



UNIVERSIDAD NACIONAL DE COLOMBIA

# Un Método Automático Para la Caracterización Morfológica Digital de Hojas

**Roger Enrique Gúzman Avendaño**

Universidad Nacional de Colombia, Universidad Nacional de Colombia

Maestría en Ingeniería de Sistemas y Computación

Bogotá, Colombia

2016



# Un Método Automático Para la Caracterización Morfológica Digital de Hojas

**Roger Enrique Guzmán Avendaño**

Tesis presentada como requisito parcial para obtener el título:  
**Maestría en Ingeniería de Sistemas y Computación**

Director:

Francisco Gómez Ph.D. (Departamento de Matemáticas, U. Nacional de Colombia)

Co-Director:

Fabio González Ph.D. (Departamento de Ingeniería de Sistemas, U. Nacional de Colombia)

Universidad Nacional de Colombia

Facultad de Ingeniería

Bogotá, Colombia

2016



## Agradecimientos

Quiero dar mis sinceros agradecimientos a todos aquellos que hicieron posible este logro.

Agradezco enormemente a mis madre Flor Mariela Avendaño, mi padre Jaime Guzmán y a mi hermana Jennifer Guzmán por su paciencia y apoyo en esta ardúa labor académica que brinda un gran impulso a mi vida profesional.

Mi mas grande agradecimiento a mi Director de Tesis Francisco Gomez por su gran apoyo, paciencia y conocimiento compartido conmigo en el campo de la investigación, en el campo de las ciencias, incentivar en mi el deseo por la solución de problemas, de igual manera por la formación recibida por el. También quiero agradecer a mi Codirector Fabio Gonzalez por los conocimientos impartidos en el aula de clase y la formación recibida en el área de ciencias de la computación.

Extiendo mi agradecimiento a mi compañera Mary Lee Berdugo quien me apoyo incondicionalmente con su orientación y su tiempo, al realizar los procesos de creación de base de datos, el apoyo en la clasificación de las muestras y el conocimiento compartido en la etapa de experimentación.



## Resumen

En esta tesis se propone un modelo computacional para la clasificación automática de las formas de hojas tropicales en términos de sus características morfológicas. El modelo de representación morfológica se basa en el concepto de descriptores elípticos de Fourier, momentos de Hu y características geométricas simples. La estrategia para la representación de características de la estructura morfológica de la hoja se apoya en métodos de clasificación. Para la evaluación de la estrategia propuesta se utilizó un nuevo sistema de categorización de formas foliares, y una nueva base de datos de muestras foliares de vegetación tropical anotadas por un experto. Los resultados sugieren que el método propuesto es capaz de caracterizar correctamente las categorías propuestas. Sin embargo, en categorías con alto grado de similitud el método presenta desempeños bajos.

**Palabras clave:** Vegetación tropical, lamina foliar, descriptores morfológicos, sistemas de clasificación visual.

# Lista de Figuras

<b>2-1.</b> Estrategia propuesta. En primer lugar, se construyó un sistema de categorización específico para muestras en vegetación tropical. Un conjunto de datos de vegetación tropical fue utilizado. Las muestras foliares fueron digitalizadas, binarizadas y el contorno fue extraído. Un conjunto de características morfológicas cuantitativas fue utilizado para la descripción. Seguidamente, un método clasificación fue empleado para la determinación de las diferentes categorías. Finalmente, el método fue evaluado en el conjunto de datos, previamente, etiquetado por el experto. . . . .	10
<b>2-2.</b> Representación del contorno. La gráfica representa la imagen original, y la reconstrucción de la forma de la hoja utilizando varios conjuntos de armónicos, tomada de [1]. . . . .	12
<b>3-1.</b> Ejemplos de muestras foliares definidas para cada categoría del sistema de clasificación de láminas foliares. . . . .	15
<b>3-2.</b> Matriz de confusión máquina de clasificación multiclase (1 - Abovada, 2 - Acorazonada, 3 - Elíptica, 4 - Lanceolada, 5 - Oblanceolada, 6 - Oblonga, 7 - Ovalada, 8 - Ovalada-Lanceolada y 9 - Redonda), basada en árboles aleatorios con 150 estimadores sobre el conjunto completo de características de forma (geométricos, elípticos de Fourier y momentos de Hu). . . . .	17



# Contenido

<b>Resumen</b>	<b>vii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Categorización de la forma foliar . . . . .	1
1.2. Morfología foliar digital . . . . .	2
1.2.1. Preprocesamiento . . . . .	3
1.2.2. Extracción de características morfológicas . . . . .	3
1.2.3. Caracterización del rasgo foliar . . . . .	7
1.3. Objetivos de la tesis . . . . .	8
1.3.1. Contribuciones . . . . .	8
1.3.2. Contribuciones científicas . . . . .	8
<b>2. Materiales y Métodos</b>	<b>9</b>
2.1. Muestras foliares . . . . .	9
2.2. Sistema de categorización de formas foliares . . . . .	10
2.3. Pre-procesamiento . . . . .	11
2.4. Caracterización de los bordes . . . . .	11
2.4.1. Descriptores geométricos simples . . . . .	11
2.4.2. Representaciones elípticas de Fourier . . . . .	11
2.4.3. Descripción del área basada en momentos . . . . .	13
2.5. Clasificación automática de las formas foliares . . . . .	13
2.6. Evaluación de la estrategia de clasificación . . . . .	14
2.7. Selección de características . . . . .	14
<b>3. Resultados</b>	<b>15</b>
3.1. Sistema de categorización de formas foliares para vegetación tropical . . . . .	15
3.2. Muestras foliares categorizadas . . . . .	16
3.3. Clasificación automática de las categorías de lámina foliar . . . . .	16
3.3.1. Selección de características . . . . .	17
<b>4. Discusión</b>	<b>20</b>
<b>5. Conclusiones</b>	<b>22</b>
5.1. Recomendaciones . . . . .	22

<b>A. Contribución Congreso Colombiano de Computación</b>	<b>23</b>
<b>B. Congreso de Innovación y Tendencias en Ingeniería</b>	<b>24</b>
<b>Bibliografía</b>	<b>25</b>

# 1. Introducción

Las plantas juegan un rol importante en el sistema de vida del planeta. Los ciclos de carbono y nutrientes están profundamente influenciados por las plantas a través del proceso de fotosíntesis realizado por las hojas. La comprensión de la morfología de las plantas y, en particular, la forma de las hojas (fisonomía) es un reto importante en los estudios de ciencias de la vida [2]. La descripción de forma de hoja es una de las alternativas más populares utilizadas para estudiar la morfología de las plantas [2]. Usualmente, estas descripciones se realizan de forma manual por un experto entrenado en la tarea de clasificación, apoyado en diccionarios botánicos o manuales [3]. Este proceso, es altamente especializado y costoso respecto a tiempo y recursos. En esta tesis se proponen un conjunto de métodos computacionales para realizar las tareas de clasificación de formas foliares de manera automática. Los métodos desarrollados se adaptan al contexto de la vegetación tropical, y permiten automatizar algunas de las tareas de caracterización morfológica realizadas por botánicos expertos.

## 1.1. Categorización de la forma foliar

Las hojas son órganos laterales, que brotan del tallo o de las ramas. Estos órganos están expuestos continuamente a variedad de condiciones ambientales, y presiones por selección natural [4, 3, 5]. Como resultado de la evolución, se han generado diferentes adaptaciones fenotípicas, que a largo plazo se ven reflejadas en cambios permanentes en la forma de las hojas. La descripción de la forma foliar es una tarea fundamental en botánica, en la medida en que permite el estudio de dichas adaptaciones.

La descripción de la forma foliar, generalmente, se realiza a través de manuales botánicos. Los manuales botánicos corresponden a sistemas de categorización visual utilizados para clasificar diferentes formas foliares [6, 3, 7]. En general, los manuales suelen estar compuestos por dos componentes: concepto botánico (referente a un identificador textual del objeto) y objeto visual (correspondiente a una imagen representativa del objeto). Los manuales botánicos poseen un rico conjunto de términos clave, diseñados específicamente, para la identificación y caracterización de diferentes especies de plantas, o para la descripción de algunos componentes específicos de las mismas, tales como, hojas, folíolos, flores, venación de las hojas y raíces, entre otros[8].

Las guías de clasificación botánica son definidas por expertos utilizando su experiencia en campo y los datos disponibles para el dominio [6, 7]. Estos sistemas de categorización juegan un papel clave en la descripción del funcionamiento de la planta, en el establecimiento de

relaciones entre diferentes especies, así como, para la determinación de la relación de las plantas con su medio ambiente [2].

