BEYOND SCALAR DISTANCES: GEOMETRY-AWARE QUANTUM EVIDENTIAL C-MEANS VIA CONTRASTIVE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Traditional distance-based clustering methods lack the representational capacity for intricate uncertainty. To address this, evidential c-means (ECM) clustering introduces the concept of credal partition, providing a more flexible and expressive framework for uncertain data modeling. However, ECM fundamentally relies on a nondirectional scalar distance metric, focusing solely on how far samples are from centroids while ignoring their spatial distribution relative to overlapping clusters. To overcome these limitations, we build upon generalized quantum evidence theory to propose a new method, quantum evidential c-means based on contrastive learning (CL-QECM). It not only integrates a fine-grained contrastive attraction-repulsion mechanism, but also encodes these spatial relationships into a geometry-aware angular representation within the complex credal space. Utilizing a temperature-scaled softmax function, the proposed method explicitly maximizes the similarity of positive pairs and minimizes the similarity of negative pairs in terms of distance. Extensive experiments on twelve benchmark datasets show that CL-QECM outperforms state-of-the-art methods in Purity, NMI, and ARI by up to 7.42%, 3.43%, and 7.30% respectively, exhibiting superior robustness and stability under noise and multi-cluster overlap.

1 INTRODUCTION

Clustering is one of the most fundamental tasks in machine learning and data mining (Zhang et al., 2024). It aims to partition a set of objects into homogeneous groups by maximizing the similarity between objects in the same group and minimizing the similarity between objects in different groups. However, real-world data are often characterized by overlapping clusters, boundary ambiguity, and noise, making uncertainty an inherent property of clustering tasks (Masson & Denoeux, 2008). If such uncertainty is ignored, ambiguous samples are inevitably forced into a single cluster, which may lead to unreliable partitioning, degraded clustering quality, and reduced interpretability. Therefore, accurately characterizing and managing uncertainty has become a fundamental issue in clustering, particularly for target recognition (Li et al., 2025), and decision-making (Song & Fei, 2026) applications where reliable confidence information is as important as the clustering result itself. To address this uncertainty, numerous theories and methods have been proposed, including random permutation sets (Deng et al., 2024; Zhou et al., 2025; Chen et al., 2026a), evidence reasoning (Yang & Xu, 2025), Dempster-Shafer evidence (DSE) theory (Dempster, 1967; Shafer, 1976; Deng & Jiang, 2025), multi-source information fusion (Sun et al., 2026), open set recognition (Liu et al., 2022), fuzzy clustering(Nie et al., 2025), and others.

Among them, the DSE theory has a strong ability to handle uncertainty and incomplete information through the basic probability assignment function (Li et al., 2024b), which has been deeply studied and extended to belief rule bases (Yang et al., 2023), entropy (Zhao et al., 2024; Zhan et al., 2026), divergence (Huang et al., 2024; Zhang & Xiao, 2024; Xiao et al., 2025a), complex evidence theory (Xiao et al., 2024), quantum evidence theory (Xiao, 2023; Aug. 2025) and so on (Yu et al., 2025). In addition to its theoretical development, DSE theory has been successfully applied in multiple fields, such as decision-making (Fan et al., 2025), risk assessment (Chen et al., 2025), reliability evaluation (Su et al., 2026), uncertainty estimation (Chen et al., 2024; Li et al., 2024a), and clustering (Zhang et al., 2026). Unfortunately, DSE theory is limited by its strict expression

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

in the real-number space, making it difficult to effectively characterize the spatial distributions of uncertain data.

To overcome this inherent challenge, complex evidence theory (CET) (Xiao et al., 2024) was initially proposed to handle uncertainty on the complex plane, which was subsequently formalized into quantum evidence theory (QET) by extending uncertainty into a Hilbert space (Xiao, 2023; Aug. 2025). Within this framework, the generalized quantum basic probability amplitude (GQPBA) models uncertainty via magnitude and phase. It provides a fine-grained representation of uncertainty, including partial ignorance, complete ignorance and outlier. By exploiting quantum interference, the generalized quantum evidential combination rule (GQECR) aggregates multi-source GQPBAs to derive the generalized quantum belief (GQBel) and plausibility (GQPl) functions, establishing robust lower and upper probability bounds (Xiao, Aug. 2025). This expressive quantum-evidential paradigm has recently been successfully applied in various advanced fields. For instance, it has facilitated the design of adaptive quantum circuits for efficient classification under uncertain environments (Xiao et al., 2025b), enabled robust one-dimensional time series classification pipelines (Yu et al., 2026), and inspired trustworthy quantum evidential neural networks to overcome the dual limitations of incomplete feature representation and insufficient uncertainty modeling (Chen et al., 2026b).

Recently, a multitude of clustering methods have been proposed to tackle uncertain data. As a pioneering baseline based on DSE, the evidential c-means (ECM) algorithm serves as a representative and foundational evidential model, which addresses this issue by distributing basic belief assignments over the power set of classes (Masson & Denoeux, 2008). This distinctive credal partition establishes a unifying paradigm that extends crisp, fuzzy, and possibilistic partitions, thereby offering exceptional flexibility in modeling ambiguous cluster assignments and detecting outliers (Jiao et al., 2023). However, ECM primarily relies on a nondirectional distance metric for cluster assignments. That is, it solely focuses on how far the samples are from the centroids, but largely ignores where and how they distribute relative to other overlapping clusters.

Motivated by these challenges, we propose a new method, quantum evidential c-means based on contrastive learning (CL-QECM), which integrates a contrastive learning mechanism into the quantum evidential clustering framework. CL-QECM constructs a novel contrastive loss by evaluating the spatial relationship between data samples and subset barycenters. Particularly, the data sample and its subset barycenter with high quantum mass assignment are defined as a positive pair, while the sample and its remaining subset barycenters constitute negative pairs. Utilizing a temperature-scaled softmax function, the proposed method explicitly maximizes the similarity of positive pairs and minimizes the similarity of negative pairs in terms of distance. Furthermore, rather than being confined to a nondirectional real-valued space, CL-QECM encodes these optimized spatial relationships into a geometry-aware angular representation within the complex credal plane. This enables CL-QECM to explicitly establish an angular margin, effectively alleviating the boundary overlap problem and learning more accurate centroids to improve the overall clustering performance.

2 Quantum evidential clustering based on contrastive learning

In this paper, we propose a novel method that integrates quantum evidential clustering with contrastive learning to learn more discriminative and robust cluster centroids. The overview of CL-QECM is illustrated in Fig. 1. Our method consists of three jointly optimized components, namely, quantum credal partition generation (QCPG), instance-centroid contrastive learning (ICCL), and cluster centroid optimization (CCO). In brief, QCPG maps the raw data features to the complex-valued space to construct the quantum credal partition. After that, ICCL formulates an instance-centroid contrastive loss based on the spatial relationship between the instances and subset barycenters to capture global discriminative information. Finally, by coupling the linear system solver and the momentum gradient descent, the cluster centroids are iteratively updated by the CCO model. These three components are alternately optimized until convergence to yield the clustering results.

2.1 Quantum evidential clustering

The quantum evidential clustering is a novel algorithm based on generalized quantum evidence theory, which extends classical evidential clustering to support complex-valued credal representation and explicit modeling of multi-dimensional uncertainty. By mapping the representation of uncertainty into the complex number plane, partial knowledge regarding the class membership of objects is comprehensively expressed by GQBPA.

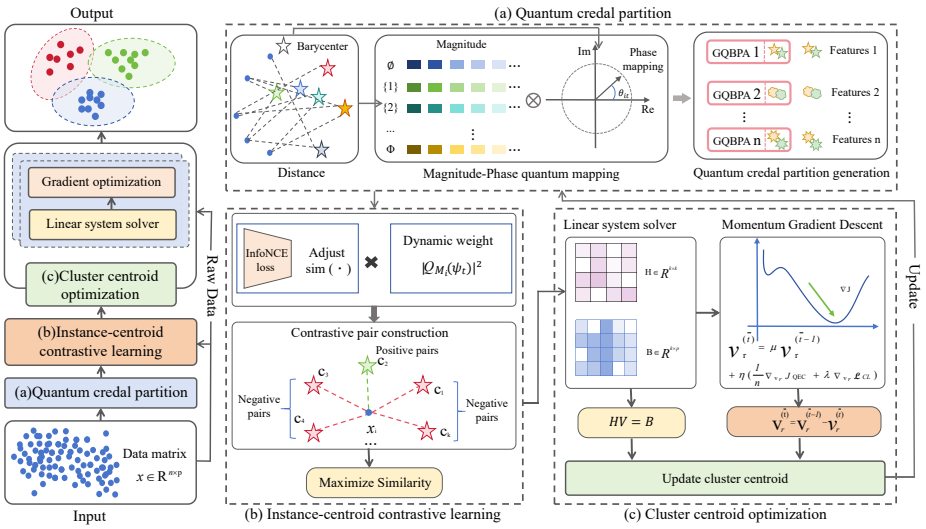


Figure 1: The framework of quantum evidential c-means clustering based on contrastive learning.

Given a dataset $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^p$ describing n objects to be clustered into k groups $\Phi = \{\phi_1, \ldots, \phi_k\}$. The cluster membership uncertainty and ambiguity of object $\mathbf{x}_i$ for $i \in \{1, \ldots, n\}$ can be represented by a generalized quantum basic probability amplitude (GQBPA) $\mathbb{Q}_{\mathbb{M}_i}$. Specifically, $\mathbb{Q}_{\mathbb{M}_i}(\psi)(1 < |\psi| < k)$ characterizes partial ignorance by assigning belief to a subset of classes, $\mathbb{Q}_{\mathbb{M}_i}(\Phi)$ represents complete ignorance, and $\mathbb{Q}_{\mathbb{M}_i}(\emptyset)$ accounts for outliers. The n-tuple $\mathbb{C}_{\mathbb{M}} = \{\mathbb{Q}_{\mathbb{M}_1}, \ldots, \mathbb{Q}_{\mathbb{M}_n}\} \in \mathbb{C}^{n \times 2^k}$ denotes a quantum credal partition of object set X. $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ indicates the degree of belief that object $\mathbf{x}_i$ belongs to the subset $\psi_t \subseteq \Phi$.

