos de los objetos o la escena [24].

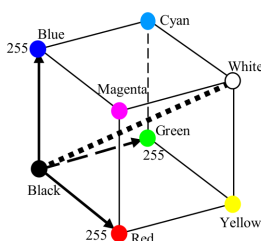


**Figura 2-7.:** Sistema de visión de máquina y los niveles del procesamiento de imágenes [33].

El color por ejemplo, es uno de los atributos característicos de este tipo de imágenes, y depende de la intensidad de luz y su absorción o reflexión en los objetos. En los productos agrícolas el color es uno de los atributos principales que puede determinar la calidad de los productos. Por medio de ésta característica los consumidores realizan una evaluación rápida para determinar su aceptación o rechazo. Por tanto, en el área de productos agrícolas, la medición rápida y objetiva de elementos apreciables en el espectro visible, como el color; permite la implementación de tecnologías que ayuden a salvaguardar mayormente el producto en sus mejores propiedades. Esto cobra relevancia en la medida que el color, apreciable por la simple inspección visual, pueda traducirse en atributos que poseen los productos; los cuales son importantes para los consumidores.

### 2.2.1. Espacio de color RGB.

Las imágenes a color RGB, Figura 2-8, hacen parte de nuestra vida cotidiana. Éstas son usadas diariamente por una proporción muy alta de seres humanos. Sin embargo, mediante los métodos de visión por computador, estas imágenes han cobrado importancia ya que han sido útiles y, comúnmente usadas en muchas industrias, entre ellas la industria de alimentos. Estas imágenes son capturadas mediante diferentes dispositivos: cámaras, celulares, dispositivos móviles; todos ellos con diferentes funcionalidades y resoluciones. A pesar de esto, las imágenes siguen siendo constituidas por la combinación de las intensidades en términos de la reflectancia de las tres componentes principales R (Red o Rojo), G (Green o Verde) y B (Blue o Azul).



**Figura 2-8.:** Modelo de color RGB [113].

Para el tratamiento de las imágenes en términos de color en las industrias, no sólo es posible la obtención de los valores numéricos de cada canal RGB, sino también es posible operarlos como muestra la Figura 2-9, discretizando el color. El valor de intensidad de cada canal antes de la discretización varía de 0 a 255 (Figura 2-9a), por lo tanto sumando los tres planos que componen la imagen en total se tendrían  $256 \times 256 \times 256 = 2^{24} = 16'777'216$  colores diferentes. Luego, utilizando 4 divisiones (Figura 2-9b), se puede representar cada canal de color en una menor categorización, por ejemplo denotando 0, 1, 2, 3 intensidades 0-63, 64-127, 128-191, y 192-255, respectivamente para cada canal. Este paso es denotado como “discretización de color”. Luego de ello, se tienen en total solamente  $4 \times 4 \times 4 = 64$  tipos diferentes de colores [124].



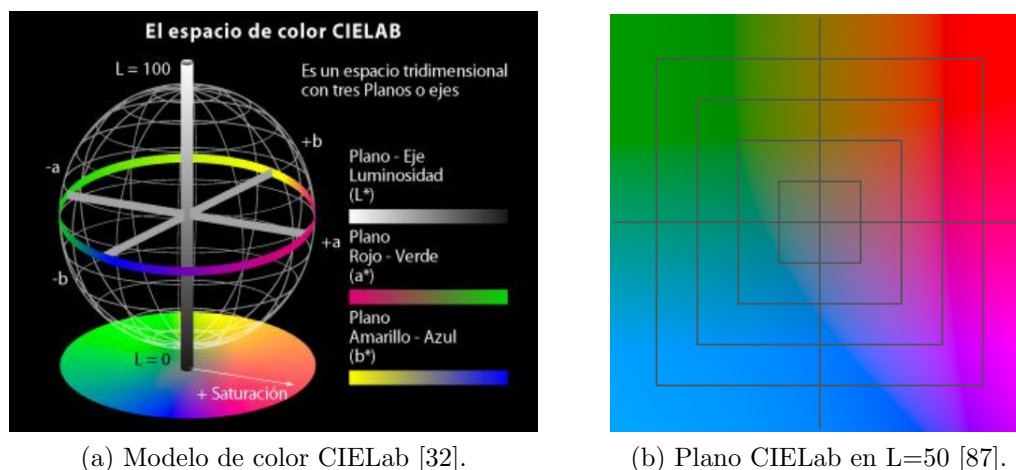
**Figura 2-9.:** Discretización de RGB reduciendo de (a) 16'777.216 niveles de color hacia (b) 64 niveles de color [124].

### 2.2.2. Espacio de color CIELab.

Una transformación de color muy utilizada en el análisis de imágenes es el CIELab. El espacio de color  $L^*a^*b^*$  o CIELab, es una norma internacional para la medición de color, adoptado por la Comisión Internationale d'Eclairage (CIE) en 1976.  $L^*$  es la componente de luminancia en el rango de 0 a 100, y los parámetros  $a^*$  (de verde a rojo) y  $b^*$  (de azul a amarillo), son los dos componentes cromáticos que varían entre -120 y 120 [118].

En este sistema de representación del color, los valores  $L^*$ ,  $a^*$ ,  $b^*$  describen un espacio de color tridimensional uniforme, en el que el valor  $L^*$  corresponde a una escala de oscuro a brillante,  $a^*$  es negativo para el verde y positivo para el rojo y  $b^*$  es negativo para el azul y positivo para el amarillo, Figura 2-10 [103].

Aunque existen diferentes espacios de color, uno de los más utilizados en la medición del color en alimentos (frutas, hortalizas y granos) es el espacio de color  $L^*a^*b^*$ , debido a la distribución uniforme de los colores y, también, a que está muy cerca de la percepción humana del color. Como se mencionó anteriormente, existen en la actualidad soluciones comerciales (colorímetros), para obtener las medidas del espacio de color  $L^*a^*b^*$ , pero estos miden generalmente en pequeñas áreas, lo que resulta en áreas no representativas de unos pocos centímetros cuadrados [66].



**Figura 2-10.:** Modelo de color Lab.

Existen otros espacios de color, algunos de ellos comúnmente conocidos (HSV, YCrCb, HIS, etc) y otros adaptados a partir del análisis deseado para una aplicación específica. A partir de las diferentes necesidades, algunos de ellos son utilizados para la cuantización de aspectos como luminosidad y tonalidad, permitiendo así el nacimiento y uso de nuevos espacios de color, para la representación de dichas variaciones.

### 2.2.3. Constancia de color.

Hace ya algunas décadas se realizaron los primeros experimentos para determinar la constancia del color en imágenes. La idea principal era poder percibir el color sobre un objeto de manera constante a pesar de los cambios en la iluminación que pudieran existir. En esta forma, surgieron modelos para estimar dichos iluminantes y atenuar los problemas de la constancia de color [44]. Uno de los métodos de constancia de color bien conocidos es el de Mundo-gris (Grey-World en inglés), en este se asume que el promedio de la luz reflejada en la escena es acromático. Aunque existen otros métodos, éste es uno de los métodos ampliamente utilizado debido a su bajo costo computacional.

En el análisis de la constancia de color, generalmente está implícita que la fuente de iluminación no se conoce directamente por el observador, Figura 2-11. Esta presunción no es tan natural como podría parecer, así la(s) fuente(s) de luz sean conocidas (por ejemplo: sol, cielo despejado, nublado, incandescente o incluso una lámpara). Sin embargo, a partir de inspeccionar directamente las condiciones de la escena o realizar un análisis matemático es posible inferir la representación de las componentes del espectro de luz que afecta al objeto. El reto entonces, para la constancia de color se basa, en que algunas de esas componentes como la reflectancia espectral de una superficie, o la irradiancia espectral de la iluminación incidente se pueden estimar fácilmente desde el patrón de radiancia espectral reflejada desde la superficie hacia la cámara [44].



**Figura 2-11.:** Tres adquisiciones de la misma escena bajo diferentes condiciones de iluminación [109].

Los valores de la imagen en términos del espacio de color RGB,  $f = (R, G, B)^T$ , para determinada superficie son dependientes de la fuente de luz  $e(\lambda)$ , donde  $\lambda$  es la longitud de onda, la reflectancia de la superficie del objeto  $s(\lambda)$  y las funciones de sensibilidad de la cámara como  $c(\lambda) = (R(\lambda), G(\lambda), B(\lambda))$ , Ecuación 2-2. Sabiendo que  $w$  es el espectro visible, el objetivo de la constancia de color es estimar la fuente de luz  $e(\lambda)$  o su proyección en cada uno de los canales RGB.

$$f = \int_w e(\lambda)s(\lambda)c(\lambda)d\lambda \quad (2-2)$$

En este sentido, es posible utilizar distintos algoritmos existentes de constancia de color para generar un conjunto de los posibles iluminantes en una escena [110]. Dos métodos de constancia de color ampliamente utilizados pueden ser unificados, proporcionando una demostración de que los métodos son en realidad dos instancias de la norma de Minkowski (2-3) en una imagen, donde  $i$  es el contador sobre los  $N$  píxeles  $f_i$  y  $k$  es una constante que se elige de modo que el color iluminante  $c$  tiene unidades de longitud (Shades of Grey). Para  $p = 1$  la estimación del iluminante es igual a la constancia de color derivada de la hipótesis Grey-World, la cual supone que la reflectancia promedio en una escena debe ser gris. Usando  $p = \infty$ , la estimación del iluminante es igual al método max-RGB, el cual asume que la respuesta máxima de los canales separados debe ser igual al color iluminante.

$$\left( \sum_{i=1}^N (f_i(X))^p \right)^{\frac{1}{p}} = kc. \quad (2-3)$$

El Grey-World es uno de los métodos de constancia de color, que utiliza el balance de blancos, suponiendo que la escena, en promedio, posee un gris neutro. Así, los supuestos del Grey-World asumen que se tiene una buena distribución de los colores en la escena. Suponiendo dicha buena distribución de los colores, el promedio de color reflejado es asumido como el color de la luz. Por lo tanto, se puede estimar el color de iluminación proyectada, analizando el color medio y comparándolo con el gris.

El algoritmo de Grey-World produce una estimación de la iluminación mediante el cálculo de la media de cada canal de la imagen. Uno de los métodos de la normalización, es que la media de los tres componentes se utiliza como estimación de iluminación de la imagen. Para normalizar la imagen del canal  $i$ , el valor de píxel se escala por:  $S_1 = \frac{avg}{avg_i}$ , donde  $avg$  es la media del canal y  $avg_i$  es el estimado de la iluminación.

### 2.2.4. Matriz de co-ocurrencia de niveles de gris.

La matriz de co-ocurrencia de niveles de gris (MCNG o GLCM por sus siglas en inglés Gray Level Co-occurrence Matrix); se puede considerar como una estimación de la probabilidad conjunta de encontrar pares de niveles de gris en una imagen o región de la misma [98, 111]. La matriz es en general simétrica. La MCNG es comúnmente implementada con algún grado de invarianza a la rotación. Esto se consigue mediante la combinación de los resultados de un subconjunto de ángulos. Si la MCNG es calculada con simetría, solo los ángulos hasta  $180^\circ$  son necesarios teniendo como opción efectiva ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  y  $135^\circ$ ).

Las texturas que pueden ser analizadas mediante esta técnica, incluyen detalles finos en alta resolución así como estructuras gruesas en baja resolución. Para capturar la mayor parte de la información de la textura, es posible definir una distancia para el análisis entre píxeles, definido como  $d$ , generalmente en el rango de 1 a 4. Dado  $P(i, j, d, \theta)$  una frecuencia normalizada de la ocurrencia de un par  $(i, j)$  de niveles de gris, a una distancia  $d$  y un ángulo  $\theta$ ; teniendo  $N_g$  niveles de gris; es posible definir algunas relaciones de interés, tales como Entropía (Ecuación 2-4), Homogeneidad (Ecuación 2-5), Contraste (Ecuación 2-6), Energía (Ecuación 2-7) y Correlación (Ecuación 2-8) [106].

$$Ent_{d,\theta} = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j, d, \theta) \cdot \log P(i, j, d, \theta). \quad (2-4)$$

$$Hom_{d,\theta} = \sum_{i=1}^{N_g} \frac{P(i, j, d, \theta)}{1 + |i - j|}. \quad (2-5)$$

$$Cont_{d,\theta} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 \cdot P(i, j, d, \theta). \quad (2-6)$$

$$En_{d,\theta} = \sum_{i=1}^{N_g} P(i, j, d, \theta)^2. \quad (2-7)$$

$$Corr_{d,\theta} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i - \mu)(j - \mu) \cdot P(i, j, d, \theta)^2}{\sigma^2}. \quad (2-8)$$

### 2.2.5. Momentos de Hu.

Los descriptores de forma son una poderosa herramienta para las tareas de clasificación. Describir un objeto de interés en términos de sus proporciones, puede llegar a ser útil, ya que esta información obtenida puede ser relacionada con mediciones reales del mismo. De esta manera, características como el ancho, alto, área y perímetro han sido usadas generalmente como descriptores de la forma y el tamaño de los objetos visibles en imágenes. Estas medidas tiene un significado geométrico claro, su comportamiento es bien entendido y se podría decir que pueden ser bien discriminados en términos de la aplicación.



Otras herramientas de análisis de forma, no poseen claramente una propiedad agradable o fácilmente definible, sin embargo han sido usadas extensivamente en algunos casos específicos para medir la forma elíptica [128] o circular de objetos [127]. Estos últimos son conocidos como los momentos invariantes de Hu, llamados así debido a su capacidad de ser constantes a la traslación, escala y rotación de los objetos en una imagen. Los momentos centrales  $\mu_{ij}$  son de por sí invariantes a la traslación. Los momentos invariantes a la escala  $n_{ij}$  pueden construirse en función de los momentos centrales a partir de la Ecuación 2-9, dividiendo a través de una adecuada escala el  $n$ -ésimo momento central.

$$n_{ij} = \frac{\mu_{ij}}{\mu_{00}^{1+\frac{i+j}{2}}}. \quad (2-9)$$

A partir de los momentos centrales invariantes a la traslación  $\mu_{ij}$  y de los momentos invariantes a la escala  $n_{ij}$ , es posible entonces construir los llamados momentos de Hu, invariantes a la traslación, escala y rotación:  $Hu_1$  (Ecuación 2-10),  $Hu_2$  (Ecuación 2-11),  $Hu_3$  (Ecuación 2-12),  $Hu_4$  (Ecuación 2-13),  $Hu_5$  (Ecuación 2-14),  $Hu_6$  (Ecuación 2-15) y  $Hu_7$  (Ecuación 2-16). El primero de ellos  $Hu_1$  es análogo al momento de inercia alrededor del centroide de la imagen, donde las intensidades de los píxeles son análogas a la densidad física. El último de ellos  $Hu_7$  permite distinguir entre imágenes espejo o simétricas en función de un eje, es decir imágenes idénticas. Sin embargo, el resto de momentos invariantes no pueden ser definidos de una manera clara, debido a ser combinacionales lineales de los anteriores, como se ha definido previamente.

$$Hu_1 = n_{20} + n_{02}. \quad (2-10)$$

$$Hu_2 = (n_{20} - n_{02})^2 + 4n_{11}^2. \quad (2-11)$$

$$Hu_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2. \quad (2-12)$$

$$Hu_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2. \quad (2-13)$$

$$Hu_5 = (n_{30} - 3n_{12})^2(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} - n_{03})^2] + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]. \quad (2-14)$$

$$Hu_6 = (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}). \quad (2-15)$$

$$Hu_7 = (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] - (n_{30} + 3n_{12})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]. \quad (2-16)$$

## 2.3. Técnicas matemáticas y computacionales para correlación de parámetros

### 2.3.1. Análisis por componentes principales.

El Análisis por Componentes Principales (ACP o PCA por sus siglas en inglés) es una técnica para la reducción de dimensionalidad capturando, de forma óptima, la varianza máxima de un conjunto de datos llevándolo a un espacio de pocas dimensiones [73, 89, 102, 120]. Los datos normales de ciertos procesos son dispuestos en una matriz  $X^0 \in \mathbb{R}^{n \times m_0}$ , donde  $m_0$  es el número de variables que han sido medidas y  $n$  es el número de muestras. Las variables con varianza nula, con problemas de signo o faltantes son eliminadas, teniendo que  $X^0$  sea  $X^p \in \mathbb{R}^{n \times m}$ , con  $m \leq m_0$ . Las medias de estas  $m$ -variables restantes son contenidas en el vector  $b \in \mathbb{R}^m$ , y las desviaciones estándar de las  $m$ -variables son contenidas en la sumatoria de la Ecuación 2-17.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \cdot & \cdot & \dots & 0 \\ \cdot & \cdot & \dots & 0 \\ \cdot & \cdot & \dots & 0 \\ 0 & 0 & \dots & \sigma_m \end{bmatrix} \in \mathbb{R}^{m \times m}. \quad (2-17)$$

Los datos ( $X$ ), son normalizados con el objetivo de evitar problemas en la reducción de dimensión, teniendo en cuenta el problema de escalamiento. Luego, la matriz de covarianza  $S \in \mathbb{R}^{m \times m}$ , es calculada como:  $S = X^T X / (n - 1)$ . Entonces, es realizada una descomposición espectral de  $S$  así:  $S = V \Lambda V^T$ , donde  $\Lambda \in \mathbb{R}^{m \times m}$  es una matriz diagonal que contiene los valores propios ordenados de manera descendente en función de su magnitud; y  $V \in \mathbb{R}^{m \times m}$  es una matriz que contiene los correspondientes vectores propios [85].

### 2.3.2. Análisis discriminante lineal.

El Análisis Discriminante Lineal (ADL o LDA por sus siglas en inglés Linear Discriminant Analysis) es otra técnica para la reducción de dimensionalidad, capturando aquí la mayor discriminancia del conjunto de datos llevándolo a un espacio de pocas dimensiones [45, 73, 102, 120]. Es definido mediante la maximización de la relación de las matrices de dispersión entre y dentro de cada clase ( $S_b$  Ecuación 2-18 y  $S_w$  Ecuación 2-19), en un futuro espacio lineal; donde cada clase será un grupo de muestras semejantes dentro de todo un espacio muestral.

$$S_b = \sum_{i=1}^c P(\omega_i) (\psi_i - \psi)(\psi_i - \psi)^T. \quad (2-18)$$

$$S_w = \sum_{i=1}^c P(\omega_i) \left[ \frac{1}{n_i} \sum_{\gamma \in \omega_i} (\omega - \psi_i)(\omega - \psi_i)^T \right]. \quad (2-19)$$

Donde  $\omega_i$  denota la  $i$ -ésima clase con  $n_i$  muestras y medida de clase  $\psi_i$ .  $P(\omega_i)$  denota la probabilidad a priori de la clase  $\omega_i$ . El ADL busca un sub-espacio lineal  $W$  de  $c-1$  componentes, en el que las proyecciones de las diferentes clases se separan mejor, tal como se define mediante la maximización de los criterios discriminantes  $J(W)$  Ecuación 2-20.

$$J_w = \frac{|W^T S_b W|}{|W^T S_w W|}. \quad (2-20)$$

Teniendo la restricción de ortonormalidad para  $W$ , el problema puede ser resuelto como un problema general de valores y vectores propios,  $S_b W_i = \lambda_i S_w W_i$ , con  $W_i$  y  $\lambda_i$  como el  $i$ -ésimo valor propio y vector propio de  $S_b$  con respecto a  $S_w$ . La solución de ADL  $W$  contiene todos los  $c-1$  vectores propios valores propios no-cero ( $S_b$  con rango máximo de  $c-1$ ) [45].

### 2.3.3. Regresión Lineal Múltiple.

Uno de los métodos comúnmente utilizados para la correlación de variables es la regresión lineal múltiple (RLM o MLR por sus siglas en inglés Multiple Linear Regression). Este análisis permite establecer las relaciones entre una variable dependiente ( $Y$ ) con cierto grupo de una(s) variable(s) independiente(s) ( $X_1, X_2, X_3, \dots, X_n$ ). A partir de la aplicación de este método matemático, es posible obtener una ecuación o modelo que describe la relación entre la variable dependiente y la(s) variables independientes [107]. Este método puede ser definido como:  $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m + \epsilon$ ; en donde  $Y$  es la variable a predecir, el vector de coeficientes  $\vec{\alpha}$ 's contiene cada uno de los parámetros  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$  asociados a cada una de las variables independientes y  $\epsilon$  el error de la correlación realizada.

Para realizar la RLM es necesario un sencillo pero demandante manejo matemático de las variables dependiente e independientes [86]. Cierta variable dependiente puede ser definida mediante un vector  $\vec{Y}$ ; los valores asociados a cada una de las variables independientes ( $X$ ) es posible organizarlos en una matriz  $[X]$ , los  $\alpha$  coeficientes definidos en el vector  $\vec{\alpha}$  y el error  $\epsilon$  puede ser definido en términos de la variable estimada  $Y_{estimada}$  y la variable dependiente  $Y_{medida}$ , lo anterior se puede evidenciar en la Ecuación 2-21.

$$\vec{Y} = [X]\vec{\alpha} + \epsilon. \quad (2-21)$$

Es posible definir las relaciones existentes entre la variable  $Y_{medida}$  a predecir y cada una de las  $m$ -variables independientes  $X_m$ ; donde  $\vec{\alpha}$  será el objetivo del cálculo matemático, con el fin de obtener el resultado de los coeficientes  $\alpha_0, \alpha_1, \dots, \alpha_m$  asociados a cada una de las  $m$ -variables independientes con que se disponga [52]. La matriz  $[X]$  contendrá los datos asociados a cada una de las  $X_m$  variables independientes, con la primera columna compuesta por un vector de 1's, por medio del cual es posible calcular el parámetro  $\alpha_0$  llamado el término independiente de la regresión; y  $\epsilon$  como el error estándar de la misma, mediante la Ecuación 2-22.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdot \\ \cdot \\ \cdot \\ y_{i-1} \\ y_i \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,m-1} & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,m-1} & x_{2,m} \\ 1 & x_{3,1} & x_{3,2} & \dots & x_{3,m-1} & x_{3,m} \\ 1 & x_{4,1} & x_{4,2} & \dots & x_{4,m-1} & x_{4,m} \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 1 & x_{i-1,1} & x_{i-1,2} & \dots & x_{i-1,m-1} & x_{i-1,m} \\ 1 & x_{i,1} & x_{i,2} & \dots & x_{i,m-1} & x_{i,m} \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{m-1} \\ \alpha_m \end{bmatrix} + \epsilon. \quad (2-22)$$

Definir el modelo de RLM para un problema específico se basa entonces en el cálculo matemático para determinar los  $\alpha$  coeficientes. Para tal fin, es posible calcular las matrices  $S$  y  $T$ , Ecuación 2-23; donde  $S = X^T X$  y  $T$  es la interacción entre el valor real medido de la variable de interés, por cada una de las  $m$ -variables, respectivamente, es decir, el primer término del vector  $T$  será la sumatoria de todos los  $m$ -ésimos valores de la variable dependiente  $Y$  a raíz del vector de 1's en la matriz  $[X]$ ; el segundo término estará compuesto por la sumatoria de multiplicar cada uno de los  $m$ -ésimos valores de la variable independiente  $Y$  por el valor  $i$ -ésimo de la  $m$ -ésima variable independiente  $X$ , Ecuación 2-24.

$$\vec{\alpha} = S^{-1}T. \quad (2-23)$$

$$T = \begin{bmatrix} \sum_{j=1}^i Y_j \\ \sum_{j=1}^i Y_j \cdot X_{1,j} \\ \sum_{j=1}^i Y_j \cdot X_{2,j} \\ \sum_{j=1}^i Y_j \cdot X_{3,j} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{j=1}^i Y_j \cdot X_{m-1,j} \\ \sum_{j=1}^i Y_j \cdot X_{m,j} \end{bmatrix}. \quad (2-24)$$

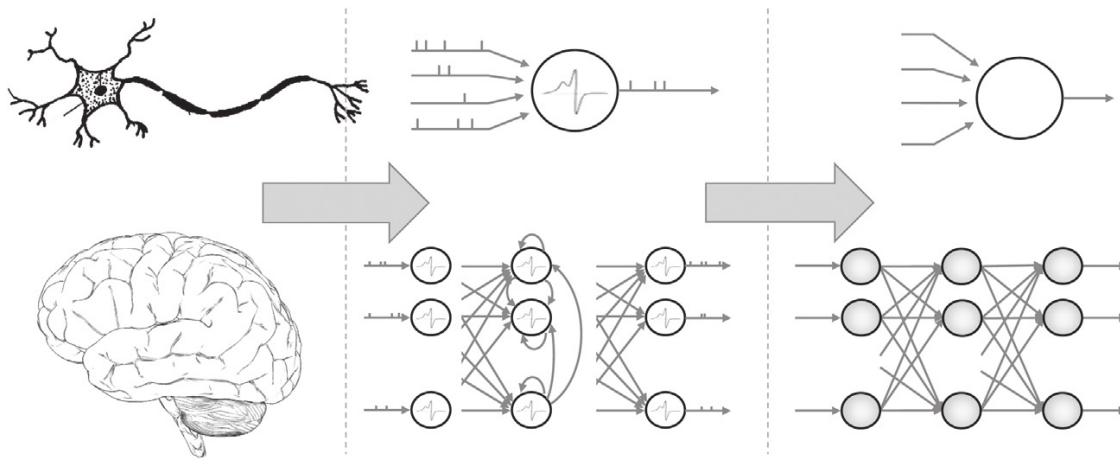
El error estándar de la regresión  $\epsilon$ , es posible medirlo mediante la Ecuación 2-25, a partir de la relación entre el ajuste obtenido a partir de los coeficientes  $\alpha$  calculados y los valores reales esperados. Donde  $N$  es el número de mediciones de la variable dependiente  $Y$  utilizados para el análisis;  $Y_i$  los valores reales y  $\hat{Y}_i$  los valores estimados mediante la regresión, en función de las  $m$ -variables utilizadas para obtener el modelo de RLM.

$$\epsilon = \sqrt{\frac{1}{N - m - 1} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}. \quad (2-25)$$

### 2.3.4. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA o ANN por sus siglas en inglés Artificial Neural Networks), han sido utilizadas como una herramienta de modelado poderosa en comparación con los métodos estadísticos o numéricos como la RLM. Por esta razón, las RNA han sido utilizadas en muchas aplicaciones de ingeniería para predecir, optimizar o clasificar determinadas variables [74, 98].

Éstas tiene una estructura interconectada emulando el proceso realizado por las neuronas en el cerebro humano, utilizando determinado número de las mismas, dispuestas en diferentes capas de la red, Figura 2-12 [107]. Las redes están compuestas básicamente por una serie de neuronas de entrada (información conocida o variables independientes), neuronas de salida (variable(s) de interés a predecir) y una serie de neuronas dispuestas en una o más capas ocultas en medio de la entrada y la salida.



**Figura 2-12.:** Transición desde el proceso biológico a las redes neuronales artificiales a través de la simplificación del modelo de la neurobiología y la neurociencia computacional [5].

Las neuronas dispuestas en la red neuronal, se encuentran generalmente interconectadas entre capas y se ajustan en función de factores de ponderación ( $W_{ij}$ ). Una neurona ( $j$ ) en cierta capa, recibe información ( $X_i$ ) a partir de la totalidad de las neuronas de la capa inmediatamente anterior. La información recolectada es resumida como ( $net_j$ , ver Ecuación 2-26), ponderada por los pesos de cada una de las conexiones y un término de bias asociado a cada una de las capas ( $\theta_j$ ), [62, 107]. Finalmente estos pesos transfieren su valor de salida ( $Y_j$ , ver Ecuación 2-27), a través de la aplicación de cierta función matemática ( $f(net)$ ), a todas las neuronas de la siguiente capa.

$$net_j = \sum_{i=1}^n X_i W_{ij} - \theta_j. \quad (2-26)$$

$$Y_j = f(net_j) = \frac{1}{1 + e^{-net_j}}. \quad (2-27)$$

La definición de cada uno de los pesos en la red neuronal artificial, es realizado a partir de un algoritmo conocido comúnmente como back-propagation [124]. La esencia de este algoritmo es básicamente establecer una óptima relación entre los datos de entrada y salida, ajustando los pesos en las neuronas a partir de lo aprendido en una iteración anterior, con el fin de poder realizar una regla de generalización. Lo anterior se puede ver claramente en la curva de aprendizaje (Figura 2-13), en la cuál a medida que el número de repeticiones aumenta, el error cuadrático medio disminuye. Sin embargo es necesario evaluar el comportamiento tanto en el conjunto de entrenamiento, como en un nuevo conjunto de datos de validación a fin de obtener el número óptimo de iteraciones, en el cuál la curva para el conjunto de validación encuentra un mínimo en función del error cuadrático medio. Este proceso es conocido como validación cruzada [42].

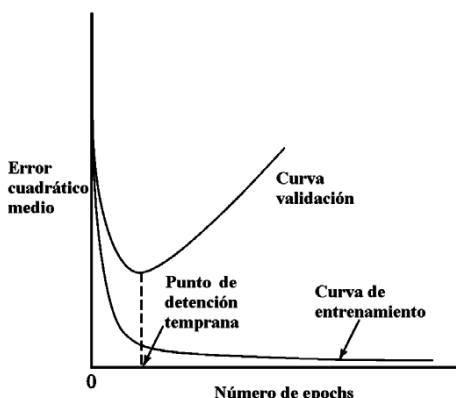


Figura 2-13.: Regla de la detención temprana basada en la validación cruzada de la curva de aprendizaje [42].

## Resumen del capítulo

En este capítulo se presenta la fundamentación teórica referente a los aspectos principales utilizados en la metodología: medición de características fisicoquímicas, procesamiento de imágenes y técnicas de correlación. Algunas de las características fisicoquímicas más utilizadas y comunes son: el color, la masa, el volumen, los sólidos solubles totales, la acidez, el pH y la firmeza. Para su medición suele requerirse de la destrucción del fruto y uso de equipos especializados en laboratorio. Utilizando técnicas de visión de máquina es posible estimar dichas propiedades mediante imágenes, eliminando estos aspectos. Mediante herramientas de procesamiento de imágenes es posible atenuar aspectos como la iluminación, medir las componentes de color y segmentar la región de interés. Aplicando técnicas como RLM, ACP o ADL es posible analizar estadísticamente el descriptor de características extraído de las imágenes mediante relaciones lineales, de variabilidad o discriminancia de clases. Por último mediante RNA's se puede estimar las características en función del descriptor resultante de aplicar alguno de los métodos anteriores. A continuación se presenta la metodología propuesta para la identificación de características fisicoquímicas mediante técnicas de visión de máquina.

### **3. Metodología para la estimación de características fisicoquímicas en productos agrícolas a partir del análisis de imágenes digitales.**

Establecer una metodología que permita identificar las características fisicoquímicas en productos agrícolas por medio de atributos obtenidos en imágenes es útil, ya que evita el uso de métodos destructivos y optimiza tiempos en los procesos de cosecha, poscosecha y logística [4, 13, 24, 33, 56, 90, 96, 100, 115]. Para tal fin, se evaluaron diferentes métodos de análisis para el procesamiento de imágenes y correlación de los parámetros en los productos agrícolas (variables físicas y/o químicas), con las características extraídas de las imágenes.

Una revisión bibliográfica evidencia que determinar y/o estimar las propiedades fisicoquímicas en los productos agrícolas es de gran utilidad tanto para productores como consumidores [34, 41, 68, 58, 84, 94, 95, 96, 126]. La apariencia física, por ejemplo, es un elemento predominante en el consumidor, ya que a raíz de sus experiencias previas le es posible “concluir” de manera intuitiva, algunas de las propiedades intrínsecas del fruto tales como masa, volumen, dulzura, acidez, etc. Los productores también se apoyan, regularmente, en estos criterios subjetivos mediante un razonamiento similar, para determinar los tiempos de cosecha o realizar clasificaciones en la poscosecha de los productos agrícolas.

Por lo anterior, previo a realizar un análisis experimental con un caso de estudio y una característica específica medida, se estableció una metodología general, que permitiera investigar más a fondo en algunos de los métodos de visión por computador existentes; y así definir métodos específicos para cada uno de los pasos de la metodología. De esta manera, fue posible identificar un sistema que permitiera obtener los mejores resultados de correlación, es decir un sistema automático de extracción de características en productos agrícolas por medio del análisis de imágenes digitales.

Con este fin, la visión de máquina puede ayudar al desarrollo de un sistema que permita realizar de manera análoga y sistemática la estimación y/o predicción de características fisicoquímicas. En dicho sistema, los parámetros visibles en imágenes serán normalizados, estandarizados y correlacionados con mediciones reales de la variable de interés en el producto agrícola. Así, se estableció una metodología general para realizar dicha correlación y diseñar un sistema computacional automático para estimar características fisicoquímicas en productos agrícolas, el cual está compuesto por los siguientes cuatro pasos o etapas fundamentales:

1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.
2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.
3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.
4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.

A continuación se presenta en detalle cada uno de los pasos enumerados.

### **3.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.**

Los productos agrícolas frescos son presentados al consumidor en lotes. Se espera que esta etapa se caracterice por ser aquella en la que se alcancen los valores óptimos en términos de madurez y calidad, asociados generalmente a parámetros fisicoquímicos. Existe una gran variedad de éstos parámetros, algunos de ellos resumidos en la Tabla **3-1**, que utilizan y definen diversas propiedades del producto. Sin embargo, tanto para el productor, distribuidor y consumidor es complicado determinar con exactitud y precisión el estado real de los frutos. De allí que las características físicas sean las de mayor prelación en el momento de elegir un producto para su consumo, evaluando aspectos como: color, textura, tamaño y forma; en general lo que se podría denominar: apariencia.

Los frutos, en su mayoría, requieren de cuidados especiales en la manipulación posterior a la cosecha, para garantizar la conservación de sus propiedades al momento de llegar al consumidor. Para esto, es necesario determinar adecuadamente parámetros como masa y volumen con el objetivo de empacar, almacenar y transportar adecuadamente cada lote de frutos, agrupándolos en función de características similares.



**Tabla 3-1.:** Propiedades fisicoquímicas en productos agrícolas.

<b>Propiedades Químicas</b>	<b>Propiedades Físicas</b>
Humedad	<u>Color</u>
Cenizas	Área
Proteínas	Superficie de la cáscara
Grasa cruda	<u>Longitud</u>
Fibra cruda	<u>Grosor</u>
Carbohidratos totales	<u>Ancho</u>
Fósforo	Peso
Calcio	<u>Volumen</u>
Magnesio	Densidad
Sodio	Diámetro geométrico promedio
Potasio	Esfericidad
Hierro	Porosidad
Cobre	Superficie
Manganeso	Epicarpio
Cobalto	Mesocarpio
Cinc	Semillas
Molibdeno	<u>Masa</u>
<u>Sólidos solubles</u>	
Sacarosa	
Glucosa	
Fructosa	
Vitamina C	
Fenoles totales	
Flavonoides totales	
Antioxidantes	
Vitamina A	
Vitamina B	
Vitamina C	
Vitamina E	
Tocoferoles	
Selenio	
<u>pH</u>	
<u>Acidez titulable</u>	

A pesar de esto, realizar la medición de sólo estos parámetros, puede llegar a ser tedioso, teniendo en cuenta la cantidad de producto generado sólo en una pequeña cosecha. Adicionalmente, estas propiedades físicas pueden ser usadas también para garantizar el tiempo óptimo de la cosecha, aprovechando por ejemplo: el momento en el cuál el producto posea mayor masa o volumen [58].

