

APLICACIÓN DE LA TÉCNICA DE ANÁLISIS DE IMÁGENES EN EL ESPECTRO  
VISIBLE PARA LA OPTIMIZACIÓN Y CONTROL DEL  
TOSTADO DEL CAFÉ

JUAN CAMILO SARRIA GONZÁLEZ

Código: 20122112772

UNIVERSIDAD SURCOLOMBIANA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA ELECTRÓNICA  
NEIVA, HUILA.

2018

APLICACIÓN DE LA TÉCNICA DE ANÁLISIS DE IMÁGENES EN EL ESPECTRO  
VISIBLE PARA LA OPTIMIZACIÓN Y CONTROL DEL  
TOSTADO DEL CAFÉ

JUAN CAMILO SARRIA GONZÁLEZ

TRABAJO DE GRADO PRESENTADO PARA OPTAR POR EL TÍTULO DE  
INGENIERO ELECTRÓNICO

Director:

Ingeniero Electrónico Albeiro Cortés Cabezas Ph.D

Codirectores:

Ingeniero Electrónico Eugenio Ivorra Ph.D

Ingeniero Agrónomo Joel Girón Hernández Ph.D

UNIVERSIDAD SURCOLOMBIANA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA ELECTRÓNICA  
NEIVA, HUILA.

2018

Nota de aceptación:

Aprobado por el Comité de Grado en cumplimiento de los requisitos exigidos por la Universidad Surcolombiana para optar al título de Ingeniero Electrónico.

José de Jesús Salgado Patrón

---

Jurado

Diego Fernando Sendoya Losada

---

Jurado

Mis más sinceros agradecimientos para todas las personas que han estado en este camino y de alguna manera aportaron partes importantes que me permiten hoy, entregar este proyecto desarrollado, especialmente a mi familia por ser siempre incondicional y en especial a mi director Joel Girón Hernández, sin el cual no hubiera podido llevar a cabo la realización del mismo.

## TABLA DE CONTENIDO

	<b>Pág.</b>
<b>GLOSARIO</b>	<b>5</b>
<b>RESUMEN</b>	<b>9</b>
<b>ABSTRACT</b>	<b>10</b>
<b>INTRODUCCIÓN</b>	<b>11</b>
<b>1. MARCO TEÓRICO</b>	<b>13</b>
1.1 EL TOSTADO DEL CAFÉ	13
1.2 VISIÓN ARTIFICIAL	15
1.3 LENGUAJE DE PROGRAMACIÓN	15
1.4 PROCESAMIENTO DE LA IMAGEN	16
1.5 SEGMENTACIÓN (PROCESAMIENTO DE IMÁGENES)	16
1.6 PROCESAMIENTO DE DATOS	16
1.6.1 ANOVA (ANalysis Of VAriance)	17
1.6.2 Análisis de Componentes Principales (PCA)	18
1.6.3 Mínimos Cuadrados Parciales (PLS – Partial Least Squares)	19
1.7 DETERMINACIONES ANÁLITICAS	20
1.7.1 Cromatografía	20
1.7.2 HPLC.	20
1.8.	ANTECEDENTES
20	
<b>2. OBJETIVOS</b>	<b>21</b>

2.1.	OBJETIVO GENERAL	21
2.2.	OBJETIVOS ESPECÍFICOS	22
<b>3.</b>	<b>MATERIALES Y METODOLOGÍA</b>	<b>22</b>
3.1	preparación de las muestras	23
3.2	ADAPTACIÓN de los sensores respectivos	24
3.3	CAPTURA de Imagen	25
3.4	DETERMINACIONES analíticas	26
3.4.1	Determinación de ACG y cafeína.	26
3.4.2	Determinación de pH y acidez	27
3.4.3	Determinación física de las coordenadas de color L*a*b	27
3.5	DESARROLLO del algoritmo para el análisis de imagen	27
3.5.1	Lectura dinámica del directorio de imágenes	28
3.5.2	Segmentación de las capturas para eliminar del fondo de la imagen	29
3.5.3	Obtención de la información de cada imagen	30
3.5.4	Filtrado, extracción de características y obtención de datos	31
3.6	ANÁLISIS estadístico:	31
<b>4.</b>	<b>RESULTADOS Y DISCUSIÓN</b>	<b>32</b>
4.1	IMÁGENES adquiridas	32
4.2	DETERMINACIONES analíticas:	33
4.3	ANÁLISIS estadístico:	35
4.3.1.	Análisis de Componentes Principales (PCA).	35
4.3.2.	Regresión de Mínimos Cuadrados Parciales (PLS)	36
<b>5.</b>	<b>CONCLUSIONES</b>	<b>37</b>
<b>6.</b>	<b>RECOMENDACIONES</b>	<b>38</b>

## INDICE DE TABLAS

	Pág.
<b>Tabla 1.</b> Media más la desviaciones estándar de las coordenadas de color L*b*a* para cada tiempo de tosti3n.....	33
<b>Tabla 2.</b> pH, Acidez, Ácido clorogénico y cafeína con sus respectivas desviaciones estándar durante el proceso de tosti3n del café.....	34
<b>Tabla 3.</b> Análisis de varianza en los componentes principales.....	35
<b>Tabla 4.</b> Variaci3n total en las variables dependientes (Y) e independientes (X). y el % de predicci3n del modelo estadístico conforme aumenta el número de componentes.....	36
<b>Tabla 5.</b> Resultados estadísticos de los modelos de PLS para la optimizaci3n y control de la etapa de tostado de café. ....	37

## ÍNDICE DE FIGURAS

	Pág.
<b>Figura 1.</b> Imagen SEM de la sección transversal de un grano de café antes (a) y después (b) de tostar. (fuente: Mondelez Internacional). ....	14
<b>Figura 2.</b> Diagrama esquemático de la metodología empleada. ....	23
<b>Figura 3.</b> Sistema de captura de imagen. ....	25
<b>Figura 4.</b> Laboratorio de imagen UPV de la universidad politécnica de Valencia, España. ....	26
<b>Figura 5.</b> Diagrama de flujo para desarrollo del algoritmo. ....	25
<b>Figura 6.</b> (a) Imagen Real. (b) Máscara ( <i>Imagen segmentada dinámicamente</i> ). ....	29
<b>Figura 7.</b> Imagen real con el fondo sustraído(a). Aplicación del algoritmo de watersheed (b). Puntos de información obtenidos de cada grano de café (c) ....	29
<b>Figura 8.</b> Curva de temperatura en la que salieron las muestras de la tostadora con la curva gráfica de tosti3n en los granos de café para los diferentes tiempos de tosti3n. ....	32
<b>Figura 9.</b> Dispersi3n componentes principales. ....	36



## RESUMEN

La creciente competitividad durante los últimos años en los diversos sectores industriales obliga a las empresas a mejorar la eficiencia y calidad de sus procesos para garantizar que sus productos satisfagan las expectativas de sus clientes (Prats, 2005). La industria del café no es ajena a estos avances y también busca técnicas de control en la transformación del grano. Dentro del procesamiento del café, el tostado, es una etapa importante que ayuda a definir la calidad final del producto; en esta fase se potencian los aromas y sabores que se verán reflejados en la evaluación de la taza.

Así, en el presente trabajo se utilizaron las técnicas de análisis de imagen (espectro visible, coordenadas de color ( $L^*a^*b^*$ )) y técnicas de cromatografía (HPLC), con el fin de obtener un algoritmo encargado de predecir el tiempo de tostado ideal para granos de café. El trabajo se realizó utilizando muestras de café pertenecientes a la región Huilense en Colombia. Las muestras se analizaron mediante cromatografía y análisis de imagen siguiendo metodologías establecidas para cada una de las técnicas. Con los datos obtenidos se realizaron análisis de componentes principales (PCA) para observar si los métodos discriminaban las muestras y el nivel de discriminación, además se empleó el método estadístico de regresión de mínimos cuadrados (PLS) para correlacionar los datos obtenidos en el laboratorio con los adquiridos de las imágenes, los resultados preliminares se muestran satisfactorios.

**Palabras clave:** tostado del café, análisis de conglomerados, espectro visible.

## ABSTRACT

The increasing competitiveness of the various industrial sectors in recent years compels companies to improve the efficiency and quality of their processes to ensure that their products meet the expectations of their customers (Prats, 2005). The coffee industry is not alien to these advances and also seeks control techniques in grain processing. Within the processing of coffee, roasting is an important step that helps define the final quality of the product; in this phase the aromas and flavors that will be reflected in the evaluation of the cup are enhanced.

In this work, the techniques of image analysis (visible spectrum, color coordinates ( $L^*a^*b^*$ )) and chromatography techniques (HPLC) were used, in order to obtain an algorithm to predict the ideal roasting time for coffee beans. The work was carried out using samples of coffee belonging to Huila-Colombia. The samples were analyzed by chromatography and image analysis following established methodologies for each of the techniques. With the data obtained, principal component analysis (PCA) was performed to observe if the methods discriminated the samples and the level of discrimination. In addition, the statistical least squares regression (PLS) method was used to correlate the data obtained in the laboratory with the acquired results of the images, the preliminary results are satisfactory.

**Keywords:** Coffee roasting, conglomerate analysis, visible spectrum

## INTRODUCCIÓN

El Huila es un productor importante de cafés especiales de alta calidad, pero aún se conservan procedimientos artesanales en la transformación del grano; por lo que es necesario buscar técnicas que impliquen bajos costos, y que permitan optimizar y controlar el procesamiento del café, ya que esto generará incrementos en la eficiencia de la producción. Se han desarrollado diferentes proyectos para mejorar la etapa del tostado, tales como la automatización de las tostadoras (Álvarez, 2014), el desarrollo de controladores en las variables del tostado (Acero, 2007), además de estudios en análisis de imagen en el espectro visible (Hernández et al, 2008), en la región del espectro infrarroja (Alessandrini et al, 2008), y en la región hiperespectral (Nansen et al, 2016).

Actualmente las tostadoras de café disponibles en el mercado están diseñadas para llegar a una temperatura que se mantiene constante, y es el catador quien establece el tiempo necesario para cumplir con un tostado óptimo, el cual se encuentra en un intervalo que permite desarrollar las características deseables en el producto. Los granos sometidos al proceso que no logran entrar en el intervalo específico para cada tipo de café, presentarán resultados que deterioran la calidad de la taza, por lo tanto, mediante el desarrollo de un sistema de medida se busca obtener una solución económica y precisa que ayude al catador a determinar el intervalo de tiempo adecuado.

Gracias a la introducción de sensores multispectrales en algunos sectores industriales se ha propiciado que las técnicas de análisis de imagen mejoren diferentes procesos, sobre todo en aquellos en los que se establecen comparaciones con un patrón bien definido, de carácter más bien determinista, a partir de extracción de características de la imagen registrada (Song et al, 1995).

