

Shape Signature Features of Healthy and Diseased Tomato Leaves Using Contour Metrics [†]

Jazzie R. Jao ^{1,2,*}  and Edgar A. Vallar ² ¹ Department of Software Technology, College of Computer Studies, De La Salle University, 2401 Taft Avenue, Manila, Philippines² Department of Physics, College of Science, De La Salle University, 2401 Taft Avenue, Manila, Philippines; edgar.vallar@dlsu.edu.ph

* Correspondence: jazzie.jao@dlsu.edu.ph

[†] Presented at the 12th International Electronic Conference on Sensors and Applications (ECSA-12), 12–14 November 2025; Available online: <https://sciforum.net/event/ECSA-12>.

Abstract

This study investigates the role of leaf shape in detecting disease in tomato plants, grounded in the observation that plant leaves often undergo structural changes in response to infection. Healthy and diseased tomato leaves are characterized by extracting shape signature features from images and analyzing their spectral characteristics. Leaf images were captured using a Sony ZV-E10 Mark II mirrorless camera equipped with a Sigma 16mm f/1.4 DC DN lens. Each leaf was placed flat on a matte white surface under a controlled overhead photography setup. The camera was mounted at a fixed height on a tripod, and uniform illumination was achieved using two symmetrically positioned LED spotlight lamps, minimizing shadows and glare. The dataset comprises 200 samples: 100 healthy and 100 diseased tomato leaves, representing a range of morphological and pathological variations. Three primary shape metrics were extracted from the images to characterize the structural differences. (1) The Centroid Contour Distance measured the radial distances from the leaf centroid to its outer contour, (2) The Hausdorff Distance quantified the geometric dissimilarity between contours, and (3) Dice Similarity Index assessed the degree of overlap. In addition, spectral characteristics were derived from the RGB channels: mean intensities of red, green, blue, and the Excess Green Index. Results show that both shape and spectral features are valuable for detecting plant diseases: PCA show clustering patterns between the two classes of leaves and correlation analysis highlights the relationship between several pairs of geometric and color features. In conclusion, shape is an essential aspect of plant health as it reflects the structural changes that occur as a result of disease. When combined with spectral data, can form the basis for an effective, automated disease detection system.

Keywords: shape analysis; leaf morphology; tomato leaf analysis; centroid contour distance; Fourier transform



Academic Editor: Firstname

Lastname

Published:

Citation: Jao, J.R.; Vallar, E.A. Shape Signature Features of Healthy and Diseased Tomato Leaves Using Contour Metrics. *Eng. Proc.* **2025**, *1*, 0. <https://doi.org/>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plant leaves are highly sensitive indicators of health status and environmental stress. In tomato (*Solanum lycopersicum*), one of the most widely cultivated vegetable crops globally, leaf morphology is directly affected by biotic stresses such as viral, bacterial, and fungal infections. Plant pathologists have long noted that disease-induced anomalies manifest not only as color changes but also as structural irregularities such as leaf curling, vein

distortion, holes from necrosis, or asymmetric growth patterns [1,2]. These morphological alterations provide critical diagnostic cues that complement spectral signatures and are particularly useful when color differences are subtle or confounded by environmental variability. Traditional plant disease recognition approaches have primarily emphasized texture- and color-based features extracted from leaf images. While spectral indices remain valuable, they are often sensitive to illumination conditions, sensor calibration, and background noise [3]. In contrast, shape-based descriptors capture inherent geometric properties of leaves that are invariant to lighting and thus less susceptible to extrinsic noise.

Shape analysis in plant studies is well-established [4]. Early work such as Mahdikhani and Ebrahimnezhad [5] employed centroid distance and the axis of least inertia method for leaf classification, demonstrating the utility of boundary-based descriptors. More recent research has explored hybrid approaches that integrate shape with color features: for instance, Yigit et al. [6] combined geometric descriptors (e.g., length, area, perimeter) with RGB statistics to train AI-based leaf identification models. Similarly, the MASS framework (Morphological Analysis of Size and Shape) introduced by Chuanromanee et al. [7] formalized morphometric analysis to enable streamlined comparisons across botanical datasets. CCD-based methods have successfully distinguished canonical leaf classes such as elliptic, cordate, and ovate with high recognition accuracy [8,9], and have also been applied in automated plant identification systems [10,11]. Extensions of CCD incorporate Fourier analysis, where normalized radial distance signatures are transformed into the frequency domain to extract rotation-, translation-, and scale-invariant features [12,13]. Such Fourier-based descriptors allow fine-grained capture of contour variability beyond what is achievable with geometric indices alone.

Several studies have explored frequency-domain approaches for contour-based shape analysis. Lee et al. [13] developed an FFT-based leaf recognition system in which shape features—extracted as centroid-to-boundary distances—were transformed into the frequency domain. From this, they derived a compact feature set of ten descriptors incorporating both magnitude and phase information. More recently, Wu et al. [14] introduced a novel elliptic Fourier descriptor (EFD) normalization method which ensures invariance to translation, rotation, scaling, starting point selection, and symmetry transformations. The authors also released a user-friendly software tool called *ElliShape*, which combines interactive contour extraction with the new normalization approach. Earlier work by Neto et al. [15] employed elliptic Fourier analysis and discriminant function analysis to classify plant species specifically soybean, sunflower, redroot pigweed, and velvet leaf—based on leaf shape. They generated EFDs from chain-coded boundary contours and used these features in species classification tasks. Beyond CCD and Fourier techniques, morphometric frameworks have also been applied to broader taxonomic contexts. Oso and Jayeola [16] investigated exploratory and confirmatory landmark-based geometric morphometrics within the Cucurbitaceae family, a group with high leaf-shape diversity. Using *MorphoLeaf*, they generated a dataset of 140 specimens across seven species through landmark extraction, contour reparameterization, and normalization. Similarly, Prasetyo et al. [17] applied boundary-based descriptors, including CCD-derived Boundary Moments, along with texture and color features, for mango leaf variety classification. Haque and Haque [18] applied leaf image analysis to extract geometric parameters including area, perimeter, length, and width, and used these features for plant identification. Collectively, these studies highlight that morphometric descriptors whether centroid-based, frequency-based, or landmark-based, provide a robust toolkit for capturing taxonomically and biologically relevant variation in leaf morphology.