La identificación de características en productos agrícolas suele variar de acuerdo al criterio de la persona que lo realiza y el lugar donde se lleva a cabo, a raíz de esto terminan siendo ambiguos e imprecisos. Sin embargo, estos criterios subjetivos pueden homogeneizarse a partir de su correlación con características fisicoquímicas. Por ejemplo: para algunos frutos cítricos la relación de color R/G, su forma y textura, así como el peso, han ayudado a predecir el contenido de azúcar o pH [62]. Adicionalmente, algunos de éstos parámetros extraídos desde imágenes a productos agrícolas, pueden componer un espacio de características para tareas clasificación mediante redes neuronales artificiales o máquinas de vectores de soporte, con una exactitud superior al 80 % [122].

Así, los criterios visibles, entre ellos las escalas de color, suelen ser una poderosa herramienta utilizada en la agricultura para la estimación de dichas características y, finalmente, el estado de madurez o calidad de los frutos. Entre éstos últimos, el estado de madurez de un fruto es entonces un parámetro importante para determinar el tiempo óptimo de la cosecha y su tiempo de vida en la etapa de la poscosecha. Normalmente, para determinar el grado de madurez, un experto humano asocia algunos aspectos visuales presentes en el fruto, con colores definidos previamente en escalas estandarizadas. Algunos métodos de visión de máquina permiten generar dichas escalas para seguir adecuadamente la evolución del color en el proceso de maduración de un fruto, proporcionando una herramienta en la definición de las diferentes etapas (precosecha, cosecha y poscosecha) [8].

Es posible entonces, definir una serie de características en términos de color, textura, tamaño y forma de un fruto con objetivos de estimación de los parámetros tanto físicos como químicos más representativos, subrayados en la Tabla **3-1**. Como se mencionó anteriormente, dos de las características asociadas a la madurez o calidad de los productos agrícolas pueden ser: la masa y el volumen. De allí que su estimación se haya convertido en un objetivo de investigación para diferentes productos agrícolas [20, 37, 65, 81, 85, 92, 93, 99, 104, 121].

Por su parte, el análisis químico aunque es muy rico y suele tener información valiosa respecto al producto, no necesariamente puede llegar a ser fácil su medición para los eslabones de la cadena (productor, intermediario y consumidor). Algunos de los más ampliamente usados son los sólidos solubles totales y la acidez titulable relacionada también con el pH del producto. Los sólidos solubles pueden ser medidos fácilmente mediante un refráctometro. La acidez por su parte, suele necesitar de un equipo más complejo en laboratorio pero, es comúnmente usado ya que es determinante para valorar procesos de maduración y sobremaduración de los frutos. Sin embargo los dos métodos requieren de la destrucción del producto.

A raíz de lo anterior, las características fisicoquímicas de mayor representación suelen ser cuatro. Dos de ellas son físicas: masa, volumen y las otras dos son químicas: sólidos solubles y acidez titulable. A continuación se analiza lo que sucede al estimar estas características mediante el procesamiento de imágenes digitales; utilizando los pasos 2, 3 y 4 de la metodología descrita anteriormente.

### 3.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

Las aplicaciones de visión por computador para la evaluación de productos agrícolas, han centrado principalmente la investigación en la clasificación y predicción del estado de los frutos. Sin embargo, la segmentación y selección más precisa de acuerdo a características específicas del fruto, ha generado nuevos campos de estudio [19].

De esta manera la implementación, uso y caracterización de modelos específicos bajo parámetros estandarizados, hace que cada nuevo desarrollo sea único y particular, dependiendo de las propiedades de cada fruto. Lo anterior, debido a que las frutas, hortalizas y granos suelen presentarse al consumidor final en lotes, donde la homogeneidad y apariencia de los mismos tienen un efecto radicalmente significativo para el consumidor. Por esta razón, en la presentación de los productos agrícolas al consumidor es relevante la limpieza y clasificación del producto.

Otras investigaciones han utilizado sistemas inteligentes, automáticos e integrados, para realizar tareas de inspección. La visión artificial entonces, de la mano de sistemas robóticos, manejados con precisión, pueden coayudar a la identificación de propiedades físicas: tamaño, color y determinación de ciertas características del producto tales como detección de defectos externos [20].

Generalmente, esto se realiza bajo ambientes controlados, donde el fruto es ubicado de manera independiente a su entorno, para simplificar el análisis de la imagen, específicamente en el proceso de segmentación, Figura 3-1. Todo esto ligado al estudio del fruto desde la parte agrícola, utilizando imágenes en el espectro visible, puede ayudar a la correlación de variables utilizando sus propiedades visibles, una de ellas como es evidente, el color del fruto.

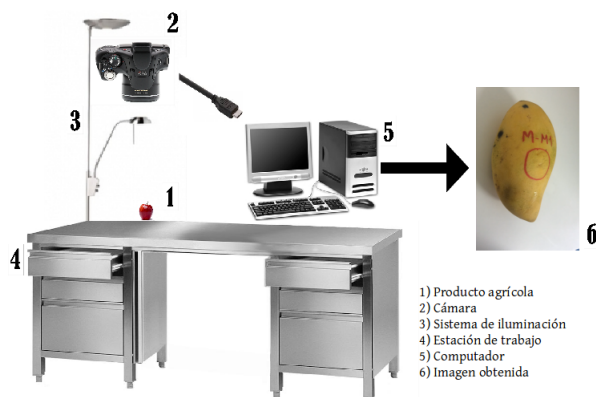
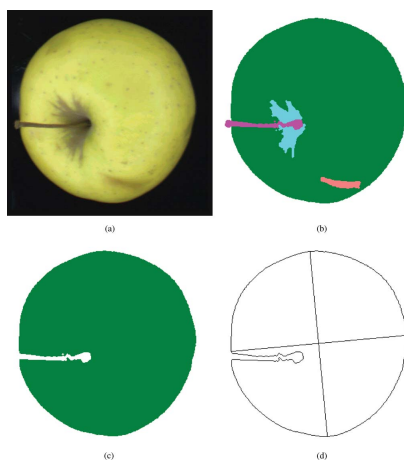


Figura 3-1.: Tipos de imágenes de productos agrícolas obtenidas.

El procesamiento y análisis de las imágenes puede ser considerado como el núcleo de la visión por computador, en donde se pueden utilizar algoritmos y métodos disponibles para llevar a cabo la medición de parámetros y una clasificación específica, ver Figura 3-2. Por ejemplo, algunos algoritmos como umbralización adaptativa y análisis de textura para el objeto de interés, pueden otorgar información útil para la estimación de la masa [70].



**Figura 3-2.:** (a) Imagen original capturada por la cámara; (b) Imagen segmentada que muestra algunas regiones como la piel, el tallo y las regiones de daño; (c) Región que contiene todas las clases, excepto el vástago y el fondo, que se utiliza para calcular el tamaño; y (d) Imagen que muestra cómo se calcula el tamaño [13].

Para el desarrollo de un sistema de visión de máquina, es entonces necesario utilizar como insumo un banco de imágenes. En el caso de los productos agrícolas, obtener las imágenes en campo puede llegar a ser complicado debido a la ubicación del fruto en los árboles o el cultivo. Por tal motivo es recomendable realizar la captura de las imágenes luego de la cosecha utilizando un ambiente controlado en laboratorio. Algunos de los aspectos críticos a controlar suelen ser la distancia del fruto a la cámara y la cantidad de luz en el entorno.

Es aconsejable también, que la cantidad de imágenes sea lo suficientemente grande con el objetivo de poder generalizar el proceso realizado. Algunos autores que han trabajado aspectos de color suelen utilizar al menos 9 imágenes por producto para hacer algún tipo de generalización [72], por lo cual es recomendable que el banco de imágenes por producto sea mínimo este valor y lo más grande posible. A mayor cantidad de información el sistema de visión de máquina podrá realizar mejores estimaciones de las características fisicoquímicas en nuevos frutos. En adelante se presentarán algunas técnicas útiles para el procesamiento de las imágenes y la extracción de características en imágenes con el fin de estimar algunas características de interés como: masa, volumen, sólidos solubles y acidez.

### 3.2.1. Constancia de color.

La luz reflejada por un objeto que entra al ojo o a una cámara, es un producto de las propiedades de reflectancia y el espectro iluminante [43]. La constancia de color es una habilidad atribuida al sistema de visión humano, que se define como la habilidad de percibir los colores de los objetos sin importar el color de la fuente de luz. Las diferencias en la iluminación causan medidas sesgadas del color del objeto hacia el color de la fuente de luz [69], por lo que la tarea de la constancia de color es determinar la reflectancia del espectro iluminante, permitiendo reconocer los colores de los objetos independiente del color de la iluminación.

Las fuentes de luz muchas veces introducen matices de color en las imágenes digitales capturadas. Para resolver este problema, un método estándar es estimar el color de la luz que prevalece y luego, en la segunda etapa, retirarlo. Una vez que se obtiene el color de la luz en los canales individuales de cada píxel de color, se puede normalizar por medio de un factor de escala. Dos de las técnicas más utilizadas para estimar el color de la luz son los algoritmos Gray-World y Max-RGB. Estos dos métodos funcionan bien en la práctica si el color medio en la escena es de color gris o el máximo es de color blanco.

---

#### Algoritmo 1 Constancia de color: Gray-World

---

**Entrada:**  $Imagen_{Original}[w \times h]$

**Salida:** Imagen  $Constancia_{Color} \leftarrow Imagen [w \times h]$ .

- 1: Dividir ( $Imagen_{Original}$ ) en (Plano R, Plano G, Plano B)
  - 2: **Para**  $i = 0$  hasta  $w$  **hacer**
  - 3:   **Para**  $j = 0$  hasta  $h$  **hacer**
  - 4:      $R \leftarrow Plano\ R(i,j); G \leftarrow Plano\ G(i,j); B \leftarrow Plano\ B(i,j)$
  - 5:      $Suma_R += R; Suma_G += G; Suma_B += B; Suma_{RGB} += R+G+B$
  - 6:   **Fin Para**
  - 7: **Fin Para**
  - 8:  $AVG_R \leftarrow Suma_R/(w \times h); AVG_G \leftarrow Suma_G/(w \times h); AVG_B \leftarrow Suma_B/(w \times h)$
  - 9:  $AVG \leftarrow Suma\ RGB/(w \times h)$
  - 10:  $S_R \leftarrow AVG/AVG_R; S_G \leftarrow AVG/AVG_G; S_B \leftarrow AVG/AVG_B$
  - 11: **Para**  $i = 0$  hasta  $w$  **hacer**
  - 12:   **Para**  $j = 0$  hasta  $h$  **hacer**
  - 13:      $R_{nuevo} \leftarrow R \times S_R; G_{nuevo} \leftarrow G \times S_G; B_{nuevo} \leftarrow B \times S_B$
  - 14:     Imagen  $Constancia_{Color}(i,j) = (R_{nuevo}, G_{nuevo}, B_{nuevo})$
  - 15:   **Fin Para**
  - 16: **Fin Para**
  - 17: **Devolver** Imagen  $Constancia_{Color}$
-

La constancia de color analizada de manera computacional es relevante para muchas tareas de visión por computador, como el reconocimiento de objetos, rastreo y vigilancia [9], por lo que en la extracción de características fisicoquímicas en productos agrícolas, no es la excepción. Típicamente, los métodos computacionales para estimar el color de la fuente de luz están basados en solo señales de bajo nivel de imagen. Los valores de cada píxel son usados para estimar los iluminantes y establecer métodos sencillos denotados anteriormente como por ejemplo el Gray-World, pero existen otros más elaborados como el Max-RGB o la estimación de color por correlación o Gamut mapping.

**Tabla 3-2.:** Constancia de color a partir de diferentes iluminaciones sobre un producto agrícola.

Exterior	Iluminación controlada	Sin iluminación	Lab/RGB
			L = 60.76 ± 7.57 a = 17.26 ± 4.05 b = 19.10 ± 5.40 R = 194.73 ± 30.31 G = 130.76 ± 16.31 B = 105.45 ± 11.88
			L = 57.03 ± 3.62 a = 18.22 ± 0.76 b = 20.33 ± 0.96 R = 186.39 ± 12.15 G = 119.83 ± 8.41 B = 93.24 ± 6.73

Al implementar el Algoritmo 1, sobre un producto agrícola como el tomate bajo diferentes condiciones de iluminación, se pueden ver los efectos de la constancia de color utilizando la técnica Gray-World. En la última columna de la Tabla 3-2 es posible observar de manera cuantitativa dichos efectos sobre las imágenes antes y después de aplicar el algoritmo. Como se observa en cada una de las componentes L, a, b, R, G y B la desviación estándar disminuye luego de realizar la constancia de color. Lo anterior es un indicativo de la atenuación de los efectos iluminantes que se obtienen a partir implementar este tipo de técnicas.

### 3.2.2. Segmentación de Regiones.

Una de las primeras tareas en el procesamiento de imágenes es aislar la región de interés en la escena. Esto se realiza con el objetivo de utilizar solo la información relevante del producto en la imagen, y no agregar ruido al análisis. Para realizar esta tarea en imágenes de productos agrícolas, es posible utilizar técnicas de segmentación de regiones tales como OHTA. En comparación con los tradicionales espacios de color HIS y HSV, la relación de conversión de la técnica OHTA y el espacio de color RGB es lineal. OHTA es definido mediante expresiones definidas a partir del tratamiento de los valores píxel a píxel obtenidos mediante el espacio de color RGB, a partir de las Ecuaciones 3-1, 3-2 y 3-3. Donde R, G y B son los tres colores primarios del espacio de color RGB e  $I_1$ ,  $I_2$  e  $I_3$  la definición de OHTA.

$$I_1 = (R + G + B)/3. \quad (3-1)$$

$$I_2 = (R - B)/2. \quad (3-2)$$

$$I_3 = (2G - R - B)/4. \quad (3-3)$$

Sin embargo, no toda la información irrelevante para el análisis de la imagen es eliminada a partir de este conjunto de ecuaciones. Debido a esto, es posible determinar una definición alternativa para el conjunto de expresiones de la técnica de segmentación de regiones OHTA, mediante las Ecuaciones 3-4, 3-5 y 3-6. Los dos conjuntos de expresiones tienen un significado similar y pueden ser utilizadas acorde al tipo de elementos que se encuentran en la imagen acompañando al objeto de interés, en este caso el fruto o producto agrícola.

$$I'_1 = R - G. \quad (3-4)$$

$$I'_2 = R - B. \quad (3-5)$$

$$I'_3 = (2G - R - B)/2. \quad (3-6)$$

Usando éstas expresiones, una imagen a color puede ser convertida desde el espacio RGB hacia OHTA. De acuerdo a las características del fruto en la imagen, las características  $I_1$ ,  $I'_1$ ,  $I_3$  e  $I'_3$  no se utilizan comúnmente para extraer frutos en una imagen que posee un fondo con bastante paisaje o aspectos de la agricultura en el fondo. Por otra parte las características  $I_2$  e  $I'_2$  pueden ser usadas para extraer los frutos en un entorno rodeado por otros elementos [115]. Para verificar lo anterior, se ha aplicado cada una de las expresiones sobre imágenes de diferentes productos agrícolas bajo un ambiente controlado en laboratorio.

---

### Algoritmo 2 Segmentación de Regiones OHTA: Transformación $I_2$

---

**Entrada:** Imagen  $Constancia_{Color}$










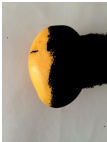























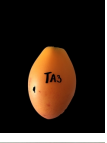








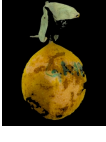






**Salida:** Imagen  $I_2 \leftarrow$  Imagen  $[w*h]$ .

- 1: Dividir Imagen  $Constancia_{Color}$  en: Plano R, Plano G, Plano B
  - 2: **Para**  $i = 0$  hasta  $w$  **hacer**
  - 3:   **Para**  $j = 0$  hasta  $h$  **hacer**
  - 4:     R  $\leftarrow$  Plano R (i,j); G  $\leftarrow$  Plano G (i,j); B  $\leftarrow$  Plano B (i,j)
  - 5:      $I_2 \leftarrow (R-B)/2$
  - 6:     Imagen  $I_2(i,j) \leftarrow I_2$
  - 7:   **Fin Para**
  - 8: **Fin Para**
  - 9: **Devolver** Imagen  $I_2$
-

Si bien, las condiciones de campo y laboratorio son diferentes; el interés particular es poder identificar las características fisicoquímicas de los productos agrícolas. Por tal motivo, los frutos aquí presentados, han sido cosechados y capturado su imagen en laboratorio. Lo anterior ha permitido concluir que en un ambiente controlado, la transformación de la imagen en términos de las expresiones OHTA tiene un comportamiento similar.

Aplicando el Algoritmo 2 se puede obtener la transformación  $I_2$  de la técnica de segmentación de regiones OHTA. De manera similar pueden obtenerse las demás transformaciones a partir de las interacciones entre los valores R, G y B de cada píxel definidas en las ecuaciones de la técnica OHTA.

**Tabla 3-3.:** Visibilidad de la Región de Interés(VRI) al aplicar la técnica de segmentación OHTA en diferentes productos agrícolas.

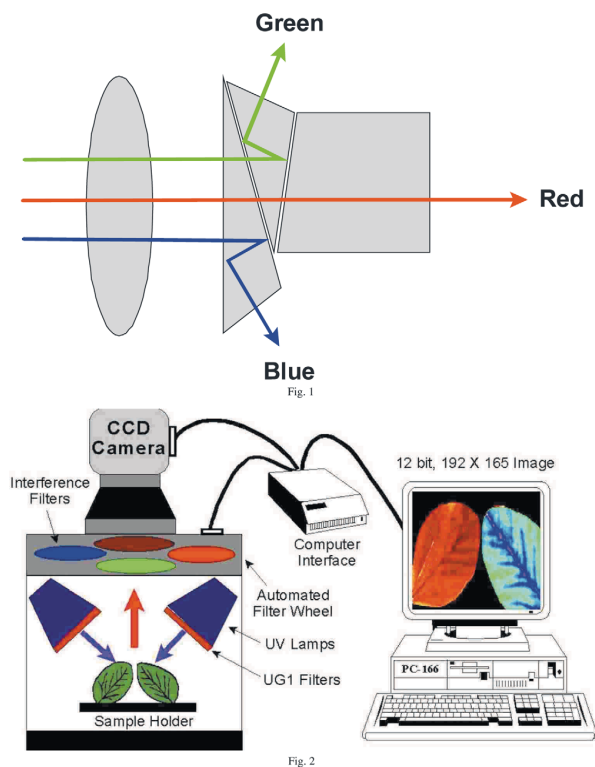
OHTA/Fruto	Granadilla	Guayaba	Mango Común	Mango Nam doc Mai	Maracuyá	Tomate de Árbol	Tomate
Original							
$I_1$ VRI=20 %							
$I'_1$ VRI=75 %							
$I_2$ VRI=95 %							
$I'_2$ VRI=95 %							
$I_3$ VRI=40 %							
$I'_3$ VRI=90 %							



En particular esta transformación  $I_2$ , como se aprecia en la Tabla. 3-3 puede resultar bastante útil para el aislamiento y segmentación del producto agrícola. Comparando en cada una de las transformaciones la visibilidad de la región de interés (VRI), es decir el producto agrícola luego de la segmentación, se ratifica que porcentualmente la cantidad de producto visible en esta transformación es mayor, cercana al 95 %. Si bien, la transformación  $I'_3$  posee también un VRI alto, debido al tratamiento de la información RGB de la imagen se sufren variaciones en la tonalidad del objeto de interés.

### 3.2.3. Medición de componentes de color.

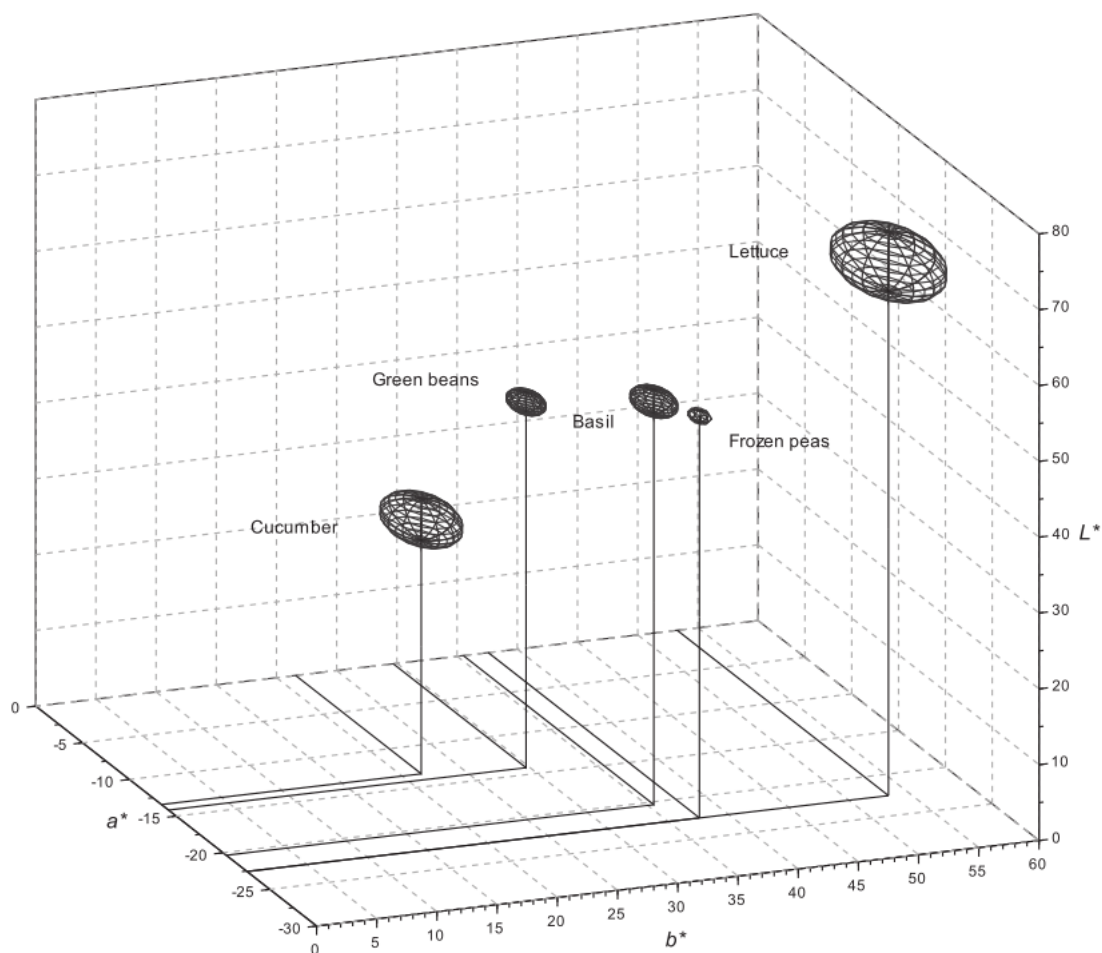
Luego de garantizar la constancia de color, realizar el procesamiento de las imágenes y obtener la región de interés, es posible parametrizar algunos atributos de manera análoga al proceso realizado por expertos en términos de: color, textura, tamaño y forma. Por ejemplo, la identificación del color en los frutos por medio de visión de máquina ha sido una rama de gran interés. Esto es posible a partir de la separación de cada uno de los planos de interés. Es ampliamente conocido que las imágenes que obtenemos están categorizadas en el espacio de color RGB, y como se muestra en la Figura 3-3 es de interés separar cada una de sus componentes, con el fin de identificar algunos aspectos como las variaciones de color rojo o verde de manera independiente.



**Figura 3-3.:** Elementos de un sistema de visión de máquina con imágenes a color y separación de planos RGB en la agricultura [24].

Hacia el año 1992, fue presentado uno de los primeros sistemas de análisis de imagen para evaluar el color de las frutas. Los algoritmos fueron desarrollados para la clasificación en la madurez de melocotones. En estas primeras clasificaciones por análisis de imagen se coincidió proporcionalmente con la clasificación manual en un 46 % de los casos, y con respecto al estándar de color en un 75 % de las muestras.

La necesidad de mejorar los resultados obtenidos, aumentó significativamente el estudio del desarrollo de técnicas de análisis de imagen, para medir algunos parámetros que definieran por ejemplo, parámetros de calidad, no solo en los melocotones, sino en diferentes frutas y vegetales. De esta forma, en otras investigaciones, las componentes del color en tomates y feijos ha logrado correlacionarse de manera satisfactoria con respecto a los valores colorimétricos estándar y clasificaciones realizadas por expertos en términos de madurez.

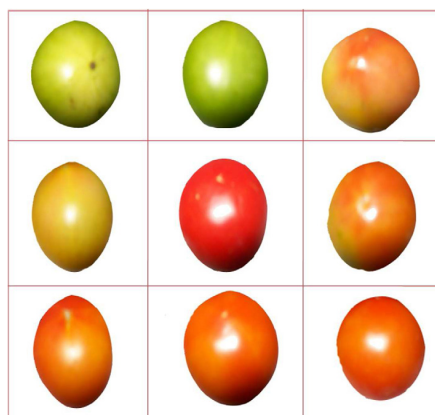


**Figura 3-4.:** Los valores medios de coordenadas de color (centros de las esferas) y desviaciones estándar (límites de las superficies de las esferas) de diferentes lotes de hortalizas (nueve vegetales medidos por lote), representados en el espacio 3-D [72].

Como se evidencia, en el análisis de imagen es común utilizar características asociadas al color, para el caso de los frutos es posible calcular las intensidades de color en imágenes de racimos de fruta fresca (RFF o FFB por sus siglas en inglés Fresh Fruit Bunches), no segmentadas y segmentadas. Los valores de color se pueden calcular para obtener relaciones numéricas del tipo R/G o R/B. Por ejemplo para frutos de palma de aceite es posible enfatizar la diferencia entre la clasificación de maduras y verdes, utilizando los valores máximos al calcular  $R/G \times R/B$ . El valor de salida obtenido a partir de esta manipulación de datos se conoce como índice de madurez. Este índice de madurez, como ha sido denotado anteriormente, es uno de los principales parámetros utilizados para la clasificación según el grado de madurez de los frutos [91].

Por otro lado, analizando las diferencias de color entre distintas hortalizas analizadas en un espacio de color como el CIELab, éstas varían de 9.47 (lechuga - pepino) a 3.04 (albahaca - arvejas congeladas). Las desviaciones máximas en las coordenadas de color en diferentes lotes de cada vegetal, varían entre 3.90 (lechuga) y 0.80 (arvejas congeladas). Los valores medios calculados y sus desviaciones de color en términos de las componentes L, a y b se pueden observar en la Figura 3-4, allí se evidencia que para productos con superficies irregulares como la lechuga y el pepino, las variaciones del color dentro de las muestras son significativas [72].

El color es entonces, un importante factor de calidad que ha sido ampliamente estudiado. Algunas frutas tienen un color homogéneamente distribuido en la superficie de la piel, llamado comúnmente color primario. El color promedio primario de la superficie puede ser un buen indicador de calidad en algunos productos agrícolas como se puede ver en la Figura 3-5. En otros casos, no es posible basarse sólo en el color global como un parámetro determinante de alguna propiedad del producto. Por ejemplo, en algunas variedades de melocotones, manzanas o tomates, se dispone de un color secundario que se puede utilizar como un indicador de la madurez o de su estado de desarrollo.



**Figura 3-5.:** Variaciones del color durante el proceso de maduración en productos agrícolas [38].

Teniendo en cuenta lo evidenciado, se utilizaron diferentes tipos de productos agrícolas para determinar las componentes de color en términos de la separación en planos de los espacios RGB y CIELab, ampliamente utilizados en la literatura. El proceso que se realiza consiste en dos pasos fundamentales, convertir la imagen RGB al espacio CIELab y extraer píxel a píxel cada una de las componentes R, G, B, L, a y b de manera independiente. Ya que la conversión entre los espacios RGB y CIELab no es directa, es necesario realizar una transformación intermedia conocida como XYZ.

El espacio RGB posee una referencia de blancos para capturar las coordenadas de cada uno de sus planos. CIELab por su parte, requiere de un punto de blancos de referencia conocido y definido en función de la herramienta de captura de imágenes, los más conocidos como D50 o D65. Para realizar la transformación de RGB a XYZ es utilizada la Ecuación 3-7, donde los coeficientes de la matriz  $[M]$  son definidos numéricamente en términos de la referencia de blancos.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [M] \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (3-7)$$

Para la conversión del espacio XYZ al espacio CIELab son utilizadas la Ecuaciones 3-8, 3-9, 3-10, 3-11, donde  $X_n$ ,  $Y_n$  y  $Z_n$  son los valores del punto referencia de blancos. Cada una de las componentes encontradas de manera independiente: R, G, B, L, a y b, es representada numéricamente mediante el valor promedio y desviación estándar de dicha componente para la totalidad de píxeles en la imagen. Dicho proceso de medición de color en las imágenes, es mostrado en la Tabla 3-4, para algunos frutos comunes del país como la maracuya, granadilla, mango y tomate.

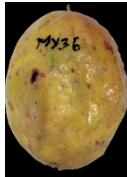



$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 & \text{si } \frac{Y}{Y_n} > 0.008856. \\ 903.3\left(\frac{Y}{Y_n}\right) & \text{si } \frac{Y}{Y_n} \leq 0.008856. \end{cases} \quad (3-8)$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right]. \quad (3-9)$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right]. \quad (3-10)$$

$$\text{donde } f(t) = \begin{cases} t^{1/3} & \text{si } t > 0.008856 \\ 7.787t + \left(\frac{16}{116}\right) & \text{si } t \leq 0.008856. \end{cases} \quad (3-11)$$





**Tabla 3-4.:** Características de color extraídas en diferentes frutos.

Maracuyá	Lab/RGB	Granadilla	Lab/RGB
	L=54.78 ± 4.83 a=-1.28 ± 2.29 b=27.37 ± 4.57 R=160.05 ± 15.99 G=140.80 ± 12.38 B=66.58 ± 9.18		L=59.51 ± 4.25 a=7.31 ± 1.78 b=23.25 ± 3.56 R=181.58 ± 12.02 G=134.51 ± 11.58 B=77.01 ± 6.84
Mango	Lab/RGB	Tomate	Lab/RGB
	L=51.38 ± 3.81 a=-7.92 ± 2.35 b=22.82 ± 3.33 R=121.42 ± 14.78 G=127.42 ± 9.01 B=57.73 ± 5.29		L=57.81 ± 6.90 a=15.59 ± 14.53 b=20.29 ± 4.65 R=182.85 ± 28.85 G=122.89 ± 20.46 B=94.87 ± 15.83

### 3.2.4. Análisis de textura MCNG.

Aspectos como la textura en los frutos mediante imágenes aún no ha sido trabajado ampliamente. Uno de los descriptores de textura usados comúnmente es la matriz de co-ocurrencia de niveles de gris (MCNG o GLCM por sus siglas en inglés Gray-Level Co-occurrence Matrix), a partir de la cuál es posible calcular las características estadísticas de segundo orden en las imágenes. Considerando una imagen que tiene tamaño  $w \times h$  con  $N_g$  niveles de gris, las variaciones de la textura se calculan utilizando la matriz de dependencia espacial de los tonos de gris, donde los píxeles  $i, j$  están separados por una distancia  $d$  a un ángulo  $\theta$  con diferentes nivel de gris [111].

**Tabla 3-5.:** Características de textura MCNG extraídas en diferentes frutos.

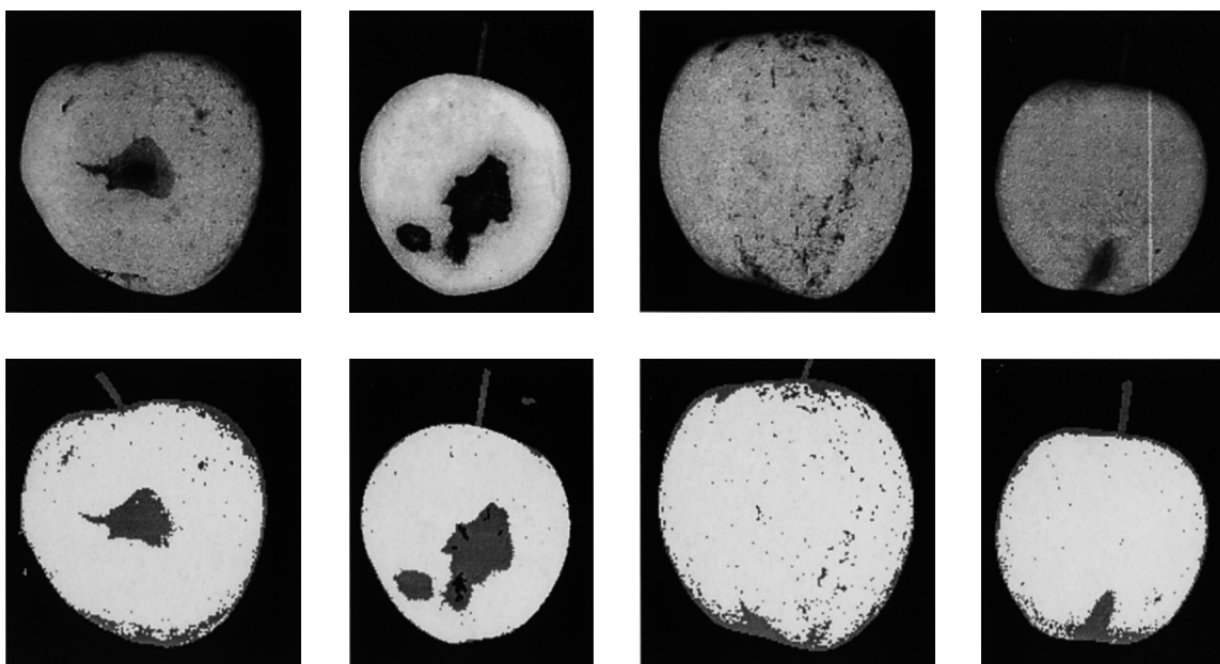
Maracuyá	MCNG	Granadilla	MCNG
	Ent.=5.77 ± 0.36 Hom.=0.47 ± 0.04 Con.=312.89 ± 53.99 Ene.=0.11 ± 0.04 Cor.=0.96 ± 0.01		Ent.=5.72 ± 0.41 Hom.=0.45 ± 0.05 Con.=683.63 ± 104.15 Ene.=0.13 ± 0.04 Cor.=0.94 ± 0.01
Mango	MCNG	Tomate	MCNG
	Ent.=5.75 ± 0.40 Hom.=0.45 ± 0.05 Con.=582.89 ± 94.63 Ene.=0.12 ± 0.04 Cor.=0.92 ± 0.02		Ent.=6.29 ± 0.21 Hom.=0.48 ± 0.04 Con.=375.00 ± 98.97 Ene.=0.05 ± 0.01 Cor.=0.95 ± 0.01

**Algoritmo 3** Análisis de textura: MCNG**Entrada:** Imagen  $I_2$  [w×h]**Salida:** Entropía, Homogeneidad, Contraste, Energía, Correlación1: Transformar Imagen  $I_2$  a:  $Imagen_{gris}$ 2: MCNG  $\leftarrow Imagen_{gris}$ ,  $d$ ,  $\theta$ 3: Entropía, Homogeneidad, Contraste, Energía, Correlación  $\leftarrow$  MCNG4: **Devolver** Entropía, Homogeneidad, Contraste, Energía, Correlación.

