

UNIVERSIDAD PRIVADA ANTENOR ORREGO
FACULTAD DE CIENCIAS AGRARIAS
ESCUELA PROFESIONAL DE INGENIERÍA EN
INDUSTRIAS ALIMENTARIAS



**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO EN INDUSTRIAS ALIMENTARIAS**

**“Modelos de visión por computadora para la determinación de
características superficiales y categorías de calidad del Pallar
(*Phaseolus lunatus* L.) lqueño seco”**

Área de Investigación:

Visión Artificial - Tecnología de procesos analíticos y métodos
no destructivos de la industria alimentaria.

Autor(es):

Br. Ever Omar Cortez Agreda

Jurado Evaluador:

Presidente: Dr. Vásquez Villalobos, Víctor Javier

Secretario: Dr. Pérez Azahuanche, Freddy Romel

Vocal: Ms. Márquez Villacorta, Luis Francisco

Asesor:

PhD. Lescano Anadón, Carlos Eduardo

Código Orcid: <https://orcid.org/0000-0003-4147-9655>

Co-asesor:

Dr. Mendoza Vilcarromero, Fernando Antonio

Código Orcid: <https://orcid.org/0000-0001-6481-2778>

**TRUJILLO – PERÚ
2021**

Fecha de sustentación: 2020/10/23

La presente tesis ha sido revisada y aprobada por el siguiente Jurado:



Dr. Víctor Javier Vásquez Villalobos
PRESIDENTE



Dr. Freddy Romel Pérez Azahuanche
SECRETARIO



Ms. Luis Francisco Márquez Villacorta
VOCAL



PhD. Carlos Eduardo Lescano Anadón
ASESOR

DEDICATORIA

A mi madre Maximina Dionicia Agreda Valverde, a mi hijo Ever Eduardo Cortez Díaz, a mi padre Francisco Javier Cortez Marino, a mis hermanos Harly, Francisco y Thalia por ser fuente de inspiración, y a mi recordado maestro Dr. Héctor Eduardo Flores Cevallos por sus consejos.

AGRADECIMIENTO

A Dios, en primer lugar y por sobre todas las cosas, por este logro, camino de aprendizaje y superación, a mi madre por su confianza incondicional, y a mi familia en general, por ser fuente de superación.

En forma particular al PhD Carlos Lescano, por darme la oportunidad de realizar esta investigación, con su asesoría y dirección; al Dr. Fernando Mendoza por su asesoría, enseñanzas y consejos; al Sr. Luis Miguel Guerrero por su colaboración en este proyecto, a quien le estoy muy agradecido.

Así mismo, quiero reconocer el apoyo profesional de la Ing. María Luisa Hayayumi Valdivia, de los docentes y compañeros de estudios de la Escuela Profesional de Ingeniería en Industrias Alimentarias de la Facultad de Ciencias Agrarias.

ÍNDICE GENERAL

Pág.

Autor(es):.....	i
Jurado Evaluador:	i
Asesor:.....	i
2021	i
APROBACIÓN POR EL JURADO DE TESIS	ii
DEDICATORIA	iii
AGRADECIMIENTO	iv
ÍNDICE GENERAL	v
ÍNDICE DE CUADROS.....	ix
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE ANEXOS	xii
RESUMEN	xiii
ABSTRACT	xv
I. INTRODUCCIÓN.....	1
II. REVISIÓN DE BIBLIOGRAFÍA.....	5
2.1. El pallar.....	5
2.1.1. Origen.....	5
Cuadro 1. Hallazgos arqueológicos de especies cultivadas del género Phaseolus.....	6
2.1.2. Clasificación botánica según cultigrupo	6
Cuadro 2. Clasificación taxonómica del pallar.....	7
Cuadro 3. Clasificación de pallar según cultigrupo	7
2.1.3. Clasificación de pallares comerciales en el Perú, tipo Big Lima.....	7
Cuadro 4. Pallares comerciales tipo Big Lima en el Perú	9
2.1.4. Características sanitarias y de aspecto del pallar	9
Cuadro 5. Tipos de defectos del pallar.....	10
2.2. Visión por computadora (VC).....	10
2.2.1. Conceptos básicos	10
2.2.1.1. Imagen digital	10

2.2.1.2.	Modelos de descripción numérica de espacios de color.....	11
	Figura 1. Descripción numérica de espacios de color para el análisis de imágenes	12
2.2.1.3.	Textura de imagen	12
2.2.1.4.	Modelos de caracterización y clasificación de imágenes.....	13
	Figura 2. Proceso de reconocimiento y clasificación de patrones por Visión por Computadora.	14
	Figura 3. Configuración general de un modelo de clasificación por Visión por Computadora.	15
2.2.2.	Elementos básicos de un sistema de visión por computadora.....	15
2.2.3.	Visión por computadora aplicada a los alimentos.....	17
2.3.	Análisis Discriminante.....	19
III.	MATERIALES Y MÉTODOS	22
3.1.	Lugar de ejecución.....	22
3.2.	Materiales y equipos	22
3.2.1.	Pallar lqueño seco.....	22
3.2.2.	Sistema de visión por computadora	22
	Figura 4. Cámara de iluminación para la VC.....	24
	Figura 5. Cámara digital y estabilizador de corriente.....	25
	Figura 6. Vista en computador de captura remota de imagen con software especializado.	26
3.3.	Metodología Experimental.....	26
3.3.1.	Identificación y caracterización de pallares según defectos de calidad mediante inspección visual	27
	Figura 7. Procedimiento experimental para la investigación.....	28
3.3.2.	Calibración y manejo de la cámara digital	29
	Figura 8. Colorchecker clásica X-Rite	30
	Cuadro 6. Parámetros de captura de imagen.....	32
	en la cámara digital	32
3.3.3.	Procesamiento y análisis de imágenes.....	33
	Figura 9. Transformación de imágenes RGB a imágenes binarias.....	33
	Figura 10. Proceso de segmentación de imágenes.....	34
	Figura 11. Conteo del número de pallares por cada imagen digital.....	35
3.3.4.	Identificación descriptiva de tipos de defectos del pallar mediante visión por computadora	36

3.3.5. Clasificación en categorías de calidad del pallar mediante modelos discriminantes	38
3.4. Métodos de análisis.....	41
3.4.1. Inspección visual humana de las muestras	41
Figura 12. Identificación de defectos del Pallar lqueño seco por inspección visual humana.....	42
3.4.2. Evaluación de descriptores superficiales del pallar mediante visión por computadora.....	44
3.5. Análisis estadístico.....	45
IV. RESULTADOS Y DISCUSIÓN	47
4.1. Identificación de pallares según defectos de calidad mediante inspección visual humana.....	47
Cuadro 7. Distribución del pallar muestra según defectos de calidad	48
4.2. Calibración de cámara digital.....	48
Figura 13. Correlación L^* de la cámara digital vs Colorchecker	49
4.3. Descriptores y clases discriminantes de clasificación por visión por computadora	49
Cuadro 8. Descriptores extraídos del análisis de imagen del pallar	51
Cuadro 9. Definición de clases discriminantes de clasificación.....	53
4.4. Caracterización descriptiva del pallar mediante visión por computadora.....	54
4.4.1. Medidas de parámetros de calidad del pallar	54
Cuadro 10. Medida de descriptores superficiales de tipos de defectos del pallar	57
Figura 14. Box-plot de parámetros de color $L^*a^*b^*$, contraste, homogeneidad, largo y redondez de los defectos del Pallar lqueño seco.....	58
4.4.2. Análisis estadísticos de parámetros de calidad del pallar.....	59
Cuadro 11. ANVA de parámetros de color, apariencia, tamaño y forma.....	59
de los tipos de defectos del Pallar lqueño seco.	59
Cuadro 12. LSD de Fisher de parámetros de color, apariencia, tamaño y forma de los tipos de defectos del Pallar lqueño seco	60
4.4.3. Identificación descriptiva de defectos del pallar mediante color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez.	62
Figura 15. Límites en la gráfica Box-plot del color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez	63
Cuadro 13. Valores de descriptores de calidad del pallar lqueño seco.....	65
Figura 16. Cartilla de identificación y caracterización del Pallar lqueño seco	66
Figura 17. Carta de matices de blanco del Pallar lqueño.....	67

4.5. Clasificación del pallar por imágenes digitales.....	68
4.5.1. Clasificación en 2 Categorías de calidad.....	68
Figura 18. Modelos discriminantes de clasificación en 2 categorías de calidad.....	69
Figura 19. Gráfico 3D del modelo LDA de la clasificación en 2 Categorías de Calidad.....	70
4.5.2. Clasificación en 3 Categorías de Calidad.....	71
Figura 20. Modelos discriminantes de clasificación en 3 categorías de calidad.....	72
Figura 21. Gráfico 3D del modelo LDA de la clasificación en 3 Categorías de Calidad.....	73
4.5.3. Clasificación en 5 Categorías de Calidad.....	73
Figura 22. Modelos discriminantes de clasificación en 5 categorías de calidad.....	75
Figura 23. Gráfico 3D del modelo LDA de la clasificación en 5 Categorías de Calidad.....	76
4.5.4. Performance de los modelos de clasificación del pallar.....	76
Cuadro 14. Eficiencia General y Eficiencia de Clase de los modelos de clasificación.....	77
4.5.5. Sensibilidad de los modelos de clasificación del pallar.....	78
Figura 24. Sensibilidad de Eficiencia General de modelos LDA y QDA.....	78
V. CONCLUSIONES.....	79
VI. RECOMENDACIONES.....	80
VII. BIBLIOGRAFÍA.....	81
VIII. ANEXOS.....	88

ÍNDICE DE CUADROS

	Pág.
Cuadro 1. Hallazgos arqueológicos de especies cultivadas del género Phaseolus.....	6
Cuadro 2. Clasificación taxonómica del pallar	7
Cuadro 3. Clasificación de pallar según cultigrupo	7
Cuadro 4. Pallares comerciales tipo Big Lima en el Perú	9
Cuadro 5. Tipos de defectos del pallar	10
Cuadro 6. Parámetros de captura de imagen en la cámara digital	32
Cuadro 7. Distribución del pallar muestra según defectos de calidad..	48
Cuadro 8. Descriptores extraídos del análisis de imagen del pallar	51
Cuadro 9. Definición de clases discriminantes de clasificación	53
Cuadro 10. Medida de descriptores superficiales de tipos de defectos del pallar.....	57
Cuadro 11. ANVA de parámetros de color, apariencia, tamaño y forma de los tipos de defectos del Pallar Iqueño seco.	59
Cuadro 12. LSD de Fisher de parámetros de color, apariencia, tamaño y forma de los tipos de defectos del Pallar Iqueño seco	60
Cuadro 13. Valores de descriptores de calidad del pallar Iqueño seco.....	65
Cuadro 14. Eficiencia General y Eficiencia de Clase de los modelos de clasificación	77

ÍNDICE DE FIGURAS

	Pág.
Figura 1. Descripción numérica de espacios de color para el análisis de imágenes	12
Figura 2. Proceso de reconocimiento y clasificación de patrones por Visión por Computadora.	14
Figura 3. Configuración general de un modelo de clasificación por Visión por Computadora	15
Figura 4. Cámara de iluminación para la VC	24
Figura 5. Cámara digital y estabilizador de corriente	25
Figura 6. Vista en computador de captura remota de imagen con software especializado.....	26
Figura 7. Procedimiento experimental para la investigación	28
Figura 8. Colorchecker clásica X-Rite	30
Figura 9. Transformación de imágenes RGB a imágenes binarias	33
Figura 10. Proceso de segmentación de imágenes	34
Figura 11. Conteo del número de pallares por cada imagen digital	35
Figura 12. Identificación de defectos del Pallar lqueño seco por inspección visual humana.....	42
Figura 13. Correlación L^* de la cámara digital vs Colorchecker	49
Figura 14. Box-plot de parámetros de color $L^*a^*b^*$, contraste, homogeneidad, largo y redondez de los defectos del Pallar lqueño seco.	58
Figura 15. Límites en la gráfica Box-plot del color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez	63
Figura 16. Cartilla de identificación y caracterización del Pallar lqueño seco	66
Figura 17. Carta de matices de blancura del Pallar lqueño seco	67
Figura 18. Modelos discriminantes de clasificación en 2 categorías de calidad	69

Figura 19. Gráfico 3D del modelo LDA de la clasificación en 2 Categorías de Calidad	70
Figura 20. Modelos discriminantes de clasificación en 3 categorías de calidad	72
Figura 21. Gráfico 3D del modelo LDA de la clasificación en 3 Categorías de Calidad	73
Figura 22. Modelos discriminantes de clasificación en 5 categorías de calidad	75
Figura 23. Gráfico 3D del modelo LDA de la clasificación en 5 Categorías de Calidad	76
Figura 24. Sensibilidad de Eficiencia General de modelos LDA y QDA.....	78

ÍNDICE DE ANEXOS

	Pág.
ANEXO 1. Algoritmos de procesamiento de imágenes	88
ANEXO 2. Algoritmos de análisis de imágenes.....	90
ANEXO 3. Algoritmos de aplicación de modelos de clasificación discriminante	94

RESUMEN

En esta investigación se ha desarrollado, a nivel de laboratorio un Sistema de Visión por Computadora (SVC) que capta imágenes con una cámara fotográfica digital a color conectada a una computadora personal portátil que, utilizando software especializado, algoritmos estadísticos multivariantes y aquellos disponibles en MATLAB, realiza el procesamiento, análisis y clasificación de manera no-invasiva, objetiva y automática de los granos del Pallar (*Phaseolus lunatus* L.) lqueño seco. La clasificación en 2, 3 y 5 categorías de calidad del pallar, el SVC lo realiza utilizando modelos discriminantes simples: lineal (LDA) y cuadrático (QDA).

Los modelos de clasificación discriminante lineal (LDA) y cuadrático (QDA), aplicados a 3169 imágenes de granos de Pallar lqueño seco, agrupados en 2, 3 y 5 categorías de calidad, para todos los 6 casos estudiados, mostraron performances con una Eficiencia General mayor al 80% para los modelos discriminantes lineal y cuadrático, demostrando un buen desempeño.

El mejor performance general lo obtuvo el modelo LDA, clasificación en 3 categorías de calidad, con una Eficiencia General de 85.6%, el que también mostró una Eficiencia de Clase de 95.6%, 97.5% y 96.7% en la clasificación por cada categoría individual: ÓPTIMO, TOLERANTE y DESCARTE.

Así mismo, todos los modelos de clasificación estudiados, demostraron ser robustos al mantener sin variación importante los valores de Eficiencia General, los que estuvieron por encima del 75% en un balance de muestras entre 10% a 80% para la validación cruzada de 2, 3 y 5 categorías de calidad.

Los datos de los descriptores de defectos superficiales del grano, obtenidos por visión por computadora permitieron elaborar una carta sobre la identificación y caracterización del Pallar Iqueño en tres clases, lo mismo que una cartilla con matices de blancura de utilidad para los interesados.

Los resultados de la investigación muestran que los modelos discriminantes de clasificación lineales y cuadráticos en visión por computadora son eficaces en emular la visión humana experta, en la evaluación y clasificación objetiva de la calidad de muestras de Pallar Iqueño seco.

ABSTRACT

In this research, a Computer Vision System (SVC) has been developed at the laboratory level that captures images with a color digital camera connected to a portable personal computer that, using specialized software, multivariate statistical algorithms and those available in MATLAB, it performs the processing, analysis and classification, in a non-invasive, objective and automatic way of the grains of the dry Pallar (*Phaseolus lunatus* L.) Iqueño. The classification in 2, 3 and 5 categories of quality of the pallar, the SVC performs it using simple discriminant models: linear (LDA) and quadratic (QDA).

The linear (LDA) and quadratic (QDA) discriminant classification models, applied to 3169 images of dry Pallar Iqueño grains, grouped into 2, 3 and 5 quality categories, for all the 6 cases studied, showed performances with a General Efficiency greater than 80% for the linear and quadratic discriminant models, showing good performance.

The best general performance was obtained by the LDA model, classified in 3 quality categories, with a General Efficiency of 85.6%, which also showed a Class Efficiency of 95.6%, 97.5% and 96.7% in the classification for each individual category: OPTIMAL, TOLERANT and DISCARD.

Likewise, all the classification models studied proved to be robust by maintaining, without significant variation, the general efficiency values, which were above 75% in a balance of samples, between 10% and 80% for validation. cross of 2, 3 and 5 quality categories.

The data of the descriptors of surface defects of the grain, obtained by computer vision, allowed to elaborate a letter on the identification and

characterization of Pallar Iqueño in three classes, as well as a primer with shades of whiteness, useful for those interested.

The research results show that the discriminant linear and quadratic classification models in computer vision are effective in emulating expert human vision, in the objective evaluation and classification of the quality of samples from Pallar Iqueño seco.

I. INTRODUCCIÓN

El grano de pallar (*Phaseolus lunatus* L) tiene gran importancia agroecológica, económica y culinaria en el Perú y el mundo, principalmente por su alto nivel nutritivo y energético y como fuente de proteína vegetal. La aceptación y valor comercial del pallar, se basa en la inspección visual del color, en atributos como su apariencia o aspecto general, tono de blancura y homogeneidad de color; daños o picaduras y en características morfológicas de tamaño y forma; es decir, en su calidad externa que representa la primera impresión que tienen comerciantes y consumidores de la calidad del grano, que es un factor crítico y decisivo para la aceptabilidad del mercado, estabilidad del almacenamiento, calidad del procesamiento y la aceptabilidad general del consumidor (Vithu y Moses, 2016).

La industria de granos comestibles, particularmente la del pallar requieren de estándares de clasificación acordes con las exigencias del mercado nacional e internacional y con técnicas de medición objetiva no-destructivas, que permitan una evaluación precisa, consistente y rápida en la inspección de calidad externa de las muestras (Mahajan y otros, 2015). Una alternativa interesante, son las técnicas de Visión por Computadora (VC) usando cámaras digitales a color (CDC), que son eficaces en proporcionar información precisa, reproducible y rápida de varios aspectos de la calidad externa y, en especial de aspectos superficiales de los alimentos, de coloración no homogénea y/o de formas curvadas que requieren ser definidas, como sucede en las frutas, hortalizas y granos (Mendoza, 2005; Mahajan y otros, 2015; Mendoza y otros, 2016; Vithu y Moses, 2016). La inspección del color mediante un colorímetro o un espectrofotómetro comparada con una imagen a color tiene la desventaja de dar un valor promedio que no describe adecuadamente el color de la

muestra, debido a una distribución no uniforme del color, es decir, estos equipos no realizan segmentación ni clasificación de las imágenes muestreadas (Padron, 2009); sin embargo, la técnica de VC, por si sola no provee información objetiva de los atributos de calidad interna de las muestras evaluadas (Mendoza y otros, 2016).

La técnica VC en el área de alimentos permite básicamente la caracterización de superficies complejas con gran detalle y precisión, el reconocimiento y segmentación de regiones de interés en la muestra y, finalmente, la predicción y clasificación de muestras en categorías de calidad mediante modelos estadísticos multivariantes (análisis de componentes principales, análisis de factores, análisis discriminante, redes neuronales, entre otros) (Mendoza y otros, 2006). Varios estudios han mostrado que las técnicas de VC superan la eficacia de la inspección visual de la calidad de frutas y hortalizas en línea; lo que, sumado a una clasificación automática por categorías de calidad mejoran la productividad y competitividad de las empresas (Larcher y otros, 2013).

En los laboratorios de investigación de alimentos, es frecuente el uso de técnicas de VC estacionarias que comprenden la adquisición de imágenes a color en condiciones estándar de iluminación y captura de imagen. La técnica permite mediciones rápidas y reproducibles de aspectos físicos, tales como: tamaño, forma, atributos de color y apariencia, entre otros parámetros, para el control de calidad de los alimentos (Mendoza y otros, 2016); así mismo, permite una valoración objetiva y no destructiva de la calidad de los productos, con un amplio futuro en la industria alimentaria (Sanchez y otros, 2004).

En las últimas décadas, las técnicas de VC y el análisis de imágenes han mostrado avances importantes en la evaluación experimental de la calidad de alimentos en laboratorios; sin embargo, a nivel industrial la

mayoría de las aplicaciones en línea para el control de calidad de alimentos frescos provenientes del campo o del mar es aún limitado; ello debido principalmente a la complejidad, especificidad y variabilidad biológica de cada alimento; y, además, al desconocimiento de índices de clasificación óptimos obtenidos desde una imagen del alimento y de modelos matemáticos o estadísticos eficientes de clasificación por calidad. Es así que es necesario más trabajos de investigación científica y de evaluación práctica con diversos alimentos, a fin de mejorar la aplicabilidad de la técnica para su adopción efectiva y sostenible en la industria alimentaria (Mahajan y otros, 2015).

La aplicación de esta tecnología podría también otorgar una ventaja competitiva a productores y comercializadores de pallares en el Perú, quienes disponiendo de un sistema de VC para la clasificación automática podrían definir estándares para la tipificación de parámetros objetivos (Venora y otros, 2009), acordes con las exigencias del mercado nacional e internacional, para facilitar y garantizar la comercialización y la calidad del grano.

En esta investigación se pretende implementar y configurar un sistema de visión por computadora para la captura de imágenes de color y su procesamiento, para la extracción de características únicas y el desarrollo de un modelo estadístico de clasificación automática en clases o categorías de calidad, mediante la evaluación simultánea del color, apariencia, forma y tamaño del pallar. La investigación considera aplicar la definición de los tipos de calidad del pallar mediante descriptores de imagen en la construcción de una Cartilla de identificación y clasificación visual, según su color, apariencia, tamaño y forma, y en la elaboración de una Carta de matices de blancura, según los tonos de blanco del grano, que permitan una descripción y clasificación objetiva, por categorías de calidad del pallar. Una consecuencia de este estudio será contribuir a la

mejora de las condiciones de intercambio en el mercado de acopio y transacciones de exportación de este grano alimenticio, mediante el uso de imágenes de color y sistemas inteligentes de manera que garantice, verifique y registre la calidad de las muestras de pallar.

Por tanto, esta investigación busca resolver el siguiente problema: ¿Un sistema de visión por computadora es capaz de clasificar y caracterizar muestras de pallar (*Phaseolus lunatus* L.) lqueño seco en categorías de calidad usando técnicas estadísticas discriminantes simples?

Los objetivos de la presente investigación, son:

- a. Desarrollar un modelo para la clasificación del pallar lqueño seco en categorías de calidad, mediante características superficiales extraídas por un sistema de visión por computadora, usando modelos estadísticos discriminantes.
- b. Evaluar mediante estadísticas descriptivas las características de color, apariencia, tamaño y forma, definidas por descriptores superficiales obtenidos por un sistema de visión por computadora, para la propuesta de una “cartilla de identificación y categorización” y una “carta de matices de blancos” del Pallar lqueño seco.

II. REVISIÓN DE BIBLIOGRAFÍA

2.1. El pallar

Es una leguminosa de origen sudamericano, de plantas erectas con una altura aproximada de 0.5 m. y rastreras con una longitud de 4 m; es de tallo vellosa, con pequeñas hojas y flores purpúreas dispuestas en espigas; sus frutos son vainas, de 5 a 15 cm, que contienen de 3 a 5 semillas aplanadas, sub-globosas, alargadas, de color blanco jaspeado o cremoso, con alto valor nutritivo por su contenido de proteínas, y de sabor agradable, cuando cocidas. El Perú, en particular la región de la costa, tiene condiciones agroecológicas favorables para la producción de pallar con buenos rendimientos agrícolas que permiten satisfacer el consumo nacional y disponer de una mayor oferta exportable (Vásquez, 1997). El Perú posee un gran potencial genético en este cultivo, pues es uno de sus centros de origen (Gutiérrez-Salgado y otros, 1995; mencionados por Matos y Zúñiga, 2002). El pallar de Ica, debido fundamentalmente a su menor contenido de ácido cianhídrico, comparado con el de otras zonas de producción, presenta un sabor agradable, textura suave y aspecto cremoso al cocerse, las que son algunas de sus cualidades diferenciadas, apreciadas por los consumidores, que lo prefieren frecuentemente en la alimentación familiar (Espinoza, 2005).

2.1.1. Origen

Las especies del género *Phaseolus* tienen su origen en América tropical (México, Guatemala y Perú), cuyo material genético y los hallazgos arqueológicos (Cuadro 1) existentes en la región, prueban su antigüedad. El pallar es una especie domesticada en época prehispánica que se cultivaba 7.500 A.C. en Ancash, y 5800 A.C. en la costa norte del Perú, donde se la ha encontrado en forma silvestre (Ospina, 1980).

Cuadro 1. Hallazgos arqueológicos de especies cultivadas del género *Phaseolus*.

Espece	Edad Arqueológica (años AC)	Lugar excavación
<i>P. vulgaris</i>	7000	México (Tehuacán)
<i>P. lunatus, tipo Big lima</i>	5300	Perú (Chilcas)
<i>P. lunatus, tipo Steve</i>	1400	México (Tehuacán)
<i>P. coccineus</i>	2200	México (Tehuacán)
<i>P. ecutifollus</i>	5000	México (Tehuacán)

Fuente: Ospina (1980).

2.1.2. Clasificación botánica según cultigrupo

Altamirano (2011) menciona que la clasificación botánica utilizando la noción del cultigrupo (cv-gr), es una subdivisión creada por el hombre para las distintas formas cultivadas de pallar, según sus caracteres, evolución, adaptación en las diferentes zonas ecológicas, usos y preferencias de cada región desde épocas primitivas, según los datos arqueológicos y cronológicos, pero que no se incluye en la clasificación taxonómica (Cuadro 2); sin embargo se en la NTP 205.019.2015 se caracteriza los cultigrupo-cv de pallares más importantes: Pallar grande (Big Lima), Pallar Sieva o Bebe y el Pallar tipo papa (potato) (Cuadro 3), los dos primeros irradian desde el hilium hasta el dorso de la semilla y el tipo papa tiene semillas redondas y sus estrías no son fácilmente observadas, el hilium es de forma ovoide y de germinación epigea.

Cuadro 2. Clasificación taxonómica del pallar

Subreino	Fanerógamas
División	Magnoliophyta (Angiospermas)
Subdivisión	Magnoliopsida (Dicotiledóneas)
Clase	Dicotiledóneas
Subclase	Rosidae
Orden	Leguminosales (Fabales)
Familia	Papilionaceae (Fabaceae o Leguminosae)
Subfamilia	Papilionoideae
Tribu	Phaseoleae
Subtribu	Phaseolus
Especie	<i>Phaseolus lunatus</i> L.

