IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 October 2025, accepted 17 December 2025, date of publication 23 December 2025, date of current version 8 January 2026.

Digital Object Identifier 10.1109/ACCESS.2025.3647525

RESEARCH ARTICLE

MDAM: A Multidimensional Discriminant Analysis-Based Method for Time Series Modality Testing

IPELENG LABIUS MACHELE[1], **ADEIZA J. ONUMANYI**[2], (Member, IEEE), **ADNAN M. ABU-MAHFOUZ**[1,2], (Senior Member, IEEE), **AND ANISH KURIEN**[1], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Faculty of Engineering and the Built Environment, Tshwane University of Technology (TUT), Pretoria 0183, South Africa
[2]Next Generation Enterprises and Institutions, Council for Scientific and Industrial Research (CSIR), Pretoria 0001, South Africa

Corresponding author: Ipeleng Labius Machele (LMachele@csir.co.za)

This work was supported by the Council for Scientific and Industrial Research (CSIR) under Project KR6EDNP.

ABSTRACT In this paper, we introduce a multidimensional discriminant analysis-based method (MDAM), which is a modality testing method designed to determine whether an unknown input multidimensional time series is unimodal or multimodal. Existing unimodality testing methods face several key limitations: 1) they are primarily designed for unidimensional data and struggle with multidimensional extensions, 2) they rely on probability density function (PDF)-based approaches that fail in the presence of overlapping distributions, skewed data, and noise, and 3) they often misinterpret multimodal structures due to misleading PDF-based marginal analysis. To address these challenges, MDAM leverages a novel function that integrates the between-class mean and variance variables using a discriminant analysis approach. This distribution-independent method effectively detects modality variations across both mean and variance parameters, making it well-suited for high-dimensional and complex datasets. Comparative analysis based on synthetic and real datasets revealed that MDAM consistently outperformed five state-of-the-art techniques such as Folding, Runt, KS, DAT, and Dip, across unidimensional, multidimensional, balanced, unbalanced, unimodal, and multimodal datasets. Notably, MDAM achieved a high average accuracy of 99.8% across all dataset types, with a 20% to 40% accuracy improvement over the next-best algorithms in multimodal and mixed distributions. Its robustness across various evaluation metrics, including precision, recall, and F1 score, further establishes MDAM as a reliable tool for modality testing in time series datasets.

INDEX TERMS Automatic, distributions, multidimensional, modality, statistics.

I. INTRODUCTION

Multidimensional time-series modality testing is crucial in many scientific and engineering fields [1]. For example, problems in areas such as fault detection in voltage power distribution networks [2], multidimensional time-series classification [3], and pattern recognition in drone detection [4] often require determining whether an input dataset behaves according to a single homogeneous statistical regime or exhibits evidence of multiple distinguishable regimes [5], [6]. Such decisions influence whether clustering or mixture-based modelling is appropriate, guide anomaly detection strategies, and help determine the most suitable modelling framework for the data [7], [8], [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali.

Although several unimodality tests exist for unidimensional datasets, many rely on analysing the probability density function (PDF) and are primarily effective under i.i.d. (independent and identically distributed) assumptions. PDF-dependent unimodality tests such as the Dip statistic [10], KS-based tests [11], and discriminant analysis-based tests such as DAT [12] often struggle when the underlying distribution experiences localised changes in mean or

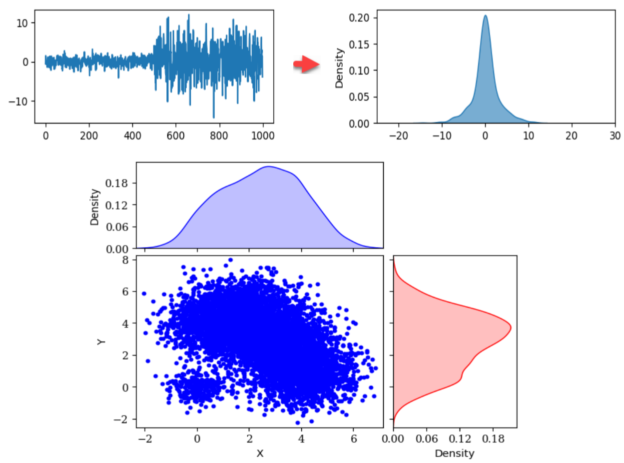


FIGURE 1. Examples demonstrating that PDF-based methods can fail in both 1D and 2D input datasets. Top: Time-series PDF reveals unimodality whereas the actual dataset is multimodal due to sudden changes in variance. Bottom: Multidimensional data is multimodal, but marginal PDFs appear unimodal, masking the true structure.

variance, or when the data exhibits overlapping substructures. As illustrated in Fig.1 (top), abrupt changes in variance may yield a unimodal PDF estimate despite the presence of within-window distributional heterogeneity. Similarly, for multidimensional data, marginal projections frequently mask multimodal structure present in the joint distribution (Fig.1, bottom), leading to misleading conclusions when relying solely on density-based or marginal analyses [13]. These limitations become even more pronounced in complex or skewed datasets [14], where noise and distributional irregularities degrade the performance of classical PDF-based methods.

Many existing approaches are also inherently restricted to 1D settings [10], making them difficult to extend to multidimensional or temporally dependent datasets. Consequently, there is a need for a modality test capable of capturing distributional heterogeneity within a time-series window, particularly changes in both the mean and variance that may indicate transitions between two distinct sub-regimes. Addressing this gap served as a central motivation for this work.

Consequently, we propose MDAM (**M**ultidimensional **D**iscriminant **A**nalysis-based **M**ethod), a window-based unimodality test for both uni- and multidimensional time-series data. Unlike classical multimodality tests that focus on detecting multiple peaks in an estimated PDF, MDAM formally defines multimodality as the presence of two statistically distinguishable distributional regimes within a window. In time-ordered data, multimodal empirical distributions typically arise from internal regime changes, hence MDAM leverages this equivalence by detecting multimodality through direct analysis of within-window distributional shifts. Drawing on principles of discriminant analysis commonly used in image segmentation, MDAM computes discriminative means and variances across all possible bipartitions of a window and aggregates them into a discriminant function whose maximum value serves as evidence against unimodality.

This formulation allows MDAM to detect unimodality violations arising from regime shifts, mixture-like behaviour, evolving system states, or abrupt parameter changes conditions under which PDF-based or marginal approaches often fail. MDAM remains distribution-agnostic by computing p-values under multiple null distributions and selecting the distribution with the largest cut-off value as a robust general reference.

Based on the above, the specific contributions of this paper are summarized as follows:

1) Unified unimodality testing: MDAM supports unimodality testing in both unidimensional and multidimensional time-series windows without requiring PDF estimation or i.i.d. assumptions.
2) New discriminant-based modality function: A novel discriminant function is proposed to accurately distinguish unimodal windows from those containing two discriminable sub-regimes, even under simultaneous changes in mean and variance conditions that challenge traditional methods.
3) Superior empirical performance: Across various datasets including unidimensional, multidimensional, balanced, unbalanced, unimodal, and multimodal types MDAM significantly outperforms five state-of-the-art techniques (Folding, Runt, KS, DAT, and Dip), achieving an average accuracy of 99.8%.

The rest of the paper is organized as follows. Section II reviews existing unimodality testing methods across uni- and multidimensional settings. Section III presents the formulation and statistical properties of the proposed MDAM approach. Section IV reports comparative experimental results demonstrating the superior performance of MDAM, and Section V concludes the paper and outlines avenues for future work.

II. RELATED WORK

This section reviews related work in the literature with focus on both unidimensional and multidimensional modality testing methods. By examining these methods, the need for our proposed approach is established.

A. UNIDIMENSIONAL-BASED METHODS

Silverman's test [15] is a classic method for unimodality assessment using kernel density estimation to find bandwidths where the number of modes changes. It indicates multimodality at smaller bandwidths and unimodality at larger ones. However, its accuracy is highly sensitive to bandwidth choice, particularly in noisy or small datasets. Building on this, the excess mass estimate [16] quantifies the deviation from a unimodal density via optimization over potential modes. Like Silverman's test, it is sensitive to noise and tuning parameters, thus limiting its use in complex datasets.

The dip test [10] measures the deviation between the empirical cumulative distribution function (CDF) and a unimodal CDF using the dip statistic. While being popular, its power drops with small samples and it assumes a uniform unimodal reference, which may not always be valid. To further advance the field, a few recent methods have been proposed including the folding test [13], which compares variances across folds, and the UU-test [8], which approximates the ECDF with a piecewise linear function. Both methods struggle with noisy, skewed, or small datasets. On the other hand, the Kolmogorov-Smirnov (K-S) test [11], although commonly used, is less effective in high-dimensional cases and depends on the choice of a suitable reference distribution. Most recently, the discriminant analysis-based test (DAT) [12] uses only the mean to detect modality changes in unidimensional time series. Consequently, it is limited in handling variance shifts and cannot be applied to multidimensional data.

Overall, these methods are restricted to one-dimensional datasets, which limits their use for more complex multidimensional applications.

B. MULTIDIMENSIONAL-BASED METHODS

Unlike unidimensional approaches, testing for modality in multidimensional time series is more challenging due to temporal dependencies, high-dimensional noise, and complex parameter interdependencies [17]. Most existing tests rely on assumptions, which are difficult to generalize across dimensions. For example, the Runt Test [18] addresses multidimensional testing by using hierarchical clustering to analyze runt sizes, i.e. gaps between clusters, in order to infer modality. While effective, it is sensitive to outliers and noise.

The Mud-pod technique [17] improves on this by projecting high-dimensional data into random subspaces, calculating Mahalanobis distances, and applying the dip test. By aggregating results across projections, the method is able to detect modality, but the approach is computationally intensive and may require many iterations for accuracy. Extensions of unidimensional methods, such as the multidimensional K-S test [11], compare sample distributions to a reference unimodal distribution. However, this requires large datasets and careful choice of reference, and is often less effective in detecting subtle deviations in high dimensions.

These limitations highlight the need for new approaches that work across both unidimensional and multidimensional domains, including the capability to detect changes in mean and variance parameters, being robust to noise, distribution-independent, and requiring minimal hyperparameter tuning. Developing such an encompassing approach served to inspire the algorithm introduced in the current paper.

Lastly, we note that in multidimensional settings, testing based on marginal distributions may be influenced by cross-dimensional dependence, particularly when variables exhibit strong correlation or shared latent structure. Nevertheless, marginal-based analysis remains a widely adopted and theoretically justified approach in multivariate statistics, especially for detection and screening tasks, where the objective is to identify the presence of distributional heterogeneity rather than to fully characterize the joint distribution. Classical results show that changes in the joint distribution necessarily induce changes in at least one marginal under mild conditions, making marginal testing a conservative but effective strategy for multimodality and regime-change detection. A rigorous treatment of the relationship between joint distributions, marginalization, and dependence structures can be found in [19].

III. MDAM

In this section, the mathematical foundations of the MDAM approach is outlined starting with its foundation in image processing. The construction of MDAM, the hypothesis testing framework, the cutoff values required for testing, and the method's time complexity are presented.

A. MOTIVATION

MDAM was inspired based on the theoretical background behind Otsu's method for image segmentation, which uses the between-class variance (BCV) as a key metric for threshold optimization [20]. The foundational principles of the Otsu's approach are presented below.

In the context of image analysis, Otsu's method identifies an optimal threshold, K, by first constructing a histogram of the image and then maximizing the BCV, denoted as σ_B^2, which quantifies the separation between two classes: the foreground C_0 and the background C_1. The method evaluates σ_B^2 over all possible threshold values $K = 1, 2, 3, \ldots, L$, corresponding to the range of intensity levels in the image. The BCV is mathematically expressed as [20]

$$\sigma_B^2(K) = w_0(K)w_1(K)[\mu_0(K) - \mu_1(K)]^2, \tag{1}$$

where μ_0 and μ_1 represent the means of the foreground and background pixel intensities, and w_0 and w_1 are their respective weights, respectively. The optimal threshold, K^*, is then determined by maximizing σ_B^2 as

$$K^* = \underset{1 \leq K \leq L}{\arg\max}\, \sigma_B^2(K). \tag{2}$$

The above-described theoretical framework for BCV optimization serves as the basis for the MDAM algorithm. It draws on the similarities between detecting modality changes in time series data and identifying segmentation boundaries in images. In time series analysis, modality changes signify transitions between distinct segments of the data (i.e. different input distributions within the data) much like the foreground and background separation in image processing. However, directly applying the BCV concept to multidimensional time series introduces challenges, such as excessive noise in BCV values, which fluctuate significantly across time series data points, thus making the measure unsuitable for direct modality detection. Additionally, deriving a reliable test statistic for individual time series points is inherently complex since modality changes can occur both in the mean and variance parameters. These methodological limitations motivated the development of the MDAM algorithm which serves as a solution for detecting modality changes in multidimensional time series data.

B. DEFINITION OF MULTIMODALITY IN MDAM

Before we present MDAM, it is important to note that "multimodality" as considered in MDAM does not refer to evaluating the static, global property of a probability density function evaluated independently of time. Instead, MDAM adopts a time-aware, data-driven definition where data in a window is considered multimodal if it contains multiple internally consistent distributional regimes along the time index. Such regimes may differ in mean, variance, or higher-order structure and, when aggregated, induce a multi-peaked empirical distribution.

In time-ordered data, multimodal empirical distributions typically arise from internal regime changes. For example, consider a time-ordered dataset $\{x_t\}_{t=1}^{N}$ drawn from an underlying data-generating process whose distribution may change over time. If the data within a window W are generated by a single stationary distribution $p(x)$, then the empirical probability density function (PDF) of W is unimodal under mild regularity conditions. Conversely, if the window contains two or more subsegments generated by distinct distributions $p_1(x), p_2(x), \ldots,$ with differing moments (e.g., mean or variance), then the aggregated empirical distribution can be expressed as a mixture

$$p(x) = \sum_k \pi_k p_k(x), \tag{3}$$

which generically induces a multi-peaked (multimodal) density.

Therefore, in time-indexed data, multimodality of the empirical distribution is a direct consequence of internal distributional or regime changes. Consequently, detecting multimodality in the PDF is thus equivalent to detecting structural changes in the underlying dataset. MDAM exploits this equivalence by identifying statistically significant within-window distributional splits rather than explicitly estimating the PDF.