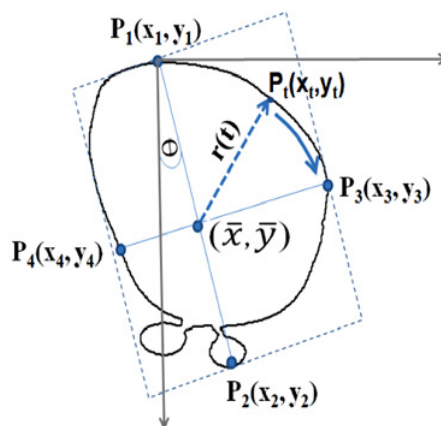
Siguiendo éstas consideraciones, las propiedades de Entropía, Energía, Correlación, Contraste y Homogeneidad fueron medidas en diferentes productos agrícolas, utilizando las Ecuaciones 2-4, 2-5, 2-6, 2-7 y 2-8 presentadas en el capítulo precedente mediante el Algoritmo 3. Los valores promedio junto con su desviación estándar, obtenidos para una serie de imágenes de niveles de gris de cada producto agrícola son mostrados en la Tabla. 3-5.

**3.2.5. Características de tamaño y forma.**

Por otra parte, se han determinado también aspectos como el tamaño, forma y defectos en productos ampliamente comercializados, siendo éstas algunas de las características apreciables en imágenes del espectro visible. En algunos frutos como naranjas, peras o manzanas, las áreas dañadas han podido ser detectadas correctamente, como se aprecia en la Figura 3-6, y la localización de aspectos como el tallo ha podido medirse con cierta exactitud.




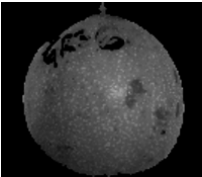

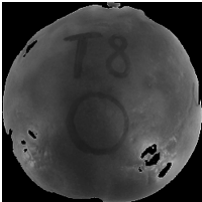
**Figura 3-6.:** Análisis de calidad y defectos en productos agrícolas [64].



**Figura 3-7.:** Extracción de características de tamaño en productos agrícolas mediante imágenes [39].

Utilizando otros aspectos como: la localización correcta de las áreas parcialmente coloreadas que aparecen en algunas variedades de frutos como el melocotón, ha sido posible lograr una estimación real de su tamaño. Estas propiedades físicas, asociadas al tamaño y la forma, han sido otro de los parámetros relacionados con las características fisicoquímicas de los productos agrícolas. Por ejemplo el tamaño, ha podido ser estimado de igual manera usando la visión por computador, ya sea mediante la medición de área del fruto[14], el perímetro [15] o el diámetro mayor del mismo [16]. Estos parámetros de tamaño pueden ser medidos como se muestra en la Figura 3-7, pero para realizar la parametrización de aspectos de forma suele ser un poco más complejo.

**Tabla 3-6.:** Características de forma (Momentos de Hu) extraídas en diferentes frutos.

Maracuyá	Hu	Granadilla	Hu
	$Hu_1=6.74e-4\pm 2.83e-5$ $Hu_2=2.93e-8\pm 2.19e-8$ $Hu_3=8.76e-12\pm 1.36e-11$ $Hu_4=2.31e-12\pm 4.13e-12$ $Hu_5=3.99e-23\pm 1.28e-22$ $Hu_6=5.01e-16\pm 1.04e-15$ $Hu_7=4.79e-25\pm 6.01e-24$		$Hu_1=6.54e-4\pm 2.25e-5$ $Hu_2=2.93e-9\pm 4.40e-9$ $Hu_3=3.05e-12\pm 9.01e-12$ $Hu_4=8.04e-13\pm 3.26e-12$ $Hu_5=1.81e-23\pm 1.25e-22$ $Hu_6=3.45e-17\pm 1.97e-16$ $Hu_7=-1.67e-24\pm 1.51e-23$
Mango	Hu	Tomate	Hu
	$Hu_1=6.98e-4\pm 3.05e-5$ $Hu_2=2.31e-8\pm 1.90e-8$ $Hu_3=5.54e-12\pm 9.47e-12$ $Hu_4=6.44e-13\pm 2.22e-12$ $Hu_5=-6.28e-24\pm 3.73e-23$ $Hu_6=-4.87e-17\pm 3.83e-16$ $Hu_7=7.32e-24\pm 4.93e-23$		$Hu_1=6.40e-4\pm 1.40e-5$ $Hu_2=6.48e-9\pm 7.43e-9$ $Hu_3=2.73e-12\pm 3.00e-12$ $Hu_4=5.45e-14\pm 1.13e-13$ $Hu_5=-1.29e-26\pm 6.63e-26$ $Hu_6=-5.00e-19\pm 7.56e-18$ $Hu_7=6.49e-27\pm 2.22e-25$

**Algoritmo 4** Características de Tamaño y Forma**Entrada:** Imagen  $I_2$  [w×h]**Salida:**  $Eje_{Mayor}$ ,  $Eje_{Menor}$ , Area, Perimetro,  $Hu_1$ ,  $Hu_2$ ,  $Hu_3$ ,  $Hu_4$ ,  $Hu_5$ ,  $Hu_6$ ,  $Hu_7$ 


---

```

1: Contornos  $\leftarrow$  Encontrar contornos de: Imagen  $I_2$ 
2: Para  $contorno = 0$  hasta  $ultimo_{contorno}$  hacer
3:    $Area_{Contorno} =$  Calcular area de: contorno
4:   Si  $Area_{Contorno} > Area$  Entonces
5:      $Area = Area_{Contorno}$ 
6:      $Contorno_{Mayor} =$  contorno
7:   Fin Si
8: Fin Para
9: Elipse  $\leftarrow Contorno_{Mayor}$ 
10:  $Eje_{Mayor} \leftarrow$  Elipse;  $Eje_{Menor} \leftarrow$  Elipse
11: Calcular perimetro de:  $Contorno_{Mayor}$ 
12: Transformar Imagen  $I_2$  a:  $Imagen_{gris}$ 
13:  $Momentos_{centrales} \leftarrow Imagen_{gris}$ 
14:  $Momentos_{Hu} \leftarrow Momentos_{centrales}$ 
15:  $Hu_1, Hu_2, Hu_3, Hu_4, Hu_5, Hu_6, Hu_7 \leftarrow Momentos_{Hu}$ 
16: Devolver  $Eje_{Mayor}$ ,  $Eje_{Menor}$ , Area, Perimetro,  $Hu_1$ ,  $Hu_2$ ,  $Hu_3$ ,  $Hu_4$ ,  $Hu_5$ ,  $Hu_6$ ,  $Hu_7$ 

```

---

Utilizando por ejemplo los momentos de  $Hu$ , es posible obtener un espacio de características de forma invariante a los cambios que pueda sufrir un objeto dentro de la imagen, como lo son la traslación, el escalamiento o la rotación. Mediante las Ecuaciones 2-10, 2-11, 2-12, 2-13, 2-14, 2-15 y 2-16, presentadas en el capítulo anterior, es posible analizar algunos productos agrícolas utilizando el Algoritmo 4 como se puede evidenciar en la Tabla 3-6. Para realizar este proceso es necesario utilizar imágenes en niveles de gris de la región segmentada, obtenidas luego del procesamiento.

### 3.2.6. Composición del vector de características.

Por otro lado, otras investigaciones han desarrollado algoritmos para el reconocimiento de manzanas en los árboles, haciendo uso de robot's tipo "cosechador" [25, 55, 100, 125]. Estos sistemas poseen un algoritmo de visión que utiliza una cámara de color para captar imágenes y un computador para procesarlas, realizando tareas de reconocimiento y localización. En éstos casos, las imágenes son analizadas usando la diferencia del color rojo, porque se ha demostrado que frutos como éste (manzana), tienen su mayor dispersión en esta componente del color. Los resultados para este caso, han mostrado una tasa de éxito del 88% con un umbral óptimo, pero la tasa de error era de un 18% en condiciones de mala iluminación.



**Tabla 3-7.:** Media y desviación estándar de las 22 características extraídas para diferentes frutos.

No.	Característica	Maracuyá	Granadilla	Mango Común	Tomate
1	L	58.78 ± 4.83	59.51 ± 4.25	51.38 ± 3.81	57.81 ± 6.90
2	a	-1.28 ± 2.29	7.31 ± 1.78	-7.92 ± 2.35	15.59 ± 14.53
3	b	27.37 ± 4.57	23.25 ± 3.56	22.82 ± 3.33	20.29 ± 4.65
4	B	66.58 ± 9.18	77.01 ± 6.84	57.73 ± 5.29	94.87 ± 15.83
5	G	140.80 ± 12.38	134.51 ± 11.58	127.42 ± 9.01	122.89 ± 28.85
6	R	160.05 ± 15.99	181.58 ± 12.02	121.42 ± 14.78	182.85 ± 20.46
7	Entropía	5.77 ± 0.36	5.72 ± 0.41	5.75 ± 0.40	6.29 ± 0.21
8	Homogeneidad	0.47 ± 0.04	0.45 ± 0.05	0.45 ± 0.05	0.48 ± 0.04
9	Contraste	312.89 ± 53.99	683.63 ± 104.15	582.89 ± 94.63	375.00 ± 98.97
10	Energía	0.11 ± 0.04	0.13 ± 0.04	0.12 ± 0.04	0.05 ± 0.01
11	Correlación	0.96 ± 0.01	0.94 ± 0.01	0.92 ± 0.02	0.95 ± 0.01
12	Eje Mayor	280.10 ± 29.28	112.16 ± 10.73	115.38 ± 8.98	243.70 ± 31.45
13	Eje Menor	205.06 ± 18.28	100.80 ± 8.48	97.62 ± 10.01	219.20 ± 28.41
14	Área	43079.40 ± 6386.36	8595.43 ± 1399.24	8598.08 ± 1388.57	42297.58 ± 10297.62
15	Perímetro	984.24 ± 157.99	402.74 ± 42.47	475.09 ± 51.69	804.95 ± 105.64
16	Hu 1	6.74e-4 ± 2.83e-5	6.54e-4 ± 2.25e-5	6.98e-4 ± 3.05e-5	6.40e-4 ± 1.40e-5
17	Hu 2	2.93e-8 ± 2.19e-8	2.93e-9 ± 4.40e-9	2.31e-8 ± 1.90e-8	6.48e-9 ± 7.43e-9
18	Hu 3	8.76e-12 ± 1.36e-11	3.05e-12 ± 9.01e-12	5.54e-12 ± 9.47e-12	2.73e-12 ± 3.00e-12
19	Hu 4	2.31e-12 ± 4.13e-12	8.04e-13 ± 3.26e-12	6.44e-13 ± 2.22e-12	5.45e-14 ± 1.13e-13
20	Hu 5	3.99e-23 ± 1.28e-22	1.81e-23 ± 1.25e-22	-6.28e-24 ± 3.73e-23	-1.29e-26 ± 6.63e-26
21	Hu 6	5.01e-16 ± 1.04e-15	3.45e-17 ± 1.97e-16	-4.87e-17 ± 3.83e-16	-5.00e-19 ± 7.56e-18
22	Hu 7	4.79e-25 ± 6.01e-24	-1.67e-24 ± 1.51e-23	7.32e-24 ± 4.93e-23	6.49e-27 ± 2.22e-25

En otros productos como las olivas de mesa (aceitunas), se han empleado sistemas de visión para determinar los parámetros relacionados a sus características fisicoquímicas, para el desarrollo de un rápido y eficiente algoritmo que fue incorporado a un sistema automático de clasificación [17, 30, 31]. El análisis colorimétrico indicó que había suficientes diferencias en las coordenadas de color e intensidad entre las clases de oliva, como para usar este sistema. En otros casos, se han utilizado técnicas más avanzadas como análisis espectral, para evaluar la firmeza y contenido de sólidos solubles [27, 75, 95], y recientemente se han realizado estudios para utilizar aspectos de textura en la imagen, como la media, el contraste, la homogeneidad, la varianza, la correlación y la entropía [11, 22, 125] para analizar productos agrícolas. Éste ha sido utilizado como elemento discriminante entre tipo de frutos, pero no como suministro para la estimación de características fisicoquímicas.

Sin embargo, según lo presentado en la metodología es posible realizar sobre las imágenes de frutos, algunas transformaciones de color, separando cada uno de sus canales o planos, de manera independiente. También es posible determinar aspectos de tamaño como áreas, perímetros, contornos y ejes principales. Y, agregando otros descriptores de aspectos visibles en las imágenes como lo son la textura y la forma, es posible componer un descriptor que permita la correlación de éstos con características fisicoquímicas medidas, asociadas a los productos agrícolas. En la Tabla 3-7 se presentan algunos de estos descriptores, obtenidas a partir de la aplicación de los Algoritmos 1, 2, 3 y 4 presentados.

### 3.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.

Componer un descriptor lo suficientemente bueno para poder ser correlacionado con cualquier característica fisicoquímica es una tarea complicada. Sin embargo, como se analizó en la sección anterior, usando características de color, textura, tamaño y forma, es posible obtener una correlación con ciertas propiedades fisicoquímicas. Al obtener estas características la información de los productos agrícolas puede ser tratada ahora como datos numéricos únicamente que necesitan ser interpretados. Una de las apreciaciones normales en estas series de datos es la desviación estándar, la cual refleja la medida de dispersión de éstos en función del promedio.

#### 3.3.1. Análisis ACP.

Al analizar el conjunto global de datos como un espacio  $R^n$  en el cual  $n$  es el número de características, éstos tendrán una variabilidad asociada en cada una de sus componentes. Una de las técnicas más utilizadas para llevar a cabo los análisis estadísticos para el tratamiento de los datos es el análisis por componentes principales (ACP), el cuál básicamente determina las componentes de mayor dispersión o variabilidad de los datos, en términos de las  $n$ -componentes. Sin embargo, no en todos los procesos las características o parámetros analizados según la variabilidad de los datos son eficaces para realizar procesos de estimación o clasificación, por lo cuál es necesario utilizar otro tipo de análisis.

Mediante la aplicación de ACP a un descriptor conformado por 22 características que incluyen aspectos de color, textura, tamaño y forma como el presentado en la sección precedente, se pueden analizar los datos en términos de su varianza. A partir del vector propio asociado al valor propio con mayor variabilidad como los presentados en la Tabla 3-8, se pueden identificar como la mayoría de estas características influye en éste vector. Multiplicando cada uno de los vectores propios por el valor obtenido de la  $m$  característica para la  $n$  muestra de determinado producto, es posible obtener la proyección sobre este nuevo espacio. Utilizando la proyección de los mayores valores propios cuya variabilidad explicada sea mayor al 95 % se conforma un nuevo conjunto de datos útil con fines de estimación mediante una RNA.

Por ejemplo, al utilizar los 5 mayores valores propios y obtener su respectiva proyección mediante sus vectores propios asociados en Mangos Nam doc mai, se obtiene una varianza explicada cercana al 96 %. Una RNA con 5 neuronas de entrada correspondientes a ésta proyección y siendo la neurona de salida la estimación de masa, volumen, sólidos solubles, acidez o pH, es posible definir mediante una regresión lineal simple, un modelo como el presentado en las Ecuación 3-12 para la estimación de dichas características.

$$Caracteristica_{fisicoquimica} = Factor \times Estimacion_{VA} + b; Error : \epsilon \quad (3-12)$$

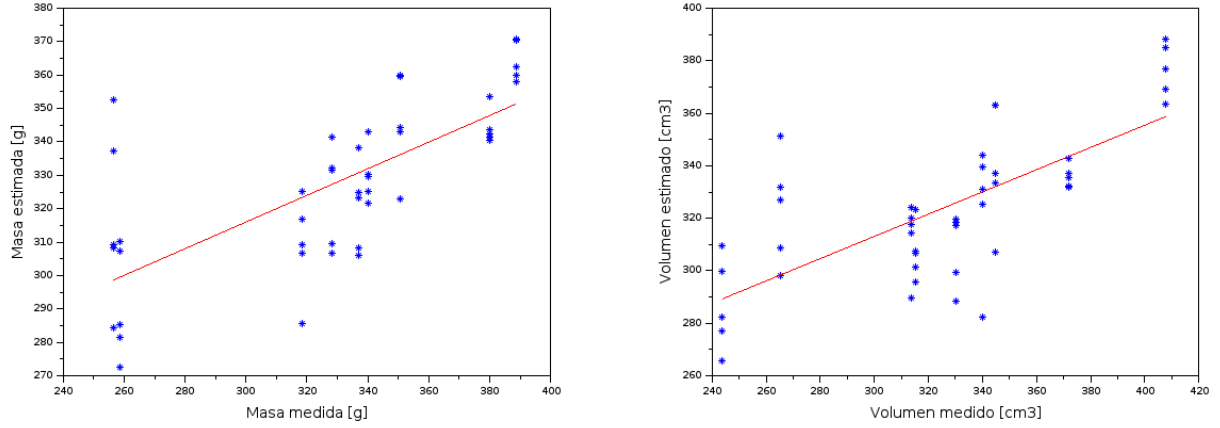
**Tabla 3-8.:** Vector propio del mayor valor propio ACP en términos de las 22 características en diferentes frutos.

No.	Característica	Maracuyá	Granadilla	Mango Común	Tomate	Mango Nam doc mai
1	L	-0.19	-0.05	-0.18	-0.20	-0.12
2	a	-0.01	-0.23	-0.04	<b>0.38</b>	-0.18
3	b	-0.25	-0.26	-0.45	-0.39	-0.17
4	B	-0.10	-0.01	0.01	0.05	-0.07
5	G	-0.18	-0.02	-0.15	-0.29	-0.10
6	R	-0.18	-0.09	-0.21	<b>0.19</b>	-0.15
7	Entropía	-0.29	-0.30	-0.40	0.04	-0.21
8	Homogeneidad	<b>0.25</b>	<b>0.37</b>	<b>0.38</b>	-0.31	<b>0.22</b>
9	Contraste	-0.05	-0.19	-0.07	0.08	-0.18
10	Energía	<b>0.40</b>	<b>0.35</b>	<b>0.49</b>	-0.06	<b>0.21</b>
11	Correlación	0.00	<b>0.28</b>	0.04	-0.28	0.07
12	Eje Mayor	<b>0.29</b>	<b>0.39</b>	0.01	<b>0.27</b>	<b>0.20</b>
13	Eje Menor	-0.02	<b>0.19</b>	0.08	<b>0.29</b>	-0.06
14	Área	0.08	<b>0.23</b>	0.04	<b>0.30</b>	0.06
15	Perímetro	<b>0.26</b>	<b>0.33</b>	<b>0.23</b>	<b>0.27</b>	<b>0.17</b>
16	Hu 1	<b>0.32</b>	0.08	<b>0.17</b>	0.03	<b>0.33</b>
17	Hu 2	<b>0.21</b>	-0.06	0.00	0.12	<b>0.33</b>
18	Hu 3	<b>0.25</b>	0.10	<b>0.13</b>	-0.13	<b>0.27</b>
19	Hu 4	<b>0.26</b>	0.10	<b>0.11</b>	0.03	<b>0.33</b>
20	Hu 5	<b>0.17</b>	0.09	-0.09	-0.01	<b>0.30</b>
21	Hu 6	<b>0.23</b>	0.09	-0.09	-0.01	<b>0.36</b>
22	Hu 7	-0.01	-0.07	0.08	0.02	0.13
	Valor Propio	0.36	0.31	0.22	0.39	0.52
	Varianza Explicada (%)	45.29	34.43	25.37	32.71	37.07

Aplicando el procedimiento sobre un conjunto de frutos de Mangos Nam doc mai, utilizando una RNA luego de aplicar ACP al conjunto de 22 características, es posible obtener la estimación por visión artificial  $Estimacion_{VA}$ . Comparando los valores reales de dichas características con la estimación obtenida se pueden definir las correlaciones existentes y por tanto definir el modelo de regresión lineal, para futuras estimaciones.

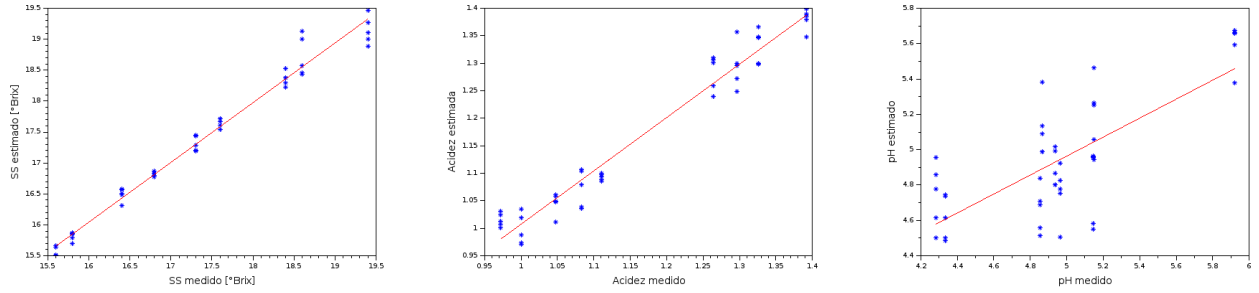
Por ejemplo, para la estimación de características físicas como la masa y el volumen se puede comparar el valor medido en gramos en laboratorio y la estimación obtenida mediante el sistema de visión artificial, como se puede apreciar en la Figura 3-8, donde cada punto está ubicado en un plano XY donde X es el valor medido y Y el valor estimado. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura 3-9.

Los modelos obtenidos mediante un sistema de regresión son presentados en las Ecuaciones 3-13, 3-14, 3-15, 3-16 y 3-17 y representados en las Figuras 3-8 y 3-9 como líneas continuas de color rojo.



(a) Estimación de masa mediante ACP.

(b) Estimación de volumen mediante ACP.

**Figura 3-8.:** Estimación de propiedades físicas mediante ACP.

(a) Estimación de SS mediante ACP. (b) Estimación de AT mediante ACP. (c) Estimación de pH mediante ACP.

**Figura 3-9.:** Estimación de propiedades químicas mediante ACP.

$$Masa = 1.26 \times Estimacion_{VA} - 86.94[g]; Error : 17.59[g] \quad (3-13)$$

$$Volumen = 1.17 \times Estimacion_{VA} - 56.27[cm^3]; Error : 20.49[cm^3] \quad (3-14)$$

$$SS = 1.01 \times Estimacion_{VA} - 0.25[Brix]; Error : 0.17[Brix] \quad (3-15)$$

$$AT = 0.98 \times Estimacion_{VA} + 0.01; Error : 0.03 \quad (3-16)$$

$$pH = 1.01 \times Estimacion_{VA} - 0.07; Error : 0.22 \quad (3-17)$$

### 3.3.2. Análisis ADL.

Mediante el uso de técnicas como el análisis discriminante lineal (ADL), es posible encontrar un hiperplano en el espacio de las componentes que permita separar entre diferentes tipos de características que no necesariamente están en función de la variabilidad de los datos, Figura 3-10. Este tipo de análisis llega a ser mucho más útil en la estimación de parámetros fisicoquímicos de los productos agrícolas, ya que al componer el espacio de características sin previo conocimiento de la dependencia o no de la variable real con éstas, no será fácil realizar el análisis ACP.

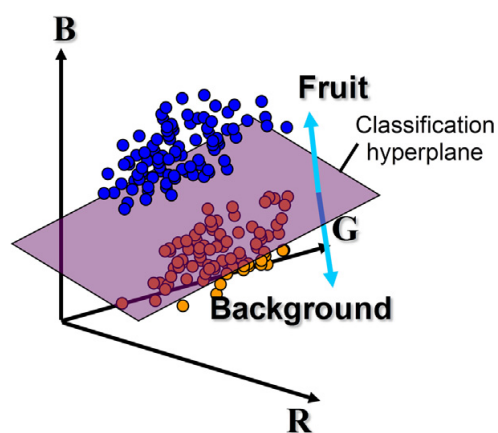


Figura 3-10.: Separación en hiperplanos mediante la técnica ADL [77].

Al analizar lo que sucede con cada una de las componentes obtenidas por ejemplo para la separación de los datos en términos de la masa o el volumen como variables dependientes, es posible observar que un conjunto de datos compuesto por características de color, textura, tamaño y forma puede determinar espacios lo suficientemente discriminantes. Analizando los resultados obtenidos para el análisis ADL resumidos en la Tabla 3-9, en términos de la componente con mayor discriminancia de los datos, se puede observar que los datos son discriminados con una proporción entre el 32 % y el 71 % para el caso de una estimación de masa en diferentes productos agrícolas. Observando las componentes del vector de características, se puede observar que para dichos productos, las componentes L, b, G, R, Contraste, Correlación, Eje Mayor, Eje Menor,  $Hu_2$ ,  $Hu_4$ ,  $Hu_5$  y  $Hu_7$ , son las que mayor información aportan a dicha discriminancia.

Otro análisis ADL similar para la estimación de volumen en diferentes productos agrícolas, permite evidenciar que las componentes L, G, R, Energía, Correlación, Eje Mayor, Eje Menor, Área,  $Hu_4$ ,  $Hu_5$  y  $Hu_6$  como se muestra en la Tabla 3-10, ayudan a discriminar los valores medidos con una proporción entre el 35 % y el 97 %. De esta manera, es posible considerar que mediante este espacio, vector o descriptor de características es posible correlacionar adecuadamente valores reales asociados a productos fisicoquímicos, mediante el uso de técnicas compuestas como la regresión lineal múltiple o las redes neuronales artificiales.

**Tabla 3-9.:** Vectores y valores propios del primer discriminante ADL en términos de las 22 características para la estimación de masa en diferentes frutos.

No.	Característica	Maracuyá	Granadilla	Mango Común	Tomate
1	L	<b>0.50</b>	-0.55	<b>0.42</b>	<b>0.21</b>
2	a	0.05	0.09	0.04	-0.60
3	b	-0.08	-0.07	<b>0.11</b>	-0.23
4	B	-0.08	-0.07	0.04	-0.21
5	G	-0.31	<b>0.64</b>	-0.40	-0.50
6	R	-0.28	0.03	-0.27	<b>0.37</b>
7	Entropía	0.00	0.00	-0.08	0.01
8	Homogeneidad	-0.01	-0.02	-0.04	0.04
9	Contraste	<b>0.16</b>	0.05	<b>0.11</b>	0.03
10	Energía	-0.12	0.02	0.01	-0.08
11	Correlación	<b>0.25</b>	0.03	0.02	0.06
12	Eje Mayor	0.09	0.00	<b>0.30</b>	0.06
13	Eje Menor	<b>0.18</b>	-0.03	<b>0.40</b>	<b>0.11</b>
14	Área	-0.27	0.04	-0.50	-0.17
15	Perímetro	-0.03	-0.03	-0.05	0.04
16	Hu 1	-0.07	0.01	-0.05	-0.01
17	Hu 2	<b>0.12</b>	0.03	0.01	0.03
18	Hu 3	-0.03	-0.17	0.05	0.04
19	Hu 4	<b>0.43</b>	-0.18	-0.04	0.02
20	Hu 5	0.00	0.02	<b>0.17</b>	<b>0.17</b>
21	Hu 6	-0.37	<b>0.44</b>	-0.15	-0.08
22	Hu 7	-0.03	0.06	0.03	<b>0.11</b>
	Valor Propio	2.04	7.03	1.34	9.80
	Discrim. Explicada (%)	68.97	63.59	32.77	71.06

**Tabla 3-10.:** Vectores y valores propios del primer discriminante ADL en términos de las 22 características para la estimación de volumen en diferentes frutos.

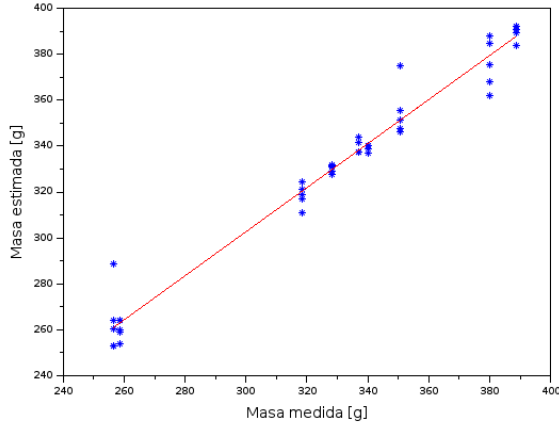
No.	Característica	Maracuyá	Granadilla	Mango Común	Tomate
1	L	-0.52	<b>0.43</b>	<b>0.72</b>	0.01
2	a	-0.09	-0.23	0.01	-0.73
3	b	0.07	-0.12	-0.02	-0.11
4	B	0.08	-0.04	-0.05	-0.15
5	G	<b>0.29</b>	-0.76	-0.55	-0.35
6	R	<b>0.34</b>	-0.32	-0.24	<b>0.62</b>
7	Entropía	-0.02	-0.01	0.01	0.07
8	Homogeneidad	0.00	-0.03	0.01	0.01
9	Contraste	-0.17	0.00	0.06	-0.24
10	Energía	<b>0.12</b>	-0.13	-0.01	0.07
11	Correlación	-0.27	-0.01	0.02	-0.16
12	Eje Mayor	-0.09	0.03	<b>0.10</b>	0.04
13	Eje Menor	-0.19	0.08	<b>0.14</b>	<b>0.22</b>
14	Área	<b>0.28</b>	-0.07	-0.18	-0.17
15	Perímetro	0.03	0.00	0.00	-0.06
16	Hu 1	0.06	-0.04	-0.06	-0.06
17	Hu 2	-0.12	0.02	0.01	0.08
18	Hu 3	0.04	-0.01	-0.01	0.04
19	Hu 4	-0.39	-0.05	<b>0.14</b>	-0.04
20	Hu 5	0.00	<b>0.11</b>	<b>0.15</b>	-0.06
21	Hu 6	<b>0.33</b>	-0.12	-0.08	0.00
22	Hu 7	0.03	-0.11	0.00	-0.06
	Valor Propio	1.79	2.71	1.24	267.63
	Discrim. Explicada (%)	69.09	44.51	35.74	96.99

Mediante la aplicación de ADL a un descriptor conformado por 22 características que incluyen aspectos de color, textura, tamaño y forma como el realizado mediante ACP, se pueden discriminar los datos en función de grupos o clases. Las clases o grupos son conformadas por un conjunto de muestras del producto agrícola con características similares. Utilizando la proyección de los mayores valores propios cuya discriminancia explicada sea mayor al 95% se conforma un nuevo conjunto de datos útil con fines de estimación mediante una RNA. El resultado para el análisis ADL para los discriminantes lineales con discriminancia explicada total de 96.8% en Mango Nam doc mai, se puede observar en la Tabla **3-11**.

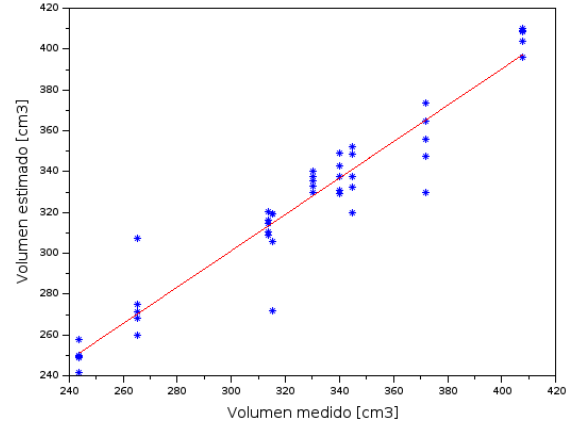
**Tabla 3-11.:** Vectores y valores propios del ADL en términos de las 22 características para la separación por clases en Mango Nam doc mai.

No.	Característica	DL1	DL2	DL3	DL4	DL5
1	L	-0.61	-0.79	<b>1.45</b>	<b>1.10</b>	<b>1.60</b>
2	a	0.06	-0.12	<b>0.23</b>	0.16	-0.24
3	b	-0.07	-0.19	-0.41	-0.38	0.13
4	B	-0.08	-0.01	-0.26	-0.23	0.07
5	G	<b>0.66</b>	<b>0.43</b>	-0.31	-0.18	-1.45
6	R	-0.06	<b>0.75</b>	-0.93	-0.75	-0.09
7	Entropía	-0.02	0.06	-0.07	0.04	0.04
8	Homogeneidad	-0.02	0.01	-0.09	-0.06	0.01
9	Contraste	0.04	-0.13	0.05	0.08	-0.01
10	Energía	-0.03	0.04	-0.14	-0.21	0.10
11	Correlación	0.03	-0.23	0.06	<b>0.21</b>	-0.04
12	Eje Mayor	-0.09	0.02	0.02	-0.29	0.12
13	Eje Menor	-0.07	-0.08	-0.08	-0.21	0.10
14	Área	0.12	0.02	0.04	<b>0.28</b>	-0.16
15	Perímetro	0.01	0.08	-0.06	0.08	-0.01
16	Hu 1	-0.24	-0.27	<b>0.20</b>	0.18	-0.18
17	Hu 2	<b>0.26</b>	0.13	-0.26	-0.08	0.16
18	Hu 3	0.06	-0.01	-0.03	0.18	-0.04
19	Hu 4	-0.05	<b>0.20</b>	0.03	-0.82	0.09
20	Hu 5	-0.07	-0.05	0.03	0.18	0.06
21	Hu 6	0.04	-0.11	0.00	<b>0.42</b>	-0.11
22	Hu 7	0.04	0.03	-0.03	0.10	-0.02
	Valor Propio	11.05	3.00	1.05	0.51	0.30
96.80%	Discrim. Explicada (%)	67.13	18.26	6.40	3.12	1.87

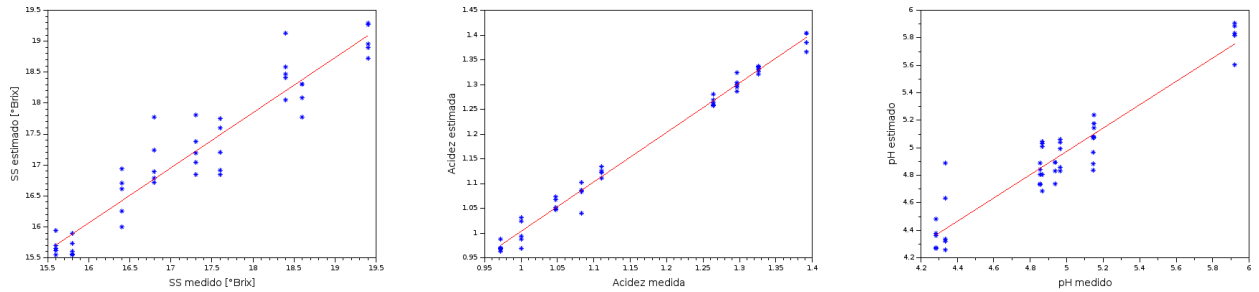
Aplicando el procedimiento sobre el mismo conjunto de frutos de Mangos Nam doc mai, utilizando para el análisis ACP al conjunto de 22 características, es posible obtener la estimación por visión artificial  $Estimacion_{VA}$  basado en el ADL. Comparando los valores reales de dichas características con la estimación obtenida se pueden definir las correlaciones existentes y por tanto definir el modelo de regresión lineal, para futuras estimaciones. Para la estimación de características físicas como la masa y el volumen se puede apreciar en la Figura **3-11** la comparación entre la medición y la estimación. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura **3-12**. Los modelos obtenidos mediante un sistema de regresión son presentados en las Ecuaciones 3-18, 3-19, 3-20, 3-21 y 3-22 y representados en las Figuras **3-11** y **3-12** como líneas continuas de color rojo.