En general, la tarea de categorización de objetos visuales utilizando sistemas de categorización es complicada, y requiere un alto grado de especialización debida a los detalles de la forma que por muy mínimo puede resultar en un tipo de clase diferente [9]. Por lo tanto, los programas de formación en botánica comúnmente, dedican grandes esfuerzos en la formación de expertos para la realización de estas tareas [9]. A pesar del alto nivel de formación y especialización de los expertos, diferentes fuentes biológicas pueden ser incorrectamente clasificadas, esto, como resultado del alto grado de subjetividad de la tarea de categorización de objetos biológicos.

Es importante resaltar que en general, no existe un sistema de clasificación visual universal que permita la categorización de todas las formas foliares. En otras palabras, los sistemas de clasificación visual son particulares para la descripción de un dominio biológico específico. De tal manera que, la tarea de categorización se hace aún más especializada al variar entre diferentes dominios biológicos [6, 7].

## 1.2. Morfología foliar digital

Con el fin de automatizar algunas de las tareas de descripción morfológica de la hoja, para reducir posibles sesgos introducidos por la evaluación manual de las formas foliares, en los últimos años se han propuesto diferentes herramientas digitales para soportar las tareas de adquisición, caracterización y análisis de la morfología foliar [10, 11, 12, 13]. Estas herramientas se han agrupado bajo el término de morfología foliar digital [8].

Actualmente, el desarrollo de tecnologías para caracterización morfológica foliar ha atraído la atención de diferentes disciplinas. Por ejemplo, desde la perspectiva biológica, la morfología digital de las hojas posiblemente puede reducir considerablemente los sesgos introducidos por los expertos. Adicionalmente, gracias al nivel de automatización que se puede lograr con este tipo de herramientas, se abre la posibilidad para procesar grandes volúmenes de muestras foliares [2, 14]. Respecto al área computacional, el desarrollo de las representaciones de señales que capturen la riqueza y variedad de la morfología foliar representa un problema abierto en el área de procesamiento de señales [15, 16]. Desde la perspectiva de la estadística, es necesaria la formulación de nuevos métodos para el descubrimiento de posibles relaciones significativas entre las propiedades cualitativas/cuantitativas de las hojas, y las condiciones ambientales de desarrollo, o entre las propiedades de diferentes especies [17, 18, 19].

Al respecto, recientemente se han propuesto métodos computacionales para la segmentación y partes de las hojas [20, 21, 22, 23, 24, 2], para la descripción de la forma, así como, para la caracterización estadística de los rasgos foliares [18, 17, 19], entre otros. Estos métodos se han utilizado con éxito en diferentes escenarios de estudio biológico, incluyendo, taxonomía automática [25, 26], comprensión del desarrollo de plantas [3], evaluación cuantitativa de la morfología de las hojas en relación con la información genómica [27], y descripción de

cambios morfológicos relacionados con las condiciones ambientales [28, 29], entre otros. A pesar de estos avances la comprensión de la naturaleza de la morfología foliar, la morfología foliar digital, aún requiere el abordaje de algunas problemáticas adicionales. En particular, la construcción de estrategias automáticas que permitan cerrar la “*brecha semántica*” entre la información visual foliar (es decir, la imagen asociada a la entidad biológica) y los términos clave definidos por los expertos, en los manuales botánicos. Aunque se han propuesto algunos métodos para abordar esta problemática, estos sistemas de clasificación no necesariamente se adaptan a las formas foliares particulares, como es el caso la vegetación tropical [2, 29]. Desde el punto de vista computacional la morfología digital foliar comprende diferentes etapas, entre las que se encuentran: 1) preprocesamiento, 2) extracción de características, y 3) caracterización del rasgo foliar de interés.

### 1.2.1. Preprocesamiento

Las imágenes de hojas pueden adquirirse en diferentes condiciones: ambientes naturales, semi-controlados y controlados. Los estudios en el área se han realizado principalmente en ambientes controlados y semi-controlados [2, 30]. No obstante, recientes estudios han abordado el problema de caracterización de la forma foliar en ambientes no controlados [31].

El primer paso para la caracterización de los rasgos morfológicos foliares es la segmentación de la hoja. Este proceso consiste en separar el objeto de interés (por ejemplo, la hoja entera o el patrón de venación), del fondo. El grado de dificultad de esta tarea depende de la complejidad de la adquisición, así como, de la complejidad del patrón visual, asociado al objeto de interés [31].

El problema de la segmentación de la hoja entera ha sido abordado para tareas de clasificación de plantas. En este caso, se han explorado diferentes alternativas, incluyendo, umbralización de los niveles de gris, detección de los bordes de la lámina foliar [25], y segmentación del órgano foliar utilizando contornos activos [32], entre otros. Cuando lo importante es obtener una descripción más detallada de las diferentes partes de la hoja (es decir, pecíolo, ápice, base, forma) [11], se debe realizar una segmentación adicional para las partes foliares de interés (por ejemplo, base, pecíolo y nervadura). Para esta tarea, se ha propuesto el uso de la simetría local de la traslación, que busca capturar las relaciones de paralelismo y simetría, usualmente, observadas para la hoja y el pecíolo [2]. Más recientemente, se ha propuesto fijar un conjunto de vértices sobre la representación del contorno poligonal como el ápice y las regiones basales, y establecerlos como referencia para la segmentación de las partes foliares utilizando la información de los ángulos alrededor de los mismos[32].

### 1.2.2. Extracción de características morfológicas

La caracterización cuantitativa de la forma de la hoja representa uno de los métodos más informativos, y empleados por los expertos, para representar la morfología foliar [11]. A

diferencia de la caracterización morfológica foliar utilizando sistemas de clasificación visual, la caracterización cuantitativa, provee medidas numéricas objetivas del rasgo foliar, las cuales reducen, considerablemente, los sesgos derivados de la evaluación del observador [33].

La cuantificación de los rasgos morfológicos representa un insumo fundamental para la caracterización de variaciones fenotípicas fundamentales en diferentes tareas, incluyendo, la identificación de especies biológicas [11], la determinación de las condiciones climáticas pasadas mediante el estudio de fósiles [34], y el estudio de la evolución de diferentes especies [35], entre otros.

En la literatura se han propuesto diferentes enfoques para abordar la problemática de caracterización cuantitativa de la forma biológica, y en particular, para la caracterización de la forma foliar [36]. Entre ellos se destacan tres estrategias de caracterización: descriptores geométricos, descriptores de contorno, y descriptores de área.

### **Descriptores geométricos**

La forma más simple de descripción de la forma foliar es a través de cantidades geométricas simples que se pueden calcular en toda la forma foliar [30, 36]. Algunas de las cantidades más importantes utilizadas en la literatura para esta descripción de la forma foliar, se detallan en la tabla 1-1.

Este tipo de caracterización asume modelos geométricos muy generales para explicar la forma de la hoja, incluyendo modelos rectangulares, circulares, elípticos y convexos. Esta simplificación en la definición de la forma puede limitar de forma drástica la capacidad de caracterización del método, dada la amplia variedad de formas esperadas para la descripción de la planta, y las partes de la misma. En el caso donde sea necesaria la descripción de muestras foliares con una alta variabilidad morfológica, esta limitación puede ser importante. En estos casos, deben usarse representaciones de las formas más ricas respecto a su capacidad de caracterización.

### **Descriptores de contorno basados en representaciones armónicas**

Una de las técnicas más comunes para realizar representaciones de formas de manera flexible, se basa en el uso de descriptores de Fourier [37]. El primer paso en estos enfoques es el cálculo de una firma de forma. Esta tarea se realiza por lo general, mediante el cálculo de una función 1D que representa la información de la frontera de la hoja. Seguidamente, se calcula una transformada de Fourier de esta señal para caracterizar variaciones de alta y baja frecuencia en el borde de la hoja. Dependiendo del tipo de firma y el tipo de transformada de Fourier utilizadas, se pueden obtener diferentes tipos de representación de forma [38, 39]. Por ejemplo, si la firma del borde corresponde a un código de cadena y sobre este se aplica una transformada de Fourier real, se obtienen coeficientes de Fourier elípticos. Este tipo de representación es adecuada para los contornos cerrados (como en el caso de hojas no sésiles),

Característica geométrica	Relación	Descripción
<i>Relación de aspecto</i>	Longitud máxima/longitud mínima	Razón de longitud máxima a longitud mínima
<i>Rectangularidad</i>	$ARDI/(L. \text{ Max} * L. \text{ min})$	Área de la región de interés sobre (máxima longitud*mínima longitud)
<i>Relación de aspecto de la envolvente convexa</i>	$ARDI/A$	Relación entre el área de la región de interés y el área de la envolvente convexa
<i>Esfericidad</i>	$RCI/RCE$	Radio de círculo interior de la región de interés y Radio círculo exterior del área de interés
<i>Circularidad</i>	$RDM/DCM$	Relación distancia media entre el centro de la region de interés y todos los puntos de borde y la desviación cuadrática media de la distancia media
<i>Excentricidad</i>	$RLIP/ LEIROI$	Relación de la longitud del eje de inercia principal y la longitud del eje de inercia del área de interés
<i>Perímetro de la relación de la envolvente convexa</i>	$PADI/PEC$	Relación del radio del perímetro de la envolvente convexa representado la relación del perímetro de la región de interés y el perímetro de la envolvente convexa