Fuente: Adaptado de Altamirano (2011)

Cuadro 3. Clasificación de pallar según cultigrupo

Cultigrupo	Big Lima	Sieva	Papa
Variedad	Iqueño precoz Sol de Ica Generoso de Ica Criollo	Baby UNALM1 Baby	Prog. Legum. UNALM
Forma de la semilla	Arriñonada Semi-arriñonada ligeramente aplanada	Arriñonada Semi-arriñonada aplanado	Redonda
Longitud del grano (mm)	18-32	10 a 16	8.7 – 14.0
Ancho del grano (mm)	12.5 – 16.0	8.5 – 10.5	7.5 – 9.5
Grosor del grano (mm)	4.5 – 7.5	4 – 5.5	6 – 7.5
Peso de 100 granos (g)	153 – 256	40 – 57	21 – 40
Nº granos en 100 gramos	40 – 65	225 – 234	280 – 281

Fuente: Adaptado de NTP. 205.019.2015. Indecopi (2015)

2.1.3. Clasificación de pallares comerciales en el Perú, tipo Big Lima

Los pallares comerciales clasificados tipo Big Lima como el Pallar Iqueño y Pallar Americano son descritos por Valladolid y Voysest (2006) las que abarcan diversas variedades, siendo los de mayor producción exportable los departamentos de la costa del Perú, tal como se muestra en el Cuadro 4.

El Pallar Iqueño (pallar tipo Big Lima, denominado comercialmente en inglés “extra large”) es originario del Perú, de ahí su nombre de “Lima bean”; entre ellos se tiene a la variedad precoz INIA y Sol de Ica, entre otros (Espinoza, 2005); se produce en los valles costeros del departamento de Ica, caracterizados por su ambiente agroecológico especial, de clima templado y de baja humedad relativa; estas variedades son de periodo vegetativo largo (mayor a 200 días) y tienen el grano más grande que cualquier otro pallar del mundo. Por su tamaño y calidad culinaria es muy demandado en el mercado japonés y europeo. De otro lado el Pallar Americano (pallar tipo Big Lima denominado comercialmente en inglés “extra large”), por su precocidad se adapta muy bien a las diferentes condiciones agroecológicas de la costa del Perú y se consume en grano seco, verde desgranado, procesado, congelado y en conserva.

Cuadro 4. Pallares comerciales tipo Big Lima en el Perú

Características del grano	Pallar Iqueño (Extra large Lima bean)	Pallar Americano (Large Lima bean)
Color	Blanco	Blanco
Forma	Arriñonada aplanada	Arriñonada aplanada
Tamaño	Extra grande 100 semillas pesan 190 - 270 g	Grande 100 semillas pesan 110-180 g
Calibre	40-52 semillas por 100 g 11-15 semillas por onza (28.35 g)	55-91 semillas por 100 g 16-26 por onza (28.35 g)
Principales Variedades	Perú: Generoso de Ica (I 1548-71) Judion, Criollo Iqueño precoz-Inia	Perú: Sol de Ica (I-450-3-71) LUN 1002, LUN 1003, UC-92 USA: WV-N, María, UC-92, Fordhood 242, Fordhood (138 g/100 semillas) UC-92, Fordhood 242
Países Productores	Perú	Perú, Estados Unidos, Madagascar
Zonas Producción de Perú	Ica Arequipa (Acarí) Lima	Ica Ancash (Casma) Lima (Huacho, Huaral, Huaura) Arequipa (Acarí) Piura, Lambayeque, La Libertad
Principales países importadores	España, Italia, Canadá Medio oriente, Japón Argentina	España, Italia, Reino Unido Medio oriente, Japón Argentina

Fuente: Adaptado de Valladolid y Voysest (2006)

2.1.4. Características sanitarias y de aspecto del pallar

La calidad comercial de los granos en el Perú, está regulada por las Normas Técnicas Peruanas (NTP); la norma vigente para el pallar de consumo humano es la NTP 205.019:2015 (Indecopi, 2015), la que es utilizada como referente de calidad en el comercio de granos a nivel nacional (Cuadro 5).

Cuadro 5. Tipos de defectos del pallar.

Tipos de pallares según defectos

Grano enfermo

Grano picado

Grano partido o quebrado

Grano arrugado y/o chupado

Grano abierto

Grano descascarado

Variedad Contraste

Fuente: Adaptado de NTP 205.019.2015. Indecopi (2015)

2.2. Visión por computadora (VC)

2.2.1. Conceptos básicos

2.2.1.1. Imagen digital

Una imagen digital es una representación numérica en 2D o 3D de un objeto físico, a partir del cual se puede obtener una representación geométrica y/o espectral (para el caso de imágenes hiperespectrales), con detalles suficientes (resolución de imagen) para su procesamiento, comprensión, almacenamiento, impresión y visualización. Una imagen digital puede ser de tipo vectorial o matricial (o mapa en bits) dependiendo de si la resolución es estática o dinámica (Mendoza y Lu, 2015).

La resolución de imagen se define como la cantidad de píxeles por unidad de dimensión (2D o 3D) utilizados en la captura de matices del objeto que se está digitalizando, siendo el píxel la menor unidad homogénea de color que forma parte de la imagen digital y pueden ser expresadas generalmente como ppi (píxeles por pulgada) (Mendoza y Lu, 2015).

2.2.1.2. Modelos de descripción numérica de espacios de color

Los modelos de color, describen numéricamente espacios de color los que pueden ser los sistemas RGB, $L^*a^*b^*$, HSV (Figura 1) y CMYK. El sistema RGB es un modelo de color basado en la síntesis aditiva que utiliza los matices de los tres colores básicos (rojo-R, verde-G y azul-B), los que se pueden mezclar para obtener todos los demás colores. La combinación de los tres colores produce la denominada escala de grises, que son combinaciones de negro y blanco de diferente intensidad. La escala de grises sigue una línea desde el negro (origen de las coordenadas) hasta el blanco. El modelo CMYK es un modelo de color sustractivo; el Cian (C), Magenta (M) y Amarillo (Y) es la base de colores y representan el complemento de los colores primarios RGB y el Negro (K) que es usado en las impresoras comerciales. El modelo de color HSV (del inglés Hue (H), Saturation (S) y Value (V), (Matiz, Saturación, Valor)) es un modelo orientado al artista, expresado en tres componentes que varían de 0 a 1. En la investigación de alimentos, el color se representa con frecuencia utilizando el sistema CIE- $L^*a^*b^*$ o simplemente $L^*a^*b^*$ (para diferenciar de otros sistemas con nombres similares como el AnLab (L_AaAb_A) y HunterLab (L,a,b) (Padron, 2009), donde L^* coincide la percepción humana de luminosidad, que va de 0 (negro) a 100 (blanco), y los parámetros de a^* (varia de verde de -120 a 0 y rojo de 0 a +120) y b^* (varia de azul de -120 a 0 y amarillo de 0 a +120) que son los dos componentes cromáticos (Mendoza y Lu, 2015).

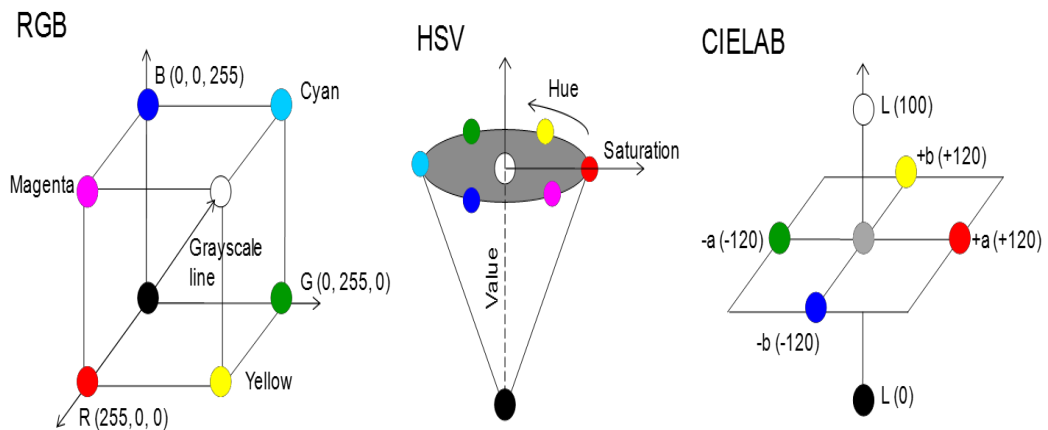


Figura 1. Descripción numérica de espacios de color para el análisis de imágenes

Fuente: Mendoza y Lu (2015)

2.2.1.3. Textura de imagen

La textura de imagen puede definirse como las propiedades estadísticas locales de un pixel en escala de grises, que son valoradas desde un punto de pixel en cuatro direcciones independientes vecinas de: 0° , 45° , 90° y 135° a una distancia (d) de pixeles vecinos, las que se pueden derivar de un histograma de imágenes propuesta por Harleick (1973) mencionado por Mendoza y Lu (2015), como la estadística de primer y segundo orden, las que se describen a continuación (Mendoza y Lu, 2015).

Energía, es una medida de la uniformidad local de los pixeles, es decir la repetición de pares de pixeles, cuyos valores son altos cuando la matriz de co-ocurrencia tiene pocas entradas de gran magnitud y son de valores bajos cuando todas las entradas son similares. Matemáticamente se define a continuación:

$$Energia = \sum_i \sum_j P_{d\theta}(i,j)^2$$

Homogeneidad, llamada también diferencia inversa de momento, es una medida local contraria al contraste en energía constante y se define matemáticamente con la ecuación:

$$\text{Homogeneidad} = \sum_i \sum_j \frac{P_{d\theta}(i,j)}{1 + (i-j)^2}$$

Correlación, es una medida de la linealidad de los pixeles de la imagen que se define matemáticamente con la expresión:

$$\text{Correlacion} = \sum_i \sum_j \frac{(i,j)P(i,j) - \mu_x \cdot \mu_y}{\sigma_x \cdot \sigma_y}$$

Contraste, es una medida de la variación local de la imagen, que se define matemáticamente de la siguiente manera:

$$\text{Contraste} = \sum_i \sum_j (i-j)^2 P_{d\theta}(i,j)$$

Donde $P_{d\theta}(i,j)$, es la probabilidad de ocurrencia de un par de niveles de grises reportado en la matriz de ocurrencia, a una distancia d y dirección θ , μ_x , μ_y son medias, y σ_x y σ_y son desviaciones estándar (Mendoza y otros, 2016).

2.2.1.4. Modelos de caracterización y clasificación de imágenes

La identificación de pallares de manera automática, usando visión por computadora puede subdividirse en las siguientes etapas: adquisición de imágenes, procesamiento digital (reconocimiento y segmentación de las imágenes, ejemplo granos de pallar), análisis de las secciones segmentadas (extracción de parámetros o características de calidad) y, finalmente, clasificación en Categorías de Calidad usando modelos estadísticos de variables múltiples. La Figura 2, grafica los pasos

mencionados en un proceso de clasificación o reconocimiento de patrones, usando un sistema de visión por computadora (SVC) (Mendoza y Lu, 2015; Leiva, 2013).

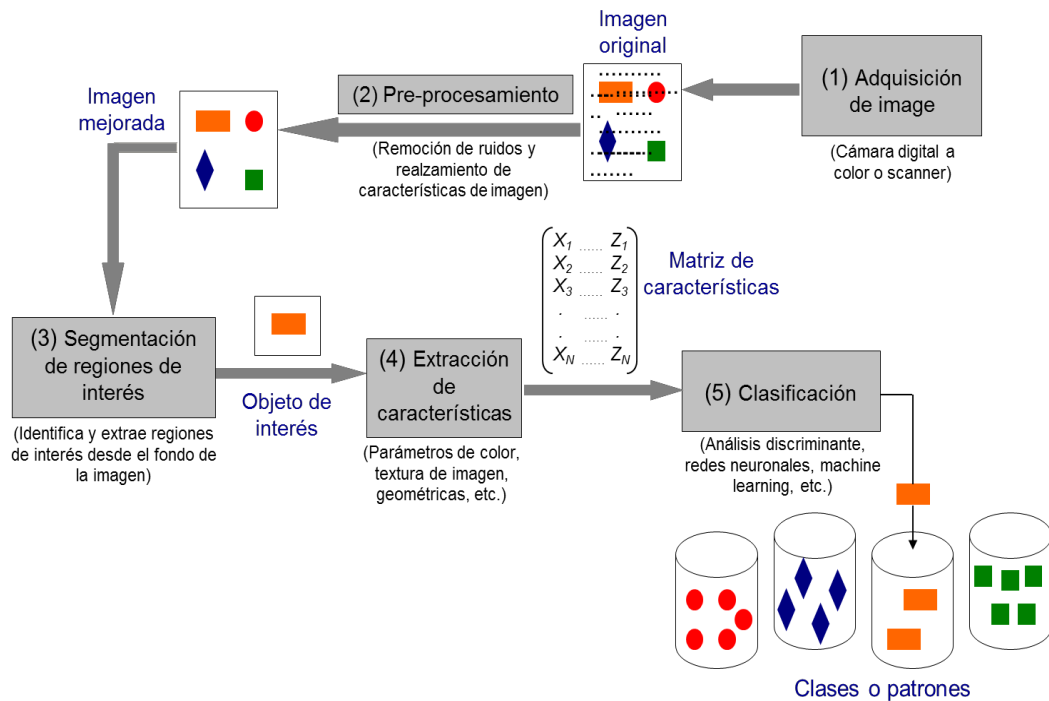


Figura 2. Proceso de reconocimiento y clasificación de patrones por Visión por Computadora.

Fuente: Modificado de Mendoza y Lu (2015)

Du y Su (2006), menciona que utilizando las técnicas de procesamiento de imágenes, las imágenes de los productos alimenticios se caracterizan cuantitativamente por un conjunto de características como tamaño, forma, color y textura (McDonald y otros, 2016), donde estas características son datos objetivos que se pueden utilizar para entrenar modelos de aprendizaje (Figura 3) y que puedan tomar decisiones de casos desconocidos mediante el uso de algoritmos de aprendizaje, decisiones como clasificar y predecir automáticamente (Vithu y Moses, 2016), emulando la decisión visual humana.

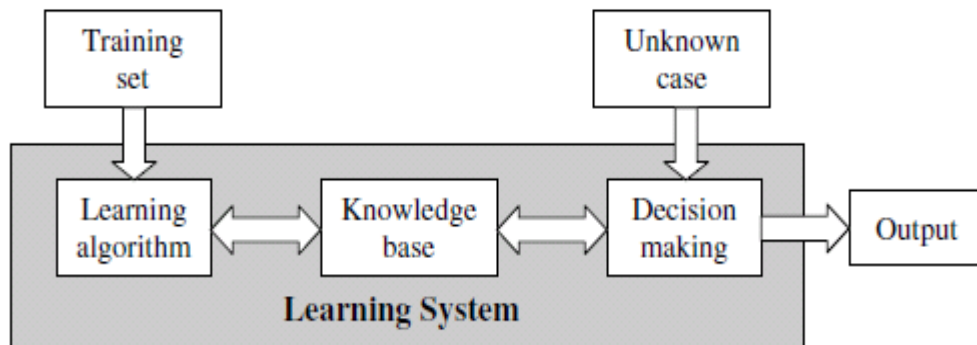


Figura 3. Configuración general de un modelo de clasificación por Visión por Computadora.

Fuente. Du y Su (2006)

2.2.2. Elementos básicos de un sistema de visión por computadora

Un SVC consta de tres elementos básicos: una fuente de luz blanca natural (o iluminantes tipo, luz día), una cámara digital a color (CDC) para la adquisición de imágenes, y un ordenador conectado a la CDC provisto de dos softwares: uno para el manejo de la CDC, visualización y adquisición de imágenes, y otro para el procesamiento y análisis de las imágenes almacenadas (Mendoza, 2005; Mendoza y otros, 2006; Padron, 2009).

a. Sistema de iluminación

Generalmente las muestras se iluminan utilizando dos o cuatro (2 a 2) lámparas paralelas (comúnmente tubos fluorescentes, tipo luz natural 18W/965) emulando preferentemente la luz del día, la cual corresponde a una temperatura de color de 6500 K (D_{65} , fuente de luz estándar comúnmente usada en investigación de alimentos), con un índice de

rendimiento de color (Ra) cerca del 95% (Vidal y otros, 2013; Mendoza, 2005).

Un aspecto crítico en visión por computadora, es la distribución homogénea de la intensidad de luz difusa sobre el fondo de captura. Por ello, las lámparas fluorescentes son preferentemente situadas por encima de la muestra y a una distancia apropiada de ésta (dependiendo del tamaño de la muestra), de forma tal, que forme un ángulo de 45° entre el eje de proyección de los iluminantes y la CDC (Francis y Clydisdale, 1975; Hutchings y otros, 2002; mencionados por Vidal y otros, 2013), por lo que se recomienda cubrir las lámparas fluorescentes con difusores plásticos de luz a fin de evitar las reflexiones de luz directa (es decir, luz especular en exceso); además del uso de estabilizadores de corriente para evitar el parpadeo de los tubos fluorescentes, asegurando así una iluminación constante y uniforme sobre la superficie de la muestra (Mendoza, 2005).

b. Cámara digital y programa de adquisición de imágenes.

Una cámara digital a color se localiza en posición vertical sobre el fondo del sistema y a una distancia que permita la adquisición de imágenes correctamente enfocadas y claras de los aspectos de interés de la muestra. El campo de visión ideal deberá cubrir un área donde la iluminación es constante y homogénea, y se eviten sombras y/o distorsiones de imagen debido a la curvatura o forma de la muestra (Francis y Clydisdale, 1975; mencionado por Vidal y otros, 2013; Mendoza, 2005; Mendoza y otros, 2006).

Se recomienda instalar el sistema de iluminación y la CDC dentro de una caja oscura, a fin de evitar las reflexiones de luz externa. Las condiciones de iluminación en la toma de la imagen son manejadas desde el ordenador mediante un software de asistencia remota, propio de la cámara digital (Mendoza, 2005; Mendoza y otros, 2006).

c. Ordenador para el procesamiento de imágenes.

Los algoritmos computacionales para el pre-procesamiento automático de las imágenes capturadas, incluyendo la segmentación del fondo y la extracción de parámetros de color de la región de interés, son desarrollados con las herramientas de procesamiento de imágenes disponibles en MATLAB (The MathWorks, Inc., USA) o usando otra plataforma de programación más sofisticada como C++, entre otras (Mendoza, 2005).

2.2.3. Visión por computadora aplicada a los alimentos

El sistema tenía dos ventajas importantes sobre los inspectores humanos. Eso produce alto rendimiento, reproducible y objetivo clasificación de muestras y también elimina las posibles clasificaciones erróneas de muestras que podrían ser realizadas por inspectores humanos (Kilic y otros, 2007).

En el procesamiento de alimentos, la aplicación de visión por computadora representa una herramienta clave en la inspección objetiva, efectiva, económica y en tiempo real, de la calidad de las frutas, hortalizas y carnes (Mahajan y otros, 2015), como rama de la inteligencia artificial para llevarlo a la Visión Industrial a través de señales de clasificación automática, donde la inspección humana es limitada (Alvarado, 2012). Sin embargo, los sistemas de visión por computadora presentan limitaciones al evaluar los alimentos, ya que proveen información superficial de la muestra; por tanto, no son capaces de medir características químicas y/o nutricionales (contenido de aceite, proteínas, minerales, vitaminas, etc.) y sensoriales, (sabor, aroma y textura), los cuales son atributos determinantes en la calidad de la mayoría de los alimentos (Mahajan y otros, 2015).

Mendoza y otros (2016), mediante visión por computadora mostraron que un modelo de regresión lineal múltiple ajustado por mínimos cuadrados parciales (PLSR, por sus siglas en inglés), puede predecir automáticamente el color y la apariencia de los frijoles negros (*Phaseolus vulgaris*) enlatados, en una escala del 1 al 5 con correlaciones para el color y apariencia de 0.937 y 0.871, y errores de Eficiencia de Clase de 0.26 y 0.38, respectivamente. Cuando se realizó una clasificación basada en características de color y apariencia, juntos, un modelo de Máquina de Vectores de Soporte (Support Vector Machine, SVM, por sus siglas en inglés) pudo clasificar las muestras de frijoles negros en dos categorías de calidad sensorial, "aceptable" e "inaceptable", con una Exactitud del 89.7%.

En frutas y hortalizas, las técnicas VC permiten mediante una sola imagen digital, estimar diversas características como: el perímetro, diámetro, área, color, textura de imagen, entre otros parámetros útiles en el reconocimiento de patrones y procesos de clasificación (Moreda y otros, 2009).

Igathinathane y otros (2009), mediante una aplicación de bajo costo midieron el tamaño de los granos de leguminosas utilizando dimensiones ortogonales a fin de clasificarlos por categorías simulando el análisis granulométrico estándar.

Valous y otros (2009) mediante visión por computadora clasificaron rebanadas de jamón cocido comercial basados en la distribución de grasa y carne magra presentes en la superficie del jamón con Eficiencia Generales de clasificación mayores al 90%, demostrando que el método de visión por computadora combinado con técnicas apropiadas de análisis de imagen, es una herramienta objetiva y eficiente para detectar y cuantificar propiedades de color y aspecto visuales de imágenes

complejas, tales como los componentes presentes en la superficie de un jamón comercial de color y textura no uniforme (por ejemplo: poros, grasa y carne magra). También se ha desarrollado la VC para seleccionar y clasificar el tamarindo por niveles de calidad relacionados a parámetros de color y tamaño, para su posterior empaque (Jarimopas y Jaisin, 2008).

Mendoza y Aguilera (2006) mostraron una técnica simple y confiable para la caracterización de las etapas de maduración en bananos (*Musa cavendish*). La técnica fue basada en la extracción de las características de color de la superficie del fruto entero en $L^*a^*b^*$, textura de imagen en niveles de grises y la cuantificación del desarrollo de las pintas café en la superficie del banano (porcentaje de área café y número de pintas por cm^2) debido a la degradación de la clorofila durante la maduración. Las características extraídas fueron comparadas con la clasificación visual realizada por expertos bananeros en la industria. Un modelo estadístico discriminante simple logró clasificar los bananos en siete etapas de maduración, con una Eficiencia General del 98%.

2.3. Análisis Discriminante

El análisis discriminante es una técnica estadística de múltiples variables, cuya finalidad es describir (si existen) las diferencias significativas entre g grupos de objetos ($g > 1$) sobre los que se observan p variables (variables discriminantes); más concretamente ayuda a distinguir entre dos o más grupos de datos basado en un conjunto de p variables cuantitativas observadas (StatPoint, 2006), de tal manera que las diferencias entre los grupos predefinidos se maximizan (Du y Su, 2006).

Amat (2016) describe al Análisis Discriminante Lineal o *Linear Discriminant Analysis* (LDA) como un método de clasificación supervisado de variables cualitativas en el que dos o más grupos son conocidos *a priori* y nuevas observaciones se clasifican en uno de ellos en función de sus características. Haciendo uso del teorema de Bayes, LDA estima la

probabilidad de que una observación, dado un determinado valor de los predictores, pertenezca a cada una de las clases de la variable cualitativa, $P(Y=k|X=x)$. Finalmente se asigna la observación a la clase k para la que la probabilidad predicha sea mayor. El clasificador cuadrático o Quadratic Discriminant Analysis-QDA se asemeja en gran medida al LDA, con la única diferencia de que el QDA considera que cada clase k tiene su propia matriz de covarianza (Σ_k) y, como consecuencia, la función discriminante toma forma cuadrática:

$$\text{Log}(P(Y = k|X = x)) = -\frac{1}{2}\text{Log}|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \text{Log}(\pi_k)$$

Para calcular la *probabilidad a posteriori* con esta ecuación discriminante es necesario estimar para cada clase, (Σ_k), μ_k y π_k a partir de la muestra. Cada nueva observación se clasifica en aquella clase para la que el valor de la *probabilidad a posteriori* sea mayor. QDA genera límites de decisión curvos por lo que puede aplicarse a situaciones en las que la separación entre grupos no es lineal.

Determinar que clasificador, LDA o QDA, es el más adecuado, depende de las implicaciones que tiene, en el balance sesgo-varianza, el asumir que todos los grupos comparten una matriz de covarianza común. LDA produce límites de decisión lineales, lo que se traduce en menor flexibilidad y, por lo tanto, menor problema de varianza. Sin embargo, si la separación de los grupos no es lineal, tendrá un sesgo grande. El método QDA produce límites cuadráticos y, por lo tanto, curvos lo que aporta mayor flexibilidad, permitiendo ajustarse mejor a los datos con menor sesgo, pero mayor riesgo de varianza. En términos generales, LDA tiende a conseguir mejores clasificaciones que QDA cuando hay pocas observaciones con las que entrenar al modelo, escenario en el que evitar la varianza es crucial. Por el contrario, si se dispone de una gran cantidad

de observaciones de entrenamiento o si no es asumible que existe una matriz de covarianza común entre clases, QDA es el más adecuado.

ángulo de 45° entre el eje del lente de la cámara y el eje de la fuente de iluminación), con carcasas compactas (Modelo BN068C LED6/NW L600 G2); luz blanca de 18W/965, con temperatura de color de 6500 K (D65) e índice de rendimiento de color (Ra) de cerca del 95%; con difusores de luz para cada carcasa (papel bond, color blanco de 75 g/m²) para evitar el brillo del pallar por el reflejo de la luz (Constante y otros, 2016).

- b. La cámara digital de color (CDC), modelo Canon PowerShot A520 (Canon, USA) (Figura 5), para la toma o captura de las imágenes a color, con software incorporado para su digitalización y transferencia de datos al computador personal. La cámara está montada en el interior de la cámara oscura, a 16.5 cm de altura en dirección vertical, sobre la base del portador de la muestra. La CDC está permanentemente conectada a una fuente eléctrica a través de un estabilizador de corriente 500 VA/250W, Marca Forza Power Technologies, Modelo NT-502U.
- c. Computador personal (Laptop Marca Lenovo, Modelo Intel Duo Core2, CPU T9400, 2.96 GB de RAM) (Figura 6), para la visualización, captura y almacenamiento automático de las imágenes y el procesamiento, análisis y clasificación de sus datos; cuenta con Software Adobe Photoshop CS6 (versión 13.0.1 x 64 bits) y Software MATLAB v7.5 (The MathWork, USA), con Image Processing Toolbox para el pre-procesamiento, análisis de imágenes y algoritmos para los modelos de clasificación; también tiene el Software ZoomBrowser Ex5.0 (Canon Inc., USA) para Windows XP que con la interfaz IFC-400PCU (suministrada con la cámara) permite realizar de manera remota (desde el computador) la toma y control de la captura directa de las imágenes por la CDC.