Let every cluster ϕ_r for $1 \leq r \leq k$ be represented by a class center $\mathbf{v}_r \in \mathbb{R}^p$. $\mathbb{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^{k \times p}$ are composed of the class centers. The magnitude $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ associated with object $\mathbf{x}_i$ and subset $\psi_t (\psi_t \neq \emptyset, \psi_t \subseteq \Phi)$ depends on the distance $d_{it} = ||\mathbf{x}_i - \overline{\mathbf{v}}_t||$ between object $\mathbf{x}_i$ and the subset barycenter $\overline{\mathbf{v}}_t$ of ψ_t. The assignment ensures that the magnitude is high when the distance d_{it} is small. The subset barycenter $\overline{\mathbf{v}}_t$ is defined as:

$$\overline{\mathbf{v}}_t = \frac{1}{c_t} \sum_{r=1}^{k} s_{rt} \mathbf{v}_r, \qquad s_{rt} = \begin{cases} 1, & \phi_r \in \psi_t, \\ 0, & \phi_r \notin \psi_t, \end{cases} \tag{1}$$

where c_t is the cardinality of the subset ψ_t, s_{rt} is the subset-class indicator function, and $\mathbf{v}_r$ is the center of the class ϕ_r.

The quantum evidential clustering (QEC) objective function with respect to the quantum credal partition $\mathbb{C}_{\mathbb{M}}$ and cluster centers $\mathbb{V}$ is defined as:

$$\min J_{\text{QEC}}(\mathbb{C}_{\mathbb{M}}, \mathbb{V}) \triangleq \sum_{i=1}^{n} \sum_{\{t | \psi_t \neq \emptyset, \psi_t \in 2^{\Phi}\}} c_t^{\alpha} \mathbb{Q}_{\mathbb{M}_i}{}^{\beta}(\psi_t) d_{it}^2 + \sum_{i=1}^{n} \delta^2 \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\emptyset), \tag{2}$$

subject to

$$\sum_{\{t | \psi_t \neq \emptyset, \psi_t \in 2^{\Phi}\}} |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^2 + |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)|^2 = 1, \qquad \forall i = 1, 2, \ldots, n, \tag{3}$$

here,

$$\mathbb{Q}_{\mathbb{M}}(\psi_t) = \varphi(\psi_t) e^{j\theta(\psi_t)}, \quad \theta \in [0, \frac{\pi}{2}], \quad \varphi(\psi_t) = |\mathbb{Q}_{\mathbb{M}}(\psi_t)|, \tag{4}$$

where $|\mathbb{Q}_{\mathbb{M}}(\psi_t)|$ represents the cluster membership, and phase θ serves as a geometry-aware angular representation. Restricting θ to $[0, \pi/2]$ monotonically maps unbounded Euclidean distances to a bounded angular space. This explicitly encodes a geometry-aware angular representation into the complex credal plane, establishing an angular margin that fundamentally resolves boundary ambiguity. As shown in Fig. 2, even when scalar distances degenerate, differing cluster spatial distributions ensure that ambiguous boundary samples are separated not only by their magnitude but crucially by their geometric orientation. Further theoretical discussions are deferred to Appendix B.2. In Eq. (2), $c_t = |\psi_t|$ denotes the number of elements in the nonempty subset ψ_t. The objective function involves three hyperparameters: α, β, and δ. Specifically, α controls the degree of punishment, the weighting exponent β controls the fuzziness of the partition and δ characterizes the distance to the noise class.

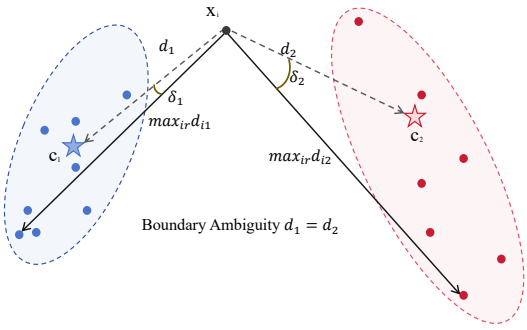


Figure 2: Geometric mechanism of distance-induced phase in CL-QECM. For an ambiguous boundary sample x_i equidistant to centroids ($d_1 = d_2$), traditional scalar distance fails to distinguish clusters. By incorporating geometric orientation deviations ($\delta_1 \neq \delta_2$) and boundary scales ($\max_{x_{ir}} d_{i1} \neq \max_{x_{ir}} d_{i2}$), CL-QECM produces distinct phases ($\theta_{i1} \neq \theta_{i2}$) via Eq. 12.

2.2 Instance-Centroid Contrastive Learning

While the quantum evidential clustering (QEC) framework establishes a geometry-aware representation via phase mapping, its fundamental objective function primarily focuses on the relationships between samples and global cluster centroids. To fully leverage the spatial distributions and enhance cluster separability, we propose an instance-centroid contrastive learning module. Driven by the principle of contrastive representation learning, we construct positive and negative pairs based on subset barycenters for the proposed loss: each sample forms a positive pair with the subset barycenter corresponding to its focal set with high magnitude $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$, and negative pairs with the barycenters of other focal sets. The instance-centroid contrastive loss is formulated as:

$$\mathcal{L}_{\text{CL}} = -\sum_{i=1}^{n} \sum_{\{t|\psi_t \neq \emptyset, \psi_t \in 2^{\Phi}\}} |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^2 \log \frac{\exp\left(\text{sim}(\mathbf{x}_i, \bar{\mathbf{v}}_t)/\tau\right)}{\sum_{\psi_s \neq \emptyset} \exp\left(\text{sim}(\mathbf{x}_i, \bar{\mathbf{v}}_s)/\tau\right)}, \tag{5}$$

where τ denotes the temperature hyperparameter to control the softness and $sim(\cdot)$ is used to measure the pair-wise similarity: $\text{sim}(\mathbf{x}_i, \bar{\mathbf{v}}_t) = -\|\mathbf{x}_i - \bar{\mathbf{v}}_t\|^2$. $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^2$ represents a dynamic weight for the contrastive optimization between the instance and the subset barycenter, which makes positive sample pairs with high belief dominant in contrastive optimization, so as to effectively mitigate the impact of ambiguous assignments and noisy representations in cluster boundary overlapping areas.

2.3 Optimization Procedure

The proposed CL-QECM achieves collaborative optimization by unifying the quantum evidential clustering objective function and the instance-centroid contrastive loss into a joint framework. The overall objective function is formulated as follows:

$$J = J_{\text{QEC}} + \lambda \mathcal{L}_{\text{CL}}, \tag{6}$$

where $\lambda \in [0, 1]$ serves as a regularization hyperparameter balancing between the contrastive loss and the quantum evidential clustering loss of CL-QECM. We restrict $\lambda \leq 1$ to prevent the contrastive loss from dominating the quantum credal assignment.

In this section, the optimization procedure of the CL-QECM is proposed. Inspired by the alternating optimization strategy, the variables to be optimized are iteratively updated, including the quantum credal partition $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ and the cluster center matrix $\mathbb{V}$.

2.3.1 Update of Quantum Credal Partition

Directly minimizing the joint objective J with respect to $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ is mathematically intractable due to non-linear coupling in the contrastive term. To maintain analytical tractability, we optimize a decoupled surrogate objective. By treating $\mathcal{L}_{\text{CL}}$ as constant during this step, the surrogate reduces to: $\tilde{J}(\mathbb{C}_{\mathbb{M}}) = J_{\text{QEC}}(\mathbb{C}_{\mathbb{M}})$.

Stage 1: Update of magnitude. By introducing Lagrange multipliers λ_i, we formulate the Lagrangian for the constrained optimization of $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ as:

$$\mathcal{L}(\mathbb{C}_{\mathbb{M}}, \lambda_1, \ldots, \lambda_n) = J - \sum_{i=1}^{n} \lambda_i \left(\sum_{\{t|\psi_t \neq \emptyset, \psi_t \in 2^{\Phi}\}} |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^2 + |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)|^2 - 1 \right). \tag{7}$$

Setting the partial derivatives of $\mathcal{L}$ with respect to $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ and $|\mathbb{Q}_{\mathbb{M}_i}(\emptyset)|$ to 0 yields:

$$\frac{\partial \mathcal{L}}{\partial |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|} = \beta c_t^{\alpha} d_{it}^2 |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^{\beta-1} - 2\lambda_i |\mathbb{Q}_{\mathbb{M}_i}(\psi_t)| = 0. \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)|} = \beta \delta^2 |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)|^{\beta-1} - 2\lambda_i |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)| = 0. \tag{9}$$

Solving Eqs.(8) and (9),

$$|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)| = \sqrt{\frac{\left(\frac{1}{c_t^{\alpha} d_{it}^2}\right)^{\frac{2}{\beta-2}}}{\sum_{\psi_r \neq \emptyset} \left(\frac{1}{c_r^{\alpha} d_{ir}^2}\right)^{\frac{2}{\beta-2}} + \left(\frac{1}{\delta^2}\right)^{\frac{2}{\beta-2}}}}, \quad |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)| = \sqrt{\frac{\left(\frac{1}{\delta^2}\right)^{\frac{2}{\beta-2}}}{\sum_{\psi_r \neq \emptyset} \left(\frac{1}{c_r^{\alpha} d_{ir}^2}\right)^{\frac{2}{\beta-2}} + \left(\frac{1}{\delta^2}\right)^{\frac{2}{\beta-2}}}}. \tag{10}$$

Notably, we decouple the contrastive module during the update of $|\mathbb{Q}_{\mathbb{M}_i}|$ to maintain analytical solvability via alternating optimization. Theoretical justifications are provided in Appendix B.1.