(a) Estimación de masa mediante ADL.



(b) Estimación de volumen mediante ADL.

**Figura 3-11.:** Estimación de propiedades físicas mediante ADL.

(a) Estimación de SS mediante ADL.(b) Estimación de AT mediante ADL.(c) Estimación de pH mediante ADL.

**Figura 3-12.:** Estimación de propiedades químicas mediante ADL.

$$Masa = 1.01 \times Estimacion_{VA} - 5.29[g]; Error : 7.52[g] \quad (3-18)$$

$$Volumen = 1.02 \times Estimacion_{VA} - 7.60[cm^3]; Error : 13.02[cm^3] \quad (3-19)$$

$$SS = 1.01 \times Estimacion_{VA} - 0.22[Brix]; Error : 0.36[Brix] \quad (3-20)$$

$$AT = 0.99 \times Estimacion_{VA} + 0.00; Error : 0.01 \quad (3-21)$$

$$pH = 1.04 \times Estimacion_{VA} - 0.22; Error : 0.14 \quad (3-22)$$



### 3.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.

Se dice que la correlación entre dos variables X e Y es perfectamente positiva cuando exactamente en la medida que aumenta una de ellas aumenta la otra. Esto sucede cuando la relación entre ambas variables es funcionalmente exacta. Difícilmente ocurrirá pero, es frecuente en las ciencias físicas donde los fenómenos se ajustan a leyes conocidas. Análogamente, se dice que la relación es perfectamente negativa cuando exactamente en la medida que aumenta una variable disminuye la otra. Igual que en el caso anterior esto sucede para relaciones funcionales exactas, propio de las ciencias físicas.

En función de una serie de atributos extraídos mediante las imágenes es posible realizar procesos matemáticos y computacionales para estimar las características fisicoquímicas de los productos agrícolas. Uno de los procesos más utilizadas para la correlación de variables es la RLM, sin embargo este proceso conlleva un demandante proceso matemático. Por otro lado, utilizando las redes neuronales artificiales es posible realizar estimaciones con una buena exactitud y menos demanda computacional y matemática. Sin embargo, es adecuado realizar una serie de tratamiento previo a los datos, para poder encontrar la mejor variabilidad o discriminancia de los mismos en términos de la variable a analizar. Dos de los métodos útiles en este caso pueden ser el ACP o el ADL.

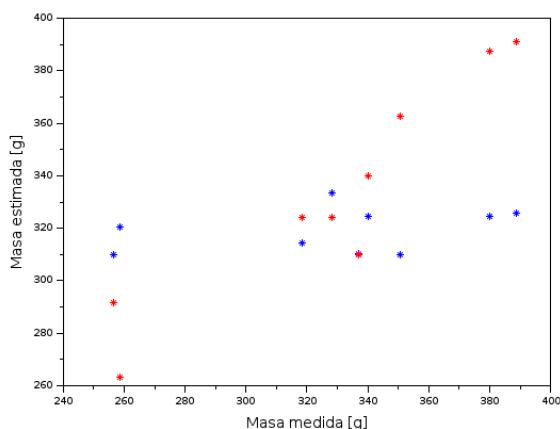
Es posible definir algunas medidas para definir la bondad del ajuste en un sistema de regresión lineal múltiple. Los cuatro parámetros son comúnmente conocidos como  $R$ ,  $R^2$ ,  $R^2$  ajustado y  $\epsilon$ , éste último definido previamente como el error de la regresión, Ecuación 2-25.  $R$ , es conocido como el coeficiente de correlación; el cual define la intensidad de la correlación entre un conjunto de  $m$ -variables independientes ( $m \geq 1$ ) y una variable independiente, Ecuación 3-23. El coeficiente de determinación o  $R^2$ , mide la variabilidad de la variable dependiente explicada mediante las  $m$ -variables independientes.

$$R = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}. \quad (3-23)$$

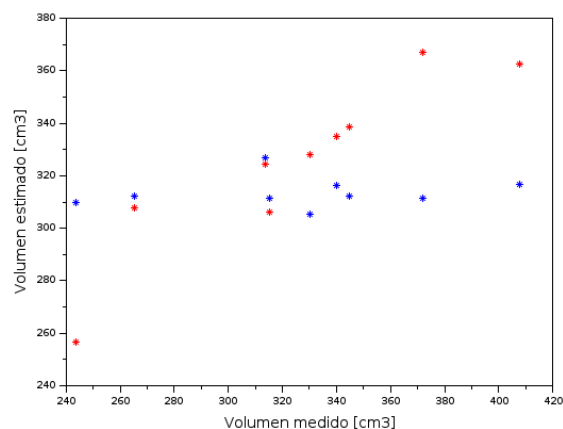
El último parámetro,  $R^2$  ajustado es análogo al anterior, sin embargo a medida que se introducen más variables para que  $R$  sea mucho más cercano a 1, como sería lo ideal, este va aumentando en función de las mismas; es posible calcularlo mediante la Ecuación 3-24 donde  $n$  es la cantidad de muestras utilizadas en el modelo de regresión y  $m$ -la cantidad de  $m$ -variables independientes utilizadas para realizar el ajuste del mismo.

$$R_{ajustado}^2 = 1 - \frac{n-1}{n-m} \cdot (1 - R^2). \quad (3-24)$$

Utilizando esta información es posible comparar por ejemplo las Ecuaciones determinadas mediante el ACP y ADL presentados en la sección precedente. Si bien, el sistema de regresión tiene asociado un coeficiente de correlación y de determinación, es necesario determinar la respuesta sobre un nuevo conjunto. Utilizando los modelos determinados por cada uno de los sistemas, se comparó la respuesta sobre la estimación de un conjunto nuevo de Mangos Nam doc mai. La estimación de características físicas como la masa y el volumen se puede apreciar en la Figura 3-13, en color azul para el ACP y rojo para el ADL. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura 3-14.

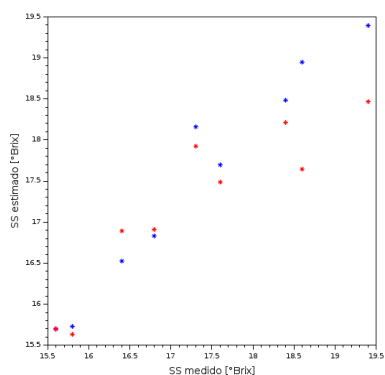


(a) Validación de masa mediante ACP y ADL.

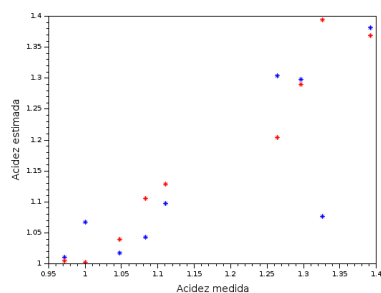


(b) Validación de volumen mediante ACP y ADL.

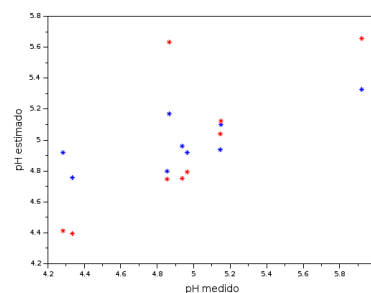
**Figura 3-13.:** Validación de la estimación de propiedades físicas mediante ACP y ADL.



(a) Validación de SS mediante ACP y ADL.



(b) Validación de AT mediante ACP y ADL.



(c) Validación de pH mediante ACP y ADL.

**Figura 3-14.:** Validación de la estimación de propiedades químicas mediante ACP y ADL.

Comparando los valores de los estadísticos descriptivos consignados en las Tablas **3-12**, **3-13**, **3-14**, **3-15** y **3-16** se puede evidenciar claramente que la separación por clases mediante el análisis ADL permite obtener resultados de correlación múltiple superiores a 0.8 para la estimación de características fisicoquímicas: masa, volumen, sólidos solubles, acidez y pH. El erro típico de los modelos obtenidos es coherente con los resultados obtenidos para el conjunto de validación. En resumen, es posible validar mediante un conjunto de prueba y los estadísticos descriptivos mostrados, los modelos obtenidos con visión artificial para la estimación de características fisicoquímicas. En el siguiente capítulo se presentarán los resultados obtenidos para algunas casos de estudio siguiendo la metodología aquí presentada.

**Tabla 3-12.:** Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.71	0.33	0.50	0.11	0.49	-0.01	17.59	48.40
ADL + RNA	0.98	0.93	0.97	0.87	0.97	0.85	7.52	17.97

**Tabla 3-13.:** Análisis estadístico de correlación entre los datos de visión y el volumen de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.70	0.14	0.49	0.02	0.48	-0.11	20.49	54.78
ADL + RNA	0.95	0.92	0.91	0.85	0.91	0.83	13.02	24.75

**Tabla 3-14.:** Análisis estadístico de correlación entre los datos de visión y los sólidos solubles de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.99	0.97	0.98	0.95	0.98	0.95	0.17	0.36
ADL + RNA	0.95	0.91	0.90	0.84	0.90	0.81	0.36	0.60

**Tabla 3-15.:** Análisis estadístico de correlación entre los datos de visión y la acidez de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.97	0.80	0.96	0.65	0.96	0.60	0.03	0.10
ADL + RNA	0.99	0.97	0.99	0.94	0.98	0.93	0.01	0.03

**Tabla 3-16.:** Análisis estadístico de correlación entre los datos de visión y el pH de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.73	0.76	0.54	0.58	0.53	0.52	0.22	0.39
ADL + RNA	0.94	0.78	0.88	0.61	0.88	0.56	0.13	0.33

## Resumen del capítulo

En este capítulo se presenta la metodología propuesta para la identificación de características fisicoquímicas en productos agrícolas mediante técnicas de visión de máquina. Se proponen cuatro etapas fundamentales: *i.* identificación y medición de las características fisicoquímicas de interés en el producto agrícola, *ii.* procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola, *iii.* estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes, *iv.* validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma. Se presentan algoritmos generales para procesar las imágenes y extraer características de color, textura, tamaño y forma. A lo largo de la presentación y sustentación de cada uno de las etapas se evidencian resultados de aplicar los algoritmos en algunos productos agrícolas. A continuación se presenta la validación de aplicar la metodología propuesta en diferentes productos agrícolas para la identificación de algunas de las características fisicoquímicas más útiles.

## 4. Validación de la metodología en casos de estudio.

Para validar la metodología propuesta, se utilizó ésta para la estimación de masa y volumen en algunos frutos. Se realizaron diferentes experimentos con este fin. Se analizaron cuatro productos agrícolas: Maracuyá (*P.Edulis*), Granadilla (*P.Ligularis*), Mango común (*M.IndicaL.*) y Tomate (*S.Lycopersicum*). Y, adicionalmente estimaron las características fisicoquímicas de mayor interés: masa, volumen, sólidos solubles, acidez y pH en frutos de Mango Nam doc mai. Para este fin, se extrajeron características de color, textura, tamaño y forma y se realizó la medición física de masa y volumen.

La masa fue medida utilizando una balanza digital con resolución de 1 [g], en cada uno de los cuatro casos estudiados. Debido a que los cuatro frutos pueden ser modelados matemáticamente como una superficie de revolución elipsoidal, utilizando las mediciones físicas realizadas en el fruto, correspondientes a la altura, ancho y profundidad del mismo; como parámetros a,b y c de la elipsoide. Las características químicas fueron medidas en laboratorio luego de la destrucción de los frutos. Los sólidos solubles totales se midieron utilizando un refractómetro, y la acidez fue determinada mediante un titulador automático por medio del cuál fue posible estimar el pH de la muestra.

Sin embargo, es de anotar que el cálculo de volumen fue realizado mediante un ajuste matemático y puede no corresponder en su totalidad al volumen real del objeto, pero ha sido considerado como base para el análisis. Para otro tipo de frutos, se debe considerar la superficie de revolución que más se ajuste al mismo y medir los parámetros necesarios en ella, para hacer el cálculo de volumen. En caso de tener una superficie muy irregular, se recomienda realizar la medición de volumen utilizando el principio de arquímedes.

## 4.1. Diseño experimental

Los cinco productos agrícolas fueron analizados justo después de la cosecha. Fue analizado un conjunto compuesto por 100 imágenes de maracuyá, 50 de mango común y granadilla, 32 de tomate, y 54 de mango nam doc mai. Para el procesamiento de imágenes se utilizó la transformación  $I_2$  del espacio de color OHTA, para segmentar la región de interés (los frutos) en la imagen. Adicionalmente, se aplicó una técnica de constancia de color conocida como Gray-world. Se determinó la región de interés en la imagen, en términos de la forma del contorno del fruto, respecto a una forma geométrica definible (círculo, esfera, elipse, etc.).

Se obtuvieron los parámetros de color en términos de los espacios de color RGB y CIELab. Características de textura mediante la matriz de co-ocurrencia de niveles de gris. Parámetros de tamaño tales como: área, perímetro y longitudes de ancho y alto en función de la forma geométrica que describe el contorno del fruto. Finalmente fueron extraídos los momentos de  $Hu$  como descriptores de forma. Lo anterior compone un espacio de 22 características por muestra: L, a, b, B, G, R, Energía, Contraste, Correlación, Homogeneidad, Entropía, Eje mayor, Eje menor, área, perímetro y los 7 momentos de  $Hu$ , como los definidos en el capítulo precedente.

El procesamiento de la imagen y la extracción de características se realizó de igual manera para todos los frutos, utilizando el mismo algoritmo. Se utilizaron herramientas de software libre, desarrollando un código bajo el sistema operativo Linux - Ubuntu 14.04 LTS, en el entorno Qt Creator 3.3.1 y el uso de las librerías de OpenCV 2.4.9 (Open Computer Vision), sirvió como base para el análisis realizado para todos los productos. Para el ADL se determinó el mejor sistema que reflejara la mejor discriminancia de los datos en función de la característica fisicoquímica medida.

Por último, se utilizó una red neuronal artificial (RNA) basando su estimación de pesos para cada neurona, en la mejor discriminancia lineal (ADL) de las características obtenidas. Dicha red neuronal artificial fue ajustada para cada uno de los frutos y características, ya que la curva de aprendizaje puede variar a partir del ADL realizado, en función del número de épocas o iteraciones de la red. A continuación, se evidencian los resultados obtenidos para cada uno de los experimentos realizados.

## 4.2. Caso de Estudio: Maracuyá - *P. edulis*.

### 4.2.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.

Para encontrar métodos específicos y útiles en cada una de las instancias de la metodología descrita; ésta, fue validada realizando un análisis experimental utilizando frutos de maracuyá - *P. Edulis*. Las características físicas de interés identificadas en este producto agrícola fueron: la masa, medida en gramos y el volumen, medido en  $cm^3$ .

Dichas características fueron escogidas, debido a que los atributos visibles en los frutos de maracuyá, como en la mayoría de los productos agrícolas, son de difícil correlación; requiriendo generalmente su medición en los procesos de intercambio (compra-venta). Esto se explica porque la masa y el volumen del fruto, puede depender tanto del epicarpio como del mesocarpio (grosor de su capa externa y contenido de producto útil en su interior).

Mediante la visión artificial los objetos en una escena fotografiada pueden ser parametrizados mediante atributos extraídos a partir de las imágenes. En el caso de los productos agrícolas, es posible parametrizar dichos atributos que generalmente son concluidos mediante la percepción de productores y consumidores.

Como se ha visto anteriormente, el color (Secciones 2.1.6, 2.2.1, 2.2.2, 2.2.3., 3.2.2. y 3.2.3), la textura (Secciones 2.1.2., 2.2.4. y 3.2.4.), el tamaño (2.1.4., 2.1.5., 2.2.5. y 3.2.5.) y la forma (Secciones 2.2.5. y 3.2.5.) del fruto; son algunos de los parámetros cualitativos que son utilizados para inferir sus propiedades fisicoquímicas y por ende tomar decisiones respecto al producto (aceptar o rechazar).

A partir de ésto, los atributos parametrizables en las imágenes de productos agrícolas, se centrarán de ahora en adelante en la extracción de dichas características. Las características físicas identificadas: los datos de masa [g] y volumen [ $cm^3$ ]; se midieron en un conjunto de 100 frutos de maracuyá. La masa fue medida utilizando una balanza digital con resolución de 1 [g]. Para el volumen, se analizó el producto, identificando que éste era posible calcularlo modelando los frutos como elipsoides mediante la Ecuación 4-1, donde  $a$ ,  $b$  y  $c$  son las mediciones físicas realizadas en los frutos, correspondientes a la altura, ancho y profundidad del mismo.

$$V = \frac{4\pi}{3} \times abc. \quad (4-1)$$

Los valores de masa y volumen (Apéndice **A-1**) son resumidos en las tablas de frecuencia **4-1** y **4-2**, donde se puede apreciar las clases realizadas para separar las mediciones, Límite inferior (L. inf), límite superior (L. sup), marca de clase (mi), frecuencia absoluta (ni), frecuencia absoluta acumulada (Ni), frecuencia relativa (fi), y frecuencia relativa acumulada

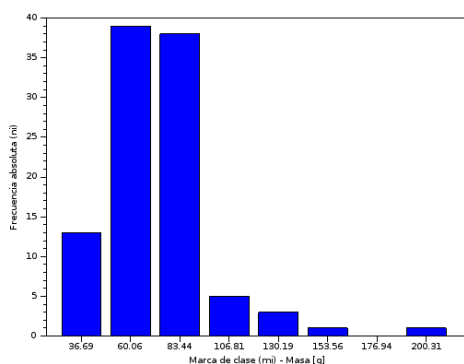
(Fi). El número de clases es definido a partir de la fórmula de Sturges ( $K$ ), correspondiente a  $K = 1 + 3.3\text{Log}(N)$  donde  $N$  es el número total de muestras. De allí, Figura 4-1, es posible inferir que el 95 % de los datos medidos de masa y volumen en maracuyá recién cosechada, se encuentra en el rango de 25.00 [g] a 118.50 [g] y 99.37 [ $\text{cm}^3$ ] a 231.64 [ $\text{cm}^3$ ]; respectivamente.

**Tabla 4-1.:** Tabla de frecuencias para la masa [g] de frutos de maracuyá.

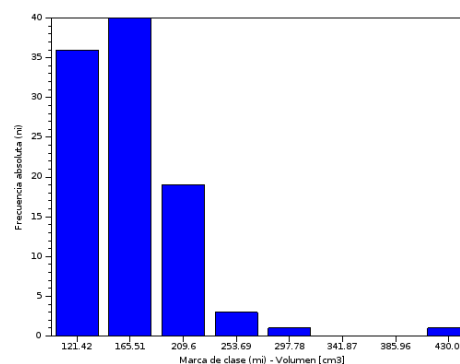
Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	25.00	48.38	36.69	13	13	0.13	0.13
2	48.38	71.75	60.06	39	52	0.39	0.52
3	71.75	95.13	83.44	38	90	0.38	0.9
4	95.13	118.50	106.81	5	95	0.05	0.95
5	118.50	141.88	130.19	3	98	0.03	0.98
6	141.88	165.25	153.56	1	99	0.01	0.99
7	165.25	188.63	176.94	0	99	0	0.99
8	188.63	212.00	200.31	1	100	0.01	1

**Tabla 4-2.:** Tabla de frecuencias para el volumen [ $\text{cm}^3$ ] de frutos de maracuyá.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	99.37	143.46	121.42	36	36	0.36	0.36
2	143.46	187.55	165.51	40	76	0.4	0.76
3	187.55	231.64	209.60	19	95	0.19	0.95
4	231.64	275.73	253.69	3	98	0.03	0.98
5	275.73	319.82	297.78	1	99	0.01	0.99
6	319.82	363.91	341.87	0	99	0	0.99
7	363.91	408.01	385.96	0	99	0	0.99
8	408.01	452.10	430.05	1	100	0.01	1



(a) mi Masa (g) vs ni.



(b) mi Volumen ( $\text{cm}^3$ ) vs ni.

**Figura 4-1.:** Gráfico de barras comparativo de masa y volumen en términos de frecuencia absoluta para maracuyá.



#### 4.2.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

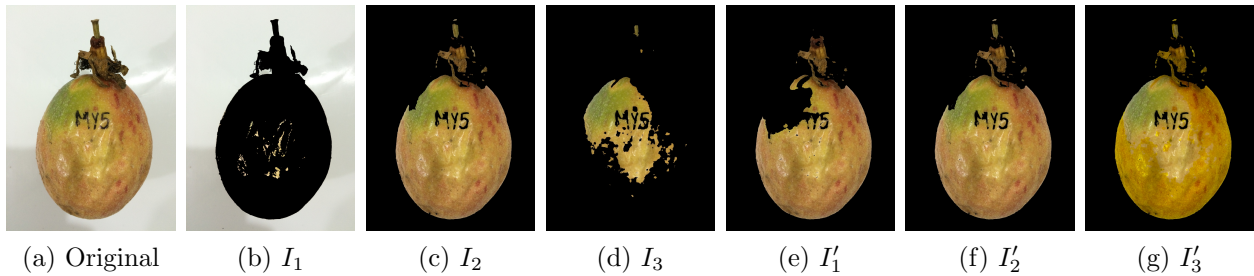
Aplicando la metodología para estimación de características fisicoquímicas en productos agrícolas, se analizó un conjunto de 100 frutos de maracuyá. Se obtuvieron imágenes en alta resolución a 8 Mpx (Tamaño imagen 4:3 (píxeles),  $2448 \times 3264$  píxeles) tales como la presentada en la Figura 4-2. Las imágenes fueron tomadas justo después de la cosecha, en una pequeña finca productora al norte del departamento del Huila en Colombia, Sur América.



**Figura 4-2.:** Imagen de maracuyá del banco de imágenes obtenido para el análisis.

El procesamiento de imágenes se refiere a la etapa de la visión artificial, en la cuál las imágenes obtenidas sufren cambios controlados con el objetivo de realzar algunos atributos o atenuar problemas en la adquisición de las mismas. Para realizar este procesamiento, se desarrollaron los códigos basados en los algoritmos de la metodología presentada, utilizando herramientas de software libre. Para lograr este fin se procedió, de la siguiente manera.

El banco de imágenes de maracuyá fue normalizado en color, empleando la técnica de constancia de color Gray-World [110]. A continuación, se aplicaron las transformaciones  $I_1$ ,  $I_2$ ,  $I_3$ ,  $I'_1$ ,  $I'_2$  e  $I'_3$ , presentadas anteriormente en la Ecuaciones 3-1, 3-2, 3-3, 3-5 y 3-6 del espacio de color OHTA [80], con el fin de determinar la transformación que permitiera segmentar el fruto de la imagen. Al finalizar la aplicación de cada una de las transformaciones, se concluye que; como se evidencia en la Figura 4-3, la transformación  $I_2$  es la más adecuada, como fue mostrado previamente en la metodología obteniendo una visibilidad de la región de interés superior al 95% y sin sufrir cambios en las componentes de color como  $I_3$ ; permitiendo así obtener la mayor información del fruto.



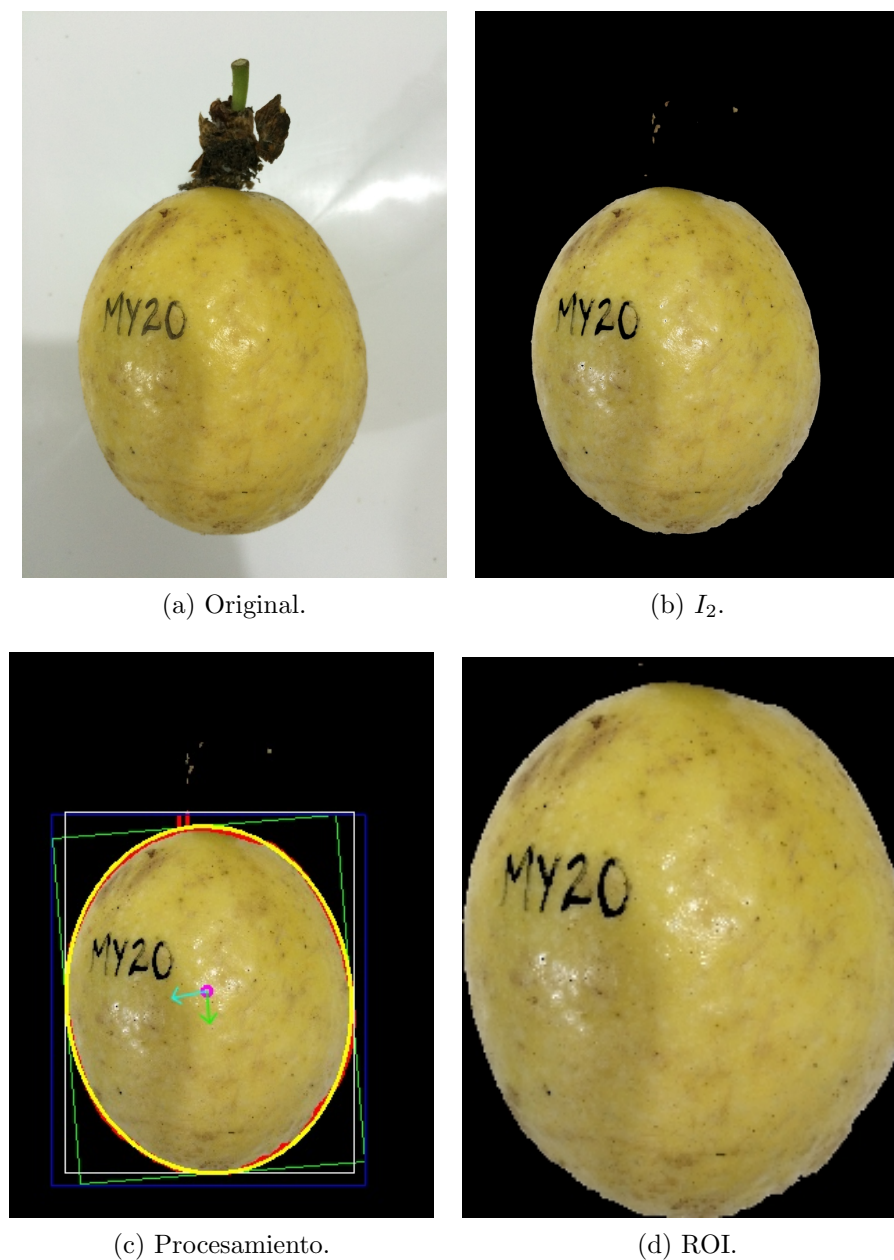
**Figura 4-3.:** Procesamiento OHTA realizado a una muestra del banco de imágenes de maracuyá.

Luego de obtener la imagen resultante  $I_2$  para el banco de imágenes; éstas fueron suavizadas utilizando un filtro de gauss ( $9 \times 9$ ). A continuación se aplicó una transformación morfológica del tipo cierre (dilatamiento seguido de erosión), utilizando un elemento estructurante tamaño 18. Ésto, con el fin de eliminar posibles cavidades dentro de la región de interés del fruto. A cada una de las imágenes resultantes, se le hallaron sus contornos. Si bien, el objeto de interés en la imagen es el fruto, no necesariamente puede ser el único contorno hallado; pero, obtener el máximo de ellos, garantizará la mayor proporción de información del fruto. Éste podrá de ser utilizado como base para los análisis posteriores.

Se determinó entonces, la elipse que mejor se ajusta a los puntos que definen dicho contorno, fijando como región de interés (ROI), el mejor rectángulo que los contenga, en la Figura 4-4 se muestra el resultado del procesamiento descrito. Por medio del algoritmo de Análisis por Componentes Principales (ACP), aplicado sobre el contorno definido según el proceso descrito anteriormente en cada una de las imágenes, fue posible determinar los ejes coordenados (alto y ancho) de la elipse ajustada al fruto en 2D, Figura 4-4c.

Luego de realizar el procesamiento de las imágenes y obtener la región de interés en cada una de ellas, se parametrizaron algunos atributos en la imagen del fruto, y relativos a los elementos parametrizables en imágenes identificados como: color, textura, tamaño y forma. Se realizó una transformación de color al espacio CIELab, y se separaron cada uno de los canales L, a y b, de manera independiente. Ésto con el fin de obtener el valor promedio de cada canal. Análogamente, se realizó el mismo análisis con la imagen RGB, separando de manera independiente cada uno de los canales R, G y B y obteniendo el promedio de los mismos sobre la totalidad de la imagen (6 características de color).

Para analizar la textura de la ROI del fruto en la imagen, se obtuvo la matriz de co-ocurrencia de niveles de gris (MCNG), a partir del conjunto de las imágenes en niveles de gris, obteniendo el promedio de las cuatro direcciones/ángulos ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ), para una distancia de desplazamiento de 1 píxel. De la matriz, se obtuvieron los valores de Entropía, Homogeneidad, Contraste, Energía y Correlación (5 características de textura).



**Figura 4-4.:** Procesamiento de imagen realizado a una muestra del banco de imágenes de maracuyá.

Las magnitudes de los ejes mayor y menor que definen la elipse circunscrita para el fruto en la región de interés, así como su área y perímetro (4 características), fueron utilizadas como medidas que describen el tamaño del fruto. Por último, se utilizaron los momentos de Hu como descriptores de forma en los frutos (7 características de forma). Lo anterior compone un total de 22 características para el banco de 100 imágenes de maracuyá.

### 4.2.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.

Con el fin de correlacionar los atributos extraídos de las imágenes con la masa y el volumen medidos físicamente, se utilizaron dos métodos de análisis estadístico, los cuales se presentarán a continuación. Para realizar un análisis a partir de la variabilidad y discriminancia de los datos mas no por su magnitud, la matriz de datos  $M_{Datos}$  ( $100 \times 22$ ) fue normalizada (entre 0 y 1), de manera independiente cada una de sus 22 características, utilizando la Ecuación 4-2, donde  $X_{i,j}$  es la característica  $j$ -ésima para el fruto  $i$ -ésimo en la matriz,  $Min_{(1,100),j}$  y  $Max_{(1,100),j}$  son los valores mínimo y máximo del conjunto de 100 frutos para cada una de las  $j$ -ésimas características.

Como se ha mostrado en la literatura [107, 86], realizar una regresión lineal múltiple (RLM), en un conjunto de datos lo suficientemente grande  $> 16$ , puede llegar a ser poco útil; ya que no necesariamente podría permitir una correlación entre los datos de la imagen y la información física medida. Por lo anterior se decidió realizar dos análisis previos a la correlación para  $M_{Datos}$ : Análisis por Componentes Principales (ACP) y Análisis Discriminante Lineal (ADL), con el fin de obtener la mayor variabilidad y mayor discriminancia de los datos; respectivamente.

$$\bar{X}_{i,j} = \frac{X_{i,j} - Min_{(1,100),j}}{Max_{(1,100),j} - Min_{(1,100),j}}. \quad (4-2)$$

El ACP se realizó sobre la matriz de datos normalizada [ $M_{Datos-norm}$ ], obteniendo la proyección de la misma, multiplicándola por la matriz de los vectores propios asociados a la los mayores valores propios que correspondan a más del 95%. Para el caso de estudio, los 8 valores propios de mayor magnitud, sumados obtienen: 0.798, correspondiente al 96.5% de la variabilidad de los datos. Utilizando los vectores propios asociados a éstos, se obtiene la matriz de proyección  $M_{ACP}$ . ( $100 \times 8$ ). Con este conjunto de datos se procedió entonces a realizar una RLM, entre esta última matriz y los datos de masa medidos previamente.

Para realizar la RLM de los valores de masa contenidos en el arreglo  $Y$ , se utilizó la Ecuación. 4-3, donde  $\vec{\alpha}$  será el resultado de los coeficientes  $\alpha_1, \alpha_2, \dots, \alpha_i$  asociados a cada una de las  $m$ -variables.  $X$  la matriz  $M_{ACP}$ . ampliada donde la primera columna estará compuesta de un vector de 1's con el fin de poder calcular el  $\alpha_0$  como término independiente; y  $\epsilon$  como el error estándar de la regresión.

$$Y = \begin{bmatrix} 79 \\ 71 \\ 59 \\ 122 \\ \cdot \\ \cdot \\ \cdot \\ 96 \\ 81 \end{bmatrix} = X\vec{\alpha} + \epsilon = \begin{bmatrix} 1 & -0.734957 & -0.851536 & \dots & -0.484906 \\ 1 & -0.458599 & -0.811877 & \dots & -0.607123 \\ 1 & -0.190305 & -0.837371 & \dots & -0.164688 \\ 1 & 0.112146 & -0.811659 & \dots & -0.461251 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & -0.492984 & -1.33633 & \dots & -0.489958 \\ 1 & -0.403307 & -1.23857 & \dots & -0.501886 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_7 \\ \alpha_8 \end{bmatrix} + \epsilon \quad (4-3)$$

El error estándar de la regresión  $\epsilon$ , fue calculado mediante la Ecuación 4-4. Donde  $N$  es el número de elementos utilizados para el análisis.  $Y_i$  los valores reales de masa y  $\hat{Y}_i$  los valores estimados mediante la regresión. Para el análisis RLM, las  $m$ -variables utilizadas corresponden a las 8 proyecciones obtenidas por ACP mediante combinaciones lineales de las 22 características obtenidas mediante imágenes. Se utilizaron 80 elementos del conjunto total para determinar la regresión, y los restantes 20, se utilizaron como conjunto de validación o test.

$$\epsilon = \sqrt{\frac{1}{N - m - 1} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}. \quad (4-4)$$

Resolviendo el sistema de RLM se calculan las matrices  $S$  y  $T$ , Ecuación 4-5. La primera se obtiene al multiplicar la matriz  $X = M_{ACP}$ , por su transpuesta. La segunda a partir de la interacción entre el valor medido real de la variable de interés (en el caso de estudio será la masa), por cada una de las  $m$ -variables, respectivamente, es decir,  $X$  transpuesta por  $Y$ . Estas matrices permitirán determinar las relaciones intra y entre cada una de las  $m$ -variables y el valor real, con el objetivo de poder determinar el vector de coeficientes  $\vec{\alpha}$ .

$$S = X^T X = \begin{bmatrix} 100 & 7.15 & -38.53 & 108.15 & \dots & \sum X_{i8} \\ 7.15 & 26.86 & \sum X_{i1}X_{i2} & \sum X_{i1}X_{i3} & \dots & \sum X_{i1}X_{i8} \\ -38.53 & \sum X_{i2}X_{i1} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} & \dots & \sum X_{i2}X_{i8} \\ 108.15 & \sum X_{i3}X_{i1} & \sum X_{i3}X_{i2} & \sum X_{i3}^2 & \dots & \sum X_{i3}X_{i8} \\ \cdot & \cdot & \cdot & \cdot & \sum X_{i4}^2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \sum X_{i8} & \sum X_{i8}X_{i1} & \sum X_{i8}X_{i2} & \sum X_{i8}X_{i3} & \dots & \sum X_{i8}^2 \end{bmatrix} T = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{i1} \\ \sum Y_i X_{i2} \\ \sum Y_i X_{i3} \\ \cdot \\ \cdot \\ \cdot \\ \sum Y_i X_{i7} \\ \sum Y_i X_{i8} \end{bmatrix} \quad (4-5)$$

La RLM para la correlación entre la masa medida y los datos obtenidos mediante imágenes, se obtienen los 9 coeficientes de la Ecuación 4-6. Éstos corresponden al término independiente  $\alpha_0$  y los coeficientes para las 8 variables utilizadas. Sin embargo, tras hacer el análisis de los estadísticos descriptivos de correlación múltiple y determinación, los cuales se encuentran resumidos en la Tabla 4-3 de la siguiente sección, se obtiene: 0.45 y 0.20 para el conjunto de entrenamiento y de 0.66 y 0.44 para el conjunto de prueba, la estimación de masa en gramos para el conjunto de entrenamiento contará con una efectividad del 20 % para el conjunto de prueba y de 44 % para el conjunto de entrenamiento lo cuál es significativamente bajo, teniendo en cuenta que se contará con un error de 24.18 [g] y 27.31 [g], respectivamente.