C. HOW DOES MDAM WORK?

The pseudo-codes for MDAM are presented in Algorithms 1 and 2. The goal of Algorithm 1 is to apply a function that computes a reliable test statistic termed the *total between-class difference* measure, $\mathcal{B}$, for multidimensional time series modality testing. Unlike other PDF-based methods that rely on examining the PDF, our proposed function detects modality directly from the data by determining changes occurring simultaneously in the mean and variance parameters of the input data. Algorithm 2 then conducts the modality test based on the output of Algorithm 1. In summary, the entire MDAM algorithm proceeds by accepting the input data, initializing necessary variables, partitioning and computing statistical parameters, calculating the test statistics, and conducting the test. These steps are outlined as follows:

1) DATASET, PARTITIONING, AND INITIALIZATION

Let N denote the total number of input data points and D the dimensionality of the input dataset. The dataset, represented as $\mathbf{X} \in \mathbb{R}^{N\times D}$, is iteratively partitioned into two subsets based on a threshold index T using the principle of discriminant analysis in [20], such that: $\mathbf{C}_L$ denotes the first T points in $\mathbf{X}$, $\mathbf{C}_L = \{x_1, x_2, \ldots, x_T\}$, corresponding to the lower class demarcated at the threshold T, and $\mathbf{C}_U$ represents the remaining $N - T$ points, $\mathbf{C}_U = \{x_{T+1}, x_{T+2}, \ldots, x_N\}$, denoting the upper class. The following metrics are then initialized for each threshold T and dimension d as follows: $\mathcal{B}(T,d) \in \mathbb{R}^{N\times D}$ represents the total between-class measure, $\beta_\mu(T,d) \in \mathbb{R}^{N\times D}$ the contribution to $\mathcal{B}$ from the mean differences, $\beta_{\sigma^2}(T,d) \in \mathbb{R}^{N\times D}$ denotes the contribution to $\mathcal{B}$ from the variance differences, $p_L(T), p_U(T) \in \mathbb{R}^N$ represent the proportion of data in $\mathbf{C}_L$ and $\mathbf{C}_U$, respectively, $\mu_L(T,d), \mu_U(T,d)$ the means of $\mathbf{C}_L$ and $\mathbf{C}_U$ for dimension d, and $\sigma_L^2(T,d), \sigma_U^2(T,d)$ the variances of $\mathbf{C}_L$ and $\mathbf{C}_U$ for dimension d.

2) STATISTICAL COMPUTATION

The algorithm advances by computing the class statistics for $\mathbf{C}_L$ and $\mathbf{C}_U$ per threshold point $T \in \{1, 2, \ldots, N\}$ as follows:

a: CLASS PROPORTIONS

$$p_L(T) = \frac{T}{N}, \tag{4}$$

$$p_U(T) = 1 - p_L(T), \tag{5}$$

b: CLASS MEANS

$$\mu_L(T,d) = \frac{1}{T}\sum_{x\in\mathbf{C}_L} x_d, \tag{6}$$

$$\mu_U(T,d) = \frac{1}{N-T}\sum_{x\in\mathbf{C}_U} x_d. \tag{7}$$

c: CLASS VARIANCES

$$\sigma_L^2(T,d) = \frac{1}{T}\sum_{x\in\mathbf{C}_L} (x_d - \mu_L(T,d))^2, \tag{8}$$

$$\sigma_U^2(T,d) = \frac{1}{N-T}\sum_{x\in\mathbf{C}_U} (x_d - \mu_U(T,d))^2. \tag{9}$$

3) THE $\mathcal{B}$-MEASURE

The algorithm then computes the $\mathcal{B}$-measure based on a combination of two components, namely the mean contribution, β_μ, given as:

$$\beta_\mu(T,d) = p_L(T)\cdot p_U(T)\cdot \|\mu_L(T,d) - \mu_U(T,d)\|^2. \tag{10}$$

and the variance contribution, β_{σ^2}, obtained as:

$$\beta_{\sigma^2}(T,d) = p_L(T) \cdot p_U(T) \cdot \|\sigma_L^2(T,d) - \sigma_U^2(T,d)\|. \quad (11)$$

Thus, based on the above dual parameters, the overall $\mathcal{B}$-measure is averaged as follows:

$$\mathcal{B}(T,d) = \frac{\beta_\mu(T,d) + \beta_{\sigma^2}(T,d)}{2}. \quad (12)$$

The rationale for (12) is based on the concern that many discriminant analysis methods typically emphasize mean separation while treating variance differences only implicitly or not at all, leaving their ability to detect multimodal structures limited [21]. However, the strength of the proposed $\mathcal{B}$-measure lies in how it explicitly aggregates and averages the mean and variance contributions in (12). Other standard discriminant criteria, such as Fisher's ratio [21] and the Otsu's approach [20], rely primarily on location shifts (μ_L, μ_U) and therefore fail to capture structural changes that arise from heteroscedasticity or variance-driven effects. Conversely, variance-sensitive methods alone cannot identify modality changes driven by shifts in central tendency.

By averaging the mean contribution, $\beta_\mu(T,d)$, and the variance contribution, $\beta_{\sigma^2}(T,d)$, the measure achieves a balanced sensitivity to both forms of structural change. This approach provides three advantages:

1) **Complementarity**: It integrates location (mean) and scale (variance) differences, thus ensuring that multimodality is detectable regardless of whether changes manifest through shifts in central tendency, dispersion, or both.
2) **Stability**: The symmetric averaging prevents domination by either mean or variance effects, yielding a robust signal across diverse datasets, even when one factor is weak.
3) **Distribution Independence**: Unlike probability-density or covariance-normalized approaches, the $\mathcal{B}$-measure does not rely on Gaussian or parametric assumptions, making it adaptable to skewed, noisy, or high-dimensional data.

From a practical perspective, high values of $\mathcal{B}(T,d)$ signal substantial differences between the lower and upper classes, reflecting potential modality changes. By evaluating $\mathcal{B}(T,d)$ across all thresholds and dimensions, the method captures subtle distributional shifts that may not be detected using traditional discriminant-analysis approaches. This enables the detection of structural changes in multidimensional data distributions, providing a robust and interpretable framework for multimodality analysis in high-dimensional settings. Through the explicit integration of both mean and variance-based contributions, MDAM achieves sensitivity to complex patterns whether arising from shifts in central tendency, dispersion, or both thereby enhancing adaptability to diverse and challenging datasets.

Algorithm 1 MDAM: $\mathcal{B}$-Measure Computation

Input: Dataset $\mathbf{X} \in \mathbb{R}^{N\times D}$
Output: $\mathcal{B}(T,d)$ for all T, d
Initialization:
$\mathcal{B} \leftarrow$ zero matrix of size $N \times D$
$\beta_\mu \leftarrow$ zero matrix of size $N \times D$
$\beta_{\sigma^2} \leftarrow$ zero matrix of size $N \times D$
$p_L, p_U \leftarrow$ zero arrays of size N
$\mu_L, \mu_U \leftarrow$ zero matrices of size $N \times D$
$\sigma_L^2, \sigma_U^2 \leftarrow$ zero matrices of size $N \times D$
for $d = 1$ to D **do**
  for $T = 1$ to N **do**
    $N_L \leftarrow T$
    $N_U \leftarrow N - T$
    Data partitioning:
    $\mathbf{C}_L \leftarrow \{x_1, x_2, \ldots, x_T\}$
    $\mathbf{C}_U \leftarrow \{x_{T+1}, x_{T+2}, \ldots, x_N\}$
    if $N_L > 0$ **then**
      Compute $p_L(T)$ and $\mu_L(T,d)$ using (4), (5), (6), and (7)
      Compute $\sigma_L^2(T,d)$ using (8) and (9)
    end if
    if $N_U > 0$ **then**
      Compute $p_U(T)$ and $\mu_U(T,d)$ using (4), (5), (6), and (7);
      Compute $\sigma_U^2(T,d)$ using (8) and (9)
    end if
    $\mathcal{B}$-measure computation:
    Compute $\beta_\mu(T,d)$ using (10)
    Compute $\beta_{\sigma^2}(T,d)$ using (11)
    Compute total $\mathcal{B}(T,d)$ using (12)
  end for
end for
Return: $\mathcal{B}(T,d)$ for all T, d

4) MODALITY TESTING FRAMEWORK

The MDAM test framework, $\mathcal{M}_{\text{test}}$, is constructed in this section to evaluate whether the input dataset $\mathbf{X} \in \mathbb{R}^{N\times D}$ exhibits unimodal or multimodal characteristics. The proposed hypothesis test is performed differently for unidimensional $(D = 1)$ and multidimensional $(D > 1)$ datasets as described below, thus ensuring that MDAM is applicable to both unidimensional and multidimensional input datasets.

a: FOR D = 1

In this case, it is assumed that the dataset is unidimensional and the variable measured represents a single dimension, thus the test statistic $\mathcal{B}_{\text{stat}}$ is defined as:

$$\mathcal{B}_{\text{stat}} = \frac{\max(\mathcal{B}(T))}{\sigma_X^2}, \quad (13)$$

where σ_X^2 denotes the variance of the dataset $\mathbf{X}$, and $\max(\mathcal{B}(T))$ represents the maximum $\mathcal{B}(T,d)$ value across all possible T thresholds for $d = 1$. It is noted that (13) presents a normalized test statistic since dividing $\max(\mathcal{B}(T))$ by σ_X^2 ensures that $\mathcal{B}_{\text{stat}}$ is bounded between 0 and 1, since $\max(\mathcal{B}(T))$ cannot exceed the total variance of the input data [20].

Consequently, the unidimensional hypothesis test is defined as:

$$\mathcal{M}_{\text{test}}^{\text{UNI}} = \begin{cases} \text{Multimodal}, & \text{if } \mathcal{B}_{\text{stat}} \geq \mathcal{B}_{\text{cutoff}}, \\ \text{Unimodal}, & \text{if } \mathcal{B}_{\text{stat}} < \mathcal{B}_{\text{cutoff}}, \end{cases} \tag{14}$$

where $\mathcal{B}_{\text{cutoff}}$ is a predefined threshold discussed in the next subsection.

b: FOR $D > 1$

In this case, a multidimensional dataset is considered as the input to MDAM. The test is applied to each dimension individually (marginal analysis) and the test statistic for the d-th dimension is given by:

$$\mathcal{B}_{\text{stat}}^{(d)} = \frac{\max(\mathcal{B}(T, d))}{\sigma_{X_d}^2}, \tag{15}$$

where $\max(\mathcal{B}(T, d))$ and $\sigma_{X_d}^2$ are the maximum $\mathcal{B}(T, d)$ value across all possible T thresholds and variance respectively for the d-th dimension. Each dimension is labeled as either "Unimodal" or "Multimodal" based on the criterion $\mathcal{B}_{\text{stat}}^{(d)} \geq \mathcal{B}_{\text{cutoff}}$. Hence, the overall modality of the dataset can be mathematically expressed as:

$$\mathcal{M}_{\text{test}}^{\text{MULTI}} = \begin{cases} \text{Multimodal}, \\ \quad \text{if } \exists\, d \in \{1, 2, \ldots, D\} \text{ such that } \mathcal{B}_{\text{stat}}^{(d)} \geq \mathcal{B}_{\text{cutoff}}, \\ \text{Unimodal}, \\ \quad \text{otherwise}. \end{cases} \tag{16}$$

Using (16), MDAM checks if any dimension is classified as "Multimodal", and if at least one dimension meets this criterion, the dataset is classified as "Multimodal"; otherwise, it is classified as "Unimodal." Using this approach, MDAM can effectively identify multimodality in multidimensional datasets, thus ensuring robustness and interpretability in the decision-making process. This approach leverages the principle of marginal analysis, wherein each dimension of the dataset is evaluated individually for its modality using a combination of equations (15) and (16) in Algorithm 2. By combining these per-dimension results, the method generalizes the concept of unimodality testing from unidimensional to multidimensional datasets. This is rooted in the observation that in multidimensional datasets, modality in any one dimension can significantly affect the overall structure of the data [22].

Algorithm 2 MDAM: Modality Test Framework

Require: $\mathbf{X} \in \mathbb{R}^{N \times D}$, $\mathcal{B}$, $\mathcal{B}_{\text{cutoff}}$
Ensure: $\mathcal{M}_{\text{test}}$: "Unimodal" or "Multimodal"
  unidimensional Case
if $D = 1$ **then**
  Compute $\mathcal{B}_{\text{stat}}$ using (13)
  if $\mathcal{B}_{\text{stat}} \geq \mathcal{B}_{\text{cutoff}}$ **then**
    $\mathcal{M}_{\text{test}} \leftarrow$ "Multimodal"
  else
    $\mathcal{M}_{\text{test}} \leftarrow$ "Unimodal"
  end if
  Multidimensional Case
else
  Initialize $\mathcal{M}_{\text{test}} \leftarrow [\,]$ and $\mathcal{B}_{\text{stat}} \leftarrow [\,]$
  for $d = 1$ to D **do**
    Compute $\mathcal{B}_{\max}^{(d)} \leftarrow \max(\mathcal{B}[:, d])$
    Compute $\sigma_{X_d}^2 \leftarrow \text{var}(\mathbf{X}[:, d])$
    Compute $\mathcal{B}_{\text{stat}}^{(d)} \leftarrow \mathcal{B}_{\max}^{(d)} / \sigma_{X_d}^2$ as in (15)
    Apply 16 as follows:
    if $\mathcal{B}_{\text{stat}}^{(d)} \geq \mathcal{B}_{\text{cutoff}}$ **then**
      Append "Multimodal" to $\mathcal{M}_{\text{test}}$
    else
      Append "Unimodal" to $\mathcal{M}_{\text{test}}$
    end if
  end for
  exists $\leftarrow$ False
  for item $\in \mathcal{M}_{\text{test}}$ **do**
    if item = "Multimodal" **then**
      exists $\leftarrow$ True
      break
    end if
  end for
  if exists **then**
    $\mathcal{M}_{\text{test}} \leftarrow$ "Multimodal"
  else
    $\mathcal{M}_{\text{test}} \leftarrow$ "Unimodal"
  end if
end if
return $\mathcal{M}_{\text{test}}$

D. OBTAINING $\mathcal{B}_{\text{CUTOFF}}$ WITH STATISTICAL SIGNIFICANCE

To apply the MDAM algorithm with varying sample sizes (N) and user-specified p-values P, it is necessary to estimate cutoff values with statistical significance. This was accomplished using the Monte Carlo (MC) simulation approach from [23]. The pseudocode for estimating P associated with a specified cutoff value $\mathcal{B}_{\text{cutoff}}$ and N is provided in Algorithm 3.

Following Algorithm 3, in each MC iteration, a random sample of size N (i.e., the data length) is first generated and processed through the MDAM algorithm to compute the value of $\mathcal{B}$. This random sample may be drawn from any distribution of choice, e.g., Gaussian, exponential, uniform, or Chi-square. The process is repeated for a specified number of iterations, N_{mc}. After completing the MC process over N_{mc} iterations, the p-value P is calculated as the ratio $\frac{r}{N_{\text{mc}}}$, where r represents the number of iterations in which the computed

Algorithm 3 MDAM: Monte Carlo Simulation for $\mathcal{B}_{\text{cutoff}}$

Input: N_{mc} (number of Monte Carlo iterations), $\mathcal{B}_{\text{cutoff\_range}}$ (range of cutoff values), N_{values} (different dataset sizes)
Output: cutoff_results (dictionary with p-values for each N and $\mathcal{B}_{\text{cutoff}}$)
Initialize cutoff_results $\leftarrow$ [ ]
for each $N \in N_{\text{values}}$ **do**
  $r \leftarrow [\,]$
  for $i = 1$ **to** N_{mc} **do**
    $\mathbf{X} \leftarrow$ Generate random sample of size N `Any distribution type can be used, i.e., either Gaussian, exponential, uniform or Chi-square`
    $\mathcal{B} \leftarrow$ MDAM_algorithm($\mathbf{X}$) `i.e. using Algorithm 1 and 2`
    for each $\mathcal{B}_{\text{cutoff}} \in \mathcal{B}_{\text{cutoff\_range}}$ **do**
      if $\mathcal{B} \geq \mathcal{B}_{\text{cutoff}}$ **then**
        $r \leftarrow$ append{1}
      else
        $r \leftarrow$ append{0}
      end if
    end for
  end for
  $P \leftarrow \frac{\text{sum}(r)}{N_{\text{mc}}}$
  cutoff_results $\leftarrow \{P, N, \mathcal{B}_{\text{cutoff}}\}$
end for
Return cutoff_results

$\mathcal{B}$ exceeds a specified cutoff value $\mathcal{B}_{\text{cutoff}}$. Repeating this procedure for different values of N and across a range of cutoff values establishes the relationship between P and sample size, thereby yielding statistically significant cutoff thresholds. The resulting p-values are used to determine the corresponding cutoff values for different distribution types, as summarized in Tables 1–4. A similar procedure can be applied to generate lookup tables for other P and N values, though $P \leq 0.05$ is generally accepted as a threshold for statistical significance [24].