While CCD, Fourier technique, and other geometric shape descriptors have demonstrated strong utility in botanical classification tasks, their application to distinguishing

healthy versus diseased tomato leaves remains under-explored. In particular, pairwise contour similarity metrics such as the Hausdorff distance and Dice similarity index, commonly used in computer vision for object comparison, have not been widely leveraged to quantify disease-associated shape distortions or boundary irregularities in plants. To address these gaps, this study establishes a pipeline that spans controlled image acquisition, automated segmentation, and extraction of shape-based features. Shape characterization is conducted using CCD, Fourier-based descriptors, and pairwise similarity metrics, while complementary spectral indices derived from RGB channels provide color-based cues. To assess discriminability, dimensionality reduction and unsupervised embedding techniques are applied, including Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). The primary contributions of this work are as follows:

- Construction of a new dataset of tomato leaves, acquired under controlled imaging conditions, and segmentation using the Segment Anything Model (SAM).
- Comprehensive analysis of shape descriptors (CCD, FRS, Hausdorff distance, Dice similarity) to characterize healthy vs. diseased leaves.
- Evaluation of combined shape and spectral feature sets through PCA and t-SNE embeddings to investigate class separability and clustering behavior.

2. Materials and Methods

Figure 1 illustrates the pipeline adopted in this study, designed to extract shape features from tomato leaf images and enable quantitative shape comparisons across health conditions. The framework is structured around four stages: image acquisition, segmentation and contour extraction, shape extraction, and leaf-level comparison.

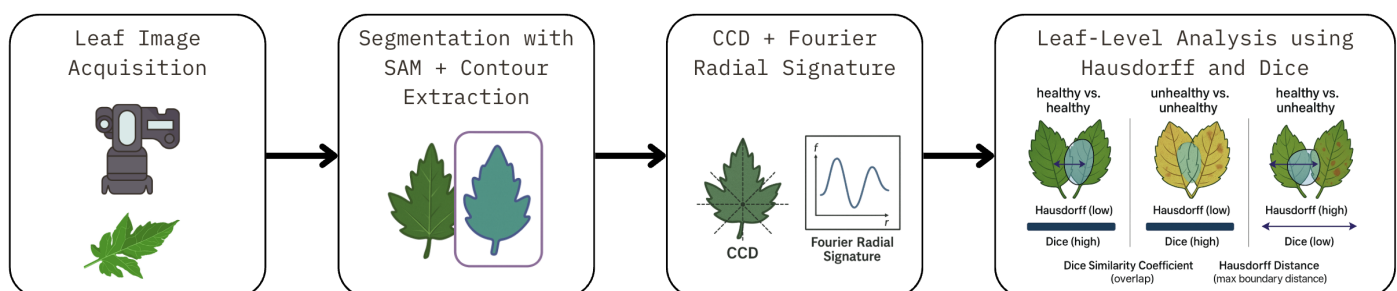


Figure 1. The Process Framework.

2.1. Data Collection

The tomato species used is *Solanum lycopersicum* L. 1753. Leaf samples are initially collected using Sony ZV-E10 Mark II mounted with Sigma 1.4 fixed aperture lens. The leaf sample is illuminated using two led light boxes to minimize shadowing and nonuniform light and is 20 cm away from the end part of the lens hood. The schematic diagram of this setup is illustrated in Figure 2. After acquisition, the leaf samples are segmented from background using Segment Anything Model algorithm specifically the `sam_vit_h`. Binary contour images were generated by segmenting each leaf to isolate the primary outline. Figure 3 shows sample instances of healthy and unhealthy leaves and the masking + contour extraction process.

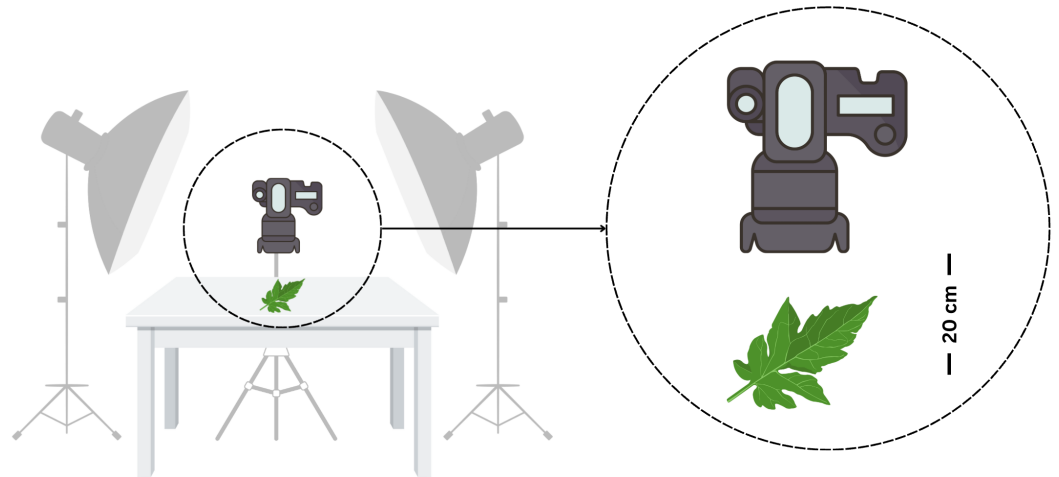


Figure 2. The Leaf Image Acquisition Setup.

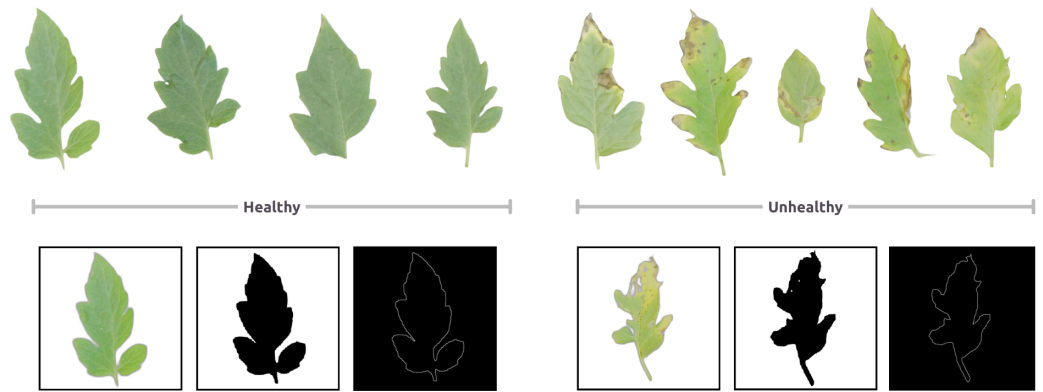


Figure 3. Several Instances of Healthy and Unhealthy Leaves and the Binary Masks.