In practice, direct calculation of Eq. (10) may be numerically unstable. To avoid division by zero when $d_{it}^2 = 0$, a small positive constant ϵ is added to $c_t^{\alpha} d_{it}^2$. Moreover, in high-dimensional spaces, large distances may cause $(c_t^{\alpha} d_{it}^2)^{-\frac{2}{\beta-2}}$ to underflow since $\beta > 2$. We therefore normalize both the numerator and denominator by $D^{\frac{2}{\beta-2}}$, where $D = \min_r(c_r^{\alpha} d_{ir}^2)$, yielding the following numerically stable formulation:

$$|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)| = \sqrt{\frac{\left(\frac{D}{c_t^{\alpha} d_{it}^2 + \epsilon}\right)^{\frac{2}{\beta-2}}}{\sum_{\psi_r \neq \emptyset} \left(\frac{D}{c_r^{\alpha} d_{ir}^2 + \epsilon}\right)^{\frac{2}{\beta-2}} + \left(\frac{D}{\delta^2}\right)^{\frac{2}{\beta-2}}}}, \quad |\mathbb{Q}_{\mathbb{M}_i}(\emptyset)| = \sqrt{\frac{\left(\frac{D}{\delta^2}\right)^{\frac{2}{\beta-2}}}{\sum_{\psi_r \neq \emptyset} \left(\frac{D}{c_r^{\alpha} d_{ir}^2 + \epsilon}\right)^{\frac{2}{\beta-2}} + \left(\frac{D}{\delta^2}\right)^{\frac{2}{\beta-2}}}}. \tag{11}$$

Stage 2: Update of phase. Since the phase term is not bounded by the simplex constraint, strictly deriving it jointly with the magnitude leads to intractable non-convexity. Instead, we introduce a distance-induced phase reparameterization. We explicitly formulate θ as a deterministic mapping of the spatial distance. Consequently, as the cluster centroids are iteratively updated to minimize spatial distances, the phase representation dynamically adapts to establish the expected angular margin. Guided by principles of scale invariance and structural regularization, the mapping is defined as:

$$\theta_{it} = \frac{\pi}{2}\left(1 - \frac{d_{it}^2}{\max_r d_{ir}^2}\right), \qquad \theta_{i\emptyset} = \frac{\pi}{2}\left(1 - \frac{d_{i\emptyset}^2}{\max_r d_{ir}^2}\right), \tag{12}$$

where $d_{i\emptyset}$ represents the specific phase mapping for the empty set, $\max_r d_{ir}$ denotes the maximum distance from x_i to instances within the cluster, geometrically represented as the cluster boundary span $\max_{x_{ir}} d_{it}$ in Fig. 2.

Finally, we synthesize the robust amplitude $|\mathbb{Q}_{\mathbb{M}}|$ and the geometrically-aware phase θ to obtain the necessary condition of optimality for $\mathbb{C}_{\mathbb{M}}$:

$$\mathbb{Q}_{\mathbb{M}_i}(\psi_t) = \varphi(\psi_t) \cdot e^{(j\theta_{it})}, \qquad \mathbb{Q}_{\mathbb{M}_i}(\emptyset) = \varphi(\emptyset)| \cdot e^{(j\theta_{i\emptyset})}. \tag{13}$$

2.3.2 Update of cluster centers

Fixing the quantum credal partition $\mathbb{C}_{\mathbb{M}}$, we update the cluster center matrix $\mathbb{V}$. To ensure stable convergence and enhance the discriminability of overlapping clusters across disparate datasets, we introduce a two-stage optimization strategy with a warm-up mechanism to optimize $\mathbb{V}$.

Stage 1: Linear system solver. The partial derivatives of J_{QEC} with respect to the cluster centers $\mathbf{v}_l$ are obtained:

$$\frac{\partial J_{\text{QEC}}}{\partial \mathbf{v}_l} = \sum_{i=1}^{n} \sum_{\psi_t \neq \emptyset} c_t^{\alpha} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) \frac{\partial d_{it}^2}{\partial \mathbf{v}_l}, \qquad \frac{\partial d_{it}^2}{\partial \mathbf{v}_l} = 2 s_{lt}(\mathbf{x}_i - \overline{\mathbf{v}}_t)(-\frac{1}{c_t}). \tag{14}$$

From Eq. (14), it follows that:

$$\frac{\partial J_{\text{QEC}}}{\partial \mathbf{v}_l} = -2 \sum_{i=1}^{n} \sum_{\psi_t \neq \emptyset} c_t^{\alpha-1} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt}(\mathbf{x}_i - \overline{\mathbf{v}}_t) = -2 \sum_{i=1}^{n} \sum_{\psi_t \neq \emptyset} c_t^{\alpha-1} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt}(\mathbf{x}_i - \frac{1}{c_t} \sum_{r=1}^{k} s_{rt} \mathbf{v}_r). \tag{15}$$

Then setting the derivative to zero results in l linear equations in $\mathbf{v}_r$, expressible as:

$$\sum_i \mathbf{x}_i \sum_{\psi_t \neq \emptyset} c_t^{\alpha-1} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt} = \sum_r \mathbf{v}_r \sum_i \sum_{\psi_t \neq \emptyset} c_t^{\alpha-2} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt} s_{rt}. \tag{16}$$

To reformulate the system of linear equations in Eq. (16) into a concise matrix notation, we define the elements of the matrices $\mathbb{Q}_{\mathbb{B}} \in \mathbb{C}^{k \times p}$ and $\mathbb{Q}_{\mathbb{H}} \in \mathbb{C}^{k \times k}$ as:

$$\mathbb{Q}_{\mathbb{B}_{lq}} = \sum_{i=1}^{n} x_{iq}^{\dagger} \sum_{\psi_t \neq \emptyset} c_t^{\alpha-1} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt} = \sum_{i=1}^{n} x_{iq}^{\dagger} \sum_{\psi_t \ni \phi_l} c_t^{\alpha-1} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t), \tag{17}$$

$$\mathbb{Q}_{\mathbb{H}_{lr}} = \sum_{i=1}^{n} \sum_{\psi_t \neq \emptyset} c_t^{\alpha-2} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t) s_{lt} s_{rt} = \sum_{i=1}^{n} \sum_{\psi_t \supseteq \{\phi_r, \phi_l\}} c_t^{\alpha-2} \mathbb{Q}_{\mathbb{M}_i}^{\beta}(\psi_t), \tag{18}$$

where $q = 1, 2, \ldots, p$ and $r = 1, 2, \ldots, k$. Substituting Eqs. (17) and (18) back into Eq. (16) yields a compact matrix form. Thus, the optimal cluster center matrix $\mathbb{V}$ is updated by solving the following linear system:

$$\mathbb{Q}_{\mathbb{H}} \mathbb{V} = \mathbb{Q}_{\mathbb{B}}. \tag{19}$$

In the initial iterations, structural matrices may exhibit singularity in sparse or high-dimensional regimes. We employ Tikhonov regularization to acquire a stable solution:

$$(\mathbb{Q}_{\mathbb{H}} + \gamma I) \mathbb{V} = \mathbb{Q}_{\mathbb{B}}, \tag{20}$$

where I denotes the k-dimensional identity matrix, and γ is a regularization parameter.

Stage 2: Contrastive Gradient Refinement. In the subsequent iterations, the cluster centroids are fine-tuned by incorporating the gradient of the contrastive loss. Specifically, the gradient updates are driven by both the quantum clustering loss ∇J_{QEC} and the contrastive loss $\nabla \mathcal{L}_{\text{CL}}$. To accelerate convergence during this refinement phase, a momentum-accelerated gradient descent scheme is employed to update the cluster centroids:

$$\boldsymbol{\nu}_l^{(t')} = \mu \boldsymbol{\nu}_l^{(t'-1)} + \eta \left(\frac{1}{n} \nabla_{\mathbf{v}_l} J_{\text{QEC}} + \lambda \nabla_{\mathbf{v}_l} \mathcal{L}_{\text{CL}} \right), \qquad \mathbf{v}_l^{(t')} = \mathbf{v}_l(t'-1) - \boldsymbol{\nu}_l^{(t')}, \tag{21}$$

where η is learning rate hyperparameter, t' represents the iteration step, and μ denotes momentum coefficient hyperparameter. See Appendix D for details on our algorithm's pseudocode flow.

2.4 Computational complexity analysis

The overall computational complexity of CL-QECM is $O(t \cdot n \cdot p \cdot n_f)$, where t is the number of iterations, n is the sample size, p is the feature dimension, and n_f is quantum focal elements. In real-world scenarios, restricting $O(t \cdot n \cdot p \cdot 2^k)$ reduces the complexity to $O(t \cdot n \cdot p \cdot k^2)$, making it highly scalable. A detailed complexity analysis is provided in Appendix C.

3 Experiments

3.1 Datasets and evaluation metrics

In this paper, twelve real-world datasets[1] were used for experiments, including Wireless Indoor Localization (WIL), Predict Students' Dropout and Academic Success (PSDAS), Phishing Websites (PW), Abalone, Magic, Australian, Yale32, Yale64, COIL20, USPS, Isolation of Substituted Letter (Isolet) and Yeast. Among them, USPS is a set of classic handwritten images; Australian, Magic, Abalone, PW, PSDAS, Yeast, WIL are from UCI Machine Learning Repository; COIL20 and Yale are face image datasets; and Isolet is a dataset of audio samples of people pronouncing the English letters (A to Z) in English. Table 4 of Appendix E provides a detailed overview of these datasets. To provide a more comprehensive illustration of the superiority of our algorithm, we utilize three standard clustering evaluation metrics: Purity, Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI). The detailed definitions of these metrics are provided in Appendix E.