A natural concern is how cutoff values derived from unidimensional cases extend to multidimensional data, since the joint behavior of dimensions may not follow the same distributional assumptions. It is important to emphasize that MDAM adopts a marginal approach such that the $\mathcal{B}$-measure is computed independently for each dimension, and multimodality is assessed by aggregating evidence across these marginals rather than assuming a joint parametric model. This design avoids restrictive distributional assumptions in high-dimensional data, where dependencies can be complex or non-Gaussian. Consequently, the unidimensional cutoff values $\mathcal{B}_{\text{cutoff}}$ remain directly applicable at the dimension level, with statistical significance preserved in each marginal test while still enabling detection of multimodal structures across dimensions.

Furthermore, to guarantee distribution independence, the largest $\mathcal{B}_{\text{cutoff}}$ value for a given P and N across all tested distributions was adopted as the default preset in MDAM. This conservative strategy minimizes the likelihood of false positives in multidimensional settings and ensures robust statistical guarantees without requiring explicit modeling of cross-dimensional dependencies.

E. TIME COMPLEXITY

To compute the overall time complexity (TC) of the MDAM algorithm, we consider its two main components, namely the computation of the $\mathcal{B}$-measure (Algorithm 1) and the modality test framework (Algorithm 2). These are derived as follows:

1) $\mathcal{B}$-MEASURE COMPUTATION

Algorithm 1 computes the $\mathcal{B}$-measure for all thresholds $T \in \{1, \ldots, N\}$ across all dimensions $d \in \{1, \ldots, D\}$ and its key operations include: (1) Data Partitioning: For each threshold T, the dataset is divided into two parts, $\mathbf{C}_L$ and $\mathbf{C}_U$, which involves slicing operations requiring a TC of $\mathcal{O}(N)$. (2) Statistics Computation: The proportions (p_L, p_U) are computed in $\mathcal{O}(1)$, while means (μ_L, μ_U) and variances (σ_L^2, σ_U^2) require $\mathcal{O}(N)$ per threshold. (3) $\mathcal{B}$-measure Computation: Computing $\beta_\mu(T, d)$, $\beta_{\sigma^2}(T, d)$, and $\mathcal{B}(T, d)$ involves constant-time operations of $\mathcal{O}(N)$. Thus, for N thresholds and D dimensions, the overall time complexity of Algorithm 1 is given as $\mathcal{O}(N \cdot D \cdot N) = \mathcal{O}(N^2 \cdot D)$.

2) MODALITY TEST FRAMEWORK

In Algorithm 2, the TC for determining the modality of an input dataset using the $\mathcal{B}$-measure and its operations for unidimensional and multidimensional use cases is obtained as follows:

1) *Unidimensional Case ($D = 1$):* It computes $\mathcal{B}_{\text{stat}}$ by finding $\max(\mathcal{B})$ in $(\mathcal{O}(N))$ and the variance in $(\mathcal{O}(N))$, which is followed by a comparison with $\mathcal{B}_{\text{cutoff}}$ in $(\mathcal{O}(1))$. In this case, the total time complexity is $\mathcal{O}(N)$.
2) *Multidimensional Case ($D > 1$):* For each dimension d, the algorithm computes $\mathcal{B}_{\text{stat}}^{(d)}$ by finding $\max(\mathcal{B}[:, d])$ in $(\mathcal{O}(N))$, as well as computing variance $\sigma_{X_d}^2$ in $(\mathcal{O}(N))$, and performing the comparison in $(\mathcal{O}(1))$. Consequently, the total time complexity is $\mathcal{O}(N \cdot D)$.

3) OVERALL TIME COMPLEXITY OF MDAM

Thus, the overall time complexity of MDAM combines the complexities of Algorithm 1 and Algorithm 2 as follows: $\mathcal{O}(N^2 \cdot D) + \mathcal{O}(N \cdot D) = \mathcal{O}(N^2 \cdot D)$.

Since the $\mathcal{B}$-measure computation is the dominant contributor to the overall complexity, its quadratic dependence on N leads to higher computational costs for very large datasets. This challenge can, however, be alleviated by using smaller window sizes as demonstrated in the results

TABLE 1. MDAM cutoff values for Gaussian (mean = 0, variance = 1), $D = 1$.

N	0.2	0.1	0.05	0.025	0.01
50	0.116	0.133	0.148	0.162	0.181
100	0.078	0.089	0.099	0.107	0.117
200	0.054	0.061	0.068	0.073	0.081
500	0.034	0.039	0.043	0.047	0.051
1000	0.024	0.027	0.030	0.033	0.036

TABLE 2. MDAM cutoff values for Exponential (scale = 1), $D = 1$.

N	0.2	0.1	0.05	0.025	0.01
50	0.185	0.219	0.252	0.283	0.331
100	0.137	0.160	0.181	0.200	0.228
200	0.100	0.117	0.132	0.145	0.162
500	0.065	0.075	0.084	0.094	0.104
1000	0.046	0.054	0.060	0.066	0.074

TABLE 3. MDAM cutoff values for Uniform (0, 10), $D = 1$.

N	0.2	0.1	0.05	0.025	0.01
50	0.097	0.112	0.126	0.141	0.157
100	0.060	0.068	0.075	0.082	0.093
200	0.038	0.043	0.048	0.052	0.058
500	0.023	0.026	0.028	0.031	0.033
1000	0.016	0.018	0.020	0.021	0.023

TABLE 4. MDAM cutoff values for Chi-square (df = 2), D=1.

N	0.2	0.1	0.05	0.025	0.01
50	0.184	0.218	0.250	0.281	0.325
100	0.135	0.159	0.181	0.202	0.231
200	0.099	0.116	0.132	0.146	0.165
500	0.065	0.076	0.086	0.095	0.107
1000	0.047	0.054	0.060	0.066	0.074

section on real datasets, thereby reducing processing time. Furthermore, the linear dependence on D can allow MDAM to efficiently handle high-dimensional data since it does not scale exponentially in D.

IV. RESULTS AND DISCUSSION

This section compares the performance of the MDAM with other popular methods using datasets (synthetic and real) that include both multidimensional (unimodal and multimodal) and unidimensional (unimodal and multimodal) cases. The analysis evaluates MDAM against various competing methods, over a range of performance metrics and use cases as described below.

A. COMPETING TECHNIQUES

The performance of the MDAM approach was compared with the following methods: DAT [12], Folding [13], Runt [18], KS [11], and Dip [15]. These methods were chosen based on their unique modality test measures for both multidimensional and unidimensional cases and their publicly available source codes which reduces the risk of implementation bias. The parameter configurations reported in the respective papers were adopted to ensure reproducibility, as summarized in Table 5. Unless otherwise specified, the significance level was fixed for all the methods at $\alpha = 0.05$. For the DAT method, we used the closed-form, N-dependent critical values corresponding to $\alpha \in \{0.001, 0.01, 0.05, 0.10, 0.25\}$. For both the Dip and KS tests, the decision rule was based on comparing the computed p-values against the chosen α. The Folding test applied the standard folding statistic threshold $4\,\mathrm{Var}(|X - S|)/\mathrm{Var}(X) \geq 1 - q$, with $q = \alpha$. For MDAM, the cutoff $\mathcal{B}_{\text{cutoff}}$ was estimated via Monte Carlo simulation for each sample size N. In the multidimensional case ($D > 1$), MDAM was applied marginally to each dimension, and the final decision was determined using an "any-multimodal $\Rightarrow$ multimodal" rule.

B. EVALUATION METRICS

Five criteria were used for evaluation purposes namely accuracy (ACC), recall (RE), precision (PR), F-score, and area under the curve (AUC). Their definitions according to modality testing are as follows:

1) ACCURACY (ACC)

Accuracy measures the proportion of correctly classified unimodal and multimodal instances relative to the total number of data samples. It provides a comprehensive measure of the reliability of a method's ability to distinguish between unimodal and multimodal data samples [25]. It is computed as: $ACC = \frac{TP+TN}{TP+FP+TN+FN}$, where TP denotes true positives (correctly identified multimodal samples), TN true negatives (correctly identified unimodal samples), FP the false positives (unimodal samples misclassified as multimodal), and FN the false negatives (multimodal samples misclassified as unimodal).

2) RECALL (RE)

Recall, or sensitivity/true positive rate (TPR), represents the ability of the method to correctly identify all multimodal data samples. It measures the proportion of true multimodal samples that are successfully detected as such [25]. It is computed as:$RE = \frac{TP}{TP+FN}$, where a higher recall indicates fewer multimodal samples are missed by the method.

3) PRECISION (PR)

Precision evaluates the proportion of data samples classified as multimodal that are truly multimodal. This metric assesses the accuracy of the method in distinguishing multimodal samples, emphasizing its robustness against false positives [25]. It is calculated as:$PR = \frac{TP}{TP+FP}$, where a high precision value indicates that the method effectively minimizes false multimodal classifications.

4) F1-SCORE

The F1-Score combines precision and recall into a single metric to assess the accuracy of a method for detecting multimodal data samples [26]. It is particularly effective when the dataset has imbalanced proportions of unimodal and multimodal data samples. It is computed as: $F1\text{-}Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$, where a higher F1-Score suggests a

TABLE 5. Parameter settings and decision rules for all baselines and MDAM used in the experimental comparison.

Method	Test statistic & decision rule	Key parameters	Values used in experiments
MDAM	$\mathcal{B}_{\max} = \max_T \mathcal{B}_T$; $\mathcal{B}_{\max}/\operatorname{Var}(X) \geq \mathcal{B}_{\text{cutoff}}$	$\mathcal{B}_{\text{cutoff}}$; marginal rule ($D > 1$)	N; any-dim. crossing = M
DAT	$C = \max_k \text{BCV}(k)/\operatorname{Var}(X)$; $C \geq c_\alpha(N)$	α; $c_\alpha(N)$	$\alpha \in \{0.001, 0.01, 0.05, ...\}$; $c_\alpha(N)$
Folding	$F = 4\operatorname{Var}(\lvert X - S\rvert)/\operatorname{Var}(X)$; $F \geq 1 - q \Rightarrow$ U or M	q	$q = 0.05$
Dip	D; $p < \alpha \Rightarrow$ M or U	α	$\alpha = 0.05$
KS	D_{KS}; $p < \alpha \Rightarrow$ M, else U	α; $(\hat{\mu}, \hat{\sigma})$	$\alpha = 0.05$; $(\hat{\mu}, \hat{\sigma})$

Unless otherwise stated, $\alpha = 0.05$. For multidimensional MDAM, a marginal "any-dimension" rule is applied.

good balance between identifying multimodal samples and avoiding false positives.

5) AREA UNDER THE CURVE (AUC)

The AUC quantifies the trade-off between true positive rate (TPR) and false positive rate (FPR) over various thresholds by measuring the area under the receiver operating characteristic (ROC) curve [27]. It is typically computed as: $AUC = \int_0^1 TPR(FPR)\, d(FPR)$. An AUC of 1 represents a perfect method for distinguishing between unimodal and multimodal samples, while an AUC of 0.5 indicates performance equivalent to random guessing. Computationally, AUC is often determined using numerical methods such as the trapezoidal rule.

6) VISUALIZATION OF MDAM'S β-MEASURE WITH SYNTHETIC DATASETS

In this section, we visualize how MDAM determines the modality of an input data. For this purpose, a sample unimodal time series as well as the two problematic datasets that were shown earlier in Fig. 1 of Section I were used.

The visual representation in Fig. 2 highlights MDAM's robustness and effectiveness in detecting sample unimodal and multimodal distributions across different time series structures. In the unimodal case shown in Fig. 2(a), the method correctly identifies the absence of significant modality changes as reflected by the relatively stable $\mathcal{B}$-measure values below the $\mathcal{B}_{\text{cutoff}}$ threshold. Conversely, in the multimodal case of Fig. 2(b), the MDAM process successfully captures variations in the data structure as shown by the $\mathcal{B}$-measure surpassing the $\mathcal{B}_{\text{cutoff}}$ threshold, hence providing a reliable indication of modality changes.

The 2D case of Fig. 2(c) further emphasizes MDAM's adaptability to higher-dimensional data, where modality changes may occur along different axes. By computing marginal $\mathcal{B}$-measures and applying the modality test independently to each dimension and then subjecting to the testing framework, MDAM demonstrates its capability to detect modality changes under complex multidimensional datasets effectively.

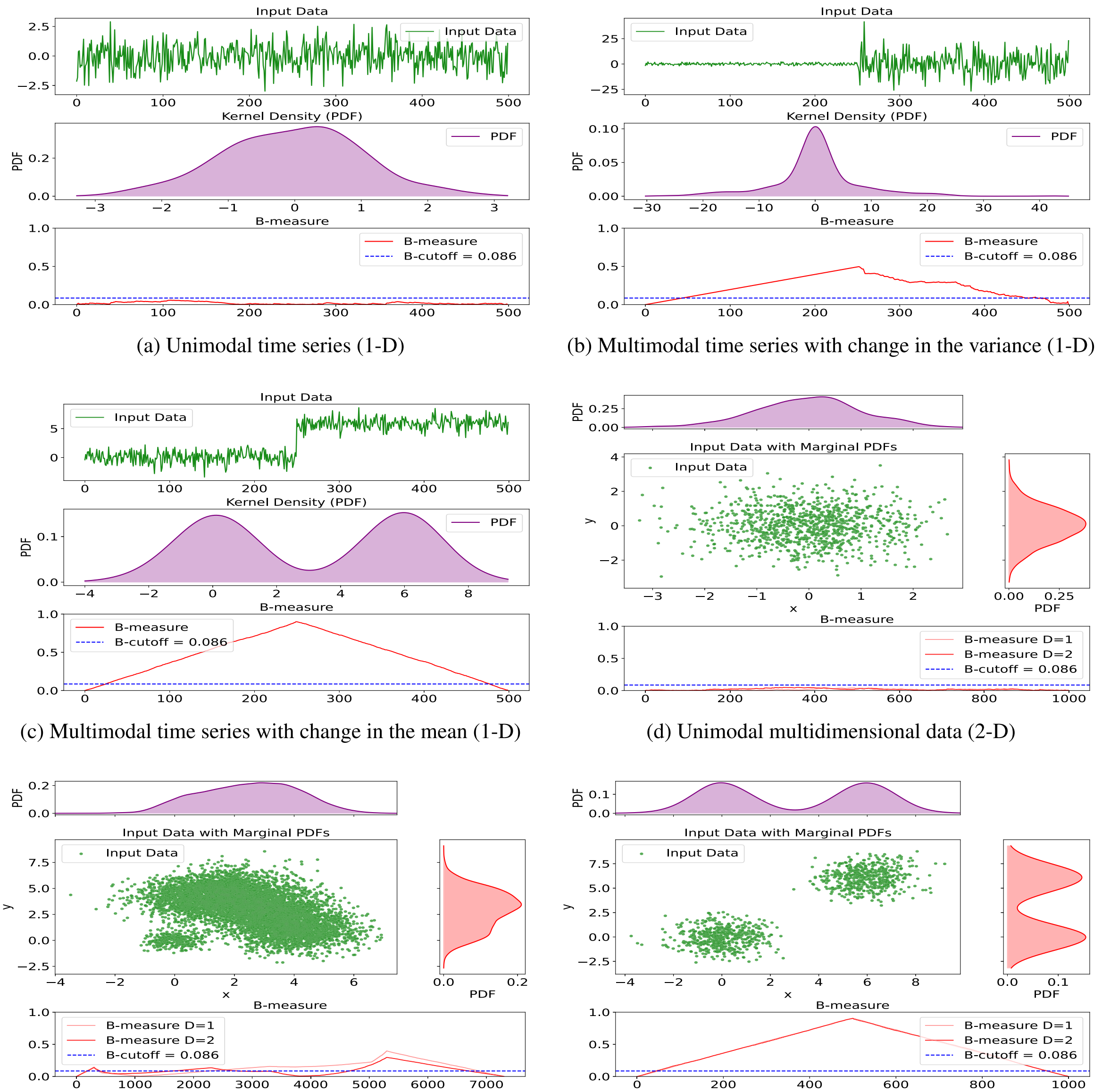
These visuals demonstrate that MDAM provides a comprehensive framework for unimodality and multimodality testing offering key advantages such as (i) Sensitivity to distribution changes, enabling accurate detection of modality shifts; (ii) Applicability to both unidimensional and multidimensional data making it suitable for real-world applications; (iii) Robustness to varying mean and variance conditions, ensuring consistent performance across different data structures. Essentially, Fig. 2 effectively validates the practical utility of MDAM in diverse scenarios, highlighting its superiority over existing methods as will be shown in the next section.