2.2. Centroid Contour Distance (CCD) and Fourier Signatures

Given a closed contour represented as a set of 2D points $\{(x_i, y_i)\}_{i=1}^N$ and its centroid (c_x, c_y) , the CCD function $r(\theta)$ measures the distance from the centroid to the contour at uniformly sampled angles. Formally,

$$r(\theta) = \max_{\substack{(x,y) \in \text{contour} \\ \angle(x-c_x, y-c_y) = \theta}} \sqrt{(x - c_x)^2 + (y - c_y)^2}, \quad \theta \in [0, 2\pi).$$

We sample $N_\theta = 360$ angles to produce a vector $\mathbf{r} \in \mathbb{R}^{360}$. To ensure **scale invariance**, we normalize this curve by its mean radius:

$$\tilde{r}(\theta) = \frac{r(\theta)}{\bar{r}}, \quad \bar{r} = \frac{1}{2\pi} \int_0^{2\pi} r(\theta) d\theta.$$

To enforce rotation invariance, we circularly shift $\tilde{r}$ such that the maximum radius occurs at a fixed angle, aligning the longest direction across leaves. This produces normalized CCDs that reflect pure shape, independent of size or orientation. Figure 4 illustrates the process. The normalized CCD $\tilde{r}(\theta)$ is transformed via the 1D discrete Fourier transform (DFT), yielding complex coefficients F_k :

$$F_k = \frac{1}{N_\theta} \sum_{n=0}^{N_\theta-1} \tilde{r}(\theta_n) e^{-i2\pi kn/N_\theta}, \quad k = 0, \dots, N_\theta/2.$$

We exclude the DC component F_0 (mean already normalized), and retain the magnitudes $|F_k|$ of the first $K = 30$ harmonics, forming the **Fourier radial signature vector**:

$$\mathbf{s} = [|F_1|, |F_2|, \dots, |F_{30}|].$$

This representation captures periodic patterns along the contour, encoding global features (low frequencies) and local margin irregularities (high frequencies). Using magnitudes ensures phase invariance, allowing shape comparison regardless of rotational alignment.

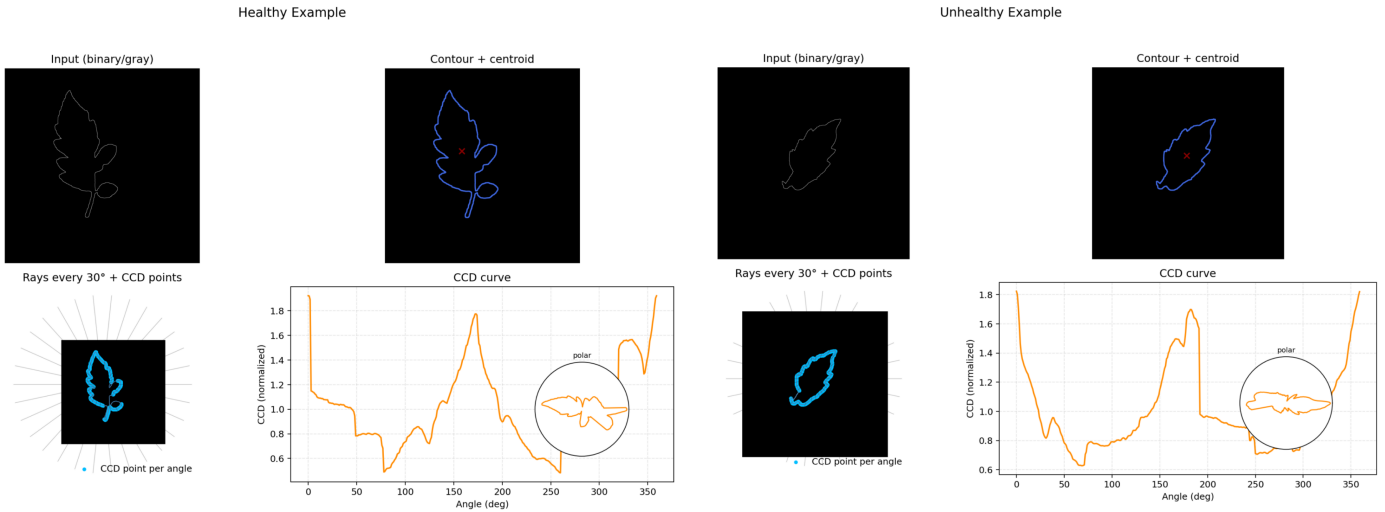


Figure 4. Normalized Centroid Contour Distance of Healthy vs. Unhealthy Leaf.

2.3. Elliptic Fourier Descriptors (EFDs)

While CCD captures radial fluctuations, it remains sensitive to local noise. To provide a complementary and robust global shape descriptor, we employed *Elliptical Fourier Descriptors* (EFD). Given a closed contour parameterized by sequential boundary coordinates $\{(x(t), y(t))\}_{t=1}^N$, the boundary can be expressed as a truncated Fourier series:

$$x(t) = a_0 + \sum_{n=1}^K [a_n \cos(\frac{2\pi nt}{T}) + b_n \sin(\frac{2\pi nt}{T})],$$

$$y(t) = c_0 + \sum_{n=1}^K [c_n \cos(\frac{2\pi nt}{T}) + d_n \sin(\frac{2\pi nt}{T})],$$

where K is the harmonic order, and (a_n, b_n, c_n, d_n) are the Fourier coefficients. Each harmonic captures progressively finer details of the outline. In this study, we fixed $K = 15$ harmonics to balance reconstruction fidelity and descriptor compactness. To ensure comparability across leaves, EFD coefficients were normalized: translation invariance by removing a_0, c_0 , scale invariance by dividing by the first harmonic's amplitude, and rotation invariance by aligning the first ellipse to the x -axis. This yields a canonical representation of each outline that is independent of position, size, or orientation. To quantify intra- and inter-group variability, we compared pairs of normalized outlines using two complementary metrics:

- **Hausdorff distance (HD):** Measures the maximum deviation between two point sets. This reflects the worst-case dissimilarity between contours.

- **Dice coefficient (DC):** Evaluates area overlap between two binary masks. For masks M_A, M_B ,

$$DC(M_A, M_B) = \frac{2|M_A \cap M_B|}{|M_A| + |M_B|}.$$

The Dice coefficient ranges from 0 (no overlap) to 1 (perfect match), capturing volumetric similarity rather than boundary extremes.