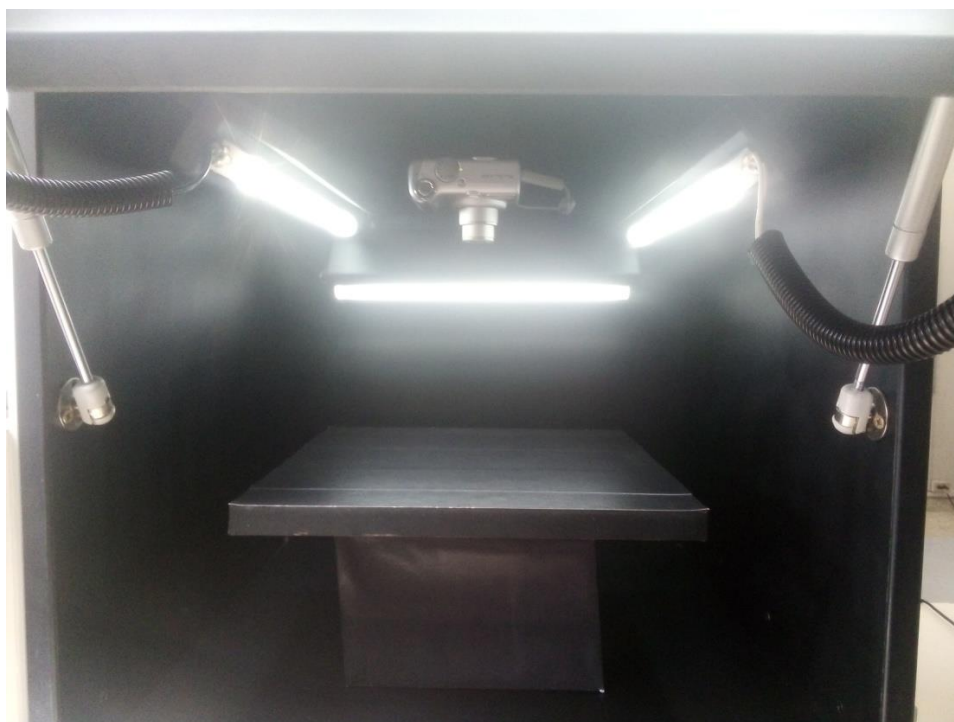


Figura 4. Cámara de iluminación para la VC

Fuente. Elaboración propia

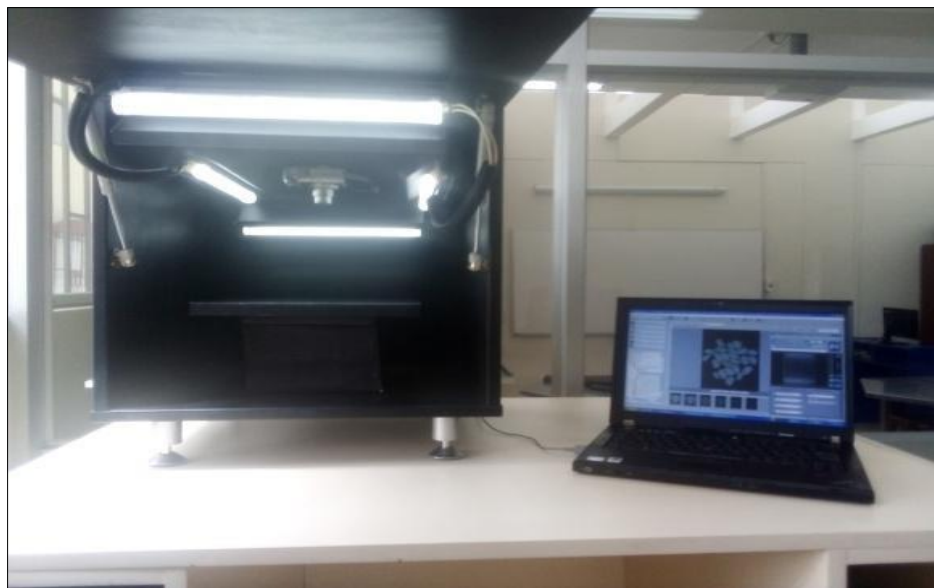


a. cámara digital de color (CDC)

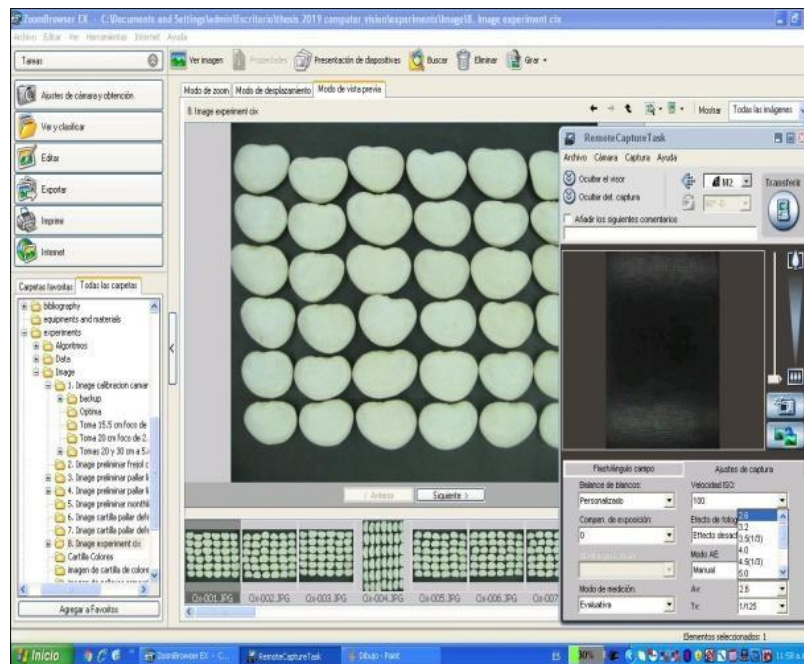


Figura 5. Cámara digital y estabilizador de corriente.

Fuente. Elaboración propia



a. Computador personal y cámara oscura.



- b. Software zoomBrowser Ex5.0 de Canon para la captura remota de imágenes.

Figura 6. Vista en computador de captura remota de imagen con software especializado.

Fuente: Elaboración propia

3.3. Metodología Experimental

La Figura 7 muestra gráficamente las etapas que comprendió la investigación: identificación y caracterización por defectos de la muestra del pallar por inspección visual humana; calibración del sistema de visión por computadora (SVC); captura de las imágenes digitales; optimización, procesamiento y análisis de las imágenes digitales; caracterización y clasificación del pallar mediante visión por computadora. La metodología de la investigación, en general, se orientó por los trabajos de visión por computadora en granos de leguminosas realizados por Kilic y otros (2007), Venora y otros (2009) y McDonald y otros (2016).

3.3.1. Identificación y caracterización de pallares según defectos de calidad mediante inspección visual

A la muestra de 19.17 kg de Pallar Iqueño seco, proveniente del Mercado de Mochoqueque y recepcionada en el Laboratorio de Ingeniería de Alimentos de la UPAO, se la sometió a una selección preliminar para eliminar materias extrañas y pallares de otras variedades medianos y pequeños, quedando 14.37 kg. Luego, se agruparon los granos por defectos presentes y se caracterizó la muestra, según su contenido en peso de granos con defectos predominantes. Se siguió el siguiente procedimiento:

1. Se seleccionó la materia extraña, granos enteros de tamaño menor a 18 mm (pallar mediano y pequeño), según el protocolo de la norma técnica peruana NTP.205.029.1982. Rev.2016 (Inacal, 2016), que detalla el procedimiento para el análisis físico de granos alimenticios, entre ellos las leguminosas, como el pallar.

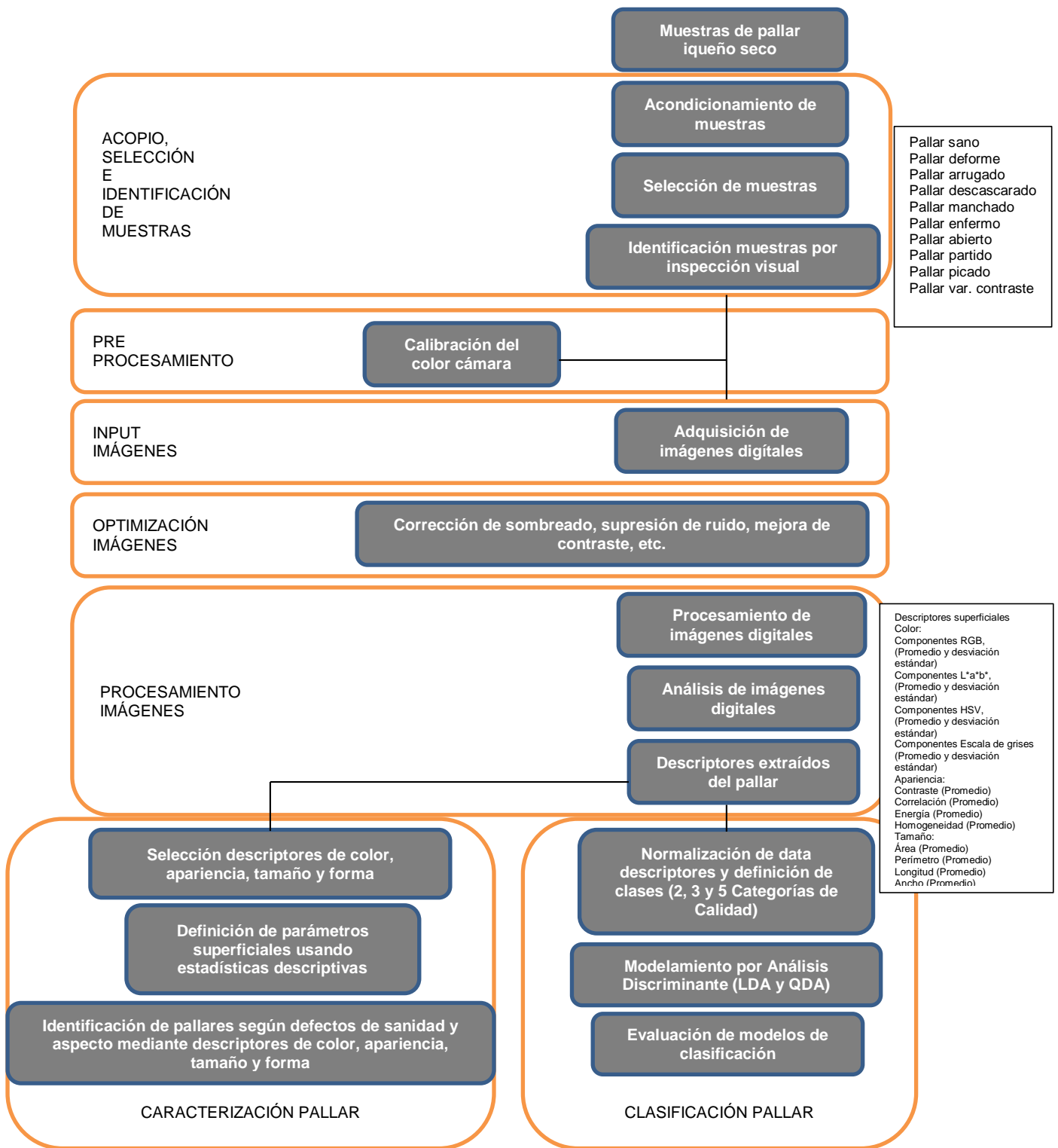


Figura 7. Procedimiento experimental para la investigación.

Fuente. Elaboración propia

2. A partir de los 14.37 kg del pallar, se inspeccionaron todos los granos manual y visualmente, uno por uno, y se agruparon en 10 tipos de pallares defectuosos incluyendo el pallar sano (pallares sin defectos). El defecto predominante en cada grano, se estableció de acuerdo a la especificación de calidad del pallar descrita en la norma técnica peruana NTP.205.019.2015 (Indecopi, 2015) y a la experiencia del candidato a ingeniero (ejecutor de esta tesis) en la industria de pallares en el Perú apoyado de la opinión especializada de un grupo de comercializadores
3. Se acondicionó la muestra experimental y se almacenaron los granos de pallar en bolsas de polietileno por grupos de defectos de calidad a las condiciones ambientes del laboratorio para la captura posterior de las imágenes digitales.

3.3.2. Calibración y manejo de la cámara digital

La calibración de la cámara, por tanto, la calibración del sistema de visión por computadora, se refiere a la configuración del sistema de imagen a un estado conocido de color para reproducir valores adecuados, consistentes y reproducibles (Mendoza, 2005). Esta calibración, que se utiliza con frecuencia para la caracterización de las sensibilidades espectrales de las cámaras fotográficas digitales y dispositivos de grabación, se realiza mediante una correlación de la sensibilidad del sensor de la cámara digital con respecto a un patrón, carta o paleta estándar de colores, conocido como “Colorchecker”, bajo condiciones de iluminación artificial (Westland y otros, 2012; Hong y otros, 2001).

La calibración de la cámara digital para esta investigación se realizó emulando la visión humana y siguiendo el procedimiento adaptado de Valous y otros (2009), detallado a continuación.

1. Con la cámara digital de color Canon A250, se fotografiaron los seis colores, entre blanco y negro, de la paleta de colores Colorchecker Clásica X-rite (Gretag- Macbeth, EE.UU.) (Figura 8).
2. A los valores RGB de las imágenes de los seis colores, entre blanco y negro del Colorchecker Clásica, se les aplicó el procesamiento y análisis de imágenes, y se obtuvieron valores de color de las imágenes segmentadas, expresados en el espacio de color $L^*a^*b^*$ y se correlacionaron los valores de L de la CDC con los del Colorchecker Clásica.

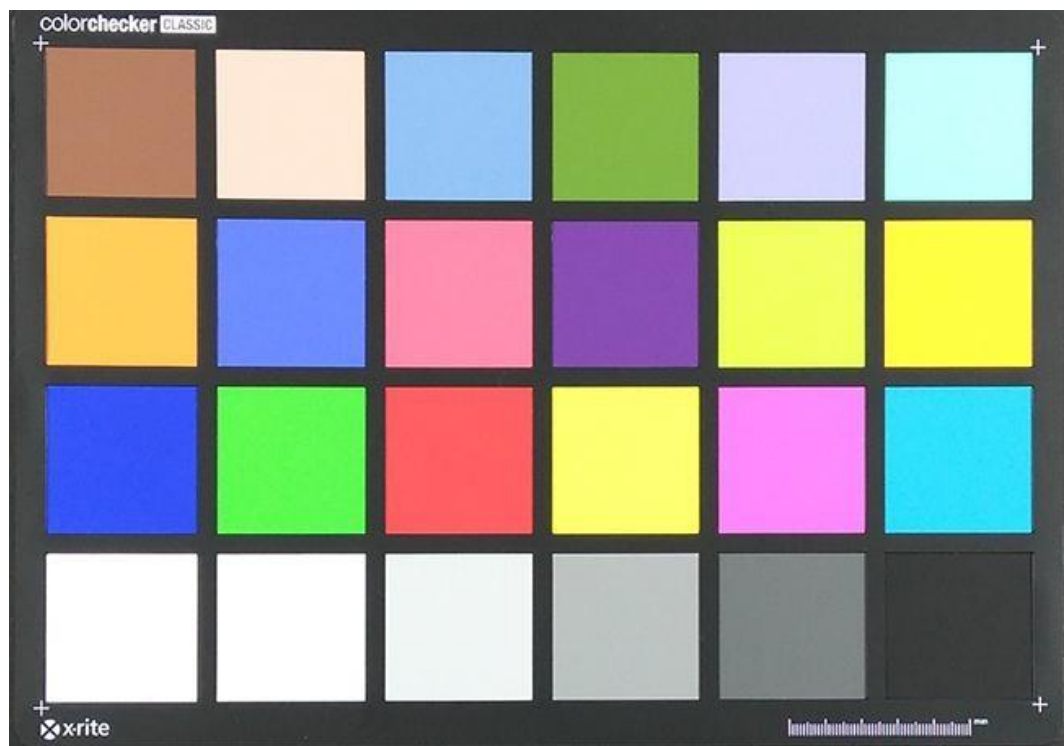


Figura 8. Colorchecker clásica X-Rite

Fuente. <https://www.xrite.com/categories/calibration-profiling/colorchecker-classic>

Los pasos para operar la cámara digital y adquirir las imágenes digitales del pallar para su posterior procesamiento y análisis, se detallan a continuación.

1. Se conectó el cable USB de la cámara al computador personal, y se revisó la estabilidad de las lámparas fluorescentes antes de iniciar la adquisición de las imágenes.
2. Se encendió la cámara en "Modo Manual" evitando cualquier vibración o movimiento de la cámara oscura instalada.
3. Se encendió el computador (PC) y se abrió el software Zoom Browser Ex5.0 (ícono ubicado en el escritorio de la PC), activando en el menú el ajuste de "cámara y obtención", y luego el de "disparo remoto".
4. Se configuraron los parámetros de captura de imagen como ISO, Obturación y Apertura de diafragma (Cuadro 7), para un Balance de Blancos personalizado (configurado antes, apuntando la cámara digital hacia una carta de color gris estándar (papel o tela blanca), con un 18% de reflectancia, de modo que ocupe el recuadro del centro de la pantalla LCD o todo el visor). Acto seguido se procedió a realizar el disparo y a grabar los parámetros, fijando el estado de la función (Canon A250, 2004), como condición estándar para cualquier captura de imagen, sin flash, en calidad extrafina y sin ningún efecto fotográfico, con lo que se aseguró el ajuste permanente de la sensibilidad del sensor de la cámara a las condiciones de iluminación de la cámara oscura.

Cuadro 6. Parámetros de captura de imagen
en la cámara digital

Altura:	16,5 cm
Resolución:	1024 x 768 pixeles
Compresión:	Calidad extrafina
Pto. enfoque:	Automático
Macro:	Apagado
Luz auxiliar AF:	Apagado
Operación AF:	Desbloqueado AF
Flash:	Apagado
Calidad flash:	Cero
Balances blancos:	Personalizado
Compensación exposición:	Cero
Modo medición:	Evaluativa
Efecto fotografía:	Desactivado
ISO:	100
Obturador:	1/100
Diafragma:	f2,6

Fuente: Elaboración propia

5. Se colocó la muestra en la base de la plataforma de la cámara oscura manteniendo la distancia entre cada uno de los granos, a fin de poder conseguir una buena segmentación y conteo del pallar por el algoritmo del procesamiento de imágenes a una distancia del lente de la cámara de 16.5 cm, similar a la 15 cm, usada por Kilic y otros (2007), a fin de tener una imagen equilibrada (con apoyo de la herramienta fotográfica “histograma”) para extraer la mayor cantidad de información de las muestras.
6. Acto seguido, se realizó el disparo de la cámara digital y la captura de la imagen de cada muestra, y este paso se repitió el número de veces que requirió el experimento para el total de las muestras.
7. La imagen adquirida se guardó en una carpeta asignada en la PC para su almacenamiento, en formato JPEG o JPG (este último formato

facilita un mejor manejo de la información y la reducción del tamaño de almacenamiento de las imágenes) y con una resolución 1024 x 760 pixeles (180 dpi), que permitió detectar los cambios de color y resolución independiente de la medida del color.

3.3.3. Procesamiento y análisis de imágenes

Las imágenes capturadas por la cámara digital fueron pre-procesadas para eliminar los efectos de sombra causados por un alto espesor del pallar, a fin de distinguir y segmentar del fondo (Venora y otros, 2009), que permita procesar, analizar y medir fácilmente el pallar de manera automática (Anexo 1), mediante el desarrollo de algoritmos en MATLAB (Kilic y otros, 2007), cuyo procedimiento se detalla a continuación.

1. La imagen digital a color RGB, fue transformada en imagen de escala de grises (también llamada imagen de intensidad), mediante el algoritmo `rgb2gray` de MATLAB y luego transformada en imagen binaria. Esta imagen binaria consiste en pixeles blancos y negros, donde “0” es el fondo negro y “1” es el pallar en blanco (Figura 9).

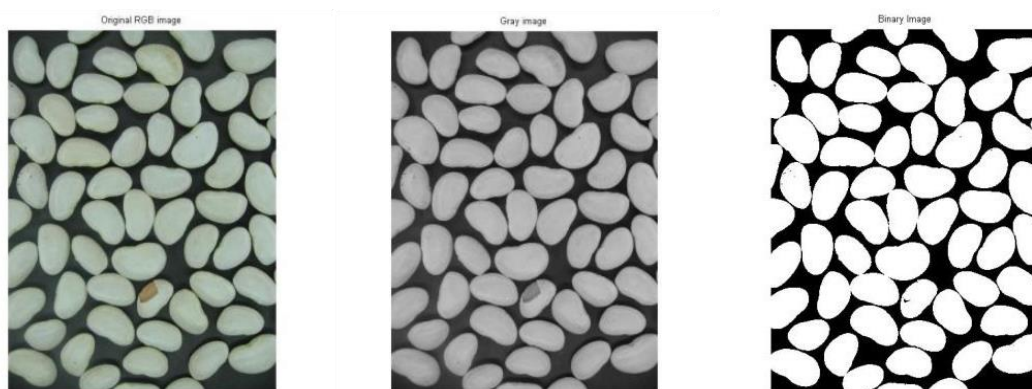


Figura 9. Transformación de imágenes RGB a imágenes binarias

Fuente. Elaboración propia

2. A las imágenes binarias de los pallares se las separó desde el fondo de estas y entre estas a fin de bordear las imágenes y realizar una adecuada segmentación individual del pallar. La segmentación es una operación en la cual se separa la imagen del fondo negro (denominado background, en inglés). Las imágenes bordeadas en blanco y negro, luego fueron transformadas en imágenes a color segmentadas, sin que se toquen los bordes (Figura 10).

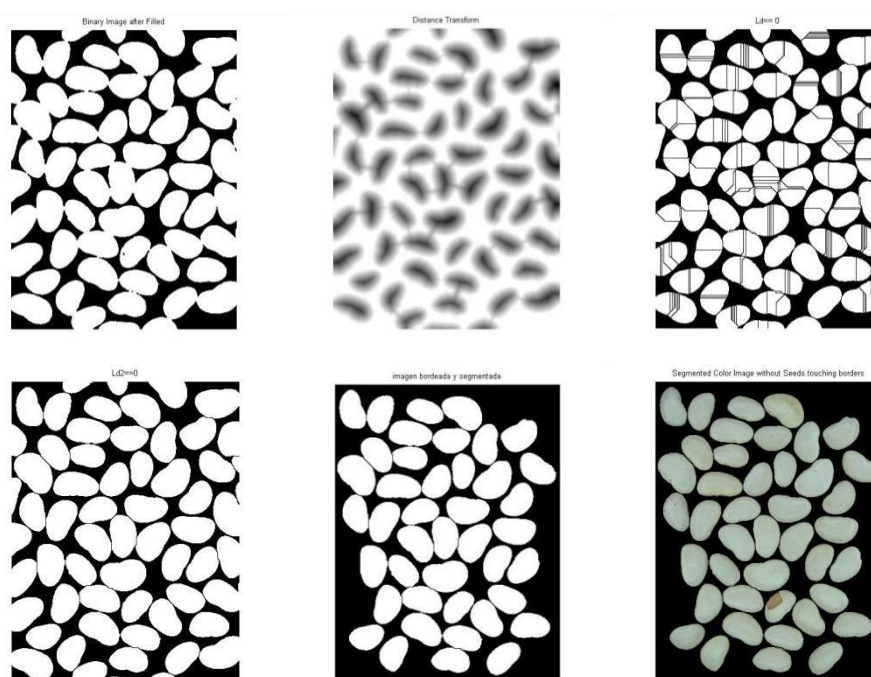


Figura 10. Proceso de segmentación de imágenes

Fuente. Elaboración propia

3. Las imágenes bordeadas y segmentadas a color, mediante la función “NumObjets” fueron localizadas y se contaron los pallares enteros, uno por uno, y se almacenaron para su posterior análisis (Figura 11).

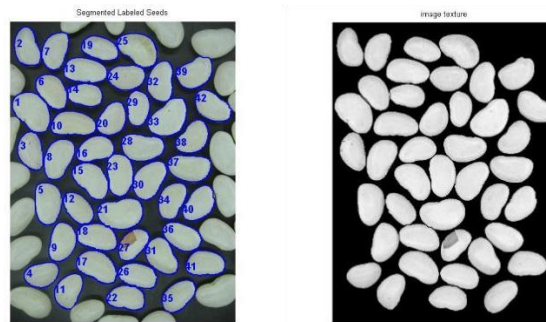


Figura 11. Conteo del número de pallares por cada imagen digital.

Fuente. Elaboración propia

- Las imágenes segmentadas se analizaron y se calcularon los descriptores de color, apariencia (textura de imagen), tamaño y forma usando funciones de análisis de imágenes de MATLAB (Mendoza, 2005, Mendoza y otros, 2006; 2016; Kilic y otros, 2007) (Anexo 2), mediante las siguientes funciones:

```
stats = regionprops(double(BW_unit), 'Area', 'Perimeter',
'MajorAxisLength',
'MinorAxisLength'); Área =
stats.Area;
Perimeter = stats.Perimeter;
MajorAxis =
stats.MajorAxisLength; MinorAxis
= stats.MinorAxisLength;
Roundness = 4*pi*Area/Perimeter^2; %Computetheroundness.metric
Calculando los valores de los pallares segmentados expresados en
píxeles a excepción de la redondez
```

`function [Gray, RGB, Lab, HSV] = colour2pallar(segRGB)`

Calculando los valores promedio y desviación estándar mediante la función “size” de Matlab

`function [Contrast, Correlation, Energy, Homogeneity] =`

`texture2pallar(gray, dist)ima = double(gray);`

Calculando los valores como el promedio de 4 direcciones independientes (0, 45, 90, 135°, con una distancia $d=1$), para garantizar mediciones de rotación de invariancia (Mendoza y otros, 2016).

5. Las 3169 imágenes segmentadas y analizadas con sus 29 descriptores superficiales de cada una de ellas fueron consolidadas en una matriz y almacenados para usos posteriores en la investigación (Anexo 3).
6. La selección de los pallares por defectos de calidad y la caracterización del pallar por defectos utilizaron datos de siete descriptores del color, apariencia (textura de imagen), tamaño y forma.
7. Los 29 descriptores fueron parámetros de entrada para los modelos de clasificación de pallares.