**Tabla 1-1.:** Características geométricas simples comúnmente utilizadas para la descripción de la forma foliar.

y además permite compactar la forma general de la hoja en los coeficientes de baja frecuencia de la representación, y los detalles específicos de la forma en coeficientes de alta frecuencia. En algunos casos, donde la forma de la hoja debe describirse necesariamente utilizando curvas abiertas (por ejemplo, hojas de arroz o en vegetación de páramo andino), la aproximación de transformada de Fourier elíptica es limitada [40]. Principalmente, debido a las condiciones periódicas impuestas por el método. En estos de números complejos, utilizando una transformada de Fourier compleja para calcular los armónicos correspondientes [41]. Este enfoque también compacta información de forma en pocos armónicocasos, pueden usarse representaciones alternativas, como la transformada P-Type de Fourier [40]. En esta transformación, los segmentos del borde de la hoja se codifican como una secuencias, pero evita inestabilidades en los extremos de la curva que representa la forma foliar [40]. Por esta razón, esta transformación puede ser útil para la descripción de la formas de diferentes partes de la hoja, por ejemplo, ápice, pecíolo y lámina.

Otro método de descripción de los bordes propuesto en la literatura para la caracterización de los órganos foliares es el análisis estadístico de la forma [42, 43]. Para cada contorno, un conjunto de puntos guía son localizados sobre el borde del contorno, delimitando regiones clave, por ejemplo, base y pecíolo. El conjunto de puntos guía son representados como puntos en un espacio vectorial sobre los cuales se aplican métodos estadísticos, como por ejemplo, análisis de componentes principales, para estimar los patrones de forma del conjunto de objetos bajo estudio. La principal limitación de esta aproximación es la necesidad de localizar los puntos guía sobre el contorno, tarea que usualmente se realiza manualmente.

## **Descriptores de área basados en momentos**

La fase de extracción de las características puede dar como resultado un conjunto de descriptores de muy alta dimensión. En muchos casos, la mayoría de estas características no son informativas para la descripción de la forma. Por lo tanto, métodos de selección de características y de reducción de la dimensionalidad se pueden utilizar para elegir un subconjunto reducido de características relevantes que permitan explicar la forma o visualizar las relaciones de forma en las muestras. Comúnmente, se espera que este método de reducción de dimensionalidad provea representaciones interpretables [44], que permitan al botánico interpretar visual y geométricamente los rasgos morfológicos foliares.

En el dominio de descripción de la forma foliar, uno de los métodos más populares para realizar, simultáneamente, ambas tareas es el llamado SHAPE [45]. En este método, primero se representa la forma de un conjunto de hojas usando armónicos elípticos de Fourier. A continuación, se calcula un análisis de componentes principales (ACP) sobre esta representación. El ACP, calcula el conjunto de bases ordenadas ortogonales que explican la mayor parte de la variabilidad de forma en la muestra de hojas, estas bases pueden interpretarse como posibles “explicaciones” o “causas” de forma [45]. Normalmente, el conjunto de componentes principales (CP) que capturan la mayor parte de la variabilidad de la forma, se seleccionan



para un análisis estadístico posterior [40]. Estos CP, también se pueden utilizar en tareas de visualización, en este caso, típicamente se seleccionan los dos primeros CPs [40]. El enfoque SHAPE ha demostrado ser de gran utilidad y aceptación para la descripción de la forma en diferentes dominios biológicos, incluyendo, botánica, paleobotánica, y la agricultura, entre otros. No obstante, este método presenta ciertas limitaciones, por ejemplo, el ACP, asume la existencia de relaciones ortogonales entre los componente principales, de esta forma el número de posibles explicaciones atribuibles para la forma está restringido por la dimensión del espacio de descripción [40]. Esta restricción limita la capacidad explicativa del método. Otras estrategias de visualización de la forma de las hojas se han explorado en la literatura, principalmente, mediante el cálculo de embebidos de baja dimensión utilizando, por ejemplo, isomap [46] o proyecciones lineales locales [47].

### 1.2.3. Caracterización del rasgo foliar

Un paso crítico para la categorización de la morfología foliar es la construcción de un conjunto de categorías de referencia de forma. En esta fase, la opinión del experto desempeña un papel fundamental, en la medida en que éste define las categorías morfológicas basadas en su conocimiento, y el problema biológico en cuestión.

Como se describió anteriormente, diferentes enfoques automáticos han sido propuestos recientemente para la categorización de la forma de la lámina foliar. Mzoughi, propuso utilizar la información visual extraída del manual botánico de la arquitectura foliar [3]. Por otro lado, Cerutti, buscaba discriminar entre las categorías descriptivas de margen definido para una región geográfica específica, en Francia [48]. En ambos casos, los sistemas de categorización propuestos pueden resultar insuficientes para describir la morfología foliar de muestras de regiones geográficas inexploradas. Por ejemplo, en el caso de vegetación tropical donde se espera una alta variabilidad morfológica dados sus altos índices de biodiversidad, los dos sistemas de categorización, no necesariamente se pueden utilizar para la tarea de descripción de la forma para este tipo de muestras. Por esta razón, es importante construir nuevos sistemas de categorización aplicables a otros escenarios biológicos, no necesariamente descritos en la literatura.

Después de construir un sistema de categorías de referencia para la aplicación particular se debe determinar la categoría morfológica. Para esta tarea, Mzoughi, propuso un enfoque jerárquico para reducir la superposición entre categorías, mediante el filtrado de diferentes entradas del vector de características. Este enfoque se evaluó en un entrenamiento y prueba de partición [2]. Por otra parte, Cerutti, [32] propuso utilizar un algoritmo de fuzzy C means para la tarea de clasificación.

En este contexto, esta tesis aborda los siguientes objetivos:

## 1.3. Objetivos de la tesis

*En esta tesis se proponen un conjunto de métodos para la caracterización automática de categorías de forma foliar para vegetación tropical, a partir de imágenes digitales.*

Como objetivos específicos se propone:

1. Formular un método computacional para la categorización automática de categorías de forma foliar, para un sistema de categorías adaptado a la vegetación tropical.
2. Implementar el método propuesto.
3. Evaluar el método propuesto en la tarea de clasificación de categorías de formas foliares, para un conjunto de muestras de vegetación tropical.

### 1.3.1. Contribuciones

En este trabajo se propuso... base de datos seleccion de características sistema de clasificación biológico método de clasificación automático

### 1.3.2. Contribuciones científicas

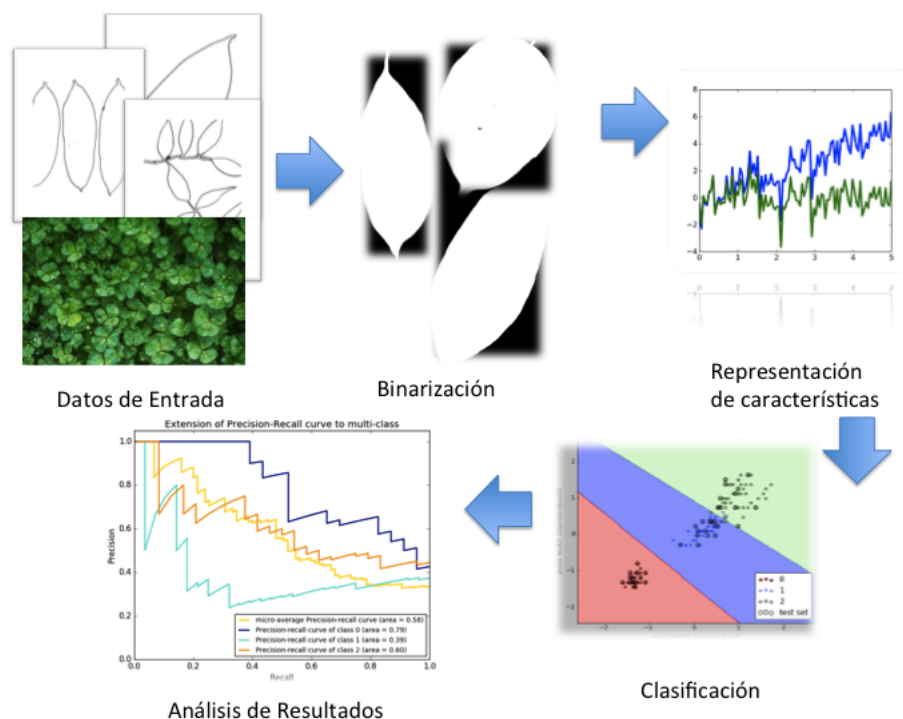
Como resultado de este trabajo se publicaron dos contribuciones científicas: un artículo de conferencia en el Congreso Colombiano de Computación (ver anexo A), y una contribución en el Congreso de Innovación y Tendencias en Ingeniería (ver anexoB).

