Concentric Cluster Analysis: A Radial Approach for Quantifying Clustering and Scatter in Two-Dimensional Data

Eduardo Jr Piedad
Advanced Science and Technology Institute
DOST, Quezon City 1101 Philippines
eduardojr.piedad@asti.dost.gov.ph

***Abstract*—Assessing the spatial clustering or scatter of points in two-dimensional data is essential for applications across scientific, engineering, and multimedia domains. Conventional metrics, such as variance, convex hull area, and pairwise distances, provide global measures of spatial dispersion but may overlook local or radial structural nuances. This paper introduces Concentric Cluster Analysis (CCA), a novel approach that segments data into concentric rings around the centroid to quantify radial distribution patterns. The methodology is validated on both synthetic datasets and a real-world shrimp feed tray image, representing practical aquaculture monitoring. CCA generates interpretable sub-metrics from both histogram-based and cumulative-based analyses, offering granular insights into the degree and nature of clustering. Comparative experiments show that CCA complements and extends conventional spatial metrics, effectively distinguishing between compact and dispersed point patterns and revealing subtle features that may be masked by aggregate measures. The results establish CCA as a valuable tool for spatial data analysis, with demonstrated applicability to real-world multimedia image analysis and potential for broader use in pattern recognition, spatial clustering, and environmental monitoring.**

***Keywords*—Spatial analysis, clustering, data scatter, cluster metrics, radial distribution, aquaculture, feed density, shrimp**

I. Introduction

Analyzing the spatial distribution of data points plays a fundamental role in numerous fields, including spatial ecology, geography, epidemiology, and data mining. Understanding whether points are clustered, uniformly distributed, or randomly dispersed provides critical insights for decision-making, resource allocation, and further exploratory analyses [1], [2]. Various conventional metrics have been established to quantify these spatial patterns, each capturing different aspects of spatial structure and dispersion.

Classical methods such as variance and standard deviation of distances from centroids, average pairwise distances, and convex hull areas offer straightforward and intuitive measures for evaluating spatial dispersion [2]–[4]. Additionally, clustering-based techniques, including density-based clustering (DBSCAN) [5], Silhouette analysis [6], Ripley's K-function [1], and K-means inertia [7], have been widely adopted due to their capability to identify and quantify clustering structures at varying scales. Despite their utility, these methods often have limitations, such as sensitivity to parameter selection, computational complexity with large datasets, and the necessity for prior assumptions about the data structure.

To address these limitations and provide a complementary approach, we propose the Concentric Cluster Analysis (CCA), an innovative spatial analytical method. CCA evaluates spatial patterns by systematically segmenting the study region into concentric rings around a centroid and examining the radial distribution of points. Unlike conventional methods that typically provide global measures, CCA offers detailed insights into the radial spatial structure, highlighting nuances in point distributions that might be overlooked by conventional global metrics.

This study demonstrates the utility and effectiveness of CCA by comparing its performance against conventional metrics using two synthetic datasets with distinct spatial characteristics (Data A and Data B). Data A exemplifies a tightly clustered pattern, while Data B represents a more dispersed spatial configuration. The comparative analysis highlights how CCA provides additional insights into spatial distributions, complementing and enhancing conventional methodologies.

The remainder of this paper is structured as follows. Section II presents a detailed description of the Concentric Cluster Analysis methodology, including its implementation steps and sub-metrics. Section III provides a comparative analysis between the proposed CCA approach and conventional spatial metrics, supported by both synthetic and real-world datasets. In particular, CCA is demonstrated using an image of a shrimp feed tray from an aquaculture setting, highlighting its practical application for multimedia image analysis and feed monitoring. Finally, Section IV summarizes the main findings and outlines directions for future research.