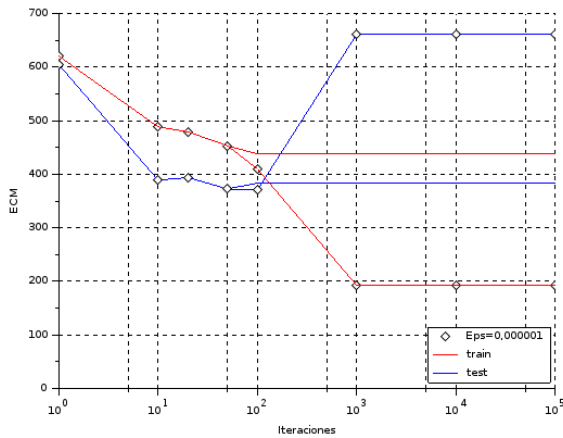
$$\vec{\alpha} = S^{-1}T = \begin{bmatrix} 28.66 \\ 0.19 \\ 18.47 \\ 12.26 \\ 24.73 \\ 6.38 \\ 11.34 \\ -30.44 \\ -58.88 \end{bmatrix} + \epsilon. \quad (4-6)$$

Con el fin de encontrar una herramienta de correlación automática, con una expectativa de mejores resultados respecto a su comparación con la variabilidad de los datos y la correlación lineal, se utilizaron redes neuronales artificiales (RNA), como herramienta de estimación. Si bien, las redes neuronales artificiales, son utilizadas ampliamente para clasificación de patrones, éstas pueden ser usadas con el objetivo de predecir determinado comportamiento específico (para el caso, las variables físico-químicas) en función de parámetros (atributos obtenidos mediante las imágenes).

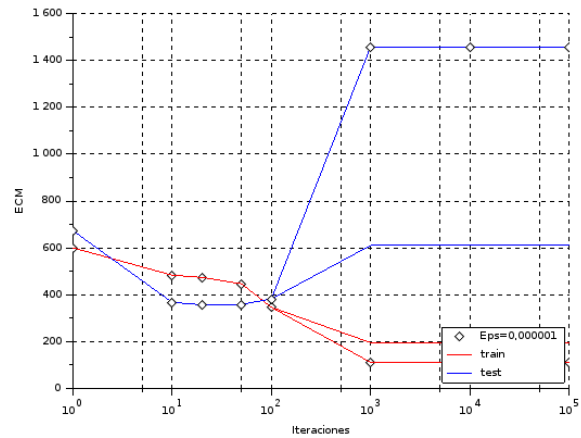
Un problema común al usar las RNA como herramienta de predicción, radica en determinar la curva de aprendizaje en función del error con que se estiman los pesos de cada neurona, el número de iteraciones o épocas por medio de las cuáles la red neuronal converge al menor error cuadrático medio, y obviamente a la cantidad de neuronas utilizadas en la capa oculta de la red. En la Figura 4-5 es posible observar el análisis de este procedimiento, conocido comúnmente como la curva de aprendizaje de la red neuronal.

La arquitectura de red propuesta, obedece al uso de las 8 proyecciones obtenidas mediante el análisis ACP y contenidos en la matriz  $M_{ACP}$  ( $100 \times 8$ ). Utilizando un sistema heurístico para determinar el número óptimo de neuronas en la capa oculta, se utilizó la ley de los 2/3, la cual reza que el número adecuado de neuronas en la capa oculta estará definido como los 2/3 de la suma de la cantidad de neuronas de entrada y salida. Para el caso de estudio, se tienen 8 neuronas en la capa de entrada y 1 neurona en la capa de salida, correspondiente a la predicción de la masa de los frutos.

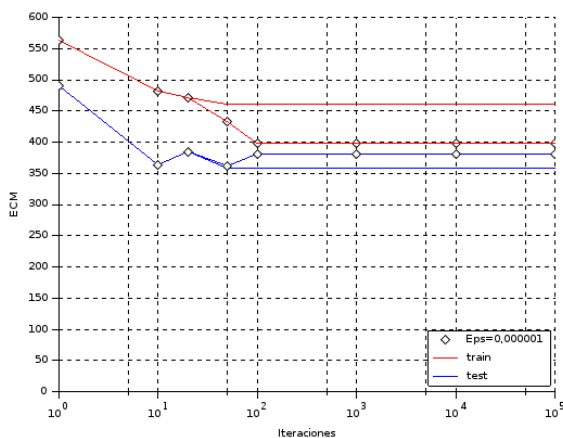
Para la implementación computacional de la RNA, se utilizaron las 8 variables de la matriz resultante del ACP  $M_{ACP}$  que usó para su análisis la matriz de datos normalizada  $M_{Datos-norm}$ . Para el caso de la neurona de salida, que en este caso hace referencia a la variable medida masa [g], los valores fueron normalizados entre -1 y 1, debido a que el ajuste de los pesos se realiza en la RNA mediante una función sigmoïdal; la cuál finalmente determinará la salida de la estimación o predicción, encontrándose en este rango los valores de salida.



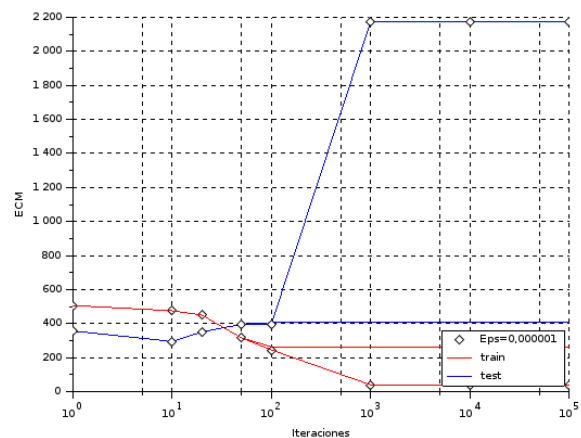
(a) 5 Neuronas en la capa oculta.



(b) 6 Neuronas en la capa oculta.



(c) 7 Neuronas en la capa oculta.



(d) 19 Neuronas en la capa oculta.

**Figura 4-5.:** Curva de aprendizaje para la RNA luego del ACP; Iteraciones (Épocas) vs. E.C.M. (Error cuadrático medio)

#### 4.2.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.

Se realizaron cuatro predicciones de masa, mediante los diferentes métodos de análisis propuestos. RLM después de realizar ACP y ADL, y RNA después de realizar ACP y ADL. En la Tabla 4-3, se pueden observar los resultados en términos de los descriptores estadísticos después de realizar cada una de las estimaciones. Allí es posible apreciar que el sistema compuesto por una red neuronal artificial basada en un análisis discriminante lineal, obtiene los mejores resultados para la estimación de masa, con un coeficiente de correlación de 0.77 y 0.83 para el conjunto de imágenes de entrenamiento y prueba, respectivamente; con un error típico de 16.59 y 15.97 gramos, respectivamente.

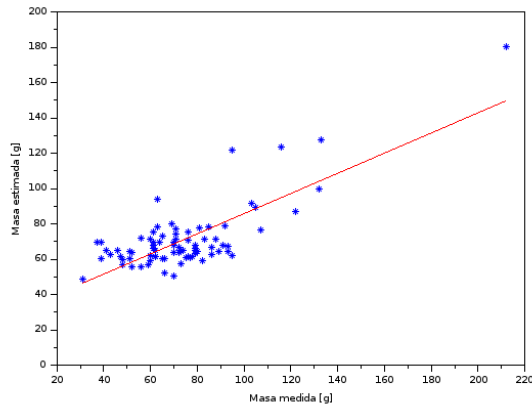
**Tabla 4-3.:** Análisis estadístico de correlación entre los datos de visión y la masa de frutos de maracuyá.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RLM	0.45	0.66	0.20	0.44	0.12	0.03	24.18	27.31
ADL + RLM	0.65	0.81	0.43	0.66	0.40	0.57	19.87	21.03
ACP + RNA	0.53	0.76	0.29	0.59	0.21	0.29	23.14	23.16
<b>ADL + RNA</b>	<b>0.77</b>	<b>0.83</b>	<b>0.60</b>	<b>0.69</b>	<b>0.59</b>	<b>0.65</b>	<b>16.59</b>	<b>15.97</b>

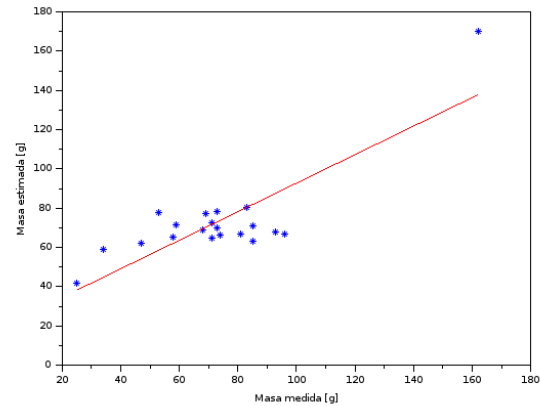
A partir de los resultados observados para la estimación de la masa, se implementó un sistema de RNA a partir de un ADL, para la estimación de volumen en las imágenes de maracuyá. Encontrando que la estimación para ésta característica, responde de manera similar al desarrollado para la masa. En la Figura 4-6 se puede apreciar la respuesta en términos de la comparación medición vs. estimación, para el caso de la masa. En la Figura 4-7 se puede observar una buena correlación entre la información medida y estimada para el volumen, obteniendo un coeficiente de correlación de 0.75 y 0.73 para el conjunto de imágenes de entrenamiento y prueba, respectivamente; con un error típico de 34.44 y 31.58  $cm^3$ , respectivamente.

De los anteriores resultados en maracuyá, es posible inferir que un sistema de visión de máquina realizado a partir de la implementación de los siguientes métodos, puede determinar una buena correlación para la estimación de al menos, la masa y el volumen. Inicialmente obtener un banco de imágenes lo suficientemente grande para poder obtener las suficientes discriminaciones entre las características del fruto y la variable de interés. Medir la(s) característica(s) física(s) o química(s), para el caso masa o volumen. Y, realizar un procesamiento de imágenes basando en  $I_2$  de la transformación de color OHTA.

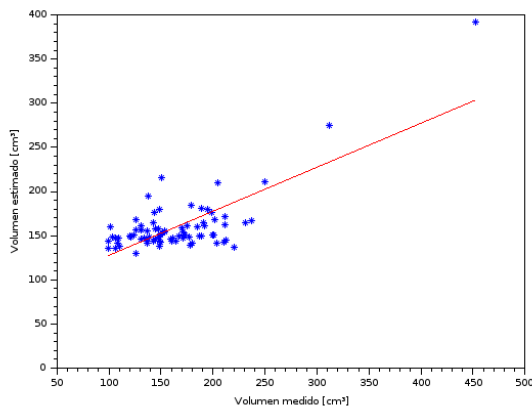




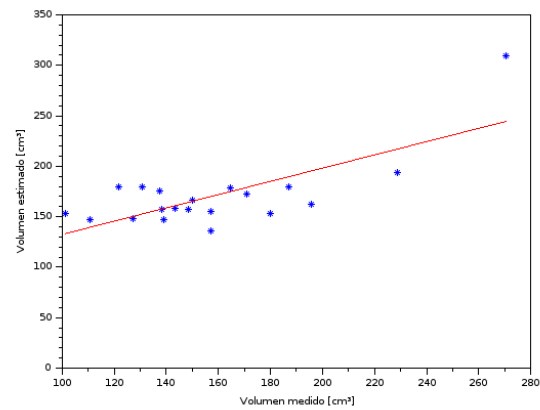
(a) Conjunto de entrenamiento.



(b) Conjunto de prueba.

**Figura 4-6.:** Masa medida [g] vs. Masa estimada [g] utilizando una RNA luego de aplicar un ADL en maracuyá

(a) Conjunto de entrenamiento.



(b) Conjunto de prueba.

**Figura 4-7.:** Volumen medido [ $cm^3$ ] vs. Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en maracuyá

Implementar un sistema de constancia de color para poder garantizar repetibilidad en la obtención de los parámetros de color. Determinar la región de interés en la imagen, es decir el fruto, en función de la forma del mismo. Obtener los parámetros de color, textura, tamaño y forma, pueden ayudar a describir los aspectos visibles del fruto. Finalmente, para obtener el mejor sistema de estimación de características es recomendable utilizar una RNA basando su estimación de pesos para cada neurona, en la mejor discriminancia lineal a partir de realizar combinaciones lineales de las características normalizadas obtenidas. Lo anterior, debido a que comparando el sistema de RNA contra la RLM, los resultados suelen ser mayores debido al procesamiento computacional de la primera.

### 4.3. Caso de Estudio: Granadilla - *P. Ligularis*.

#### 4.3.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.

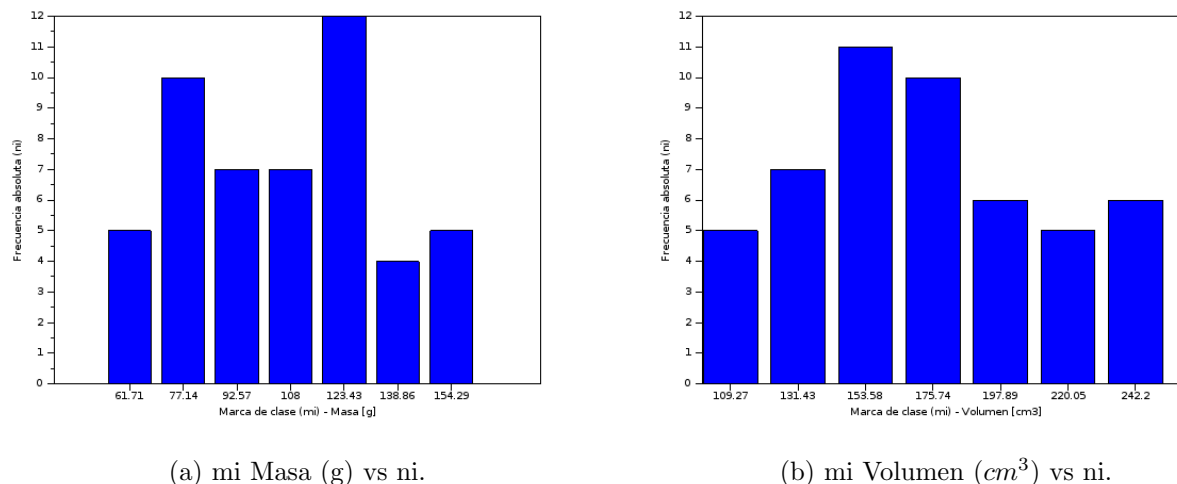
Para otros productos agrícola: Granadilla (*P. Ligularis*), se utilizó una muestra de 50 frutos de granadilla analizados justo después de la cosecha. Fueron obtenidas imágenes a resolución de 2 Mpx (Tamaño imagen 4:3 (píxeles), 1080×1920 píxeles). Se recopiló un banco de imágenes correspondiente al mismo número de frutos, y se midió su masa utilizando una balanza (Tabla 4-4 y Figura 4-8a) y las características de dimensionamiento (ancho, alto y profundo) mediante un calibrador pie de rey para el cálculo de volumen (Tabla 4-5 y Figura 4-8a Figura 4-8b).

**Tabla 4-4.:** Tabla de frecuencias para la masa [g] de frutos de granadilla.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	54.00	69.43	61.71	5	5	0.1	0.1
2	69.43	84.86	77.14	10	15	0.2	0.3
3	84.86	100.29	92.57	7	22	0.14	0.44
4	100.29	115.71	108.00	7	29	0.14	0.58
5	115.71	131.14	123.43	12	41	0.24	0.82
6	131.14	146.57	138.86	4	45	0.08	0.9
7	146.57	162.00	154.29	5	50	0.1	1

**Tabla 4-5.:** Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de granadilla.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	98.19	120.35	109.27	5	5	0.1	0.1
2	120.35	142.50	131.43	7	12	0.14	0.24
3	142.50	164.66	153.58	11	23	0.22	0.46
4	164.66	186.81	175.74	10	33	0.2	0.66
5	186.81	208.97	197.89	6	39	0.12	0.78
6	208.97	231.12	220.05	5	44	0.1	0.88
7	231.12	253.28	242.20	6	50	0.12	1

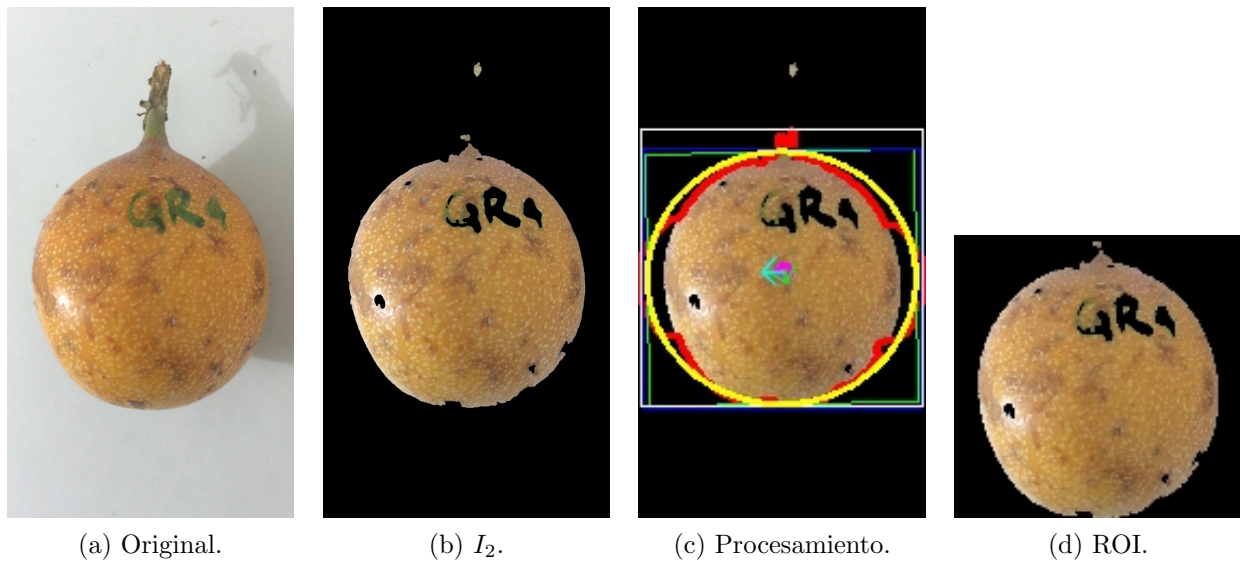


**Figura 4-8.:** Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para granadilla.

El volumen fue estimado de manera análoga al maracuyá, aproximando el fruto a una elipsoide. Es de anotar que a pesar de que este fruto tiene una perspectiva esférica sin embargo, pueden existir ligeras variaciones entre cada uno de los ejes del fruto. Sin embargo, si fueran iguales, el cálculo del volumen sería válido ya que la esfera puede ser considerada como un caso particular del elipsoide, en el cuál todos sus ejes tienen la misma magnitud. Para realizar la estimación de masa y volumen, la matriz de datos obtenida fue normalizada entre 0 y 1, incluídos los valores medidos físicamente (Apéndice A-3).

#### 4.3.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

Al total del banco de imágenes fue aplicado el mismo procesamiento de imagen descrito en la sección anterior: constancia de color, transformación  $I_2$  OHTA y ajuste de elipse al contorno del fruto para extracción de ésta como la región de interés a analizar, ver Figura 4-9. Posteriormente fueron extraídas las características de color, canales R, G y B de manera independiente, así como L, a y b. Se obtuvieron los parámetros de textura, entropía, homogeneidad, contraste, energía y correlación. La longitud de los ejes mayores y menores del fruto, así como su área y perímetro y finalmente los 7 momentos de Hu.



**Figura 4-9.:** Procesamiento de imagen realizado a una muestra del banco de imágenes de granadilla.

### 4.3.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.

Al normalizar los valores medidos, es posible realizar el ADL utilizando la magnitud entre 0 y 1, para determinar la mayor masa o volumen. Es decir, un valor cercano a 0, para una determinada imagen, corresponderá a una masa baja dentro del conjunto datos y viceversa. Para el análisis de masa, por ejemplo, las mediciones obtenidas se encuentran entre 54 y 162 gramos para el total de 50 frutos, con una media de 106.54 gramos y una desviación estándar de 29.04 gramos. En este caso el valor de 0 corresponderá a los frutos muy cercanos a 54 gramos y 1 corresponderá a los cercanos a 162 gramos.

El ADL arroja como resultados una matriz de valores y vectores propios, correspondientes a la mejor discriminancia en función de la masa medida. Para este caso se obtuvieron 10 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22. Utilizando los vectores propios asociados a los seis mayores valores propios, es posible obtener el 96.5% de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 6 vectores se utilizará como insumo de entrada a la red neuronal.

A partir de lo anterior, la RNA contó con 6 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de masa o volumen. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge en 100 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

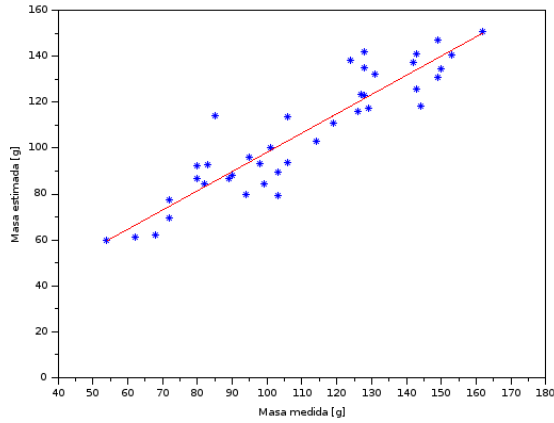
Para el análisis de volumen, se realizó el mismo proceso. Las mediciones obtenidas se encuentran entre  $98.19 \text{ cm}^3$  y  $253.28 \text{ cm}^3$  para el total de 50 frutos, con una media de  $173.88 \text{ cm}^3$  y una desviación estándar de  $40.58 \text{ cm}^3$ . En este caso el valor de 0 corresponderá a los frutos muy cercanos a  $98.19 \text{ cm}^3$ s y 1 corresponderá a los cercanos a  $253.28 \text{ cm}^3$ . El ADL arrojó en este caso 10 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22. Utilizando los vectores propios asociados a los seis mayores valores propios, es posible obtener el 96.6 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 6 vectores se utilizará como insumo de entrada a la red neuronal.

En este caso, la RNA contó con 6 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de volumen. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge en 10 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

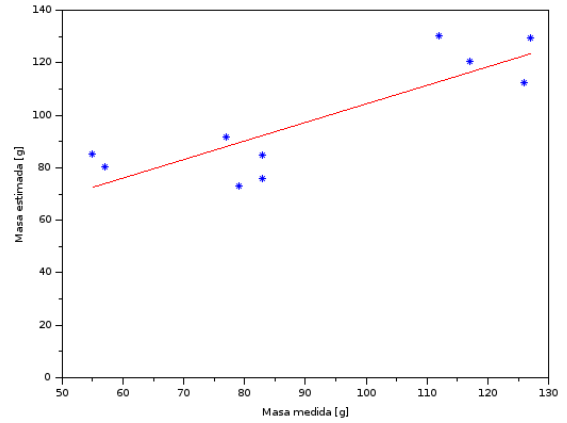
#### **4.3.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.**

Definiendo los anteriores parámetros, fue posible determinar la mejor cantidad de neuronas en la capa oculta para el análisis de masa; determinando que 5 neuronas para este caso, es la que menor error cuadrático medio presenta. Teniendo en cuenta la media de 106.54 gramos, y la desviación estándar de 29.04 gramos, la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. En la Figura 4-10 se puede apreciar, los resultados en términos de correlación para la masa medida y estimada. Obteniendo un coeficiente de correlación de 0.92 y 0.85 respectivamente para cada conjunto, con un error típico de 16.98 y 11.91 gramos, respectivamente.

Para el caso del volumen, fue posible determinar la mejor cantidad de neuronas en la capa oculta para la estimación; determinando que 10 neuronas para este caso, es la que menor error cuadrático medio presenta. Teniendo en cuenta la media de  $173.88 \text{ cm}^3$ , y la desviación estándar de  $40.58 \text{ cm}^3$ , la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. Como se puede apreciar en la Figura 4-11, los resultados en términos de correlación para el volumen medido y estimado son muy buenos. Obteniendo un coeficiente de correlación de 0.87 y 0.76 respectivamente para cada conjunto, con un error típico de 26.27 y  $20.84 \text{ cm}^3$ , respectivamente.

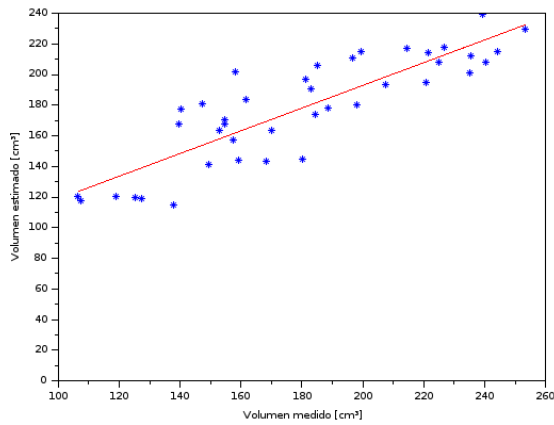


(a) Conjunto de entrenamiento.

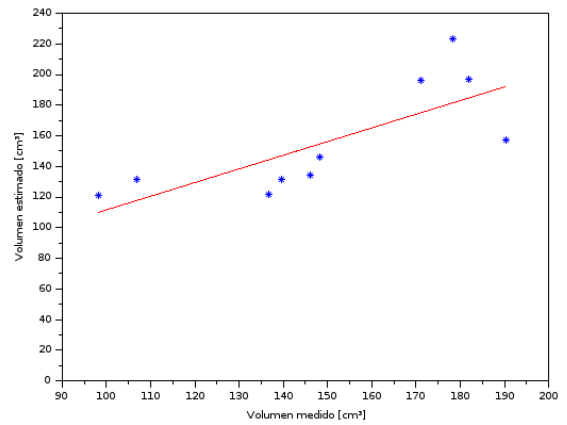


(b) Conjunto de prueba.

**Figura 4-10.:** Masa medida [g] vs. Masa estimada [g] utilizando una RNA luego de aplicar un ADL en granadilla.



(a) Conjunto de entrenamiento.



(b) Conjunto de prueba.

**Figura 4-11.:** Volumen medido [ $cm^3$ ] vs Volumen estimado [ $cm^3$ ] utilizando una RNA luego de aplicar un ADL en granadilla.

## 4.4. Caso de Estudio: Mango Común - M. Indica L.

### 4.4.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.

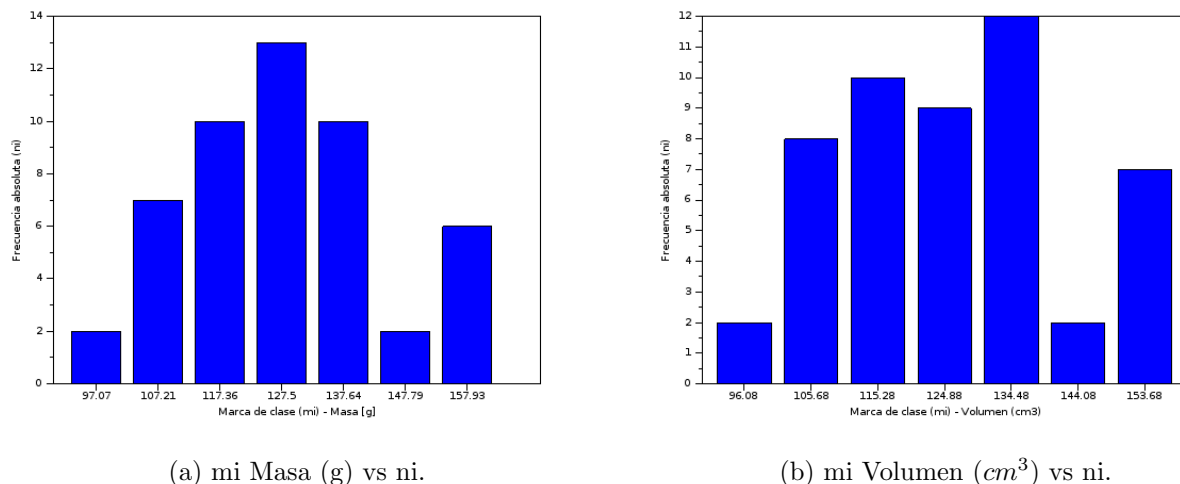
Para este experimento, se utilizaron de igual forma 50 frutos de mango común analizados justo después de la cosecha. Se compiló un banco de imágenes correspondiente al mismo número de frutos, y se midió su masa y características de dimensionamiento (ancho, alto y profundo), de manera análoga a lo realizado con la granadilla (Apéndice A-2). El volumen fue calculado nuevamente, aproximando el fruto a una elipsoide, obteniendo mediciones entre  $91.28 \text{ cm}^3$  y  $158.47 \text{ cm}^3$ , con una media de  $126.21 \text{ cm}^3$  y una desviación estándar de  $16.52 \text{ cm}^3$  (Tabla 4-7 y Figura 4-12b). Para el análisis de masa, las mediciones obtenidas se encuentran entre 92 y 163 gramos para el total de 50 frutos, con una media de 128.58 gramos y una desviación estándar de 16.43 gramos (Tabla 4-6 y Figura 4-12a).

**Tabla 4-6.:** Tabla de frecuencias para la masa [g] de frutos de mango común.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	92.00	102.14	97.07	2	2	0.04	0.04
2	102.14	112.29	107.21	7	9	0.14	0.18
3	112.29	122.43	117.36	10	19	0.2	0.38
4	122.43	132.57	127.50	13	32	0.26	0.64
5	132.57	142.71	137.64	10	42	0.2	0.84
6	142.71	152.86	147.79	2	44	0.04	0.88
7	152.86	163.00	157.93	6	50	0.12	1

**Tabla 4-7.:** Tabla de frecuencias para el volumen [ $\text{cm}^3$ ] de frutos de mango común.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	91.28	100.88	96.08	2	2	0.04	0.04
2	100.88	110.48	105.68	8	10	0.16	0.2
3	110.48	120.08	115.28	10	20	0.2	0.4
4	120.08	129.68	124.88	9	29	0.18	0.58
5	129.68	139.28	134.48	12	41	0.24	0.82
6	139.28	148.88	144.08	2	43	0.04	0.86
7	148.88	158.48	153.68	7	50	0.14	1



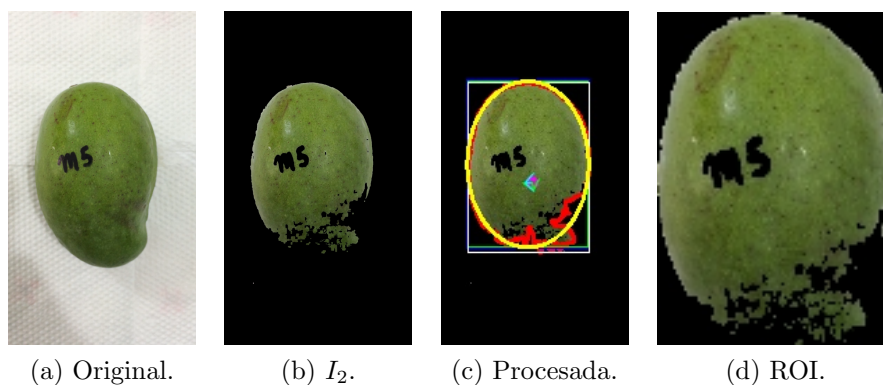
(a) mi Masa (g) vs ni.

(b) mi Volumen ( $cm^3$ ) vs ni.

**Figura 4-12.:** Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para mango común.

#### 4.4.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

Al total del banco de imágenes le fue aplicado el mismo procesamiento de imagen aplicado a granadilla y maracuyá, descrito y mostrado ya en detalle previamente, Figura 4-13. Se realizó un proceso de constancia de color, transformación  $I_2$  OHTA y ajuste de elipse al contorno del fruto para extracción de ésta como la región de interés a analizar. Posteriormente, fueron extraídas las características de color, canales R, G y B de manera independiente así como L, a y b. Se obtuvieron los parámetros de textura, entropía, homogeneidad, contraste, energía y correlación. La longitud de los ejes mayores y menores del fruto, así como su área y perímetro y finalmente los 7 momentos de Hu.



(a) Original.

(b)  $I_2$ .

(c) Procesada.

(d) ROI.

**Figura 4-13.:** Procesamiento de imagen realizado a una muestra del banco de imágenes de mango común.



#### **4.4.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.**

Para la masa, el ADL arrojó como resultados una matriz de valores y vectores propios, correspondientes a la mejor discriminancia en función de la medición realizada. Para este caso se obtuvieron igualmente, 10 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22.

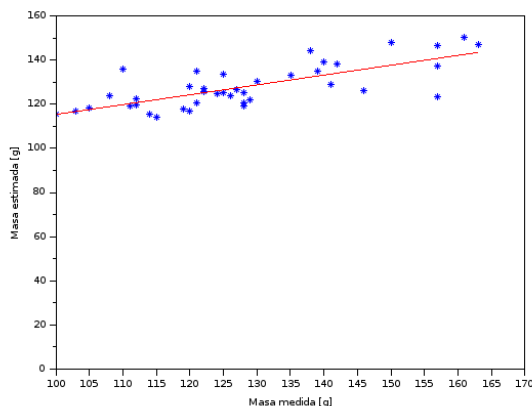
Utilizando los vectores propios asociados a los siete mayores valores propios, es posible obtener el 97 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 7 vectores se utilizó como insumo de entrada a la red neuronal. A partir de lo anterior, la RNA contó con 7 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de masa. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge entre 0 y 10 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

En el caso del volumen, se realizó un proceso similar. El ADL arrojó en este caso 10 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22. Utilizando los vectores propios asociados a los siete mayores valores propios, es posible obtener el 97.1 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 7 vectores se utilizara como insumo de entrada a la red neuronal.