## 2. Materiales y Métodos

En este capítulo, se introduce una estrategia para la caracterización automática de diferentes categorías de forma foliar en vegetación tropical, a partir de imágenes digitales. La estrategia propuesta se ilustra en la Figura 2-1. En primer lugar, y junto con la colaboración de un experto, se construyó un sistema de categorización particular para la descripción de las formas foliares apropiado para la descripción morfológica en vegetación tropical. Seguidamente, dicho sistema de categorización fue utilizado por el experto para etiquetar un conjunto de muestras foliares, correspondientes a vegetación tropical (datos de entrada). A continuación las muestras foliares fueron binarizadas y el contorno de cada muestra fue extraído. Un conjunto de características morfológicas fue calculado para la descripción de la lámina foliar y sus bordes, incluyendo, descriptores geométricos simples, representaciones de contorno basadas en armónicas, y momentos geométricos. Un método de clasificación multiclase basado en aprendizaje de máquina, fue utilizado, para la tarea de clasificación de las formas. El método propuesto fue evaluado en un conjunto de 3285 muestras etiquetadas por un experto, en un esquema de entrenamiento, validación y prueba. Finalmente, se estudiaron el conjunto de características con mayor contribución a la varianza de los datos observados. A continuación, se describen cada uno de estos aspectos.

### 2.1. Muestras foliares

Para este proyecto se utilizó un subconjunto de 3285 muestras foliares, correspondientes a vegetación tropical obtenidas por el grupo Biodiversidad y Conservación dentro del proyecto *Patrón biotipológico foliar en un gradiente de precipitación en la región caribe de Colombia* [49]. Los inventarios de información para la toma de muestras se definieron según [50, 51], de la siguiente forma: arbóreo superior (As) > 25 m, arbóreo inferior (Ai) 25 - 12 m, subarbóreo (Ar) 12 - 5 m, arbustivo (Arb) 5 - 1,5 m, herbáceo (H) 1.5 - 0.25 m, rasante (R) < 0.25 m. Para la caracterización foliar se colectaron de 5 a 12 hojas por individuo. En el campo se tomó el peso fresco de las hojas con dos pesolas y una balanza electrónica portátil (Balanza Lexus, precisión 0.1 gr.), se procuró tomar el peso en el momento de obtener la muestra, para evitar la rápida pérdida de agua. Después, el material vegetal fue prensado y preservado en alcohol al 70 %, con el fin de minimizar el riesgo de variaciones morfológicas generadas por procesos de deshidratación [49]. Los detalles sobre las localidades geográficas de las cuales las muestras fueron obtenidas se describen en [29].



**Figura 2-1.:** Estrategia propuesta. En primer lugar, se construyó un sistema de categorización específico para muestras en vegetación tropical. Un conjunto de datos de vegetación tropical fue utilizado. Las muestras foliares fueron digitalizadas, binarizadas y el contorno fue extraído. Un conjunto de características morfológicas cuantitativas fue utilizado para la descripción. Seguidamente, un método clasificación fue empleado para la determinación de las diferentes categorías. Finalmente, el método fue evaluado en el conjunto de datos, previamente, etiquetado por el experto.

## 2.2. Sistema de categorización de formas foliares

Para la definición de un sistema de categorización particular para vegetación tropical, un experto estableció un criterio de búsqueda sobre manuales botánicos relacionados con la descripción de vegetación tropical desde el año 1900. Como conclusión de la revisión los dos sistemas de descripción más completos fueron el manual de Hickey [3] y el trabajo de Tafur y Rangel [52]. El trabajo de Tafur y Rangel, resume muy bien las posibles formas de hoja que pueden aparecer en varios ambientes a nivel Colombia, el de Hickey provee objetividad al decidir si una hoja entra o no, en una categoría específica. Los criterios utilizados por estos dos sistemas de clasificación, fueron unificados para definir las categorías del nuevo sistema de referencia.

## 2.3. Pre-procesamiento

Para la extracción de los contornos foliares a partir de las imágenes digitales se utilizó un algoritmo de binarización, seguido por un algoritmo de trazado de contornos [53].

## 2.4. Caracterización de los bordes

Después de la extracción de los bordes, fue realizada una caracterización morfológica cuantitativa de la hoja. Para esta tarea, se usaron tres aproximaciones diferentes. Primero, un conjunto de características previamente utilizadas para la descripción de morfología foliar [49]. Segundo, descriptores elípticos de Fourier para representar los contornos de las hojas [38]. Tercero, descriptores basados en momentos, previamente utilizados para la caracterización de muestras foliares, más específicamente, momentos de Hu [54].

### 2.4.1. Descriptores geométricos simples

Con el fin de describir la información geométrica de las muestras foliares, se utilizaron un conjunto de características geométricas básicas, incluyendo, relación de aspecto de la envolvente convexa de la hoja, excentricidad, factor de angostura, relación eje mayor a eje menor, circularidad, rectangularidad, y razón entre el perímetro [29].

### 2.4.2. Representaciones elípticas de Fourier

Para la representación de la información del contorno se utilizó representación de contornos elípticas de Fourier, en ésta, el contorno es codificado al utilizar un código de cadena de Freeman [38]. Seguidamente, una transformada discreta de Fourier es calculada sobre la representación, resultando en un conjunto de armónicos, indexados por  $n$ , como sigue:

$$A_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^k \frac{\Delta x_p}{\Delta y_p} \left( \cos\left(\frac{2\pi n t_p}{T}\right) - \cos\left(\frac{2\pi n t_{p-1}}{T}\right) \right)$$

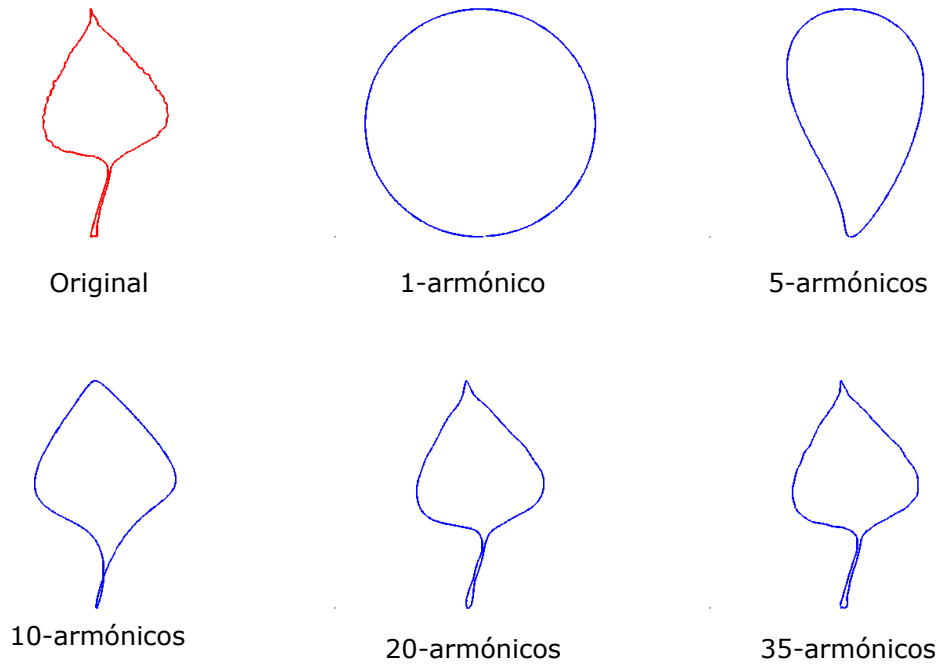
$$B_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^k \frac{\Delta x_p}{\Delta y_p} \left( \sin\left(\frac{2\pi n t_p}{T}\right) - \sin\left(\frac{2\pi n t_{p-1}}{T}\right) \right)$$

$$C_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^k \frac{\Delta y_p}{\Delta y_p} \left( \cos\left(\frac{2\pi n t_p}{T}\right) - \cos\left(\frac{2\pi n t_{p-1}}{T}\right) \right)$$

$$D_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^k \frac{\Delta y_p}{\Delta y_p} \left( \sin \left( \frac{2\pi n t_p}{T} \right) - \sin \left( \frac{2\pi n t_{p-1}}{T} \right) \right)$$

donde  $k$ , es el número de puntos utilizados para la codificación del borde;  $\Delta x$  y  $\Delta y$ , los desplazamientos entre el punto  $p$  y el punto  $p+1$  a lo largo del eje  $x$  e  $y$ , respectivamente;  $\Delta t_p$ , la distancia entre el punto  $p$  y el punto  $p+1$ ;  $t_p$  la longitud acumulada de los segmentos hasta  $p$ , y  $T$ , es el tamaño total del borde.  $A_n$  y  $B_n$ , describe el aspecto de forma del contorno en el eje  $x$ , y  $C_n$  y  $D_n$ , en el eje  $y$ .