II. The Concentric Cluster Analysis (CCA)

The Concentric Cluster Analysis (CCA) is proposed as an innovative spatial analytical technique that evaluates the distribution of points by their radial positioning relative to a central centroid. This method systematically divides the spatial region into concentric rings around the centroid,

979-8-3315-6946-4/25/$31.00 ©2025 IEEE

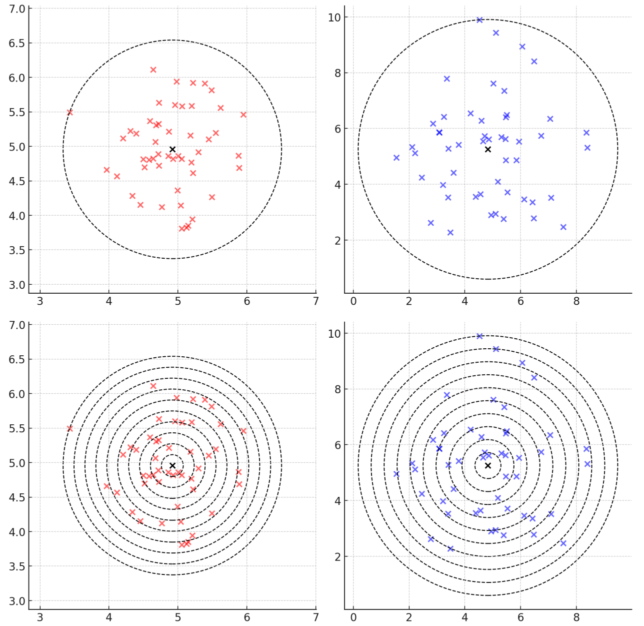
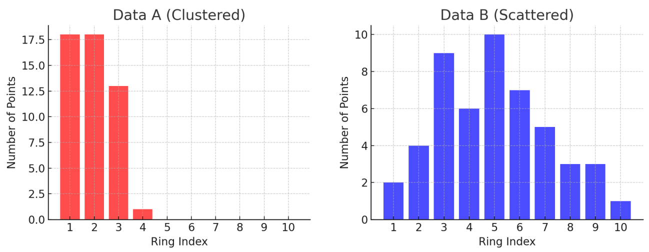


Fig. 1: Stepwise illustration of Concentric Cluster Analysis (CCA) for two synthetic datasets. Top row: Each dataset with its centroid and the largest enclosing circle. Bottom row: Division into 10 concentric rings centered at the centroid, showing the spatial segmentation used for CCA metrics.



(a) Data A

(b) Data B

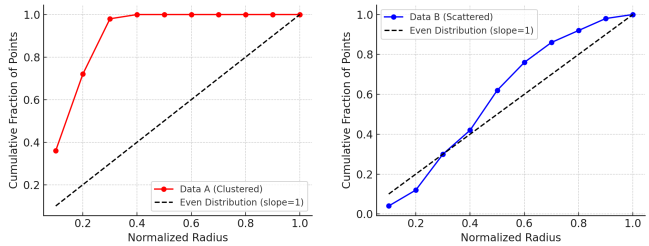
Fig. 2: CCA based on histogram for Data A and B.

enabling a nuanced understanding of spatial dispersion and clustering patterns beyond conventional global dispersion metrics. CCA provides a detailed radial distribution profile, complementing existing methods by explicitly focusing on spatial density gradients.

A. CCA Methodology

The implementation of CCA proceeds through a few steps, as illustrated in Fig. 1 with synthetic datasets Data A and Data B demonstrating different spatial characteristics– Data A exemplifies a tightly clustered spatial distribution, whereas Data B illustrates a more dispersed spatial configuration. These datasets are used throughout to illustrate the methodology.

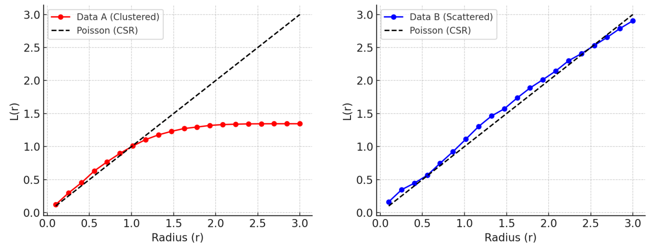
First, the centroid of the point set is computed as the mean of the coordinates. Next, the largest circle is drawn, centered at the centroid, with a radius equal to the maximum distance from the centroid to any point in the dataset. The region is then subdivided into n concentric rings (in this study, $n = 10$) by drawing circles at equally spaced radii between zero and the maximum radius. Each point is assigned to a ring based on its distance from the centroid. This stepwise process allows a visual and quantitative breakdown of radial spatial structure, providing the foundation for the subsequent histogram- and cumulative-based CCA metrics. Fig. 1 demonstrates these steps for both a tightly clustered (Data A) and a more scattered (Data B) synthetic dataset.



(a) Data A

(b) Data B

Fig. 3: CCA based on cumulative distribution for Data A and B.