El abaratamiento en la adquisición de imágenes en color en los procesos industriales ha ampliado y mejorado la posibilidad de controlar los mismos a partir de la inspección visual automatizada, lo que da mayor viabilidad a estas técnicas desde un punto de vista económico. El análisis de imagen en el espectro visual, utiliza imágenes en color que caracterizan proceso en condiciones normales de operación y se puede crear un modelo que acumula la mayor cantidad de información relevante posible y, a partir del mismo, se determina si nuevas

imágenes se adecuan o no al modelo (Prats, 2005). La versatilidad de la técnica de análisis de imagen hace que sea aplicable a la etapa de tostado de café debido a que permite cuantificar los cambios que se generan en los granos a medida que se generan diferenciales en temperatura y tiempo.

Diferentes estudios se han realizado para la optimización en el proceso de tuestión del café con ayuda del análisis de imagen, muchos de estos trabajos se fundamentan en propiedades que no son observables a simple vista si no en el rango infrarrojo del espectro (NIR) (Alessandrini et al, 2008. Santos et al, 2016a. Santos et al, 2016b. De Luca et al, 2016), lo que hace que se necesiten equipos de un costo elevado. La presente propuesta está enfocada en trabajar en el espectro visible; se intentará correlacionar los parámetros fisicoquímicos (Análisis de coordenadas ( $L^*a^*b^*$ ), ácido clorogénico, Cafeína, pH, Acidez) del grano los cuales varían en el proceso de tuestión con la información obtenida a través de las imágenes coordenadas de color (RGB) y características físicas en los granos. Estos parámetros fisicoquímicos son medidos con equipos de laboratorio (Colorímetro ( $L^*a^*b^*$ ), HPLC (ácido clorogénico, cafeína), Sartre-5000 (pH y acidez)).

## **1. MARCO TEÓRICO**

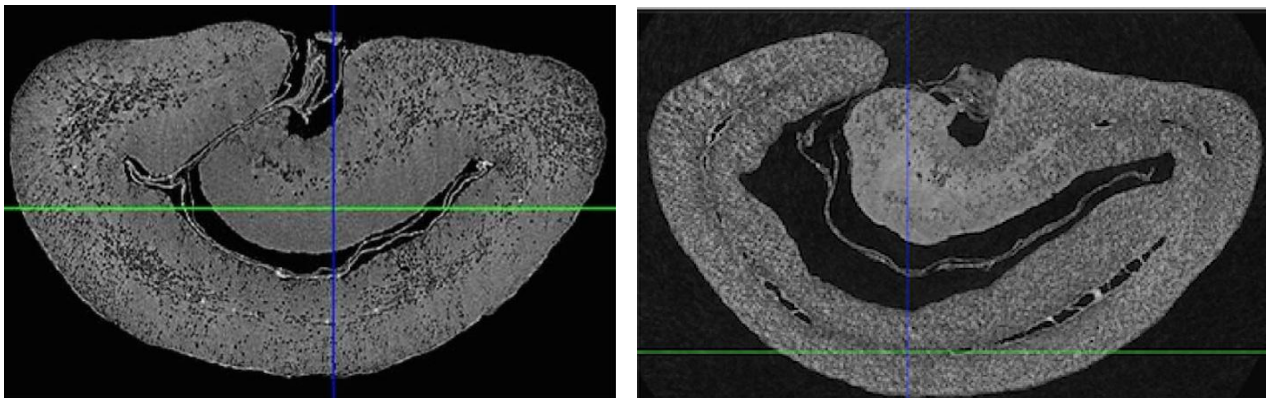
La automatización de la cadena productiva del café se ve reflejada en la calidad del producto final, lo que permite obtener mayor eficiencia en la producción; específicamente en el tostado del grano, anteriormente se mencionó como la automatización de la maquinaria y el desarrollo de controladores de las variables, permiten mejorar la operación de los factores que intervienen en el tostado.

### **1.1 EL TOSTADO DEL CAFÉ**

Cada grano y cada cosecha de café tienen atributos que se desarrollan durante el beneficio y que, una vez realizada la etapa de tostado, permiten la obtención de óptimos resultados en taza. En el tostado de los granos, por el aumento de la temperatura se incrementan diferentes compuestos volátiles resultantes de la dinámica del proceso. Por ejemplo, a mayor concentración de furfural se generan sabores suaves y acaramelados, los cuales se pueden obtener incluso a niveles de tostado muy cortos. Por el contrario, la acidez decrece gradualmente. Por lo anteriormente expuesto, se sabe que el grado de tueste tiene una influencia decisiva sobre la liberación y concentración de los agentes aromáticos (Schenker et al, 2002). El conocimiento de los procesos químicos que influyen en el sabor y aroma del café demanda el desarrollo e implementación de tecnologías para asegurar una excelente calidad, la cual se pueda reproducir.

Después de la trilla del café, los granos se someten al proceso de tostado, para esto, son comúnmente utilizadas las tostadoras de tambor rotatorio y las de lecho fluidizado (Alessandrini et al, 2011). En las tostadoras de tambor, los granos de café se colocan en un cilindro giratorio unido a una entrada de calor, el cual se transfiere a los granos por conducción del grano con las paredes del tambor y por contacto grano a grano. En las tostadoras del lecho fluidizado, se sopla una corriente vertical de aire caliente desde la base de la cámara, la cual pasa entre los granos de café, esto hace que el grano se suspenda en el aire, en este tipo de tuestión, la convección es el mecanismo utilizado para transferir el calor.

El proceso de tostado del café se divide en tres fases: secado, desarrollo de atributos deseables (color, sabor) y enfriamiento. La fase de secado suele ser la más larga del proceso, en esta, el contenido del agua en la estructura del grano se evapora y desciende la humedad del 12 al 2% (Fadai et al, 2017). La temperatura del grano aumenta rápidamente y se produce un incremento de su volumen. Después de la fase de secado se da inicio al desarrollo de sabor y el color en el grano. Una vez que la temperatura del café alcanza los 170°C aproximadamente, se da inicio a las reacciones de Maillard. Estas reacciones generan el color, el sabor y los aromas típicos de los granos de tostados (Ho et al, 2013). Adicionalmente, se genera dióxido de carbono dentro del grano, lo que produce un hinchamiento. En algún momento crítico en esta etapa, se puede oír un sonido distintivo conocido como el "primer crack", posteriormente el color del grano sigue oscureciendo y sus aromas siguen aumentando, al oírse un "segundo crack" en alguno de los granos de café, se debe suspender el suministro de calor. Para el café tostado industrialmente, el proceso en mención termina entre el primer y el segundo crack. Antes de primer crack, los aromas y sabores del grano no se han desarrollado completamente y después de segundo crack, se quema y se pierden los aromas y sabores deseados. Una vez obtenido el tostado adecuado, los granos se transfieren a la fase de enfriamiento (Rovira, 2015). Por lo general, se agrega agua a la cámara para detener rápidamente el proceso de tueste. Sin embargo, este procedimiento puede ser remplazado aplicando una corriente de aire frío. La **(Figura 1)**, muestra una imagen de barrido microscópico para los cambios en la estructura porosa del grano durante el tueste (Fadai et al, 2017). Se observa que el grano trillado presenta una estructura de poros pequeños, mientras que en el grano tostado los poros son más grandes y la mayoría de estos están llenos de gas.



(a)

(b)

**Figura 1.** Imagen SEM de la sección transversal de un grano de café antes (a) y después (b) de tostar. (fuente: Mondelez Internacional).

Como se mencionó en los apartados anteriores, el uso de sensores espectrales ha permitido el desarrollo de múltiples industrias como la farmacéutica, de materiales

y agrícola. Muchos de los conocimientos previos son aplicables en los procesos de transformación del café y en el caso específico de la etapa del tostado, es posible relacionar la información adquirida en la imagen con los cambios que ocurren en los parámetros fisicoquímicos del grano. Esto se logra mediante el desarrollo de un algoritmo, que permite correlacionar los cambios en la imagen con sus correspondientes cambios en el proceso, para poder predecir el tiempo y la temperatura a la cual se deberá exponer el grano para el tipo de café que se desee obtener, dependiendo de sus características.

## **1.2 VISIÓN ARTIFICIAL**

La visión artificial o visión por computador es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser procesadas por un computador. Tal y como los humanos usan sus ojos y cerebros para comprender el mundo que los rodea, la visión por computador trata de producir el mismo efecto para que las computadoras puedan percibir y comprender una imagen o secuencia de imágenes y actuar según convenga en una determinada situación. Esta comprensión se consigue gracias a distintos campos como la geometría, matemáticas, la estadística, la física y otras disciplinas. La adquisición de los datos se consigue por varios medios, como secuencias de imágenes, vistas desde varias cámaras de video o datos multidimensionales desde un escáner médico.

## **1.3 LENGUAJE DE PROGRAMACIÓN**

Según la Real Academia Española, un lenguaje de programación informático es un conjunto de signos y reglas que permite la comunicación con un ordenador.

El fin de los lenguajes de programación es dar a entender (impartir comandos) a la máquina, generalmente una computadora, los pasos y/o procedimientos para ejecutar tareas específicas. Es la forma cómo el programador se comunica con la máquina en un mismo idioma. Como todo lenguaje, tiene reglas y conceptos que permiten dar orden al código y/o tareas que se vayan a realizar: Sintaxis, Gramática y Semántica. Cada lenguaje tiene sus propias reglas. Las cuales deben ser entendidas y acatadas por el programador para lograr la comunicación entre la máquina y él. Los lenguajes de programación pueden clasificarse en distintas familias, una de la más importante es basada en el modelo de computación en

donde aparece a los lenguajes imperativos y su principal clase: los lenguajes orientados a objetos.

#### **1.4 PROCESAMIENTO DE LA IMAGEN**

El procesamiento de la imagen consiste en analizar y procesar la información adquirida a través de los sensores. Se utiliza un software que convierte la imagen obtenida en el espectro visible a valores numéricos, se almacena la información a modo de vectores bidimensionales (matrices), en el que cada elemento de la matriz corresponde a un píxel, cada píxel contiene información que se utilizará para realizar la correlación en los cambios ocurridos en el proceso de tostión del café. El procesamiento de la imagen comprende la adquisición de la imagen, el almacenamiento, lectura y procesado de los datos. En el procesado de datos se encuentran las etapas de segmentación de la imagen, que consiste en seleccionar el área de interés (ROI); la definición de los bordes de los objetos que componen la imagen; la selección de un objeto de interés para determinar los componentes de color (RGB). La información obtenida se compara con las coordenadas de color ( $L^*a^*b^*$ ) obtenidas en la misma muestra a través de un colorímetro digital para relacionar las coordenadas RGB (sensor CCD) con  $L^*a^*b^*$  (colorímetro).