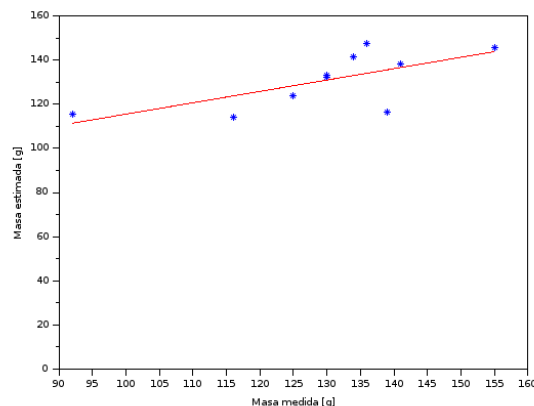
La RNA contó entonces con 7 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de volumen. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge en 10 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

#### **4.4.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.**

Fue posible entonces, determinar la mejor cantidad de neuronas en la capa oculta; encontrando que 11 neuronas para el caso de la masa es la que menor error cuadrático medio presenta. Teniendo en cuenta la media de 128.58 gramos, y la desviación estándar de 16.43 gramos, la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. Como se puede apreciar en la Figura 4-14, los resultados en términos de correlación para la masa medida y estimada son buenos. Obteniendo un coeficiente de correlación de 0.74 y 0.69 respectivamente para cada conjunto, con un error típico de 13.08 y 11.45 gramos, respectivamente.



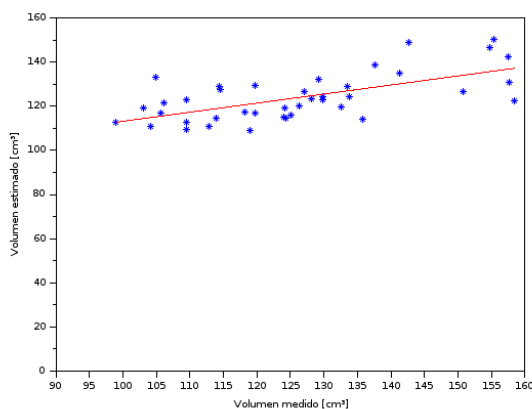
(a) Conjunto de entrenamiento.



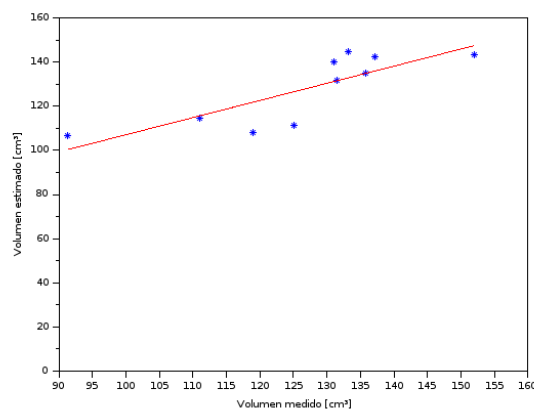
(b) Conjunto de prueba.

**Figura 4-14.:** Masa medida [g] vs Masa estimada [g] utilizando una RNA luego de aplicar un ADL en mango común.

Para el caso del volumen se determinó que 13 neuronas para este caso, es la que menor error cuadrático medio presenta. Teniendo en cuenta la media de  $126.21 \text{ cm}^3$ , y la desviación estándar de  $16.52 \text{ cm}^3$ , la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. Como se puede apreciar en la Figura 4-15, los resultados en términos de correlación para el volumen medido y estimado, mediante el coeficiente de correlación de 0.64 y 0.81 respectivamente para cada conjunto, con un error típico de  $13.20$  y  $10.48 \text{ cm}^3$ , respectivamente.



(a) Conjunto de entrenamiento



(b) Conjunto de prueba

**Figura 4-15.:** Volumen medido [ $\text{cm}^3$ ] vs. Volumen estimado [ $\text{cm}^3$ ] utilizando una RNA luego de aplicar un ADL en mango común

## 4.5. Caso de Estudio: Tomate - *S. lycopersicum*.

### 4.5.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.

En el tomate, se realizó una experimentación diferente, con el objetivo de buscar mejores resultados en términos de la estimación utilizando el mismo conjunto de características. Basados en el análisis de color propuesto por [72], se analizaron 9 frutos de tomate, adquiriendo imágenes y midiendo paralelamente su masa (Tabla 4-8 y Figura 4-16a) y dimensiones (ancho, alto y profundo) que sirvieran como estimadores del volumen (Apéndice A-4).

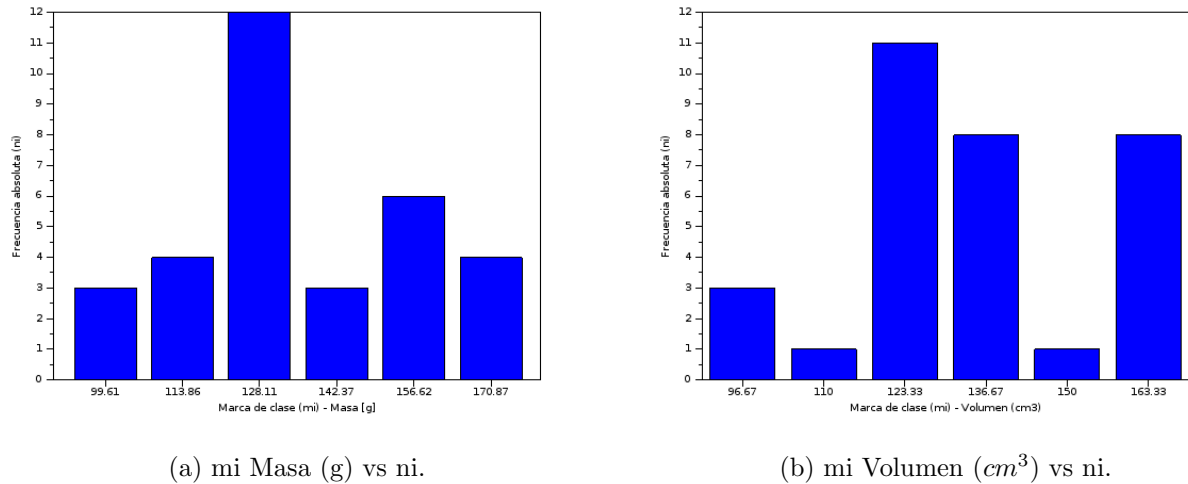
El volumen (Tabla 4-9 y Figura 4-16b) fue estimado de manera análoga a lo realizado en maracuyá, granadilla y mango común, aproximando el fruto a una elipsoide. Nótese que este fruto tiene una perspectiva esférica, pero puede ser generalizado como un elipsoide en el cuál todos sus ejes tienen la misma magnitud. Para realizar la estimación de masa y volumen, la matriz de datos obtenida fue normalizada entre 0 y 1, incluídos los valores medidos físicamente.

**Tabla 4-8.:** Tabla de frecuencias para la masa [g] de frutos de tomate.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	92.48	106.73	99.61	3	3	0.09	0.09
2	106.73	120.99	113.86	4	7	0.12	0.22
3	120.99	135.24	128.11	12	19	0.37	0.59
4	135.24	149.49	142.37	3	22	0.09	0.68
5	149.49	163.75	156.62	6	28	0.18	0.87
6	163.75	178.00	170.87	4	32	0.12	1

**Tabla 4-9.:** Tabla de frecuencias para el volumen [ $cm^3$ ] de frutos de tomate.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	90.00	103.33	96.67	3	3	0.09	0.09
2	103.33	116.67	110.00	1	4	0.03	0.12
3	116.67	130.00	123.33	11	15	0.34	0.46
4	130.00	143.33	136.67	8	23	0.25	0.71
5	143.33	156.67	150.00	1	24	0.03	0.75
6	156.67	170.00	163.33	8	32	0.25	1



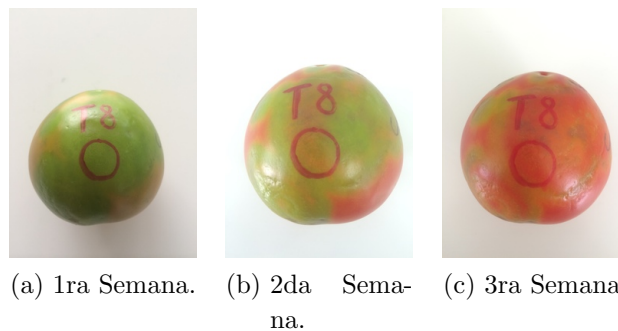
(a) mi Masa (g) vs ni.

(b) mi Volumen ( $cm^3$ ) vs ni.

**Figura 4-16.:** Gráfico de barras para la comparación de masa -izquierda- y volumen -derecha- en términos de su frecuencia absoluta para tomate.

#### 4.5.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

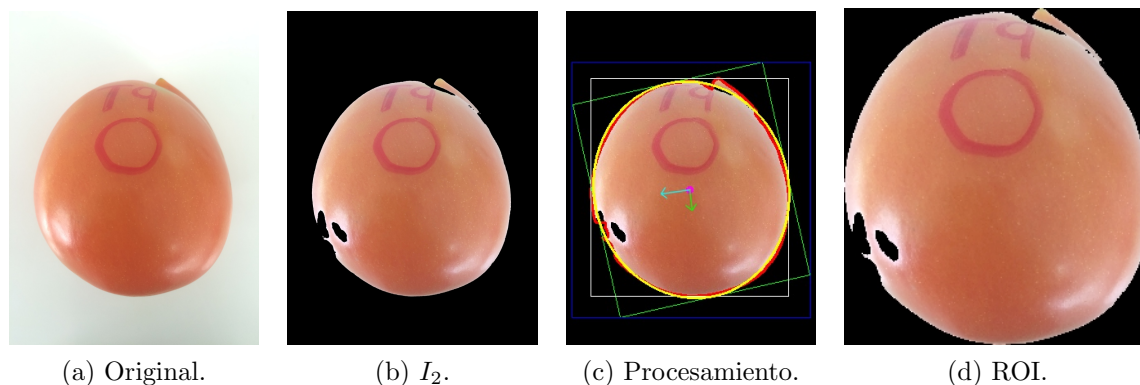
Se utilizó una muestra de 14 frutos de tomate analizados justo después de la cosecha. Se analizaron los primeros 9 frutos para el desarrollo del sistema de visión de máquina a partir de sus correspondientes 9 mediciones por cada semana durante 3 semanas, es decir 27 mediciones. Se compiló un banco de imágenes durante el proceso, midiendo también su masa y características de dimensionamiento (ancho, alto y profundo) cada semana, en la Figura 4-17, se puede apreciar los cambios en un fruto durante cada una de las semanas. Al término del ajuste del sistema, se capturaron 5 frutos nuevos, midiendo su masa y volumen para corroborar el resultado de la estimación.



(a) 1ra Semana. (b) 2da Sema-  
na. (c) 3ra Semana.

**Figura 4-17.:** Evolución de un fruto de tomate evidenciado en imágenes.

Al total del banco de 32 imágenes le fue aplicado el procesamiento de imagen descrito y aplicado para los productos anteriormente analizados (maracuyá, granadilla y mango común), ver Figura 4-18. Para este análisis de masa, las mediciones obtenidas se encuentran entre 92.48 y 178.00 gramos para el total de 32 mediciones realizadas sobre los 9 frutos, con una media de 135.45 gramos y una desviación estándar de 22.26 gramos.



**Figura 4-18.:** Procesamiento de imagen realizado a una muestra del banco de imágenes de tomate.

### 4.5.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.

Para la masa, el ADL arrojó como resultados una matriz de valores y vectores propios, correspondientes a la mejor discriminancia en función de la medición realizada. Para este caso se obtuvieron nuevamente, 5 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22.

Utilizando los vectores propios asociados a los tres mayores valores propios, es posible obtener el 94 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 3 vectores se utilizó como insumo de entrada a la red neuronal. A partir de lo anterior, la RNA contó con 3 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de masa. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge entre 50 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

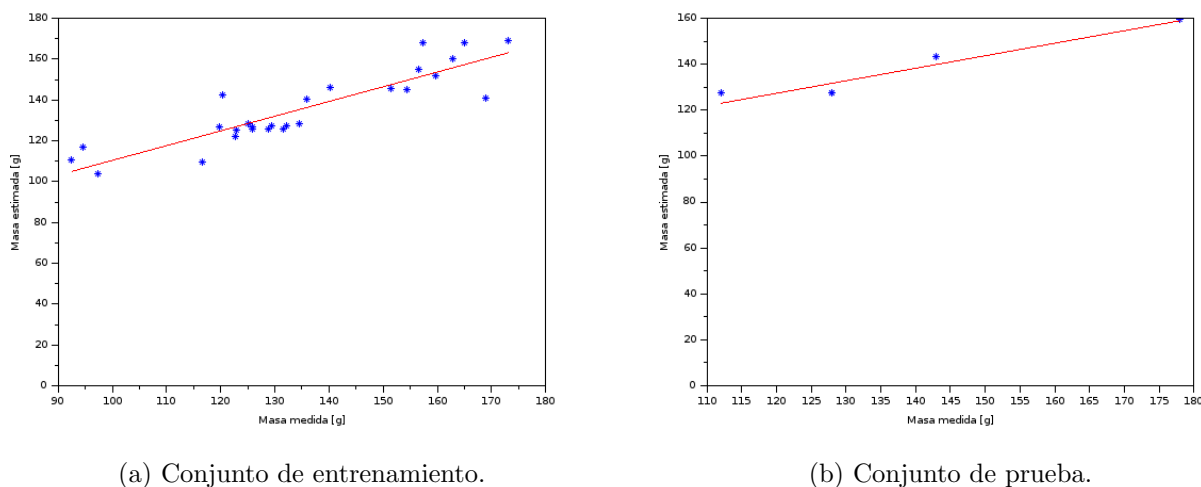
Para el análisis de volumen, se realizó el mismo proceso. Las mediciones obtenidas se encuentran entre  $90 \text{ cm}^3$  y  $170 \text{ cm}^3$  para el total de 32 mediciones realizadas, con una media de  $133.70 \text{ cm}^3$  y una desviación estándar de  $21.75 \text{ cm}^3$ . El ADL arrojó en este caso 5 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22.

Utilizando los vectores propios asociados a los tres mayores valores propios, es posible obtener el 97 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 3 vectores se utiliza como entrada a la red neuronal.

Según lo anterior, la RNA contó con 3 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de volumen. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge en 500 iteraciones o épocas, encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento.

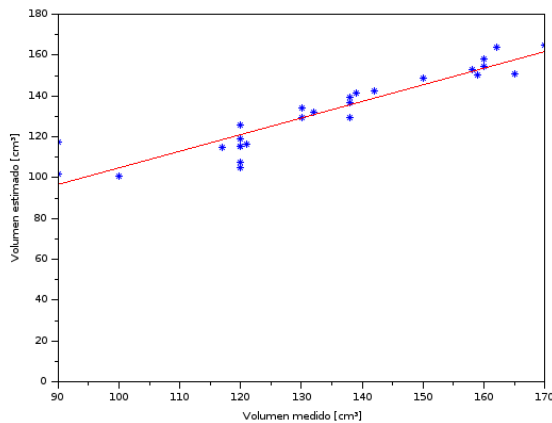
#### 4.5.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.

La RNA para la estimación de masa, arrojó que la mejor cantidad de neuronas en la capa oculta; es 8 neuronas para este caso, en la que menor error cuadrático medio presenta. Teniendo en cuenta la media de 135.45 gramos, y la desviación estándar de 22.26 gramos, la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. Como se puede apreciar en la Figura 4-19, los resultados en términos de correlación para la masa medida y estimada son muy buenos. Obteniendo un coeficiente de correlación de 0.89 y 0.95 respectivamente para cada conjunto, con un error típico de 14.03 y 10.52 gramos, respectivamente.

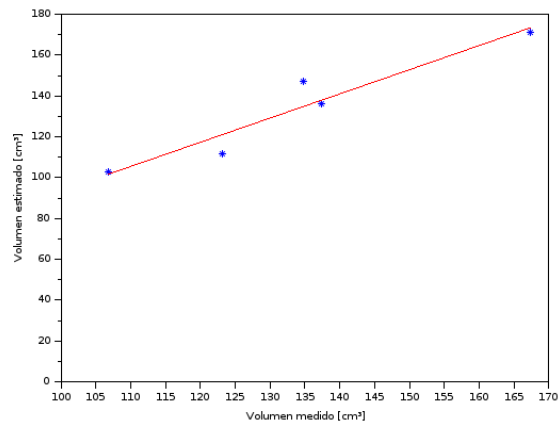


**Figura 4-19.:** Masa medida [g] vs Masa estimada [g] utilizando una RNA luego de aplicar un ADL en tomate.

Para el volumen, fue posible determinar la mejor cantidad de neuronas en la capa oculta para la estimación; determinando que 10 neuronas para este caso, es la que menor error cuadrático medio presenta. Teniendo en cuenta la media de  $133.70 \text{ cm}^3$ , y la desviación estándar de  $21.75 \text{ cm}^3$ , la estimación se espera encontrar por lo menos con error típico menor a la desviación estándar de los datos, y un coeficiente de correlación lo más cercano a 1 posible. Como se puede apreciar en la Figura 4-20, los resultados en términos de correlación para el volumen medido y estimado son buenos. Obteniendo un coeficiente de correlación de 0.93 y 0.96 respectivamente para cada conjunto, con un error típico de  $10.10$  y  $8.55 \text{ cm}^3$ , respectivamente.



(a) Conjunto de entrenamiento.



(b) Conjunto de prueba.

**Figura 4-20.:** Volumen medido [ $\text{cm}^3$ ] vs Volumen estimado [ $\text{cm}^3$ ] utilizando una RNA luego de aplicar un ADL en tomate.

## 4.6. Caso de Estudio: Mango - Nam doc mai

### 4.6.1. Identificación y medición de las características fisicoquímicas de interés en el producto agrícola.

Se utilizó una muestra de 54 frutos de Mango Nam doc mai analizados justo después de la cosecha. Fueron obtenidas imágenes a resolución de 2 Mpx (Tamaño imagen 4:3 (píxeles), 1080×1920 píxeles). Se recopiló un banco de imágenes correspondiente al mismo número de frutos, y se midió su masa utilizando una balanza (Tabla 4-10 y Figura 4-21a) y las características de dimensionamiento (ancho, alto y profundo) mediante un calibrador pie de rey para el cálculo de volumen (Tabla 4-11 y Figura 4-21a Figura 4-21b). Los sólidos solubles (Tabla 4-12 y Figura 4-21c) fueron medidos mediante un refractómetro. Para la estimación de la acidez (Tabla 4-13 y Figura 4-21d), se utilizó un titulador automático permitiendo determinar de igual manera el pH (Tabla 4-14 y Figura 4-21e) de cada una de las muestras. Las tablas de Frecuencia fueron construídas mediante la fórmula de Struges como fue explicado en casos de estudio precedentes.

**Tabla 4-10.:** Tabla de frecuencias para la masa [g] de frutos de mango nam doc mai.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	256.40	275.30	265.85	12	12	0.22	0.22
2	275.30	294.20	284.75	0	12	0.00	0.22
3	294.20	313.10	303.65	0	12	0.00	0.22
4	313.10	332.00	322.55	12	24	0.22	0.44
5	332.00	350.90	341.45	18	42	0.33	0.78
6	350.90	369.80	360.35	0	42	0.00	0.78
7	369.80	388.70	379.25	6	48	0.11	0.89
8	388.70	407.60	398.15	6	54	0.11	1.00

**Tabla 4-11.:** Tabla de frecuencias para el volumen [ $cm^3$ ] de mango nam doc mai.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	243.67	267.11	255.39	12	12	0.22	0.22
2	267.11	290.55	278.83	0	12	0.00	0.22
3	290.55	313.99	302.27	6	18	0.11	0.33
4	313.99	337.43	325.71	12	30	0.22	0.56
5	337.43	360.87	349.15	12	42	0.22	0.78
6	360.87	384.31	372.59	6	48	0.11	0.89
7	384.31	407.75	396.03	6	54	0.11	1.00



**Tabla 4-12.:** Tabla de frecuencias para sólidos solubles [*Brix*] de mango nam doc mai.

Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	15.60	16.14	15.87	12	12	0.22	0.22
2	16.14	16.69	16.41	6	18	0.11	0.33
3	16.69	17.23	16.96	6	24	0.11	0.44
4	17.23	17.77	17.50	12	36	0.22	0.67
5	17.77	18.31	18.04	0	36	0.00	0.67
6	18.31	18.86	18.59	12	48	0.22	0.89
7	18.86	19.40	19.13	6	54	0.11	1.00

**Tabla 4-13.:** Tabla de frecuencias para acidez de mango nam doc mai.

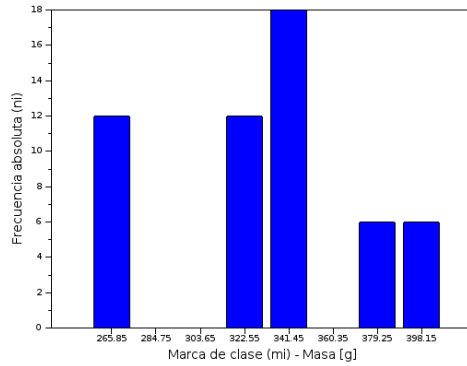
Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	0.97	1.03	1.00	12	12	0.22	0.22
2	1.03	1.09	1.06	12	24	0.22	0.44
3	1.09	1.15	1.12	6	30	0.11	0.56
4	1.15	1.21	1.18	0	30	0.00	0.56
5	1.21	1.27	1.24	6	36	0.11	0.67
6	1.27	1.33	1.30	12	48	0.22	0.89
7	1.33	1.39	1.36	6	54	0.11	1.00

**Tabla 4-14.:** Tabla de frecuencias para pH de mango nam doc mai.

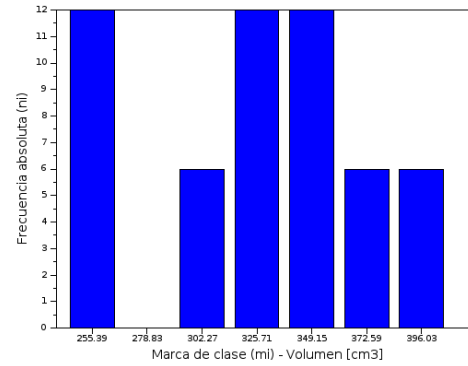
Clase	L. inf.	L. sup.	mi	ni	Ni	fi	Fi
1	4.28	4.52	4.40	12	12	0.22	0.22
2	4.52	4.75	4.63	0	12	0.00	0.22
3	4.75	4.98	4.87	24	36	0.44	0.67
4	4.98	5.22	5.10	12	48	0.22	0.89
5	5.22	5.45	5.34	0	48	0.00	0.89
6	5.45	5.69	5.57	0	48	0.00	0.89
7	5.69	5.92	5.80	6	54	0.11	1.00

#### 4.6.2. Procesamiento de imágenes y extracción de características a un grupo de imágenes del producto agrícola.

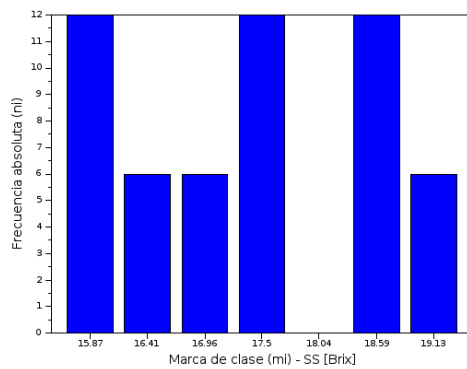
El banco de imágenes de Mango Nam doc mai fue normalizado en color, empleando la técnica de constancia de color Gray-World [110], se aplicó la transformación  $I_2$  OHTA y ajuste de elipse al contorno del fruto para extracción de ésta como la región de interés a analizar, ver Figura 4-22. Posteriormente fueron extraídas las características de color, canales R, G y B de manera independiente, así como L, a y b. Se obtuvieron los parámetros de textura, entropía, homogeneidad, contraste, energía y correlación. La longitud de los ejes mayores y menores del fruto, así como su área y perímetro y finalmente los 7 momentos de Hu.



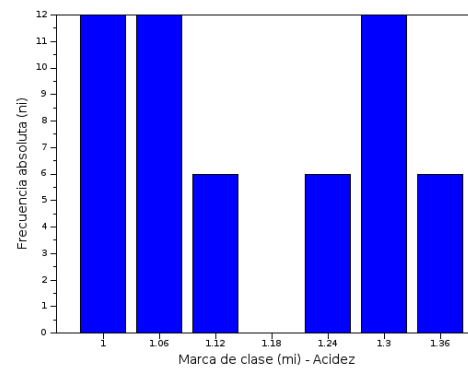
(a) mi Masa (g) vs ni.



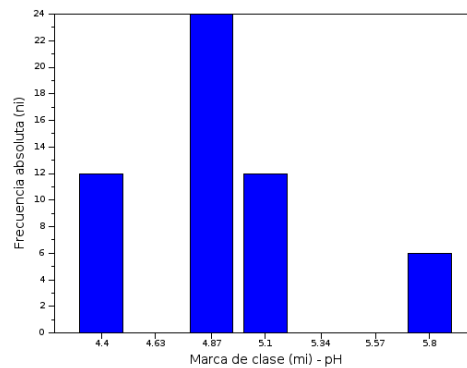
(b) mi Volumen ( $cm^3$ ) vs ni.



(c) mi SS (Brix) vs ni.

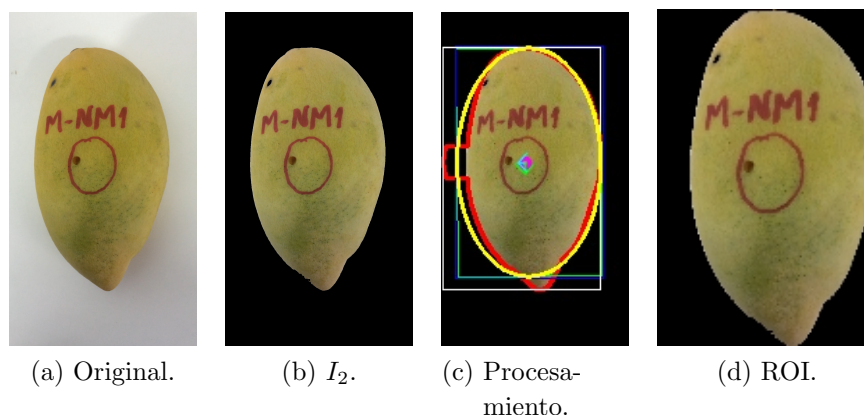


(d) mi Acidez vs ni.



(e) mi pH vs ni.

**Figura 4-21.:** Gráfico de barras para la comparación de masa, volumen, SS, acidez y pH en términos de su frecuencia absoluta para mango nam doc mai.



**Figura 4-22.:** Procesamiento de imagen realizado a una muestra del banco de imágenes de Mango Nam doc mai.

### 4.6.3. Estimación de las características fisicoquímicas medidas a partir de los atributos extraídos en las imágenes.

Al analizar el conjunto global de datos como un espacio  $R^n$  en el cual  $n$  es el número de características, éstos tendrán una variabilidad asociada en cada una de sus componentes. Una de las técnicas más utilizadas para llevar a cabo los análisis estadísticos para el tratamiento de los datos es el análisis por componentes principales (ACP), el cuál básicamente determina las componentes de mayor dispersión o variabilidad de los datos, en términos de las  $n$ -componentes. Sin embargo, no en todos los procesos las características o parámetros analizados según la variabilidad de los datos son eficaces para realizar procesos de estimación o clasificación, por lo cuál es necesario utilizar otro tipo de análisis.

Mediante la aplicación de ACP a un descriptor conformado por 22 características que incluyen aspectos de color, textura, tamaño y forma, se pueden analizar los datos en términos de su varianza. A partir del vector propio asociado al valor propio con mayor variabilidad como los presentados en la Tabla 4-15, se pueden identificar como la mayoría de estas características influye en éste vector. Multiplicando cada uno de los vectores propios por el valor obtenido de la  $m$  característica para la  $n$  muestra de determinado producto, es posible obtener la proyección sobre este nuevo espacio. Utilizando la proyección de los mayores valores propios cuya variabilidad explicada sea mayor al 95 % se conforma un nuevo conjunto de datos útil con fines de estimación mediante una RNA.

Por ejemplo, al utilizar los 5 mayores valores propios y obtener su respectiva proyección mediante sus vectores propios asociados en Mangos Nam doc mai, se obtiene una varianza explicada cercana al 96 %. Una RNA con 5 neuronas de entrada correspondientes a ésta proyección y siendo la neurona de salida la estimación de masa, volumen, sólidos solubles, acidez o pH, es posible definir mediante una regresión lineal simple, un modelo como el presentado en las Ecuación 4-7 para la estimación de dichas características.

$$Caracteristica_{físicoquímica} = Factor \times Estimacion_{VA} + b; Error : \epsilon \quad (4-7)$$

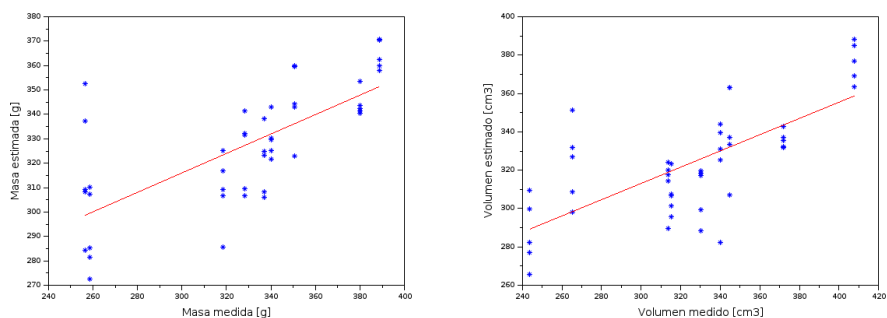
**Tabla 4-15.:** Vectores y valores propios del ACP en términos de las 22 características para la separación por variabilidad en Mango Nam doc mai.

No.	Característica	CP1	CP2	CP3	CP4	CP5
1	L	-0.12	-0.23	<b>0.34</b>	-0.16	-0.04
2	a	-0.18	-0.19	0.07	0.12	<b>0.80</b>
3	b	-0.17	-0.24	0.05	-0.15	0.10
4	B	-0.07	-0.23	<b>0.35</b>	0.12	-0.33
5	G	-0.10	-0.22	<b>0.36</b>	-0.19	-0.19
6	R	-0.15	-0.24	<b>0.28</b>	-0.12	0.24
7	Entropía	-0.21	-0.26	-0.25	<b>0.22</b>	-0.02
8	Homogeneidad	<b>0.22</b>	<b>0.25</b>	<b>0.27</b>	-0.28	0.07
9	Contraste	-0.18	-0.36	0.06	0.08	-0.12
10	Energía	<b>0.21</b>	<b>0.27</b>	<b>0.21</b>	-0.18	0.09
11	Correlación	0.07	0.11	<b>0.34</b>	-0.26	0.16
12	Eje Mayor	<b>0.20</b>	-0.06	0.09	<b>0.34</b>	0.05
13	Eje Menor	-0.06	0.21	<b>0.31</b>	<b>0.39</b>	0.01
14	Área	0.06	0.10	<b>0.26</b>	<b>0.50</b>	0.02
15	Perímetro	0.17	0.07	<b>0.20</b>	<b>0.29</b>	0.07
16	Hu 1	<b>0.33</b>	-0.17	-0.12	-0.06	0.18
17	Hu 2	<b>0.33</b>	-0.16	-0.12	-0.06	0.16
18	Hu 3	<b>0.27</b>	-0.28	-0.01	-0.09	-0.10
19	Hu 4	<b>0.33</b>	-0.25	0.02	0.02	-0.05
20	Hu 5	<b>0.30</b>	-0.20	0.02	0.07	0.00
21	Hu 6	<b>0.36</b>	-0.22	0.03	0.06	-0.05
22	Hu 7	0.13	0.00	0.07	0.13	0.01
	Valor Propio	0.52	0.27	0.25	0.11	0.08
	Varianza Explicada (%)	37.07	19.23	17.85	8.04	5.91

Aplicando el procedimiento sobre un conjunto de frutos de Mangos Nam doc mai, utilizando una RNA luego de aplicar ACP al conjunto de 22 características, es posible obtener la estimación por visión artificial  $Estimacion_{VA}$ . Comparando los valores reales de dichas características con la estimación obtenida se pueden definir las correlaciones existentes y por tanto definir el modelo de regresión lineal, para futuras estimaciones.

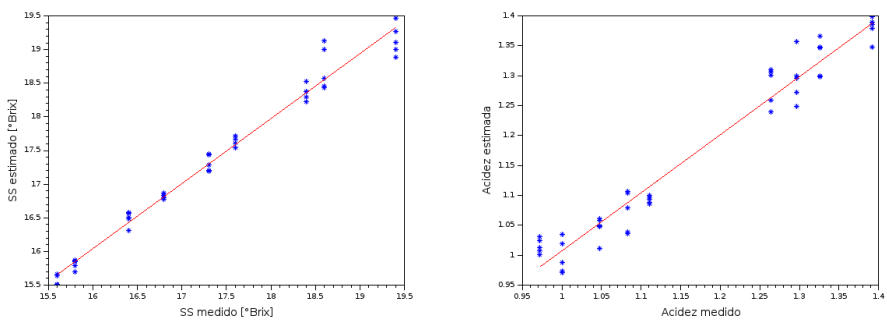
Por ejemplo, para la estimación de características físicas como la masa y el volumen se puede comparar el valor medido en gramos en laboratorio y la estimación obtenida mediante el sistema de visión artificial, como se puede apreciar en la Figura 4-23, donde cada punto está ubicado en un plano XY donde X es el valor medido y Y el valor estimado. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura 4-24.

Los modelos obtenidos mediante un sistema de regresión son presentados en las Ecuaciones 4-8, 4-9, 4-10, 4-11 y 4-12 y representados en las Figuras 4-23 y 4-24 como líneas continuas de color rojo.

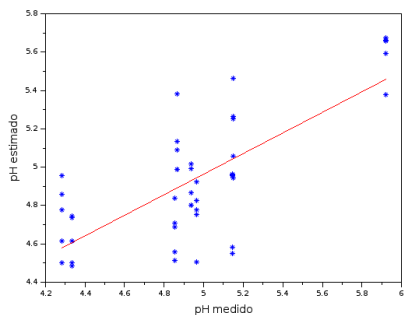


(a) Estimación de masa mediante ACP. (b) Estimación de volumen mediante ACP.

**Figura 4-23.**: Estimación de propiedades físicas mediante ACP.



(a) Estimación de SS mediante ACP. (b) Estimación de AT mediante ACP.



(c) Estimación de pH mediante ACP.

**Figura 4-24.**: Estimación de propiedades químicas mediante ACP.

$$Masa = 1.26 \times Estimacion_{VA} - 86.94[g]; Error : 17.59[g] \quad (4-8)$$

$$Volumen = 1.17 \times Estimacion_{VA} - 56.27[cm^3]; Error : 20.49[cm^3] \quad (4-9)$$

$$SS = 1.01 \times Estimacion_{VA} - 0.25[Brix]; Error : 0.17[Brix] \quad (4-10)$$

$$AT = 0.98 \times Estimacion_{VA} + 0.01; Error : 0.03 \quad (4-11)$$

$$pH = 1.01 \times Estimacion_{VA} - 0.07; Error : 0.22 \quad (4-12)$$

Al normalizar los valores medidos, es posible realizar el ADL utilizando la magnitud entre 0 y 1, para determinar el mayor valor de la característica de interés. Es decir, un valor cercano a 0, para una determinada imagen, corresponderá a una valor bajo en la medición dentro del conjunto datos y viceversa. Para el análisis de masa, por ejemplo, las mediciones obtenidas se encuentran entre 256 y 389 gramos para el total de 54 frutos, con una media de 329 gramos y una desviación estándar de 44 gramos. En este caso el valor de 0 corresponderá a los frutos muy cercanos a 256 gramos y 1 corresponderá a los cercanos a 389 gramos.