Estos armónicos determinan una serie de formas elípticas de detalle incremental, su combinación describe la forma completa de la hoja. Para esta experimentación, 20 armónicos fueron utilizados. Los coeficientes elípticos de Fourier se normalizaron con el fin de hacerlos invariantes a transformaciones de escala y rotación [37]. La figura 2-2 ilustra la reconstrucción obtenida al utilizar un número incremental de coeficientes elípticos de Fourier.



**Figura 2-2.:** Representación del contorno. La gráfica representa la imagen original, y la reconstrucción de la forma de la hoja utilizando varios conjuntos de armónicos, tomada de [1].

### 2.4.3. Descripción del área basada en momentos

En la categorización de formas de láminas foliares, comúnmente, se incluyen descripciones referentes a la distribución del área sobre la lámina foliar [37]. Por ejemplo, la forma de hoja cordada se describe como una lámina con forma de corazón, o la reniforme que corresponde a láminas con forma de riñón. Las descripciones textuales, buscan capturar propiedades morfológicas relacionadas con la distribución del material biológico sobre la lámina. Esta característica puede ser capturada por los momentos geométricos, en particular, los momentos geométricos invariantes [55]. Los momentos de la imagen han sido previamente utilizados para la caracterización de la morfología foliar. Estos momentos, capturan diferentes propiedades de la distribución espacial de los puntos. En este trabajo se tomaron los primeros 8 momentos geométricos de Hu [54], con el fin de capturar diferentes patrones de distribución de área al interior de la lámina foliar, se definen de la siguiente manera:

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2].$$

$$I_8 = \eta_{11}[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] - (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})$$

Donde:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}}$$

con  $\mu_{00}$ , el área del interior de hoja,  $\mu_{ij}$ , los momentos centrales definidos en términos de los momentos simples de la región [54]. Estos momentos describen diferentes propiedades del área. Por ejemplo, el primer momento  $I_1$ , es análogo al momento de inercia, donde la intensidad de los pixeles es análoga a la densidad física. El último momento  $I_7$ , es invariante al sesgo y permite describir las propiedades de simetría del área.

## 2.5. Clasificación automática de las formas foliares

Después de la descripción de los contornos, y de la distribución del área al interior de la lámina foliar, un método para la clasificación multiclase (una por cada categoría de forma) de la forma, fue propuesto. Para esta tarea, tres máquinas de aprendizaje fueron utilizadas: máquinas con vectores de soporte uno versus uno, máquinas con vectores de soporte uno

versus todos y árboles aleatorios [56]. Para las maquinas con vectores de soporte, se utilizó un kernel lineal, el parámetro de margen  $C$  se exploró en una escala logarítmica en base 10 con potencias de  $\{-2, 9\}$ . Respecto a los arboles aleatorios, el número de árboles se exploró en un conjunto de estimadores de  $\{10, 200\}$ .

## 2.6. Evaluación de la estrategia de clasificación

Para las tareas de entrenamiento, validación y prueba. El conjunto de datos se subdividió en tres particiones disyuntas: entrenamiento (30 %), validación (30 %), y prueba (30 %). Con el fin de compensar sesgos por imbalance de clases, esta subdivisión se realizó utilizando nuestro estratificado [57]. Para el entrenamiento de cada máquina, cuatro conjuntos de características morfológicas fueron empleadas: descriptores geométricos simples, descriptores basados en contornos, descriptores basados en momentos, y un conjunto con todos los descriptores. Se trabajaron tres medidas de desempeño del clasificador: precisión (*precision*), recuperación (*recall*), y medida F1 . Utilizando el conjunto de validación se determinó la máquina de aprendizaje con mayor desempeño usando la medida F1 [58]. Una vez seleccionada la máquina de aprendizaje, esta se evaluó sobre el conjunto de prueba.

## 2.7. Selección de características

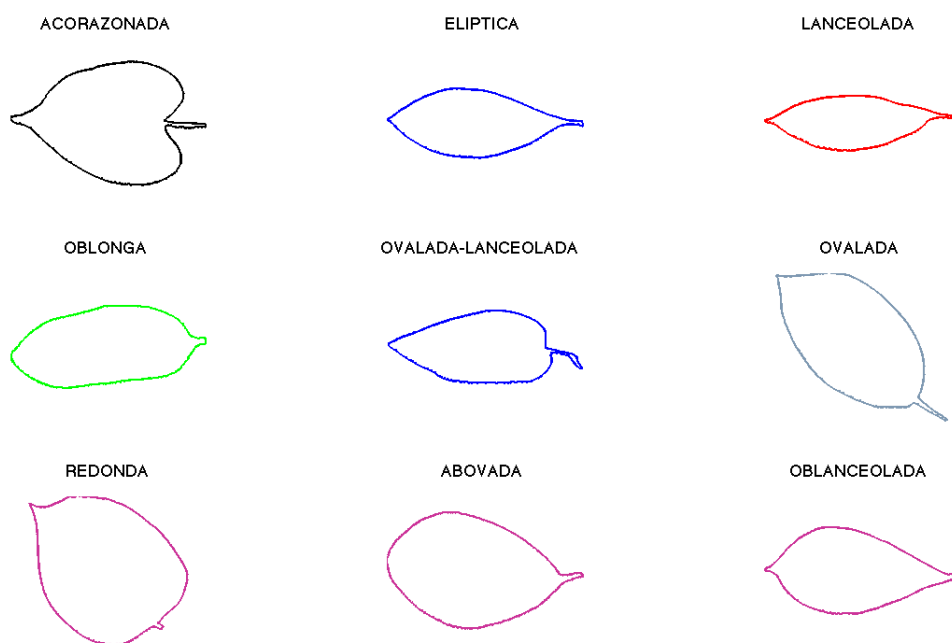
Finalmente, se estudio el sub-conjunto de características que mejor describieron la forma foliar. Esta tarea fue posible gracias a la naturaleza interpretable de la caracterización utilizada para la descripción de la forma foliar. En particular, se utilizó un método de selección de características basado en análisis de componentes principales []. Para esto, el conjunto de datos se particionó de manera estratificada en dos sub-conjuntos: entrenamiento (70 %), utilizado para el cálculo de los componentes principales, y prueba (70 %) utilizado para estudiar la proyección de un conjunto de datos independientes sobre el conjunto de componentes principales previamente calculados. Para la selección de las características de mayor contribución, se escogió el vector propio asociado al mayor valor propio en valor absoluto. Las entradas con mayor valor de los componentes del vector propio fueron seleccionada como las características de mayor contribución a la caracterización morfológica foliar.



## 3. Resultados

### 3.1. Sistema de categorización de formas foliares para vegetación tropical

Como se describe en la sección 2.2, los manuales botánicos actuales pueden ser insuficientes para la descripción de las categorías de forma de las láminas foliares en vegetación tropical. Del resultado del análisis realizado por el experto al respecto se sugiere el uso de nueve categorías de forma para la descripción: “acorazonada”, “elíptica”, “lanceolada”, “oblonga”, “ovalada-lanceolada”, “ovalada”, “redonda”, “abovada”, y “oblanceolada”. En la figura 3-1, se puede observar un ejemplo de cada una de las categorías sugeridas por el experto.



**Figura 3-1.:** Ejemplos de muestras foliares definidas para cada categoría del sistema de clasificación de láminas foliares.

## 3.2. Muestras foliares categorizadas

Utilizando el conjunto de categorías previamente descritas, un experto realizó el etiquetado de las 3285 muestras foliares. La distribución de las muestras por tipo de forma de lámina foliar puede encontrarse en la tabla **3-1**.