#### **1.5 SEGMENTACIÓN (PROCESAMIENTO DE IMÁGENES)**

La segmentación en el campo de la visión artificial es el proceso de dividir una imagen digital en varias partes (grupos de píxeles) u objetos. El objetivo de la segmentación es simplificar y/o cambiar la representación de una imagen en otra más significativa y más fácil de analizar. La segmentación se usa tanto para localizar objetos como para encontrar los límites de estos dentro de una imagen. Más precisamente, la segmentación de la imagen es el proceso de asignación de una etiqueta a cada píxel de la imagen, de forma tal que los píxeles que compartan la misma etiqueta también tendrán ciertas características visuales similares.

#### **1.6 PROCESAMIENTO DE DATOS**



El volumen de datos resultantes en las imágenes de cada muestra es una característica de las técnicas de análisis de imagen, por tanto, se utilizaron métodos estadísticos de análisis multivariante para manejar la información. A continuación, se describen los métodos utilizados: Análisis de Componentes principales (PCA) y Regresión por Mínimos Cuadrados Parciales (PLS). También se presenta los análisis de varianza (ANOVA) realizados a los datos obtenidos en las pruebas fisicoquímicas.

**1.6.1 ANOVA (ANalysis Of VAriance).** La tabla ANOVA descompone la varianza de la variable dependiente en dos componentes: un componente entre-grupos y un componente dentro-de-grupos.

El **valor-P** es una probabilidad que mide la evidencia en contra de la hipótesis nula. Las probabilidades más bajas proporcionan una evidencia más fuerte en contra de la hipótesis nula. Cuando el (valor-P < 0.05) existe una diferencia estadísticamente significativa entre la media de las variables entre un nivel de tiempo y otro con un nivel del 95.0% de confianza, en las tablas se agregó un nivel de significancia estadístico ( **$\alpha$** ): (n.s. = no significativo); (\* = P<0.05); (\*\* = P<0.01); (\*\*\*) = P<0,001).

Por otro lado la **razón-F** es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos y también da un indicativo de la diferencia estadística entre las medias de las muestras. Cuando la (razón-F >>1) significa que existe una diferencia estadística significativa entre los valores medios de un tiempo de tostión a otro.

Con el método de **diferencia mínima significativa (LSD)** de Fisher se realizó la prueba de múltiples rangos para determinar cuáles medias son significativamente diferentes de otras. No existen diferencias estadísticamente significativas entre aquellos niveles que compartan una misma letra.

**1.6.2 Análisis de Componentes Principales (PCA).** El análisis de componentes principales (PCA) es una técnica de análisis multivariante que consiste en encontrar un nuevo sistema de referencia que permita reducir la dimensión de nuestro espacio de variables independientes dado que cuando el número de variables independientes es mayor de tres resulta difícil representar las muestras en función de las medidas (Jolliffe, I.T., 1986). El PCA puede emplearse como una herramienta de clasificación no supervisada o como paso previo a otros análisis, ya que permite tener una visión general del conjunto de los datos recogidos. El nuevo sistema de referencia debe cumplir las siguientes condiciones:

- Los ejes de este nuevo sistema de referencia deben ser ortogonales entre sí, de forma que la información aportada por cada uno de ellos sea diferente y exclusiva.
- La proyección de los datos sobre el nuevo sistema de referencia debe contener la información completa del conjunto de los datos, concentrando en los primeros ejes la información más relevante.
- A los ejes de este nuevo sistema de referencia se les denomina componentes principales.

Cuando se lleva a cabo un análisis de componentes principales se asegura que la primera componente principal contenga el máximo de la varianza de los datos. A continuación, se elige la segunda componente principal de forma que contenga el máximo de la varianza restante (es decir, la varianza que no haya sido explicada por la primera componente principal) y que sea ortogonal a la primera componente. Y de igual modo se procede con el resto de componentes principales (Brezmes, J., 2001). Una vez se ha construido el nuevo sistema de referencia, una representación de las muestras proyectadas sobre las dos primeras componentes principales, daría, a priori, la máxima información que se puede obtener sobre las medidas trabajando en dos dimensiones. Por lo general, el PCA se utiliza para poder visualizar matrices de datos que contienen un gran número de variables independientes en dos o tres dimensiones. A partir de estas representaciones se pueden desarrollar hipótesis para agrupar y clasificar las muestras que pueden ser estudiadas por métodos estadísticos apropiados (Girón, 2013).

**1.6.3 Mínimos Cuadrados Parciales (PLS – Partial Least Squares).** El procedimiento de mínimos cuadrados está diseñado para construir un modelo estadístico que relaciona múltiples variables independientes  $X$  (imágenes) con múltiples variables dependientes  $Y$  (físicoquímicas) (Geladi et al, 1986). El procedimiento es mucho más útil cuando hay muchos factores y el objetivo principal es predecir las variables respuesta, este método presenta el siguiente modelo estadístico:

$$Y = X\beta + E$$

Donde  $Y$  es una matriz  $n$  por  $m$  que contiene los  $n$  valores estandarizados de las  $m$  variables dependientes,  $X$  es una matriz  $n$  por  $p$  que contiene los valores estandarizados de las  $p$  variables predictoras,  $\beta$  es una matriz  $p$  por  $m$  de parámetros del modelo, y  $E$  es una matriz  $n$  por  $m$  de errores. A diferencia de la regresión múltiple, el número de observaciones  $n$  puede ser menor que el número de variables independientes  $p$ .

Sin embargo, en vez de estimar  $\beta$  directamente, se extraen primero  $c$  componentes. Luego los coeficientes se calculan a partir del producto de dos matrices:

$$\beta = WQ$$

Donde  $W$  es una matriz  $p$  por  $c$  de pesos que transforman a  $X$  en una matriz  $T$  de valores de los factores de acuerdo con

$$T = XW$$

y  $Q$  es una matriz de coeficientes de regresión (cargas) que expresan la dependencia entre  $Y$  y los valores de los factores:

$$Y = TQ + E$$

La matriz de variables independientes también puede representarse en términos de  $P$  una matriz  $c$  por  $p$  de cargas de factores como

$$X = TP + F$$

Donde  $F$  es una matriz  $n$  por  $p$  de desviaciones.

Parte de la tarea al realizar un análisis de mínimos cuadrados parciales (PLS) es determinar el número apropiado de componentes  $c$ . Si se establece  $c$  demasiado bajo o demasiado alto, el modelo puede no dar buenas predicciones para las futuras observaciones (Helland, 2001).

## 1.7 DETERMINACIONES ANÁLITICAS

**1.7.1 Cromatografía.** La cromatografía es un método físico de separación para la caracterización de mezclas complejas, la cual tiene aplicación en todas las ramas de la ciencia; es un conjunto de técnicas basadas en el principio de retención selectiva, cuyo objetivo es separar los distintos componentes de una mezcla, permitiendo identificar y determinar las cantidades de dichos componentes. Diferencias sutiles en el coeficiente de partición de los compuestos dan como resultado una retención diferencial sobre la fase estacionaria y, por tanto, una separación efectiva en función de los tiempos de retención de cada componente de la mezcla.

**1.7.2 HPLC.** La cromatografía líquida de alta eficacia o high performance liquid chromatography (HPLC), es un tipo de cromatografía en columna utilizada frecuentemente en bioquímica y química analítica. El HPLC es una técnica utilizada para separar los componentes de una mezcla basándose en diferentes tipos de interacciones químicas entre las sustancias analizadas y la columna cromatográfica.

## 1.8 ANTECEDENTES

Desde finales de los años 80, la introducción de cámaras en color en diferentes procesos industriales, ha derivado en la aparición de nuevas técnicas basadas en la comprensión de la información subyacente en una imagen mediante técnicas de proyección sobre estructuras latentes, en lo que se conoce como Análisis Multivariable de Imágenes (*Multivariate Image Analysis, MIA*) (Geladi et al, 1989).

Estas técnicas se muestran bastante potentes en situaciones donde existen multitud de canales de color asociados a cada imagen registrada, es decir, en su aplicación a imágenes recogidas por medio de cámaras multiespectrales, en las que cada canal puede recoger información de un número más reducido de fenómenos y, por lo tanto, poder relacionar unos pocos fenómenos de interés con unos pocos canales que los caractericen (Prats, 2005).

El análisis de imagen muestra ser una técnica innovadora y precisa, que permite desarrollar avances tecnológicos y científicos, ejemplos prácticos de su aplicación en el control de procesos alimenticios como: análisis del jamón curado (Girón et al, 2014), análisis en el tostado de café (Hernández, Heyd y Trystram, 2008), análisis en horneado de pan (Verdú et al, 2016), análisis de la vida útil en salmón (Ivorra et al, 2013) demuestran la efectividad en la predicción del comportamiento del producto a diferentes condiciones.

## **2. OBJETIVOS**

### **2.1. OBJETIVO GENERAL**

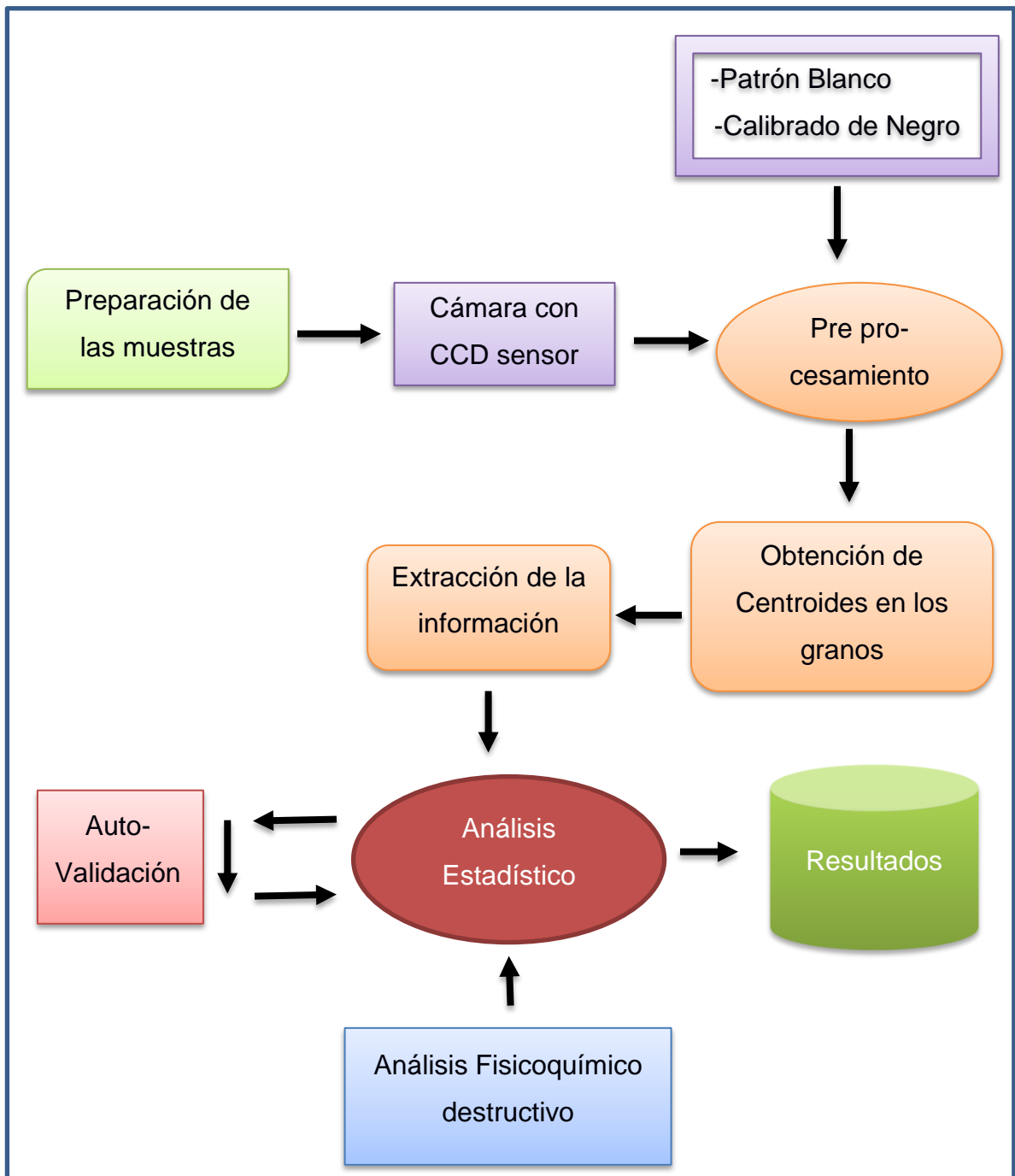
- Aplicar análisis de imagen en el espectro visible para la optimización y el control del tostado de café.