3.3.4. Identificación descriptiva de tipos de defectos del pallar mediante visión por computadora

La identificación de los 10 tipos de defectos, a partir de los datos de las 3169 imágenes de los pallares, permitió caracterizarlos estadísticamente, mediante descriptores superficiales de color, apariencia, tamaño y forma, tal como se detalla a continuación:

1. Los atributos de color, apariencia, tamaño y forma son importantes y de uso frecuente en la evaluación de la calidad de los granos alimenticios entre ellos el pallar. Para estos atributos se seleccionaron

7 descriptores superficiales de imagen para representar la calidad según defectos para los 10 tipos del pallar, los que se detallan seguidamente:

Color	:	Espacio de color L*a*b*	(3 parámetros)
Apariencia	:	Contraste y Homogeneidad	(2 parámetros)
Tamaño	:	Longitud	(1 parámetro)
Forma	:	Redondez	(1 parámetro)

2. Los 7 descriptores superficiales, seleccionados como parámetros de calidad del pallar, fueron evaluados mediante el paquete estadístico Statgraphics Centurion v17.2.07, determinando el valor promedio, máximos y mínimos de las medidas de color L*a*b*, contraste y homogeneidad, longitud y redondez de las imágenes de pallar analizadas.
3. También se evaluó mediante análisis de varianza (ANVA) y la prueba de comparaciones múltiples LSD de Fisher (LSD, Least Significant Difference, por sus siglas en inglés) con un 95% de confianza, las diferencias estadísticamente significativas de las imágenes de los granos de pallar de cada grupo de tipo de defecto, y entre los 10 tipos de defectos de calidad.
4. Finalmente, se consideró de utilidad práctica relacionar la información cuantitativa de los 10 tipos de defectos del pallar, con los datos de imágenes de fotografías a color de granos de pallar para diferenciar tres clases de calidad de granos: ÓPTIMO (pallar sin defectos), TOLERANTE y DESCARTE, para lo cual se ideó y construyó una “Cartilla de identificación y categorización del Pallar Iqueño mediante visión por computadora”, que puede ser más discriminante en el matiz del color blanco característico por medio de una “Carta de tonos de

blanco del Pallar Iqueño”, con imágenes a color y sus parámetros de calidad, obtenidos mediante el software Adobe PhotoShop CS6.

3.3.5. Clasificación en categorías de calidad del pallar mediante modelos discriminantes

La clasificación en categorías de calidad de las 3169 imágenes de pallar, agrupadas en 10 tipos de defectos de calidad, se realizó utilizando modelos discriminantes de clasificación (LDA y QDA, por sus siglas en ingles), por ser rápidos, simples, exactos y los más usados para clasificar cuando se conocen los objetos de los diferentes grupos (Abdullah y otros, 2001, mencionado por Mendoza y otros, 2006). Para la clasificación en 2,3 y 5 categorías de calidad, se desarrollaron algoritmos de MATLAB (Anexo 3).

Los 10 tipos de defectos, identificados y caracterizados cuantitativamente por sus respectivos 29 descriptores, se agruparon en 3 modelos de clasificación de calidad, cada uno de ellos para la clasificación en 2, 3 y 5 clases discriminantes, cuyas características se detallan seguidamente:

- A) Dos clases discriminantes, que son las categorías de calidad ACEPTABLE y la NO ACEPTABLE, para la comercialización del pallar.
- La Categoría de calidad ACEPTABLE comprende los tipos de defectos del pallar: Sano, Deforme, Arrugado, Descascarado y Manchado.
 - La Categoría de calidad NO ACEPTABLE comprende los tipos de defectos del pallar: Enfermo, Picado, Abierto, Partido y Variedad Contraste.

B) Tres clases discriminantes, que son las categorías de calidad ÓPTIMO, TOLERANTE y DESCARTE para la comercialización del pallar.

- La Categoría de calidad ÓPTIMO considera solo el pallar clasificado sin defectos, o sano.
- La categoría de calidad TOLERANTE considera los defectos de los pallares: Arrugado, Deforme, Descascarado y Manchado.
- La categoría de calidad DESCARTE considera los defectos de los pallares: Enfermo, Picado, Abierto, Partido y Variedad Contraste.

C) Cinco clases discriminantes, que son las categorías de calidad según la severidad de defectos: ÓPTIMO, TOLERANTE, DESCARTE FISIOLÓGICO, DESCARTE MORFOLÓGICO y DESCARTE VARIEDAD.

- La categoría de calidad ÓPTIMO considera solo el pallar clasificado como sin defectos, o Sano.
- La categoría de calidad TOLERANTE, abarca los defectos: Arrugado, Deforme, Descascarado y Manchado.
- La categoría DESCARTE FISIOLÓGICO comprende a los defectos Enfermo y Picado.
- Categoría DESCARTE MORFOLÓGICO se refiere a los defectos Abierto y Partido.
- Categoría DESCARTE VARIEDAD considera a variedades diferentes.

La clasificación en 2, 3 y 5 categorías de calidad, se realizó siguiendo el siguiente procedimiento:

1. Los 29 descriptores obtenidos de las imágenes de los pallares identificados por defectos de calidad, se agruparon en una matriz

individual de 2, 3 y 5 clases (Categorías de Calidad), respectivamente, para cada modelo de clasificación.

2. Se etiquetó y guardó, en formato xls, la matriz de descriptores, conforme el algoritmo de clasificación, que sigue:
Descascarado. Matriz de descriptores para la clasificación en 2 Categorías de Calidad.
Descascarado. Matriz de descriptores para la clasificación en 3 Categorías de Calidad.
Descascarado. Matriz de descriptores para la clasificación en 5 Categorías de Calidad.
3. Se abrió el software MATLAB e importó la matriz de datos etiquetados en formato xls para cada clasificación, y se guardó en una carpeta etiquetada en la PC (carpeta de algoritmos de clasificación), en formato mar (extensión de lenguaje de programación del MATLAB).
4. Ejecución del algoritmo de clasificación, mediante la función "Run" de MATLAB, para la clasificación en 2, 3 y 5 Categorías de Calidad, respectivamente, con la data previamente normalizada mediante la función "normData", para todos los modelos de clasificación.
5. Los algoritmos "DLA2linear" y "DLA2quadratic" de MATLAB se utilizaron para la clasificación lineal y cuadrática, respectivamente, siendo el "perc_testing" el algoritmo usado para determinar el porcentaje de muestras usadas para la calibración y la validación del modelo, respectivamente, y la función "cal2val" para estructurar las muestras del modelo.
6. Se obtuvo la matriz de confusión para cada modelo de clasificación (LDA y QDA, respectivamente) y el gráfico 3D de la clasificación LDA (Área, contraste, a^*).

7. Los resultados de los modelos de clasificación fueron presentados en una matriz de confusión, usando el algoritmo “confusionmat” de MATLAB, y de manera gráfica usando la función “scatter3”.

3.4. Métodos de análisis

3.4.1. Inspección visual humana de las muestras

La inspección visual se siguió mediante el protocolo adaptado de la NTP.205.029.1982 rev 2016 (Inacal, 2016), para el análisis físico de Cereales y Menestras, las mismas que se describen a continuación.

1. Acondicionamiento de muestras.

La muestra total recolectada, se selecciona por medio de un tamizado manual o mecánico, realizando un promedio de 30 movimientos circulares, eliminándose las partículas extrañas, granos pequeños y partidos, materia extraña, insectos, piedrillas, etc., y quedando los granos limpios y enteros en reposo en el tamiz (o tamices), para su posterior selección y clasificación. Finalmente, se determinó el porcentaje en masa de los granos enteros acondicionados en relación a la muestra acopiada total.

2. Selección y clasificación de la muestra.

La muestra acondicionada total, considerada como masa inicial, se dividió por medio de un divisor o por cuarteo manual y se obtuvo la muestra para análisis, la que debe tener como mínimo 500 g, y sobre la que se efectuaron los procedimientos de identificación de defectos del pallar (Figura 12), los que se realizaron hasta terminar el total del lote de la muestra, y que se detallan a continuación.

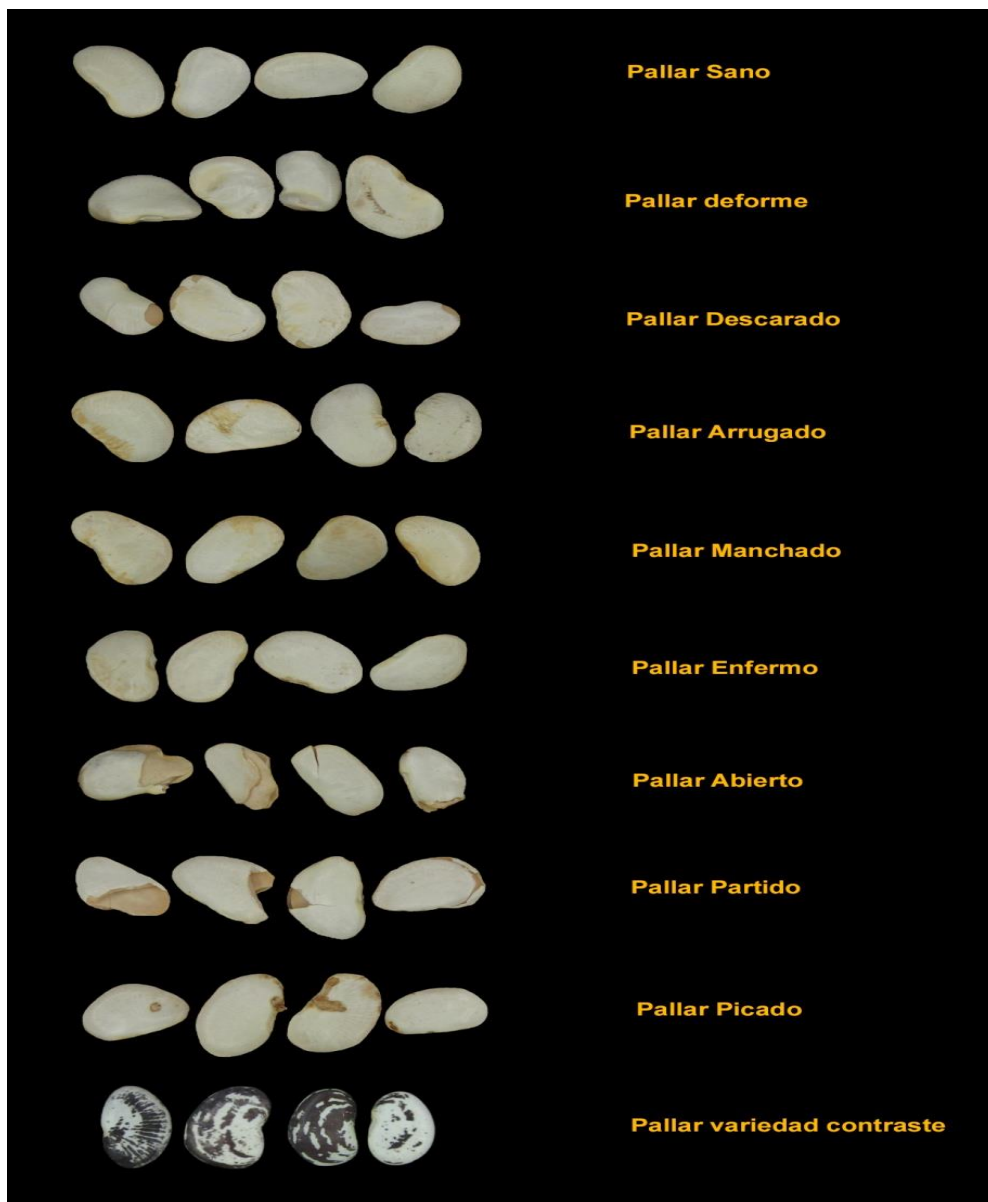


Figura 12. Identificación de defectos del Pallar Iqueño seco por inspección visual humana.

Fuente. Elaboración Propia

- a. Determinación de granos partidos. Los granos partidos se separaron en forma manual de la muestra acondicionada y limpia, quedando el saldo de los granos enteros en la bandeja para determinar su porcentaje de masa respecto al valor inicial.

- b. Determinación de las variedades contrastantes. Se realizó con la muestra remanente, después de realizados los análisis anteriores, se separaron en forma manual las variedades contrastantes (pallar de otras variedades). Luego se determinó el porcentaje con respecto a la masa inicial.

- c. Determinación de granos chupados o deshidratados, deformes, manchados, descascarados, infestados, abiertos y picados. Se realizó con la ayuda de una lupa y se procedió a clasificar en forma manual, quedando como remanente el pallar sano comercial, entero y óptimo. Luego se determinaron los porcentajes de cada defecto, con respecto a la masa inicial de la muestra empleada para el análisis, y finalmente, se sumaron estos porcentajes parciales para obtener el porcentaje total de granos dañados.

- d. Determinación de la calidad del grano. Se efectuó tomando en cuenta el valor del componente de granos defectuosos presente en la muestra inicial, cuyo valor corresponde a la mayor tolerancia de la Tabla de requisitos de la NTP 205.019:2015, y después de haber realizado todos los análisis anteriores, cuya descripción desarrollada para esta investigación se describe a continuación:

La Categoría ÓPTIMO, de máxima calidad, corresponde al Pallar Sano comercializable, entero, de color blanco característico, sin ningún daño o defecto en la integridad de su superficie.

La Categoría TOLERANTE, que según la NTP 205.019:2015 (Indecopi, 2015) corresponde a aquellos pallares que tienen algún tipo de defecto o daño de menor grado, que pueden ser considerados, con ciertos límites máximos, en una muestra comercial. Comprenden a los granos manchados, deformes, arrugados y descascarados.

La Categoría DESCARTE, son aquellos pallares que, por ninguna razón pueden estar presentes en una muestra comercial, ya que son pallares que tienen daños en su integridad superficial, defectos de color, de tamaño y forma, producto de un deficiente manejo fitosanitario, de postcosecha y transporte, y comprenden los granos enfermos, abiertos, picados, partidos, y con presencia de otras variedades.

La muestra acopiada y seleccionada proveniente de la ciudad de Chiclayo, Lambayeque, Perú, donde se acopia de todo el país, se considera representativa y suficiente para realizar la clasificación visual en Categorías de Calidad obteniendo un lote seleccionado y acondicionado de 14.37 kg de pallar por clasificar.

3.4.2. Evaluación de descriptores superficiales del pallar mediante visión por computadora.

La medida del color, apariencia, tamaño y forma, se realizó mediante la extracción de descriptores superficiales de las imágenes del pallar, procesadas y analizadas mediante algoritmos en MATLAB, realizado mediante los siguientes pasos:

1. Se colocó 100 g de muestra en la plataforma de objeto de disparo, a 16.5 cm de la cámara.
2. Se realizó la captura remota de la imagen de las muestras de pallar y se guardó la data en una carpeta etiquetada de la PC.
3. Con el software MATLAB, se importó el algoritmo de procesamiento y análisis de imagen (Anexo x), a la carpeta etiquetada de la PC.
4. Se ejecutó el algoritmo de procesamiento de imágenes de MATLAB, mediante la función F9, para la medida de color (media y desviación estándar de espacios de color RGB, L*a*b*, HSV), apariencia (media de textura de imagen: Contraste, Correlación, Energía y Homogeneidad),

tamaño (métrica de área, perímetro, longitud, ancho, expresado en pixeles) (Anexo 3 y 4) y forma (métrica adimensional de redondez) (Anexo 1) de cada unidad de pallar, de la muestra fotografiada, mediante los siguientes algoritmos:

COLOR= [COLOR; i k Gray RGB Lab HSV].

TEXTURE=[TEXTURE;Contrast,Correlation,Energy,Homogeneity]

MORPHO=[MORPHO;Area,Perimeter,MajorAxis,MinorAxis,Round
ness].

3.5. Análisis estadístico

La evaluación estadística del experimento se realizó utilizando el software Statgraphics Centurion v17.2.07, versión de prueba (StatPoint technologies, Inc., USA.), para la caracterización del pallar, y MATLAB v7.5 (The MathWorks, Inc., Natick, MA, USA.), para la discriminación de los modelos de clasificación.

El método estadístico de múltiples variables aplicado, comprendió las pruebas del análisis de varianza (ANVA) y la de la Mínima Diferencia Significativa (LSD, por sus siglas en inglés) de Fisher, para determinar las diferencias significativas de los atributos de calidad (color, apariencia, tamaño y forma) entre los tipos de pallar y cuáles son estas; con un nivel de confianza del 95%.

El análisis discriminante, lineal y cuadrático, de las categorías de calidad, en 2, 3 y 5 grupos, se realizó mediante validación cruzada con 10 repeticiones, con muestras escogidas al azar, con un 50% de muestra usada para el entrenamiento y 50% de muestra usada para la predicción, y representadas en una matriz de confusión, en base a 3169 imágenes de pallar con 29 descriptores: promedio, desviación estándar de escala de grises, RGB, $L^*a^*b^*$, HSV, Valor adimensional de Contraste, Correlación, Homogeneidad y Energía, Valor en pixeles de Área, Perímetro, Largo,

Ancho y transformada en mm mediante el factor experimental de 14.2 px/mm y un valor adimensional de redondez.

IV. RESULTADOS Y DISCUSIÓN

4.1. Identificación de pallares según defectos de calidad mediante inspección visual humana

Como resultado de la aplicación de la inspección visual humana, según protocolo adaptado de la NTP 205.029:1982 Rev. 2016 (Inacal, 2016) y las especificaciones de sus atributos sanitarios y de aspecto, de acuerdo a la NTP 205.019:2015 (Indecopi, 2015), el Cuadro 7 detalla la distribución porcentual (masa) de la muestra experimental del Pallar Iqueño seco diferenciados en 10 grupos o tipos de acuerdo al defecto predominante del grano, dentro de los cuales se ha incluido al grupo de pallar sano y sin defectos. La presencia de estos defectos determina la calidad o grado de calidad comercial del pallar cuyos estándares comerciales están definidos por la oferta y la demanda (Leiva, 2013), donde el pallar sano es considerado un pallar comercial sin defectos.

Cabe resaltar que la identificación de los defectos del pallar, así como su evaluación presenta limitaciones debido a la operatividad manual, que requiere personal experto o semicalificado y que dificulta el proceso de una evaluación (Kannur y otros, 2011, mencionado por Mahajan y otros, 2015) rápida, debido a la variabilidad de las características visuales de naturaleza multivariantes (color, apariencia, tamaño y forma) causada por la diversidad de contrastes, variedades y condiciones ambientales (Brosnan y Sun, 2004), y al mismo tiempo no existe estándares cuantitativos de color, apariencia, tamaño y forma en la industria del Pallar Iqueño seco.

Cuadro 7. Distribución del pallar muestra según defectos de calidad

Tipo Pallas según atributos sanitarios y aspecto	Criterios sanitarios y de aspecto para seleccionar pallares	Peso (g)	Distribución (%)
Pallar Sano	Comercial	1128.2	7.9%
Pallar Manchado	Defecto color	3089.9	21.5%
Pallar Deforme	Defecto color y forma	127.5	0.9%
Pallar Arrugado	Defecto forma	835.1	5.8%
Pallar Descascarado	Daño físico	439.9	3.1%
Pallar Enfermo	Daño fisiológico	607.8	4.2%
Pallas Abierto	Daño fisiológico	1979.4	13.8%
Pallar Partido	Daño morfológico	3090.6	21.5%
Pallar Picado	Daño morfológico	1365.4	9.5%
Pallar Var. Contraste	Otra variedad de pallar	1687.1	11.8%
Total		14350.9	100%

Fuente: Elaboración propia

4.2. Calibración de cámara digital

En la Figura 13, se muestra la caracterización de la Luminosidad del color de la cámara digital respecto a los valores grises del Colorchecker Clásica con una correlación lineal con coeficiente de regresión de 0.9966 y un valor inicial de 15.023 para el negro absoluto en el espacio de color CIE L*a*b*.

La sensibilidad relativa al valor negro absoluto de la cámara digital se debe a diversos factores, como las características propias del sensor de la cámara, la temperatura de calor de la fuente de iluminación (Mendoza y otros, 2006, Leiva, 2013, Valous y otros, 2009) y/o el color del difusor del iluminante empleado (papel blanco), que tiene un color distinto al color blanco del monitor de la computadora (Padron, 2009).

La alta correlación encontrada, demuestra la predictibilidad consistente y objetiva de las lecturas de color de la cámara digital en el espectro gris y, por tanto, en los espacios de color, superando las limitaciones propias del sistema (Valous y otros, 2009).

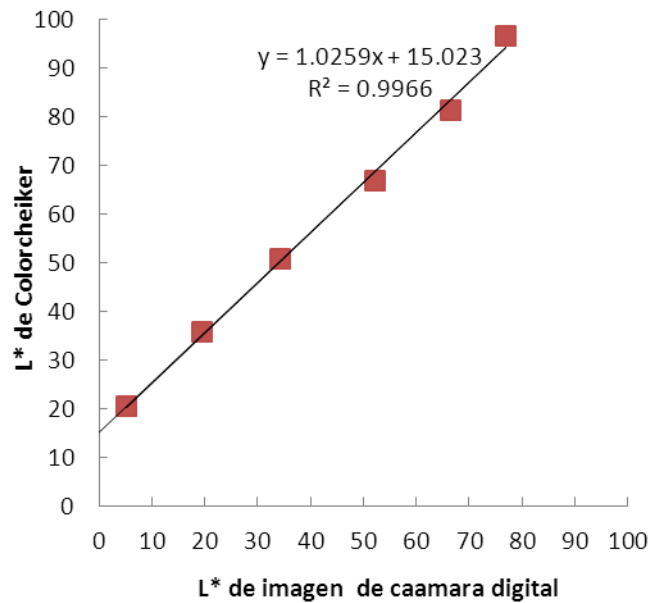


Figura 13. Correlación L* de la cámara digital vs Colorchecker

Fuente: Elaboración propia

4.3. Descriptores y clases discriminantes de clasificación por visión por computadora

La aplicación del sistema de visión por computadora implementado ha permitido recopilar los valores de los 29 descriptores por cada imagen, medidos mediante el análisis de imágenes (Majumdar y Jayas, 2000 y Javadikia y otros, 2017) de 3169 pallares segmentados, distribuidos en 10 tipos de pallares, identificados por sus defectos de calidad. Estos valores, se organizaron y almacenaron en una matriz de doble entrada que constituye la base de datos experimentales de la investigación.

Los 29 descriptores son los siguientes:

- 20 descriptores de color. Con valor promedio y desviación estándar de componentes de RGB, L*a*b*, RGB y escala grises.

- 4 descriptores de apariencia. Con valor adimensional de parámetros de textura de imagen: contraste, correlación, energía y homogeneidad.
- 4 descriptores de tamaño. Con valor de longitud (largo) medido en píxeles.
- 1 descriptor de forma. Con valor adimensional.

Estos descriptores, para cada uno de los 10 tipos de pallares con defectos de calidad se muestran en el Cuadro 8 en el que se nota que, de los 3169 granos de la muestra total se han identificado 346 como pallares sanos, 392 manchados, 390 arrugados, 34 deformes, 423 descascarados, 43 enfermos, 389 picados, 409 abiertos, 383 partidos y 360 de variedad contraste.

Cuadro 8. Descriptores extraídos del análisis de imagen del pallar

Pallar caracterizado por inspección visual		Sano	Manchado	Arrugado	Deforme	Descascarado	Enfermo	Picado	Abierto	Partido	Var Contraste	Total	
Nº muestras segmentadas		346	392	390	34	423	43	389	409	383	360	3169	
Descriptores de imágenes del pallar iqueño seco	Color	Gris (Media y St)	2	2	2	2	2	2	2	2	2	20	
		R,G,B (Media y st)	6	6	6	6	6	6	6	6	6	60	
		L*a*b* (media y st)	6	6	6	6	6	6	6	6	6	60	
		H,S,V (media y st)	6	6	6	6	6	6	6	6	6	60	
	Apariencia (textura de imagen)	Contraste	1	1	1	1	1	1	1	1	1	1	10
		Correlación	1	1	1	1	1	1	1	1	1	1	10
		Homogeneidad	1	1	1	1	1	1	1	1	1	1	10
		Energía	1	1	1	1	1	1	1	1	1	1	10
	Tamaño	Área	1	1	1	1	1	1	1	1	1	1	10
		Perímetro	1	1	1	1	1	1	1	1	1	1	10
Largo		1	1	1	1	1	1	1	1	1	1	10	
Ancho		1	1	1	1	1	1	1	1	1	1	10	
Forma	Redondez	1	1	1	1	1	1	1	1	1	1	10	
Nº total descriptores		29	29	29	29	29	29	29	29	29	29	290	

Fuente: Elaboración propia

Los datos recopilados de los 29 descriptores por cada uno de los 10 tipos de defectos de pallares fueron agrupados en 3 clases discriminantes de clasificación, a las que se les definió como de 2 (ACEPTABLE y NO ACEPTABLE), 3 (OPTIMO, TOLERANTE y DESCARTE) y 5 (OPTIMO, TOLERANTE, DESCARTE DEFECTOS FISIOLÓGICOS, DESCARTE DEFECTOS MORFOLÓGICOS y VARIEDAD CONTRASTE) categorías de calidad (Cuadro 9) según su nivel de defectos, contemplados en la NTP 205.019.2015 (Indecopi, 2015).

Cuadro 9. Definición de clases discriminantes de clasificación

Clases discriminantes de clasificación del pallar					
Tipos de pallares según defectos	Criterios de sanitarios y de aspecto en la caracterización visual	Nº muestras segmentadas	2 clases discriminante	3 clases discriminante	5 clases discriminante
Sano	Comercial	346		Categoría Óptimo (346 muestras)	Categoría Óptima (346 muestras)
Manchado	Defecto color	392	Categoría aceptable (1585 muestras)	Categoría Tolerante (1239 muestras)	Categoría Tolerante (1239 muestras)
Arrugado	Defecto color y forma	390			
Deforme	Defecto forma	34			
Descascarado	Daño físico	423			
Enfermo	Daño fisiológico	43	Categoría no aceptable (1584 muestras)	Categoría Descarte (1584 muestras)	Categoría por defectos fisiológicos (432 muestras)
Picado	Daño fisiológico	389			
Abierto	Daño morfológico	409			
Partido	Daño morfológico	383			
Var Contraste	Otra variedad de pallar	360			Categorías por defecto var. de pallar (360 muestras)

Fuente: Elaboración propia

Las clases discriminantes definidas fueron evaluadas en 2 modelos discriminantes de clasificación: Análisis Discriminante Lineal (LDA, por sus siglas en inglés) y los modelos de Análisis Discriminante Cuadrático (QDA, por sus siglas en inglés), de manera similar al criterio empleado con modelos discriminantes para clasificar: frejoles italianos por Venora y otros (2009) y Lo Bianco y otros (2015); arveja (*Pisum sativum* L.) con defectos y sin defectos; y en diferentes grados de calidad comercial por McDonald y otros (2016); frutas cítricas por Iqbal y otros (2016); naranjas por Javadikia y otros (2017) y en la detección de semillas de arroz enfermas y descoloridas por Baek y otros (2019).

4.4. Caracterización descriptiva del pallar mediante visión por computadora.

4.4.1. Medidas de parámetros de calidad del pallar

Diaz y otros (2004) consideran que la variabilidad de los descriptores extraídos presentan una complejidad y especificidad de los datos que dificulta la discriminación de los muestras y, en particular, la identificación cuantitativa debido a la superposición de la distribución y reducidos rangos de valores de los descriptores superficiales de las imágenes, por lo que es necesario reducir el número de descriptores a los más significativos mostrando valores promedio, máximos y mínimos, y la distribución de las muestras a través de gráficas Box-plot. La Figura 14 muestra los gráficos Box-plot para color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez, para las 3169 imágenes de pallar, agrupadas en los 10 tipos de defectos de calidad.