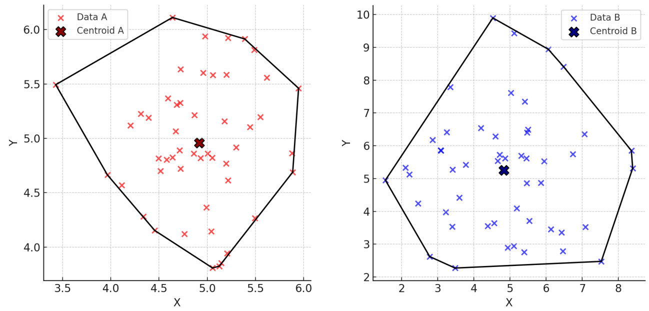
(a) Data A

(b) Data B

Fig. 4: Ripley's L-function for Data A and B

B. Histogram-based CCA

Histogram-based CCA quantifies spatial distribution by counting points within predefined concentric rings around the centroid. This approach provides discrete measurements of radial density, allowing intuitive interpretation of spatial patterns (Fig. 2).



(a) Data A

(b) Data B

Fig. 5: Convex hull for Data A and B

TABLE I: Comparison of clustering/scatter metrics for Data A (clustered) and Data B (scattered), including conventional and CCA methods with corresponding plots.

Method / Metric	Data A	Data B	Remarks
Variance	0.12	1.08	Higher = more spread from centroid
Std Dev	0.34	1.04	Higher = more spread from centroid
Avg Pairwise Distance	0.98	3.07	Higher = more scatter
Convex Hull Area	7.24	22.41	Larger = more scatter
Convex Hull Plot	Fig. 5a	Fig. 5b	Visual envelope of data
DBSCAN Clusters	1	1	More clusters if data is denser and less scattered
DBSCAN Outliers	0	10	More outliers = more scatter
Silhouette Score	0.34	0.31	Lower = less clear clustering, more scatter
Ripley's L-function	Fig. 4a	Fig. 4b	Curve close to diagonal = even spread, high curve = cluster
K-means Inertia	17.89	182.21	Higher = more spread from centers
Histogram-based CCA	Fig. 2a	Fig. 2b	CCA using points per concentric ring
Entropy	1.16	2.14	Higher = points more evenly spread across rings
Weighted-AvgRing	1.94	4.98	Higher = points farther from center on average
PeakRingIndex	1	5	Peak closer to 1 = cluster near center; higher = outer ring
Cumulative-based CCA	Fig. 3a	Fig. 3b	CCA using cumulative distribution across rings
SSE	1.71	0.10	Lower = distribution closer to even (diagonal)
RMSE	0.41	0.10	Lower = closer to even
BestFitSlope	0.47	1.16	Slope $\approx$ 1: even, $<$1: clustered center, $>$1: outer-heavy
KSDistance	0.68	0.16	Max difference from even; higher = less even

Several histogram-based sub-metrics were introduced: entropy, weighted average ring, and peak ring index. Entropy measures the randomness or uniformity of points distributed across the concentric rings, with higher entropy values indicating more uniform radial distribution. The weighted average ring metric computes the average radial distance weighted by ring indices, reflecting the average spatial dispersion of points. The peak ring index identifies the ring containing the maximum number of points, indicating the radial zone of greatest concentration.

C. Cumulative-based CCA

The cumulative-based CCA approach builds upon the cumulative distribution of points across radial distances, providing a continuous representation of radial dispersion. It involves plotting the cumulative fraction of points as a function of normalized radial distance from the centroid and comparing this curve against an ideal uniform distribution line (Fig. 3).

Several sub-metrics derived from cumulative-based analysis were considered. The sum of squared errors (SSE) and root mean squared error (RMSE) quantify the deviation from a perfect diagonal line representing uniform distribution, with lower values indicating a distribution closer to uniformity. The best-fit slope provides an intuitive measure of radial dispersion: a slope near unity indicates even radial distribution, slopes less than one suggest central clustering, and slopes greater than one suggest peripheral dispersion. Furthermore, the Kolmogorov-Smirnov (KS) distance, a widely-used metric in spatial statistics and hypothesis testing [8], [9], measures the maximum absolute difference between the empirical cumulative distribution and the theoretical uniform distribution, providing robust insights into the degree of non-uniformity.