## 2.2. OBJETIVOS ESPECÍFICOS

- Adaptar sensores CCD para captura de imagen en el espectro visible.
- Desarrollar algoritmo para procesar las imágenes adquiridas.
- Desarrollar una metodología que permita correlacionar la información extraída en las imágenes con las características fisicoquímicas y sensoriales del café, obtenidas a partir de equipos de laboratorio especializados.

## 3. MATERIALES Y METODOLOGÍA

La **(Figura 2)** refleja el diagrama esquemático de la metodología empleada para predecir el tostado del café desde la adquisición hasta el resultado:



**Figura 2.** Diagrama esquemático de la metodología empleada.

### 3.1 PREPARACIÓN DE LAS MUESTRAS

Se recolectaron 20 kg de café (*Coffea arabica*) variedad Castillo y Colombia, las muestras fueron procesadas en la Universidad Surcolombiana, antes de realizar el tostado se realizó la selección de la almendra (SCAA, 2017).

### 3.2 ADAPTACIÓN DE LOS SENSORES RESPECTIVOS

Para el montaje del sistema de captura se seleccionó como sistema de medida una cámara digital Canon T3, la cual entrega imágenes en formato RAW (sin compresión de información) y también permite manipular los parámetros de captura (balance de blancos, tiempo de exposición y distancia focal), se desarrolló un soporte para el iluminar las muestras de café, que consistió en dos bombillas LED dicroicas con iluminación difusa de 620 Lumen, con una potencia de 7.5W y una temperatura de color de 6500°K (Lumek, Alemania), después de adaptar la iluminación a una base metálica se adecuó también la cámara (**Figura 3**).

Se prosiguió con el sistema de toma de datos, para esto se realizó la interconexión entre el sensor y el equipo de cómputo, en el que se almacenarán las imágenes adquiridas; para la toma de imágenes se calibró la apertura del lente en f/8, para tener una distancia focal de 25 mm en el lente y una profundidad de campo de más o menos 5 cm, para ajustar el tiempo de obturación (tiempo necesario para que el lente de la cámara se cierre); se ajustó conforme a la iluminación que se tiene; se observó que el rango en el cual las imágenes quedan con mejor resolución es entre 1/80 a 1/100 de seg., teniendo el equipo listo para iniciar la toma de datos, se utilizó un software en su versión libre (Capture One PRO 10 desarrollado por "Phase One") para tomar las imágenes de manera remota y así no descalibrar la precisión del sistema.





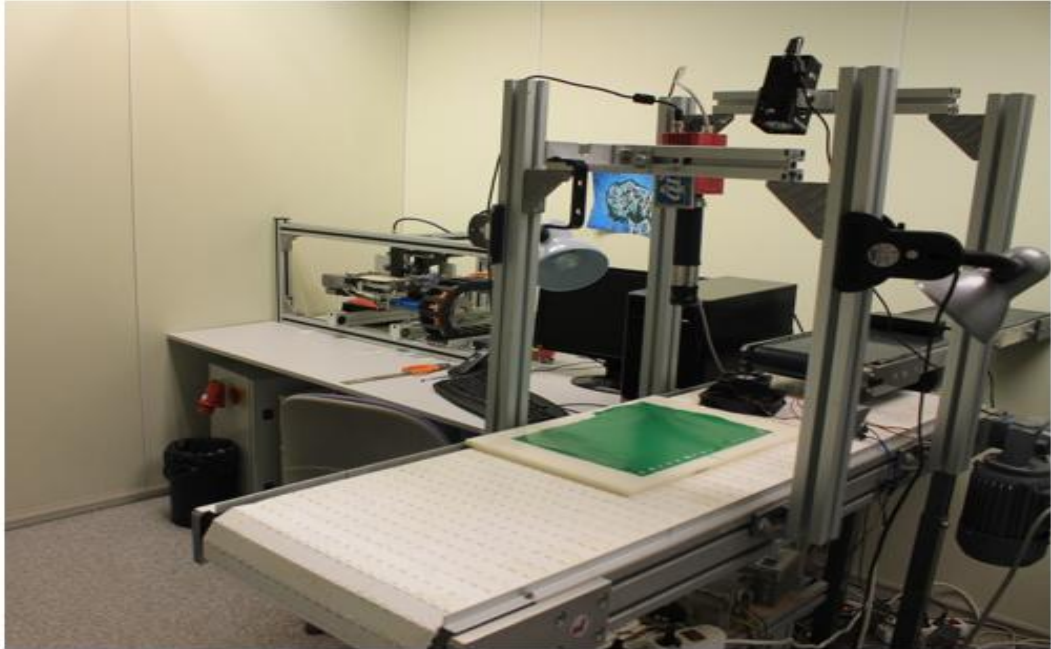
**Figura 3.** Sistema de captura de imagen.

### **3.3 CAPTURA DE IMAGEN**

Para iniciar la captura de imágenes se dejó previamente encendida la iluminación hasta que está llegó a una temperatura de color estable de 5500°K.

Las imágenes se obtuvieron a partir de 150 gramos de café, capacidad máxima para la tostadora DIBER M-D 150, se realizaron seis capturas por cada una de las muestras tostadas en diversos intervalos de tiempo (45 y 60 segundos), para un total de 256 imágenes, 24 imágenes para cada tiempo de tuestión. Las imágenes obtenidas se almacenaron en una base de datos para su posterior análisis.

Se realizaron pruebas experimentales en el laboratorio de imagen en el instituto de sistemas y automática industrial (isa) de la Universidad Politécnica de Valencia, España (**Figura 4**); en muestras de café de la misma variedad (Castilla húmedo, Colombia semiseco, Colombia húmedo, Castilla semiseco) analizadas en el laboratorio de Agroindustria USCO. Se utilizó el mismo sensor CCD e iluminación empleada en los experimentos anteriores.



**Figura 4.** Laboratorio de imagen UPV de la universidad politécnica de Valencia, España.

### **3.4 DETERMINACIONES ANALÍTICAS**

**3.4.1 Determinación de ACG y cafeína.** La determinación de ácido clorogénico y cafeína se realizó mediante cromatografía de líquidos utilizando un cromatógrafo LaChrom Elite (Hitachi Ltd., Tokyo, Japan) equipado con un auto-muestreador (modelo L-2200) y un detector de UV (modelo L-2400). La separación se realizó durante 24 minutos con una columna C18 (150 × 4.6 mm i.d., 5 μm) (Scharlab, Barcelona, Spain). La fase móvil utilizada consistió en una mezcla de MeOH y agua acidificada. Se realizaron lecturas para el ácido clorogénico a 324nm y para la cafeína a 280nm, en ambos casos se inyectaron 20 μL con un flujo de 1.2 mL min<sup>-1</sup>.

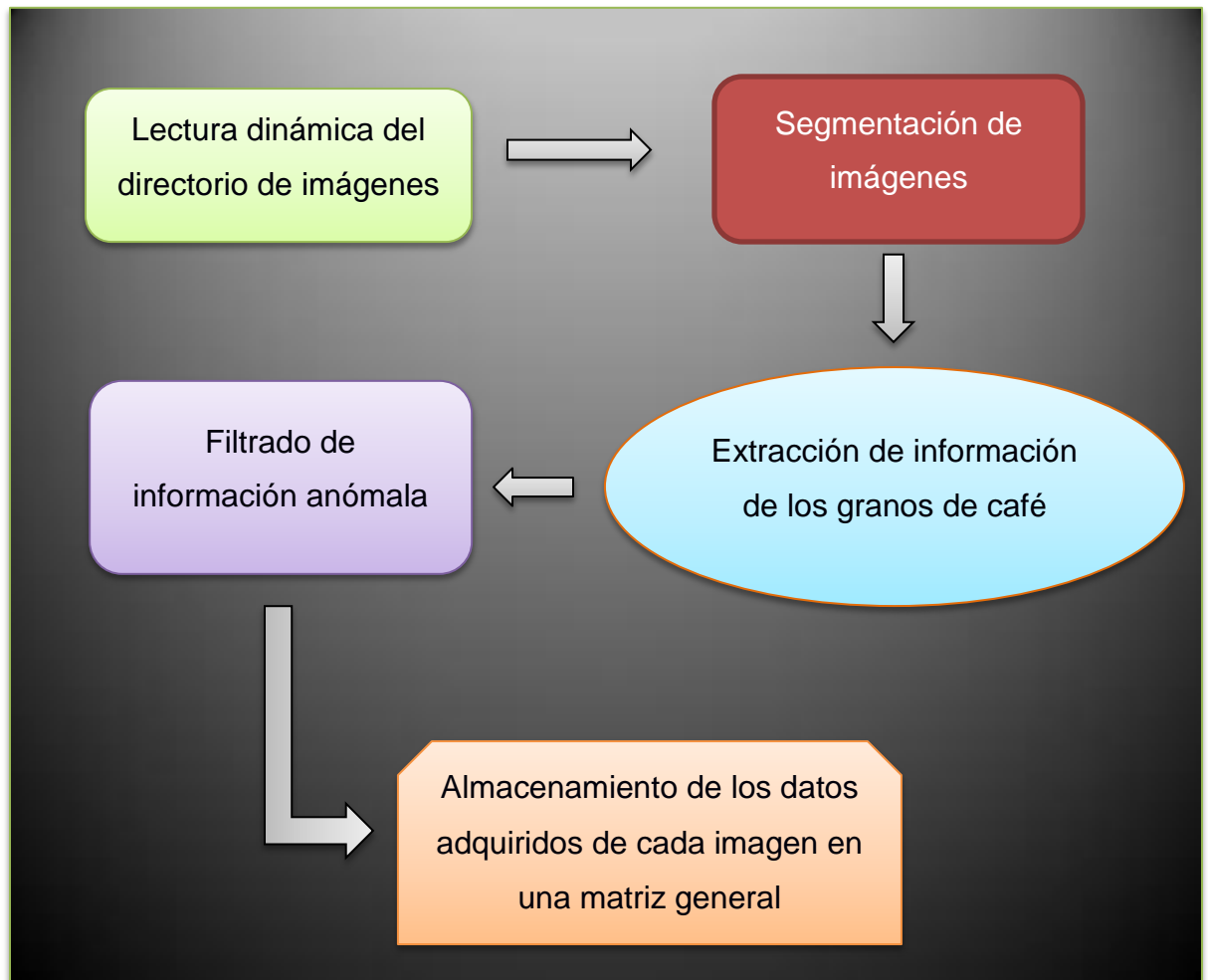
**3.4.2 Determinación de pH y acidez.** La determinación del pH y acidez en cada muestra de café se realizó con el equipo Starter ST-5000 de Chauz, USA. Para la medición de pH se tomaron 2.5 gramos de cada muestra y se molieron, después se sumergieron en 50mL de agua destilada a 80°C y se dejó reposar hasta que estuviera en temperatura ambiente. Para determinar la acidez se tomaron 5 gramos de café molido y se agregaron a 100mL de agua destilada a 90°C, se agitó constantemente por 3 minutos y después se filtró, el filtrado fue realizado en un balón aforado de 100mL y se completó el volumen a 100mL con agua destilada, se tomaron 50 mL para realizar la acidez titulable con hidróxido de sodio (NaOH) a 0.1normal, se utilizó como indicador el método de potenciómetro con un pH a 6.5; se calculó la acidez de las muestras empleando la ecuación del gasto de NaOH. El volumen de hidróxido de Sodio (VNaOH) se multiplicó por el equivalente del ácido predominante en la muestra (eqAcido) y se dividió por el volumen de la muestra de café (Vmuestra), los resultados se expresan en mg de Ácido clorogénico (ACG) por gramos de café.