Both metrics complement each other: Hausdorff emphasizes boundary discrepancies, while Dice reflects overall shape overlap. For robust comparisons, we rasterized all EFD-reconstructed contours onto a common canvas with shared global scaling, ensuring fair Dice evaluation across groups. Figure 5 shows instance of how to leaf images of same scales are normalized for FT analysis. For both healthy and unhealthy groups, representative leaves are shown with radial rays emanating from the centroid. These rays sample distances from the centroid to the contour at fixed angular intervals, forming a 1D signature vector for each leaf.

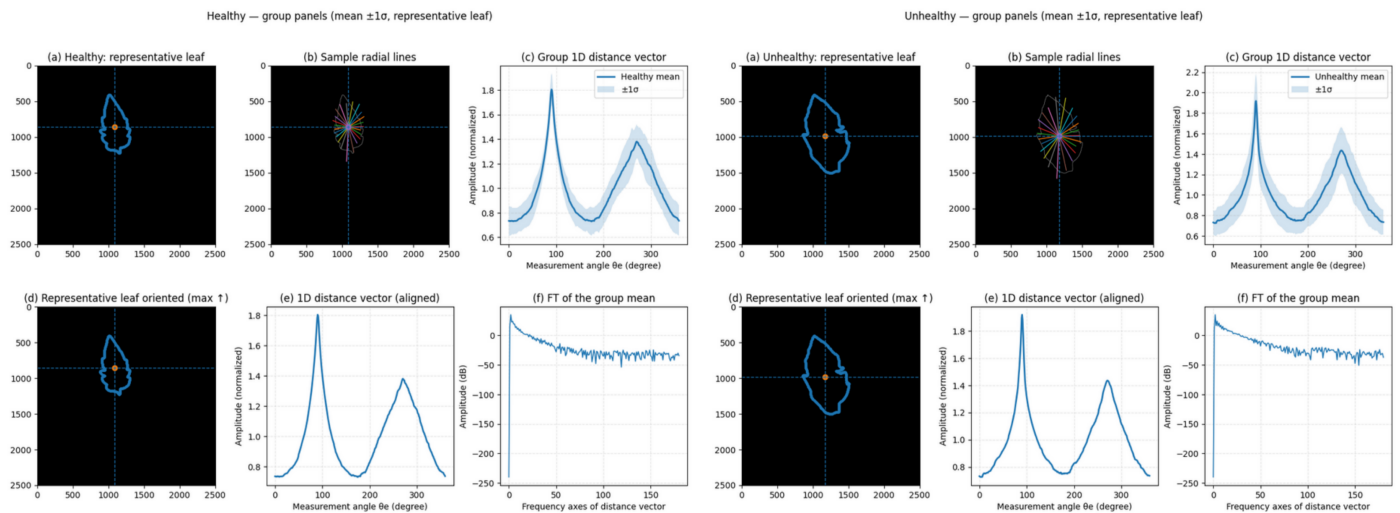


Figure 5. Two sample leaves of healthy vs. unhealthy class and their shape signal characteristics (left to right, top to bottom): (a) Representative leaf contour with centroid marked, (b) Radial rays sampled uniformly around the centroid, illustrating distance measurements to the contour. (c) Group mean radial signature (solid line) with $\pm 1\sigma$ variability (shaded region), representing shape consistency across samples, (d) Representative contour re-oriented such that the maximum radius aligns to 90° (upward), (e) Aligned group mean radial signature, and (f) Fourier magnitude spectrum of the mean signature, revealing dominant spatial frequencies associated with margin irregularities.

3. Results

3.1. CCD and EFD

Figure 7 presents that unhealthy leaves deviate from healthy leaves across wide angular ranges; the polar view emphasizes relatively larger radii for the unhealthy group in apical and basal directions. Using normalized CCD curves and their Fourier radial signatures, we evaluated differences in shape between healthy and unhealthy tomato leaves. To assess global morphological differences in the CCD domain, we performed a functional max- $|t|$ permutation test across the 360-sample CCD curves. The global p -value was $p = 0.01245$, indicating statistically significant shape differences localized at specific angular perspective. Fourier radial signatures (FRS) further enabled a frequency-domain analysis of contour shape. We computed magnitude spectra from the first 30 harmonics after discarding the DC term, capturing both global and local shape characteristics. Using the energy distance metric, we observed a substantial distributional separation between healthy and unhealthy FRS vectors ($E = 0.0043$, $p = 5.000 \times 10^{-5}$, permutation test). Additionally,

a multivariate Hotelling-like T^2 test across the 30D Fourier signature vectors revealed a highly significant group difference ($T^2 = 100.5222$, $p = 5.000 \times 10^{-5}$, permutation-based). Together, these analyses confirm that the overall radial structure of unhealthy leaves deviates significantly from healthy counterparts, both in direct spatial comparison (CCD) and in harmonic decomposition (FRS), capturing disease-related shape alterations in a scale-, rotation-, and phase-invariant manner.

Figure 6 further quantifies the differences. The leftmost image shows the aligned mean radial signature curves across all samples. The rightmost displays the corresponding Fourier magnitude spectra, where unhealthy leaves tend to exhibit stronger amplitudes at certain harmonics. A global permutation test was conducted on the full Fourier spectra to determine whether spectral profiles of the two groups differ significantly. The test yielded a global statistic of $T_{\text{obs}} = 0.550$, with a permutation p -value of 0.0096, confirming statistically significant group-level differences in the frequency domain.