C. EXPERIMENTAL ANALYSES OF DIFFERENT SYNTHETIC CASE STUDIES

This section presents the experimental case studies used to assess the performance of the different approaches. By applying MDAM to various scenarios, the objective is to showcase its robustness, accuracy, and versatility in solving different challenges. Each case study reflects a specific set of potential conditions or complexities, which ensures that the results obtained should be both relevant and insightful. These are described as follows:

D. UNIDIMENSIONAL DATASETS

In this section, the performance of MDAM and other unidimensional-based methods are examined under different uni-dimensional input conditions considering both balanced and imbalanced (i.e. skewed) distributions. The term "balance" indicates that each subset of the data have similar sample sizes and the distribution have approximately equal spread on both sides of the central value. In the first case, the dataset contained only unimodal unidimensional cases, which allowed the methods to be tested on their ability to accurately identify unimodality across varying sample sizes, distribution types (e.g., Gaussian, uniform), and statistical parameter changes (e.g., mean, variance). Similarly, in the second case, the dataset contained only multimodal unidimensional samples to assess the methods' capability to reliably detect multimodality under the same varying conditions. The third case considered datasets containing a mixture of both unimodal and multimodal data samples unknown *a priori* to the different methods. The fourth and fifth case studied the effects of imbalanced unidimensional unimodal and multimodal datasets, respectively and in the sixth case, a combination of all the datasets applied to the different methods. These different cases ensured a focused analysis of



(a) Unimodal time series (1-D)

(b) Multimodal time series with change in the variance (1-D)

(c) Multimodal time series with change in the mean (1-D)

(d) Unimodal multidimensional data (2-D)

(e) Multimodal multidimensional data with inseparable marginals (2-D)

(f) Multimodal multidimensional data with separable marginals (2-D)

FIGURE 2. Visualization of the MDAM B-measure under different sample cases.

each method's strengths and limitations in detecting different unimodal and multimodal patterns.

1) CASE 1: "BALANCED UNIMODAL UNIDIMENSIONAL DATASETS"

This analysis focuses exclusively on unimodal datasets ($M = 1$), where M represents the number of modes with varying statistical parameter values to comprehensively evaluate the methodology's performance under different conditions of both shifting means and variances, with sample snippets illustrated in Fig. 3. To investigate the impact of data length on the results, datasets with varying means but constant variance were tested to assess sensitivity to changes in central tendency while datasets with a constant mean and varying variance were analyzed to evaluate robustness against fluctuations in data spread. The data was categorized into four segments based on different data lengths ($N = 50, 100, 500, 1000$). For instance, taking $N = 50$ as shown in the topmost sample of Fig. 3, 100 different samples were tested for $N = 50$ with linearly increasing mean (0, 4, 6, 8) and constant variance (var $= 1$), as well as constant mean ($\mu = 2$) with progressively increasing variances $(0.1, 2, 4, 6)$. These configurations were then systematically repeated and applied across all data lengths (i.e. $N = 100, 500, 1000$) to ensure consistency and provide a thorough understanding of the methodology's behavior in response to variations in data length and distribution characteristics.

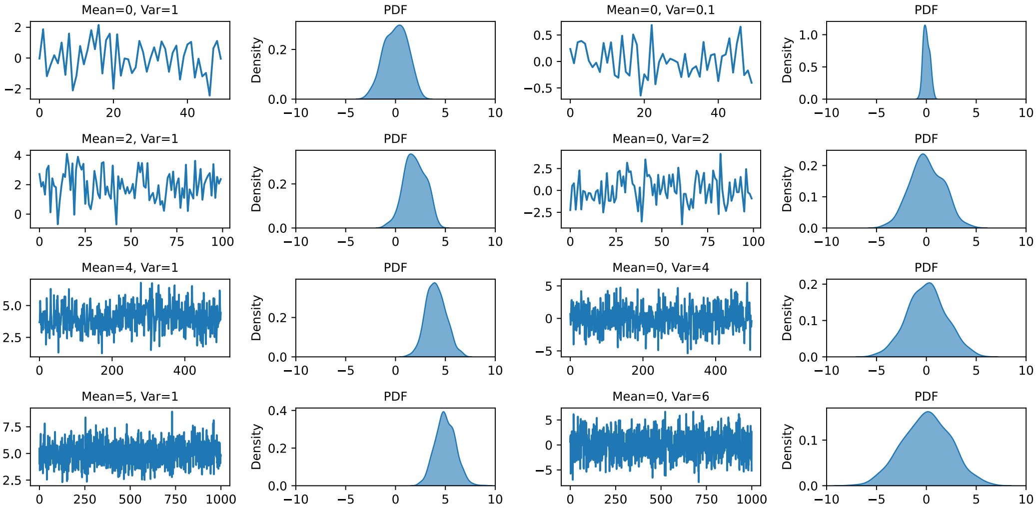
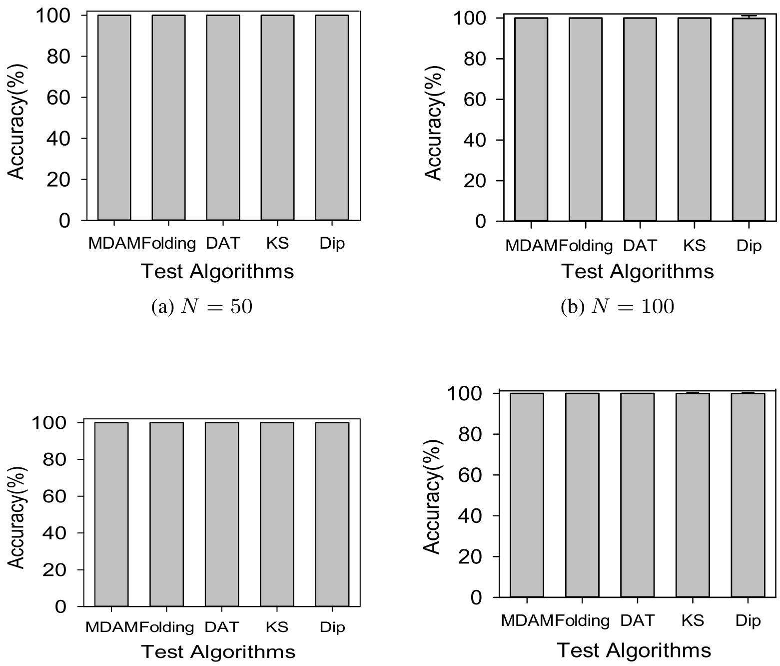


FIGURE 3. **Balanced unimodal unidimensional samples with varying characteristics and data lengths (on the x-axis), where each sample consists of a unique combination of mean and variance values.**



(a) $N = 50$ (b) $N = 100$ (c) $N = 500$ (d) $N = 1000$

FIGURE 4. **Accuracy under balanced unimodal unidimensional samples with the following dataset lengths: (a) 50 (b) 100 (c) 500 (d) 1000.**

The results obtained are shown in Fig. 4, which demonstrates that all test algorithms (MDAM, Folding, DAT, KS, and Dip) achieved near-perfect accuracy. The MDAM, Folding, and DAT methods consistently maintained 100% accuracy across all dataset sizes, highlighting their robustness. In contrast, KS and Dip displayed slight sensitivity to dataset length. For instance, the Dip's accuracy marginally decreased to 99.8% for a dataset length of 100, and both KS and Dip dropped to 99.9% for a dataset length of 1000. These minor performance variations indicate that while KS and Dip remain highly effective, they may be slightly influenced by longer datasets, suggesting a need for further investigation in scenarios with subtle data variations.

Overall, all algorithms performed well in detecting datasets with unimodal characteristics of different data lengths, as well as shifting means and variances having no significant impact on their performance. Next, the case of strictly multimodal datasets is explored to evaluate the discriminatory power of the algorithms.

2) CASE 2: "BALANCED MULTIMODAL UNIDIMENSIONAL DATASETS"

The performance of the various methods was evaluated using balanced datasets comprising only bimodal ($M = 2$) and trimodal ($M = 3$) distributions with snippets as shown in Fig. 5. Each dataset contained 1000 data points, generated by concatenating samples from normal distributions with fixed variance ($\sigma^2 = 1$) but varying means. The bimodal datasets contained equal-sized segments with mean pairs (0, 3), (0, 4), (0, 5), (0, 8), while the trimodal datasets comprised mean combinations such as (−3, 3, 10), (20, 4, 10), (0, 9, 20), (0, 8, −15), and different variances as evidenced in Fig. 5. The objective of this evaluation was to determine how effectively each algorithm could identify the underlying distribution as being multimodal across different configurations. The results obtained over 1000 Monte Carlo trials, as illustrated in Fig. 6, show that MDAM consistently outperformed the other algorithms in terms of accuracy achieving the highest accuracy of identifying the datasets as multimodal across both bimodal and trimodal samples. The MDAM's perfect accuracy remained robust even as the complexity of the dataset was increased with additional modes and shifting mean and variance values, indicating its superior capability in capturing multimodal distributions with varying statistical properties. This was followed by the Folding method with a performance of 70.06%, demonstrating relatively stable accuracy but slightly lower effectiveness, particularly in correctly identifying the trimodal distributions compared to the bimodal cases.

In Fig. 6, the DAT and KS methods showed moderate accuracy levels (i.e. 60.14% and 66.05%, respectively), with noticeable declines in performance when applied to trimodal datasets. This suggests that these methods are less effective in detecting more complex multimodal structures, potentially due to their reliance on specific statistical assumptions that do not generalize well to datasets with smaller mean and variance separations or multiple close clusters. The Dip test, while widely used in unimodality testing, exhibited average accuracy performance across all dataset configurations. Its performance was observed to be particularly poor in both the bi- and trimodal datasets where it struggled to differentiate between multiple modes effectively.

Essentially, the experimental results in Fig. 6 show the effectiveness of MDAM in detecting multimodality across different multimodal dataset configurations. The consistent performance of MDAM, particularly in scenarios with increasing complexity, was due to its ability to integrate both the mean and variance features of the dataset in its decision function (noted in eq. 12). These findings suggest that MDAM can serve as a reliable tool for multimodal detection offering a significant advantage over traditional methods such as KS and Dip tests, which show limitations in handling datasets with higher modality and diverse statistical characteristics.

TABLE 6. Performance under mixture of balanced unimodal and multimodal unidimensional datasets.

Evaluation Metrics (%)	MDAM	Folding	DAT	KS	Dip
Accuracy	100	75	66.67	75	70
Precision	100	50	33.33	50	40
Recall	100	50	33.33	50	40
F1_score	100	50	33.33	50	40
AUC	100	100	66.67	75	68

3) CASE 3: "MIXTURE OF BALANCED UNIMODAL AND MULTIMODAL UNIDIMENSIONAL DATASETS"

In this case study, datasets were randomly combined including the balanced unimodal and multimodal datasets from previous experiments to create a mixed dataset containing both unimodal and highly multimodal samples for $N = 1000$. The objective of this experiment was to evaluate the methods' performance in detecting modality changes when faced with varying statistical complexities and at a substantial long data length i.e. $N = 1000$ as commonly encountered in real-world streaming data applications such as cognitive radio monitoring stations.

The evaluation results presented in Fig. 7 on balanced mixed datasets (i.e., a mixture of 50% unimodal and 50% multimodal samples) revealed significant performance differences among the evaluated algorithms. Notably, MDAM consistently achieved perfect accuracy (100%) across this mixed scenario highlighting its robustness and adaptability in accurately detecting modality structures regardless of the dataset's complexity or composition.

Fig. 7 shows a modest improvement in accuracy for most algorithms compared to the results from the purely multimodal dataset in Fig.6 due to the augmentation with unimodal datasets. The Folding and KS methods yielded a slight performance increase, reaching an accuracy of 75.00%, followed by the Dip method at 70.00% and the DAT algorithm at 66.67%. Although these improvements indicate some adaptability to mixed distributions, a significant performance gap remains when compared to the MDAM's high accuracy, thus revealing the limitations of the other methods in handling the increased dataset complexity.

The quantitative performance analysis, summarized in Table 6, further reinforces MDAM's superiority over the other methods across all evaluation metrics. MDAM achieved the highest precision, recall, and F1-score (100%), indicating its ability to detect modality changes without false positives or negatives. In contrast, the Folding and KS methods achieved moderate performance with precision, recall, and F1-scores of 50%, while the Dip and DAT methods lagged further behind at 40% and 33.33%, respectively. Notably, MDAM also achieved a perfect area under the curve (AUC) score of 100%, compared to the next-best Folding method at 100%, while other methods trailed with AUC

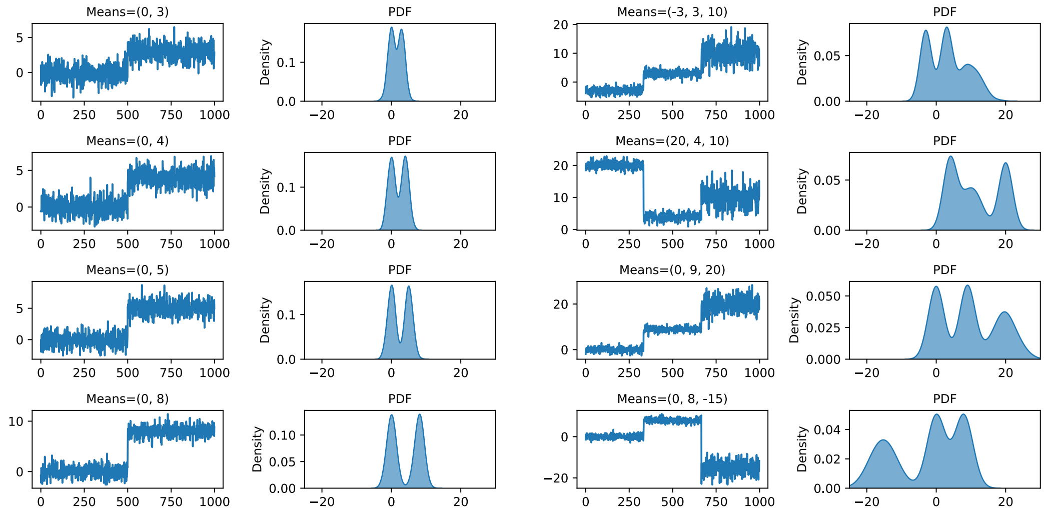


FIGURE 5. Balanced multimodal unidimensional samples with varying characteristics consisting of different combinations of mean and variances values.