$$\frac{VNaOH * eqAcido}{Vmuestra} = \frac{mg ACG}{g café}$$

**3.4.3 Determinación física de las coordenadas de color L\*a\*b.** El color se midió en coordenadas de color L\*a\*b\* con un colorímetro Konica Minolta, “Chroma meter” CR-410, Japón. Para determinar las coordenadas de color L\*a\*b se utilizó un volumen de 40  $cm^3$  de cada muestra de café, el instrumento se calibró con patrón blanco (87 .316 .3231).

### 3.5 DESARROLLO DEL ALGORITMO PARA EL ANÁLISIS DE IMAGEN

Se desarrolló un algoritmo de procesamiento de esta las imágenes adquiridas utilizando en el software MATLAB 8.0 (2012b) (The Mathworks, Natick, Massachusetts, USA), el diagrama de flujo se refleja en la **(Figura 5)**.

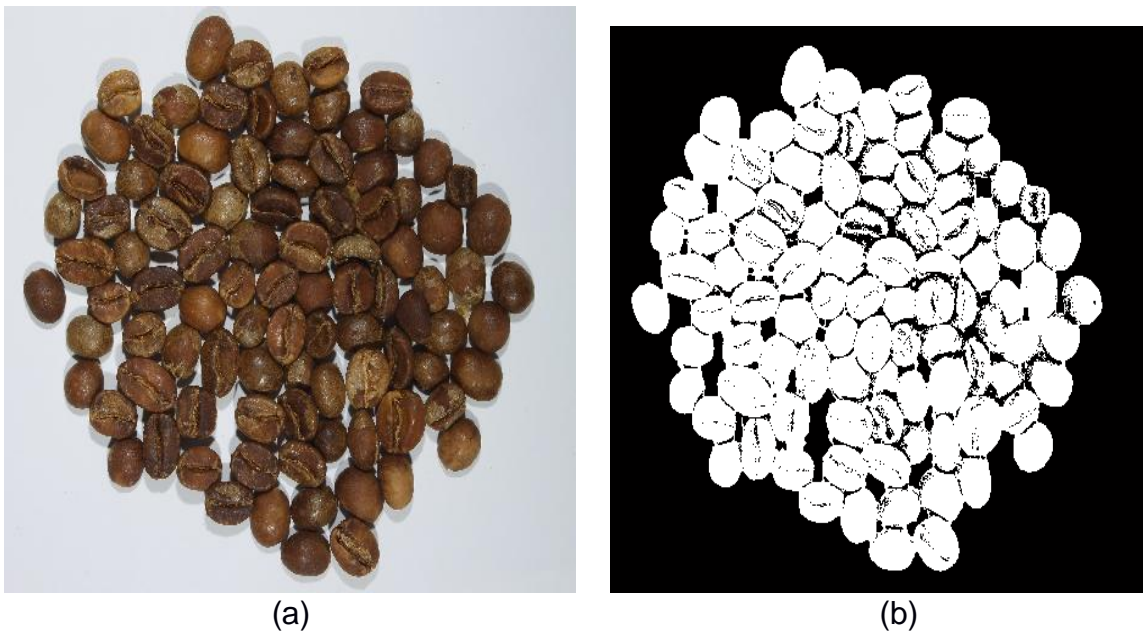


**Figura 5.** Diagrama de flujo para desarrollo del algoritmo.

**3.5.1 Lectura dinámica del directorio de imágenes.** Se leyeron las imágenes en el directorio (carpeta) según la variedad de café, luego se creó un vector (“*vectornombres*”) que almacenó los datos principales de cada imagen leída como (“nombre”, “fecha”, “Tipo de archivo”). Posteriormente se creó un bucle que dependió de un contador, este se definió por el número de imágenes leídas en el directorio, este se encargó de buscar, leer y analizar de manera independiente cada una de las imágenes almacenadas en el “*vectornombres*”.

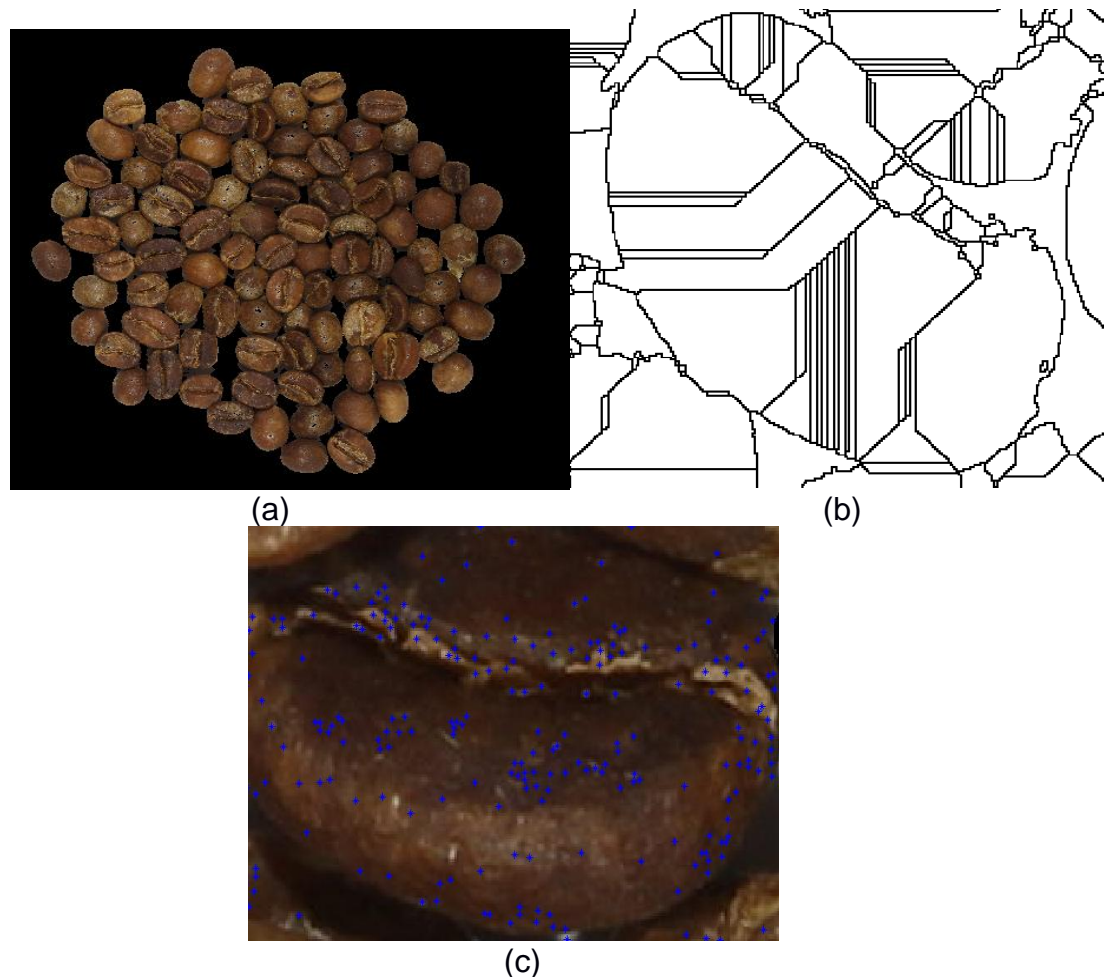
### 3.5.2 Segmentación de las capturas para eliminar del fondo de la imagen.

Para la segmentación de la imagen se transformó la información de la matriz de coordenadas RGB a escala de grises, se aplicó una función que permitió la binarización dinámica, luego se filtraron las áreas pequeñas (ruido), además se creó un filtro de disco para cerrar los bordes en la imagen. Después se rellenó todo el interior del perímetro binarizado que conforman los granos de café, de esta manera se diferenció correctamente los objetos en la imagen del fondo (máscara de la imagen). Este procedimiento se realizó a cada una de las imágenes de manera dinámica. La **Figura 6 (a)** muestra la imagen real de los granos de café con un tiempo de tuestión de 6min, en la **Figura 6 (b)** se muestra la máscara binaria obtenida a partir de la imagen real.



**Figura 6.** (a) Imagen Real. (b) Máscara (*Imagen segmentada dinámicamente*).

**3.5.3 Obtención de la información de cada imagen.** Después de obtener la máscara, esta se superpuso a la imagen real y así se eliminó el fondo de la imagen (segmentación) **Figura 7 (a)**. Una vez segmentada la imagen se pasó a una escala de grises, después se empleó el método de Otsu para obtener un umbral dinámico de binarización de la imagen, se determinó la distancia euclidiana en cada imagen y se igualó a infinito, esto, para poder utilizar el algoritmo de Watersheed. El algoritmo en mención permitió crear un modelo “topográfico” en 3D de la imagen, en donde cada nivel de gris representa una altura en el relieve topográfico, así se diferenció cada grano de café en la imagen, seguidamente se dividieron en múltiples regiones más pequeñas **Figura 7 (b)**, a partir de éstas se obtuvieron los centroides o punto central del perímetro de la **Figura 7 (c)**. Concretamente, se obtienen los valores RGB y HSV medios obtenidos de una ventana entorno al centroíde de cada grano (Beucher & Meyer 1993), a mayor cantidad de centroides obtenidos por grano, el algoritmo presenta mayor eficiencia debido a que la información en cada imagen será más confiable.