El Cuadro 10 muestra los valores de los descriptores superficiales de los atributos de calidad, que corresponden a los componentes de espacio de color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez, de los 10 tipos de pallares defectuosos (pallar sano, manchado, arrugado,

deforme, descascarado, enfermo, picado, abierto, partido y variedad contraste). Estos valores de color, contraste, homogeneidad, longitud y redondez, describen de manera específica, objetiva, cuantitativa y reproducible a los 10 tipos de defectos de calidad, los que se han obtenido mediante el análisis de imágenes digitales, de manera similar al que registró Verona y otros (2009), al caracterizar frejoles italianos con una resolución de 100 dpi.

En las gráficas Box-plot, el pallar sano se distingue en su distribución y promedio del resto de pallares, debido a las diferencias de los componentes de color $L^*a^*b^*$, contraste, longitud y redondez; así mismo, los valores de los pallares considerados como TOLERANTE (pallar manchado, deforme, arrugado y descascarado, respectivamente) se distinguen por los componentes de color a^* y b^* , longitud y redondez, de los pallares DESCARTE (pallar enfermo, abierto, partido, picado y de variedad contraste), haciendo la consideración del pallar “variedad contraste” con una mayor diferencia significativa, respecto en los espacios de color del pallar “sano”.

Comparando los resultados obtenidos en esta investigación con los de Venora y otros (2009) referidos a frijoles italianos de variedades diversas, se encuentran valores similares en el color superficial, que en el Pallar Iqueño sano (muestra de valor comercial de color) son de valores $L^*a^*b^*$ de 80, -8.9, 8.5, respectivamente ($R=186$, $G=203$, $B=182$, para RGB) frente a los frijoles italianos blancos como el frejol Bianco di Pigna (181.69, 183.55, 157.26, para valores de RGB), Cannellino di San Cinese (191.09, 193.95, 167.07 para valores de RGB), Cannellino di Pisa (198.69, 199.56, 176.65 para valores de RGB), Piatella Pisana (203.18, 207.03, 182.13 para valores de RGB), Purgatorio Lazio (188.69, 182.66, 159.86 para valores de RGB), Purgatorio Umbria (186.59, 177.06, 151.87 para valores de RGB), pero con diferencias significativas de tamaño, ya que

estas pertenecen a la variedad *Phaseolus vulgaris* L, con longitudes menores a 16 mm, en comparación con el pallar (*Phaseolus lunatus* L.) lqueño seco sano con longitud promedio mayores a 23 mm; la longitud es un factor importante en operaciones de calibrado, tal como lo demostró Kilic y otros (2007), encontrando correlación lineal de $r= 0.984$ entre las medidas experimentales en milímetros y las medidas de calibrado teórico en pixeles; Majumdar y Jayas (2000), consideran que la característica morfológica más significativa es la textura y el color al clasificar trigo CWRS (Canadá Western Red Spring, por sus siglas en inglés), trigo CWAD (Canadá Occidental Amber Durum, por sus siglas en inglés), cebada, avena y centeno.

Cuadro 10. Medida de descriptores superficiales de tipos de defectos del pallar

Descriptores de color, apariencia (textura imagen), tamaño y forma		Sano	Manchado	Arrugado	Deforme	Descascarado	Enfermo	Picado	Abierto	Partido	Variedad contraste
Recuento		346	392	390	34	423	43	389	409	383	360
L*	Promedio	80.0	77.0	77.5	78.2	78.3	78.4	76.3	77.2	76.5	63.0
	Mínimo	75.2	66.9	69.5	73.1	72.0	68.7	64.8	70.2	68.6	24.3
	Máximo	82.4	80.7	80.8	81.0	81.5	80.6	80.7	81.7	81.6	77.4
a*	Promedio	-8.9	-7.6	-7.1	-6.3	-7.4	-7.1	-7.1	-7.4	-7.1	-3.7
	Mínimo	-10.8	-9.1	-8.4	-7.1	-8.6	-8.2	-9.1	-8.8	-9.2	-6.1
	Máximo	-7.6	-5.9	-5.8	-5.6	-6.4	-6.3	-4.6	-6.0	-4.6	8.3
b*	Promedio	8.5	8.2	7.4	6.9	6.7	6.1	7.6	7.1	6.7	3.7
	Mínimo	5.1	2.6	3.1	4.1	2.8	3.5	2.8	2.2	2.4	-1.4
	Máximo	12.5	22.0	20.7	11.5	13.3	9.2	20.4	19.5	13.7	10.1
Contraste	Promedio	41.04E-3	38.49E-3	36.60E-3	36.53E-3	35.76E-3	35.30E-3	37.63E-3	38.04E-3	38.68E-3	36.91E-3
	Mínimo	34.27E-3	27.02E-3	26.14E-3	29.14E-3	22.64E-3	29.89E-3	28.27E-3	27.14E-3	24.64E-3	27.64E-3
	Máximo	47.53E-3	48.15E-3	44.78E-3	42.65E-3	44.15E-3	41.52E-3	51.53E-3	51.65E-3	46.90E-3	48.53E-3
Homogeneidad	Promedio	999.27E-3	999.31E-3	999.35E-3	999.35E-3	999.36E-3	999.37E-3	999.33E-3	999.32E-3	999.31E-3	999.34E-3
	Mínimo	999.15E-3	999.14E-3	999.20E-3	999.24E-3	999.21E-3	999.26E-3	999.08E-3	999.08E-3	999.16E-3	999.13E-3
	Máximo	999.39E-3	999.52E-3	999.53E-3	999.48E-3	999.60E-3	999.47E-3	999.50E-3	999.52E-3	999.56E-3	999.51E-3
Longitud (píxeles)	Promedio (pxl)	81.5	75.6	71.5	70.9	69.7	69.4	73.1	73.6	74.9	72.5
	Mínimo (pxl)	65.6	51.1	50.9	50.3	37.7	58.1	51.0	43.8	42.4	52.3
	Máximo (pxl)	94.5	94.8	89.9	84.6	88.6	82.4	89.1	94.0	92.5	100.6
Longitud (mm)	Promedio (mm)	23.0	21.3	20.2	20.0	19.7	19.6	20.6	20.8	21.1	20.5
	Mínimo (mm)	18.5	14.4	14.4	14.2	10.7	16.4	14.4	12.4	12.0	14.7
	Máximo (mm)	26.7	26.8	25.4	23.9	25.0	23.3	25.2	26.5	26.1	28.4
Área (píxeles ²)	Promedio (pxl ²)	3263	2865	2610	2568	2474	2452	2678	2739	2801	2648
	Mínimo (pxl ²)	2289	1392	1323	1601	1038	1761	1468	1223	1166	1501
	Máximo (pxl ²)	4323	4316	3777	3545	3770	3457	3854	4000	3955	4295
Área (mm ²)	Promedio (mm ²)	260.0	228.2	207.9	204.6	197.1	195.4	213.4	218.2	223.1	210.9
	Mínimo (mm ²)	182.4	110.9	105.4	127.6	82.7	140.3	117.0	97.4	92.9	119.6
	Máximo (mm ²)	344.4	343.9	300.9	282.4	300.4	275.4	307.0	318.7	315.1	342.2
Forma (adimensional)	Promedio	923E-3	920E-3	927E-3	916E-3	921E-3	935E-3	904E-3	903E-3	894E-3	923E-3
	Mínimo	872E-3	817E-3	847E-3	722E-3	799E-3	886E-3	700E-3	708E-3	709E-3	844E-3
	Máximo	966E-3	977E-3	990E-3	972E-3	993E-3	977E-3	970E-3	977E-3	963E-3	995E-3

Fuente. Elaboración propia

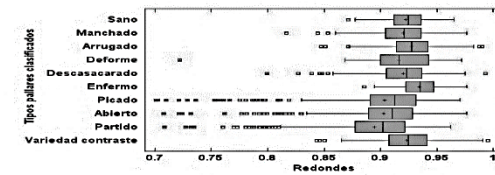
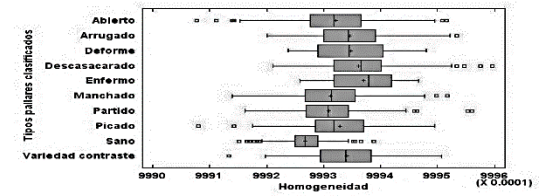
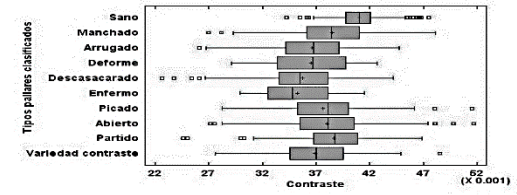
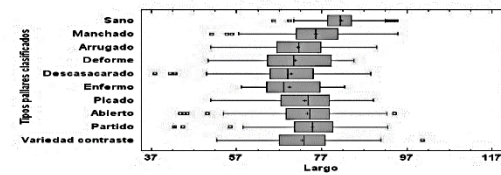
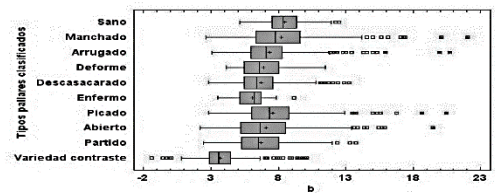
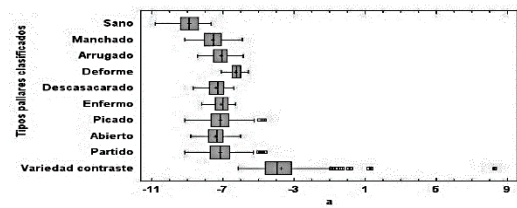
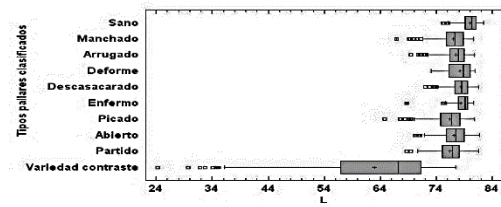


Figura 14. Box-plot de parámetros de color $L^*a^*b^*$, contraste, homogeneidad, largo y redondez de los defectos del Pallar Iqueño seco.

Fuente. Elaboración propia

4.4.2. Análisis estadísticos de parámetros de calidad del pallar

El análisis de varianza (ANVA), que se presenta en el Cuadro 11, muestra resultados específicos, identificando diferencias significativas entre los valores promedio de cada parámetro de calidad, medido por descriptores de color $L^*a^*b^*$, contraste, homogeneidad, largo y redondez, para cada tipo de defecto del Pallar lqueño seco.

Así mismo, la prueba LSD de Fisher muestra diferencias mínimas, estadísticamente significativas, entre los pallares defectuosos, evaluados en cada atributo de calidad, resaltando una diferencia significativa del pallar “sano” del resto de tipo de pallares defectuosos, pero con algunas semejanzas en $L^* a^* b^*$, contraste, homogeneidad, longitud y redondez entre alguno de los pallares considerados TOLERANTE y el DESCARTE (Cuadro 12).

Cuadro 11. ANVA de parámetros de color, apariencia, tamaño y forma de los tipos de defectos del Pallar lqueño seco.

Fuente		Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-p
L*	Entre grupos	70709.8	9	7856.64	452.26	0
	Intra grupos	54878.3	3159	17.3721		
	Total (Corr.)	125588	3168			
a*	Entre grupos	5398.06	9	599.785	1031.56	0
	Intra grupos	1836.75	3159	0.581434		
	Total (Corr.)	7234.81	3168			
b*	Entre grupos	5558.94	9	617.659	131.77	0
	Intra grupos	14807.8	3159	4.68748		
	Total (Corr.)	20366.7	3168			
Contraste	Entre grupos	0.00709008	9	0.00078779	65.91	0
	Intra grupos	0.0377603	3159	1.1953E-05		
	Total (Corr.)	0.0448504	3168			
Homogeneidad	Entre grupos	2.26087E-06	9	2.51E-07	65.91	0
	Intra grupos	1.20409E-05	3159	3.81E-09		
	Total (Corr.)	1.43018E-05	3168			
Largo	Entre grupos	33190	9	3687.77	69.27	0
	Intra grupos	168174	3159	53.2363		
	Total (Corr.)	201363	3168			
Redondes	Entre grupos	0.415946	9	0.0462162	46.05	0
	Intra grupos	3.1707	3159	0.0010037		
	Total (Corr.)	3.58665	3168			

Fuente. Elaboración propia

Cuadro 12. LSD de Fisher de parámetros de color, apariencia, tamaño y forma de los tipos de defectos del Pallar Iqueño seco

Tipos de defectos del pallar		Casos	Media	Grupos Homogéneos
L*	Variedad contraste	360	63.0164	X
	Picado	389	76.3316	X
	Partido	383	76.4648	XX
	Manchado	392	77.0308	XX
	Abierto	409	77.1553	XX
	Arrugado	390	77.481	XX
	Deforme	34	78.1787	XXX
	Descascarado	423	78.3054	X
	Enfermo	43	78.4153	XX
	Sano	346	79.9745	X
a*	Sano	346	-8.91714	X
	Manchado	392	-7.55645	X
	Abierto	409	-7.40167	X
	Descascarado	423	-7.35661	X
	Picado	389	-7.14214	X
	Partido	383	-7.13779	X
	Arrugado	390	-7.08699	X
	Enfermo	43	-7.08195	X
	Deforme	34	-6.2877	X
	Variedad contraste	360	-3.72732	X
b*	Variedad contraste	360	3.69604	X
	Enfermo	43	6.08026	X
	Descascarado	423	6.67548	X
	Partido	383	6.70423	X
	Deforme	34	6.8697	XXX
	Abierto	409	7.06596	X
	Arrugado	390	7.35379	XX
	Picado	389	7.59249	X
	Manchado	392	8.1927	X
	Sano	346	8.49443	X
Contraste	Enfermo	43	0.0353	X
	Descascarado	423	0.03576	X
	Deforme	34	0.03653	XXX
	Arrugado	390	0.0366	X

Continúa...

			Continuación
	Variedad contraste	360	0.03691	X
	Picado	389	0.03763	XX
	Abierto	409	0.03804	XX
	Manchado	392	0.03849	XX
	Partido	383	0.03868	X
	Sano	346	0.04104	X
Homogeneidad	Sano	346	0.99927	X
	Partido	383	0.99931	X
	Manchado	392	0.99931	XX
	Abierto	409	0.99932	XX
	Picado	389	0.99933	XX
	Variedad contraste	360	0.99934	X
	Arrugado	390	0.99935	X
	Deforme	34	0.99935	XXX
	Descascarado	423	0.99936	X
	Enfermo	43	0.99937	X
Longitud	Enfermo	43	69.3711	XX
	Descascarado	423	69.7457	X
	Deforme	34	70.8959	XXXX
	Arrugado	390	71.4935	XX
	Variedad contraste	360	72.4802	XX
	Picado	389	73.1099	XX
	Abierto	409	73.581	X
	Partido	383	74.8602	X
	Manchado	392	75.6036	X
	Sano	346	81.4964	X
Redondez	Partido	383	0.89431	X
	Abierto	409	0.90281	X
	Picado	389	0.90377	X
	Deforme	34	0.91645	XX
	Manchado	392	0.91992	X
	Descascarado	423	0.92053	X
	Sano	346	0.92278	X
	Variedad contraste	360	0.92288	X
	Arrugado	390	0.92748	XX
	Enfermo	43	0.93463	X

Fuente. Elaboración propia

4.4.3. Identificación descriptiva de defectos del pallar mediante color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez.

Demostradas las diferencias significativas entre los tipos de pallares defectuosos y las pocas similitudes con el pallar sano, se procedió a desarrollar una “Cartilla de identificación y caracterización del Pallar Iqueño”, con ayuda del programa Adobe Photoshop CS6, y definir rangos de color a través de una “Carta de matices de blanco del Pallar Iqueño”, a partir de las estadísticas descriptivas evaluadas con el programa Statgraphics Centurion v17.2.07, como se detalla a continuación.

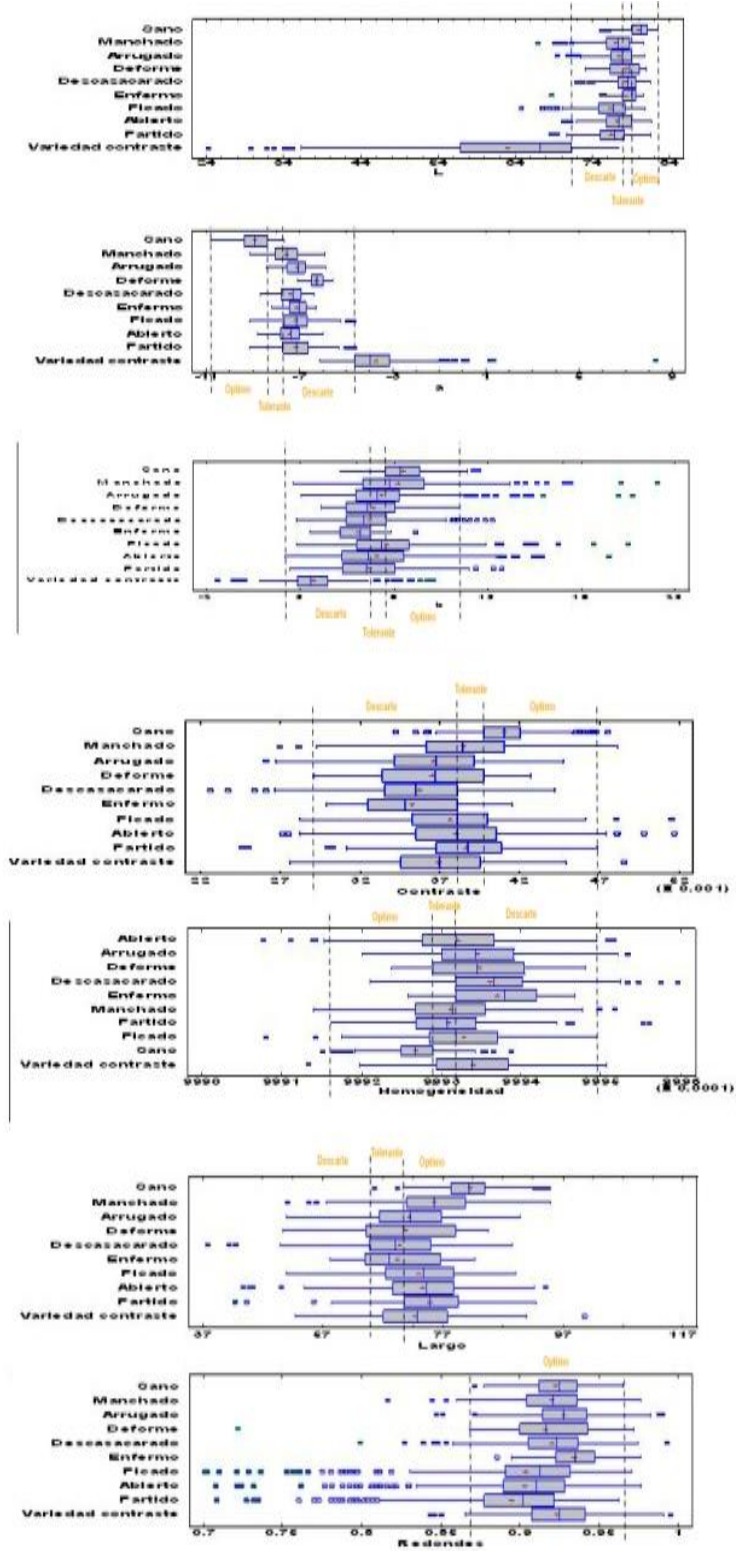


Figura 15. Límites en la gráfica Box-plot del color $L^*a^*b^*$, contraste, homogeneidad, longitud y redondez

Fuente. Elaboración propia

Utilizando la gráfica Box-plot de los tipos de pallares defectuosos, se establecieron rangos, con límites mínimos y máximos para los valores de los descriptores de cada defecto de calidad. Teniendo en cuenta la comercialización de granos y los grados de calidad referentes, se consideró de importancia poner al alcance de los interesados, información basada en los rangos de defectos de calidad medibles por visión por computadora, agrupados en tres clases, similares a como se presentan los grados de calidad de las normas técnicas, y el contenido del Cuadro 9: pallares sanos (ÓPTIMO), pallares manchados, deformes, arrugados y descarados (TOLERANTE) y pallares picados, abiertos, partidos y enfermos (DESCARTE) y su presentación en cartillas que podrían tener un uso práctico en función a la diferenciación del pallar sano respecto al resto de grupos de pallares, a excepción del pallar variedad contraste por ser de otra variedad, considerando al pallar sano como el pallar sin defectos, óptimo para la categorización. La Figura 15 muestra los valores de los rangos característicos de los descriptores que definen cuantitativamente el color, apariencia, tamaño y forma, mediante visión por computadora (Cuadro 13) de los pallares ÓPTIMO, TOLERANTE y DESCARTE, que se muestran en la “Cartilla de identificación y caracterización del pallar lqueño seco por visión por computadora” (Figura 16) y estableciendo el color, de manera particular, mediante una “Carta de matices de blancura para el Pallar lqueño seco” (Figura 17), sin considerar los descriptores del pallar variedad contraste, ya que no pertenecen a la variedad estudiada.

Cuadro 13. Valores de descriptores de calidad del pallar iqueño seco

Descriptores superficiales extraídos por visión por computadora	Descriptores de calidad del pallar iqueño seco								
	Óptimo			Tolerante			Descarte		
L*	(82.4	a	79.1)	(79.1	a	78.0)	(78.0	a	71.2)
a*	(-10.8	a	-8.4)	(-8.4	a	-7.8)	(-7.8	a	-4.6)
b*	(11.5	a	7.5)	(7.5	a	6.7)	(6.7	a	2.2)
Contraste	(46.90E-3	a	39.77E-3)	(39.77E-3	a	38.15E-3)	(38.15E-3	a	29.14E-3)
Homogeneidad	(999.16E-3	a	999.29E-3)	(999.29E-3	a	999.32E-3)	(999.32E-3	a	999.50E-3)
Largo (mm)	(>19.9)	(19.9	a	18.5)	(<18.5)
Redondez	(966E-3	a	722E-3)						

Fuente. Elaboración propia

CARTILLA DE IDENTIFICACIÓN Y CARACTERIZACIÓN DEL PALLAR IQUEÑO SECO POR VISIÓN POR COMPUTADORA



Pallar Sano

Superficie arrifonda íntegra, con estruvas radiales desde el hilum, blanco marfil característico

CATEGORÍA ÓPTIMA

L*	(79.1	82.4)
a*	(-8.4	-10.8)
b*	(7.5	11.5)
Contraste	(39.77E-3	46.90E-3)
Homogeneidad	(999.29E-3	999.16E-3)
Longitud	>19.9 mm	
Redondes	(777E-3	966E-3)



Pallar Manchado

Cambio de coloración con tonalidades amarillentas en su área superficial, 1% tolerancia max en muestra comercial



Pallar Deforme

Deformación forma arrifonda con blanco marfil característico, 1% tolerancia max en muestra comercial



Pallar Arrugado

Superficie deshidratada en mas 50% de su superficie, 1% tolerancia max en muestra comercial



Pallar Descascarado

Parcialmente sin cotiledon hasta 25% de su área superficial, 1% tolerancia max en muestra comercial

CATEGORÍA TOLERANTE

L*	(78.0	79.1)
a*	(-7.8	-8.4)
b*	(6.7	7.5)
Contraste	(38.15E-3	39.77E-3)
Homogeneidad	(999.32E-3	999.29E-3)
Longitud	(18.5 mm 19.9mm)	
Redondes	sin tolerancia	



Pallar Enfermo

Coloración en la parte superficial del hilum por infestación de microorganismos, 1% tolerancia max en muestra comercial



Pallar Abierto

Cotiledones separados parcialmente por daños fisiológicos y/o mecánicos, 1% tolerancia max en muestra comercial



Pallar Picado

Cotiledones dañados por ataques de insectos u otros microorganismos, 1% tolerancia max en muestra comercial



Pallar Partido

Grano entero sin el 10% a mas de su tamaño original, 1% tolerancia max en muestra comercial

CATEGORÍA DESCARTE

L*	(71.2	78.0)
a*	(-4.6	-7.8)
b*	(2.2	6.7)
Contraste	(29.14E-3	38.15E-3)
Homogeneidad	(999.50E-3	999.32E-3)
Longitud	<18.5 mm	
Redondes	Valores fuera del rango optimo	

* Imagen RGB

Figura 16. Cartilla de identificación y caracterización del Pallar Iqueño seco

Fuente. Elaboración propia.