3.2 Comparison with State of the Arts

To more convincingly evaluate the effectiveness of the proposed method, we compare it against eight competitive baselines, including evidential-based approaches and deep contrastive models. Detailed baseline descriptions and experimental setup are provided in Appendices F and G, respectively. Run each algorithm independently 20 times on each dataset, and record the mean and standard deviation

[1] https://anonymous.4open.science/r/CL-QECM-Data-49FB/

Table 1: Clustering performance comparison of CL-QECM with eight baselines on twelve benchmark datasets concerning clustering Purity

Datasets	Ckm	PCL(MLP)	IDEC	DEC	DFKM	DTEC	EGMM	ECM	CL-QECM
WIL	0.8275±0.03	0.7339±0.11	0.9588±0.01	0.8770±0.06	0.9556±0.00	0.2755±0.00	0.6680±0.20	**0.9771±0.00**	0.9460±0.00
PSDAS	0.6608±0.01	0.5578±0.04	0.6144±0.03	0.6271±0.03	0.6323±0.01	0.4993±0.00	0.4664±0.19	0.5007±0.00	**0.6834±0.00**
PW	0.5569±0.00	0.5610±0.01	0.5569±0.00	0.5569±0.00	0.5569±0.00	0.5569±0.00	0.5481±0.13	0.6872±0.05	**0.6909±0.02**
Abalone	**0.5354±0.00**	0.4958±0.02	0.5236±0.01	0.5261±0.00	0.4955±0.01	0.3692±0.00	0.4131±0.11	0.5288±0.00	0.5318±0.01
Magic	0.6484±0.00	0.6504±0.01	0.6573±0.02	0.6485±0.00	0.6484±0.00	0.6519±0.00	**0.6900±0.03**	0.6484±0.00	0.6763±0.03
Australia	0.8392±0.01	0.6444±0.07	0.7793±0.09	0.8133±0.06	0.7525±0.03	0.5551±0.00	0.5497±0.20	0.6246±0.00	**0.8520±0.00**
Yale32	0.5082±0.04	0.3958±0.03	0.4818±0.02	0.4830±0.04	0.4364±0.02	0.1939±0.00	0.0643±0.05	0.2876±0.02	**0.5152±0.00**
Yale64	0.5682±0.04	0.5206±0.04	0.4327±0.03	0.4448±0.03	0.3721±0.02	0.1818±0.00	0.3879±0.00	0.2694±0.04	**0.6406±0.04**
COIL20	0.6436±0.03	0.5181±0.02	0.6214±0.03	0.5490±0.05	0.5578±0.01	-	0.1581±0.03	0.4007±0.04	**0.7178±0.02**
USPS	0.6537±0.03	0.4133±0.03	0.6648±0.02	0.6876±0.03	0.5937±0.03	-	0.1684±0.00	0.4379±0.00	**0.7413±0.01**
Isolet	0.5405±0.02	0.3416±0.03	0.6576±0.02	0.5342±0.04	0.4654±0.02	-	0.0899±0.01	0.1358±0.02	**0.6726±0.02**
Yeast	0.5004±0.01	0.4847±0.02	0.5063±0.03	0.5026±0.03	0.5372±0.00	0.3127±0.00	0.4958±0.02	0.5145±0.01	**0.5470±0.01**
Avg. Rank	3.67	5.75	3.83	4.25	5.00	8.08	7.50	5.50	**1.42**

Table 2: Clustering performance comparison of CL-QECM with eight baselines on twelve benchmark datasets concerning clustering NMI

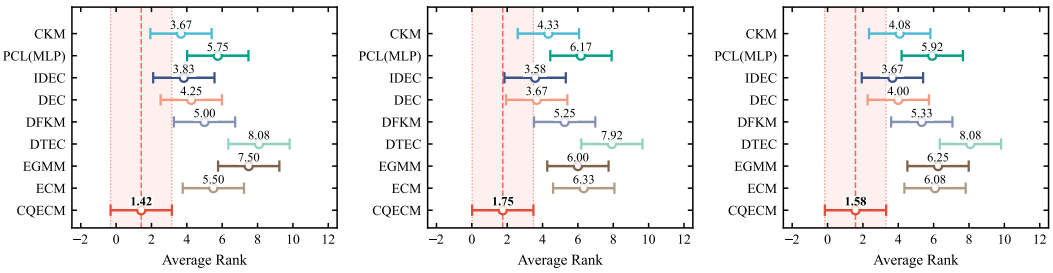
Datasets	Ckm	PCL(MLP)	IDEC	DEC	DFKM	DTEC	EGMM	ECM	CL-QECM
WIL	0.6486±0.02	0.5457±0.13	0.8776±0.02	0.8168±0.05	0.8619±0.01	0.0795±0.00	0.7101±0.15	**0.9205±0.00**	0.8432±0.00
PSDAS	0.1570±0.01	0.0592±0.03	0.1106±0.04	0.1222±0.03	0.1387±0.01	0.0051±0.00	0.0601±0.07	0.0043±0.00	**0.1720±0.01**
PW	0.0024±0.00	0.0071±0.01	0.0026±0.00	0.0049±0.00	0.0016±0.00	0.0013±0.00	0.0239±0.04	**0.1127±0.05**	0.1035±0.00
Abalone	0.1640±0.00	0.1581±0.01	0.1647±0.00	0.1624±0.01	0.1561±0.00	0.0442±0.00	0.0863±0.06	0.1542±0.00	**0.1718±0.00**
Magic	0.0003±0.00	0.0203±0.03	0.0202±0.04	0.0094±0.01	0.0011±0.00	0.0247±0.00	**0.1014±0.05**	0.0004±0.00	0.0520±0.04
Australia	0.3599±0.01	0.0825±0.07	0.2703±0.12	0.3300±0.11	0.1927±0.06	0.3451±0.00	0.0539±0.10	0.0435±0.00	**0.3942±0.00**
Yale32	0.5477±0.03	0.4302±0.02	0.5541±0.02	**0.5742±0.03**	0.4784±0.02	0.0604±0.00	0.2118±0.02	0.3283±0.02	0.5508±0.00
Yale64	0.6298±0.03	0.5430±0.03	0.5486±0.02	0.5599±0.02	0.4315±0.05	0.2891±0.00	0.4483±0.00	0.3176±0.03	**0.6515±0.03**
COIL20	0.7553±0.02	0.5931±0.02	0.7338±0.02	0.7293±0.02	0.6526±0.01	-	0.2949±0.08	0.5825±0.03	**0.7763±0.01**
USPS	0.5346±0.02	0.2908±0.03	0.6068±0.01	**0.6519±0.02**	0.4555±0.03	-	0.0218±0.00	0.4143±0.00	0.6166±0.00
Isolet	0.6836±0.01	0.3929±0.03	**0.7737±0.01**	0.7523±0.03	0.5801±0.01	-	0.1546±0.01	0.4079±0.02	0.7668±0.01
Yeast	0.2432±0.01	0.2093±0.02	0.2458±0.02	0.2477±0.03	0.2646±0.01	0.0607±0.00	0.2689±0.03	0.2390±0.00	**0.2718±0.01**
Avg. Rank	4.33	6.17	3.58	3.67	5.25	7.92	6.00	6.33	**1.75**

Table 3: Clustering performance comparison of CL-QECM with eight baselines on twelve benchmark datasets concerning clustering ARI

Datasets	Ckm	PCL(MLP)	IDEC	DEC	DFKM	DTEC	EGMM	ECM	CL-QECM
WIL	0.6474±0.04	0.5101±0.16	0.8939±0.02	0.7749±0.08	0.8859±0.01	0.0018±0.00	0.6210±0.21	**0.9404±0.00**	0.8618±0.00
PSDAS	0.2065±0.01	0.0654±0.04	0.1503±0.04	0.1646±0.03	0.1629±0.01	-0.0029±0.00	0.0434±0.05	0.0053±0.00	**0.2413±0.03**
PW	-0.0017±0.00	0.0089±0.01	-0.0007±0.00	-0.0015±0.00	-0.0029±0.00	0.0047±0.00	0.0184±0.03	**0.1509±0.07**	0.1455±0.00
Abalone	0.1845±0.00	0.1249±0.02	0.1582±0.01	0.1609±0.01	0.1003±0.01	0.0021±0.00	0.0494±0.07	0.1428±0.00	**0.1901±0.01**
Magic	0.0012±0.00	0.0241±0.04	0.0294±0.04	0.0236±0.02	0.0059±0.00	0.0062±0.01	**0.0824±0.06**	-0.0014±0.00	0.0728±0.09
Australia	0.4595±0.01	0.1014±0.09	0.3418±0.16	0.4069±0.13	0.2569±0.06	0.3108±0.00	0.0597±0.12	0.0607±0.00	**0.4946±0.00**
Yale32	0.2830±0.04	0.1509±0.03	0.2988±0.03	**0.3245±0.04**	0.2000±0.02	-0.0017±0.00	0.0212±0.01	0.0852±0.01	0.2989±0.00
Yale64	0.3816±0.03	0.2823±0.04	0.2752±0.03	0.2883±0.03	0.1609±0.04	0.0280±0.00	0.1223±0.00	0.0845±0.02	**0.4007±0.05**
COIL20	0.5426±0.03	0.3625±0.03	0.5185±0.03	0.4845±0.05	0.4044±0.02	-	0.0394±0.04	0.3152±0.03	**0.6156±0.02**
USPS	0.4648±0.02	0.1761±0.03	0.4999±0.02	0.0200±0.03	0.3756±0.04	-	-0.0001±0.00	0.3144±0.00	**0.5454±0.01**
Isolet	0.4374±0.02	0.1657±0.02	0.5555±0.02	0.4431±0.05	0.3066±0.02	-	0.0006±0.00	0.0815±0.01	**0.5591±0.01**
Yeast	0.1244±0.01	0.0969±0.01	0.1312±0.02	0.1332±0.02	0.1430±0.00	0.0030±0.00	**0.1602±0.04**	0.1182±0.00	0.1462±0.01
Avg. Rank	4.08	5.92	3.67	4.00	5.33	8.08	6.25	6.08	**1.58**