**Figura 7.** (a) Imagen real con el fondo sustraído. (b) Aplicación del algoritmo de watersheed. (c)Puntos de información obtenidos de cada grano de café.

**3.5.4 Filtrado, extracción de características y obtención de datos.** Una vez analizadas las imágenes, los datos se almacenan en matrices numéricas. Las matrices de datos obtenidas en ocasiones contienen información redundante o errónea, para corregir este problema se emplean bucles condicionales que se encargan de filtrar y eliminar de cada canal (variable) los centroides (puntos) que no cumplen con las condiciones del bucle respectivo. Se aplicaron operaciones estadísticas a cada una de las matrices de datos obtenidos en las ventanas de los centroides, de esta matriz se obtiene un solo vector de datos que contenga toda la información de interés (tanto física como de color) por cada imagen, en total **44 valores diferentes**, 30 valores que caracterizaron los cambios en el color : Media R, Media G, Media B, Máximo R, Máximo G, Máximo B, Mínimo R, Mínimo G, Mínimo B, Mediana R, Mediana G, Mediana B, Desviación R, Desviación G, Desviación B, Media H, Media S, Media V, Máximo H, Máximo S, Máximo V, Mínimo H, Mínimo S, Mínimo V, Mediana H, Mediana S, Mediana V, Desviación H, Desviación S, Desviación V; y 14 valores que caracterizaron los cambios físicos: área, perímetro, distancia euclidiana, excentricidad, entre otras. Finalmente los vectores obtenidos por cada imagen se almacenan en una sola matriz para su posterior análisis estadístico.

### 3.6 ANÁLISIS ESTADÍSTICO

A los datos fisicoquímicos se les realizó el análisis de varianza (ANOVA simple), con un nivel de confianza del 95%. Con el método de diferencia mínima significativa (LSD) de Fisher, se realizó la prueba de múltiples rangos para discriminar las medias que son significativamente diferentes de otras. El procedimiento fue realizado con el paquete estadístico Statgraphics centurión XVI. (Manugistics Inc., Rockville, MD, EE.UU.).

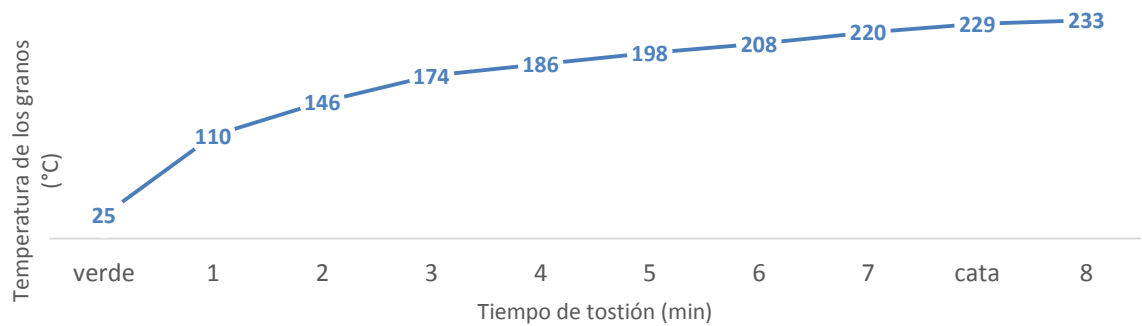
Se utilizó el método de análisis de componentes principales (PCA), para reducir la varianza los datos obtenidos en las imágenes, lo que permite comprimir un grupo voluminoso de datos con la máxima variabilidad a un número menor de variables (componentes principales). El análisis fue realizado con el paquete estadístico Statgraphics centurión XVI. (Manugistics Inc., Rockville, MD, EE.UU.).

El análisis de Mínimos Cuadrados Parciales (PLS) se utilizó para evaluar la correlación entre las matrices de datos los fisicoquímicos y datos de imagen. Se utilizó el procedimiento de validación cruzada del modelo para predecir observaciones excluidas del ajuste del modelo. Se empleó la mitad de las muestras para entrenamiento y la otra mitad para predicción. El análisis fue realizado con el paquete estadístico Statgraphics centurión XVI. (Manugistics Inc., Rockville, MD, EE.UU.).

## 4. RESULTADOS Y DISCUSIÓN

### 4.1 IMÁGENES ADQUIRIDAS

La **Figura 8** muestra el cambio en el color presentado por las muestras de café a los diferentes tiempos de tuestión, se anexo el respectivo intervalo de tiempo al que pertenece cada muestra y la temperatura a la que salieron los granos de la tostadora, esta curva inició con los granos en verde y prosiguió en intervalos de tuestión de 1 minuto. Una vez la temperatura del café alcanza los 170°C aproximadamente, se da inicio a las reacciones de caramelización y Maillard, estas reacciones generan el color, el sabor y los aromas típicos de los granos de tostados (Ho et al, 2013). Se observa que a medida que se exponen por más tiempo los granos al proceso de tuestión se presenta un pardeamiento más notorio y su volumen aumenta gradualmente.



**Figura 8.** Curva gráfica del proceso de tuestión y temperatura de los granos.



## 4.2 DETERMINACIONES ANALÍTICAS

La **Tabla 1** muestra la media más la desviación estándar de las coordenadas de color  $L^*a^*b^*$ , se presentaron diferencias significativas ( $p < 0.05$ ) para todas las coordenadas. Se observó que los valores de la luminancia de los granos ( $L^*$ ) disminuye a medida que aumenta la temperatura. La coordenada  $a^*$  presentó la mayor varianza de los datos al clasificar las muestras en nueve grupos diferentes. Los valores de  $b^*$  presentaron un máximo aproximadamente entre los tres y cuatro minutos tiempo en el que los granos de café llegan a una temperatura de 180 hasta 240 °C, los azúcares se combinan con los aminoácidos, péptidos y proteínas siguiendo el proceso de caramelización conocido como la reacción de Maillard (Budryn G. et al, 2006), lo que se refleja en el cambio que se da en el color verde-amarillo al pardo-naranja, posteriormente los valores decaen al volverse más oscuros los granos.

**Tabla 1.** Media más la desviaciones estándar de las coordenadas de color  $L^*a^*b^*$  para cada tiempo de tostión.

Nivel de significancia estadístico ( $\alpha$ ): (n.s. = no significativo); (\* =  $p < 0.05$ ); (\*\* =  $p < 0.01$ ); (\*\*\*) =  $p < 0,001$ ).

Letras iguales dentro de la misma columna indican la pertenencia a grupos homogéneos; letras distintas indican la existencia de diferencias significativas.

Tiempo (min)	$L^*$	$a^*$	$b^*$
0	43.81 ± 0.26 <sup>g</sup>	1.68 ± 0.27 <sup>ab</sup>	11.78 ± 0.54 <sup>d</sup>
1	44.75 ± 0.40 <sup>g</sup>	1.67 ± 0.22 <sup>a</sup>	12.48 ± 0.26 <sup>d</sup>
2	46.49 ± 0.56 <sup>h</sup>	2.07 ± 0.33 <sup>b</sup>	14.52 ± 0.50 <sup>e</sup>
3	48.45 ± 0.56 <sup>i</sup>	2.92 ± 0.22 <sup>c</sup>	20.46 ± 0.28 <sup>g</sup>
4	43.76 ± 0.56 <sup>f</sup>	10.65 ± 0.27 <sup>h</sup>	23.65 ± 0.49 <sup>h</sup>
5	36.11 ± 0.2 <sup>e</sup>	11.55 ± 0.12 <sup>i</sup>	17.16 ± 0.38 <sup>f</sup>

<b>6</b>	31.96 ± 0.61 <sup>d</sup>	8.75 ± 0.10 <sup>g</sup>	12.45 ± 0.26 <sup>d</sup>
<b>7</b>	29.4 ± 0.30 <sup>c</sup>	7.9 ± 0.10 <sup>f</sup>	10.47 ± 0.14 <sup>c</sup>
<b>7.4</b>	28.27 ± 0.35 <sup>b</sup>	7.54 ± 0.42 <sup>de</sup>	9.37 ± 0.59 <sup>b</sup>
<b>8</b>	28.66 ± 0.55 <sup>b</sup>	7.53 ± 0.08 <sup>ef</sup>	9.52 ± 0.19 <sup>b</sup>
<b>9</b>	27.62 ± 0.55 <sup>a</sup>	7.28 ± 0.25 <sup>d</sup>	8.39 ± 0.39 <sup>a</sup>
<b>α</b>	***	***	***

En la **Tabla 2** se observan los resultados de las determinaciones analíticas de pH, acidez, ACG y cafeína. Se presenta un descenso en los valores de pH a medida que transcurre el proceso de tuestión. La disminución en el pH fue causada por la formación de los ácidos fórmico, acético, glicólico y láctico en los granos de café a medida que estos se tuestan (Ginz et al., 2000). Se observó un aumento rápido de la acidez titulable ya que los granos de café verde se tuestan hasta el final del primer crack (tostado medio). Este cambio en la acidez titulable puede deberse a la formación de ácidos alifáticos totales a un nivel máximo (Clarke y Macrae 1988). Se observa que la concentración del ácido clorogénico (ACG) decrece cerca del 50% ya que a medida que es expuesto al proceso de tuestión la transferencia de calor genera la degradación de estos ácidos, dando como resultado la formación de productos de caramelización y condensación (Ginz et al. 2000). La cafeína también decrece a medida que transcurre la tuestión, pero esto no presenta pérdidas significativas (Ramalakshmi & Raghavan 1999).

Tiempo (min)	pH	Acidez	ACG	Cafeína
--------------	----	--------	-----	---------

Por otro lado, la prueba de múltiples rangos aplicada en la Tabla 2 dispersa en mayor medida los datos obtenidos de la acidez titulable de las muestras, al agruparlos en 8 grupos distintos, lo que comprueba lo dicho por (Voilley et al. 1981; Brollo et al. 2008) de que la acidez titulable es un indicador más confiable para correlacionar la acidez del café que el valor de pH. Las otras variables presentan la misma cantidad de grupos de dispersión, 7 respectivamente.

**Tabla 2.** pH, Acidez, Ácido clorogénico y cafeína con sus respectivas desviaciones estándar durante el proceso de tuestión del café.

Nivel de significancia estadístico (α): (n.s. = no significativo); (\* = p<0.05); (\*\* = p<0.01); (\*\*\*) = p<0,001).