Both histogram-based and cumulative-based approaches offer complementary perspectives. Histogram-based methods emphasize discrete radial concentrations, whereas cumulative-based methods offer smoother and continuous insights into spatial distribution patterns. Integrating both approaches enhances the robustness and interpretability of spatial distribution analyses, making CCA a valuable tool in spatial data exploration and clustering studies.

III. Comparative Demonstration of Conventional and Proposed Spatial Clustering Metrics

Figure 1 shows two synthetic datasets, designated as Data A and Data B, generated to illustrate varying degrees of spatial dispersion. Data A consists of 50 points closely clustered around a central centroid, exemplifying a dense and compact spatial configuration. In contrast, Data B, comprising another set of 50 points, displays a more dispersed pattern around a similar centroid, indicating a higher degree of spatial scatter.

To quantify and compare the spatial characteristics of these datasets, conventional spatial metrics were applied (Table I). Metrics such as variance and standard deviation from the centroid provided direct measures of point dispersion, where higher values clearly indicated greater scatter [2]. Additionally, the average pairwise distance effectively differentiated the datasets, confirming a significant increase in scatter from Data A to Data B [3]. The convex hull area further supported this observation, revealing a notably larger envelope for Data B compared to Data A, highlighting its dispersed nature (Fig. 5) [4]. Clustering methods including DBSCAN [5] and K-means inertia [7] similarly distinguished the datasets, with DBSCAN identifying more outliers in the scattered dataset and K-means demonstrating significantly higher inertia for Data B, reinforcing the observed disparity. Ripley's L-function (Fig. 4) provided insights into clustering behavior across multiple spatial scales, clearly showing Data A as clustered at smaller radii, whereas Data B approached a more random distribution [1]. Additionally, the Silhouette Score [6] indicated relatively clearer clustering for Data A.

Building upon these conventional approaches, the proposed Concentric Cluster Analysis (CCA) offered additional nuanced insights into spatial structure by analyzing data distribution across concentric rings around the centroid. Two variants of CCA were employed: histogram-based and

TABLE II: Qualitative Comparison of Spatial Metrics for Measuring Data Scatter and Clustering

Method	Description	Pros	Cons
Variance or Standard Deviation [2]	Quantifies spread by calculating variance or standard deviation of point distances from the centroid.	Simple, quick, intuitive measure of spread.	Sensitive to outliers, may misrepresent true dispersion.
Average Pairwise Distance [3]	Computes mean distance between every pair of points to assess global dispersion.	Effectively captures global scatter.	Computationally expensive for large datasets.
Convex Hull Area [4]	Measures area enclosed by smallest convex polygon covering all points; larger areas imply greater scatter.	Easy visual interpretation, intuitive spatial metric.	Sensitive to peripheral points and outliers.
DBSCAN [5]	Identifies clusters based on point density, separating tightly packed points from outliers.	Clearly identifies clusters and outliers.	Results depend heavily on parameters (eps, min_samples).
Silhouette Score [6]	Quantifies how distinctly points belong to clusters, ranging from -1 (poor) to +1 (excellent clustering).	Provides clear and widely accepted clustering quality measure.	Requires predefined cluster count (k), limited effectiveness with non-spherical clusters.
Ripley's K-function [1]	Evaluates spatial clustering across multiple distance scales, comparing observed distributions to random distributions.	Multi-scale spatial insights, detailed clustering evaluation.	Complex interpretation, dependent on spatial assumptions and edge corrections.
K-means Inertia [7]	Calculates sum of squared distances of points to nearest cluster centroids; lower inertia implies tighter clustering.	Fast, simple implementation, widely utilized.	Strongly dependent on cluster count selection, assumes spherical clusters.
Concentric Cluster Analysis (this method)	Assesses spatial dispersion by analyzing point distributions within concentric rings around the centroid.	Provides intuitive radial distribution insights, easy visual interpretation.	Sensitive to centroid positioning, requires careful ring boundary definition.

cumulative-based metrics. The histogram-based metrics (Fig. 2), including entropy, weighted average ring, and peak ring index, directly highlighted radial distribution differences, clearly delineating Data A as tightly clustered near the centroid, and Data B as significantly more dispersed across rings. Meanwhile, the cumulative-based CCA metrics (Fig. 3), such as sum squared error (SSE), root mean square error (RMSE), best-fit slope, and Kolmogorov-Smirnov distance, quantified how closely each dataset approximated a uniform radial distribution. Data A notably deviated from uniformity, reflecting its high central clustering, while Data B more closely resembled an even or outer-heavy distribution, indicative of its scattered nature.