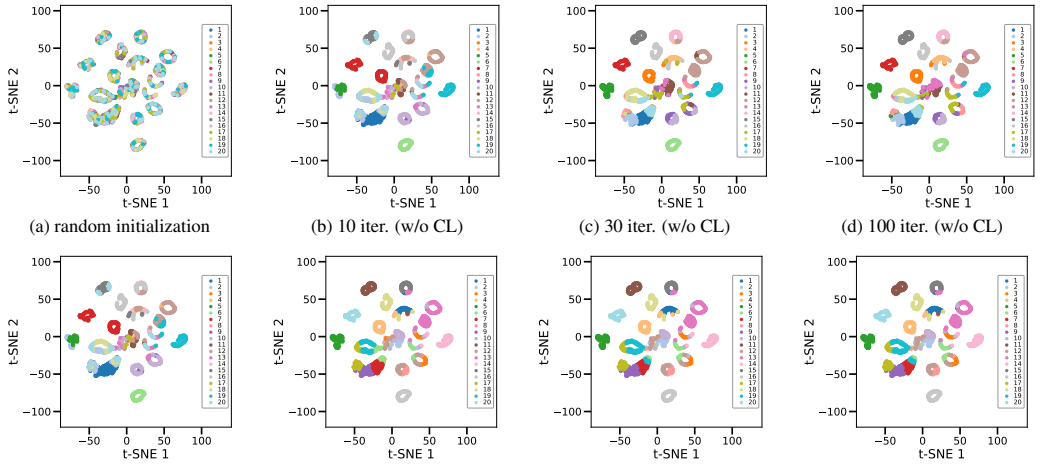
of the three evaluation metrics: Purity, NMI, and ARI. Detailed experimental results are shown in Tables 1, 2 and 3. Evidently, CL-QECM achieves state-of-the-art or highly competitive clustering performance across most benchmark datasets. Specifically, the algorithm has obtained the highest average ranking on the three evaluation indicators, which are 1.42, 1.75 and 1.58 respectively.

Additionally, to evaluate the computational feasibility of CL-QECM, we recorded the execution times of all algorithms independently over 20 trials. In order to ensure the fairness of comparison, all iterative algorithms run under the same hardware configuration until they meet their respective convergence standards. The results in Table 5 (detailed in Appendix H) show that CL-QECM converges rapidly on most small-to-medium datasets, matching lightweight models like ECM and outperforming deep baselines. Remarkably, it outpaces ECM on small datasets, as the enhanced inter-class separation significantly reduces convergence iterations. However, on high-dimensional datasets (e.g., COIL20, USPS), it becomes significantly slower than ECM. This overhead inherently stems from larger focal-sets generation and iterative phase mappings. Ultimately, CL-QECM pro-



(a) Purity (b) NMI (c) ARI

Figure 3: Nemenyi test for different clustering algorithms



(a) random initialization (b) 10 iter. (w/o CL) (c) 30 iter. (w/o CL) (d) 100 iter. (w/o CL)

(e) 10 iter. (CL-QECM) (f) 30 iter. (CL-QECM) (g) 100 iter. (CL-QECM) (h) 200 iter. (CL-QECM)

Figure 4: T-SNE visual experiments on the COIL20 dataset

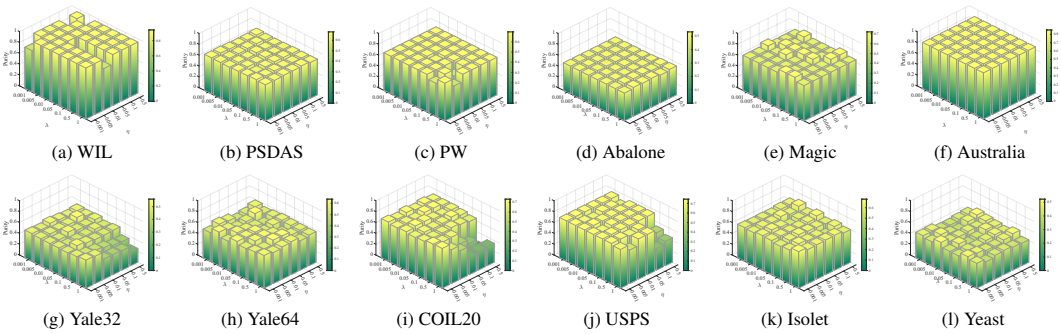
vides a practical trade-off between expressive uncertainty modeling and computational efficiency.

In order to verify whether there are significant performance differences between the proposed algorithm and the compared algorithms across multiple datasets, rigorous Friedman and Nemenyi statistical tests further confirm the significant superiority of our method (Fig. 3). The formal definitions and calculation procedures for these statistical metrics are deferred to Appendix I. The results demonstrate that CL-QECM is slightly better than DEC, CKM and IDEC in Purity indicators, and shows a statistically significant improvement compared with all other evaluation methods; in terms of NMI and ARI indicators, CL-QECM is slightly ahead of DEC, IDEC and CKM, and significantly better than all other baseline algorithms.

3.3 Ablation Study

To validate the effectiveness of the key components in our proposed CL-QECM, specifically the instance-centroid contrastive learning (CL) module and the phase information, we conduct comprehensive ablation studies. We design two sets of experiments: a quantitative assessment of clustering performance across benchmark datasets, and a qualitative topological analysis via t-distributed stochastic neighbor embedding (t-SNE). To ensure a fair comparison and minimize the effects of randomness, all models employ the same random initialization for cluster centroids.

Quantitative Evaluation. We evaluate four structural variants: a degraded baseline omitting both modules (**w/o CL & Phase**), a variant strictly without phase information (**w/o Phase**), a variant devoid of the contrastive learning (**w/o CL**), and our full model (**CL-QECM**). Detailed quantitative results (mean and standard deviation of Purity, NMI, and ARI) are provided in Table 6 of Appendix H. Empirical results reveal a consistent performance degradation upon the removal of either the CL module or the phase information on nearly all datasets, with the joint removal (**w/o CL & Phase**) yielding the most severe drop. This superiority is particularly evident on WIL and Yeast, where CL-QECM achieves substantial performance gains across all three metrics. This substantiates



(a) WIL (b) PSDAS (c) PW (d) Abalone (e) Magic (f) Australia

(g) Yale32 (h) Yale64 (i) COIL20 (j) USPS (k) Isolet (l) Yeast

Figure 5: The Purity of CL-QECM with different λ and η values on 12 datasets.

that the synergistic integration of the contrastive loss and phase information effectively establishes the angular margin within the complex credal space. This geometry-aware separation fundamentally resolves the boundary ambiguity, thereby learning more accurate cluster centroids.

Qualitative Visualization. To intuitively demonstrate the efficacy of contrastive learning in optimizing the spatial distribution, we employ t-SNE to visualize the clustering evolution of CL-QECM and its variant without contrastive learning on the COIL20 dataset, as shown in Fig. 4. As illustrated in Fig. 4a, both models start from random initialization. However, during optimization, CL-QECM explicitly reshapes the cluster morphology. Specifically, it encourages highly compact intra-cluster distributions and strictly separated inter-cluster margins. By iteration 100 (Figs. 4d, 4g), CL-QECM yields more distinguishable cluster boundaries with reduced inter-class overlap. At convergence (iteration 200, Fig. 4h), the representations are mapped into a well-separated topological state, confirming that the contrastive objective establishes the expected angular margin in the latent space.

3.4 Sensitivity analysis

We evaluate the algorithmic sensitivity to two crucial hyperparameters governing the optimization process: contrastive weight λ and the learning rate η. In order to systematically evaluate the impact of λ and η on the clustering performance, we select λ from $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0\}$, and η from $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$. For each set of parameter combinations, we ran CL-QECM 20 times under random initialization and recorded the mean clustering evaluation metrics. As shown in Fig.5, PSDAS, Abalone, Magic, Australia and Yeast are insensitive to λ and η, whereas Yale and Isolet exhibit slight sensitivity. In contrast, WIL, PW, COIL20, and USPS are more sensitive, with performance degrading when λ and η become excessively large. Across all datasets, λ has a greater impact on clustering performance because it controls the weight of contrastive loss. If the λ setting is too small, the model will degrade into the QECM, so that it cannot make full use of the contrastive discrimination information between the instance and the subset barycenter. On the contrary, if it is too large, it will obscure the basic clustering target and destroy the original quantum belief assignment structure.

4 Conclusion and future work

In this paper, a quantum evidential c-means based on contrastive learning(CL-QECM) clustering algorithm was proposed to address the fundamental limitations of nondirectional distance metrics and severe boundary overlap in traditional evidential clustering. By synergistically integrating an instance-centroid contrastive mechanism with a geometry-aware angular representation, the proposed algorithm is highly suitable for complex data scenarios with significant noise and multi-cluster overlap. In evaluations across twelve real-valued benchmark datasets, compared with other most advanced evidence clustering and contrastive clustering algorithms, the proposed CL-QECM performs best in terms of overall performance and robustness. In the future, the research direction can be further explored by integrating the CL-QECM framework with advanced deep learning backbones, such that the clustering algorithm can be effectively applied to more complex, high-dimensional raw data like images and texts.