<i>Categoría de forma</i>	<i>Número de muestras</i>
Abovada	145
Acorazonada	41
Elíptica	819
Lanceolada	606
Oblanceolada	389
Oblonga	477
Ovalada	295
Ovalada-lanceolada	479
Redonda	34

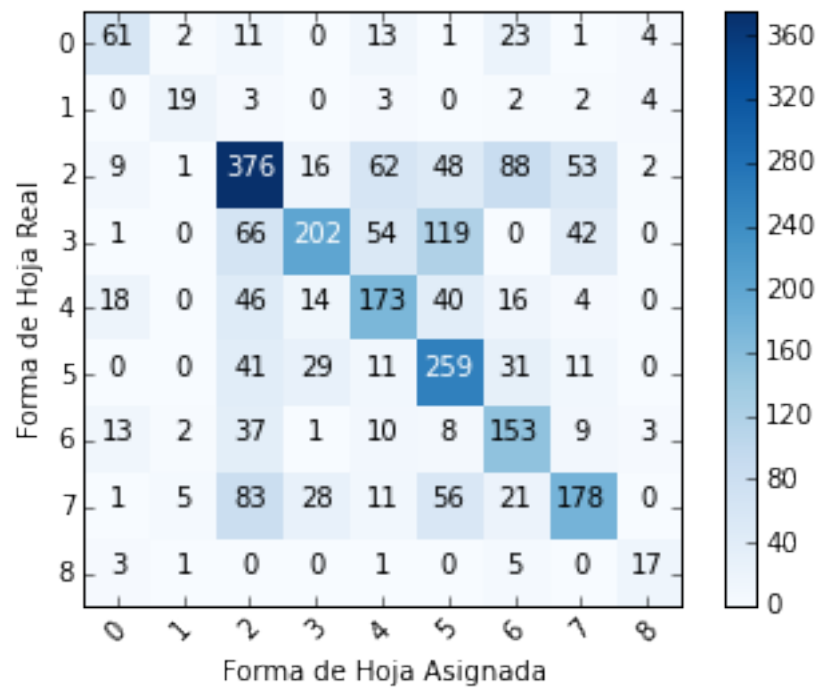
**Tabla 3-1.:** Distribución de las muestras etiquetadas por tipo de lámina foliar

## 3.3. Clasificación automática de las categorías de lámina foliar

La tabla **3-2**, muestra los resultados del proceso de validación para cada uno de los métodos explorados, en este caso, se reportan los resultados asociados al mejor conjunto de parámetros. Como puede observarse el método de árboles aleatorios provee el mejor desempeño sobre las muestras de entrenamiento. Este desempeño se obtuvo utilizando el conjunto completo de descriptores de forma (geométricos, elípticos de Fourier y momentos de Hu).

Dados los resultados del proceso de entrenamiento y validación, se seleccionó como mejor método de clasificación árboles aleatorios con 150 estimadores sobre el conjunto completo de descriptores de forma. La tabla **3-3** reporta el desempeño obtenido por este clasificador sobre el conjunto de datos de prueba. Como puede observarse en datos de prueba el método propuesto obtiene un desempeño del 52 % promedio en precisión (precision) y del 51 % promedio en recuperación (recall). El mejor desempeño en clasificación se obtuvo para hojas de tipo Lanceolada (F1 de 0.65), y el peor desempeño para hojas redondas (F1 de 0.25).

La figura **3-2**, muestra la matriz de confusión asociada al mejor clasificador. Como puede observarse redonda presenta el desempeño más bajo, con un grado alto de confusión con los otros tipos de forma.



**Figura 3-2.:** Matriz de confusión máquina de clasificación multiclase (1 - Abovada, 2 - Acorazonada, 3 - Elíptica, 4 - Lanceolada, 5 - Oblanceolada, 6 - Oblonga, 7 - Ovalada, 8 - Ovalada-Lanceolada y 9 - Redonda), basada en árboles aleatorios con 150 estimadores sobre el conjunto completo de características de forma (geométricos, elípticos de Fourier y momentos de Hu).

### 3.3.1. Selección de características

Aca van los resultados!!

<i>Método</i>	<i>Categoría de forma</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
OnVsAll	Abovada	0	0	0
	Acorazonada	0.68	0.7	0.69
	Elíptica	0.41	0.85	0.56
	Lanceolada	0.57	0.85	0.68
	Oblanceolada	1	0	0.01
	Oblonga	0.56	0.47	0.51
	Ovalada	0.41	0.28	0.33
	Ovalada-Lanceolada	0.48	0.04	0.07
	Redonda	0.8	0.15	0.25
	avg/total	0.53	0.48	0.39
OnVsOne	Abovada Clase 0	0.58	0.53	0.55
	Acorazonada	0.63	0.58	0.6
	Elíptica	0.57	0.57	0.57
	Lanceolada	0.57	0.85	0.68
	Oblanceolada	0.7	0.42	0.52
	Oblonga	0.49	0.68	0.57
	Ovalada	0.45	0.65	0.53
	Ovalada-Lanceolada	0.59	0.46	0.52
	Redonda	0.57	0.63	0.6
	avg/total	0.57	0.55	0.55
Random Forest	Abovada	0.7	0.41	0.52
	Acorazonada	0.92	0.73	0.81
	Elíptica	0.59	0.78	0.67
	Lanceolada	0.68	0.83	0.75
	Oblanceolada	0.64	0.46	0.55
	Oblonga	0.68	0.54	0.61
	Ovalada	0.59	0.52	0.56
	Ovalada-Lanceolada	0.74	0.65	0.69
	Redonda	0.73	0.41	0.52
	avg/total	0.66	0.65	0.65

**Tabla 3-2.:** Medidas de desempeño de los métodos de clasificación en validación

<i>Categoría de forma</i>	<i>Precision</i>	<i>Recall</i>	F1
Abovada	0.24	0.24	0.29
Acorazonada	0.44	0.50	0.47
Elíptica	0.49	0.64	0.55
Lanceolada	0.59	0.72	0.65
Ob lanceolada	0.43	0.33	0.37
Oblonga	0.49	0.39	0.44
Ovalada	0.54	0.49	0.51
Ovalada-Lanceolada	0.64	0.44	0.52
Redonda	1.00	0.14	0.25
avg/total	0.52	0.51	0.51

**Tabla 3-3.:** Desempeño de una máquina de clasificación multiclase basada en árboles aleatorios con 150 estimadores sobre el conjunto completo de características de forma (geométricos, elípticos de Fourier y momentos de Hu), para la clasificación de diferentes categorías de forma de lámina foliar en vegetación tropical.

## 4. Discusión

En este trabajo se propuso un nuevo método para la categorización automática de formas de lámina foliar en vegetación tropical, a partir de imágenes digitales. El objetivo de este método, es reducir los sesgos inherentes a la caracterización de este rasgo morfológico, introducidas por el etiquetado manual. Para esto, se propuso, junto con un experto, un nuevo sistema de categorías para la descripción de diferentes tipos de forma de lámina foliar. El contorno, la geometría y la distribución del área a lo largo de la lámina fueron utilizados para la caracterización de la forma. Un método de clasificación multiclase fue utilizado para resolver la tarea de categorización de forma automática. Este sistema de categorización automática se utilizó para la descripción del tipo de forma en una muestra de láminas foliares etiquetadas por un experto.

En los últimos años, diversos trabajos se han propuesto para caracterizar automáticamente la forma de la hoja [59, 60, 61]. La mayoría de estos trabajos se centra en el problema de la clasificación de especies de plantas, es decir, dado un conjunto de muestras foliares de especies predefinidas se busca construir un clasificador de diferentes especies. Una suposición subyacente importante en estas obras, es la existencia de un conjunto previamente definido de categorías (especies) de la forma. Esta hipótesis, tal vez no es válida en escenarios exploratorios biológicos donde las categorías de forma tal vez no son fijadas previamente. Por ejemplo, en el escenario de construcción de manuales de la forma de la hoja [6, 7]. El problema de clasificación de especies de plantas previamente se ha abordado usando modelos no lineales de clasificación supervisada, por ejemplo, deep learning [62], métodos basados en kernel [63, 64]. Estos métodos, proporcionan resultados competitivos desde la perspectiva de la clasificación. Sin embargo, estos métodos no necesariamente se adaptan a las particularidades morfológicas de la vegetación tropical colombiana. En este trabajo se propuso un nuevo sistema de clasificación de formas para la lámina foliar. El sistema está compuesto por 9 categorías de lámina foliar: “*acorazonada*”, “*elíptica*”, “*lanceolada*”, “*oblonga*”, “*ovalada-lanceolada*”, “*ovalada*”, “*redonda*”, “*abovada*”, y “*oblanceolada*”. Estas categorías toman en cuenta algunas de las características principales de la morfología foliar tropical. Este sistema de categorización potencialmente puede ser utilizado en estudios posteriores relacionados con este tipo de vegetación.

Utilizando el sistema de categorización propuesto, se construyó una nueva base de datos de láminas foliares, curadas y anotada por un experto, sobre el sistema de categorías propuesto. Esta base de datos puede representar un insumo importante en el estudio del problema de

categorización de formas [29, 49], y el problema de caracterización de la morfología foliar. Es importante anotar que la generación de este tipo de bases de datos para el ámbito biológico es de gran importancia para el desarrollo y estandarización de nuevos métodos en botánica [29, 49].