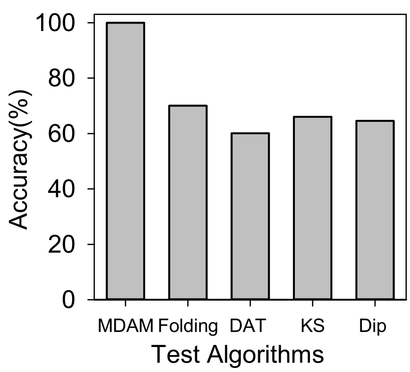


FIGURE 6. Accuracy under balanced multimodal unidimensional samples ($N = 1000$) based on a combination of 2 and 3 modes.

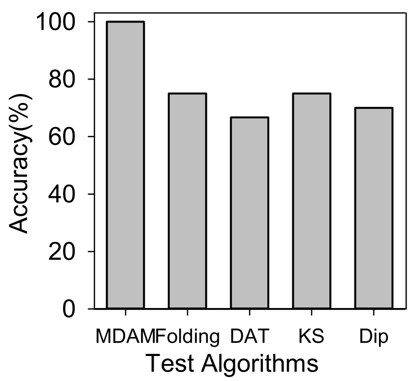


FIGURE 7. Accuracy under mixture of balanced unimodal and multimodal unidimensional samples.

values ranging from 66.67% to 75%. These results highlight MDAM's reliability in distinguishing between unimodal and multimodal structures more effectively than its counterparts.

The lower performance of the other methods can be attributed to several factors, including their sensitivity to variations in data length, inability to handle complex modality transitions, and reliance on strong parametric assumptions that may not hold in real-world scenarios. Furthermore, their limited ability to exploit statistical properties, such as distribution shape and variance, contributes to their lower accuracy as compared to MDAM, which employs a more dynamic and discriminative approach to modality detection. These findings emphasize MDAM's improved capacity to effectively handle diverse dataset conditions, hence making it a more reliable solution for applications involving complex, mixed data distributions.

4) CASE 4: "IMBALANCED UNIMODAL UNIDIMENSIONAL DATASETS"

In this section, the case of imbalanced distributions (i.e. skewed distributions) for unidimensional unimodal samples is reported on with data snippets as shown in Fig. 8. The datasets consisted of unimodal distributions ($M = 1$) with varying skewness, generated using the skew-normal distributions. Different data lengths varying mean and variance combinations. The skewness of the distributions was controlled using different shape parameters, thus allowing the creation of left- and right-skewed distribution profiles while maintaining unimodality.

The results obtained for this case are as shown in Fig. 9 considering different data lengths (50, 100, 500, and 1000). Across all dataset sizes, the MDAM algorithm consistently outperformed the other methods in terms of accuracy, achieving near-perfect classification in larger sample sizes. As the data length was increased from 50 to 1000, the accuracy of MDAM remained consistently high, showcasing its robustness in detecting unimodality even with smaller datasets. Similar performance as obtained for the Folding

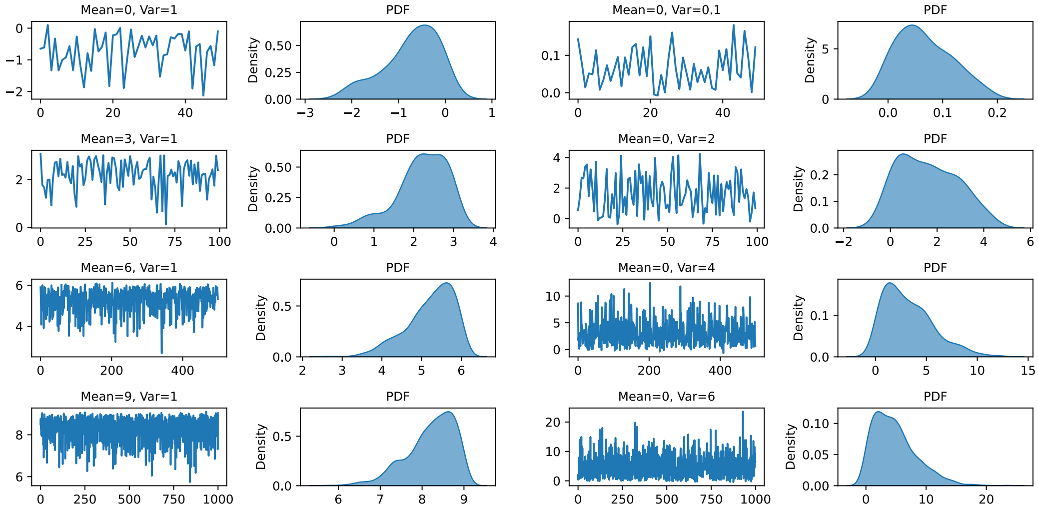
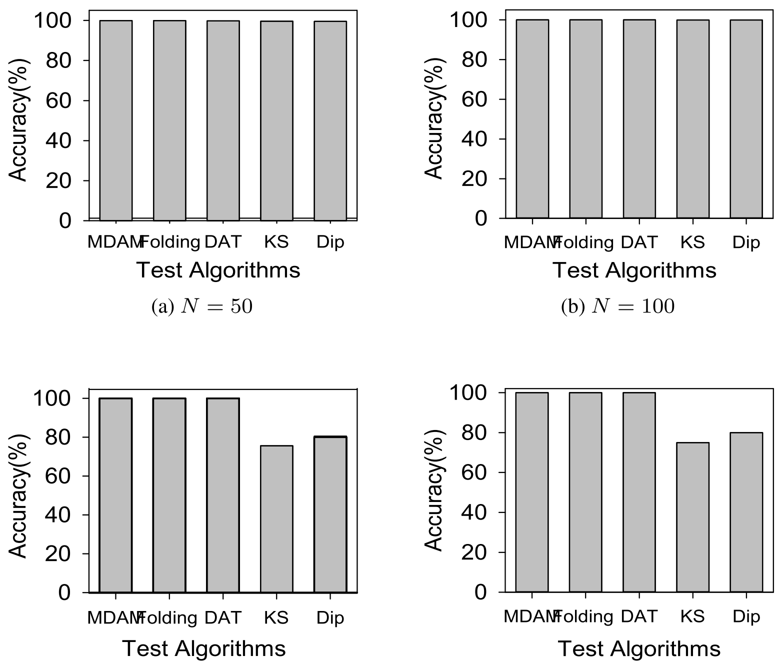


FIGURE 8. **Imbalanced (i.e. skewed) unimodal unidimensional samples with varying characteristics and data lengths, where each sample consists of a unique combination of shifting mean and variance values.**



(a) $N = 50$ (b) $N = 100$ (c) $N = 500$ (d) $N = 1000$

FIGURE 9. **Accuracy under imbalanced unimodal unidimensional datasets with the following dataset lengths: (a) 50 (b) 100 (c) 500 (d) 1000.**

and DAT highlighting their capability in identifying skewed unimodal distributions under varying statistical conditions.

On the other hand, the KS and Dip methods although highly accurate for small sample sizes showed a significant decline in accuracy with larger datasets. This drop in performance is likely due to the longer distribution tails at longer data lengths, which may have caused these methods to lean towards identifying multimodality rather than unimodality. In contrast, the MDAM, Folding, and DAT methods consistently showed improved accuracy in classifying unimodal distributions as data length increased. However, this trend was not observed with the KS and Dip methods. Overall, MDAM demonstrated superior performance across various dataset sizes, making it a reliable choice for unimodality testing in scenarios with diverse sample sizes and skewed unimodal distributions.

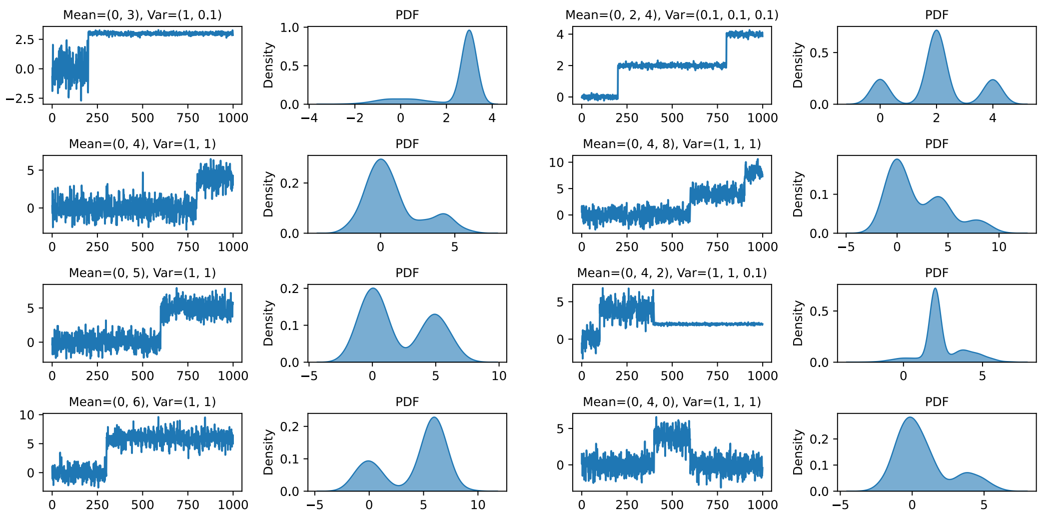


FIGURE 10. **Imbalanced multimodal unidimensional samples with varying modes ($M = 2\&3$), different statistical characteristics and data lengths, where each sample presents a unique combination of shifting mean and variance values.**

5) CASE 5: "IMBALANCED MULTIMODAL UNIDIMENSIONAL DATASETS"

This case consisted of multimodal samples, including both bi-modal ($M = 2$) and tri-modal ($M = 3$) distributions, with unequal sample proportions in each mode (i.e. imbalanced), as shown in Fig. 10. In the case of bi-modal distributions, the data was split into two segments with varying proportions. For instance, the combination ($\mu = 0, 3$) had a split ratio of 20% : 80%, while other combinations included ($\mu = 0, 4$) with 60% : 40%, ($\mu = 0, 5$) with 70% : 30%, and ($\mu = 0, 6$) with 40% : 60%. Similarly, for tri-modal distributions, the data was divided into three segments, with splits such as 20% : 60% : 20% for ($\mu = 0, 2, 4$), 60% : 30% : 10% for ($\mu = 0, 4, 8$), 10% : 30% : 60% for ($\mu = 0, 4, 2$), and 40% : 20% : 40% for ($\mu = 0, 4, 0$).

The results obtained as shown in Fig. 11 demonstrates that MDAM produced the highest accuracy, achieving near 100% detection performance. This further suggests that MDAM is highly effective in identifying multimodal structures within skewed datasets, due to its robust statistical decision function that effectively captures variations in mean and variance across the data samples. In contrast, the Folding and DAT methods exhibited moderate performance, with accuracy levels around 60-70%. These methods may face challenges in accurately detecting modes under varying skewness, as their underlying statistical assumptions are not well suited to handle such complexities.

The KS and Dip methods yielded relatively lower accuracy as compared to MDAM with performance hovering around 50-60%. This indicates potential limitations in their ability to handle skewed distributions, particularly when multiple modes with overlapping characteristics are present. One possible reason for their lower performance is their sensitivity to changes in data distribution, which may lead to misclassification when the modes are not well separated. Furthermore,

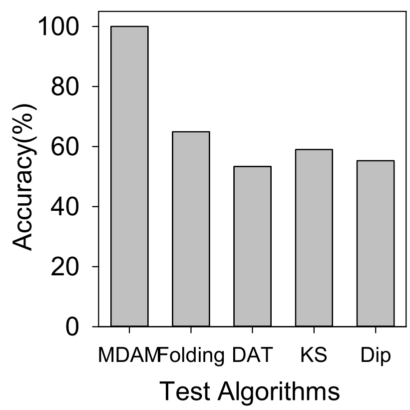


FIGURE 11. **Accuracy under imbalanced multimodal unidimensional samples ($N = 1000$) based on a combination of 2 and 3 modes.**

the presence of noise and variations in sample lengths might further impact their effectiveness in accurately identifying unimodal versus multimodal structures.

These findings further emphasize the superior performance of MDAM in handling skewed multimodal data as compared to traditional methods, which lack robustness in such scenarios.

6) CASE 6: "MIXTURE OF IMBALANCED UNIMODAL AND MULTIMODAL UNIDIMENSIONAL DATASETS"

This case presents the evaluation results obtained from testing a combination of both skewed unimodal and multimodal unidimensional datasets as presented in Fig. 12 and Table 7. Fig. 12 shows that the MDAM algorithm achieved the highest accuracy of 100%, significantly outperforming other methods such as Folding, DAT, KS, and Dip tests. The Folding and DAT tests achieved moderate performances, with accuracy levels of 79.8% and 73.13%, respectively, whereas KS and

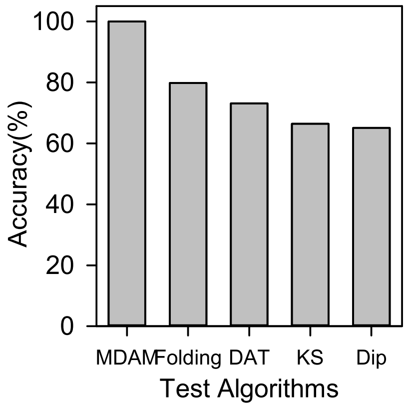


FIGURE 12. Accuracy under combination of imbalanced unimodal and multimodal unidimensional samples.

Dip tests demonstrated even lower accuracy of 66.42% and 65.1%, respectively, indicating challenges in distinguishing between unimodal and multimodal distributions when they are combined within a dataset. The superior performance of MDAM can be attributed to its robust statistical framework, which effectively captures variations in both the mean and variance parameters directly from the time series. The other methods may struggle since they operate on data from the distribution function which may consist of complex distributional characteristics and overlapping features.

Table 7 further provides a detailed breakdown of the evaluation metrics, reinforcing the findings observed in the accuracy results. It shows that MDAM achieved a perfect score across all metrics, including precision, recall, F1-score, and AUC-ROC, highlighting its capability to accurately and consistently classify samples without false positives or negatives. In contrast, the Folding and KS methods exhibited higher accuracy but suffer from lower precision and recall values, indicating potential issues with misclassification. The DAT and Dip tests showed the lowest performance across most metrics, with precision, recall, and F1-scores of 33.33% and 40%, respectively, indicating significant difficulties in handling the complexities of mixed distributions. The AUC-ROC values highlight the ability of MDAM to achieve perfect separability between unimodal and multimodal distributions, while the other methods struggle to maintain a balanced trade-off between sensitivity and specificity. These results demonstrate the effectiveness of MDAM as a robust and highly accurate solution for detecting modality in time series datasets with mixed distributions, while other methods may require further enhancements to improve their discriminatory power under such conditions.