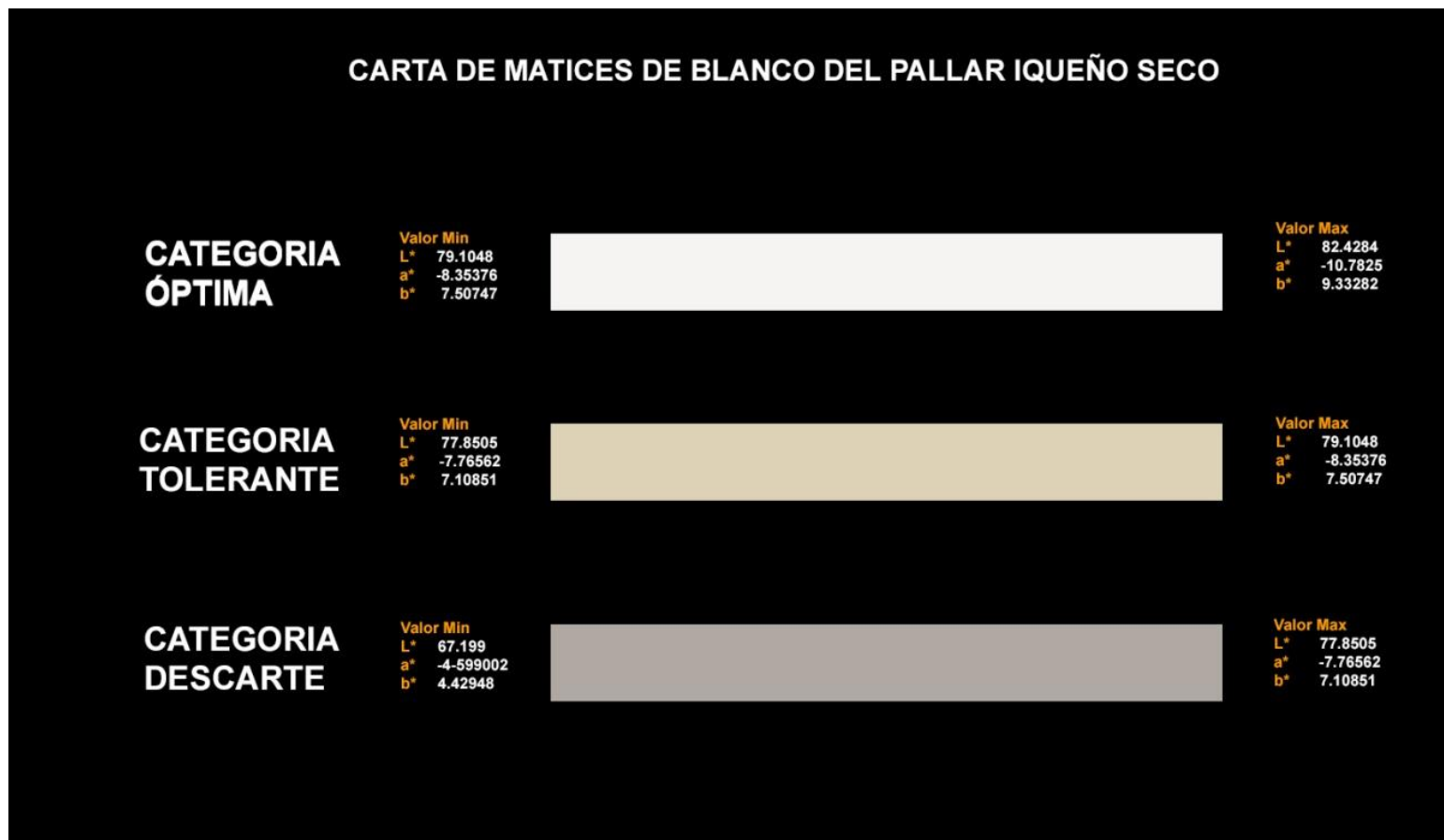


Figura 17. Carta de matices de blanco del Pallar Iqueño

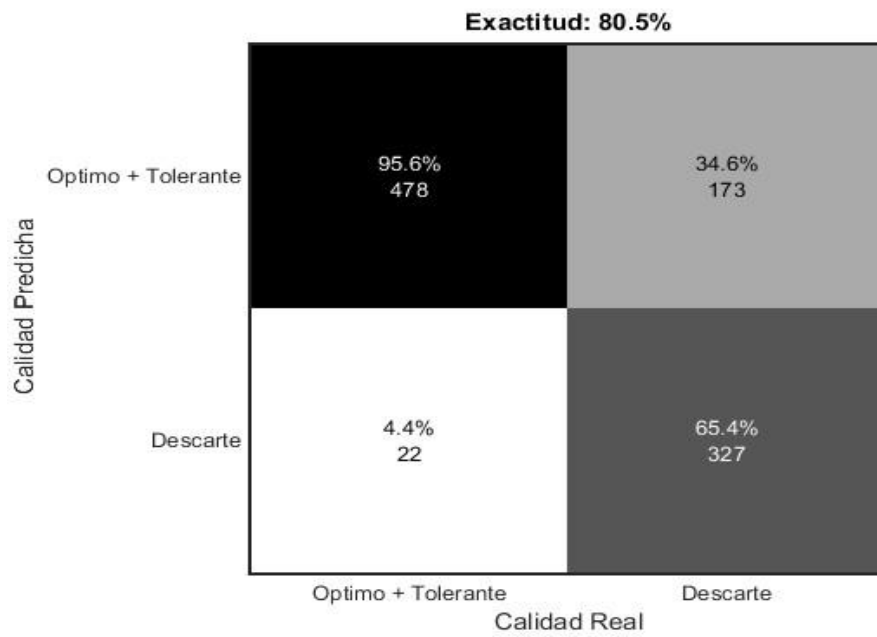
Fuente. Elaboración Propia

4.5. Clasificación del pallar por imágenes digitales

La Eficiencia General (Exactitud) y la Eficiencia por Categorías de clasificación, usando cada uno de los modelos LDA y QDA en 2, 3 y 5 categorías de calidad (clases discriminantes), se presentan en una matriz de confusión obtenida mediante la técnica de validación aleatoria cruzada con 10 repeticiones (Leiva, 2013). Los algoritmos para la clasificación del pallar fueron desarrollados utilizando el software MATLAB v7.5 con Static Toolbox, de manera similar a lo realizado por Venora y otros (2009) en la clasificación de frejoles italianos, y McDonald y otros (2016) en la clasificación de arvejas.

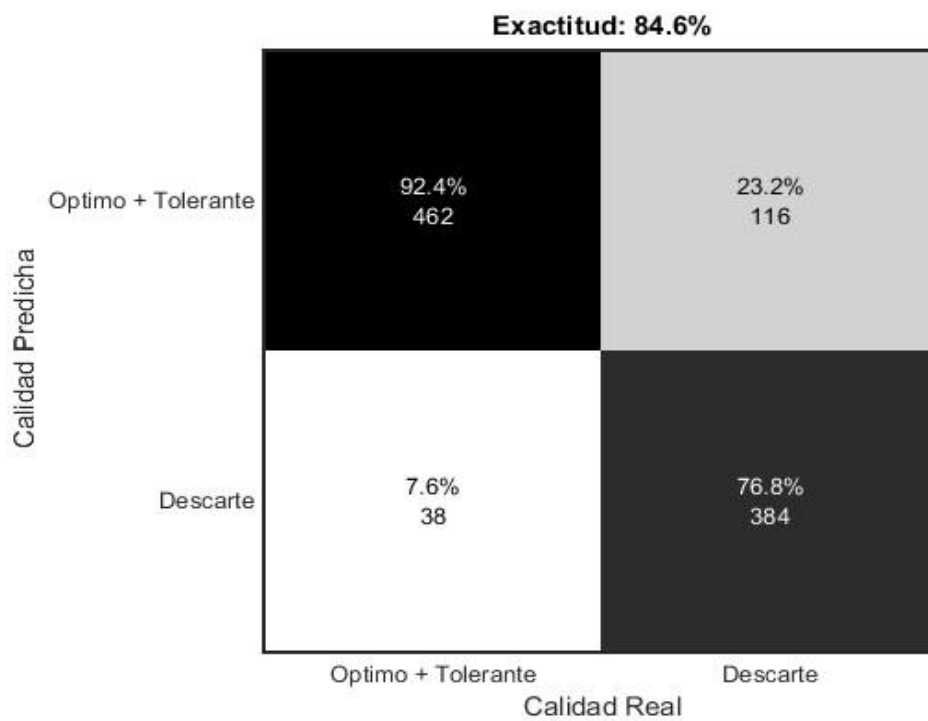
4.5.1. Clasificación en 2 Categorías de calidad

La matriz de confusión de la Figura 18, muestra el resultado de la clasificación en 2 clases discriminantes: categoría de calidad ACEPTABLE (1585 imágenes) y categoría de calidad NO ACEPTABLE (1584 imágenes), respectivamente, empleando las funciones algorítmicas de “numCal” y “numVal” con 500 muestras para la “calibración” y 500 muestras para la “validación” del modelo, de un total de 3169 muestras, con una Eficiencia General de 80.5% en el modelo LDA, valor similar al encontrado a la clasificación automática por PLSR de categorías de calidad ACEPTABLE y NO ACEPTABLE en frejoles negros enlatados con una Eficiencia General de 89.7% (Mendoza y otros, 2016), y con una Eficiencia General (Exactitud) de 84.6% en el modelo QDA, valor similar al encontrado por Baek y otros (2019) en la clasificación por color mediante QDA y LDA entre arroz sano y enfermo, con valores de Eficiencia General (Exactitud) mayores al 90%.



a.

Matriz confusión LDA,



b. Matriz confusión QDA,

Figura 18. Modelos discriminantes de clasificación en 2 categorías de calidad.

Fuente. Elaboración Propia

Así mismo, la Eficiencia de Clase de la categoría ACEPTABLE del modelo LDA, presenta un valor de verdadero positivo de 95.6%, frente a un valor de 92.4% del modelo QDA, y la Eficiencia de Clase de la categoría NO ACEPTABLE, muestra un valor de verdadero positivo de 65.4% para el modelo LDA, frente a un valor de 76.8% del modelo QDA; esto quiere decir, que la Eficiencia General (Exactitud) del modelo LDA se ve afectada por un menor valor en la Eficiencia de Clase de la categoría ACEPTABLE frente al modelo QDA, sin embargo, el modelo LDA, discrimina efectivamente las 2 categorías de calidad tal como se visualiza en el gráfica tri-dimensional (3D) de la Figura 19.

Clasificación en 2 Categorías de Calidad

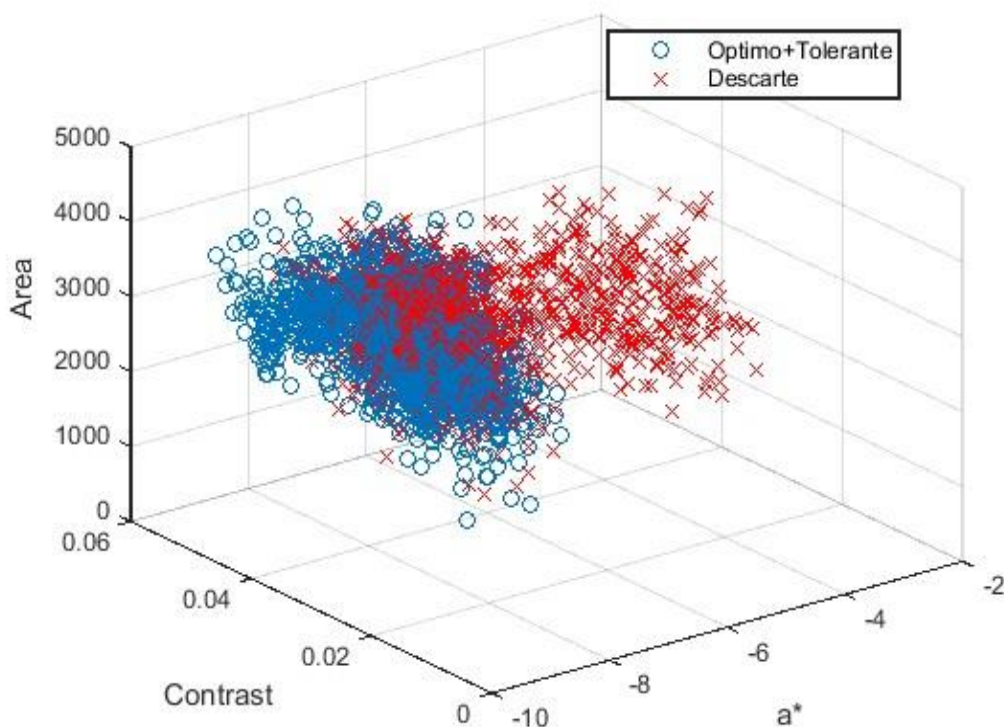


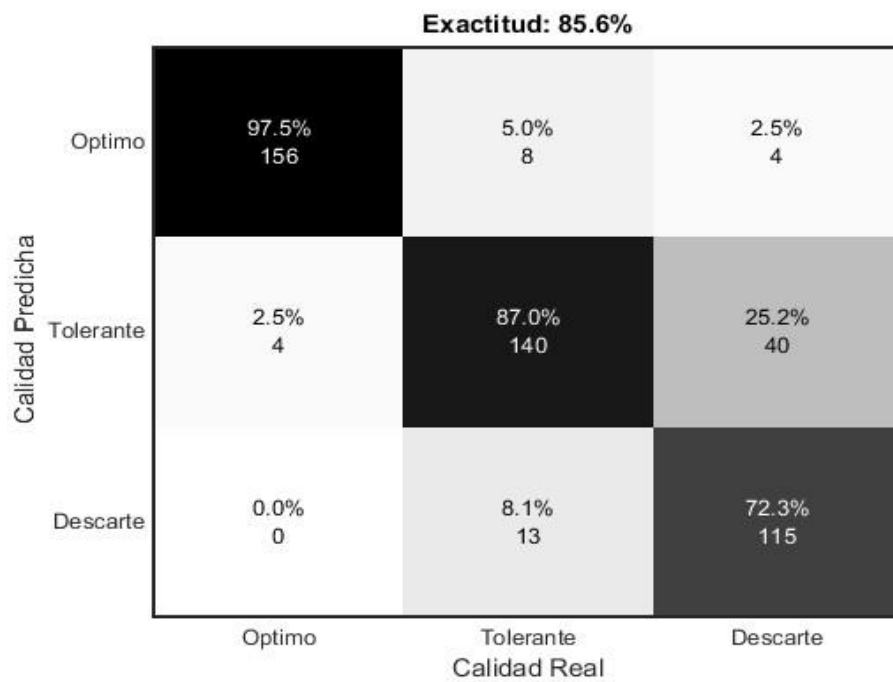
Figura 19. Gráfico 3D del modelo LDA de la clasificación en 2 Categorías de Calidad.

Fuente. Elaboración Propia

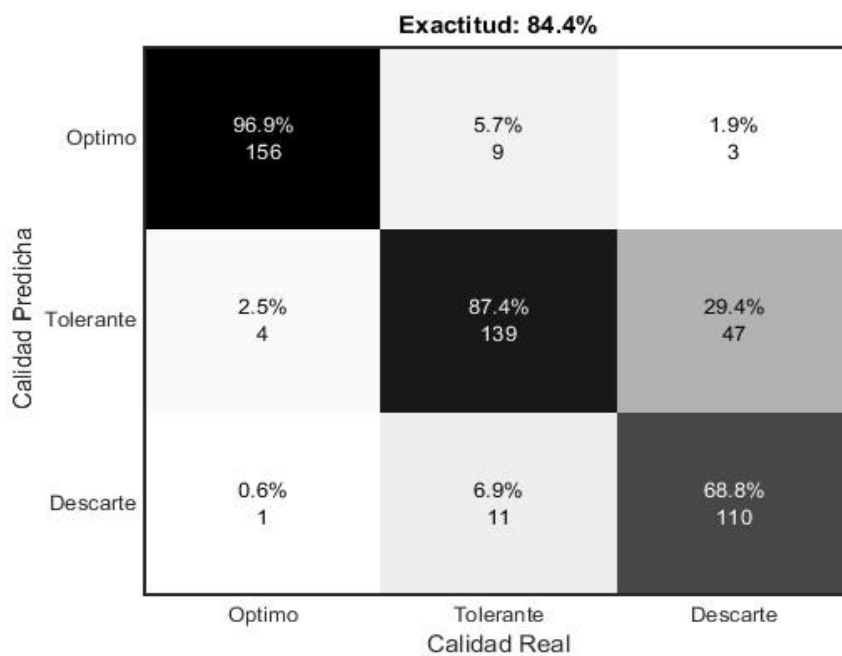
4.5.2. Clasificación en 3 Categorías de Calidad

La matriz de confusión de la Figura 20, muestra el resultado de la clasificación en 3 clases discriminantes, Categoría de Calidad “ÓPTIMO” (346 muestras), Categoría de Calidad TOLERANTE (1239 muestras) y Categoría de Calidad DESCARTE (1584 muestras), respectivamente, con 160 muestras para el entrenamiento y 160 muestras para la validación del modelo, de un total de 3169 muestras, mediante las funciones algorítmicas de “numCal” y “numVal”, respectivamente, con una Eficiencia General de 85.6% para el modelo LDA y una Eficiencia General (Exactitud) de 84.4% para el modelo QDA, valor cercano a la clasificación por redes neuronales del esparrago verde en 3 piezas de corte con una Eficiencia General (Exactitud) de 90.2% (+/- 2.1%) (Donis-Gonzalez y Guyer, 2016), teniendo en cuenta el menor grado de complejidad del tratamiento de los parámetros de entrada.

Así mismo, la Eficiencia de Clase de la categoría de calidad ÓPTIMO en el modelo LDA, presenta un valor 97.5% de verdadero positivo, frente a un valor de 92.4% del modelo QDA; la Eficiencia de Clase de la categoría de calidad TOLERANTE, en el modelo LDA presenta un valor de 87% de verdadero positivo, frente a un valor de 87.4% del modelo QDA; y finalmente la Eficiencia de Clase de la categoría DESCARTE en el modelo LDA, presenta un valor de 72.3% de verdadero positivo frente a un valor de 68.8% del modelo QDA; esto quiere decir, que la Eficiencia General del modelo QDA se ve afectada por un menor valor en la Eficiencia de Clase de la categoría DESCARTE en comparación con el modelo LDA, debido a similitudes en los atributos de color, apariencia, tamaño y forma de las categoría TOLERANTE; sin embargo el modelo LDA discrimina mejor las Categorías ÓPTIMO, TOLERANTE y DESCARTE respecto al modelo QDA, tal como se visualiza en el gráfico tri-dimensional (3D) de la Figura 21.



a. Matriz confusión LDA



b. Matriz confusión QDA

Figura 20. Modelos discriminantes de clasificación en 3 categorías de calidad.

Fuente. Elaboración Propia

Clasificación en 3 Categorías de Calidad

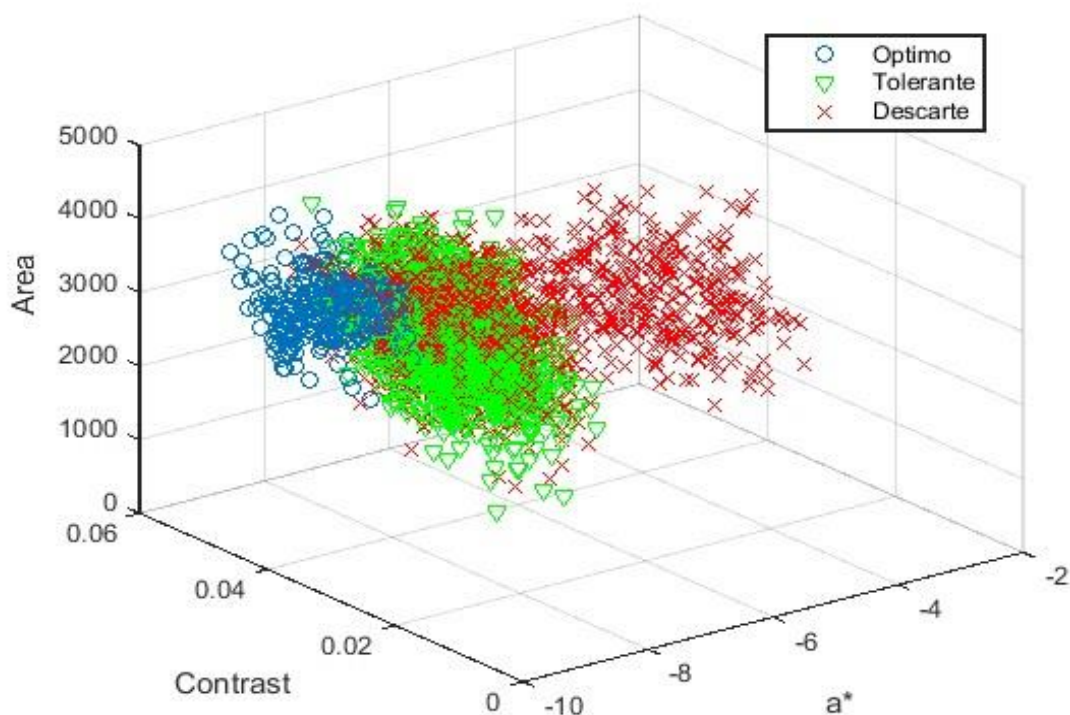


Figura 21. Gráfico 3D del modelo LDA de la clasificación en 3 Categorías de Calidad.

Fuente. Elaboración Propia

4.5.3. Clasificación en 5 Categorías de Calidad

La matriz de confusión de la Figura 22, muestra el resultado de la clasificación en 5 clases discriminantes, categoría de calidad ÓPTIMO (346 muestras), categoría de calidad TOLERANTE (1239 muestras), categoría de calidad DESCARTE FISIOLÓGICO (432 muestras), categoría de calidad DESCARTE MORFOLÓGICO (792 muestras), y categoría de calidad DESCARTE VARIEDAD (360 muestras), respectivamente, con 150 muestras para el “entrenamiento” y 150 muestras la para la “validación” del modelo, de un total de 3169 muestras,

mediante las funciones algorítmicas de “numCal” y “numVal”, respectivamente, con una Eficiencia General (Exactitud) de 87% para el modelo LDA y con una Eficiencia General (Exactitud) de 85.2% para el modelo QDA, valor similar a la clasificación de cítricos por LDA en 5 clases con una Eficiencia General (Exactitud) de 93% (Iqbal y otros, 2016).

Así mismo, la Eficiencia de Clase de la categoría de calidad ÓPTIMA del modelo LDA, muestra un valor de verdadero positivo de 96.7%, frente a un valor de 96.7% del modelo QDA; la categoría de calidad TOLERANTE del modelo LDA muestra un valor de verdadero positivo de 82.6%, frente a un 80.1% valor mayor comparado con el del modelo QDA; la categoría de calidad DESCARTE FISIOLÓGICO del modelo LDA muestra un valor de verdadero positivo de 66.7%, frente a un 60.7% del modelo QDA; para la categoría de calidad DESCARTE MORFOLÓGICO del modelo LDA, muestra un valor de verdadero positivo de 89.3%, frente a un valor de 89.3% del modelo QDA; y finalmente la categoría de calidad DESCARTE VARIEDAD del modelo LDA, muestra un valor de verdadero positivo de 100% frente a un 99.3% del modelo QDA; esto quiere decir, que la Eficiencia general (Exactitud) del modelo QDA se ve afectada por un menor valor de la Eficiencia de Clase de la categoría calidad DESCARTE FISIOLÓGICO, respecto a la Efectividad del modelo LDA, debido a algunas semejanzas en los atributos de color, apariencia, tamaño y forma con la categoría calidad TOLERANTE, pero con una buena discriminación de la categoría calidad DESCARTE FISIOLÓGICO y de la categoría calidad DESCARTE VARIEDAD al mismo tiempo; por tanto, el modelo LDA discrimina mejor las categorías de calidad ÓPTIMA, TOLERANTE, DESCARTE FISIOLÓGICO, DESCARTE MORFOLÓGICO Y DESCARTE VARIEDAD, respecto al modelo QDA, tal como se visualiza en la gráfica tri-dimensional (3D) de la Figura 23.

Exactitud: 87.0%

Calidad Predicha	Optimo	96.7% 145	4.0% 6	2.7% 4	0.0% 0	0.0% 0
	Tolerante	3.3% 5	82.6% 123	30.7% 46	10.7% 16	0.0% 0
	Descarte -	0.0% 0	10.1% 15	66.7% 100	0.0% 0	0.0% 0
	Descarte --	0.0% 0	3.4% 5	0.0% 0	89.3% 134	0.0% 0
	Descarte ---	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 150
		Optimo	Tolerante	Descarte -	Descarte --	Descarte ---
		Calidad Real				

a. Matriz confusión LDA,

Exactitud: 85.2%

Calidad Predicha	Optimo	96.7% 146	5.3% 8	2.7% 4	0.7% 1	0.0% 0
	Tolerante	2.0% 3	80.1% 121	35.3% 53	9.3% 14	0.7% 1
	Descarte -	0.7% 1	10.6% 16	60.7% 91	0.7% 1	0.0% 0
	Descarte --	0.7% 1	4.0% 6	1.3% 2	89.3% 134	0.0% 0
	Descarte ---	0.0% 0	0.0% 0	0.0% 0	0.0% 0	99.3% 149
		Optimo	Tolerante	Descarte -	Descarte --	Descarte ---
		Calidad Real				

b. Matriz confusión QDA,

Figura 22. Modelos discriminantes de clasificación en 5 categorías de calidad.

Fuente. Elaboración Propia

Clasificación en 5 Categorías de Calidad

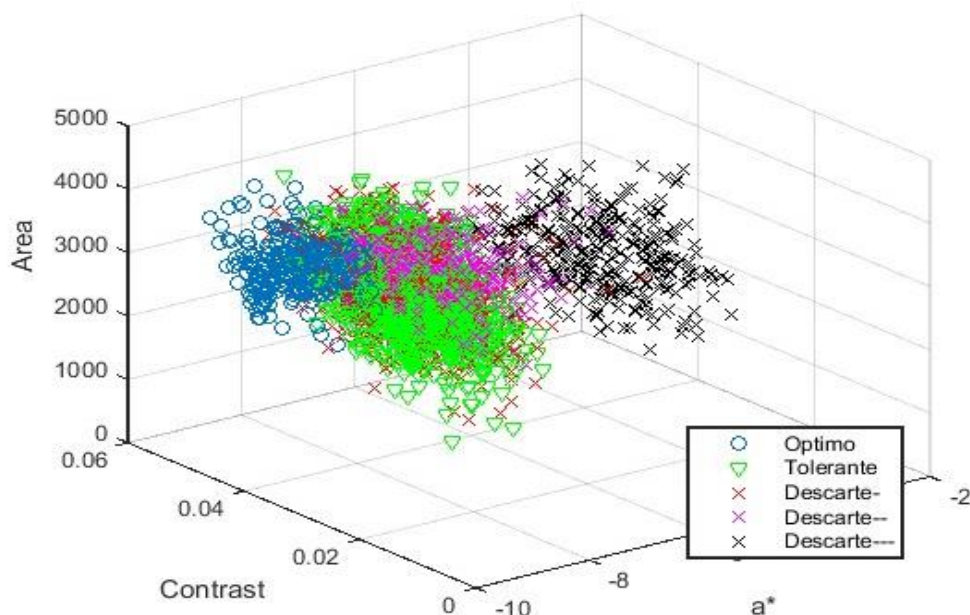


Figura 23. Gráfico 3D del modelo LDA de la clasificación en 5 Categorías de Calidad.

Fuente. Elaboración Propia

4.5.4. Performance de los modelos de clasificación del pallar

En el Cuadro 14, se muestra el comparativo del performance de los modelos LDA y QDA evaluados en 2, 3 y 5 categorías de calidad (clases discriminantes), con un mejor desempeño del modelo LDA evaluado en 3 categorías de calidad, en cuanto muestra una Eficiencia General de 85.6% y una mejor Eficiencia de Clase del pallar sano (categoría TOLERANTE) de 97.5%, que son valores mayores a los modelos LDA clasificados en 2 y 5 categorías de calidad, y a todos los modelos QDA evaluados, y teniendo en cuenta la superposición de las clases (Diaz y otros, 2004) dan resultados falsos positivos bajos los valores son consistentes a los encontrados en la clasificación de trigo siciliano mediante el modelo LDA realizado por Grillo y otros (2017) con una Eficiencia General de 89.7%, en la clasificación de frejoles negros mediante modelo por regresión de mínimos cuadrados (PLSR, por sus siglas en inglés) realizado por Mendoza y otros (2016) con

una Eficiencia General de un 89.7% y al encontrado por Kilic y otros (2007) en la clasificación de frejoles nativos turcos aplicando modelos de redes neuronales artificiales (ANN, por sus siglas en inglés) con Eficiencia General global de 90.6%.

Así mismo el modelo de clasificación LDA en 3 categorías muestra una mejor clasificación de la categoría TOLERANTE, pallar sano de uso comercial con un valor de 97.5% de Eficiencia de Clase respecto a los otros modelos de clasificación evaluados, cuyo valor es superior el encontrado por Constante y otros (2016) con valores de 92.5%, 90%, 90%, y 100% para la clasificación de Fresas en Categoría Extra, Consumo, Materia Prima y Objetos extraños, respectivamente; demostrando que el modelo presenta resultados significativos y confiables para la clasificación del pallar lqueño seco.