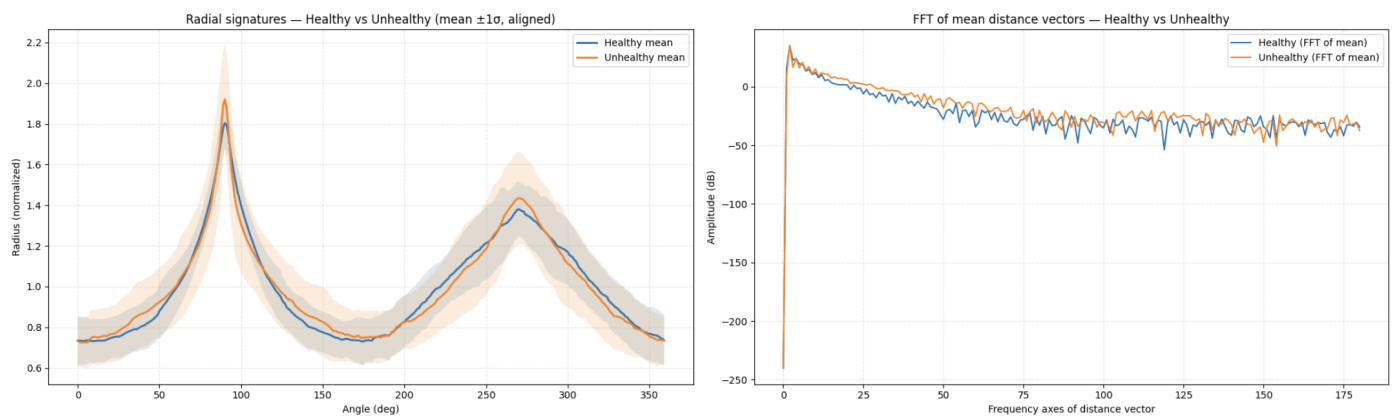


Figure 6. Radial Signature and FFT Plots (left to right): (a) Comparison of mean radial signatures between healthy (blue) and unhealthy (orange) tomato leaves, and (b) Fourier spectra of mean radial signatures for healthy and unhealthy groups. Amplitudes are shown in decibels as a function of spatial frequency (cycles per 360°).

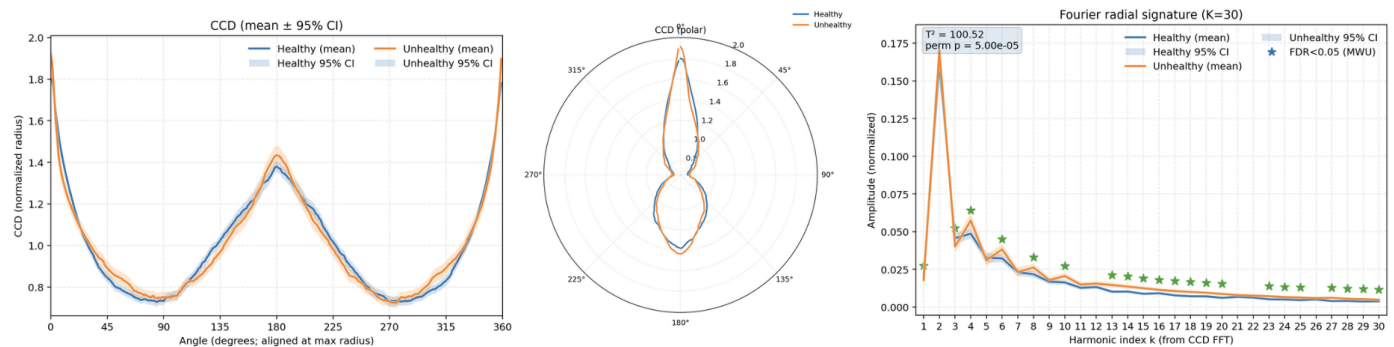


Figure 7. Mean CCD Curves (from left to right): Normalized CCD vs. Angle, CCD in Polar Coordinates, and Fourier Radial Signature of Healthy vs. Unhealthy Leaves.

We decomposed the CCD curves into Fourier descriptors (Figure 7, right) to further quantify these differences in a compact harmonic space. The first few harmonics captured the majority of the contour variation, with unhealthy leaves showing consistently larger amplitudes at low- to mid-frequency components. Energy distance testing revealed a highly significant distributional difference between groups ($E = 0.0043$, $p = 5.0 \times 10^{-5}$), confirming that the global shape signatures of unhealthy leaves are distinct. Moreover, several individual harmonics were identified as significantly different after multiple-comparison correction (green stars), suggesting that unhealthy leaves not only differ in overall shape

magnitude but also exhibit characteristic deviations in specific harmonic modes. Taken together, the CCD and Fourier analyses provide convergent evidence that unhealthy leaves exhibit systematic geometric distortions. The CCD approach highlights continuous contour-level deviations, while the Fourier descriptor analysis confirms that these deviations persist when shapes are summarized in a lower-dimensional harmonic space. These results demonstrate that leaf health status is reliably associated with quantifiable differences in contour geometry. Because CCDs are normalized by each leaf's mean radius and anchored at the centroid, the reported effects reflect shape rather than absolute size or position; uniform rescaling or translation does not affect conclusions. The Fourier-magnitude descriptors additionally remove dependence on in-plane rotation. Consequently, the detected differences implicate anisotropic deformations and localized boundary irregularities characteristic of unhealthy leaves, rather than trivial imaging factors.

3.2. Pairwise Shape Cohesion and Cross-Group Divergence

We quantified outline similarity using EFD. Prior to reconstruction we removed translation, rotation, uniform scale, and starting-point effects, so that downstream comparisons reflect shape alone. From these reconstructions we computed pairwise Hausdorff distance and Dice similarity both *within* and *across* health states as seen in Table 1 and associated box plot visualization in Figure 8. Within the healthy group, shapes were tightly clustered (Hausdorff: mean 0.305, SD 0.076; Dice: mean 0.847, SD 0.040). Unhealthy shapes were more dispersed (Hausdorff: mean 0.355, SD 0.093; Dice: mean 0.783, SD 0.068). Cross-group comparisons (healthy–unhealthy) lay between these regimes (Hausdorff: mean 0.339; Dice: mean 0.807). In practical terms, a random healthy pair was ~16% closer in Hausdorff distance than a random unhealthy pair (0.305 vs. 0.355) and overlapped by ~8% more Dice (0.847 vs. 0.783). Cross-group similarity was closer to the unhealthy distribution than to the healthy one, consistent with greater morphological heterogeneity among unhealthy leaves.

Table 1. Summary statistics for Hausdorff distance (lower is better) and Dice coefficient (higher is better).

Group	n	Mean	Std	25%	Median	75%
Hausdorff distance						
H-H	4950	0.3052	0.0759	0.2528	0.3014	0.3537
U-U	4950	0.3553	0.0932	0.2901	0.3482	0.4104
H-U	10,000	0.3388	0.0852	0.2782	0.3338	0.3924
Dice coefficient						
H-H	4950	0.8465	0.0400	0.8196	0.8492	0.8754
U-U	4950	0.7829	0.0676	0.7470	0.7921	0.8300
H-U	10,000	0.8067	0.0600	0.7750	0.8146	0.8479

These findings agree with the EFD summary: healthy leaves occupy a tighter region of normalized shape space: higher Dice, lower Hausdorff, whereas unhealthy leaves are more dispersed. Fourier-domain summaries showed larger mean amplitude and variability for unhealthy leaves (FD mean 3.23×10^5 vs. 2.67×10^5 ; FD std 1.21×10^6 vs. 1.00×10^6), and spatial-domain CCD statistics were both higher and more variable on average (centroid–distance mean 380.25 vs. 358.08; std 117.56 vs. 102.89). Together, the evidence indicates that unhealthy outlines deviate more from a compact, centrally balanced form, exhibiting greater radial unevenness and stronger low-to-mid frequency content—while healthy leaves remain more mutually consistent. Importantly, the unhealthy cohort is not uniformly “more complex” at high frequencies; it is *more variable*, that increase dispersion.