Utilizando el sistema propuesto como referencia, se propuso un nuevo sistema de categorización automática basado en aprendizaje de máquina. Cabe destacar que algunos trabajos previos se han propuesto para caracterizar formas de lámina foliar [29, 49]. No obstante, estos trabajos no necesariamente se adaptan a la morfología de la vegetación tropical, y en particular, al sistema de categorización de formas propuesto. El sistema automático utiliza un conjunto de descriptores de forma previamente utilizado para la descripción de láminas foliares, y un clasificador multiclase de árboles aleatorios. En general, el método propuesto presenta buenos desempeños para la caracterización de los distintos tipos de formas con desempeños muy por encima del nivel obtenido usando selección aleatoria (0.12). Sin embargo, el sistema presenta limitaciones respecto a la categorización de dos tipos de formas específicas, redonda y abobada. Estas limitaciones pueden derivarse de dos aspectos; primero, el alto grado de similaridad visual entre la forma abovada y el resto de las categorías; y segundo, el número reducido de muestras para la categoría redonda. Es importante resaltar que en categorías bien definidas, por ejemplo, acorazonada, el método provee un buen desempeño, independientemente del número reducido de categorías.

Respecto a las características de forma utilizadas, el mejor desempeño estuvo asociada a la combinación de los tres descriptores de forma utilizados: geométricos, elípticos de Fourier y momentos de Hu. Al respecto, una gran variedad de descriptores de forma se han propuesto para la caracterización de la lámina foliar. En general, estos descriptores son de difícil interpretación por parte de los expertos. Para la caracterización, propuesta se utilizaron descriptores con un alto grado de interpretabilidad biológica, como es el caso de las características geométricas y de los descriptores elípticos de Fourier. El resultado de la caracterización propuesta, puede ser posteriormente interpretada por los botánicos complementando el poder del método.

Este trabajo presenta algunas limitaciones. En primer lugar, el conjunto de muestras presenta un alto grado de imbalance de clases. Esta restricción se deriva de la naturaleza del problema, en la medida en que las formas foliares emergen con patrones heterogéneos. Esta restricción limita el poder del método de clasificación. Respecto a la naturaleza del sistema de categorización, en general, es propuesto por los expertos. No obstante, esta forma de construcción del sistema de categorización también puede introducir sesgos. En este caso, puede ser recomendable el uso de métodos automáticos para el descubrimiento de categorías. En cuanto a la caracterización de la forma, métodos alternativos basados en aprendizaje de características pueden ser utilizados. Sin embargo, en este caso es necesario ampliar el conjunto de datos.

## 5. Conclusiones

En este trabajo se propuso un método de clasificación automática de la morfología de hojas tropicales. El método propuesto se basa en un nuevo sistema de categorización para láminas foliares, descriptores de forma cuantitativos, y técnicas de clasificación multiclase. El método fue evaluado en una nueva base de datos anotada por un experto. Los resultados sugieren que el método propuesto presenta desempeños competitivos para diferentes tipos de forma, no obstante, el alto grado de similitud morfológica dentro de algunas categorías propuestas por el experto, limitan el desempeño de la estrategia propuesta.

### 5.1. Recomendaciones

Se recomienda una experimentación mucho mas profunda con valores logaritmicos de base más pequeña para encontrar mejor respuesta en terminos de medidas de desempeño de clasificación, además es importante notar que de una u otra forma las representacion es de características que se usan quizás no aportan la suficiente información para obtener unas medidas de desempeño mas altas influida también por la alta variabilidad de formas de hoja que pertenecen a una misma clase.

Se sugiere ampliar la cantidad de hojas de la base de datos construida actualmente se cuenta con 3284, para compensar en imbalance de clases, etiquetadas por expertos biologos y botánicos de la Universidad Nacional de Colombia.



## **A. Contribución Congreso Colombiano de Computación**

*P-type based dimensionality reduction for open contours of Paramo leaves. Congreso colombiano de Computación. 2014.*

## **B. Congreso de Innovación y Tendencias en Ingeniería**

*Un método automático para la caracterización morfológica digital de hojas tropicales de Colombia. Congreso de Innovación y Tendencias en Ingeniería. 2015.*

# Bibliografía

- [1] L. Olivares, “Modelo de caracterización de la morfología foliar a partir de imágenes digitales,” 2016.
- [2] O. Mzoughi, I. Yahiaoui, N. Boujemaa, and E. Zagrouba, “Automated semantic leaf image categorization by geometric analysis,” in *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, Jul. 2013, pp. 1–6.
- [3] B. Ellis, D. Daly, L. Hickey, J. Mitchell, K. Johnson, P. Wilf, and S. Wing, *Manual of Leaf Architecture*, Dec. 2009, vol. 34, no. 4.
- [4] J. S. Wright, “Plant diversity in tropical forests: a review of mechanisms of species coexistence,” *Oecologia*, vol. 130, no. 1, pp. 1–14, 2002.
- [5] A. B. Nicotra, A. Leigh, C. K. Boyce, C. S. Jones, K. J. Niklas, D. L. Royer, and H. Tsukaya, “The evolution and functional significance of leaf shape in the angiosperms,” *Functional Plant Biology*, vol. 38, no. 7, pp. 535–552, 2011.
- [6] H. Ball, A. Exell, J. Harding, J. L’eonard, J. Lewis, A. Melderis, R. Melville, F. Stafleu, S. Walters, P. Duvigneaud, E. Petit, R. Tournay, and P. V. der Veken, “Systematics association committee for descriptive biological terminology. II. Terminology of simple symmetrical plane shapes (Chart 1),” *Taxon*, vol. 41, no. 11, pp. 145,156, 1962.
- [7] H. Beentje, *The Kew Plant Glossary: An Illustrated Dictionary of Plant Terms*. Kew Publishing, 2012, vol. 34.
- [8] J. Cope, D. Corney, J. Clark, P. Remagnino, and P. Wilkin, “Plant species identification using digital morphometrics: A review,” *Expert Systems with Applications*, vol. 39, no. 8, pp. 7562–7573, Jun. 2012.
- [9] M. Aronoff and K. Fudeman, *What is morphology*. John Wiley & Sons, 2011, vol. 8.
- [10] C. A. Price, O. Symonova, Y. Mileyko, T. Hilley, and J. S. Weitz, “Leaf extraction and analysis framework graphical user interface: segmenting and analyzing the structure of leaf veins and areoles,” *Plant Physiology*, vol. 155, no. 1, pp. 236–245, 2011.
- [11] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, “Plant species identification using digital morphometrics: A review,” *Expert Systems with Applications*, vol. 39, no. 8, pp. 7562–7573, 2012.

- [12] D. P. Corney, H. L. Tang, J. Y. Clark, Y. Hu, and J. Jin, “Automating digital leaf measurement: The tooth, the whole tooth, and nothing but the tooth,” *PloS one*, vol. 7, no. 8, p. e42112, 2012.
- [13] D. H. Chitwood, R. Kumar, L. R. Headland, A. Ranjan, M. F. Covington, Y. Ichihashi, D. Fulop, J. M. Jiménez-Gómez, J. Peng, J. N. Maloof *et al.*, “A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines,” *The Plant Cell*, vol. 25, no. 7, pp. 2465–2481, 2013.
- [14] D. Koenig, J. M. Jiménez-Gómez, S. Kimura, D. Fulop, D. H. Chitwood, L. R. Headland, R. Kumar, M. F. Covington, U. K. Devisetty, A. V. Tat *et al.*, “Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 28, pp. E2655–E2662, 2013.
- [15] H. Goëau, P. Bonnet, A. Joly, V. Bakic, D. Barthélémy, N. Boujemaa, and J.-F. Molino, “The imageclef 2013 plant identification task,” in *CLEF*, 2013.
- [16] S. Mouine, I. Yahiaoui, A. Verroust-Blondet, L. Joyeux, S. Selmi, and H. Goëau, “An android application for leaf-based plant identification,” in *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*. ACM, 2013, pp. 309–310.
- [17] C. J. Ter Braak, “The analysis of vegetation-environment relationships by canonical correspondence analysis,” in *Theory and models in vegetation science*. Springer, 1987, pp. 69–77.
- [18] W. ACH and R. Spicer, “Canonical correspondence analysis of leaf physiognomy: a contribution to the development of a new palaeoclimatological tool,” 1996.
- [19] J. Osnas, J. Lichstein, P. Reich, and S. Pacala, “Global leaf trait relationships: mass, area, and the leaf economics spectrum,” *Science (New York, N.Y.)*, vol. 340, no. 6133, pp. 741–4, May 2013.
- [20] N. Valliammal and S. Geethalakshmi, “Performance analysis of various leaf boundary edge detection algorithms,” in *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*. ACM, 2010, p. 34.
- [21] —, “Hybrid image segmentation algorithm for leaf recognition and characterization,” in *Process Automation, Control and Computing (PACC), 2011 International Conference on*. IEEE, 2011, pp. 1–6.
- [22] —, “Automatic recognition system using preferential image segmentation for leaf and flower images,” *Computer Science & Engineering*, vol. 1, no. 4, p. 13, 2011.