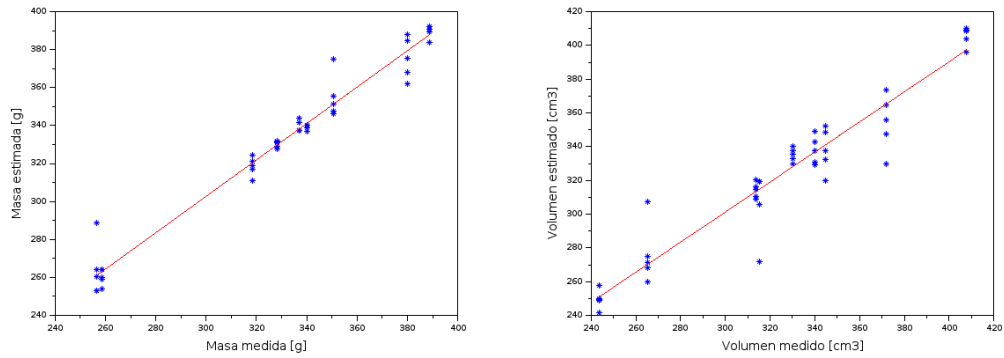
El ADL arroja como resultados una matriz de valores y vectores propios, correspondientes a la mejor discriminancia en función de la masa medida. Para este caso se obtuvieron 10 valores propios con sus correspondientes vectores propios de tamaño  $1 \times 22$ , debido al conjunto de características extraídas: 22. Utilizando los vectores propios asociados a los cinco mayores valores propios, es posible obtener el 96.8 % de la discriminancia de los datos, por lo cuál, la proyección de las 22 características utilizando estos 5 vectores se utilizará como insumo de entrada a la red neuronal. Las clases o grupos son conformadas por un conjunto de muestras del producto agrícola con características similares. Utilizando la proyección de los mayores valores propios cuya discriminancia explicada sea mayor al 95 % se conforma un nuevo conjunto de datos útil con fines de estimación mediante una RNA. El resultado para el análisis ADL para los discriminantes lineales con discriminancia explicada total de 96.8 % en Mangos Nam doc mai, se puede observar en la Tabla 4-16.

A partir de lo anterior, la RNA contó con 5 neuronas en la capa de entrada, y una neurona en la capa de salida, correspondiente a la estimación de la característica fisicoquímica de interés. Al analizar la curva de aprendizaje del sistema con un epsilon de 0.000001 se encontró que la red converge en 500 y 1000 iteraciones o épocas encontrando el menor error cuadrático medio, tanto para el conjunto de prueba como de entrenamiento, en todas las características fisicoquímicas.

**Tabla 4-16.:** Vectores y valores propios del ADL en términos de las 22 características para la separación por clases en Mango Nam doc mai.

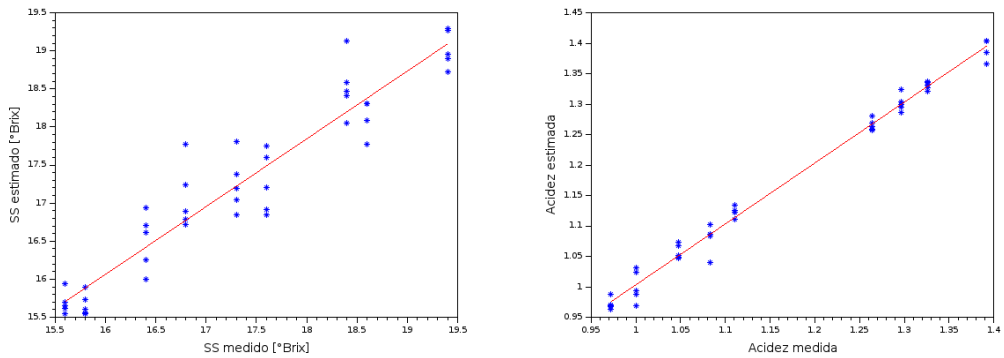
No.	Característica	DL1	DL2	DL3	DL4	DL5
1	L	-0.61	-0.79	<b>1.45</b>	<b>1.10</b>	<b>1.60</b>
2	a	0.06	-0.12	<b>0.23</b>	0.16	-0.24
3	b	-0.07	-0.19	-0.41	-0.38	0.13
4	B	-0.08	-0.01	-0.26	-0.23	0.07
5	G	<b>0.66</b>	<b>0.43</b>	-0.31	-0.18	-1.45
6	R	-0.06	<b>0.75</b>	-0.93	-0.75	-0.09
7	Entropía	-0.02	0.06	-0.07	0.04	0.04
8	Homogeneidad	-0.02	0.01	-0.09	-0.06	0.01
9	Contraste	0.04	-0.13	0.05	0.08	-0.01
10	Energía	-0.03	0.04	-0.14	-0.21	0.10
11	Correlación	0.03	-0.23	0.06	<b>0.21</b>	-0.04
12	Eje Mayor	-0.09	0.02	0.02	-0.29	0.12
13	Eje Menor	-0.07	-0.08	-0.08	-0.21	0.10
14	Área	0.12	0.02	0.04	<b>0.28</b>	-0.16
15	Perímetro	0.01	0.08	-0.06	0.08	-0.01
16	Hu 1	-0.24	-0.27	<b>0.20</b>	0.18	-0.18
17	Hu 2	<b>0.26</b>	0.13	-0.26	-0.08	0.16
18	Hu 3	0.06	-0.01	-0.03	0.18	-0.04
19	Hu 4	-0.05	<b>0.20</b>	0.03	-0.82	0.09
20	Hu 5	-0.07	-0.05	0.03	0.18	0.06
21	Hu 6	0.04	-0.11	0.00	<b>0.42</b>	-0.11
22	Hu 7	0.04	0.03	-0.03	0.10	-0.02
	Valor Propio	11.05	3.00	1.05	0.51	0.30
96.80 %	Discrim. Explicada (%)	67.13	18.26	6.40	3.12	1.87

Aplicando el procedimiento sobre el mismo conjunto de frutos de Mangos Nam doc mai, utilizando para el análisis ACP al conjunto de 22 características, es posible obtener la estimación por visión artificial  $Estimacion_{VA}$  basado en el ADL. Comparando los valores reales de dichas características con la estimación obtenida se pueden definir las correlaciones existentes y por tanto definir el modelo de regresión lineal, para futuras estimaciones. Para la estimación de características físicas como la masa y el volumen se puede apreciar en la Figura 3-11 la comparación entre la medición y la estimación. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura 3-12. Los modelos obtenidos mediante un sistema de regresión son presentados en las Ecuaciones 4-13, 4-14, 4-15, 4-16 y 4-17 y representados en las Figuras 4-25 y 4-26 como líneas continuas de color rojo.

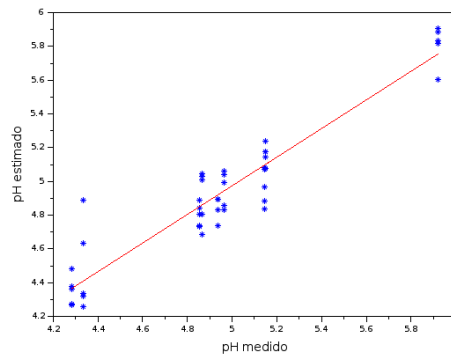


(a) Estimación de masa mediante ADL. (b) Estimación de volumen mediante ADL.

**Figura 4-25.:** Estimación de propiedades físicas mediante ADL.



(a) Estimación de SS mediante ADL. (b) Estimación de AT mediante ADL.



(c) Estimación de pH mediante ADL.

**Figura 4-26.:** Estimación de propiedades químicas mediante ADL.



$$Masa = 1.01 \times Estimacion_{VA} - 5.29[g]; Error : 7.52[g] \quad (4-13)$$

$$Volumen = 1.02 \times Estimacion_{VA} - 7.60[cm^3]; Error : 13.02[cm^3] \quad (4-14)$$

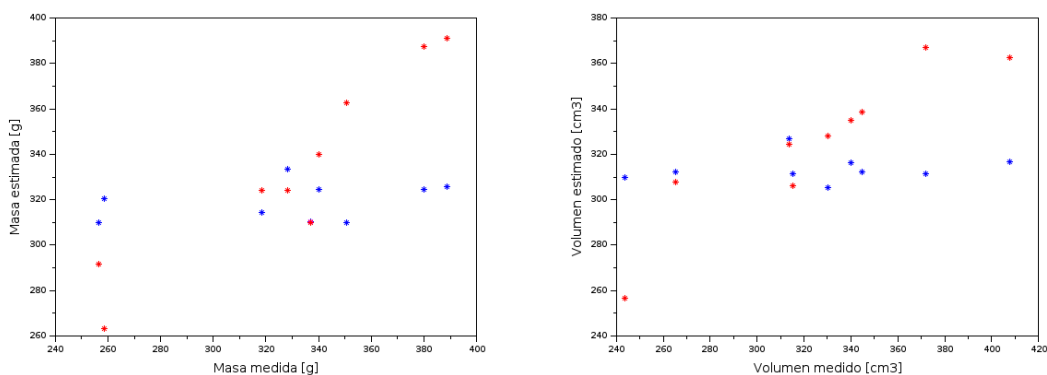
$$SS = 1.01 \times Estimacion_{VA} - 0.22[Brix]; Error : 0.36[Brix] \quad (4-15)$$

$$AT = 0.99 \times Estimacion_{VA} + 0.00; Error : 0.01 \quad (4-16)$$

$$pH = 1.04 \times Estimacion_{VA} - 0.22; Error : 0.14 \quad (4-17)$$

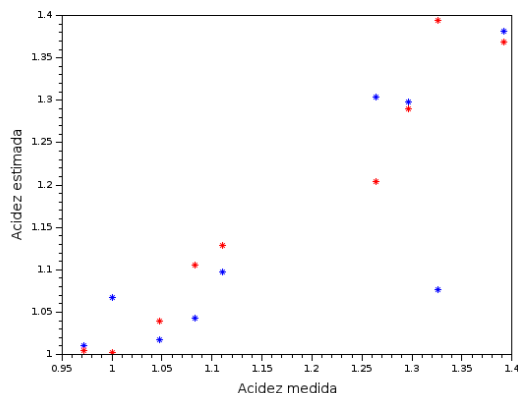
#### 4.6.4. Validación del método de correlación mediante la evaluación del desempeño de la estimación y el error de la misma.

Utilizando esta información es posible comparar por ejemplo las Ecuaciones determinadas mediante el ACP y ADL presentados en la sección precedente. Si bien, el sistema de regresión tiene asociado un coeficiente de correlación y de determinación, es necesario determinar la respuesta sobre un nuevo conjunto. Utilizando los modelos determinados por cada uno de los sistemas, se comparó la respuesta sobre la estimación de un conjunto nuevo de Mangos Nam doc mai. La estimación de características físicas como la masa y el volumen se puede apreciar en la Figura 4-27, en color azul para el ACP y rojo para el ADL. Del mismo modo, para el caso de algunas características químicas de interés, como los sólidos solubles, la acidez y el pH se pueden observar las correlaciones obtenidas en la Figura 4-28.

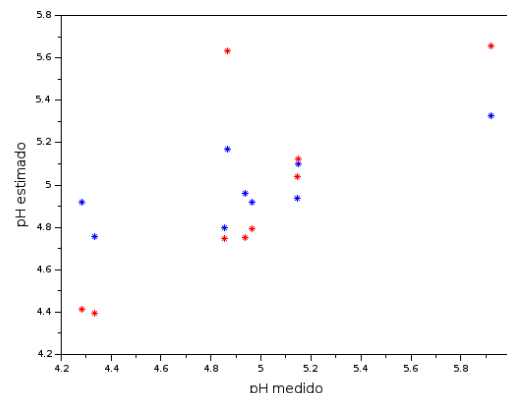


(a) Validación de masa mediante ACP y ADL. (b) Validación de volumen mediante ACP y ADL.

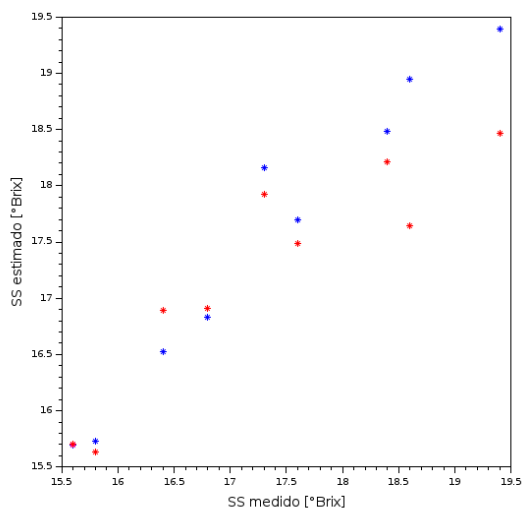
**Figura 4-27.:** Validación de la estimación de propiedades físicas mediante ACP y ADL.



(a) Validación de AT mediante ACP y ADL.



(b) Validación de pH mediante ACP y ADL.



(c) Validación de SS mediante ACP y ADL.

**Figura 4-28.:** Validación de la estimación de propiedades químicas mediante ACP y ADL.

Comparando los valores de los estadísticos descriptivos consignados en las Tablas 4-17, 4-18, 4-19, 4-20 y 4-21 se puede evidenciar claramente que la separación por clases mediante el análisis ADL permite obtener resultados de correlación múltiple superiores a 0.8 para la estimación de características fisicoquímicas: masa, volumen, sólidos solubles, acidez y pH. El error típico de los modelos obtenidos es coherente con los resultados obtenidos para el conjunto de validación. En resumen, es posible validar mediante un conjunto de prueba y los estadísticos descriptivos mostrados, los modelos obtenidos con visión artificial para la estimación de características fisicoquímicas.

**Tabla 4-17.:** Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.71	0.33	0.50	0.11	0.49	-0.01	17.59	48.40
ADL + RNA	0.98	0.93	0.97	0.87	0.97	0.85	7.52	17.97

**Tabla 4-18.:** Análisis estadístico de correlación entre los datos de visión y el volumen de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.70	0.14	0.49	0.02	0.48	-0.11	20.49	54.78
ADL + RNA	0.95	0.92	0.91	0.85	0.91	0.83	13.02	24.75

**Tabla 4-19.:** Análisis estadístico de correlación entre los datos de visión y los sólidos solubles de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.99	0.97	0.98	0.95	0.98	0.95	0.17	0.36
ADL + RNA	0.95	0.91	0.90	0.84	0.90	0.81	0.36	0.60

**Tabla 4-20.:** Análisis estadístico de correlación entre los datos de visión y la acidez de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.97	0.80	0.96	0.65	0.96	0.60	0.03	0.10
ADL + RNA	0.99	0.97	0.99	0.94	0.98	0.93	0.01	0.03

**Tabla 4-21.:** Análisis estadístico de correlación entre los datos de visión y el pH de frutos de Mango Nam doc mai.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
ACP + RNA	0.73	0.76	0.54	0.58	0.53	0.52	0.22	0.39
ADL + RNA	0.94	0.78	0.88	0.61	0.88	0.56	0.13	0.33

## 4.7. Análisis y discusión de los resultados.

Los resultados de las estimaciones para masa y volumen en términos generales, pueden ser considerados como buenos teniendo como coeficientes de correlación en la estimación de masa superiores a: 0.74 y 0.69 para el conjunto de entrenamiento y prueba respectivamente, como se puede evidenciar en la Tabla 4-22. Éstos valores corresponden a la estimación de masa en el mango común, en los demás frutos, la estimación aumenta a 0.77 y 0.83, lo cuál puede ser considerado como bueno.

Por lo anterior, es posible inferir que la estimación de masa en mango común puede estar sufriendo problemas de discriminancia de color, debido a que su análisis se realizó aún en estado verde o no maduro. Por esto, un análisis como el realizado para el tomate, en el cuál se usó un conjunto menor de datos, pero adquiriendo muestras durante la totalidad de la evolución del fruto, puede garantizar una mejor estimación.

**Tabla 4-22.:** Estadísticos descriptivos de correlación entre la masa estimada mediante una RNA luego de ADL y la masa real medida en frutos de maracuyá, granadilla, mango y tomate.

Estad./Prod.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
Maracuyá	0.77	0.83	0.60	0.69	0.59	0.65	16.59	15.97
Granadilla	0.92	0.85	0.85	0.72	0.85	0.69	16.98	11.91
Mango	0.74	0.69	0.55	0.47	0.54	0.40	13.08	11.45
Tomate	0.89	0.95	0.79	0.91	0.78	0.88	14.03	10.52

Para la estimación de volumen, los valores de correlación son superiores a 0.64 y 0.73 para el conjunto de entrenamiento y prueba, como se puede evidenciar en la Tabla 4-23. Análogamente a la estimación de masa, éstos valores corresponden a la estimación de volumen en el mango común, en los demás frutos, la estimación aumenta a 0.75 y 0.73. En el caso de maracuyá, según se observo en los datos medidos de volumen al calcular la media y desviación estándar se obtiene un valor de  $166.56 \pm 48 \text{ cm}^3$ , por lo cual comparando el error típico obtenido para éste caso de 34.44 y  $31.58 \text{ cm}^3$ , es disminuido en un 29%.

**Tabla 4-23.:** Estadísticos descriptivos de correlación entre el volumen estimado mediante una RNA luego de ADL y el volumen real medido en frutos de maracuyá, granadilla, mango y tomate.

Estad./Prod.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
Maracuyá	0.75	0.73	0.56	0.54	0.55	0.51	34.44	31.58
Granadilla	0.87	0.76	0.75	0.58	0.74	0.53	26.27	20.84
Mango	0.64	0.81	0.41	0.66	0.39	0.62	13.20	10.48
Tomate	0.93	0.96	0.86	0.92	0.84	0.68	10.10	8.55

Analizando en términos comparativos la estimación realizada, comparado con la media y desviación estándar de la muestra medida es posible concluir lo siguiente. Como se puede apreciar en la Tabla 4-24, el error típico de cada una de las estimaciones, en comparación con la media y desviación estándar muestrales es mejor ya que el error típico en función de la media de las mediciones es constante alrededor del 25 % lo cuál garantiza estimaciones con una exactitud superior al 75 % reduciendo un 20 % la desviación estándar.

Si se tiene en cuenta que, por ejemplo para el caso de la granadilla, el error típico de la estimación para el conjunto de entrenamiento y prueba se encuentra entre 11.91 y 16.98 gramos, y la media de la muestra es de 106.54 gramos en promedio para los 50 frutos, es posible concluir que la estimación de masa realizada con el sistema de visión de máquina puede obtener una respuesta entre el 84 % y 89 %.

Por otra parte, comparando el error típico obtenido mediante la estimación con la desviación estándar de las mediciones, para este mismo caso, es posible concluir que el error puede ser atenuado entre un 42 % y 59 % realizando la estimación con el método de visión de máquina.

Al realizar el análisis para los demás productos agrícolas, se encuentra una consistencia en estos resultados, manteniendo siempre una proporción de respuesta mayor al 77 %. Teniendo en cuenta que la proporción entre la desviación y la media muestral, para todos los casos en la medición de masa es menor al 80 %, la estimación puede considerarse en términos generales buena.

**Tabla 4-24.**: Análisis de resultados en proporción para la estimación de masa [g].

Estad./Prod.	Corr. Múlt.		Error Típ.		Medición		Respuesta	
	Entren.	Prueba	Entren.	Prueba	Media	Desv. Est.	Media	Desv. Est.
Maracuyá	0.77	0.83	16.59	15.97	73.67	25.99	77 %-78 %	36 %-39 %
Granadilla	0.92	0.85	16.98	11.91	106.54	29.04	84 %-89 %	42 %-59 %
Mango	0.74	0.69	13.08	11.45	128.58	16.43	90 %-91 %	20 %-30 %
Tomate	0.89	0.95	14.03	10.52	135.45	22.26	90 %-92 %	37 %-53 %

En el caso del volumen, como se puede apreciar en la Tabla 4-25, el error típico de cada una de las estimaciones, es comparación con la media y desviación estándar muestrales es también mejorado en términos porcentuales, siendo superior en la maracuyá y en los análisis de los demás productos agrícolas a 79 %. Si se tiene en cuenta que por ejemplo, para el caso del tomate, el error típico de la estimación para el conjunto de entrenamiento y prueba se encuentra entre 10.10 y 8.55  $cm^3$ , y la media de la muestra es de 133.70  $cm^3$  en promedio para las 32 mediciones realizadas sobre los 14 frutos, es posible concluir que la estimación de volumen realizada con el sistema de visión de máquina puede obtener una respuesta entre el 92 % y 94 %.

Comparando el error típico obtenido mediante la estimación con la desviación estándar de las mediciones, para este mismo caso, es posible concluir que el error puede ser atenuado entre un 54 % y 61 % realizando la estimación con el método de visión de máquina. Teniendo en cuenta que la proporción entre la desviación y la media muestral, para todos los casos en la medición de volumen es también menor al 80 %, la estimación de volumen puede considerarse en términos generales buena.

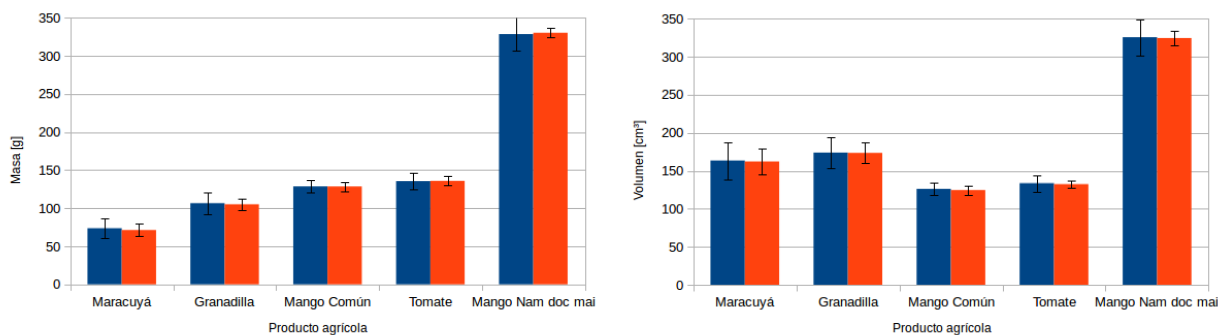
**Tabla 4-25.:** Análisis de resultados en proporción para la estimación de volumen [ $cm^3$ ].

Estad./Prod.	Corr. Múlt.		Error Típ.		Medición		Respuesta	
	Entren.	Prueba	Entren.	Prueba	Media	Desv. Est.	Media	Desv. Est.
Maracuyá	0.75	0.73	34.44	31.58	163.55	48.70	79 %-81 %	29 %-35 %
Granadilla	0.87	0.76	26.27	20.84	173.88	40.58	85 %-88 %	35 %-49 %
Mango	0.64	0.81	13.20	10.48	126.21	16.52	90 %-92 %	20 %-37 %
Tomate	0.93	0.96	10.10	8.55	133.70	21.75	92 %-94 %	54 %-61 %

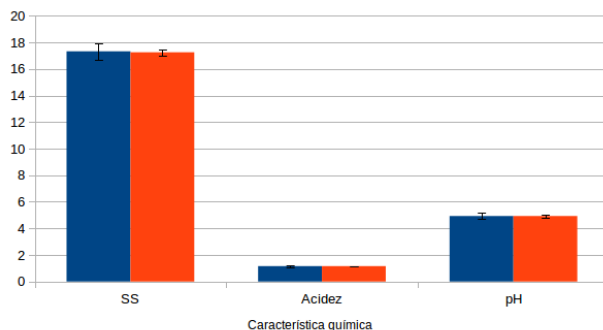
En el caso de la estimación de características fisicoquímicas en mango nam doc mai, como se puede apreciar en la Tabla 4-26, el error típico de cada una de las estimaciones, en comparación con la media y desviación estándar muestrales es también mejorado en términos porcentuales, como se evidencia de igual manera en las Figura 4-29. Los resultados obtenidos en términos del coeficiente de correlación múltiple es superior a 0.91, lo cual indica una correlación buena entre los datos medidos y estimados. La desviación estándar como se puede validar en las barras de error de la Figura 4-29 son siempre disminuidos en todas las estimaciones sin afectar la media muestral.

**Tabla 4-26.:** Análisis estadístico de correlación entre los datos de visión y la masa de frutos de Mango Nam doc mai en diferentes características fisicoquímicas usando una RNA luego de ADL.

Estad./Anál.	Corr. Múlt.		Det. $R^2$		$R^2$ ajustado		Error Típ.	
	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba	Entren.	Prueba
Masa	0.98	0.93	0.97	0.87	0.97	0.85	7.52	17.97
Volumen	0.95	0.92	0.91	0.85	0.91	0.83	13.02	24.75
SS	0.95	0.91	0.90	0.84	0.90	0.81	0.36	0.60
Acidez	0.99	0.97	0.99	0.94	0.98	0.93	0.01	0.03
pH	0.94	0.78	0.88	0.61	0.88	0.56	0.13	0.33



(a) Media y desviación estándar para estimación de masa. (b) Media y desviación estándar para estimación de volumen.



(c) Media y desviación estándar para estimación de características químicas en mango nam doc mai.

**Figura 4-29.:** Comparación entre la media y desviación estándar medidos vs estimados para diferentes productos agrícolas.

## Resumen del capítulo

En este capítulo se presenta la aplicación de la metodología propuesta para la identificación de características fisicoquímicas mediante técnicas de visión de máquina, en diferentes productos agrícolas. Los productos agrícolas analizados fueron: maracuyá, granadilla, mango común, tomate y mango nam doc mai. Se estimó la masa y volumen en todos los productos agrícolas. Para el caso del mango nam doc mai, fue añadido la estimación de las características químicas más utilizadas: sólidos solubles, acidez y pH. Los resultados obtenidos validan la metodología propuesta, obteniendo correlaciones superiores a 0.8 en la totalidad de los casos, mediante la estimación a partir de ADL y RNA's.

## 5. Conclusiones y trabajo futuro

La metodología analizada y propuesta en función de investigaciones previas realizadas en la agroindustria, responde bastante bien para la estimación de masa y volumen en términos generales superiores al 80 %. Teniendo en cuenta, que dichas investigaciones trabajan la extracción de algunas características y/o la estimación de algunos parámetros de interés en la imagen, mas no la correlación de éstos con variables físicas reales, es posible concluir que utilizar un sistema de visión de máquina que extraiga las características análogas al análisis realizados a partir de la experiencia de humanos, tales como: color, textura, tamaño y forma, puede ayudar de manera significativa en la determinación de cualidades específicas en los productos agrícolas.

Los métodos para análisis de variabilidad y discriminancia de datos, tales como análisis por componentes principales (ACP) o análisis discriminante lineal (ADL) pueden ayudar a resolver problemas de estimación de características fisicoquímicas en productos agrícolas, siendo más útil el segundo, en una proporción superior al 20 % en comparación al primero. Lo anterior, es debido a que en problemas como el analizado, no necesariamente la variabilidad de los datos puede respresentar la mejor separación entre cada una de las clases, siendo cada clase un conjunto de datos con una masa similar por ejemplo. Esto se soporta también al momento de utilizar la red neuronal artificial (RNA) para estimar la característica medida, a partir de la información obtenida mediante el ADL a partir de las características extraídas de las imágenes.

El uso de modelos matemáticos y teorías de análisis ha sido ampliamente estudiada para la correlación de parámetros. Si bien, por ejemplo los sistemas de regresión lineal múltiple (RLM) han sido utilizados en gran medida para la estimación de parámetros en función de características; este tipo de aproximaciones puede no ser tan útil en procesos en los cuáles no sea evidente o no exista la relación lineal entre las variables de interés. Como se analizó para la estimación de masa en maracuyá, este método en comparación con técnicas de visión de máquina tales como: análisis por componentes principales (ACP), análisis discriminante lineal (ADL) o redes neuronales artificiales (RNA) puede tener una respuesta menor, debido a la complejidad del problema.



La validación realizada mediante diferentes productos agrícolas para estimar diferentes características fisicoquímicas utilizando la metodología planteada, evidencia que: El coeficiente de correlación múltiple para la correlación de parámetros físicos como masa y volumen fue superior a 0.74 y 0.64 respectivamente. Para la estimación del conjunto de características fisicoquímicas: masa, volumen, sólidos solubles, acidez y pH, los resultados en términos del coeficiente de correlación múltiple para frutos de nam doc mai, es siempre superior a 0.94; con una disminución significativa del error típico en la medición. Estos resultados obtenidas, permiten concluir que el uso de un descriptor de imágenes compuesto por características de color, textura, tamaño y forma; y discriminado mediante un análisis ADL conjunto con una RNA permite estimar correctamente este tipo de atributos fisicoquímicos en productos agrícolas.

Con el objetivo de utilizar la metodología como herramienta de estimación, se propone como trabajo futuro la creación de un sistema complejo inteligente para la estimación de características fisicoquímicas, independientemente del producto agrícola. Es posible proponer a partir de los resultados obtenidos la realización de un banco lo suficientemente grande de imágenes y mediciones fisicoquímicas, que permita la estimación de dichos parámetros no solo para un conjunto específico de frutos, llamando conjunto a: tipo de fruto, variedad, tamaño, etc. El sistema puede llegar a contar con una sola red neuronal por medio de la cual cada una de las neuronas de la capa de salida, correspondiera a la estimación de cada uno de los parámetros fisicoquímicos de interés.

## A. Datos físicos medidos en productos agrícolas.

No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]
1	79	171,91	26	80	160,31	51	78	124,51	76	103	202,17
2	71	138,04	27	82	188,67	52	60	142,98	77	61	120,28
3	59	220,19	28	46	131,15	53	61	125,69	78	95	204,94
4	122	211,10	29	60	144,28	54	62	136,53	79	93	211,69
5	52	149,71	30	48	133,43	55	73	149,20	80	41	130,88
6	48	126,04	31	132	231,05	56	72	147,18	81	74	138,85
7	70	126,27	32	39	102,95	57	66	153,09	82	73	137,48
8	56	106,80	33	47	107,95	58	71	149,01	83	162	270,62
9	133	312,23	34	212	452,10	59	39	121,29	84	69	149,95
10	77	131,20	35	70	167,40	60	92	148,61	85	47	138,41
11	63	190,53	36	116	249,74	61	69	179,45	86	34	110,57
12	76	147,76	37	62	137,15	62	76	170,01	87	53	121,92
13	56	138,51	38	81	210,88	63	86	191,58	88	25	101,08
14	51	173,17	39	79	164,51	64	79	177,79	89	68	157,08
15	72	148,96	40	70	144,19	65	65	143,63	90	71	171,05
16	70	109,58	41	93	199,63	66	60	142,65	91	85	164,77
17	85	171,56	42	75	177,59	67	73	151,63	92	93	195,62
18	95	237,40	43	66	153,33	68	86	200,82	93	71	127,24
19	107	187,22	44	91	175,38	69	74	179,61	94	59	143,25
20	89	203,71	45	63	150,71	70	43	106,01	95	58	130,71
21	83	194,72	46	31	99,37	71	52	110,48	96	73	148,65
22	61	198,81	47	88	189,17	72	37	101,82	97	85	157,07
23	51	130,93	48	79	171,01	73	105	184,62	98	83	186,93
24	62	161,61	49	76	149,92	74	71	99,86	99	96	228,79
25	61	159,83	50	65	212,64	75	64	136,68	100	81	179,89

Tabla A-1.: Datos medidos de masa [g] y volumen [ $cm^3$ ] en maracuyá.