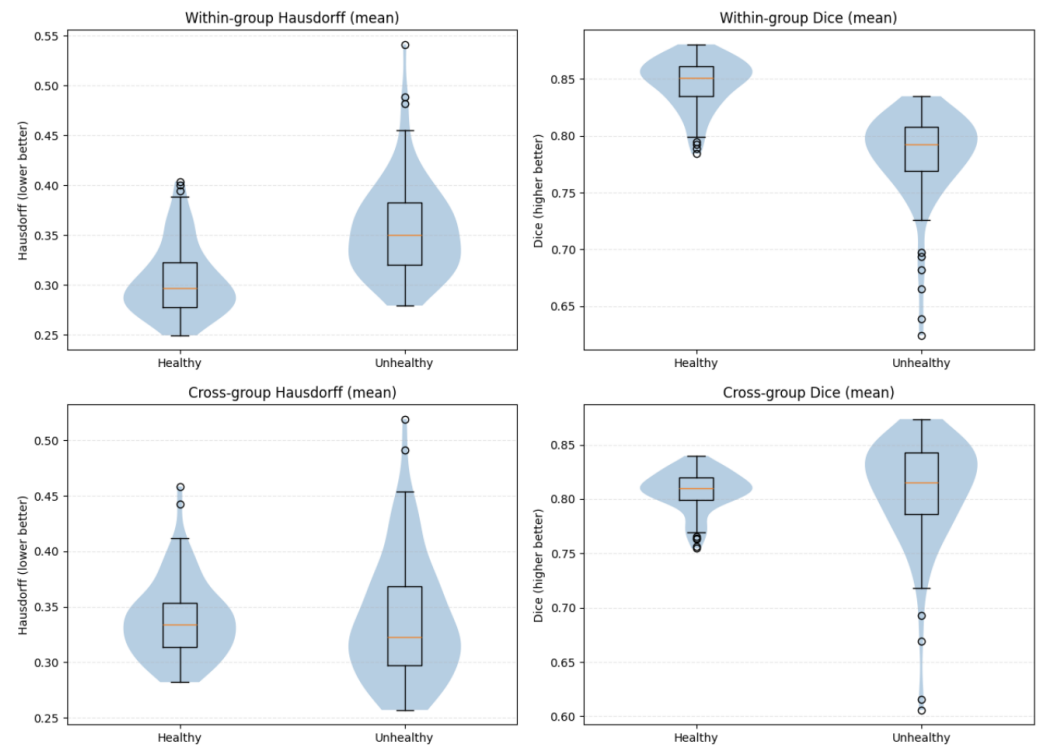


Figure 8. Box Plot of Dice and Hausdorff Distances for Different Groups.

For each leaf we also computed its mean Hausdorff and Dice to all other leaves within the same class (within-group) and to leaves from the opposite class (cross-group), yielding four per-leaf distributions. This reduces dependency from pairwise counting and supports valid inference across groups. Results are reported in Table 2. In within-group similarity, healthy leaves are more cohesive. Healthy leaves exhibited significantly higher structural cohesion, as reflected in both lower Hausdorff distances and higher Dice similarity scores. Effect sizes were large (e.g., $\delta = 0.907$ for Dice), suggesting that healthy leaf shapes are more uniform. In contrast, diseased leaves showed greater structural variability, indicating that disease disrupts shape regularity. For cross-group similarity: average proximity to the opposite class is similar. Cross-group comparisons revealed no significant differences in pairwise similarity metrics, with near-identical mean Hausdorff and Dice values across groups. Although there was a marginal trend toward higher Dice similarity within the healthy group ($p = 0.055$), overall, the shapes were not clearly separable based on these metrics alone. While shape-based descriptors capture intra-group consistency effectively, especially among healthy leaves, they offer limited discriminability between healthy and diseased groups when used in isolation. Overall, morphological stress manifests primarily as increased within-class variability rather than a uniform displacement in shape space. This has methodological implications: classifiers that rely solely on distance to a single class prototype may underperform, whereas approaches that capture shape regularity, boundary smoothness, and dispersion—potentially combined with spectral features or texture—are better aligned with the observed data.

Table 2. Statistical comparison of Healthy (A) vs. Unhealthy (B) leaves using Hausdorff distance (lower is better) and Dice coefficient (higher is better).

Setting	Group	n	Mean	Median	SD	MW U (p)	Effect Sizes/Other Tests
Within-group							
Hausdorff	A (Healthy)	100	0.3043	0.2970	0.0358	1799.0 (5.28×10^{-15})	$r_{rb} = 0.640, \delta = -0.640$ [CI: $-0.748, -0.519$],
	B (Unhealthy)	100	0.3547	0.3504	0.0471		$g = -1.202$; BM = 10.907 ($p = 1.42 \times 10^{-21}$)
Dice	A (Healthy)	100	0.8465	0.8512	0.0220	9537.0 (1.49×10^{-28})	$r_{rb} = -0.907, \delta = 0.907$ [CI: 0.848, 0.955],
	B (Unhealthy)	100	0.7829	0.7921	0.0385		$g = 2.019$; BM = -33.688 ($p = 1.32 \times 10^{-69}$)
Cross-group							
Hausdorff	A (Healthy)	100	0.3381	0.3343	0.0349	5357.0 (0.384)	$r_{rb} = -0.071, \delta = 0.071$ [CI: $-0.091, 0.233$],
	B (Unhealthy)	100	0.3381	0.3230	0.0545		$g = 0.000$; BM = -0.855 ($p = 0.394$)
Dice	A (Healthy)	100	0.8073	0.8099	0.0187	4214.0 (0.0549)	$r_{rb} = 0.157, \delta = -0.157$ [CI: $-0.327, 0.009$],
	B (Unhealthy)	100	0.8073	0.8156	0.0488		$g = 0.000$; BM = 1.826 ($p = 0.0702$)