E. MULTIDIMENSIONAL DATASETS

In this section, the results of experiments evaluating the performance of MDAM and other methods using multidimensional balanced and imbalanced samples are presented. Specifically, bi-dimensional datasets are used because they provide clear and interpretable visualizations of data distributions, enabling effective assessment of each method's ability to detect unimodal and multimodal patterns. Additionally, 2-D datasets capture essential structural complexity while maintaining computational efficiency, and dimensionality reduction techniques can always be applied to handle higher-dimensional datasets without losing critical information. Importantly, MDAM is not restricted to the 2-D case; the proposed discriminant formulation generalizes naturally to higher-dimensional time-series windows because it is based on multidimensional means and variances rather than density estimation, which allows the method to remain applicable and robust even as dimensionality increases.

TABLE 7. Performance under mixture of imbalanced unimodal and multimodal unidimensional datasets.

Evaluation Metrics (%)	MDAM	Folding	DAT	KS	Dip
Accuracy	100	75	66.67	75	70
Precision	100	50	33.33	50	40
Recall	100	50	33.33	50	40
F1_score	100	50	33.33	50	40
AUC	100	100	66.67	75	68

In the first case, the dataset consisted solely of unimodal multidimensional samples thus allowing the methods to be assessed on their capability to accurately identify unimodality across varying sample sizes, distribution shapes, and statistical parameter variations (e.g., mean vector, covariance matrix). Similarly, in the second case, the datasets contained exclusively multimodal multidimensional samples to evaluate the methods' robustness in reliably detecting multimodality under the same varying conditions.

The third case presents the effects of imbalanced multidimensional multimodal datasets, respectively and in the fourth case, a combination of both imbalanced unimodal and multimodal datasets were applied to the different methods. The inclusion of these cases enabled a comprehensive analysis of the strengths and limitations of each method in handling multidimensional data, hence providing insights into their effectiveness in detecting structural complexities inherent in real-world datasets.

1) CASE 1: "BALANCED MULTIDIMENSIONAL UNIMODAL DATASETS"

This experiment consisted of different unimodal distributions with varying mean and variance parameters, designed to represent different data characteristics across multiple dimensions. Snippets of the datasets used are shown in Fig. 13, which displays five distributions having distinct mean values, specifically $[0, 0]$, $[2, 2]$, $[4, 4]$, $[6, 6]$, and $[8, 5]$, with a consistent variance of $[1, 1]$. These distributions contained varying sample sizes, ranging from 50 to 1000 data points, which allowed for analysis of small- and large scale datasets with different dispersion. The diversity in mean values enabled exploration of different spatial arrangements of data points, which provided a basis for understanding clustering tendencies in datasets with distinct central tendencies.

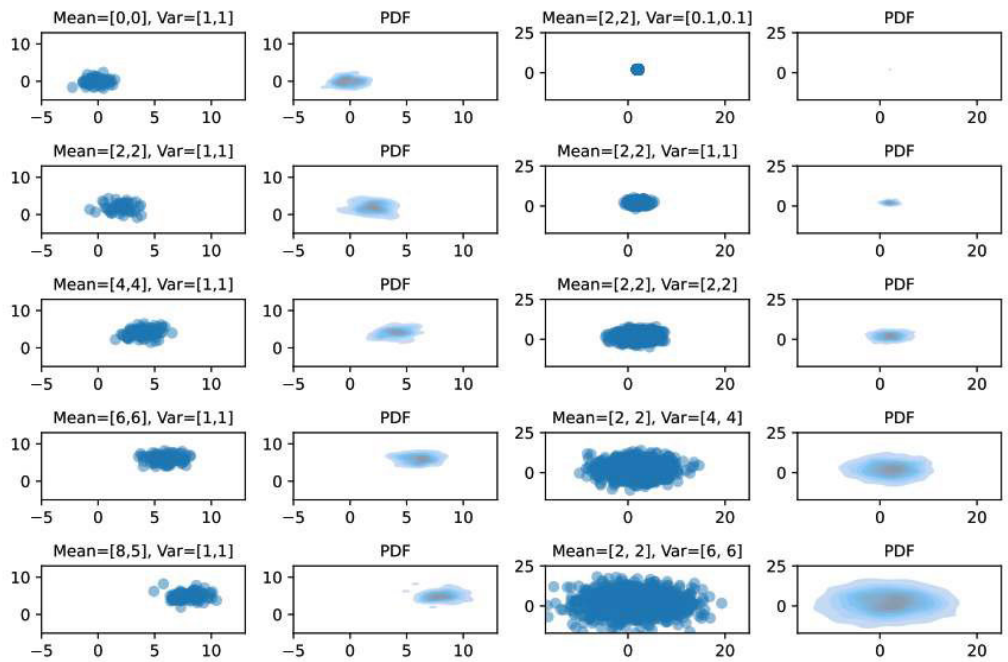


FIGURE 13. Balanced multidimensional unimodal samples (i.e. 2-dimensional) with varying characteristics and data lengths, where samples consist of shifting mean and variance values.

In Fig. 13, we display distributions across all samples but introduces variability in the variance, ranging from $[0.1, 0.1]$ to $[6, 6]$. These datasets contained sample sizes, ranging from 50 to 1000 data points, thus capturing the effects of increasing dispersion on data clustering and spread. The increasing variance in these distributions allowed for testing robustness in modality detection algorithms by progressively adding more spread to the data, which can affect overlap and density estimation. This dataset structure facilitated a comprehensive evaluation of unimodal data characteristics across different scales and dispersion levels, hence making it suitable for studying the effects of scale and variance in clustering and distribution analysis.

The results of the analysis of four multidimensional-based test methods (MDAM, Folding, Runt, and KS) on the balanced multidimensional datasets with varying data lengths are shown in Fig. 14. This result demonstrates that all the evaluated methods achieved near-perfect accuracy, thus reflecting their effectiveness in handling multidimensional unimodal datasets.

At a data length of $N = 50$, all algorithms, including MDAM, Folding, Runt, and KS, exhibited high accuracy levels, with MDAM and Folding achieving 100% accuracy, while Runt and KS yielded slightly lower accuracies at approximately 98%. This indicates that, even with limited data samples, the algorithms are highly reliable in detecting unimodality with minimal performance differences. As the dataset size was increased to $N = 100$, MDAM and Folding maintained their perfect accuracy whereas Runt and KS experienced a slight improvement reaching accuracy levels of approximately 99%. This highlights an increasing reliability with larger data samples.

In Fig. 14, at data lengths of $N = 500$ and $N = 1000$, the performance remained consistent, with MDAM and Folding consistently achieving 100% accuracy across all dataset sizes. The Runt and KS methods also exhibited high performance, with accuracy values stabilizing around 99%, showing their ability to handle large datasets with minimal loss in accuracy. The results suggest that MDAM and Folding are particularly robust across varying data lengths providing consistently perfect detection capabilities. In contrast, while Runt and KS performed slightly below these two leading methods, they still maintained exceptional accuracy, thus confirming their effectiveness for multidimensional unimodal detection.

This performance analysis reveals that MDAM and Folding are the most reliable methods achieving perfect accuracy across all dataset sizes while Runt and KS remained highly competitive with only marginally lower accuracy. MDAM along with the other algorithms have demonstrated strong scalability with increasing data lengths, which affirms their robustness for balanced multidimensional unimodal datasets.

2) CASE 2: BALANCED MULTIDIMENSIONAL MULTIMODAL DATASETS

Samples of the datasets used are as shown in Fig. 15, which consisted of different bimodal and trimodal distributions, each characterized by distinct mean values and variances. In the bimodal cases, two clusters were present in each distribution with equal proportions meaning that each cluster contributed 50% of the data points. The cluster means

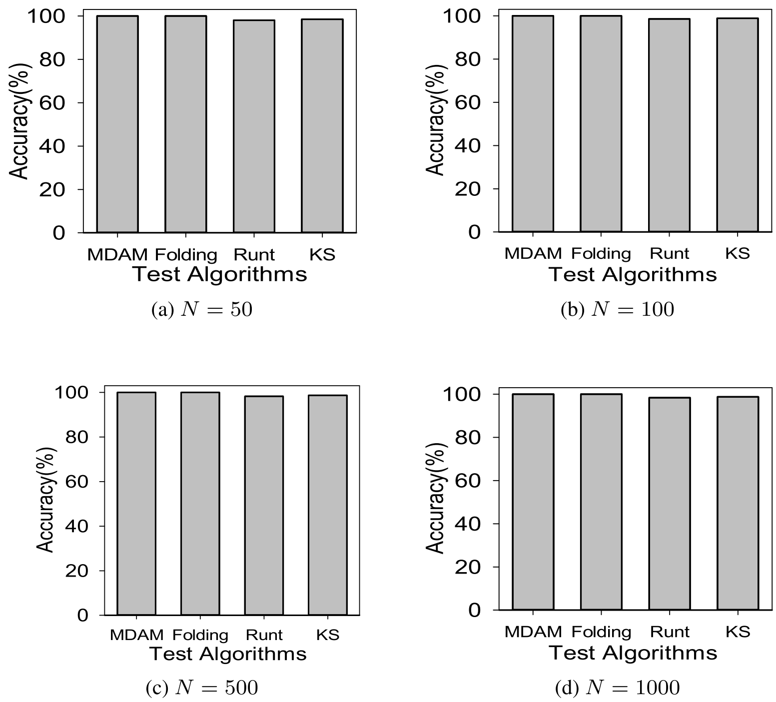


FIGURE 14. Accuracy under balanced multidimensional unimodal datasets with the following dataset lengths: (a) 50 (b) 100 (c) 500 (d) 1000.

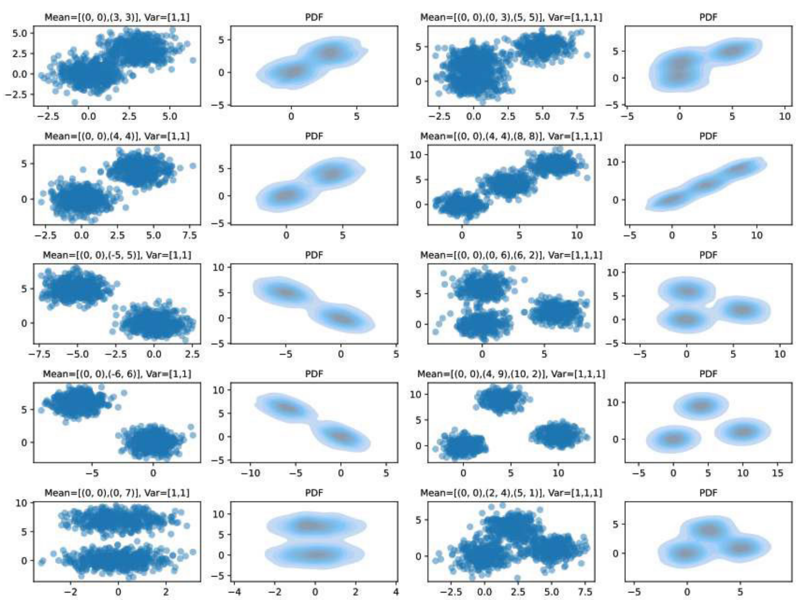


FIGURE 15. Balanced multidimensional multimodal samples ($D = 2$) for $N = 1000$ comprising a combination of 2 and 3 modes with varying characteristics, where samples consist of shifting mean and variance values.

for these bimodal distributions included pairs such as $[0, 0]$ and $[3, 3]$, $[0, 0]$ and $[4, 4]$. The variance for all clusters in the bimodal distributions was maintained at $[1, 1]$, to indicate uniform dispersion across both dimensions. This consistency ensured that differences in cluster positioning rather than variance influenced the multimodal nature of the distributions. The trimodal distributions, on the other hand, contained three distinct clusters, each contributing approximately 33.3% of the total data points. These distributions featured cluster means such as $[0, 0]$, $[0, 3]$, and $[5, 5]$,

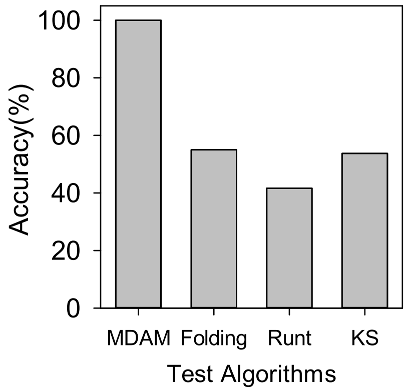


FIGURE 16. Accuracy under balanced multidimensional multimodal datasets for $N = 1000$ with a combination of 2 and 3 modes.

as well as other unique configurations like [0, 0], [4, 9], and [10, 2]. Similar to the bimodal distributions, the variance in all clusters remained constant at [1, 1, 1], which ensured uniform spread across the data dimensions. The trimodal datasets introduced a greater level of complexity, with cluster locations varying more significantly compared to the bimodal cases. This design allowed for a comprehensive evaluation of multimodal detection methods by presenting a range of cluster arrangements while maintaining a consistent variance structure across the dataset.

The analysis of the performance of the tested algorithms, as presented in Fig. 16, reveals notable differences in accuracy when applied to balanced multidimensional multimodal datasets containing both bimodal and trimodal distributions. The results indicate that the MDAM algorithm consistently outperformed the other methods, achieving near-perfect accuracy (99.8%) in identifying multimodal structures within the datasets. This suggests that MDAM is highly effective in distinguishing between distinct clusters, regardless of whether the data follows a bimodal or trimodal distribution pattern. The superior performance of MDAM can be attributed to its robust discriminant analysis-based framework, which efficiently captures directly from the datasets the variations in mean values that define the modality of the datasets.

In contrast, the Folding, Runt, and KS algorithms exhibited varying degrees of accuracy, with Folding achieving moderate performance (55%), while Runt demonstrated the lowest accuracy (41.67%) among the tested methods. The decreased performance of these methods highlights their potential limitations in effectively capturing complex cluster structures, particularly in the presence of trimodal distributions where cluster separation is less distinct. The results further emphasize the scalability and adaptability of MDAM as it maintains high accuracy across different dataset complexities and cluster arrangements. Overall, the findings reveal the importance of using advanced discriminant analysis techniques such as those employed in MDAM, to enhance the reliability of multimodal detection in multidimensional settings.

3) CASE 3: IMBALANCED MULTIDIMENSIONAL MULTIMODAL DATASETS

This case examined different multidimensional (i.e. bi-dimensional) multimodal distributions, each generated by combining multiple Gaussian components with different mean values, variances, and proportions. Some samples contain two clusters each representing bimodal distributions where the relative proportions of data points between the clusters vary. For example, in the first dataset of Fig. 17, 20% of the points are sampled from a Gaussian distribution centered at [0, 0] and the remaining 80% from [5, 5], whereas in another dataset (i.e. third row, first column of Fig. 17), 60% of the data comes from [0, 0] and 40% from [−6, 6]. These varying proportions and cluster placements allow for an in-depth analysis of how different levels of separation and imbalances influence modality detection. The sample sizes range from 50 to 1000 data points across the different samples, thus offering scenarios of increasing complexity and scale for evaluation.