IV. Qualitative Comparison of the Conventional and Proposed Spatial Clustering Metrics

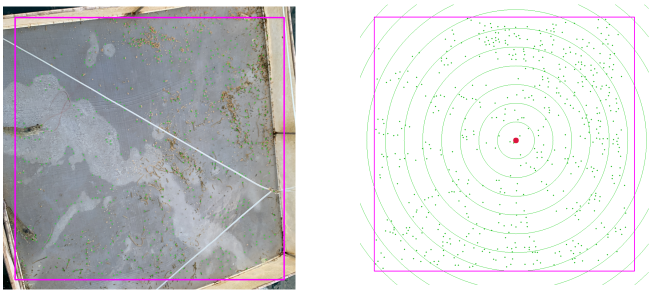
able II provides a comprehensive qualitative evaluation of several established and newly proposed metrics for quantifying the degree of clustering or scatter in two-dimensional spatial data. Conventional metrics, such as variance, standard deviation, average pairwise distance, and convex hull area, are well-regarded for their simplicity and intuitive interpretation. These approaches deliver rapid and global summaries of spatial dispersion, making them appealing for initial exploratory analysis or large-scale screening tasks. However, their reliance on aggregate measures can sometimes mask local or radial variations, especially in datasets containing outliers or non-uniform spatial patterns. For example, the convex hull area is sensitive to peripheral points, and average pairwise distance may become computationally intensive as dataset size increases.

Clustering-based methods, including DBSCAN, Silhouette Score, Ripley's K-function, and K-means inertia, further enrich spatial analysis by enabling the identification and quantification of cluster structures. DBSCAN effectively detects dense regions and outliers but is notably sensitive to parameter selection, while the Silhouette Score offers a clear and widely accepted measure of clustering quality—albeit with a reliance on a pre-specified number of clusters and optimal performance on well-separated, spherical clusters. Ripley's K-function excels at revealing multi-scale clustering but demands careful interpretation and can be affected by edge effects and spatial assumptions. K-means inertia is computationally efficient and widely used, yet its effectiveness diminishes if clusters are non-spherical or poorly separated.

In contrast, the Concentric Cluster Analysis (CCA) method advances spatial characterization by explicitly segmenting data into concentric rings centered at the centroid and analyzing the distribution of points within these segments. This radial approach affords a more nuanced view of spatial organization, providing both visual and quantitative insights into local clustering or dispersal patterns. CCA is particularly advantageous when radial distribution is of interest, as it can highlight structural features overlooked by global or aggregate statistics. Nonetheless, CCA also has limitations, including sensitivity to the placement of the centroid and the definition of ring boundaries. Despite these considerations, the complementary strengths of CCA and conventional metrics suggest that combining both approaches can lead to more robust and interpretable spatial data analyses.

V. Demonstration and Comparative Evaluation on a Shrimp Feed Tray Image

The practical value of Concentric Cluster Analysis (CCA) can be illustrated using a real-world image acquired from a shrimp feed tray, which was manually actuated and lifted above the water to monitor residual feeds. Fig. 6a presents the original tray image with detected feed particles high-



(a) Original image with detected feeds and tray border.

(b) Masked feeds, centroid, CCA rings, and tray border.

Fig. 6: Spatial analysis of feeds in a shrimp tray.

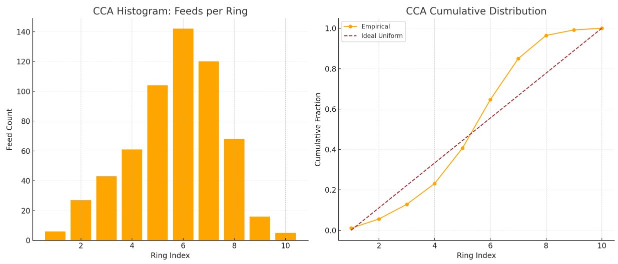


Fig. 7: CCA results for the shrimp tray image. Left: Histogram of feed counts per ring. Right: Cumulative distribution of feeds (orange) versus ideal uniform (brown dashed).

lighted in green and the tray border outlined in magenta. To provide a clearer representation of spatial distribution, Fig. 6b displays only the masked feed positions, centroid (red), concentric CCA rings (green), and tray boundary, omitting the original background.