Cuadro 14. Eficiencia General y Eficiencia de Clase de los modelos de clasificación

MODELO	CLASS	CATEGORÍAS CALIDAD	EFICIENCIA	EFICIENCIA
LDA	2 CAT	ACEPTABLE	80.5	95.6
LDA	2 CAT	NO ACEPTABLE	80.5	65.4
LDA	3 CAT	ÓPTIMO	85.6	97.5
LDA	3 CAT	TOLERANTE	85.6	87
LDA	3 CAT	DESCARTE	85.6	72.3
LDA	5 CAT	ÓPTIMA	87	96.7
LDA	5 CAT	TOLERANTE	87	82.6
LDA	5 CAT	DESCARTE FISIOLÓGICO	87	66.7
LDA	5 CAT	DESCARTE MORFOLÓGICO	87	89.3
LDA	5 CAT	DESCARTE VARIEDAD	87	100
QDA	2 CAT	ACEPTABLE	84.6	92.4
QDA	2 CAT	NO ACEPTABLE	84.6	76.8
QDA	3 CAT	ÓPTIMO	84.4	96.9
QDA	3 CAT	TOLERANTE	84.4	87.4
QDA	3 CAT	DESCARTE	84.4	68.8
QDA	5 CAT	ÓPTIMO	85.2	96.7
QDA	5 CAT	TOLERANTE	85.2	80.1
QDA	5 CAT	DESCARTE FISIOLÓGICO	85.2	60.7
QDA	5 CAT	DESCARTE MORFOLÓGICO	85.2	89.3
QDA	5 CAT	DESCARTE VARIEDAD	85.2	99.3

Fuente. Elaboración propia

4.5.5. Sensibilidad de los modelos de clasificación del pallar

La sensibilidad de los modelos de clasificación mediante LDA y QDA, para las 2, 3 y 5 categorías de calidad del pallar, se evaluó al medir la Eficiencia General en balanceando las diferentes muestras la calibración/validación, cuyos resultados son el promedio de 10 validaciones cruzadas, que demostró la robustez de los modelos discriminantes, LDA y QDA (Figura 27) con Eficiencia General mayores al 80% en los modelos LDA y con Eficiencia General mayores al 76% en los modelos QDA, demostrando la robustez del modelo, debido a una adecuada elección de los parámetros de entrada, basados en descriptores y las diferencias de estas medidas objetivamente (McDonald y otros, 2016).

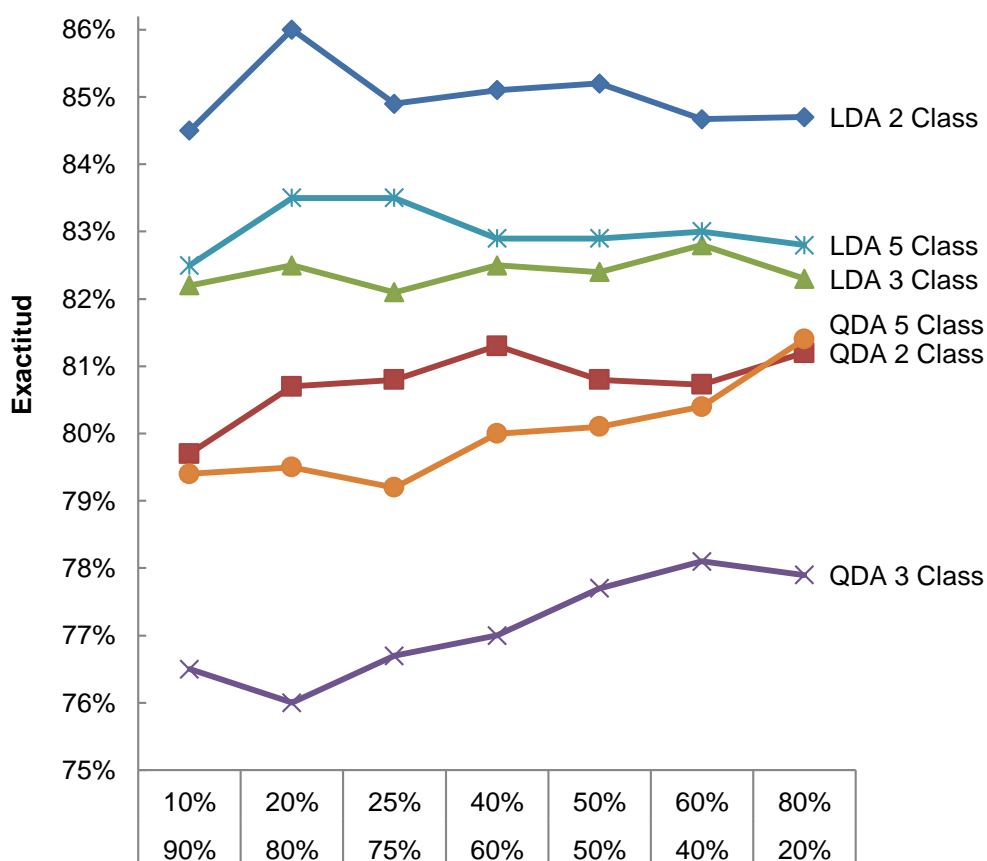


Figura 24. Sensibilidad de Eficiencia General de modelos LDA y QDA

Fuente: elaboración propia

V. CONCLUSIONES

Se ha desarrollado y puesto en operación a nivel de laboratorio un sistema de visión por computadora que utiliza una cámara digital a color, una laptop y software especializado para la captura remota de imágenes de granos de pallar, su procesamiento y análisis, y que, mediante descriptores cuantitativos (color, apariencia, tamaño y forma), identifica y caracteriza granos del Pallar lqueño seco y los clasifica automáticamente en clases de calidad por defectos.

Utilizando el sistema de visión por computadora implementado, se han investigado los modelos de clasificación por Análisis Discriminante Lineal (LDA) y Cuadrático (QDA), en una muestra de 3169 granos de Pallar lqueño seco con defectos predominantes, utilizando 2, 3 y 5 categorías de calidad, con 10 tipos de defectos, con una de las categorías de pallar sano, sin defectos. Los 6 modelos estudiados mostraron Eficiencias Generales (Exactitud) mayores de 70% y Eficiencias por Clases mayores de 70% en la calidad TOLERANTE (verdaderos positivos) y una robustez ACEPTABLE.

El Modelo LDA de 3 Categorías, para la muestra de Pallar lqueño seco, fue seleccionado como el mejor de los seis estudiados, mostrando una Eficiencia General de 85.6%, con la mayor (97.3%) Eficiencia de Clase TOLERANTE (pallar sano sin defectos) y con una muy buena robustez demostrada para diferentes composiciones de las muestras de entrenamiento/validación.

VI. RECOMENDACIONES

La clasificación automática del Pallar Iqueño seco mediante visión por computadora puede ser complementada y mejorada, usando tecnologías del espectro cercano al infrarrojo a fin de discriminar defectos fitosanitarios internos y otros atributos de calidad, que no se pueden determinar por descriptores superficiales de las imágenes digitales.

Optimizar la clasificación automática del pallar por visión por computadora, utilizando dispositivos de captura de imágenes digitales con una mejor sensibilidad, resolución y profundidad; y otros parámetros de captura que permitan incorporar otros factores, como los ambientales, entre los discriminantes.

Investigar comparativamente el comportamiento del modelo de clasificación LDA del Pallar Iqueño seco propuesto, frente a otros modelos de clasificación como redes neuronales artificiales (RNA), máquinas de vectores de soporte (en inglés, Support Vector Machine, SVM), regresión de mínimos cuadrados parciales (Partial Least Squares Regression, PLSR), análisis de componentes principales (Principal Component Analysis, PCA) y otros, con el fin de desarrollar modelos de clasificación de aplicación industrial que minimicen esfuerzos y limitaciones del ser humano mejorando la productividad y competitividad.

VII. BIBLIOGRAFÍA

Altamirano, J. 2011. Evaluación del rendimiento de grano seco de poblaciones promisorias de pallar (*Phaseolus lunatus* L.) tipo Big Lima en condiciones de La Molina. Tesis para obtener el Título Profesional de Ingeniero Agrónomo Universidad Nacional Agraria. La Molina, Perú.

Alvarado, J. 2012. Procesamiento y análisis de imágenes digitales. Notas de Clase para el curso de Grado y Postgrado. Escuela de Ingeniería Electrónica. Instituto Tecnológico de Costa Rica. Recuperado de: <http://www.ietec.org/palvarado/PAID/paid.pdf>

Amat, J. 2016. Análisis Discriminante Lineal (LDA) y Análisis Discriminante Cuadrático (QDA). Ciencia de Datos, Estadística, Programación y Machine Learning. Recuperado de: https://www.cienciadedatos.net/documentos/28_linear_discriminant_analysis_lda_y_quadratic_discriminant_analysis_qda

Baek, I., Kim, M.S., Cho, B-K., Mo, Ch., Barnaby, J.Y., McClung, A.M. y Oh, M. 2019. Selection of optimal hyperspectral wavebands for detection of discolored, diseased rice seeds. Applied Sciences, 9(5):1027. Recuperado de <https://doi.org/10.3390/app9051027>

Brosnan, T. y Sun, D-W. 2004. Improving quality inspection of food products by computer vision—a review. Journal of Food Engineering, 61: 3-16.

Canon A250. 2004. Ficha técnica PowerShot Canon A520. Cámara digital en español. Canon INC. Tokio, Japón.

Constante, P., Gordon, A., Chang, O., Pruma, E., Escobar, I. y Acuña, F. 2016. Artificial vision techniques to optimize strawberry's industrial classification. *IEEE Latin America Transactions*, 14(6):2576-2581. Recuperado de: <https://ieeexplore.ieee.org/abstract/document/7555221>

Diaz, R., Gil, L., Serrano, C., Blasco, M., Moltó, E. y Blasco, J. 2004. Comparison of three algorithms in the classification of table olives by means of computer vision. *Journal of Food Engineering*, 61:101-107.

Donis-Gonzalez, I.R. y Guyer, D.E. 2016. Classification of processing asparagus sections using color image. *Computer and Electronics in Agriculture*, 127: 236-241.

Du, Ch-J. y Su, W-S. (2006). Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering*, 72: 39-55

Espinoza, L. (2005). Mejoramiento genético del pallar (*Phaseolus lunatus* L.) por cruzamiento entre cultigrupos Big Lima, Sieva y Potato en la costa Centro-sur del Perú. Tesis para obtener el Título Profesional de Ingeniero Agrónomo. Universidad Nacional Agraria. La Molina, Perú.

Grillo, O., Blangiforti, S. y Venora, G. 2017. Wheat landraces identification through glumes image analysis. *Computers and Electronics in Agriculture*, 141: 223-231. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S0168169917307172>

Hong, G., Luo, R. M. y Rhodes, P. A. 2001. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Research and Application*, 26: 76-84.

Igathinathane, C., Pordesimo, L.O. y Batchelor, W.D. 2009. Major orthogonal dimensions measurement of food grains by machine vision using image. *J. Food Research International*, 42:76-84.

Inacal. 2016. NTP.205.029.1982 rev 2016. Cereales y Menestras. Análisis Físicos. Dirección de Normalización. Lima, Perú.

Indecopi. 2015. NTP.205.019.2015. Leguminosas. Pallar. Requisitos. Comisión de Normalización y de Fiscalización de Barreras Comerciales no Arancelarias. Lima, Perú.

Iqbal, S. M., Gopal, A., Sankaranarayanan, P. E. y Bair, A. 2016. Classification of selected citrus fruits based on color using Machine Vision System. *International Journal of Food Properties*, 19(2): 272-288.

Jarimopas, B. y Jaisin N. 2008. An experimental machine vision system for sorting sweet tamarind. *Journal of Food Engineering*, 89:291-297.

Javadikia, H., Sabzi, S. y Rabbani, H. 2017. Machine vision based expert system to estimate orange mass of three varieties. *Internati. Int J Agric & Biol Eng*, 10(2):132-139.

Larcher, L. I., Juarez, P. M., Ruggeri, A.I., BIASONI, E.M., Cattaneo, C.A. y Villalba, G.A. 2013. Ponderación de calidad en frutas usando técnicas de visión artificial para la estimación de daños. *Mecánica Computacional*, 32:2473-2484. Asociación Argentina de Mecánica Computacional. Mendoza, Argentina.

Leiva, G. 2013. Assessment of internal and external quality of blueberries using images. Tesis para obtener el Grado Académico de Doctor en

Ciencias de la Ingeniería. Escuela de Ingeniería. Pontificia Universidad Católica de Chile. Santiago de Chile, Chile.

Lo Bianco, M. Grillo, O., Cremonini, R., Sarugi, M. y Venora, G. 2015. Characterization of Italian bean landraces (*Phaseolus vulgaris* L.) using seed image analysis and texture descriptors. Australian Journal of Crop Science, 9(11):1022-1034.

Kilic, K., Boyaci, I., Köksel, H. y Kúsmenoglu, I. 2007. A classification system for beans using computer vision system and artificial neural networks. Journal of Food Engineering, 78 (3):897-904.

Mahajan, S., Das, A. y Sardana, H.K. 2015. Image acquisition techniques for assessment of legume quality. Trends in Food Science & Technology, 42:116-133

Majumdar, S. y Jayas, D. S. (2000). Classification of cereal grains using machine vision: IV. Combined morphology, color, and texture models. Food & Process Engineering Institute of ASAE, 43(6):1689-1694.

Matos, G. y Zúñiga, D. 2002. Comportamiento de cepas nativas de Rizobios aisladas de la Costa del Perú en dos cultivares de Pallar (*Phaseolus lunatus* L.). Ecología Aplicada, 1(1):19-24.

McDonald, L., Panozzo, J., Salisbury, P. y Ford, R. 2016. Discriminant Analysis of Defective and Non-Defective Field Pea (*Pisum sativum* L.) into Broad Market Grades Based on Digital Image Features. PLOS ONE, 11(5): e0155523.pmid:27176469

Mendoza, F. 2005. Characterization of surface appearance and color of some fruits and vegetables using image analysis. Tesis para obtener el

Grado Académico de Doctor. Departamento de Ingeniería Química y Bioprocesos. Pontificia Universidad Católica de Chile. Santiago de Chile, Chile. Recuperado de: http://www.lth.se/fileadmin/livsmedelsteknik/pers_hemsidor/collab/THESIS2005_FMendoza_UC-Chile.pdf

Mendoza, F., & Aguilera, J. M. (2006). Application of Image Analysis for Classification of Ripening Bananas. *Journal of Food Science*, 69(9), E471–E477.

Mendoza, F., Deje, P. y Aguilera, J.M. 2006. Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology*, 41: 285-295.

Mendoza, F. A., Kelly, J. D. y Cichy, K. A. 2016. Automated prediction of sensory scores for color and appearance in canned black beans (*Phaseolus vulgaris L.*) using machine vision. *International Journal of Food Properties*. 20(1), 83-99.

Mendoza, F y Lu, R. 2015. Chapter 2. Basics of image analysis. hyperspectral imaging technology in food and agriculture. *Food Engineering Series*. Eds Park, B. y Lu, R. Springer Science+Business Media. New York, USA.

Moreda, G.P., Ortiz-Cañavate J., García-Ramos, F.J. y Ruiz-Altisent, M. 2009. Non-destructive technologies for fruit and vegetable size determination. *Journal of Food Engineering*, 92:119-136. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S0260877408005451>

Ospina, H. 1980. Guía de Estudio. Diversidad genética de las especies cultivadas del genero *Phaseolus*. CIAT. Serie 04SB-09,02. Cali. Colombia. Recuperado de: <https://books.google.com.pe/books?id=DUG0jfrD51Gc&p>

g=PA45&lpg=PA45&dq=pallar+big+lima&source=IXbl&ots=2oTFVhPszH
&sig=Jphs8Z-1ICSwRKpmEJE4QCtbzk&hl=es419&sa=X&ved=0ahUKEw
i3vc_Evc7PAhXC0h4KHb5PAfIQ6AEIKTAF#v=onepage&q&f=false

Padron, C. 2009. Sistema de visión computarizada y herramientas de diseño gráfico para la obtención de imágenes de muestras de alimentos segmentadas y promediadas en coordenadas CIE-L*a*b*. *Agronomía Costarricense*, 33(2): 283-301.

Sánchez, C., Arizcuren, N., Abril, J. y Casp, A. 2004. Importancia de la visión artificial aplicada a la industria agroalimentaria. *Tecnología de Alimentos*. Escuela Técnica Superior de Ingenieros Agrónomos. Universidad Pública de Navarra, Campus Arrosadía. Pamplona, España.

StatPoint. 2006. Análisis Discriminante. STATGRAPHICS-rev 9/14/2006. StatPoint, Inc.

Valous, N. A., Mendoza, F., Sun, D-W. y Allen, P. 2009. Colour calibration of a laboratory computer vision system for quality valuation of pre-sliced hams. *Meat Science*, 81:132-141.

Valladolid, A. y Voysest, O. 2006. Clases comerciales de leguminosas de grano: Catálogo para orientar la comercialización en los Mercados Nacionales e Internacionales. PROMENESTRAS TEX. Chiclayo, Perú. <https://e.issuu.com/embed.html#18527847/32408671>

Venora, G., Grillo, O., Ravalli, C. y Cremonini, R. 2009. Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system. *Scientia Horticulturae*, 121(4):410-418.

Vidal A., Blasco J. y Talens, P. 2013. Measurement of color of citrus fruits using an automatic computer vision system. Tesis para obtener el Grado Académico de Master. Universidad Politécnica de Valencia. Instituto Valenciano de Investigación Agraria. España. Recuperado de: <https://riunet.upv.es/bitstream/handle/10251/27912/Tesina%20Anna%20Vidal.pdf?sequence=1>

Vithu, P y Moses, J. (2016). Machine vision system for food grain quality evaluation: A Review. *Trends in Food Science & Technology*, 56:13-20.

Westland, S., Ripamonti, C. y Cheung, V. (2012). *Computational colour science using MATLAB®*. (2nd Ed.). John Wiley and Sons, Ltd.: Chichester. United Kingdom.

VIII. ANEXOS

ANEXO 1. ALGORITMOS DE PROCESAMIENTO DE IMÁGENES

```
function [segRGB, BW_noborder, labeled, numObjects] = seg2pallar (rgb)

Gray = rgb2gray (rgb);
level = graythresh(Gray); % *Comenta aquí
BW = im2bw(Gray,level); % *Comenta aquí
se = strel('disk',1);
BW2 = imclose(BW, se);
BW2 = ~bwareaopen(~BW2, 12);
D = -bwdist (~BW2);
Ld = watershed(D);
BW3 = BW2;
BW3(Ld == 0) = 0;
mask = imextendedmin (D,2);
D2 = imimposemin (D, mask);
Ld2 = watershed(D2);
BW4 = BW2; BW4(Ld2 == 0) = 0;
BW_noborder = uint8(imclearborder(BW4, 4));
segRGB = uint8(zeros(size(rgb)));
for ii = 1:3
segRGB(:,:,ii) = rgb(:,:,ii).* BW_noborder;
end
[labeled, numObjects] = bwlabel(BW_noborder,4);
[Bor,T] = bwboundaries(BW_noborder,'noholes');
figure (1), imshow(rgb), title('Segmented Labeled Seeds')
hold on,
for k = 1:length(Bor)
boundary = Bor{k};
```



```
plot(boundary(:,2), boundary(:,1), 'b', 'LineWidth', 1.5)
end
NumPallar = regionprops(labeled, 'Centroid');
for k = 1:numel(NumPallar)
c = NumPallar(k).Centroid;
text(c(1), c(2), sprintf('%d', k), ...
'HorizontalAlignment', 'center', ...
'VerticalAlignment', 'middle');
end
```

ANEXO 2. ALGORITMOS DE ANÁLISIS DE IMÁGENES

```
clc,  
closeal  
|  
  
file = 'C:\Documents and Settings\admin\Escritorio\thesis project  
in pallar\experiments\experiments\Pallar 2018\Lote Cix\';  
COLOR = [];  
TEXTURE = [];  
MORPHO = [];  
  
for i = 099  
I = imread ([file, 'Cix-' num2str(i,'%03d') '.jpg']);  
rgb = I(1:2:end, 1:2:end, 1:3); % decreasing the resolution in 50%  
figure (1); imshow (rgb), title ('Original RGB image')  
[segRGB, BW_noborder, labeled, numObjects] = seg2pallar (rgb)  
for k = 1:numObjects  
    BW_unit = (labeled==k);  
    UnitPallar = background2black(segRGB,BW_unit);  
figure(2), imshow(UnitPallar)  
[Gray, RGB, Lab, HSV] =  
    colour2pallar(UnitPallar); COLOR = [COLOR; i k  
    Gray RGB Lab HSV];  
gray = rgb2gray(UnitPallar);  
    gray2 = gray;  
gray2(~gray2) = NaN;  
[Contrast, Correlation, Energy, Homogeneity] = texture2pallar  
(gray2,1);  
    TEXTURE = [TEXTURE; Contrast, Correlation, Energy,  
Homogeneity];
```

```

stats = regionprops(double(BW_unit),'Area','Perimeter', 'MajorAxisLength',
'MinorAxisLength');
    Area =
    stats.Area;
    Perimeter = stats.Perimeter;
    MajorAxis =
    stats.MajorAxisLength; MinorAxis
    = stats.MinorAxisLength;
    Roundness = 4*pi*Area/Perimeter^2; %Computetheroundness.metric
    MORPHO = [MORPHO; Area, Perimeter, MajorAxis, MinorAxis,
Roundness];
end end

```

```

AAResults_pallar = [COLOR TEXTURE MORPHO]

```

```

function [Gray, RGB, Lab, HSV] = colour2pallar(segRGB)
ima = segRGB;
imshow = size(ima);
gray =
rgb2gray(ima);
gray2 = nonzeros(reshape(gray,imshow(1)*imshow(2),1));
gray_mean = mean(gray2);
gray_std = std(double(gray2));
Gray = [gray_mean gray_std];
R = nonzeros(reshape(ima(:,:,1),imshow(1)*imshow(2),1)); R_mean =
mean(R); R_std = std(double(R));
G = nonzeros(reshape(ima(:,:,2),imshow(1)*imshow(2),1)); G_mean =
mean(G); G_std = std(double(G));
B = nonzeros(reshape(ima(:,:,3),imshow(1)*imshow(2),1)); B_mean =
mean(B); B_std = std(double(B));
RGB = [R_mean G_mean B_mean R_std G_std B_std];
cform = makecform('srgb2lab'); ima_Lab = applycform(ima,cform); Lab =
lab2double(ima_Lab);
L = nonzeros(reshape(Lab(:,:,1),imshow(1)*imshow(2),1)); L_mean =
mean(L); L_std = std(L);
a = nonzeros(reshape(Lab(:,:,2),imshow(1)*imshow(2),1)); a_mean =
mean(a); a_std = std(a);
b = nonzeros(reshape(Lab(:,:,3),imshow(1)*imshow(2),1)); b_mean =
mean(b); b_std = std(b);
Lab = [L_mean a_mean b_mean L_std a_std b_std];
HSV = rgb2hsv(ima);
H = nonzeros(reshape(HSV(:,:,1),imshow(1)*imshow(2),1)); H_mean =
mean(H); H_std = std(H);
S = nonzeros(reshape(HSV(:,:,2),imshow(1)*imshow(2),1)); S_mean =
mean(S); S_std = std(S);
V = nonzeros(reshape(HSV(:,:,3),imshow(1)*imshow(2),1)); V_mean =

```

```

mean(V); V_std = std(V);
HSV = [H_mean S_mean V_mean H_std S_std V_std];
function [Contrast, Correlation, Energy, Homogeneity] =
texture2pallar(gray, dist)
ima = double(gray);
d = dist; Contrast
= []; Correlation =
[]; Energy = [];
Homogeneity = [];
GLCM = graycomatrix(ima, 'Of', [0 d; -d d; -d 0; -d -d]);
stats =
graycoprops(GLCM, {'Contrast', 'Correlation', 'Energy', 'Homogeneity'});
Contrast = [Contrast; mean(stats.Contrast)];
Correlation = [Correlation; mean(stats.Correlation)];
Energy = [Energy; mean(stats.Energy)];
Homogeneity = [Homogeneity; mean(stats.Homogeneity)];

```

ANEXO 3. ALGORITMOS DE APLICACIÓN DE MODELOS DE CLASIFICACIÓN DISCRIMINANTE

Modelo discriminante lineal

```
function [IdaclassL,TestLabel,sizeTrain,sizeTest,errorL] =  
DLA2linear(normData,perc_testing)  
class = normData(:,1);  
imaData = normData(:,2:end);  
y = randperm(size(imaData,1));  
imaData = imaData(y,:);  
Labels = class(y);  
Per = perc_testing;  
Train = imaData(1:ceil(Per*size(imaData,1)),:);  
Test = imaData(1+ceil(Per*size(imaData,1)):end,:);  
TrainLabel = Labels(1:ceil(Per*size(imaData,1)),:);  
TestLabel =  
Labels(1+ceil(Per*size(imaData,1)):end,:);  
[IdaclassL,errorL,pL,logpL,coeffL] = classify(Test,Train,TrainLabel,'linear');  
errorL =  
100*sum(IdaclassL==TestLabel)/size(TestLabel,1);  
disp(['Linear accuracy is '  
int2str(100*sum(IdaclassL==TestLabel)/size(TestLabel,1))])  
sizeTrain = size(TrainLabel,1);  
sizeTest = size(TestLabel,1);
```

Modelo discriminante cuadrático

```
function [IdaclassQ,TestLabel,sizeTrain,sizeTest,errorQ] =  
DLA2quadratic(normData,perc_testing)  
class = normData(:,1);  
imaData = normData(:,2:end);
```

```

y = randperm(size(imaData,1));
imaData = imaData(y,:);
Labels = class(y);
Per = perc_testing;
Train = imaData(1:ceil(Per*size(imaData,1)),:);
Test = imaData(1+ceil(Per*size(imaData,1)):end,:);
TrainLabel = Labels(1:ceil(Per*size(imaData,1)),:);
TestLabel =
Labels(1+ceil(Per*size(imaData,1)):end,:);
[ldaclassQ,errorQ,pQ,logpQ,coeffQ] =
classify(Test,Train,TrainLabel,'Quadratic');
errorQ = 100*sum(ldaclassQ==TestLabel)/size(TestLabel,1);
disp(['Quadratic accuracy is '
int2str(100*sum(ldaclassQ==TestLabel)/size(TestLabel,1))])
sizeTrain = size(TrainLabel,1);
sizeTest = size(TestLabel,1);