Letras iguales dentro de la misma columna indican la pertenencia a grupos homogéneos; letras distintas indican la existencia de diferencias significativas.

0	6.07 ± 0.04 <sup>a</sup>	3.29 ± 0.72 <sup>a</sup>	49.93 ± 0.95 <sup>a</sup>	33.21 ± 0.49 <sup>ab</sup>
1	6.04 ± 0.02 <sup>a</sup>	2.86 ± 0.72 <sup>ab</sup>	32.7 ± 0.34 <sup>bc</sup>	26.91 ± 0.79 <sup>d</sup>
2	6.11 ± 0.01 <sup>a</sup>	2.15 ± 0.72 <sup>ab</sup>	34.32 ± 0.77 <sup>b</sup>	33.29 ± 1.17 <sup>a</sup>
3	6.16 ± 0.00 <sup>a</sup>	2.86 ± 1.43 <sup>b</sup>	26.74 ± 0.08 <sup>bc</sup>	30.46 ± 0.89 <sup>c</sup>
4	5.94 ± 0.00 <sup>b</sup>	5.01 ± 0.72 <sup>c</sup>	31.67 ± 0.20 <sup>c</sup>	30.69 ± 0.01 <sup>bc</sup>
5	5.76 ± 0.02 <sup>c</sup>	7.87 ± 1.43 <sup>d</sup>	27.38 ± 0.40 <sup>d</sup>	27.81 ± 1.05 <sup>d</sup>
6	5.29 ± 0.01 <sup>d</sup>	15.03 ± 0.72 <sup>e</sup>	21.5 ± 0.11 <sup>e</sup>	23.67 ± 0.52 <sup>e</sup>
7	4.91 ± 0.01 <sup>e</sup>	31.49 ± 0.00 <sup>f</sup>	16.61 ± 0.03 <sup>f</sup>	20.15 ± 0.86 <sup>f</sup>
7.4	4.94 ± 0.03 <sup>ef</sup>	36.50 ± 0.00 <sup>g</sup>	9.65 ± 0.28 <sup>g</sup>	16.06 ± 1.38 <sup>g</sup>
8	4.89 ± 0.03 <sup>fg</sup>	38.65 ± 0.72 <sup>g</sup>	10.37 ± 0.40 <sup>g</sup>	16.39 ± 1.60 <sup>g</sup>
9	5.01 ± 0.03 <sup>g</sup>	37.93 ± 0.00 <sup>h</sup>	8.52 ± 0.24 <sup>g</sup>	15.68 ± 1.14 <sup>g</sup>
α	***	***	***	***

### 4.3 ANÁLISIS ESTADÍSTICO

**4.3.1. Análisis de Componentes Principales (PCA).** La **Tabla 3** muestra el análisis de varianza de los componentes principales de los datos obtenidos de las imágenes adquiridas, con tres componentes se explica el 95.97% del comportamiento del modelo, el incremento de componentes por encima de este número no mejora la varianza explicada.

**Tabla 3.** Análisis de varianza en los componentes principales.

Componente	Eigenvalor	% de Varianza	% Acumulado
1	25.639	73.255	73.255
2	7.338	20.967	94.222
3	0.611	1.747	95.969

La **Figura 9** presenta el diagrama de dispersión de los componentes principales obtenidos a partir de los datos extraídos en las imágenes, se observó que las agrupaciones de datos tienen una relación no lineal. Los datos de los grupos de verde, uno, dos y tres minutos de tostión se agruparon al tener características físicas similares como color y volumen. A los cuatro minutos de tostión, cuando la temperatura interior del grano es de 170°C las muestras generan un grupo entre el color verde y el color pardo y a medida que se continúa con el proceso de tostión se van oscureciendo gradualmente hasta llegar al minuto siete cuando la temperatura en los granos llega a los 220°C, a partir de este tiempo, las muestras exhiben características más homogéneas tanto en color como en forma, lo que se refleja en la agrupación de los datos.

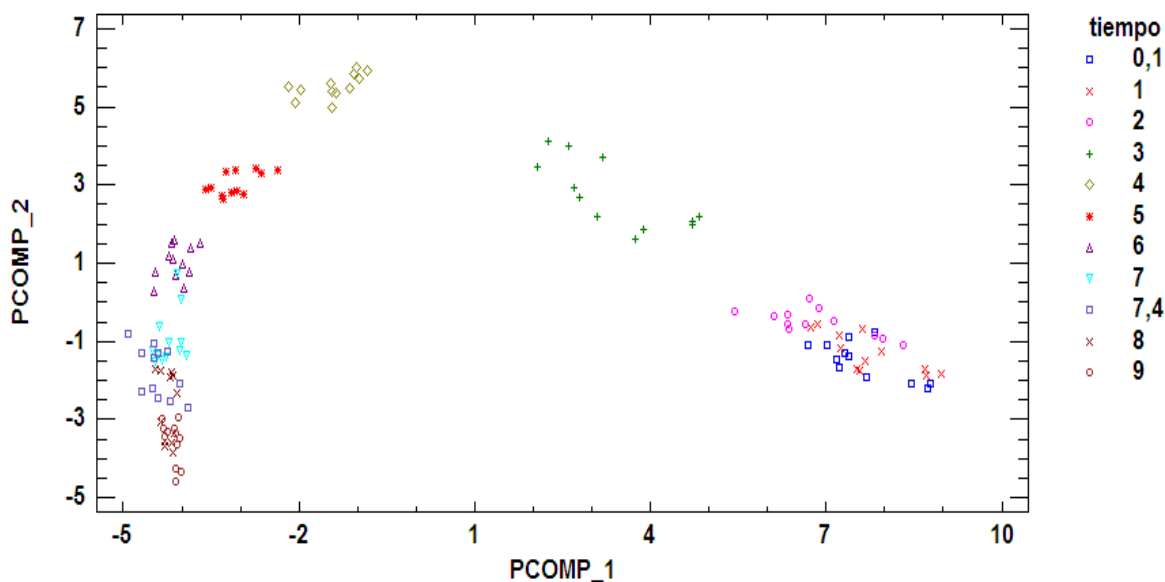


Figura 9. Dispersión componentes principales.

**4.3.2. (Regresión de Mínimos Cuadrados Parciales PLS).** La **Tabla 4** muestra los valores estadísticos de correlación del modelo de variables independientes (x) imagen y las variables dependientes (y) fisicoquímicas. Adicionalmente se muestra la relación de predicción (R-cuadrada) del modelo. Se observó que los resultados más favorables para ajustar modelos de correlación de las variables independientes del sistema ( $L^*$ ,  $a^*$ ,  $b^*$ , pH, acidez, ACG, cafeína) es con las variables de imagen que presentaron una varianza significativa ( $P < 0.05$ ), esto lo cumplen 35 de las 44 variables extraídas en las imágenes, el modelo de correlación demuestra un comportamiento confiable reflejado en el valor de la R-cuadrada de predicción (90.26%).

**Tabla 4.** Variación total en las variables dependientes (Y) e independientes (X). y el % de predicción del modelo estadístico conforme aumenta el número de componentes.

Componente	% Variación en X	% Acumulado de X	% Variación en Y	% Acumulado de Y	Predicción Promedio R-Cuadrada
1	72.21	72.21	65.03	65.03	68.78
2	22.00	94.22	24.13	89.16	89.16
3	1.515	95.73	1.20	90.37	90.26

La **Tabla 5** muestra los valores estadísticos par cada uno de los modelos establecidos para cada una de las variables fisicoquímicas utilizadas en la construcción del modelo predictivo. Los resultados presentan un buen grado de confiabilidad para la predicción debido a que los valores de r-cuadrada de validación cruzada están por encima del 85% para todas las variables, igualmente la robustez presenta valores por encima del 80%. Se presenta el Error cuadrático medio (RMSE), porcentaje del Error cuadrático medio (RMSE%), porcentaje del

Error cuadrático medio de crossvalidación (RMSECV%) y el número de componentes utilizados en el análisis (N°LV).

**Tabla 5.** Resultados estadísticos de los modelos de PLS para la optimización y control de la etapa de tostado de café.

	L*	a*	b*	pH	Acidez	ACG	Cafeína
<b>R<sup>2</sup>Cal</b>	0.978	0.968	0.941	0.954	0.957	0.852	0.800
<b>R<sup>2</sup>CV</b>	0.967	0.976	0.938	0.968	0.936	0.866	0.851
<b>RMSE</b>	1.017	0.469	0.913	0.091	2.45	3.802	2.288
<b>RMSE%</b>	0.027	0.072	0.067	0.016	0.119	0.099	0.091
<b>RMSECV%</b>	0.035	0.085	0.069	0.018	0.156	0.182	0.132
<b>N°LV</b>	8	7	8	3	3	8	8

## 5. CONCLUSIONES

- Este trabajo de investigación se centró en demostrar la efectividad de emplear el análisis de imágenes en el espectro visible con el fin de optimizar y controlar la etapa de tuestión del café.
- La aplicación de técnicas de visión artificial permitió discriminar de forma automática y objetiva las muestras de café en función del tiempo de tostado.
- Las variedades de café utilizadas en el ensayo no fueron un factor de discriminación.
- Se demostró que las imágenes adquiridas en el espectro visible presentan suficiente información para construir modelos de regresión (PLS) robustos ( $R^2_{Cal}=0.921$ ) y con una predicción confiable ( $R^2_{CV} = 0.902$ ), para los atributos fisicoquímicos relevantes en el proceso de tuestión (Coordenadas de color  $L^*a^*b$ , pH, Acidez, ACG (ácido clorogénico) y Cafeína).
- El resultado obtenido hace que esta técnica de visión artificial sea prometedora para resolver el problema planteado, ya que entregó suficiente información para crear un modelo que pudo correlacionarse con siete variables fisicoquímicas.

## 6. RECOMENDACIONES

- Aunque se demostró que las imágenes en el espectro visible contienen suficiente información para crear modelos de varianza y predicción confiables del proceso de tostión de café, sería ideal utilizar cámaras con un rango espectral mucho mayor, ya que el sensor que se utilizó no contiene longitudes de onda superiores a 720 nm, y en estas longitudes (NIR) puede encontrarse información adicional del proceso de tostión de café.
- Se debe tener en cuenta que trabajar con esos equipos especializados implicaría un costo mucho más elevado que el utilizado para la realización de este proyecto, pero sería ideal para obtener un modelo más robusto el cual contenga información de un espectro mucho más amplio.
- Futuros estudios con otras variedades son necesarios con la finalidad de generar una herramienta de clasificación aplicable a cualquier variedad para fines de control de calidad, permitiendo una determinación directa, reproducible y objetiva de la tostión en muestras de granos de café desconocidos.
- También se está considerando la posibilidad de extender este estudio a diferentes productos alimenticios con un proceso de preparación análogo al del café.

## BIBLIOGRAFÍA

Acero. W. F. 2007. Diseño e implementación de un controlador automático para la tostación del café en pequeña y media producción. Proyecto de grado. Universidad de la Salle.