The leftmost panel of Figure 9 displays a scatter–histogram matrix for six representative features (perimeter, area, mean R/G/B, and solidity) stratified by group (healthy in blue, unhealthy in orange). Two patterns are immediately apparent. First is size covariation. Perimeter and area are strongly linearly related (upper–left block), as expected for similarly shaped objects at different scales. Consistent with the univariate summaries reported earlier, the unhealthy cohort tends to occupy slightly larger values on both axes. Second is margin compactness. Solidity (area/convex–hull area) shows a clear negative association with perimeter and area—leaves with longer boundaries tend to be less compact (more lobed/indented). For a given size, unhealthy leaves are shifted toward *lower* solidity, matching the group means of unhealthy 0.816 vs. healthy 0.857, and indicating greater boundary irregularity in the diseased group. The RGB-channel histograms show only modest between-group shifts and substantial overlap. This aligns with the broader findings: healthy leaves are *more cohesive* (higher Dice, lower Hausdorff), whereas unhealthy leaves are *more dispersed* in shape. These dependencies motivate the use of normalized CCD and Fourier radial signatures—which decouple scale and orientation and summarize boundary fluctuations—rather than relying on raw perimeter/area alone.

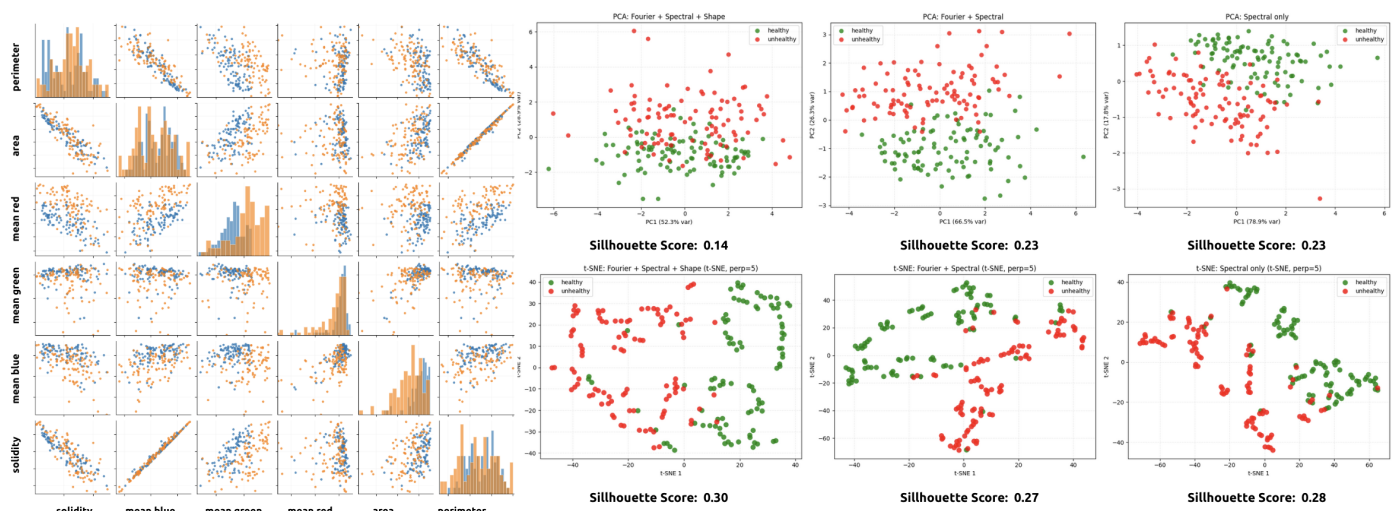
**Figure 9.** Pairplot of Select Features, PCA, and tSNE plots.

Figure 9 also illustrates the comparison of linear (PCA) and non-linear (t-SNE, perplexity = 5) embeddings for three feature sets: *Fourier+Spectral+Shape*, *Fourier+Spectral*, and *Spectral only*. In PCA, *Spectral only* attains the highest separation on the first two components (Silhouette = 0.23), matched by *Fourier+Spectral* (0.23) and exceeding *Fourier+Spectral+Shape* (0.14). PC1 explains the greatest variance for *Spectral only* (about 78.9%), followed by *Fourier+Spectral* (66.5%) and *Fourier+Spectral+Shape* (52.3%), indicating that the strongest

linear class signal lies in the RGB and ExG channels. The non-linear t-SNE projections yield moderately higher silhouettes (0.30 for *Fourier+Spectral+Shape*, 0.27 for *Fourier+Spectral*, 0.28 for *Spectral only*), showing that some class structure is non-linear and is better revealed by a manifold method than by PCA. However, the absolute silhouette values remain modest (all ≈ 0.27 – 0.30), meaning the two classes still overlap in the embedded space rather than forming perfectly separated clusters. These results indicate that there are statistically detectable group differences in radial/spectral descriptors, yet unhealthy leaves exhibit broader shape dispersion, which keeps unsupervised clusters only moderately separated.

4. Conclusions and Recommendations

This study presented a pipeline for quantifying tomato-leaf shape features from contour masks: centroid–contour distance (CCD), Fourier radial signatures (FRS), and normalized EFD reconstructions for pairwise comparisons. The analyses converge on three observations.

First, group differences exist in the radial and spectral domains. Normalized CCD curves differ by a functional max- $|t|$ permutation test ($p_{\text{global}} = 1.245 \times 10^{-2}$), and the corresponding Fourier signature vectors differ by both a Hotelling-like T^2 permutation test ($T^2 = 100.52$, $p = 5 \times 10^{-5}$) and an energy-distance test ($E = 0.0043$, $p = 5 \times 10^{-5}$). These results indicate that disease is associated with systematic changes in the radial structure and its frequency content. Second, healthy leaves are more cohesive; unhealthy leaves are more heterogeneous. Within-group EFD comparisons show markedly lower Hausdorff distances and higher Dice overlaps for healthy leaves than for unhealthy ones (e.g., mean Hausdorff 0.304 vs. 0.355; mean Dice 0.847 vs. 0.783), with large effect sizes. By contrast, *cross*-group means are similar (Hausdorff ≈ 0.338 , Dice ≈ 0.807), consistent with substantial overlap of the two populations in normalized shape space. Third, unsupervised separation is moderate. PCA on spectral statistics yields higher linear separability (silhouette ≈ 0.23) than when many shape descriptors are appended (down to ≈ 0.14). Non-linear t-SNE views increase silhouettes to ≈ 0.27 – 0.30 , but clusters remain overlapping. Pairwise feature plots further show that unhealthy leaves tend to have slightly larger perimeter/area and lower solidity, but with wide overlap.