Spatial clustering and scatter were quantitatively assessed using both the proposed CCA metrics and several conventional spatial metrics. Table III summarizes the results, including CCA sub-metrics and conventional measures such as variance of radial distances, average pairwise distance, convex hull area, and clustering indices.

The CCA histogram and cumulative distribution in Fig. 7 further reveal the spatial configuration of feeds within the tray. The histogram (left) indicates a peak concentration in rings 5 to 7, while the cumulative plot (right) shows a moderate deviation from an ideal uniform distribution, with a tendency for feeds to accumulate nearer the periphery than the center.

Conventional spatial metrics yield complementary insights. The variance of radial distances and average pairwise distance confirm an overall dispersed pattern, while the convex hull area quantifies the spread of the feed envelope. The silhouette score and DBSCAN results suggest the feed distribution forms a single main group, with minimal outlier presence. K-means inertia, computed for $k = 2$, further indicates dispersion without clear subclusters.

This comparative analysis demonstrates that CCA provides interpretable, granular metrics of radial spatial structure, distinguishing subtle distributional features that may

TABLE III: Comparative spatial metrics for feed distribution in the shrimp tray image.

Metric	**Value**	**Interpretation**
Variance (radial)	13,571	Overall radial dispersion
Avg Pairwise Distance	450.0	Global feed scatter (pixels)
Convex Hull Area	328,545	Spread of feed envelope (pixels^2)
Silhouette Score	0.29	Weak clustering (single group)
DBSCAN Clusters	1	One primary cluster (parameters: eps=30, min_samples=4)
DBSCAN Outliers	7	Few spatial outliers detected
K-means Inertia (k=2)	170,044	Dispersion relative to cluster centroids
Histogram-based CCA		
Entropy	1.97	Moderate radial spread
Weighted Avg Ring	5.71	Feeds concentrated in mid-outer rings
Peak Ring Index	6	Highest density at ring 6
Cumulative-based CCA		
SSE	0.111	Moderate deviation from uniform
RMSE	0.105	Consistent with SSE
Best-fit Slope	1.19	Outer-heavy distribution
KS Distance	0.19	Some non-uniformity present
KS Distance (CCA)	0.19	Some non-uniformity present

be masked by aggregate or cluster-based measures. The integration of both CCA and conventional metrics supports a robust, multi-perspective assessment of feed dispersal patterns relevant to aquaculture management.

VI. Conclusion

This study demonstrates the effectiveness of Concentric Cluster Analysis (CCA) for quantifying and visualizing the spatial distribution of feed particles in a real-world shrimp tray image. By systematically applying CCA, both histogram-based and cumulative-based metrics provided interpretable insights into the radial arrangement of feeds, revealing a tendency for particles to accumulate in the mid-outer regions of the tray rather than at the center.

The comparative evaluation against conventional spatial metrics, including variance, average pairwise distance, convex hull area, and clustering indices, highlights the complementary strengths of CCA. While traditional metrics summarize overall scatter and dispersion, CCA enables granular characterization of spatial structure and clustering, which are crucial for practical feed management and monitoring in aquaculture systems.

These findings confirm that CCA is a valuable addition to the toolkit for spatial pattern analysis, capable of uncovering nuanced features in two-dimensional data that may not be captured by global or cluster-based metrics alone. Integration of CCA with standard approaches supports robust, multi-perspective spatial assessments applicable not only in aquaculture but also in broader domains involving spatial pattern recognition and monitoring.

Acknowledgment

This study is supported by the ASEAN IVO 2024-04 Project Smart Aquafarm Feeder and Monitoring System in Highly Turbid Conditions (SAqFeeder), the DOST-ASTI GAA grant and the DOST S&T Fellows Program.

References

[1] B. D. Ripley, "Modelling spatial patterns," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 2, pp. 172–192, 1977.

[2] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[3] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.

[4] M. De Berg, *Computational geometry: algorithms and applications*. Springer Science & Business Media, 2000.

[5] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[6] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[7] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, vol. 5. University of California press, 1967, pp. 281–298.

[8] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[9] A. Justel, D. Peña, and R. Zamar, "A multivariate kolmogorov-smirnov test of goodness of fit," *Statistics & probability letters*, vol. 35, no. 3, pp. 251–259, 1997.