AI USE STATEMENT

In this work, we used generative AI tools for none of the tasks with required disclosure. We have not used generative AI tools for developing theoretical models, formulating mathematical claims, or implementing core methods, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for editing the research paper to improve readability (including correcting grammar errors, rewriting sentences for clarity and fluency, and unifying the writing style), and creating or editing auxiliary software code (specifically, generating data preprocessing and visualization scripts, and writing utility functions). We have reviewed all AI-assisted work. Specifically, all final text was carefully reviewed and modified by the authors to ensure it accurately reflects our research work and contributions. Furthermore, all AI-generated auxiliary and boilerplate code was manually verified for correctness, and the core implementation of the proposed algorithm was completed entirely by the authors independently. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We confirm that all authors have read and adhered to the ICLR Code of Ethics. This work complies with the ethical principles outlined therein, and no part of this study violates ethical standards.

REPRODUCIBILITY STATEMENT

We include an algorithmic implementation of CL-QECM in Algorithm 1 to clearly illustrate our proposed contrastive quantum evidential clustering framework. In all our experiments, we use standard evaluation protocols and publicly available benchmark datasets. We introduced an instance-centroid contrastive loss and distance-induced phase reparameterization into the optimization process in the different experiments that we conducted. We provide comprehensive details in the Appendix (see Appendices E, and F) for all the datasets and baseline algorithms used, and clarify all hyperparameter configurations and hardware setups to ensure complete reproducibility.

REFERENCES

Luyuan Chen, Xinghua Zhou, Peidong Gao, Zhan Deng, Yu Yang, Yin Wu, and Pierpaolo D'Urso. DDHRPS: A data-driven hierarchical method for constructing random permutation set from the perspective of layer-2 belief structure. *IEEE Transactions on Fuzzy Systems*, 34(4):1242–1255, 2026a.

Mengyuan Chen, Junyu Gao, and Changsheng Xu. R-edl: Relaxing nonessential settings of evidential deep learning. In *International Conference on Learning Representations*, volume 2024, pp. 35563–35588, 2024.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1597–1607. PMLR, 2020.

Xiaolong Chen, Fuyuan Xiao, Xiaohong Zhang, Zehong Cao, and Chin-Teng Lin. TrustworthyQENN: A quantum evidential neural network based on complex-valued contrastive learning for uncertainty pattern classification. In *The 43rd International Conference on Machine Learning*, Seoul, South Korea, 2026b.

Zhiwei Chen, Songru Zhang, and Hongyan Dui. Importance-based risk evaluation methodology in transportation cyber-physical systems. *Frontiers of Engineering Management*, 12(2):291–304, 2025.

A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967. ISSN 00034851.

Jixiang Deng, Yong Deng, and Jian-Bo Yang. Random permutation set reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. DOI: 10.1109/TPAMI.2024.3438349, 2024.

Xinyang Deng and Wen Jiang. Upper bounds of uncertainty for Dempster combination rule-based evidence fusion systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 55(1): 817–828, 2025.

Xinyu Fan, Mi Zhou, Bayi Cheng, Jian Wu, and Hamido Fujita. Two-stage heterogeneous noncooperative behavior for consensus in large-scale group decision under fuzzy distributed preference relation. *IEEE Transactions on Fuzzy Systems*, 33(11):4003–4016, 2025.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 1753–1759, 2017.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Yingcheng Huang, Fuyuan Xiao, Zehong Cao, and Chin-Teng Lin. Fractal belief Rényi divergence with its applications in pattern classification. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8297–8312, 2024.

Lianmeng Jiao, Thierry Denœux, Zhun-ga Liu, and Quan Pan. EGMM: An evidential version of the gaussian mixture model for clustering. *Applied Soft Computing*, 129:109619, 2022.

Lianmeng Jiao, Haoyu Yang, Feng Wang, Zhun-ga Liu, and Quan Pan. DTEC: Decision tree-based evidential clustering for interpretable partition of uncertain data. *Pattern Recognition*, 144: 109846, 2023.

Changbin Li, Kangshuo Li, Yuzhe Ou, Lance Kaplan, Audun Jø sang, Jin-Hee Cho, DONG HYUN JEONG, and Feng Chen. Hyper evidential deep learning to quantify composite classification uncertainty. In *International Conference on Learning Representations*, volume 2024, pp. 8826–8860, 2024a.

Chenghai Li, Wang Ke, Song Yafei, Wang Peng, and Li Lemin. Air target intent recognition method combining graphing time series and diffusion models. *Chinese Journal of Aeronautics*, pp. DOI: 10.1016/j.cja.2024.08.008, 2025.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. Prototypical contrastive learning of unsupervised representations. *arXiv preprint arXiv:2005.04966*, 2020.

Meizhu Li, Linshan Li, and Qi Zhang. A new distance measure between two basic probability assignments based on penalty coefficient. *Information Sciences*, pp. 120883, 2024b.

Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 8547–8555, 2021.

Zhun-ga Liu, Yi-min Fu, Quan Pan, and Zuo-wei Zhang. Orientational distribution learning with hierarchical spatial attention for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8757 – 8772, 2022.

Marie-Hélène Masson and Thierry Denoeux. ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397, 2008.

Feiping Nie, Runxin Zhang, Yu Duan, and Rong Wang. Unconstrained fuzzy c-means based on entropy regularization: An equivalent model. *IEEE Transactions on Knowledge and Data Engineering*, 37:979–990, 2025. ISSN 1041-4347.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Advait Parulekar, Liam Collins, Karthikeyan Shanmugam, Aryan Mokhtari, and Sanjay Shakkottai. Infonce loss provably learns cluster-preserving representations. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 1914–1961. PMLR, 12–15 Jul 2023.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976. ISBN 9780691214696.

Anlin Song and Liguo Fei. An evidential linguistic CoCoSo method for early warning system technology selection: a risk-aware emergency decision-making approach. *International Journal of Disaster Risk Science*, pp. DOI: 10.1007/s13753–026–00694–1, 2026.

Xiaoyan Su, Xuying Huang, Xiaolei Pan, and Debiao Meng. A dependence assessment method based on quantum model of mass function in human reliability analysis. *Expert Systems with Applications*, 299:129992, 2026. ISSN 0957-4174.

Kang Sun, Daijun Wei, Mingli Lei, and Ningkui Wang. Complex evidential reasoning-based multi-source information fusion for pattern classification. *Neurocomputing*, 674:132891, 2026.

Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2495–2504, 2021.

Fuyuan Xiao. Quantum x-entropy in generalized quantum evidence theory. *Information Sciences*, 643:119177, 2023.

Fuyuan Xiao. Quantum evidence theory. *viXra*, pp. 2504.0202, Aug. 2025.

Fuyuan Xiao, Junhao Wen, Witold Pedrycz, and Masayoshi Aritsugi. Complex evidence theory for multisource data fusion. *Chinese Journal of Information Fusion*, 1(2):134–159, 2024.

Fuyuan Xiao, Weiping Ding, and Witold Pedrycz. A generalized f-divergence with applications in pattern classification. *IEEE Transactions on Knowledge and Data Engineering*, 37(4):1556 – 1570, 2025a.

Fuyuan Xiao, Yu Zhou, and Witold Pedrycz. An adaptive quantum circuit of Dempster's rule of combination for uncertain pattern classification. In *The Thirty-Ninth Neural Information Processing Systems Annual Conference (NeurIPS)*, pp. 1 – 14, 2025b.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pp. 478–487. PMLR, 2016.

Jian-Bo Yang and Dong-Ling Xu. Maximum likelihood evidential reasoning. *Artificial Intelligence*, 340:104289, 2025.

Long-Hao Yang, Fei-Fei Ye, Jun Liu, and Ying-Ming Wang. Belief rule-base expert system with multilayer tree structure for complex problems modeling. *Expert Systems with Applications*, 217: 119567, 2023.

Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. Decoupled contrastive learning. In *Computer Vision – ECCV 2022*, pp. 668–684. Springer Nature Switzerland, 2022.

Junhao Yu, Fuyuan Xiao, Yi Zhang, Zehong Cao, and Chin-Teng Lin. A CDGFN-based quantum multisource information fusion with its application in time series classification. *IEEE Transactions on Knowledge and Data Engineering*, 38(3):1427–1440, 2026.

Tao Yu, Yaxian Tang, Huizi Cui, and Bingyi Kang. A novel BWM-based conflict management method for interval-valued belief structure. *Information Sciences*, pp. 122617, 2025.

Yu Yun, Geng Yang, Qin Li, and Jane You. A novel contrastive clustering algorithm based on k-means. In *2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)*, pp. 535–542. IEEE, 2023.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pp. 12310–12320. PMLR, 2021.

Tianxiang Zhan, Yuanpeng He, and Yong Deng. Ternary coding of maximum Deng entropy. *Fuzzy Sets and Systems*, pp. 109913, 2026.

Chuyu Zhang, Hui Ren, and Xuming He. P^2ot: Progressive partial optimal transport for deep imbalanced clustering. In *International Conference on Learning Representations*, volume 2024, pp. 14196–14217, 2024.

Lang Zhang and Fuyuan Xiao. Belief Rényi divergence of divergence and its application in time series classification. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):3670–3681, 2024.

Rui Zhang, Xuelong Li, Hongyuan Zhang, and Feiping Nie. Deep fuzzy k-means with adaptive loss and entropy regularization. *IEEE Transactions on Fuzzy Systems*, 28(11):2814–2824, 2019.

Zuowei Zhang, Hongpeng Tian, Jingwei Zuo, and Weiping Ding. Aggregation-based self-supervised evidential clustering for imbalanced data. *Information Fusion*, 127:103721, 2026.