A. Datos físicos medidos en productos agrícolas.

No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]
1	108	103,10	18	125	119,68	35	100	98,93
2	157	154,77	19	105	105,72	36	146	150,78
3	115	118,94	20	122	124,21	37	140	137,64
4	120	114,54	21	121	114,39	38	127	129,84
5	103	104,15	22	157	158,48	39	135	129,21
6	125	129,85	23	121	119,75	40	112	109,43
7	142	141,37	24	120	113,91	41	139	125,11
8	126	123,96	25	114	112,83	42	116	111,03
9	130	127,08	26	141	133,85	43	130	131,53
10	139	135,84	27	161	155,36	44	134	131,08
11	138	133,55	28	124	124,38	45	130	135,73
12	129	126,35	29	150	142,70	46	92	91,28
13	110	104,81	30	128	128,07	47	155	151,97
14	122	118,21	31	112	109,53	48	125	119,02
15	157	157,67	32	119	109,48	49	136	133,22
16	128	132,63	33	128	125,06	50	141	137,12
17	111	106,08	34	163	157,48			

**Tabla A-2.:** Datos medidos de masa [g] y volumen [ $cm^3$ ] en mango común.

No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]	No.	Masa [g]	Vol. [ $cm^3$ ]
1	99	170,02	18	126	182,90	35	72	125,11
2	128	221,53	19	150	235,09	36	153	235,44
3	98	184,51	20	149	239,28	37	80	140,17
4	82	154,69	21	144	220,84	38	80	139,58
5	89	157,50	22	103	168,32	39	106	137,70
6	72	127,24	23	68	118,80	40	62	107,45
7	127	198,00	24	149	240,30	41	57	98,19
8	54	106,54	25	128	181,23	42	83	146,13
9	128	214,52	26	103	180,06	43	79	136,59
10	94	152,84	27	90	159,04	44	77	148,14
11	95	158,25	28	114	185,21	45	55	106,79
12	143	226,71	29	124	199,54	46	117	178,30
13	162	253,28	30	142	225,07	47	127	181,98
14	101	161,66	31	143	244,35	48	83	139,58
15	85	147,28	32	83	149,54	49	126	190,28
16	129	188,61	33	106	154,49	50	112	171,20
17	131	207,47	34	119	196,66			

**Tabla A-3.:** Datos medidos de masa [g] y volumen [ $cm^3$ ] en granadilla.

No.	Semana 1		Semana 2		Semana 3	
	Masa [g]	Volumen [ $cm^3$ ]	Masa [g]	Volumen [ $cm^3$ ]	Masa [g]	Volumen [ $cm^3$ ]
1	140,25	138	135,82	139	131,61	138
2	128,72	130	125,92	130	122,97	120
3	97,32	100	94,63	90	92,48	90
4	125,13	120	120,29	120	116,61	120
5	157,38	159	154,34	158	151,47	142
6	162,88	160	159,77	160	156,55	150
7	125,75	121	122,67	120	119,85	117
8	134,57	132	132,08	138	129,48	120
9	173,15	165	168,87	170	165,09	162

No.	Masa [g]	Volumen [ $cm^3$ ]
10	128	123
11	178	167
12	143	137
13	128	135
14	112	107

**Tabla A-4.:** Datos medidos de masa [g] y volumen [ $cm^3$ ] en tomate.

# Referencias

- [1] ABBASGHOLIPOUR, M ; OMID, M ; KEYHANI, A ; MOHTASEBI, S S.: Color image segmentation with genetic algorithm in a raisin sorting system based on machine vision in variable conditions. En: *Expert Systems with Applications* 38 (2011), Nr. 4, p. 3671–3678. – ISBN 0957–4174
- [2] ABD LATIF, M H. ; MD. YUSOF, H ; SIDEK, S N. ; RUSLI, N: Implementation of GLCM Features in Thermal Imaging for Human Affective State Detection. En: *Procedia Computer Science* 76 (2015), Nr. Iris, p. 308–315. – ISSN 18770509
- [3] ADUBOFUOR, J ; AMANKWAH, E A. ; ARTHUR, B S. ; APPIAH, F: Comparative study related to physico-chemical properties and sensory qualities of tomato juice and cocktail juice produced from oranges , tomatoes and carrots. En: *African Journal of Food Science* 4 (2010), Nr. July, p. 427–433
- [4] AL OHALI, Yousef: Computer vision based date fruit grading system: Design and implementation. En: *Journal of King Saud University - Computer and Information Sciences* 23 (2011), Nr. 1, p. 29–36. – ISBN 1319–1578
- [5] ALMÁSI, Adela-Diana ; WOŹNIAK, Stanislaw ; CRISTEA, Valentin ; LEBLEBICI, Yusuf ; ENGBERSEN, Ton: Review of Advances in Neural Networks: Neural Design Technology Stack. En: *Proceedings of ELM-2014 Volume 1 SE - 31 3* (2015), p. 367–376. – ISBN 978–3–319–14062–9
- [6] ALZAMORA, S. M. ; GUERRERO, S. N. ; NIETO, a. B. ; VIDALES, S. L.: Conservación de frutas y hortalizas mediante tecnologías combinadas ”Manual de Capacitación”. En: *Fao* (2004), p. 69
- [7] ARIVU, C Velappan G. ; PRAKASH, G ; SARMA, A Sada S.: Online Image Capturing and Processing Using Vision Box Hardware: Apple Grading 1. 2 (2012), Nr. 3, p. 639–643
- [8] AVILA, Felipe ; MORA, Marco ; OYARCE, Miguel ; ZUÑIGA, Alex ; FREDES, Claudio: A method to construct fruit maturity color scales based on support machines for regression: Application to olives and grape seeds. En: *Journal of Food Engineering* 162 (2015), p. 9–17. – ISSN 02608774

- 
- [9] BARNARD, Kobus ; GABBUR, Prasad: Color and Color Constancy in a Translation Model for Object Recognition. En: *Eleventh Color Imaging Conference* (2003), p. 364–368
- [10] BENGTTSSON, Hanna: *The Physicochemical and Sensory Properties of Fruit and Vegetable Fibre Suspensions - The Effect of Fibre Processing and its Addition to Low-Fat Sausages*, Tesis de Grado, 2009. – 83 p.
- [11] BIANCHI, Tiago ; GUERRERO, Luis ; GRATACÓS-CUBARSI, Marta ; CLARET, Anna ; ARGYRIS, Jason ; GARCIA-MAS, Jordi ; HORTÓS, Maria: Textural properties of different melon (*Cucumis melo* L.) fruit types: sensory and physical-chemical evaluation. En: *Scientia Horticulturae* 201 (2016), p. 46–56. – ISSN 03044238
- [12] BLASCO, J. ; ALEIXOS, N. ; CUBERO, S. ; GÓMEZ-SANCHÍS, J. ; MOLTÓ, E.: Automatic sorting of satsuma (*Citrus unshiu*) segments using computer vision and morphological features. En: *Computers and Electronics in Agriculture* 66 (2009), Nr. 1, p. 1–8. – ISBN 0168–1699
- [13] BLASCO, J ; ALEIXOS, N ; MOLTÓ, E: Machine vision system for automatic quality grading of fruit. En: *Biosystems Engineering* 85 (2003), Nr. 4, p. 415–423. – ISBN 1537–5110
- [14] BLASCO, J ; CUBERO, S ; GÓMEZ-SANCHÍS, J ; MIRA, P ; MOLTÓ, E: Development of a machine for the automatic sorting of pomegranate (*Punica granatum*) arils based on computer vision. En: *Journal of Food Engineering* 90 (2009), Nr. 1, p. 27–34. – ISBN 02608774
- [15] BUITRAGO, Germán V ; LÓPEZ, Alonso P. ; CORONADO, Alfonso P. ; OSORNO, Fernando L.: Determinacion de las características físicas y propiedades mecánicas de papa cultivada en Colombia. En: *Revista Brasileira de Engenharia Agrícola e Ambiental* 8 (2004), Nr. 1, p. 102–110
- [16] BULANON, D. M. ; BURKS, T. F. ; ALCHANATIS, V.: Image fusion of visible and thermal images for fruit detection. En: *Biosystems Engineering* 103 (2009), Nr. 1, p. 12–22. – ISBN 1537–5110
- [17] CANO MARCHAL, P ; MARTÍNEZ GILA, D ; GÁMEZ GARCÍA, J ; GÓMEZ ORTEGA, J: Expert system based on computer vision to estimate the content of impurities in olive oil samples. En: *Journal of Food Engineering* 119 (2013), Nr. 2, p. 220–228. – ISBN 0260–8774
- [18] CAO, Fei ; GUAN, Chunyue ; DAI, Hongyan ; LI, Xiaoming ; ZHANG, Zhihong: Soluble solids content is positively correlated with phosphorus content in ripening strawberry fruits. En: *Scientia Horticulturae* 195 (2015), p. 183–187. – ISSN 03044238

- 
- [19] CAPITÁN-VALLVEY, Luis F. ; LÓPEZ-RUIZ, Nuria ; MARTÍNEZ-OLMOS, Antonio ; ERENAS, Miguel M. ; PALMA, Alberto J.: Recent developments in computer vision-based analytical chemistry: A tutorial review. En: *Analytica Chimica Acta* 899 (2015), p. 23–56. – ISSN 18734324
- [20] CHALIDABHONGSE, Thanarat ; YIMYAM, Panitnat ; SIRISOMBOON, Panmanas: 2D/3D vision-based mango's feature extraction and sorting. En: *9th International Conference on Control, Automation, Robotics and Vision, 2006, ICARCV '06* (2006). ISBN 1424403421
- [21] CHAUGULE, Archana: Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties. En: *Hindawi Publishing Corporation-Journal of Engineering* 2014 (2014), p. 8
- [22] CHEN, Qiwen ; AGU, Emmanuel: Exploring Statistical GLCM Texture Features for Classifying Food Images. En: *2015 International Conference on Healthcare Informatics* 3 (2015), Nr. 2006, p. 453–453. ISBN 978-1-4673-9548-9
- [23] CHEN, Quansheng ; ZHANG, Chaojie ; ZHAO, Jiewen ; OUYANG, Qin: Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety. En: *TrAC - Trends in Analytical Chemistry* 52 (2013), p. 261–274. – ISBN 0165-9936
- [24] CHEN, Yud R. ; CHAO, Kuanglin ; KIM, Moon S.: Machine vision technology for agricultural applications. En: *Computers and Electronics in Agriculture* 36 (2002), Nr. 2-3, p. 173–191. – ISBN 0168-1699
- [25] DE-AN, Zhao ; JIDONG, Lv ; WEI, Ji ; YING, Zhang ; YU, Chen: Design and control of an apple harvesting robot. En: *Biosystems Engineering* 110 (2011), Nr. 2, p. 112–122. – ISBN 1537-5110
- [26] DELMORO, Julieta ; MUÑOZ, Daniela ; NADAL, Victoria ; PRANZETTI, Valeria: El color en los alimentos: Determinación de color en mieles. En: *Invenio* 13 (2010), Nr. 25, p. 145–152. – ISBN 8771511601
- [27] DELWICHE, Stephen R. ; MEKWATANAKARN, Weena ; WANG, Chien Y.: Soluble solids and simple sugars measurement in intact mango using near infrared spectroscopy. En: *HortTechnology* 18 (2008), Nr. 3, p. 410–416. – ISBN 1063-0198
- [28] DESAI, B B. ; SALUNKHE, D K.: Fruits and vegetables. En: *Food quality and standards* II (1991), p. 301–412. ISBN 9780123741363
- [29] DI CAGNO, Raffaella ; MINERVINI, Giovanna ; RIZZELLO, Carlo G. ; DE ANGELIS, Maria ; GOBBETTI, Marco: Effect of lactic acid fermentation on antioxidant, texture, color and sensory properties of red and green smoothies. En: *Food Microbiology* 28 (2011), Nr. 5, p. 1062–1071. – ISSN 07400020

- 
- [30] DIAZ, R ; FAUS, G ; BLASCO, M ; BLASCO, J ; MOLTÓ, E: The application of a fast algorithm for the classification of olives by machine vision. En: *Food Research International* 33 (2000), Nr. 3-4, p. 305–309. – ISBN 0963–9969
- [31] DIAZ, R ; GIL, L ; SERRANO, C ; BLASCO, M ; MOLTÓ, E ; BLASCO, J: Comparison of three algorithms in the classification of table olives by means of computer vision. En: *Journal of Food Engineering* 61 (2004), Nr. 1 SPEC., p. 101–107. – ISBN 0260–8774
- [32] DOMENE R, Miguel A. ; SEGURA R, Marilo: Parámetros de calidad externa en la industria agroalimentaria. En: *Cajamar ADN Agro* 3 (2014), p. 1–12
- [33] DU, Cheng J. ; SUN, Da W.: Learning techniques used in computer vision for food quality evaluation: A review. En: *Journal of Food Engineering* 72 (2006), Nr. 1, p. 39–55. – ISBN 0260–8774
- [34] E. MARTÍNEZ-MORENO, T. CORONA-TORRES, E. AVITIA-GARCÍA, A. M. CASTILLO-GONZÁLEZ, T. TERRAZAS-SALGADO, M. T. Colinas-León.: Caracterización morfológica de frutos y semillas de Nanche (*Byrsonima crassifolia* (L.) H.B.K.). En: *Revista Chapingo Serie Horticultura* 12 (2006), Nr. 1, p. 11–17
- [35] EBNER, Marc ; TISCHLER, German ; ALBERT, Jürgen: Integrating color constancy into JPEG2000. En: *IEEE Transactions on Image Processing* 16 (2007), Nr. 11, p. 2697–2706. – ISBN 1057–7149 (Print)\r1057–7149 (Linking)
- [36] ECHEVERRIA, E.: Brix and acid determinations. En: *University of Florida, IFAS - Citrus Research and Education Center*, p. 65–75
- [37] EIFERT, Joseph D. ; SANGLAY, Gabriel C. ; LEE, Dah J. ; SUMNER, Susan S. ; PIERSON, Merle D.: Prediction of raw produce surface area from weight measurement. En: *Journal of Food Engineering* 74 (2006), Nr. 4, p. 552–556. – ISBN 0260–8774
- [38] EL-BENDARY, Nashwa ; EL HARIRI, Esraa ; HASSANIEN, Aboul E. ; BADR, Amr: Using machine learning techniques for evaluating tomato ripeness. En: *Expert Systems with Applications* 42 (2015), Nr. 4, p. 1892–1905. – ISSN 09574174
- [39] ELMASRY, Gamal ; CUBERO, Sergio ; MOLTÓ, Enrique ; BLASCO, José: In-line sorting of irregular potatoes by using automated computer-based machine vision system. En: *Journal of Food Engineering* 112 (2012), Nr. 1-2, p. 60–68. – ISBN 0260–8774
- [40] FAJARDO, BLANCA; MOLINA, DIANA; OSPINA, JULIO; GARCÍA, Hugo: Determinación de Algunas Propiedades Físicas y Mecánicas de la panela granulada. En: *Revista de Ingeniería e Investigación* 43 (1999), Nr. 43, p. 34 – 39



- 
- [41] FISCHER, G ; MARTÍNEZ, O: Calidad y madurez de la uchuva (*Physalis peruviana* L) en relación con la coloración del fruto. En: *Agronomía colombiana* 16 (1999), Nr. 1-3, p. 35–39
- [42] FISZELEW, A. ; GARCIA-MARTINEZ, R.: Generación automática de redes neuronales con ajuste de parámetros basado en algoritmos genéticos. En: *Revista electrónica de sistemas de información* (2003), p. 1–21
- [43] FORSYTH, D A.: A novel algorithm for color constancy. En: *International Journal of Computer Vision* 5 (1990), Nr. 1, p. 5–36. – ISSN 0920–5691
- [44] FOSTER, David H.: Color constancy. En: *Vision Research* 51 (2011), Nr. 7, p. 674–700. – ISBN 0042–6989
- [45] GAO, Hui ; DAVIS, James W.: Why direct LDA is not equivalent to LDA. En: *Pattern Recognition* 39 (2006), Nr. 5, p. 1002–1006. – ISSN 00313203
- [46] GARRIDO-NOVELL, Cristóbal ; PÉREZ-MARIN, Dolores ; AMIGO, Jose M. ; FERNÁNDEZ-NOVALES, Juan ; GUERRERO, Jose E. ; GARRIDO-VARO, Ana: Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras. En: *Journal of Food Engineering* 113 (2012), Nr. 2, p. 281–288. – ISBN 0260–8774
- [47] GAY, Paolo ; BERRUTO, Remigio ; PICCAROLO, Pietro: Fruit Color Assessment for Quality Grading Purposes. En: *2002 ASAE Annual International Meeting/ CIGR XVth World Congress* 0300 (2002), Nr. 02, p. 1–9
- [48] GRAJEDA-GONZALEZ, Francisco ; CONTRERAS-SALAZAR, Ernesto A. ; LUNA-MALDONADO, Alejandro I.: Sistema de Procesamiento de Imágenes para Obtener los Parámetros del Color en Frutos de dos Variedades de Tomate. (2015), p. 1156–1161
- [49] GUNASEKARAN, Sundaram: Computer vision technology for food quality assurance. En: *Trends in Food Science & Technology* 7 (1996), Nr. 8, p. 245–256. – ISBN 0924–2244
- [50] HARRILL, Rex: Using a Refractometer To Test the Quality of Fruits & Vegetables. En: *Publiseret af Pineknoll Publishing, Keedysville* (1998), p. 1–28
- [51] HERBERT, Ítalo ; CAVALCANTE, Lucena ; CAVALCANTE, Lourival F.: Physical and Chemical Characteristics of Tropical and Non-Conventional Fruits. (1999), p. 3–17
- [52] HIROSE, Hideo ; SOEJIMA, Yusuke ; HIROSE, Kei: NNRMLR: A combined method of nearest neighbor regression and multiple linear regression. En: *Proceedings of the 2012 IIAI International Conference on Advanced Applied Informatics, IIAIAAI 2012* (2012), p. 351–356. ISBN 9780769548265

- 
- [53] HUANG, Ran ; NI, Kai ; WU, Xueying ; MA, Qifei: Refractometer based on Mach-Zehnder interferometer with peanut-shape structure. En: *Optics Communications* 353 (2015), p. 27–29. – ISSN 00304018
- [54] JARIMOPAS, Bundit ; JAISIN, Nitipong: An experimental machine vision system for sorting sweet tamarind. En: *Journal of Food Engineering* 89 (2008), Nr. 3, p. 291–297. – ISSN 02608774
- [55] JI, Wei ; ZHAO, Dean ; CHENG, Fengyi ; XU, Bo ; ZHANG, Ying ; WANG, Jinjing: Automatic recognition vision system guided for apple harvesting robot. En: *Computers and Electrical Engineering* 38 (2012), Nr. 5, p. 1186–1195. – ISBN 0045–7906
- [56] JIMÉNEZ, a R. ; CERES, R ; PONS, J L.: A survey of computer vision methods for locating fruit on trees. En: *Transactions of the American Society of Agricultural Engineers* 43 (2000), Nr. 6, p. 1911–1920. – ISBN 00012351 (ISSN)
- [57] JOSHI, P ; BRIMELOW, CJB: Colour measurement of foods by colour reflectance. En: *Colour in Foods: Improving Quality* (2002), p. 80–114. ISBN 978–1–85573–590–3
- [58] JULIÁN L., Adriana P.: *Propiedades Físicas Y Químicas De Tres variedades del fruto de annona diversifolia*, Tesis de Grado, 2009. – 91 p.
- [59] KAVDIR, I ; GUYER, D E.: Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. En: *Biosystems Engineering* 89 (2004), Nr. 3, p. 331–344. – ISBN 1537–5110
- [60] KAVDIR, I. ; GUYER, D. E.: Evaluation of different pattern recognition techniques for apple sorting. En: *Biosystems Engineering* 99 (2008), Nr. 2, p. 211–219. – ISBN 1537–5110
- [61] KAY, G. ; JAGER, G. D.: A versatile colour system capable of fruit sorting and accurate\object classification. En: *Proceedings of the 1992 South African Symposium on Communications and Signal Processing* (1992), p. 145–148. ISBN 0–7803–0807–7
- [62] KONDO, Naoshi ; AHMAD, Usman ; MONTA, Mitsuji ; MURASE, Haruhiko: Machine vision based quality evaluation of Iyokan orange fruit using neural networks. En: *Computers and Electronics in Agriculture* 29 (1-2) (2000), p. 135–147.. – ISSN 01681699
- [63] LEAHU, Ana ; DAMIAN, Cristina ; OROIAN, M ; ROPCIUC, Sorina: Physico-Chemical Parameters of Fruit Juices - Evolution During Storage. En: *Lucrari Stiintifice-Seria Zootehnie* 59 (2013), p. 213–217
- [64] LEEMANS, V. ; MAGEIN, H. ; DESTAIN, M. F.: Defects segmentation on 'Golden Delicious' apples by using colour machine vision. En: *Computers and Electronics in Agriculture* 20 (1998), Nr. 2, p. 117–130. – ISSN 01681699

- 
- [65] LEMASURIER, L. S. ; PANOZZO, J. F. ; WALKER, C. K.: A digital image analysis method for assessment of lentil size traits. En: *Journal of Food Engineering* 128 (2014), p. 72–78. – ISBN 0260–8774
- [66] LEÓN, Katherine ; MERY, Domingo ; PEDRESCHI, Franco ; LEÓN, Jorge: Color measurement in L\*a\*b\* units from RGB digital images. En: *Food Research International* 39 (2006), Nr. 10, p. 1084–1091. – ISBN 0963–9969
- [67] LI, Qingzhong ; WANG, Maohua ; GU, Weikang: Computer vision based system for apple surface defect detection. En: *Computers and Electronics in Agriculture* 36 (2002), Nr. 2-3, p. 215–223. – ISBN 0168–1699
- [68] LI, Xianfeng ; ZHU, Weixing: Apple grading method based on features fusion of size, shape and color. En: *Procedia Engineering* 15 (2011), p. 2885–2891. – ISBN 8605158816
- [69] LIU, Yang ; GEVERS, Theo ; LI, Xueqing: Color Constancy by Combining Low-Mid-High Level Image Cues. En: *Computer Vision and Image Understanding* (2015). – ISSN 10773142
- [70] MAKKY, Muhammad ; SONI, Peeyush: Development of an automatic grading machine for oil palm fresh fruits bunches (FFBs) based on machine vision. En: *Computers and Electronics in Agriculture* 93 (2013), p. 129–139. – ISSN 01681699
- [71] MANICKAVASAGAN, A. ; AL-MEZEINI, N. K. ; AL-SHEKAILI, H. N.: RGB color imaging technique for grading of dates. En: *Scientia Horticulturae* 175 (2014), p. 87–94. – ISBN 0304–4238
- [72] MANNINEN, Hanna ; PAAKKI, Maija ; HOPIA, Anu ; FRANZÉN, Robert: Measuring the green color of vegetables from digital images using image analysis. En: *LWT - Food Science and Technology* 63 (2015), Nr. 2, p. 1184–1190. – ISSN 00236438
- [73] MARTIS, Roshan J. ; ACHARYA, U. R. ; MIN, Lim C.: ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. En: *Biomedical Signal Processing and Control* 8 (2013), Nr. 5, p. 437–448. – ISSN 17468108
- [74] MEIRELES, Magali R G. ; ALMEIDA, Paulo E M. ; SIMÕES, Marcelo G.: A comprehensive review for industrial applicability of artificial neural networks. En: *IEEE Transactions on Industrial Electronics* 50 (2003), Nr. 3, p. 585–601. – ISBN 0278–0046
- [75] MENDOZA, Fernando ; LU, Renfu ; CEN, Haiyan: Grading of apples based on firmness and soluble solids content using Vis/SWNIR spectroscopy and spectral scattering techniques. En: *Journal of Food Engineering* 125 (2014), Nr. 1, p. 59–68. – ISSN 02608774
- [76] MITCHAM, Beth ; CANTWELL, Marita ; KADER, Adel: Methods for determining quality of fresh commodities. En: *Perishables handling newsletter* (1996), Nr. 85, p. 1–5

- 
- [77] MIZUSHIMA, Akira ; LU, Renfu: An image segmentation method for apple sorting and grading using support vector machine and Otsu's method. En: *Computers and Electronics in Agriculture* 94 (2013), p. 29–37. – ISBN 01681699
- [78] NGUYEN, Tien T. ; VANDEVOORDE, Koenraad ; WOUTERS, Niels ; KAYACAN, Erdal ; DE BAERDEMAEKER, Josse G. ; SAEYS, Wouter: Detection of red and bicoloured apples on tree with an RGB-D camera. En: *Biosystems Engineering* (2016), p. 1–12. – ISSN 15375110
- [79] NICOLAÏ, Bart M. ; BEULLENS, Katrien ; BOBELYN, Els ; PEIRS, Ann ; SAEYS, Wouter ; THERON, Karen I. ; LAMMERTYN, Jeroen: Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. En: *Postharvest Biology and Technology* 46 (2007), Nr. 2, p. 99–118. – ISBN 0925–5214
- [80] OHTA, Y ; KANADE, T ; SAKAI, T. *Color Information for Region Segmentation*. 1980
- [81] OMID, M. ; KHOJASTEHNAZHAND, M. ; TABATABAEEFAR, A.: Estimating volume and mass of citrus fruits by image processing technique. En: *Journal of Food Engineering* 100 (2010), Nr. 2, p. 315–321. – ISSN 02608774
- [82] OU, Xiang ; PAN, Wei ; XIAO, Perry: In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). En: *International Journal of Pharmaceutics* 460 (2014), Nr. 1-2, p. 28–32. – ISSN 03785173
- [83] PAYNE, A B. ; WALSH, K B. ; SUBEDI, P P. ; JARVIS, D: Estimation of mango crop yield using image analysis - Segmentation method. En: *Computers and Electronics in Agriculture* 91 (2013), p. 57–64. – ISSN 01681699
- [84] PINZÓN, Ingrid M. ; FISHER, Gerard ; CORREDOR, Guillermo: Determination of the maturity stages of purple passion fruit (*Passiflora edulis* Sims.). En: *Agronomía Colombiana* 25 (2007), Nr. 1, p. 83–95
- [85] PORTNOY, Ivan ; MELENDEZ, Kevin ; PINZON, Horacio ; SANJUAN, Marco: An improved weighted recursive PCA algorithm for adaptive fault detection. En: *Control Engineering Practice* 50 (2016), p. 69–83. – ISSN 09670661
- [86] PREACHER, K. J. ; CURRAN, P. J. ; BAUER, D. J.: Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. En: *Journal of Educational and Behavioral Statistics* 31 (2006), Nr. 4, p. 437–448. – ISBN 1076–9986
- [87] QUAN, Zhang ; ZHENYU, Gu ; XINGUO, Ming: A modified color appearance model of CIELAB. En: *Proceedings - 2009 International Asia Conference on Informatics in Control, Automation, and Robotics, CAR 2009* (2009), p. 339–344. ISBN 9780769535197

- 
- [88] DE QUEIROZ, R L.: Color transformation for the compression of CMYK images. En: *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts V* SPIE-3963 (1999), Nr. 10, p. 221–229. – ISSN 0277786X
- [89] RIAZ, Z ; GILGITI, A ; MIRZA, S: Face Recognition: A review and comparison of HMM, PCA, ICA and Neural Networks. En: *Pakistan Institute of Engineering and Applied Sciences* (2004), p. 41–46. ISBN 0780386558
- [90] ROCHA, Anderson ; HAUAGGE, Daniel C. ; WAINER, Jacques ; GOLDENSTEIN, Siome: Automatic fruit and vegetable classification from images. En: *Computers and Electronics in Agriculture* 70 (2010), Nr. 1, p. 96–104. – ISBN 0168–1699
- [91] ROSELEENA, J ; NURSURIATI, J ; AHMED, J ; LOW, C Y.: Assessment of palm oil fresh fruit bunches using photogrammetric grading system. En: *International Food Research Journal* 18 (2011), Nr. 3, p. 999–1005. – ISSN 19854668
- [92] SA'AD, F.S.A. ; IBRAHIM, M.F. ; SHAKAFF, A.Y.Md. ; ZAKARIA, A. ; ABDULLAH, M.Z.: Shape and weight grading of mangoes using visible imaging. En: *Computers and Electronics in Agriculture* 115 (2015), p. 51–56. – ISSN 01681699
- [93] SABZI, Sajad ; JAVADIKIA, Payam ; RABANI, Hekmat ; ADELKHANI, Ali: Mass modeling of Bam orange with ANFIS and SPSS methods for using in machine vision. En: *Measurement: Journal of the International Measurement Confederation* 46 (2013), Nr. 9, p. 3333–3341. – ISSN 02632241
- [94] SALAMANCA, G. ; LONGAS, F. F. ; LOZANO, J. G. ; DÍAZ, C. ; SALAZAR, B.: Avances en la caracterización , conservación y procesamiento del mango ( *Mangifera indica* L .) en Colombia. En: *Revista Tumbaga* 2 (2007), Nr. 2007, p. 57–64
- [95] SALAZAR, S M. ; SANDOVAL, J S. ; KIRSCHBAUM, D S.: Correlación entre la medida del color del fruto y la concentración de sólidos solubles totales en frutilla o fresa ( . 35 (2015), p. 55–60
- [96] SALDAÑA ROBLES, Alberto ; SERWATOWSKI HLAWINSKA, Ryszard J. ; SALDAÑA ROBLES, Noé ; GUTIÉRREZ VACA, César ; CABRERA SIXTO, José Manuel ; GARCÍA BARRÓN, Salvador: Determinación de algunas propiedades físicas de Agave tequilana Weber para mecanizar la cosecha. En: *Revista mexicana de ciencias agrícolas* 3 (2012), Nr. 3, p. 451–465
- [97] SANTACRUZ, Liliana ; OSOSRIO, Coralia ; GONZÁLEZ-MIRET, María ; HEREIDA, Francisco: Aplicación de la colorimetría triestímulo para evaluar la estabilidad de extractos ricos en antocianinas de frutas tropicales colombianas. En: *Vitae* 19 (2012), Nr. 2, p. 93–96

- 
- [98] SANTONI, Mayanda M. ; SENSUSE, Dana I. ; ARYMURTHY, Aniati M. ; FANANY, Mohamad I.: Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks. En: *Procedia Computer Science* 59 (2015), Nr. Iccsci, p. 493–502. – ISSN 18770509
- [99] SCHULZE, Katrin ; NAGLE, Marcus ; SPREER, Wolfram ; MAHAYOTHEE, Busarakorn ; MÜLLER, Joachim: Development and assessment of different modeling approaches for size-mass estimation of mango fruits (*Mangifera indica* L., cv. ‘Nam Dokmai’). En: *Computers and Electronics in Agriculture* 114 (2015), p. 269–276. – ISSN 01681699
- [100] SEGERBLAD, Emil ; DELIGHT, B: Machine Vision in Agricultural Robotics – A short overview. En: *School of Innovation, Design and Engineering* (2011), p. 1–6
- [101] SHAHIN, M A. ; TOLLNER, E W. ; MCCLENDON, R W.: AE—Automation and Emerging Technologies. En: *Journal of Agricultural Engineering Research* 79 (2001), Nr. 3, p. 265–274. – ISBN 1537–5110
- [102] SHARMA, Alok ; PALIWAL, Kuldip K. ; ONWUBOLU, Godfrey C.: Class-dependent PCA, MDC and LDA: A combined classifier for pattern classification. En: *Pattern Recognition* 39 (2006), Nr. 7, p. 1215–1229. – ISSN 00313203
- [103] SILVA-AYALA, Tomas ; SCHNELL, Raymond J. ; MEEROW, Alan W. ; WINTERSTEIN, Michael ; CERVANTES, Cuauhtemoc ; BROWN, J. S.: Determination of Color and Fruit Traits of Half-Sib Families of Mango (*Mangifera Indica* L.). En: *Proceedings of the Florida State Horticultural Society* 118 (2005), Nr. June 1996, p. 253–257
- [104] SPREER, Wolfram ; MÜLLER, Joachim: Estimating the mass of mango fruit (*Mangifera indica*, cv. Chok Anan) from its geometric dimensions by optical measurement. En: *Computers and Electronics in Agriculture* 75 (2011), Nr. 1, p. 125–131. – ISSN 01681699
- [105] SUN, Feifei ; YU, Qingni ; ZHU, Jingke ; LEI, Lecheng ; LI, Zhongjian ; ZHANG, Xingwang: Measurement and ANN prediction of pH-dependent solubility of nitrogen-heterocyclic compounds. En: *Chemosphere* 134 (2015), p. 402–407. – ISBN 1879–1298
- [106] TAHIR, M. A. ; ROULA, M. A. ; BOURIDANE, A. ; KURUGOLLU, F. ; AMIRA, A.: An fpga based co-processor for glcm texture. (2003), p. 1006–1009. ISBN 0780381637
- [107] TIRYAKI, Sebahattin ; ÖZŞAHİN, Şükrü ; YILDIRIM, İbrahim: Comparison of artificial neural network and multiple linear regression models to predict optimum bonding strength of heat treated woods. En: *International Journal of Adhesion and Adhesives* 55 (2014), p. 29–36. – ISSN 01437496

- 
- [108] UNAY, D. ; GOSSELIN, B.: Stem and calyx recognition on 'Jonagold' apples by pattern recognition. En: *Journal of Food Engineering* 78 (2007), Nr. 2, p. 597–605. – ISBN 0260–8774
- [109] VAN DE WEIJER, Joost ; GEVERS, Theo ; GIJSENIJ, Arjan: Edge-Based Color Constancy. 16 (2007), Nr. 9, p. 2207–2214
- [110] VAN DE WEIJER, Joost ; SCHMID, Cordelia ; VERBEEK, Jakob: Using high-level visual information for color constancy. En: *Proceedings of the IEEE International Conference on Computer Vision* (2007). – ISBN 978–1–4244–1631–8
- [111] VIDYA, K. S. ; NG, E. Y K. ; ACHARYA, U. R. ; CHOU, Siaw M. ; TAN, Ru S. ; GHIS-  
TA, Dhanjoo N.: Computer-aided diagnosis of Myocardial Infarction using ultrasound  
images with DWT, GLCM and HOS methods: A comparative study. En: *Computers  
in Biology and Medicine* 62 (2015), p. 86–93. – ISBN 0010–4825
- [112] WANG, Weilin ; LI, Changying: A Multimodal Machine Vision System For Quality  
Inspection of Onions. En: *Journal of Food Engineering* 166 (2015), p. 291–301. – ISSN  
02608774
- [113] WANG, Yizhong ; CUI, Yanhua ; HUANG, George Q. ; ZHANG, Ping ; CHEN, Shaohui:  
Study on fruit quality inspection based on its surface color in produce logistics. En:  
*Proceedings - 2010 International Conference on Manufacturing Automation, ICMA  
2010* (2010), p. 107–111. ISBN 9780769542935
- [114] WATADA, A.E.: Methods of determining quality of fruits and vegetables. En: *Acta  
Horticulturae* 379 (1995), p. 559–567
- [115] WEI, Xiangqin ; JIA, Kun ; LAN, Jinhui ; LI, Yuwei ; ZENG, Yiliang ; WANG, Chunmei:  
Automatic method of fruit object extraction under complex agricultural background  
for vision system of fruit picking robot. En: *Optik - International Journal for Light  
and Electron Optics* 125 (2014), Nr. 19, p. 5684–5689. – ISSN 00304026
- [116] WEN, Zhiqing ; TAO, Yang: Building a rule-based machine-vision system for defect  
inspection on apple sorting and packing lines. En: *Expert Systems with Applications*  
16 (1999), Nr. 3, p. 307–313. – ISBN 1501575287
- [117] XIAO-BO, Zou ; JIE-WEN, Zhao ; YANXIAO, Li ; HOLMES, Mel: In-line detection of  
apple defects using three color cameras system. En: *Computers and Electronics in  
Agriculture* 70 (2010), Nr. 1, p. 129–134. – ISBN 0168–1699
- [118] YAM, Kit L. ; PAPADAKIS, Spyridon E.: A simple digital imaging method for measuring  
and analyzing color of food surfaces. En: *Journal of Food Engineering* 61 (2004), Nr.  
1 SPEC., p. 137–142. – ISSN 02608774

- 
- [119] YAM TZEC, José A. ; VILLASEÑOR PEREA, Carlos A. ; KRIUCHKOVA ROMANTCHIK, Eugenio ; SOTO ESCOBAR, Martín ; PEÑA PERALTA, Miguel A.: Una revisión sobre la importancia del fruto de Guayaba (*Psidium guajava* L.) y sus principales características en la postcosecha. En: *Revista Ciencias Técnicas Agropecuarias* 19 (2010), Nr. 4, p. 74–82
- [120] YE, Fei ; SHI, Zhiping ; SHI, Zhongzhi: A comparative study of PCA, LDA and kernel LDA for image classification. En: *Proceedings - 2009 International Symposium on Ubiquitous Virtual Reality, ISUVR 2009* (2009), p. 51–54. ISBN 9780769537047
- [121] ZHANG, Baohua ; HUANG, Wenqian ; GONG, Liang ; LI, Jiangbo ; ZHAO, Chunjiang ; LIU, Chengliang ; HUANG, Danfeng: Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. En: *Journal of Food Engineering* 146 (2015), p. 143–151. – ISSN 02608774
- [122] ZHANG, Baohua ; HUANG, Wenqian ; LI, Jiangbo ; ZHAO, Chunjiang ; FAN, Shuxiang ; WU, Jitao ; LIU, Chengliang: Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. En: *Food Research International* 62 (2014), p. 326–343. – ISSN 09639969
- [123] ZHANG, Erhu ; YU, Jia ; CHEN, Yajun ; YANG, Yang: A study of image color quality evaluation based on S-CIELAB. En: *Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010* 3 (2010), p. 1110–1114. ISBN 9781424465149
- [124] ZHANG, Yudong ; WANG, Shuihua ; JI, Genlin ; PHILLIPS, Preetha: Fruit classification using computer vision and feedforward neural network. En: *Journal of Food Engineering* 143 (2014), p. 167–177. – ISSN 02608774
- [125] ZHAO, Jun ; TOW, Joel ; KATUPITIYA, Jayantha: On-tree fruit recognition using texture properties and color data. En: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS* (2005), p. 3993–3998. ISBN 0780389123
- [126] ZULHUSIN ; AZIZ, A Hallis A. ; AHMAD, R B.: Feasibility study of a non-destructive fruit maturity testing system on banana utilizing capacitive properties. En: *2008 International Conference on Electronic Design, ICED 2008* (2008), p. 1–4. ISBN 9781424423156
- [127] ŽUNIĆ, Dragiša ; ŽUNIĆ, Joviša: Shape ellipticity based on the first Hu moment invariant. En: *Information Processing Letters* 113 (2013), Nr. 19-21, p. 807–810. – ISSN 00200190
- [128] ŽUNIĆ, Joviša ; HIROTA, Kaoru ; ROSIN, Paul L.: A Hu moment invariant as a shape circularity measure. En: *Pattern Recognition* 43 (2010), Nr. 1, p. 47–57. – ISBN 0031–3203