Datasets with trimodal distributions were also considered where three clusters were combined with different proportions increasing the complexity of the dataset. These distributions featured clusters centered at diverse locations such as [0, 0], [0, 4], and [4, 0] in Fig. 17 (i.e. first row, third column), with different proportions such as 20%, 60%, and 20%, respectively. Some datasets exhibited significant imbalances such as 10% of points from [0, 0], 30% from [0, 8], and 60% from [8, 0] in third row and third column of Fig. 17. The presence of three clusters introduces higher complexity in identifying underlying structures and assessing the performance of the testing methods. The combination of varying cluster means, proportions, and scales across bimodal and trimodal distributions makes this dataset well-suited for testing the different multimodal detection techniques across different levels of complexity and data distributions.

The accuracies presented in Fig. 18 demonstrate that the MDAM algorithm achieved the highest accuracy (99.6%) compared to the other tested methods (Folding, Runt, and KS) when applied to the imbalanced multidimensional multimodal datasets shown in Fig. 17. Again, the superior performance of MDAM can be attributed to its ability to effectively capture and analyze the varying mean and variance characteristics of the dataset, which consist of complex multimodal structures with imbalances in data distribution. The algorithm's discriminant-based approach directly applied to the actual time series datasets enables it to better identify modality changes, even in challenging scenarios where other methods may struggle due to data imbalance or overlapping distributions.

Fig. 18 reveals that the Folding and KS methods exhibited lower accuracy, likely because they do not handle the variability in mean shifts and variance changes as effectively as MDAM since they operate directly on the PDF. By depending on the PDF, these methods may be more sensitive to distribution skewness and the presence of multiple modes with varying densities leading to misclassification and lower

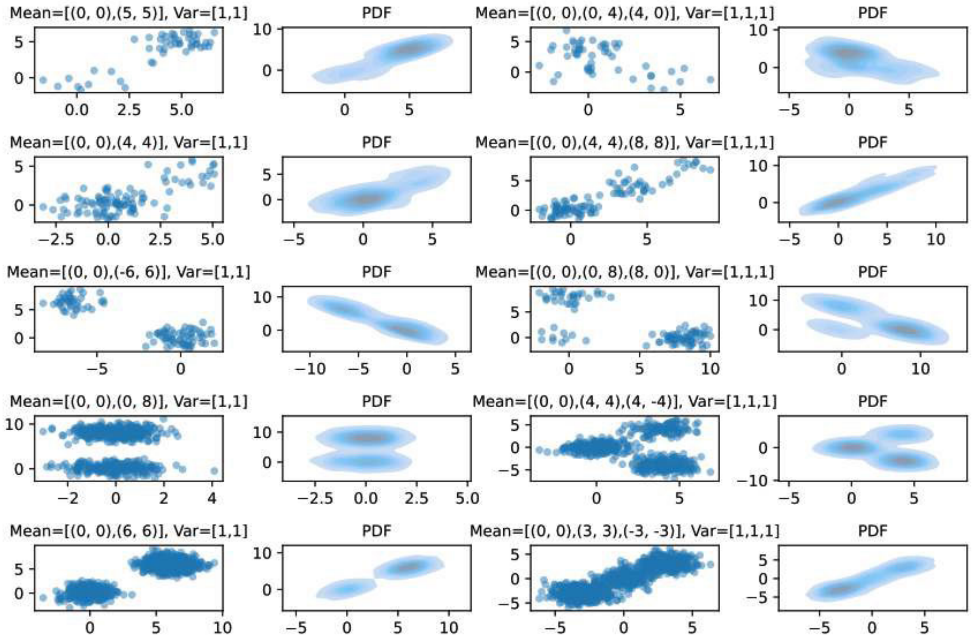


FIGURE 17. Imbalanced multidimensional multimodal samples ($D = 2$) with varying characteristics and data lengths, where samples consist of shifting mean and variance values.

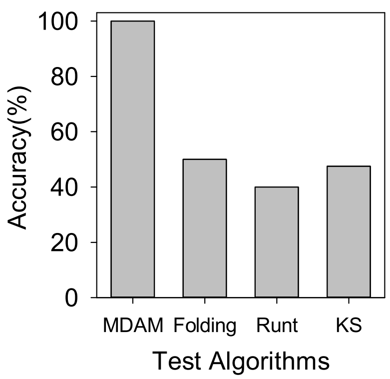


FIGURE 18. Accuracy under imbalanced multidimensional multimodal datasets.

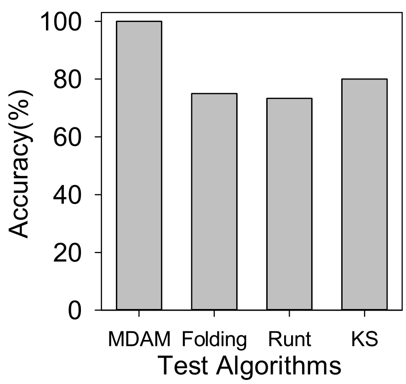


FIGURE 19. Accuracy under mixture of balanced multidimensional unimodal and imbalanced multidimensional multimodal datasets.

detection accuracy. The Runt test, while achieving slightly better performance than Folding and KS, still falls short compared to MDAM possibly due to its reliance on single linkage clusters that do not generalize well to highly imbalanced multimodal datasets.

The performance levels achieved further indicate that MDAM is better suited for datasets with shifting mean and variance values, as it provides a more robust measure of unimodality and multimodality. The lower accuracy of the other methods suggests that they might be more appropriate for balanced datasets with fewer overlapping distributions or more distinct modality separations.

4) CASE 4: MIXTURE OF BALANCED AND IMBALANCED MULTIDIMENSIONAL UNIMODAL AND MULTIMODAL DATASETS

Lastly, the mixture of both balanced multidimensional unimodal and imbalanced multidimensional multimodal datasets was examined to further test the discriminatory power of the different methods. The results presented in Fig. 19 and Table 8 illustrate the performance of the different test algorithms under this case. Once again, the MDAM algorithm consistently outperformed the other methods across all the evaluation metrics, achieving an accuracy of 99.6%, a precision of 99.8%, a recall of 99.7%, and an F1-score of 99.5%. This indicates that MDAM is highly effective in distinguishing between unimodal and multimodal distributions, further demonstrating superior robustness to the varying data characteristics.

In contrast, the Folding, Runt, and KS methods exhibited significantly lower performance as documented in Table 8. The Folding algorithm, in particular, struggled with recall achieving only 50% which suggests that it fails to detect a substantial portion of actual multimodal cases. Similarly, the Runt algorithm, while achieving a relatively better precision of 66.67%, performed poorly in recall (46.67%) indicating

TABLE 8. Performance under mixture of balanced multidimensional unimodal and imbalanced multidimensional multimodal datasets.

Evaluation Metrics (%)	MDAM	Folding	Runt	KS
Accuracy	99.6	75	73.33	80
Precision	99.8	50	66.67	75
Recall	99.7	50	46.67	60
F1_score	99.5	50	52.38	64.29

difficulty in identifying multimodal instances accurately. The KS method showed moderate performance with an accuracy of 80%, which, although better than Folding and Runt, still fell short as compared to MDAM.

The lower F1-scores observed in the Folding and Runt algorithms in Table 8 highlight their limited capability in balancing precision and recall, suggesting that they may suffer from either high false positives or false negatives in complex multimodal scenarios. The KS test, while slightly more balanced, still failed to achieve the same level of robustness and adaptability demonstrated by MDAM.

Based on the various experimental cases considered in the study, results obtained consistently demonstrated MDAM's ability to generalize across diverse data distributions. It maintained a high precision and recall performance across unimodal and multimodal datasets, whether unidimensional or multidimensional, with varying data lengths, balanced or imbalanced distributions, and changing mean and variance parameters. These findings thus emphasize the robustness and adaptability of MDAM, making it a reliable approach for detecting modality changes in complex time series data environments.

F. EXPERIMENTAL ANALYSES USING REAL DATASETS

1) DESCRIPTION OF THE DATASETS

To further evaluate the performance of the MDAM approach, we tested against a set of publicly available datasets spanning different domains, including computing, biomedical monitoring, communication systems, and finance. Each dataset is described below as follows:

a: EC2 CPU UTILIZATION DATASET

This dataset (ec2_cpu_utilization_5f5533) originates from the Numenta Anomaly Benchmark (NAB), specifically the realAWSCloudwatch collection. It contains CPU utilization traces from Amazon Web Services (AWS) Elastic Compute Cloud (EC2) instances, representing realistic server workloads under production conditions. The dataset is labeled with changes and is widely used to benchmark anomaly detection models in cloud performance monitoring, particularly for detecting sudden workload fluctuations or irregular utilization patterns [28].

b: FETAL2013 DATASE

This dataset (Fetal2013_70_6000_12000), also known as the Cardiotocography dataset, is available from the UCI Machine Learning Repository. It consists of cardiotocogram (CTG) recordings that capture fetal heart rate (FHR) and uterine contraction (UC) signals. The dataset includes more than 2,000 samples, each labeled as Normal, Suspect, or Pathological, providing a rich source for testing biomedical anomaly detection algorithms. This dataset is particularly important in prenatal healthcare applications, where early detection of abnormal fetal conditions can aid in timely intervention [29].

c: COGNITIVE RADIO DATASET

This dataset denoted as (dataset_ofdm_10dB) represents signal measurements obtained from an OFDM (Orthogonal Frequency Division Multiplexing) communication system under a controlled signal-to-noise ratio (SNR) of 10 dB. This dataset is widely used for evaluating modulation recognition, spectrum sensing, and signal classification algorithms within the context of cognitive radio networks and intelligent wireless communication systems [30].

Each record in the dataset corresponds to a sampled complex baseband signal, comprising amplitude and phase variations that reflect typical OFDM behavior under moderate noise interference (10 dB SNR). The 10 dB noise level ensures a realistic yet challenging testing scenario for signal detection and classification algorithms, particularly those that rely on statistical distribution analysis such as MDAM, Folding, or KS tests.

d: EXCHANGE-2 DATASET

This dataset (exchange-2_cpc_results) belongs to the real Known Cause collection and contains financial time series related to currency exchange rates. This dataset has been annotated with known anomalies corresponding to structural changes in exchange dynamics, making it well suited for evaluating detection methods in financial time series. It is particularly relevant for studying the ability of algorithms to capture regime changes, sudden market shocks, and other contextual anomalies [28].

Together, the above datasets provide a diverse and challenging evaluation benchmark across different application domains, hence allowing for a comprehensive validation of the different modality testing approaches under realistic conditions.

G. VISUALIZATION OF MDAM'S β-MEASURE ON REAL DATASETS

In this section, we visualize how MDAM determines the modality of the different real datasets. For this purpose, the representative datasets from the different domains described above were employed, including unimodal time series, multimodal time series with variance and mean shifts.

The illustration in Fig. 20 demonstrates MDAM's capability and reliability in accurately identifying modality structures within real-world datasets, thereby extending its validation beyond controlled synthetic environments.

TABLE 9. Performance comparison of MDAM and baseline methods on four real datasets.

Methods	CPU Utilisation Acc (%)	CPU Utilisation Time (Secs) Win. Size = 600	Financial Exchange Acc (%)	Financial Exchange Time (Secs) Win. Size = 500	Fetal Acc (%)	Fetal Time (Secs) Win. Size = 1750	Cognitive Radio Acc (%)	Cognitive Radio Time (Secs) Win. Size = 175
Folding	60	0.0075	20	0.0069	20	0.0104	80	0.0037
DAT	60	0.0510	20	0.0405	20	0.1017	80	0.0046
KS	60	0.0075	40	0.0064	80	0.0116	20	0.0017
Dip	60	0.0006	20	0.0005	40	0.0015	40	0.0004
MDAM	100	0.0147	80	0.0095	90	0.0321	100	0.0019

In this evaluation, the four distinct real datasets representing domains such as CPU utilization, biomedical signals, communication system signals, and financial exchange rates were examined to assess the method's generalization and adaptability under diverse operating conditions. These datasets were selected to include data whose positional distributions exhibited both unimodal and bimodal/multimodal characteristics, thus ensuring a balanced test of MDAM's sensitivity and discriminative power.

For each dataset, the modality classification was determined using the proposed $\mathcal{B}$-measure framework relative to the empirically derived statistical cutoff, $\mathcal{B}_{\text{cutoff}}$. This approach enables a consistent and interpretable decision rule: datasets yielding $\mathcal{B}$-measure values below the cutoff are classified as unimodal, while those exceeding the cutoff are identified as multimodal. Such a quantitative thresholding mechanism provides an objective criterion for distinguishing complex distributional behaviors that are often overlooked by classical unimodality tests.

As shown in Fig. 20(a), (c), (e), and (g), the unimodal time series maintain stable distributional dynamics, with $\mathcal{B}$-measure values persistently remaining below $\mathcal{B}_{\text{cutoff}}$. This stability reflects the absence of significant structural transitions or multiple density peaks, reaffirming MDAM's accuracy in recognizing homogeneous, single mode patterns. Conversely, Fig. 20(b), (d), (f), and (h) illustrate the datasets exhibiting clear variance- and mean-induced multimodality, wherein the $\mathcal{B}$-measure consistently surpasses the cutoff. These instances correspond to real-world conditions such as fluctuating CPU workload intensities, physiological signal variability, or cognitive radio amplitude shifts scenarios where conventional unimodality tests tend to underperform or misclassify.

Overall, these results confirm that MDAM is not only robust under synthetic evaluation but also scalable and reliable for real-world data analysis, making it a promising diagnostic tool for applications involving complex, evolving distributions across diverse application areas.

H. COMPARATIVE ANALYSES ON REAL DATASETS

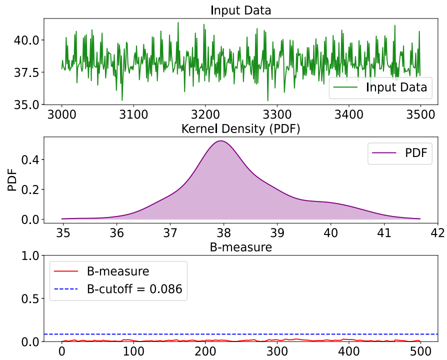
Table 9 presents a detailed comparison of the proposed MDAM algorithm against four established unimodality vs multimodality test methods (Folding, DAT, KS, and Dip) evaluated across the above four real-world datasets: CPU Utilisation, Exchange Rate, Fetal Health, and Cognitive Radio.

For the CPU Utilisation dataset, MDAM achieved a perfect accuracy of 100%, outperforming the benchmark methods considered in this study. Both the Folding and KS tests yielded moderate accuracies of 60%, while DAT and Dip showed weaker detection capabilities. In terms of computational efficiency, MDAM completed its analysis in 0.0147 seconds, slightly slower than the simpler Folding (0.0075 s) and KS (0.0075 s) tests but delivering far superior detection accuracy. This trade-off illustrates that MDAM's enhanced discriminant process incurs only a marginal computational cost while ensuring complete reliability in detecting modality.

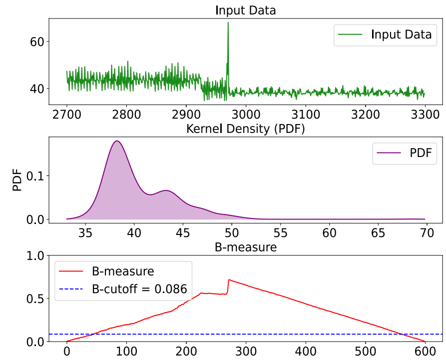
In the exchange rate dataset, MDAM maintained an accuracy of 80%, again outperforming all other tests. The folding, DAT, and Dip methods each attained low accuracies of 20%, while KS reached 40%. The strong performance of MDAM in this financial dataset highlights its robustness in capturing multimodal behaviors in data prone to noise and high variance. MDAM's computational time of 0.0095 seconds was competitive and nearly identical to the fastest method, demonstrating its scalability even in high-frequency financial sequences with fluctuating distributions.