Tong Zhao, Zhen Li, and Yong Deng. Linearity in Deng entropy. *Chaos, Solitons & Fractals*, 178: 114388, 2024.

Qianli Zhou, Witold Pedrycz, and Yong Deng. Order-2 probabilistic information fusion on random permutation set. *IEEE Transactions on Knowledge and Data Engineering*, 37(2):837–850, 2025.

APPENDIX

A Related Work

A.1 Generalized Quantum evidence theory

Generalized quantum evidence theory (GQET) (Xiao, 2023) is a formal framework for uncertainty inference from the perspective of an open world quantum framework. By leveraging complex numbers, GQET enhances representational capacity, effectively modeling multi-dimensional uncertainty through phase and magnitude information. In this framework, the fundamental event space is defined by the quantum frame of discernment (QFOD) and its associated quantum hypotheses. To quantitatively express the support degree for these hypotheses, the generalized quantum mass function (GQMF) is utilized as the core metric for uncertainty evaluation. Furthermore, the generalized quantum Pignistic transformation (GQPT) is proposed to support decision-making.

Definition 1. (Quantum frame of discernment) A quantum frame of discernment (QFOD), denoted by Φ, which consists of a set of mutually exclusive and collectively nonempty events, can be defined as follows (Xiao, 2023; Aug. 2025; Xiao et al., 2025b):

$$\Phi = \{\phi_1, \phi_2, \ldots, \phi_k\}, \tag{22}$$

where each event is represented by an orthogonal basis in a Hilbert space $\mathcal{H}$.

Definition 2. (Quantum hypothesis) The power set 2^Φ of Φ is defined as (Xiao, 2023; Aug. 2025; Xiao et al., 2025b):

$$2^\Phi = \{\emptyset, \{\phi_1\}, \ldots, \{\phi_k\}, \{\phi_1\phi_2\}, \ldots, \{\phi_1\phi_2\cdots\phi_g\}, \ldots, \Phi\}, \tag{23}$$

in which $\emptyset$ represents an unknown event.

Definition 3. (Generalized quantum mass function) The generalized quantum mass function (GQMF) $\mathbb{Q}_{\mathbb{M}}$, also called a generalized quantum basic probability amplitude (GQBPA), is used to express uncertainty; it is a mapping from 2^Φ to $\mathbb{C}$:

$$\mathbb{Q}_{\mathbb{M}}(\psi_t) = \varphi(\psi_t)e^{j\theta(\psi_t)}, \quad \psi_t \in 2^\Phi \tag{24}$$

and satisfies:

$$\sum_{\psi_t \in 2^\Phi} |\mathbb{Q}_{\mathbb{M}}(\psi_t)|^2 = 1, \tag{25}$$

where $j = \sqrt{-1}$, $\varphi(\psi_t) \in [0, 1]$ denotes the magnitude of $\mathbb{Q}_{\mathbb{M}}(\psi_t)$, θ denotes a phase term of $\mathbb{Q}_{\mathbb{M}}(\psi_t)$, and $|\mathbb{Q}_{\mathbb{M}}(\psi_t)|^2$ denotes the modulus squared of $\mathbb{Q}_{\mathbb{M}}(\psi_t)$.

Definition 4. (Quantum focal element in GQET) If the proposition $\psi_t \in 2^\Phi$ satisfies $|\mathbb{Q}_{\mathbb{M}}(\psi_t)| > 0$ (or $\varphi(\psi_t) > 0$), then ψ_t is called a quantum focal element.

Depending on the type of uncertainty information represented, the generalized quantum mass function $\mathbb{Q}_{\mathbb{M}}$ is classified into the following categories:

- **Bayesian:** The quantum focal elements of GQBPA of this type are all singletons, and satisfy $\mathbb{Q}_{\mathbb{M}}(\emptyset) = 0$.
- **Vacuous:** This type of GQBPA has only a one quantum focal set Φ, for which $\mathbb{Q}_{\mathbb{M}}(\Phi) = 1$. This GQBPA represents a situation where the information is completely unknown.

Definition 5. (Generalized quantum basic probability function) The generalized quantum basic probability function of $\mathbb{Q}_{\mathbb{M}}$, also referred to as a generalized quantum basic probability distribution (GQBPD), can be defined as (Xiao, 2023; Aug. 2025; Xiao et al., 2025b):

$$\mathrm{M} : 2^\Phi \to [0, 1], \tag{26}$$

that satisfies

$$\begin{aligned} &\mathrm{M}(\psi_t) = |\mathbb{Q}_{\mathbb{M}}(\psi_t)|^2, \qquad \psi_t \in 2^\Phi, \\ &\sum_{\psi_t \in 2^\Phi} \mathrm{M}(\psi_t) = 1, \end{aligned} \tag{27}$$

in which the value of $\mathrm{M}(\psi_t)$ represents the support degree assigned to proposition ψ_t.

Definition 6. (Generalized quantum plausibility function) The generalized quantum plausibility function in GQET, denoted as GQPl, is defined as (Xiao, 2023; Aug. 2025; Xiao et al., 2025b):

$$\text{GQPl}(\psi_t) = \begin{cases} \sum\limits_{\psi_p \cap \psi_t \neq \emptyset} |\mathbb{Q}_{\mathbb{M}}(\psi_p)|^2, & \psi_t \neq \emptyset, \\ |\mathbb{Q}_{\mathbb{M}}(\psi_p)|^2, & \psi_t = \emptyset. \end{cases} \tag{28}$$

According to Eq. (27), since $\text{M} = |\mathbb{Q}_{\mathbb{M}}|^2$, Eq. (28) can also be expressed:

$$\text{GQPl}(\psi_t) = \begin{cases} \sum\limits_{\psi_p \cap \psi_t \neq \emptyset} \text{M}(\psi_p), & \psi_t \neq \emptyset, \\ \text{M}(\psi_t), & \psi_t = \emptyset. \end{cases} \tag{29}$$

Definition 7. (Generalized quantum Pignistic transformation) Consider $\mathbb{Q}_{\mathbb{M}}$ as a GQBPA defined over QFOD Φ and ψ_t be a quantum proposition with $\psi_t \in \Phi$. The generalized quantum Pignistic probability function (GQPT), denoted as BetGQP, which is defined by (Xiao, 2023; Aug. 2025):

$$\text{BetGQP}(\psi_t) = \sum_{\psi_p \subseteq \Phi} \frac{|\mathbb{Q}_{\mathbb{M}}(\psi_p)|^2}{1 - |\mathbb{Q}_{\mathbb{M}}(\emptyset)|^2} \cdot \frac{|\psi_t \cap \psi_p|}{|\psi_p|}, \tag{30}$$

where $|\cdot|$ denotes the cardinality function. Since $\text{M} = |\mathbb{Q}_{\mathbb{M}}|^2$, Eq. (30) can also be represented:

$$\text{BetGQP}(\psi_t) = \sum_{\psi_p \subseteq \Phi} \frac{\text{M}(\psi_p)}{1 - \text{M}(\emptyset)} \cdot \frac{|\psi_t \cap \psi_p|}{|\psi_p|}. \tag{31}$$

A.2 Contrastive learning

Contrastive learning is a promising paradigm of unsupervised learning that aims to map the original data into a feature space by maximizing the similarities of positive pairs and minimizing the similarities of negative pairs (Li et al., 2021). In recent years, contrastive learning has garnered significant research attention due to its effectiveness in representation learning and universality in different fields.

The performance of contrastive learning largely depends on the loss function. Recently, a large number of contrastive loss functions have been developed. Contrastive loss (Wang & Liu, 2021) is one of the most fundamental loss functions, which aims to maximize the consistency between positive sample pairs while minimizing the consistency between negative sample pairs. Triplet loss (Schroff et al., 2015) aims to support metric learning, and its core function is to help the model learn embedded representations (i.e. feature space mapping), so that similar data points are located closer in space and the distance between different data points is farther, thereby achieving robust classification ability under different conditions. N-pair loss (Yeh et al., 2022) is an extension of triplet loss that incorporates multiple positive and negative samples for a given anchor point. InfoNCE loss (Oord et al., 2018; Parulekar et al., 2023) is one of the most commonly used loss functions in contrastive learning, which optimizes the model by treating contrastive learning problems as binary classification problems.

Furthermore, various contrastive learning frameworks have been proposed from different perspectives. SimCLR (Chen et al., 2020) leverages data augmentation and contrastive learning objectives to learn powerful feature representations, whose core idea is to maximize agreement between augmented views of the same instance while minimizing agreement between different instances. This method optimizes the model with large-batch training and Normalized Temperature-scaled Cross Entropy (NT-Xent) loss. MoCo (He et al., 2020) introduces a dynamic dictionary of negative samples and uses a momentum encoder to gradually update the representation of negative samples, thereby enhancing the learning ability of the model. BYOL (Grill et al., 2020) eliminates the reliance on negative samples and employs a dual-network architecture with a bootstrapping mechanism to stabilize representation learning. SwAV (Oquab et al., 2024) optimizes the model by swapping clustering assignments across augmented views, aiming to assign similar clustering labels to augmented views of the same instance, thereby learning more discriminative feature representations. Barlow Twins (Zbontar et al., 2021) learns representations by reducing the mutual correlation between latent representations, encouraging the model to generate distinct representations for similar instances, thereby enhancing its overall discriminative power.