- [23] X. Zheng and X. Wang, “Leaf vein extraction based on gray-scale morphology,” *International Journal of Image, Graphics and Signal Processing*, vol. 2, no. 2, p. 25, 2010.
- [24] J. Park, E. Hwang, and Y. Nam, “Utilizing venation features for efficient leaf image retrieval,” *Journal of Systems and Software*, vol. 81, no. 1, pp. 71–82, 2008.
- [25] Y. Zhang, “The reconstruction of ancestral plant proteomes reveals the dynamics of early land plant evolution,” Ph.D. dissertation, The Pennsylvania State University, 2011.
- [26] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet, “Advanced shape context for plant species identification using leaf image retrieval,” in *Proceedings of the 2nd ACM international conference on multimedia retrieval*. ACM, 2012, p. 49.
- [27] D. Chitwood, L. Headland, R. Kumar, J. Peng, J. Maloof, and N. Sinha, “The developmental trajectory of leaflet morphology in wild tomato species,” *Plant physiology*, vol. 158, no. 3, pp. 1230–40, Mar. 2012.
- [28] R. Guzman, M. Berdugo, O. Rangel, F. Gonzalez, and F. Gómez, “p-type based dimensionality reduction for open contours of páramo leaves,” in *9th Computing Colombian Conference (9CCC)*, 2014.
- [29] F. G. G. F. M. Berdugo, O. Rangel, “A review on digital leaf morphology for assessment of environmental conditions,” 2014.
- [30] A. Backhaus, A. Kuwabara, M. Bauch, N. Monk, G. Sanguinetti, and A. Fleming, “Leafprocessor: a new leaf phenotyping tool using contour bending energy and shape cluster analysis,” *New phytologist*, vol. 187, no. 1, pp. 251–261, 2010.
- [31] M. Grand-Brochier, A. Vacavant, G. Cerutti, C. Kurtz, J. Weber, and L. Tougne, “Tree leaves extraction in natural images: Comparative study of pre-processing tools and segmentation methods,” *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1549–1560, 2015. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01176699>
- [32] G. Cerutti, L. Tougne, A. Vacavant, and D. Coquin, “A parametric active polygon for leaf segmentation and shape estimation,” in *International symposium on visual computing*. Springer, 2011, pp. 202–213.
- [33] O. H. Babatunde, L. Armstrong, and J. Leng, “Dean diepeveen survey of computer-based vision systems for automatic identification of plant species,” *AGRÁRINFORMATIKA/JOURNAL OF AGRICULTURAL INFORMATICS*, vol. 6, no. 1, pp. 61–70, 2015.
- [34] D. L. Royer, P. Wilf, D. A. Janesko, E. A. Kowalski, and D. L. Dilcher, “Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record,” *American Journal of Botany*, vol. 92, no. 7, pp. 1141–1151, 2005.

- [35] W.-J. Chen, C. Bonillo, and G. Lecointre, “Phylogeny of the channichthyidae (notothenioidei, teleostei) based on two mitochondrial genes,” in *Fishes of Antarctica*. Springer, 1998, pp. 287–298.
- [36] O. Babatunde, L. Armstrong, J. Leng, and D. Diepeveen, “A computer-based vision systems for automatic identification of plant species using knn and genetic pca,” *AGRÁRINFORMATIKA/JOURNAL OF AGRICULTURAL INFORMATICS*, vol. 6, no. 2, pp. 32–44, 2015.
- [37] V. Viscosi and A. Cardini, “Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners,” *PloS one*, vol. 6, no. 10, p. e25630, 2011.
- [38] F. P. Kuhl and C. R. Giardina, “Elliptic fourier features of a closed contour,” *Computer graphics and image processing*, vol. 18, no. 3, pp. 236–258, 1982.
- [39] Y. Yoshioka, H. Iwata, R. Ohsawa, and S. Ninomiya, “Analysis of petal shape variation of *primula sieboldii* by elliptic fourier descriptors and principal component analysis,” *Annals of Botany*, vol. 94, no. 5, pp. 657–664, 2004.
- [40] E. Shamoto, N. Suzuki, and R. Hino, “Analysis of 3d elliptical vibration cutting with thin shear plane model,” *CIRP Annals-Manufacturing Technology*, vol. 57, no. 1, pp. 57–60, 2008.
- [41] F. Chen, H. Yu, R. Hu, and X. Zeng, “Deep learning shape priors for object segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1870–1877.
- [42] I. L. Dryden and K. V. Mardia, *Statistical shape analysis*. J. Wiley Chichester, 1998, vol. 4.
- [43] D. Valdes-Amaro and A. Bhalerao, “Hierarchical contour shape analysis,” *Computación y Sistemas*, vol. 19, no. 2, pp. 233–242, 2015.
- [44] J. Victorino and F. Gómez, “A comparative study of dimensionality reduction methods for p-type based contour representations,” in *Computing Colombian Conference (10CCC), 2015 10th*, Sept 2015, pp. 294–301.
- [45] H. Iwata and Y. Ukai, “SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors.” *The Journal of heredity*, vol. 93, no. 5, pp. 384–5, 2002.
- [46] M. Du, S. Zhang, and H. Wang, “Supervised isomap for plant leaf image classification,” in *International Conference on Intelligent Computing*. Springer, 2009, pp. 627–634.

- [47] J. Liu and J. Ye, "Efficient euclidean projections in linear time," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 657–664.
- [48] G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin, "Understanding leaves in natural images—a model-based approach for tree species identification," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1482–1501, 2013.
- [49] M. Berdugo and O. Rangel, "Patrón biotipológico foliar en un gradiente de precipitación en la región caribe de colombia," vol. 26, no. 5, pp. 667–681, 2012.
- [50] P. Franco, O. Rangel-Ch, and G. Lozano, "Estudios ecológicos en la cordillera oriental, ii: Las comunidades vegetales de los alrededores de la laguna de chingaza (cundinamarca)," *Caldasia*, vol. 15, no. 71-75, pp. 219–248, 1986.
- [51] O. Rangel-Ch and G. Lozano-C, "Un perfil de vegetación entre la plata (huila) y el volcán del puracé," *Caldasia*, pp. 503–547, 1986.
- [52] E. Tafur and C. Rangel, "Caracterización preliminar de la textura foliar de la vegetación en los transectos del tatamá (cordillera occidental, colombia)," 2005.
- [53] Z. Wang, Z. Chi, and D. Feng, "Shape based leaf image retrieval," *IEEE Proceedings-Vision, Image and Signal Processing*, vol. 150, no. 1, pp. 34–43, 2003.
- [54] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. 8, no. 2, pp. 179–187, 1962.
- [55] J. Galvez and M. Canton, "Normalization and shape recognition of three-dimensional objects by 3d moments," *Pattern Recognition*, vol. 26, no. 5, pp. 667–681, 1993.
- [56] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics Springer, Berlin, 2001, vol. 1.
- [57] N. Hunt and S. Tyrrell, "Stratified sampling," *Retrieved November*, vol. 10, p. 2012, 2001.
- [58] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, 2006.
- [59] H. Goëau, A. Joly, P. Bonnet, S. Selmi, J.-F. Molino, D. Barthélémy, and N. Boujemaa, "LifeCLEF plant identification task 2014," in *CLEF2014 Working Notes. Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014*. CEUR-WS, 2014, pp. 598–615.
- [60] T. Gaber, A. Tharwat, V. Snasel, and A. E. Hassanien, "Plant identification: Two dimensional-based vs. one dimensional-based feature extraction methods," in *10th International Conference on Soft Computing Models in Industrial and Environmental Applications*. Springer, 2015, pp. 375–385.

- 
- [61] H. F. Eid, “Performance improvement of plant identification model based on pso segmentation,” *International Journal of Intelligent Systems and Applications (IJISA)*, vol. 8, no. 2, p. 53, 2016.
  - [62] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, “Deep-plant: Plant identification with convolutional neural networks.”
  - [63] A. K. Reyes, J. C. Caicedo, and J. E. Camargo, “Fine-tuning deep convolutional networks for plant recognition,” in *Working notes of CLEF 2015 conference*, 2015.
  - [64] T.-L. Le, D.-T. Tran, and N.-H. Pham, “Kernel descriptor based plant leaf identification,” in *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*. IEEE, 2014, pp. 1–5.