The fetal health dataset further confirmed MDAM's superior discriminative power, achieving 90% accuracy, whereas KS followed with 80% and the other tests (Folding, DAT, and Dip) failed to exceed 40%. This result highlights MDAM's capacity to detect subtle multimodal transitions in physiological data, where overlapping and non-linear patterns are common. Although MDAM recorded a slightly higher computational time of 0.0321 seconds, the improvement in accuracy compensates for this minor delay, hence demonstrating the method's ability to maintain diagnostic precision in complex biomedical signals.

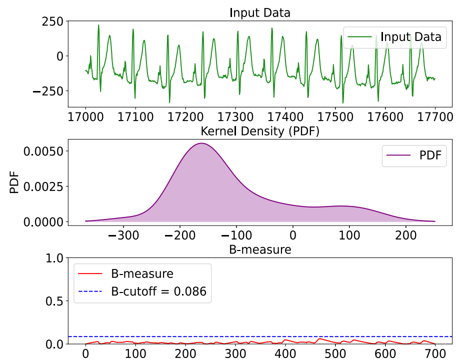
A similar trend was observed in the cognitive radio dataset, where MDAM once again achieved perfect accuracy (100%), outperforming the other baseline methods. The Dip test attained a low accuracy of 40%, while Folding and DAT reached 80%, and KS trailed with 20%. Although MDAM required slightly more computation time (0.0089 seconds) compared to DAT (0.0067 seconds), its better accuracy performance makes it particularly effective in analyzing cognitive radio signals that exhibit sudden amplitude variations and multimodal energy distributions.



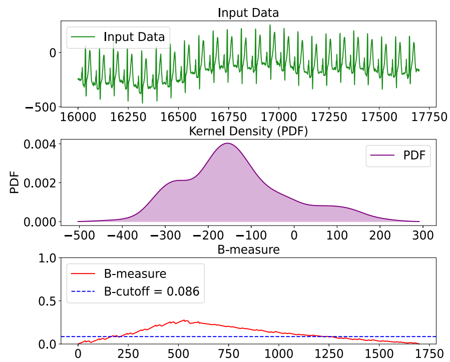
(a) CPU Utilization data(Unimodal)



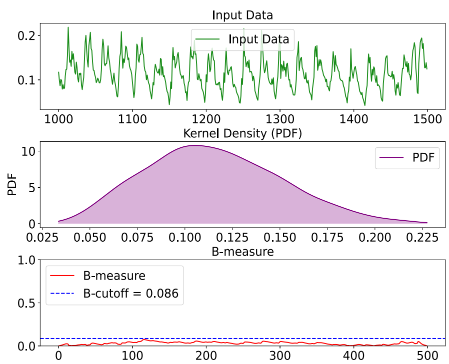
(b) CPU Utilization data(Multimodal)



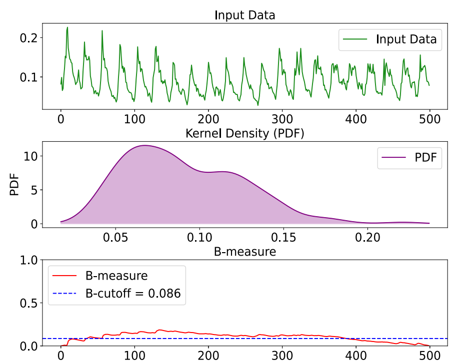
(c) Fetal data(Unimodal)



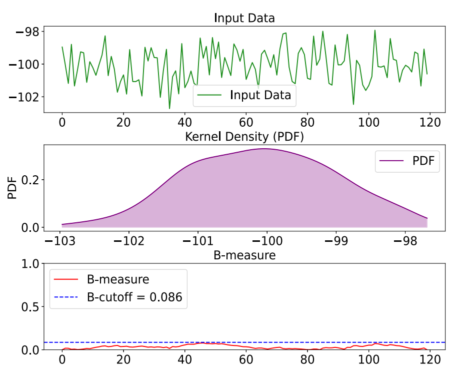
(d) Fetal data(Multimodal)



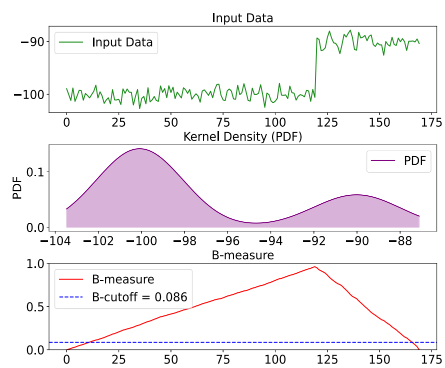
(e) Exchange data(Unimodal)



(f) Exchange data(Multimodal)



(g) Cognitive Radio data(Unimodal)



(h) Cognitive Radio data (Multimodal)

FIGURE 20. **Visualization of the MDAM B-measure on Real Datasets.**

Overall, across all the four real-world datasets, MDAM consistently achieved the highest accuracy compared to all baseline techniques. MDAM attained an average accuracy of approximately 92.5%, far exceeding the next-best performer, the KS test, which averaged around 50–55%. This consistent superiority emphasizes MDAM's discriminant analysis foundation, which effectively captures both variance changes and distributional overlap, which are two critical statistical properties that conventional unimodality tests often miss. Although MDAM's mean computation time (0.016 seconds) was marginally higher than that of other simpler methods, it remained well below 0.03 seconds, thus representing a negligible trade-off for its substantial improvement in detection reliability.

By leveraging a multidimensional discriminant framework, MDAM has been shown to effectively identify subtle structural transitions and variance driven modality changes with high performance. Consequently, MDAM emerges as a powerful and reliable modality detection tool, especially in applications where accurate detection of distributional shifts is crucial for anomaly identification, adaptive control, and decision optimization.

V. CONCLUSION

This paper has introduced a modality testing approach termed, MDAM, designed to determine whether distributions in multidimensional time series datasets are unimodal or multimodal. The proposed method is distribution-independent, enabling it to effectively identify modality in multidimensional datasets regardless of the underlying data distribution. MDAM leverages a newly derived function based on the principles of discriminant analysis, which is applied directly to the time series data rather than its probability distribution function. This direct application enhances the method's sensitivity and robustness, thus outperforming five state-of-the-art time series modality testing methods: Folding, Runt, KS, DAT, and Dip tests. The current scope of MDAM's applicability has been demonstrated across a wide range of datasets (synthetic and real), including unimodal and multimodal distributions, unidimensional and multidimensional data, varying data lengths, as well as balanced and imbalanced datasets with shifting mean and variance parameters. Additionally, MDAM consistently balances accuracy, recall, and AUC with acceptable runtime costs. This superiority highlights its practical value for real-world potential applications like anomaly detection, financial time series analysis, medical signal processing, and environmental data modeling, where understanding data modality is crucial for effective decision-making and pattern recognition. Future research will focus on extending MDAM to autonomously estimate unknown statistical parameters, hence facilitating its application in unsupervised clustering tasks.

REFERENCES

[1] M. O. Ahmed and G. Walther, "Investigating the multimodality of multivariate data with principal curves," *Comput. Statist. Data Anal.*, vol. 56, no. 12, pp. 4462–4469, Dec. 2012.

[2] P. R. Varghese, M. S. P. Subathra, G. Peter, A. A. Stonier, R. Kuppusamy, and Y. Teekaraman, "A novel MODWT–local pattern transformation feature fusion approach for high-impedance fault detection in medium voltage power distribution networks," *Neural Comput. Appl.*, vol. 37, no. 22, pp. 17457–17471, Aug. 2025.

[3] S. Xing, J. Niu, H. Wang, T. Ren, M. Cui, and X. Shi, "An enhanced data-driven framework for early kick detection based on imbalanced multivariate time series classification," *Neural Comput. Appl.*, vol. 35, no. 24, pp. 17777–17793, Aug. 2023.

[4] N. Hussen, M. Salem, A. I. Eldesouky, N. Sakr, and S. Elghamrawy, "Real-time drone detection framework based on advanced texture feature extraction and pattern recognition model using GUI," *Neural Comput. Appl.*, vol. 37, no. 5, pp. 3435–3454, Feb. 2025.

[5] B. Wang, G. Wu, X. Li, J. Gao, Y. Hu, and B. Yin, "Modality perception learning-based determinative factor discovery for multimodal fake news detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 7, pp. 12643–12654, Jul. 2025.

[6] Y. Luo, Y. Zhang, X. Ding, X. Cai, C. Song, and X. Yuan, "StrDip: A fast data stream clustering algorithm using the dip test of unimodality," in *Proc. 19th Int. Conf. Web Inf. Syst. Eng. (WISE)*, Dubai, United Arab Emirates, Nov. 2018, pp. 193–208.

[7] G. Vardakas, A. Kalogeratos, and A. Likas, "UniForCE: The unimodality forest method for clustering and estimation of the number of clusters," 2023, *arXiv:2312.11323*.

[8] P. Chasani and A. Likas, "The UU-test for statistical modeling of unimodal data," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108272.

[9] A. K. Jain, P. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, 2000.

[10] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *Ann. Statist.*, vol. 13, no. 1, pp. 70–84, Mar. 1985.

[11] I. Stoepker and E. van den Heuvel, "Testing for multimodality," Ph.D. thesis, Eindhoven Univ. Technol., 2016.

[12] A. Gupta, A. J. Onumanyi, S. Ahlawat, Y. Prasad, V. Singh, and A. M. Abu-Mahfouz, "DAT: A robust discriminant analysis-based test of unimodality for unknown input distributions," *Pattern Recognit. Lett.*, vol. 182, pp. 125–132, Jun. 2024.

[13] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouët, "Are your data gathered?" in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2210–2218.

[14] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.

[15] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed., New York, NY, USA: Routledge, 2018.

[16] D. W. Müller and G. Sawitzki, "Excess mass estimates and tests for multimodality," *J. Amer. Stat. Assoc.*, vol. 86, no. 415, pp. 738–746, Sep. 1991.

[17] P. Kolyvakis and A. Likas, "A multivariate unimodality test harnessing the dip statistic of Mahalanobis distances over random projections," 2023, *arXiv:2311.16614*.

[18] J. A. Hartigan and S. Mohanty, "The runt test for multimodality," *J. Classification*, vol. 9, no. 1, pp. 63–70, Jan. 1992.

[19] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed., New York, NY, USA: Wiley, 2003.

[20] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[21] L. Qu and Y. Pei, "A comprehensive review on discriminant analysis for addressing challenges of class-level limitations, small sample size, and robustness," *Processes*, vol. 12, no. 7, p. 1382, Jul. 2024.

[22] K. V. Mardia, J. T. Kent, and C. C. Taylor, *Multivariate analysis*, vol. 88. Oxford, U.K.: Wiley, 2024.

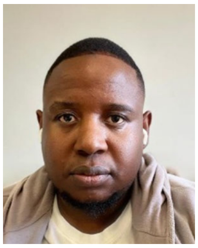
[23] S. Raychaudhuri, "Introduction to Monte Carlo simulation," in *Proc. Winter Simulation Conf.*, Dec. 2008, pp. 91–100.

[24] S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman, "Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations," *Eur. J. Epidemiol.*, vol. 31, no. 4, pp. 337–350, Apr. 2016.

[25] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[26] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," in *Proc. 1st Workshop Eval. Comparison NLP Syst.*, 2020, pp. 79–91.

[27] S.-T. Yeh, ''Using trapezoidal rule for the area under a curve calculation,'' in *Proc. 27th Annu. SAS User Group International (SUGI)*, 2002, pp. 1–5.
[28] N. Numenta, ''Numenta anomaly benchmark,'' Tech. Rep., 2015.
[29] K. Bache and M. Lichman, ''Cardiotocography data set,'' UCI Mach. Learn. Repository, Tech. Rep., 2010.
[30] J. Okonkwo, A. Onumanyi, B. A. Salihu, H. B. Salau, and S. Oyewobi, ''Performance of the recursive one-sided hypothesis testing technique under varying signal to noise ratio conditions in cognitive radio,'' Tech. Rep., 2018.



IPELENG LABIUS MACHELE received the B.Tech. and M.Comp. degrees in computer systems engineering from Tshwane University of Technology (TUT), Pretoria, South Africa, where he is currently pursuing the Ph.D. degree in electrical engineering. He is with the Council for Scientific and Industrial Research (CSIR), Pretoria. His research interests include wireless sensor networks, the Internet of Things (IoT), artificial intelligence (AI), interconnected smart transactive microgrids (ISTMGs), and anomaly detection.

ADEIZA J. ONUMANYI (Member, IEEE) received the B.Eng. degree in electrical and electronics engineering from Abubakar Tafawa Balewa University, Bauchi, Nigeria, in 2005, and the M.Eng. and Ph.D. degrees in communication engineering from the Federal University of Technology (FUT), Minna, Nigeria, in 2010 and 2014, respectively. He is currently a Researcher with the Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa. He has several research papers published in peer-reviewed journals and at IEEE flagship conferences. From 2010 to 2021, he lectured and conducted research with the Department of Telecommunication Engineering, FUT Minna, where he was involved in securing several grants, serving on several organizing committees for various conferences, including IEEE conferences, reviewing several articles for high-impact journals, and participating in various technical workshops. Among his research interests are spectrum sensing in cognitive radio, wireless sensor networks, smart transactive microgrids, dc nanogrids, radar systems, image processing, cyber-physical systems, and low-power wireless area networks.

ADNAN M. ABU-MAHFOUZ (Senior Member, IEEE) received the M.Eng. and Ph.D. degrees in computer engineering from the University of Pretoria. He is currently a Chief Researcher and the Centre Manager of the Emerging Digital Technologies for 4IR (EDT4IR) Research Centre, Council for Scientific and Industrial Research (CSIR), an Extraordinary Professor with the University of Pretoria, a Professor Extraordinaire with Tshwane University of Technology, and a Visiting Professor with the University of Johannesburg. His research interests include wireless sensor and actuator networks, low-power wide-area networks, software-defined wireless sensor networks, cognitive radio, network security, network management, and sensor/actuator node development. He is a member of many IEEE technical communities. He is a Section Editor-in-Chief of *Journal of Sensor and Actuator Networks*, and an Associate Editor of IEEE Access, IEEE Internet of Things, and IEEE Transactions on Industrial Informatics.

ANISH KURIEN (Senior Member, IEEE) received the M.Tech. and Ph.D. degrees in electrical engineering from Tshwane University of Technology, South Africa, with a specialization in telecommunication technology, and the Ph.D. degree in computing from the University of Paris-Est, France. He is an experienced Full Professor with a strong track record in the higher education sector. His research interests include wireless sensor and actuator networks, low-power wide-area networks, software-defined wireless sensor networks, cognitive radio, network security, network management, sensor/actuator node development, smart grids, and smart water systems. He focuses on next-generation wireless communication networks, the digital transformation of networks, and the digitization of economies. He also has broad expertise in computer networks, software engineering, and data analytics. He has also contributed to the development of numerous successful, large-scale, and multidisciplinary research and development proposals. He is an active member of several IEEE technical communities.