Given the results, the study recommends a supervised fusion of spectral statistics with the shape features (CCD/FRS/EFD), since spectral features provide the clearest linear signal in PCA whereas CCD/FRS capture the statistically significant radial and frequency differences between groups. Inputs should remain rotation- and scale-invariant and inference should rely on per-leaf summaries rather than raw pairwise matrices to avoid dependency bias. Because unsupervised separation is only moderate, models that can weight features with supervision and incorporate regularization are preferable to unsupervised 2D projections. To improve generalizability, we recommend increasing the dataset size beyond the current 100 healthy and 100 unhealthy leaves. Although statistical differences were detected between groups using both shape and spectral features, the sample size may limit the robustness of more complex models. Expanding the dataset to include more intra-class variability—such as different cultivars, growth stages, or mild disease cases—could improve model sensitivity and reduce overfitting. Including additional shape descriptors such as EFDs or curvature-based metrics may further enhance discrimination, especially when combined with the current centroid- and frequency-based signatures.

Author Contributions: Conceptualization, J.R.J. and E.A.V.; methodology, J.R.J.; formal analysis, J.R.J. and E.A.V.; investigation, J.R.J. and E.A.V.; resources, J.R.J.; data curation, J.R.J.; writing—original draft preparation, J.R.J. and E.A.V.; writing—review and editing, J.R.J. and E.A.V.; visualization, J.R.J. and E.A.V.; supervision, E.A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data in this study can be accessed in this <https://github.com/jazziejao/Leaf-Shape-Signatures>, Github Repository.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Schuh, M.; educator, h.I.E.; Johnson, A.; Grabowski, M.; Orshinsky, A. Tomato viruses, 2021.
- Liu, B.; Wei, S.; Zhang, F.; Guo, N.; Fan, H.; Yao, W. Tomato leaf disease recognition based on multi-task distillation learning. *Front. Plant Sci.* **2024**, *14*, 1330527. <https://doi.org/10.3389/fpls.2023.1330527>.
- Barbedo, J.G.A. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* **2016**, *144*, 52–60. <https://doi.org/10.1016/j.biosystemseng.2016.01.017>.
- Jabal. LEAF FEATURES EXTRACTION AND RECOGNITION APPROACHES TO CLASSIFY PLANT. *J. Comput. Sci.* **2013**, *9*, 1295–1304. <https://doi.org/10.3844/jcssp.2013.1295.1304>.
- Mahdikhanlou, K.; Ebrahimnezhad, H. Plant Leaf Classification using Centroid Distance and Axis of Least Inertia Method. In Proceedings of the 2014 22nd Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 20–22 May 2014; pp. 1690–1694. <https://doi.org/10.1109/iraniancee.2014.6999810>.
- Yigit, E.; Sabanci, K.; Toktas, A.; Kayabasi, A. A study on visual features of leaves in plant identification using artificial intelligence techniques. *Comput. Electron. Agric.* **2019**, *156*, 369–377. <https://doi.org/10.1016/j.compag.2018.11.036>.
- Chuanromanee, T.S.; Cohen, J.I.; Ryan, G.L. Morphological Analysis of Size and Shape (MASS): An integrative software program for morphometric analyses of leaves. *Appl. Plant Sci.* **2019**, *7*, e11288. <https://doi.org/10.1002/aps3.11288>.
- Hasim, A.; Herdiyeni, Y.; Douady, S. Leaf Shape Recognition using Centroid Contour Distance. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *31*, 012002. <https://doi.org/10.1088/1755-1315/31/1/012002>.
- Nieto-Barajas, L.E. Leaf clustering using circular densities. *arXiv* **2022**, arXiv:2211.10547. <https://doi.org/10.48550/arxiv.2211.10547>.
- Saleem, G.; Akhtar, M.; Ahmed, N.; Qureshi, W. Automated analysis of visual leaf shape features for plant classification. *Comput. Electron. Agric.* **2019**, *157*, 270–280. <https://doi.org/10.1016/j.compag.2018.12.038>.
- Khmag, A.; Al-Haddad, S.; Kamarudin, N. Recognition System for Leaf Images Based on its Leaf Contour and Centroid. In Proceedings of the 2017 IEEE 15th Student Conference on Research and Development (SCORED), Kuala Lumpur Malaysia, 13–14 December 2017; pp. 467–472. <https://doi.org/10.1109/scored.2017.8305438>.
- Siravenha, A.C.Q.; Carvalho, S.R. Exploring the use of Leaf Shape Frequencies for Plant Classification. In Proceedings of the 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Brazil, 26–29 August 2015; pp. 297–304. <https://doi.org/10.1109/sibgrapi.2015.36>.
- Lee, K.; Chung, K.W.; Hong, K.S. An Implementation of Leaf Recognition System Based on Leaf Contour and Centroid for Plant Classification. *Lect. Notes Electr. Eng.* **2013**, *214*, 109–116. https://doi.org/10.1007/978-94-007-5857-5_12.
- Wu, H.; Yang, J.J.; Li, C.Q.; Ran, J.H.; Peng, R.H.; Wang, X.Q. Reliable and superior elliptic Fourier descriptor normalization and its application software ElliShape with efficient image processing. *arXiv* **2024**, arXiv:2412.10795. <https://doi.org/10.48550/arxiv.2412.10795>.
- Neto, J.C.; Meyer, G.E.; Jones, D.D.; Samal, A.K. Plant species identification using Elliptic Fourier leaf shape analysis. *Comput. Electron. Agric.* **2006**, *50*, 121–134. <https://doi.org/10.1016/j.compag.2005.09.004>.
- Oso, O.A.; Jayeola, A.A. Digital morphometrics: Application of MorphoLeaf in shape visualization and species delimitation, using Cucurbitaceae leaves as a model. *Appl. Plant Sci.* **2021**, *9*, e11448. <https://doi.org/10.1002/aps3.11448>.
- Prasetyo, E.; Adityo, R.D.; Suciati, N.; Fatichah, C. Partial Centroid Contour Distance (PCCD) in Mango Leaf Classification. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2020**, *10*, 1920–1926. <https://doi.org/10.18517/ijaseit.10.5.8047>.
- Haque, F.; Haque, S. PLANT RECOGNITION SYSTEM USING LEAF SHAPE FEATURES AND MINIMUM EUCLIDEAN DISTANCE. *ICTACT J. Image Video Process.* **2018**, *9*, 1919–1925. <https://doi.org/10.21917/ijivp.2018.0272>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.