```

Clasificación en 2 categorías de calidad

```
clc
clear all
close all
load Quality_2
usualvalue = find(Quality_2(:,2) < 256 & Quality_2(:,4) < 256 ...
    & Quality_2(:,5) < 256 & Quality_2(:,6) < 256 & Quality_2(:,11) < -2);
class = Quality_2(usualvalue,1); imaData
= Quality_2(usualvalue,2:end);
normQuality_2 = [class imaData];
OptplusTol = (find(normQuality_2(:,1) == 1)); % Categoría Optima +
Tolerante
Descarte = find(normQuality_2(:,1) == 2); % Categoría considerada
inaceptable
OptToleran2Q = normQuality_2(OptplusTol,:);
Descarte2Q = normQuality_2(Descarte,:);
numCal = 500;
numVal = 500;
OptToleran2Q_CalVal = cal2val(OptToleran2Q,numCal,numVal);
Descarte2Q_CalVal = cal2val(Descarte2Q,numCal,numVal);
cal = genvarname(repmat({'Cal'},1,10),'Cal');
val = genvarname(repmat({'Val'},1,10),'Val');
confmat = zeros(2,2,10);
for k = 1:numel(cal)
    cal_class1 = OptToleran2Q_CalVal.(cal{k});
    cal_class2 = Descarte2Q_CalVal.(cal{k});
    calib = [cal_class1; cal_class2];
    modelLinear = fitcdiscr(calib(:,2:end),calib(:,1));
    val_class1 = OptToleran2Q_CalVal.(val{k});
    val_class2 = Descarte2Q_CalVal.(val{k});
    valid = [val_class1; val_class2];
```



```

    predClass = predict(modelLinear,valid(:,2:end));
    confmat(:,:,k) = confusionmat(predClass,valid(:,1));
end
confmat_avgL = zeros(2,2);
confmat_avgL(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgL(1,2) = sum(confmat(1,2,1:10))/10;
confmat_avgL(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgL(2,2) = sum(confmat(2,2,1:10))/10;
figure(1), plotConfMat(round(confmat_avgL), {'Optimo + Tolerante',
'Descarte'});
confmat = zeros(2,2,10);
for k = 1:numel(cal)
    cal_class1 = OptToleran2Q_CalVal.(cal{k});
    cal_class2 = Descarte2Q_CalVal.(cal{k});
    calib = [cal_class1; cal_class2];
    modelQuadratic =
fitcdiscr(calib(:,2:end),calib(:,1),'DiscrimType','quadratic');
    val_class1 = OptToleran2Q_CalVal.(val{k});
    val_class2 = Descarte2Q_CalVal.(val{k});
    valid = [val_class1; val_class2];
    predClass = predict(modelQuadratic,valid(:,2:end));
    confmat(:,:,k) = confusionmat(predClass,valid(:,1));
end
confmat_avgL = zeros(2,2);
confmat_avgL(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgL(1,2) = sum(confmat(1,2,1:10))/10;
confmat_avgL(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgL(2,2) = sum(confmat(2,2,1:10))/10;
figure(2), plotConfMat(round(confmat_avgL), {'Optimo + Tolerante',
'Descarte'});
R = normQuality_2(:,4);

```

```

a = normQuality_2(:,11); S =
normQuality_2(:,12); Contrast =
normQuality_2(:,22); Area =
normQuality_2(:,26); Classes =
normQuality_2(:,1); figure(3),
h1 = gscatter(a, Area, Classes,'rb','ov',[],'off');
h1(1).LineWidth = 1;
h1(2).LineWidth = 1;
legend('Optimo+Tolerante', 'Descarte','Location','best')
hold on
figure(4),
% a*, Area, Contrast
scatter3(OptToleran2Q(:,11),OptToleran2Q(:,22),OptToleran2Q(:,26));
hold on,
scatter3(Descarte2Q(:,11),Descarte2Q(:,22),Descarte2Q(:,26),'rx');
xlabel('a*'), ylabel('Contrast'), zlabel('Area')
legend('Optimo+Tolerante', 'Descarte','Location','best')

```

Clasificación en 3 categorías de calidad

```

clc
clear all
close all
load Quality_3
usualvalue = find(Quality_3(:,2) < 256 & Quality_3(:,4) < 256 ...
    & Quality_3(:,5) < 256 & Quality_3(:,6) < 256 & Quality_3(:,11) < -2);
class = Quality_3(usualvalue,1); imaData
= Quality_3(usualvalue,2:end);
normQuality_3 = [class imaData];
Optimo = find(normQuality_3(:,1) == 1); % Categoría Optima
Tolerante = find(normQuality_3(:,1) == 2); % Categoría Tolerante

```

```

Descarte = find(normQuality_3(:,1) == 3); % Categoria considerada
Inacceptable
Optimo3Q = normQuality_3(Optimo,:);
Tolerante3Q = normQuality_3(Tolerante,:);
Descarte3Q = normQuality_3(Descarte,:);
numCal = 160;
numVal = 160;
Optimo3Q_CalVal = cal2val(Optimo3Q,numCal,numVal);
Tolerante3Q_CalVal = cal2val(Tolerante3Q,numCal,numVal);
Descarte3Q_CalVal = cal2val(Descarte3Q,numCal,numVal); cal
= genvarname(repmat({'Cal'},1,10),'Cal');
val = genvarname(repmat({'Val'},1,10),'Val');
confmat = zeros(3,3,10);
for k = 1:numel(cal)
    cal_class1 = Optimo3Q_CalVal.(cal{k});
    cal_class2 = Tolerante3Q_CalVal.(cal{k});
    cal_class3 = Descarte3Q_CalVal.(cal{k});
    calib = [cal_class1; cal_class2; cal_class3];
    modelLinear = fitcdiscr(calib(:,2:end),calib(:,1)); val_class1 =
    Optimo3Q_CalVal.(val{k}); val_class2 = Tolerante3Q_CalVal.(val{k});
    val_class3 = Descarte3Q_CalVal.(val{k});
    valid = [val_class1; val_class2; val_class3]; predClass =
    predict(modelLinear,valid(:,2:end)); confmat(:,:,k) =
    confusionmat(predClass,valid(:,1));
end
confmat_avgL = zeros(3,3);
confmat_avgL(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgL(1,2) = sum(confmat(1,2,1:10))/10;
confmat_avgL(1,3) = sum(confmat(1,3,1:10))/10;
confmat_avgL(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgL(2,2) = sum(confmat(2,2,1:10))/10;

```

```

confmat_avgL(2,3) = sum(confmat(2,3,1:10))/10;
confmat_avgL(3,1) = sum(confmat(3,1,1:10))/10;
confmat_avgL(3,2) = sum(confmat(3,2,1:10))/10;
confmat_avgL(3,3) = sum(confmat(3,3,1:10))/10;
figure(1), plotConfMat(round(confmat_avgL), {'Optimo', 'Tolerante',
'Descarte'});
confmat = zeros(3,3,10);
for k = 1:numel(cal)
    cal_class1 = Optimo3Q_CalVal.(cal{k});
    cal_class2 = Tolerante3Q_CalVal.(cal{k});
    cal_class3 = Descarte3Q_CalVal.(cal{k});
    calib = [cal_class1; cal_class2; cal_class3];
    modelQuadratic =
fitcdiscr(calib(:,2:end),calib(:,1),'DiscrimType','quadratic');
    val_class1 = Optimo3Q_CalVal.(val{k});
    val_class2 = Tolerante3Q_CalVal.(val{k});
    val_class3 = Descarte3Q_CalVal.(val{k});
    valid = [val_class1; val_class2; val_class3];
    predClass = predict(modelQuadratic,valid(:,2:end));
    confmat(:, :,k) = confusionmat(predClass,valid(:,1));
end
confmat_avgQ = zeros(3,3);
confmat_avgQ(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgQ(1,2) = sum(confmat(1,2,1:10))/10;
confmat_avgQ(1,3) = sum(confmat(1,3,1:10))/10;
confmat_avgQ(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgQ(2,2) = sum(confmat(2,2,1:10))/10;
confmat_avgQ(2,3) = sum(confmat(2,3,1:10))/10;
confmat_avgQ(3,1) = sum(confmat(3,1,1:10))/10;
confmat_avgQ(3,2) = sum(confmat(3,2,1:10))/10;
confmat_avgQ(3,3) = sum(confmat(3,3,1:10))/10;

```

```

figure(2), plotConfMat(round(confmat_avgQ), {'Optimo', 'Tolerante',
'Descarte'});
R = normQuality_3(:,4);
a = normQuality_3(:,11); S =
normQuality_3(:,12); Contrast =
normQuality_3(:,22); Area =
normQuality_3(:,26); Classes =
normQuality_3(:,1); figure(3),
h1 = gscatter(a, Area, Classes, 'bgr', 'ovx', [], 'off');
h1(1).LineWidth = 1;
h1(2).LineWidth = 1;
h1(3).LineWidth = 1;
legend('Optimo', 'Tolera
nte', 'Descarte', 'Locatio
n', 'best') hold on
figure(4),
% a*, Area, Contrast
scatter3(Optimo3Q(:,11), Optimo3Q(:,22), Optimo3Q(:,26));
hold on,
scatter3(Tolerante3Q(:,11), Tolerante3Q(:,22), Tolerante3Q(:,26), 'gv');
scatter3(Descarte3Q(:,11), Descarte3Q(:,22), Descarte3Q(:,26), 'rx');
xlabel('a*'), ylabel('Contrast'), zlabel('Area')
legend('Optimo', 'Tolerante', 'Descarte', 'Location', 'best')

```

Clasificación en 5 categorías de calidad

```

clc
clear all
close all
load Quality_5
usualvalue = find(Quality_5(:,2) < 256 & Quality_5(:,4) < 256 ...
& Quality_5(:,5) < 256 & Quality_5(:,6) < 256 & Quality_5(:,11) < -2);

```

```

class = Quality_5(usualvalue,1); imaData
= Quality_5(usualvalue,2:end);
normQuality_5 = [class imaData];
Optimo = find(normQuality_5(:,1) == 1); % Categoria Optima Tolerante
= find(normQuality_5(:,1) == 2); % Categoria Tolerante DescarteC =
find(normQuality_5(:,1) == 3); % Categoria considerada Descarte-
DescarteD = find(normQuality_5(:,1) == 4); % Categoria considerada
Descarte--
DescarteE = find(normQuality_5(:,1) == 5); % Categoria considerada
Descarte---
Optimo5Q = normQuality_5(Optimo,:);
Tolerante5Q = normQuality_5(Tolerante,:);
Descarte5Qc = normQuality_5(DescarteC,:);
Descarte5Qd = normQuality_5(DescarteD,:);
Descarte5Qe = normQuality_5(DescarteE,:);
numCal = 150;
numVal = 150;
Optimo5Q_CalVal = cal2val(Optimo5Q,numCal,numVal);
Tolerante5Q_CalVal = cal2val(Tolerante5Q,numCal,numVal);
Descarte5Qc_CalVal = cal2val(Descarte5Qc,numCal,numVal);
Descarte5Qd_CalVal = cal2val(Descarte5Qd,numCal,numVal);
Descarte5Qe_CalVal = cal2val(Descarte5Qe,numCal,numVal);
cal = genvarname(repmat({'Cal'},1,10),'Cal');
val = genvarname(repmat({'Val'},1,10),'Val');
confmat = zeros(5,5,10);
for k = 1:numel(cal)
    cal_class1 = Optimo5Q_CalVal.(cal{k});
    cal_class2 = Tolerante5Q_CalVal.(cal{k});
    cal_class3 = Descarte5Qc_CalVal.(cal{k});
    cal_class4 = Descarte5Qd_CalVal.(cal{k});
    cal_class5 = Descarte5Qe_CalVal.(cal{k});

```

```

calib = [cal_class1; cal_class2; cal_class3; cal_class4; cal_class5];
modelLinear = fitcdiscr(calib(:,2:end),calib(:,1));
val_class1 = Optimo5Q_CalVal.(val{k});
val_class2 = Tolerante5Q_CalVal.(val{k});
val_class3 = Descarte5Qc_CalVal.(val{k});
val_class4 = Descarte5Qd_CalVal.(val{k});
val_class5 = Descarte5Qe_CalVal.(val{k});
valid = [val_class1; val_class2; val_class3; val_class4; val_class5];
predClass = predict(modelLinear,valid(:,2:end));
confmat(:,:,k) = confusionmat(predClass,valid(:,1));
end
confmat_avgL = zeros(5,5);
confmat_avgL(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgL(1,2) = sum(confmat(1,2,1:10))/10;
confmat_avgL(1,3) = sum(confmat(1,3,1:10))/10;
confmat_avgL(1,4) = sum(confmat(1,4,1:10))/10;
confmat_avgL(1,5) = sum(confmat(1,5,1:10))/10;
confmat_avgL(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgL(2,2) = sum(confmat(2,2,1:10))/10;
confmat_avgL(2,3) = sum(confmat(2,3,1:10))/10;
confmat_avgL(2,4) = sum(confmat(2,4,1:10))/10;
confmat_avgL(2,5) = sum(confmat(2,5,1:10))/10;
confmat_avgL(3,1) = sum(confmat(3,1,1:10))/10;
confmat_avgL(3,2) = sum(confmat(3,2,1:10))/10;
confmat_avgL(3,3) = sum(confmat(3,3,1:10))/10;
confmat_avgL(3,4) = sum(confmat(3,4,1:10))/10;
confmat_avgL(3,5) = sum(confmat(3,5,1:10))/10;
confmat_avgL(4,1) = sum(confmat(4,1,1:10))/10;
confmat_avgL(4,2) = sum(confmat(4,2,1:10))/10;
confmat_avgL(4,3) = sum(confmat(4,3,1:10))/10;
confmat_avgL(4,4) = sum(confmat(4,4,1:10))/10;

```

```

confmat_avgL(4,5) = sum(confmat(4,5,1:10))/10;
confmat_avgL(5,1) = sum(confmat(5,1,1:10))/10;
confmat_avgL(5,2) = sum(confmat(5,2,1:10))/10;
confmat_avgL(5,3) = sum(confmat(5,3,1:10))/10;
confmat_avgL(5,4) = sum(confmat(5,4,1:10))/10;
confmat_avgL(5,5) = sum(confmat(5,5,1:10))/10;
figure(1), plotConfMat(round(confmat_avgL), {'Optimo', 'Tolerante',
'Descarte -',...
'Descarte --','Descarte ---'});
confmat = zeros(5,5,10);
for k = 1:numel(cal)
    cal_class1 = Optimo5Q_CalVal.(cal{k});
    cal_class2 = Tolerante5Q_CalVal.(cal{k});
    cal_class3 = Descarte5Qc_CalVal.(cal{k});
    cal_class4 = Descarte5Qd_CalVal.(cal{k});
    cal_class5 = Descarte5Qe_CalVal.(cal{k});
    calib = [cal_class1; cal_class2; cal_class3; cal_class4; cal_class5];
    modelQuadratic =
fitcdiscr(calib(:,2:end),calib(:,1),'DiscrimType','quadratic');
    val_class1 = Optimo5Q_CalVal.(val{k});
    val_class2 = Tolerante5Q_CalVal.(val{k});
    val_class3 = Descarte5Qc_CalVal.(val{k});
    val_class4 = Descarte5Qd_CalVal.(val{k});
    val_class5 = Descarte5Qe_CalVal.(val{k});
    valid = [val_class1; val_class2; val_class3; val_class4; val_class5];
    predClass = predict(modelQuadratic,valid(:,2:end));
    confmat(:, :, k) = confusionmat(predClass,valid(:, 1));
end
confmat_avgQ = zeros(5,5);
confmat_avgQ(1,1) = sum(confmat(1,1,1:10))/10;
confmat_avgQ(1,2) = sum(confmat(1,2,1:10))/10;

```



```

confmat_avgQ(1,3) = sum(confmat(1,3,1:10))/10;
confmat_avgQ(1,4) = sum(confmat(1,4,1:10))/10;
confmat_avgQ(1,5) = sum(confmat(1,5,1:10))/10;
confmat_avgQ(2,1) = sum(confmat(2,1,1:10))/10;
confmat_avgQ(2,2) = sum(confmat(2,2,1:10))/10;
confmat_avgQ(2,3) = sum(confmat(2,3,1:10))/10;
confmat_avgQ(2,4) = sum(confmat(2,4,1:10))/10;
confmat_avgQ(2,5) = sum(confmat(2,5,1:10))/10;
confmat_avgQ(3,1) = sum(confmat(3,1,1:10))/10;
confmat_avgQ(3,2) = sum(confmat(3,2,1:10))/10;
confmat_avgQ(3,3) = sum(confmat(3,3,1:10))/10;
confmat_avgQ(3,4) = sum(confmat(3,4,1:10))/10;
confmat_avgQ(3,5) = sum(confmat(3,5,1:10))/10;
confmat_avgQ(4,1) = sum(confmat(4,1,1:10))/10;
confmat_avgQ(4,2) = sum(confmat(4,2,1:10))/10;
confmat_avgQ(4,3) = sum(confmat(4,3,1:10))/10;
confmat_avgQ(4,4) = sum(confmat(4,4,1:10))/10;
confmat_avgQ(4,5) = sum(confmat(4,5,1:10))/10;
confmat_avgQ(5,1) = sum(confmat(5,1,1:10))/10;
confmat_avgQ(5,2) = sum(confmat(5,2,1:10))/10;
confmat_avgQ(5,3) = sum(confmat(5,3,1:10))/10;
confmat_avgQ(5,4) = sum(confmat(5,4,1:10))/10;
confmat_avgQ(5,5) = sum(confmat(5,5,1:10))/10;
figure(2), plotConfMat(round(confmat_avgQ), {'Optimo', 'Tolerante',
'Descarte -',...
'Descarte --','Descarte ---'});
R = normQuality_5(:,4);
a = normQuality_5(:,11); S =
normQuality_5(:,12); Contrast =
normQuality_5(:,22); Area =
normQuality_5(:,26);

```

```

Classes = normQuality_5(:,1);
figure(3),
h1 = gscatter(a, Area, Classes,'bgrcm','ovxxx',[],'off');
h1(1).LineWidth = 1;
h1(2).LineWidth = 1;
h1(3).LineWidth = 1;
h1(4).LineWidth = 1;
h1(5).LineWidth = 1;
legend('Optimo','Tolerante','Descarte-','Descarte--','Descarte---',...
'Location','best'), hold on figure(4),
scatter3(Optimo5Q(:,11),Optimo5Q(:,22),Optimo5Q(:,26));
hold on,
scatter3(Tolerante5Q(:,11),Tolerante5Q(:,22),Tolerante5Q(:,26),'gv');
scatter3(Descarte5Qc(:,11),Descarte5Qc(:,22),Descarte5Qc(:,26),'rx');
scatter3(Descarte5Qd(:,11),Descarte5Qd(:,22),Descarte5Qd(:,26),'mx');
scatter3(Descarte5Qe(:,11),Descarte5Qe(:,22),Descarte5Qe(:,26),'kx');
xlabel('a*'), ylabel('Contrast'), zlabel('Area')
legend('Optimo','Tolerante','Descarte-','Descarte--','Descarte---',...
'Location','best')

```

Calibración de modelos

```

function SampleStruc = cal2val(data,num_cal,num_val)
num = size(data,1);
data1 = [rand(num,1) data];
data2 = [rand(num,1) data];
data3 = [rand(num,1) data];
data4 = [rand(num,1) data];
data5 = [rand(num,1) data];
data6 = [rand(num,1) data];
data7 = [rand(num,1) data];
data8 = [rand(num,1) data];

```

```

data9 = [rand(num,1) data];
data10= [rand(num,1) data];
[value1, index1] = sort(data1(:,1));
[value2, index2] = sort(data2(:,1));
[value3, index3] = sort(data3(:,1));
[value4, index4] = sort(data4(:,1));
[value5, index5] = sort(data5(:,1));
[value6, index6] = sort(data6(:,1));
[value7, index7] = sort(data7(:,1));
[value8, index8] = sort(data8(:,1));
[value9, index9] = sort(data9(:,1));
[value10,index10]= sort(data10(:,1));
d1 = data(index1,:);
d2 = data(index2,:); d3
= data(index3,:); d4 =
data(index4,:); d5 =
data(index5,:); d6 =
data(index6,:); d7 =
data(index7,:); d8 =
data(index8,:); d9 =
data(index9,:); d10=
data(index10,:);
Cal1 = sortrows(d1(1:num_cal,:)); % Run 1
Val1 = sortrows(d1((num_cal+1):(num_cal+num_val),:));
Cal2 = sortrows(d2(1:num_cal,:)); % Run 2
Val2 = sortrows(d2((num_cal+1):(num_cal+num_val),:));
Cal3 = sortrows(d3(1:num_cal,:)); % Run 3
Val3 = sortrows(d3((num_cal+1):(num_cal+num_val),:));
Cal4 = sortrows(d4(1:num_cal,:)); % Run 4
Val4 = sortrows(d4((num_cal+1):(num_cal+num_val),:));
Cal5 = sortrows(d5(1:num_cal,:)); % Run 5

```

```

Val5 = sortrows(d5((num_cal+1):(num_cal+num_val),:));
Cal6 = sortrows(d6(1:num_cal,:));    % Run 6
Val6 = sortrows(d6((num_cal+1):(num_cal+num_val),:));
Cal7 = sortrows(d7(1:num_cal,:));    % Run 7
Val7 = sortrows(d7((num_cal+1):(num_cal+num_val),:));
Cal8 = sortrows(d8(1:num_cal,:));    % Run 8
Val8 = sortrows(d8((num_cal+1):(num_cal+num_val),:));
Cal9 = sortrows(d9(1:num_cal,:));    % Run 9
Val9 = sortrows(d9((num_cal+1):(num_cal+num_val),:));
Cal10= sortrows(d10(1:num_cal,:));   % Run 10
Val10= sortrows(d10((num_cal+1):(num_cal+num_val),:));
SampleStruc = struct('Cal1', Cal1, 'Cal2', Cal2, 'Cal3', Cal3, 'Cal4', Cal4,
'Cal5', Cal5,...
    'Cal6', Cal6, 'Cal7', Cal7, 'Cal8', Cal8, 'Cal9', Cal9, 'Cal10', Cal10,...
    'Val1', Val1, 'Val2', Val2, 'Val3', Val3, 'Val4', Val4, 'Val5', Val5,...

```

Normalización de la data

```

function normImaData = norm2jsignal(imaData) R
    = imaData;
    [m,n] = size(R); ndata
    = zeros(m,n); for
    j=1:n
        for i=1:m
            avg = mean(R(:,j));
            stdev = std(R(:,j));
            ndata(i,j) = (R(i,j) - avg)/stdev;
        end
    end
    normImaData = ndata;

```

Ploteo

```
function plotConfMat(varargin)
switch (nargin)
    case 0
        confmat = 1;
        labels = {'1'};
    case 1
        confmat = varargin{1};
        labels = 1:size(confmat, 1);
    otherwise
        confmat = varargin{1};
        labels = varargin{2};
end
confmat(isnan(confmat))=0; % in case there are NaN elements
numlabels = size(confmat, 1); % number of labels
confpercent = 100*confmat./repmat(sum(confmat, 1),numlabels,1);
imagesc(confpercent);
title(sprintf('Eficiencia General: %.1f%%',
100*trace(confmat)/sum(confmat(:))));
ylabel('Calidad Predicha'); xlabel('Calidad Real');
colormap(flipud(gray));
textStrings = num2str([confpercent(:), confmat(:)], '%.1f%%\n%d\n');
textStrings = strtrim(cellstr(textStrings));
[x,y] = meshgrid(1:numlabels);
hStrings = text(x(:),y(:),textStrings(:), ...
    'HorizontalAlignment','center');
midValue = mean(get(gca,'CLim'));
textColors = repmat(confpercent(:) > midValue,1,3);
set(hStrings,{'Color'},num2cell(textColors,2));
set(gca,'XTick',1:numlabels,...
    'XTickLabel',labels,...
```

```
'YTick',1:numlabels,...
'YTickLabel',labels,...
'TickLength',[0 0]);
```

Correlación de parámetros

```
load Table_5_clases
figure, corrplot(Table_5_clases(:,2:3)); % Gray scale
figure, corrplot(Table_5_clases(:,2:3)); % RGB scale including STD
dataPallar = [Table_5_clases(:,2) Table_5_clases(:,4:6)
Table_5_clases(:,10:12) Table_5_clases(:,22:30)];
corrplot (dataPallar)
load Quality_5
usualvalue = find(Quality_5(:,2) < 256 & Quality_5(:,4) < 256 ...
& Quality_5(:,5) < 256 & Quality_5(:,6) < 256 & Quality_5(:,11) < -2);
class = Quality_5(usualvalue,1); imaData
= Quality_5(usualvalue,2:end); data =
[class imaData];
figure(1), plot (data(:,1), data(:,2), 'bo'); R
= corrcoef(data(:,1), data(:,2));
R_gris = R(1,2)
figure(2), plot (data(:,1), data(:,4), 'ro'); R
= corrcoef(data(:,1), data(:,4)); R_red =
R(1,2)
figure(3), plot (data(:,1), data(:,5), 'go'); R
= corrcoef(data(:,1), data(:,5)); R_green =
R(1,2)
figure(4), plot (data(:,1), data(:,6), 'bo'); R
= corrcoef(data(:,1), data(:,6)); R_blue =
R(1,2)
figure(5), plot (data(:,1), data(:,10), 'ko'); R
= corrcoef(data(:,1), data(:,10));
```

```

R_L = R(1,2)
figure(6), plot (data(:,1), data(:,11), 'ro'); R
= corrcoef(data(:,1), data(:,11));
R_a = R(1,2)
figure(7), plot (data(:,1), data(:,12), 'bo'); R
= corrcoef(data(:,1), data(:,12));
R_b = R(1,2)
figure(8), plot (data(:,1), data(:,16), 'ko'); R
= corrcoef(data(:,1), data(:,16));
R_h = R(1,2)
figure(9), plot (data(:,1), data(:,17), 'ro'); R
= corrcoef(data(:,1), data(:,17));
R_s = R(1,2)
figure(10), plot (data(:,1), data(:,18), 'bo'); R
= corrcoef(data(:,1), data(:,18));
R_v = R(1,2)
figure(11), plot (data(:,1), data(:,22), 'ko'); R
= corrcoef(data(:,1), data(:,10));
R_cont = R(1,2)
figure(12), plot (data(:,1), data(:,23), 'ro'); R
= corrcoef(data(:,1), data(:,23));
R_corr = R(1,2)
figure(13), plot (data(:,1), data(:,24), 'bo'); R
= corrcoef(data(:,1), data(:,24));
R_ener = R(1,2)
figure(14), plot (data(:,1), data(:,25), 'bo'); R
= corrcoef(data(:,1), data(:,25)); R_homo =
R(1,2)
figure(15), plot (data(:,1), data(:,26), 'bs'); R
= corrcoef(data(:,1), data(:,26));
R_area = R(1,2)

```