Alessandrini Laura. Romani Santina. Pinnavia Giangaetano. Dalla Rosa Marco. Near Infrared spectroscopy: An analytical tool to predict coffee roasting degree. University of Bologna. J. Analytica chimica acta 625 (2008) 95-102.

Álvarez. C.A. 2014. Automatización del proceso de tostación de café en lecho fluidizo. Proyecto de grado. Universidad tecnológica de Pereira.

Brezmes, J. 2001. Tesis Doctoral: Diseño de una nariz electrónica para la determinación no destructiva del grado de maduración de la fruta. Universitat Politècnica de Catalunya.

BROLLO, G. CAPPUCCI, R. & NAVARINI, L. (2008). Acidity in Coffee: Bridging the Gap Between Chemistry and Psychophysics. In 22nd Colloquium: Coffee Aroma & Flavour Chemistry Campinas, São Paulo, Brazil: Association for Science and Information on Coffee.

Budryn G., Nebesny E., Podsedek A., Żyżelewicz D., Materska M., Jankowski S., Janda B. Effect of different extraction methods on the recovery of chlorogenic acids, caffeine and Maillard reaction products in coffee beans. *European Food Research and Technology*. 228 (2009), Number 6, 913-922.

CAFÉ DE COLOMBIA. *TOSTACIÓN*. Citado el 5 de diciembre del 2016. disponible en:

[www.cafedecolombia.com/particulares/es/sobre\\_el\\_cafe/el\\_cafe/industrializacion/tostacion /](http://www.cafedecolombia.com/particulares/es/sobre_el_cafe/el_cafe/industrializacion/tostacion/)

CAFÉ LA NACIONAL. *PROCESO DE PRODUCCIÓN DEL CAFÉ (2015)*. Citado el 15 de agosto del 2016. disponible en: [www.cafelanacional.com/nosotros/proceso-de-produccion-del-cafe/](http://www.cafelanacional.com/nosotros/proceso-de-produccion-del-cafe/)

Calvini Rosalba. Ulrici Alessandro. Amigo Jose Manuel. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *J. Chemometrics and Intelligent Laboratory System* 146 (2015) 503-511.

CLARKE, R. J. & MACRAE, R. (1988). *Coffee: Physiology*. London: Elsevier Applied Science.

Contreras Calderón J.. Mejía Díaz D.. Matínez Castaño M.. Bedoya Ramírez D.. López Rojas N.. Gómez Narváez F.. Medina Pineda Y.. Vega Castro O..Evaluation



of antioxidant capacity in coffee marketed in Colombia: Relationship with the extent of non-enzymatic browning. Universidad de Antioquia. J. Food Chemistry 209 (2016) 162-170.

DIARIO EL IMPULSO. SUPERVISIÓN AUTOMATIZADA DE TOSTADORAS DE CAFÉ. POR INGENIERO DE UFT(2013). citado el 20 de agosto de 2016 disponible en: [www.elimpulso.com/noticias/actualidad/supervision-automatizada-de-tostadoras-de-cafe-por-ingeniero-uft](http://www.elimpulso.com/noticias/actualidad/supervision-automatizada-de-tostadoras-de-cafe-por-ingeniero-uft)

Esbensen K., Geladi P. 1989. *Strategy of Multivariate Image Analysis. Chemometrics and Intelligent Laboratory Systems* 7. 67-86.

Fabbri A., C. Cevoli, L. Alessandrini, S. Romani. Numerical modeling of heat and mass transfer during coffee roasting process. J. Food Eng. 105 (2011) 264–269.

Fadai N. T., Melrose J., Please C. P., Schulman A., Van Gorder R. A., A heat of mass transfer study of coffee bean roasting. Journal of Heat and Mass Transfer 104 (2017) 787-799.

Geladi P., Kowalski B. R. Partial Least Squares Regression: A Tutorial. Anal. Chim. Acta, Vol. 185, (1986) 1-17.

Gil L. 2007. Diseño de lenguas electrónicas potenciométricas construidas con electrodos metálicos y con tecnología de capa gruesa para análisis de medios complejos. Tesis Doctoral. Universidad politécnica de Valencia.

GINZ, M. BALZER, H. H. BRADBURY, A. G. W. & MAIER, H. G. (2000). Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. European Food Research and Technology 211(6), 404-410.

Girón Joel. Gil-Sánchez Luís. García-Breijo Eduardo. Pagán M<sup>a</sup>. Jesús. Barat José M.. Grau Raúl. Development of potentiometric equipment for the identification of altered dry-cured hams: A preliminary study. Meat Science. Volume 106. August 2015. Pages 1-5.

Girón L. J. 2014. Uso de sensores electrónicos para la optimización y control del jamón curado. Tesis Doctoral. Universidad Politécnica de Valencia.

Girón. L. J. 2009. Aplicación de las técnicas no destructivas de potenciometría (lengua electrónica) y espectrofotometría de infrarrojo cercano (SW-NIR) para la clasificación de cerdos de raza ibérica en función del tipo de alimentación. Tesis de máster. Universidad Politécnica de Valencia.

Gloes A. N.. Vietri A.. Wieland F.. Smrke S.. Schönbächler B.. Sánchez López J. A.. Petrozzi S.. Bongers S.. Kozirowski T.. Yeretzián C. Evidence of different flavour formation dynamics by roasting coffee from different origins: On-line analysis with PTR-ToF-MS. International Journal of Mass Spectrometry 365-366 (2014) 324-337.

Grau R.. Sánchez J. A.. Girón J.. Ivorra E.. Fuentes A.. Barat J. M. Nondestructive assessment of freshness in packaged sliced chicken breasts using SW-NIR spectroscopy. Journal of Food Research International 44 (2011) 331-337.

Hernández. J.A.. Trystam G.. Heyd B. Prediction of brightness and surface area kinetics during coffee roasting. Journal of food engineering 89 (2008) 156-163.

Ho Chi-Tang. Mussinan Cynthia. Shahidi Fereidoon. Contis Ellene Tratras. Nutrition. Functional and Sensory Properties. Publishing London of Foods Royal Society of Chemistry. 2013 - 335 páginas.

Ivorra M. E. 2016. Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroalimentario. Tesis Doctoral. Universidad Politécnica de Valencia.

Ivorra Eugenio. Girón Joel. Sánchez Antonio J. Verdú Samuel. Barat Jose M. Grau Raúl. Detection of expired vacuum-packed smoked salmon based on PLS-DA method using hyperspectral images. *Journal of Food Engineering* 117 (2013) 3 342-349.

Jeska Skowron M.. Stanisz E.. Paz de Peña M..Relationship Between antioxidant capacity. chlorogenic acids and elemental composition of green coffee. *J. LWT-Food Science and Technology* 73 (2016) 243-250.

Jimenez Ariza H.. Correa C. E.. Diezma B.. Díaz Barcos V.. Meneses B.. Oteros R.. ESTUDIO DE TÉCNICAS PARA LA SUPERVISIÓN DE CALIDAD Y CLASIFICACIÓN DE GRANOS DE CAFÉ TOSTADO: ANÁLISIS DE NUEVAS TECNOLOGIAS. Artículo publicado en el VI congreso ibérico de AgroIngeniería. Universidad de Évora (2011).

Jolliffe, I.T. 1986. *Principal Component Analysis*. Springer, New York.

Kučera L.. Papoušek R.. Kurka O.. Barták P.. Bednár P.. Study of composition of espresso coffee prepared from various roast degrees of *coffea arabica* L. coffee beans. *J. Food Chemistry* 199 (2016) 727-735.

Nansen C. et al. 2016. *Using hyperspectral imaging to characterize consistency of coffee brands and their respective roasting classes*. *Journal of food Engineering* 190 (2016) 34-39.

Morais de Oliveira E.. Samid Leme D.. Groenner B. H.. Pereira Rodarte M.. Fonseca Alvarenga R. G.. A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering* 171 (2016) 22-27.

Prats J. M. 2005. Control estadístico de procesos mediante análisis multivariante de imágenes. Tesis Doctoral. Universidad Politécnica de Valencia.

RAMALAKSHMI, K. & RAGHAVAN, B. (1999). Caffeine in Coffee: Its Removal. Why and How? *Critical Reviews in Food Science and Nutrition* 39(5), 441-456.

Ribeiro J. S.. Ferreira M. M. C.. Salva T. J. G.. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffees beverages using near infrared spectroscopy. *J. Talanta* 83 (2011) 1352-1358.

Rovira G. J. FORUM DEL CAFÉ. UN PASEO POR LA HISTORIA DEL TUESTE DEL CAFÉ. Citado el 20 de agosto del 2016. disponible en: [www.forumdelcafe.com/pdf/F-47\\_Tueste.pdf](http://www.forumdelcafe.com/pdf/F-47_Tueste.pdf)

Rovira G. J. FORUM DEL CAFÉ. LA FÍSICA Y LA QUÍMICA EN EL TUESTE DEL CAFÉ. Citado el 5 de diciembre del 2016. disponible en: [www.forumdelcafe.com/pdf/F-41\\_Fisica\\_quimica\\_tueste.pdf](http://www.forumdelcafe.com/pdf/F-41_Fisica_quimica_tueste.pdf)

Santos J. R. Lopo M. Rangel A. Almeida J. Exploiting near infrared spectroscopy as an analytical tool for on-line monitoring of acidity during coffee roasting. *Food Control* 60 (2016) 408-415.

Santos J. R.. Viegas O.. Páscoa R. N.. Ferreira I.. Rangel A.. Almeida Lopes J. In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour. *J. Food Chemistry* 208 (2016) 103-110.

Silva De Luca. Martina de Filippis. Bucci R.. Magrì A. D.. Magrì A. L.. Marini F.. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD. NIR and chemometrics. University of Roma. Microchemical Journal 129 (2016) 348-361.

Schenker S. Heinemann C. Huber M. Pompizzi R. Perren R. Escher R. Impact of Roasting Conditions on the Formation of Aroma Compounds in Coffee Beans. Journal of Food Science 67 (2002) 60-66.

Song, K.Y.. Petrou. M. & Kittler. J. 1995. Machine Vision applications. 8: 63.

Verdú Samuel. Vásquez Francisco. Ivorra Eugenio. Sánchez Antonio J. Barat José M. Grau Raúl. Hyperspectral image control of the heat-treatment process of oat flour to model composite bread properties. Journal of Food Engineering. 192 (2017) 45-52.

Virgen Navarro L.. Herrera E.. Corona González R.. Arriola Guevara E.. Guatemala Morales G. M. Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. J. Expert System with Applications 54 (2016) 162-169.

VOILLEY, A. SAUVAGEOT, F. SIMATOS, D. & WOJCIK, G. (1981). Influence of Some Processing Conditions on the Quality of Coffee Brew. Journal of Food Processing and Preservation 5(3), 135-143.

Wang Niya. Master of Science in Food (2012). Physicochemical Changes of Coffee Beans During Roasting. The University of Guelph Science.