B Detailed derivation and theoretical discussion of optimization strategy

B.1 Decoupling of Credal Partition and Contrastive Loss

When optimizing the quantum credal partition $\mathbb{C}_{\mathbb{M}}$, directly differentiating the joint objective J with respect to $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|$ is computationally intractable due to its complex non-linear presence in the contrastive loss $\mathcal{L}_{\text{CL}}$. To ensure an exact closed-form solution, we optimize a tractable decoupled surrogate objective $\tilde{J}(\mathbb{C}_{\mathbb{M}}) = J_{\text{QEC}}(\mathbb{C}_{\mathbb{M}})$. In this stage, the contrastive term $\mathcal{L}_{\text{CL}}$ is temporarily decoupled and acts as an external constant.

Although this decoupling means the quantum credal partition update minimizes a surrogate rather than the exact gradient of the joint J, the alternating optimization remains mathematically sound. The surrogate update of $\mathbb{C}_{\mathbb{M}}$ strictly guarantees a monotonic descent of the clustering energy J_{QEC}. Subsequently, in the cluster center update stage, the centroids $\mathbb{V}$ are updated using the exact joint gradient $\nabla_{\mathbf{v}} J$ (which explicitly incorporates $\nabla_{\mathbf{v}} \mathcal{L}_{\text{CL}}$). Ultimately, this alternating strategy guarantees analytical solvability while empirically driving the steady minimization of the overall joint objective.

B.2 Phase Optimization and Numerical Stability

According to the generalized quantum evidence theory (GQET), the $|\mathbb{Q}_{\mathbb{M}_i}(\psi_t)|^2$ represents the degree of support of the proposition and is directly constrained by normalization, so the analytic solution can be obtained by the Lagrange multiplier method. In contrast, the phase term does not participate in normalization constraints. It mainly portrays the geometric direction of the complex value representation, rather than supporting the intensity itself. Therefore, if the amplitude and phase are jointly optimized at the same time, it will form a highly coupled non-convex optimization problem, and it is usually difficult to obtain a parsable solution. Based on the alternating optimization strategy widely adopted in machine learning, this paper first strictly derives the optimal update for the amplitude. Subsequently, we update the geometric direction of the complex representation through the distance-induced phase reparameterization. This preserves the probabilistic semantics of the quantum mass function while yielding a numerically stable and computationally tractable optimization procedure.

Since the phase determines the geometric orientation of a complex number representation, its update should maintain the relative geometric relationship between the samples rather than its support strength. Therefore, the squared Euclidean distance is first normalized to eliminate the scale differences between different samples. Subsequently, the normalized distance is monotonously mapped to the bounded interval $[0, \pi/2]$, so that the smaller distance corresponds to the smaller phase difference, and the larger distance corresponds to the larger angle separation. As a result, the phase can maintain the neighborhood structure in the original feature space and provide a stable geometric angle representation within the complex credal space. At the same time, limiting the phase to the bounded angle range $[0, \pi/2]$ can effectively avoid excessive angular oscillation in the iterative optimization process, thus improving the numerical stability and convergence of the algorithm. More importantly, rather than discarding the spatial distribution optimized by the contrastive loss, this geometry-aware angular mapping actively encodes the spatial separability into the complex representation. It naturally establishes an angular margin, ensuring that highly confident core samples and ambiguous boundary samples are distinctly separated not only by their amplitude but also by their geometric orientation. This fundamentally improves the cluster morphology and overcomes the boundary overlap dilemma inherent in scalar evidential clustering.

C Computational Complexity Analysis

In this section, we conduct a detailed analysis of the computational complexity of the proposed algorithm. The computational complexity of CL-QECM depends primarily on the number of samples n, features p, clusters k, and quantum focal elements n_f. In each iteration of the optimization loop, to avoid redundant computations, we first compute the distance matrix between the quantum focal element centers and the instances, which requires $\mathcal{O}(n \cdot p \cdot n_f)$ multiplications. Next, updating the

Algorithm 1 Contrastive Quantum Evidential C-Means

Input: dataset X; cluster number k.
Parameter: temperature τ; contrastive weight λ; learning rate η.
Output: quantum credal partition $\mathbb{C}_{\mathbb{M}}$ and cluster centers $\mathbb{V}$.

```
Initialize cluster centers 𝕍 and t = 0.
while not converged do
    t = t + 1
    Calculate magnitude |ℚ_𝕄(ψ_t)| using Eq (11).
    Compute phase θ using Eq. (12).
    Compute quantum credal partition ℂ_𝕄 using Eq.(13).
    if t ≤ 50 then
        Compute ℚ_ℍ and ℚ_𝔹 using Eqs.(18) and (17)
        Solve (ℚ_ℍ + γI)𝕍 = ℚ_𝔹.
    else
        Compute clustering loss gradient ∇_{J_QEC} = 2(ℚ_ℍ · 𝕍 − ℚ_𝔹).
        Compute contrastive learning loss gradient ∇_{ℒ_CL}.
        Update the cluster center 𝕍 by Eq.(21).
    end if
end while
return ℂ_𝕄, 𝕍
```

quantum credal partition and computing the gradient direction requires $\mathcal{O}(n \cdot p \cdot n_f + k^3)$ operations. Assuming the number of instances is significantly larger than the number of clusters ($n \gg k$), the $\mathcal{O}(k^3)$ term introduced by matrix inversion can be omitted. In summary, the complexity of each iteration is $O(n \cdot n_f \cdot p)$. Therefore, the overall computational complexity of CL-QECM is $O(t \cdot n \cdot p \cdot n_f)$, where t is the number of iterations.

If a complete power set is generated, then $n_f = 2^k$, resulting in a total complexity of $O(t \cdot n \cdot p \cdot 2^k)$. However, generating a complete power set will lead to an exponential increase in the number of quantum focal elements, which incurs massive memory consumption and computing overhead in real scenarios. Thus, in real-world scenarios, by limiting each focal point to contain at most two clusters, the complexity can be significantly reduced from $O(t \cdot n \cdot p \cdot 2^k)$ to $O(t \cdot n \cdot p \cdot k^2)$.

D Algorithm pseudocode

The specific steps of the proposed CL-QECM algorithm are shown in Algorithm 1.

E Datasets and Evaluation metrics

E.1 Datasets

Table 4 summarizes the statistics of all benchmark datasets used in our experiments, including the number of instances, feature dimensions, and classes for each dataset.

E.2 Evaluation metrics

- Purity is a simple and straightforward evaluation method. To calculate purity, each cluster is assigned to the most frequent cluster class, and the accuracy of this assignment is measured by calculating the number of correctly classified documents and dividing it by n. Formally,

$$\text{Purity}(\Omega, \mathbb{C}) = \frac{1}{n} \sum_{k} \max_{t} |\omega_k \cap c_t|. \tag{32}$$

Table 4: Experimental datasets

Dataset	Instance	Feature	Class
WIL	2000	7	4
PSDAS	4424	36	3
PW	11055	30	2
Abalone	4177	8	3
Magic	19020	10	2
Australian	690	14	2
Yale32	165	1024	15
Yale64	165	4096	15
COIL20	1440	1024	20
USPS	9298	256	10
Isolet	1560	617	26
Yeast	1484	8	10

- NMI (Normalized Mutual Information) is utilized to measure the similarity between two clustering results, with a value range of $[0, 1]$:

$$\text{NMI} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} n^{kl}\log\left(\frac{nn^{kl}}{n^k n^l}\right)}{\sqrt{\left(\sum_{k=1}^{K} n^k \log\left(\frac{n^k}{n}\right)\right)\left(\sum_{l=1}^{L} n^l \log\left(\frac{n^l}{n}\right)\right)}}. \tag{33}$$

- ARI (Adjusted Rand Index) measures the degree of overlap between two clustering results and takes values in the range $[-1, 1]$:

$$\text{ARI} = \frac{\sum_{it}\binom{N_{it}}{2} - \frac{\left(\sum_i \binom{a_i}{2}\right)\left(\sum_t \binom{b_t}{2}\right)}{\binom{n}{2}}}{\frac{1}{2}\left(\sum_i \binom{a_i}{2} + \sum_t \binom{b_t}{2}\right) - \frac{\left(\sum_i \binom{a_i}{2}\right)\left(\sum_t \binom{b_t}{2}\right)}{\binom{n}{2}}}. \tag{34}$$

F Experiment Setup

CL-QECM involves six main tuning hyperparameters: α, β, δ, λ, τ, and learning rate η. Specifically, the value space of β is evenly divided into 5 values within $[3, 4]$, δ into 15 values within $[1, 20]$, λ into 20 values within $[0, 1]$, and the value set of η is $\{0.005, 0.01, 0.02, 0.05\}$. To achieve optimal cluster performance and determine the best combination of hyperparameters, we used random search strategies. For other fixed hyperparameters and constants, following empirical practices and prior related literature, we set the degree of punishment $\alpha = 2$, the temperature parameter $\tau = 0.5$, the momentum coefficient $\mu = 0.9$, the regularization parameter $\gamma > 0$ and the numerical stabilization constant $\epsilon = 10^{-12}$. All datasets are normalized to the range $[0, 1]$ to mitigate the impact of different data feature scales and accelerate the convergence speed of the momentum gradient descent method. All experiments were conducted on a cloud-based computing platform equipped with an 8-core virtual CPU (vCPU) and 32 GB of RAM.

For all algorithms that require a specified number of clusters, the actual number of clusters with real values of k is provided to ensure fair conditions. Furthermore, since our benchmarks primarily comprise non-spatial vector data, the original Convolutional Neural Network (CNN) architectures and spatial augmentations of deep baselines (DEC, IDEC, DFKM, and PCL) are inapplicable. To strictly isolate and evaluate the core clustering mechanisms rather than feature-extraction capabilities, we unify all deep baselines using a standardized lightweight multi-layer perceptron (MLP) $p \to 64 \to 32 \to k$. Accordingly, spatial augmentations are substituted with feature-level additive Gaussian noise ($\sigma = 0.01$). This harmonized setup ensures a fair, mechanism-focused comparison.

G Baseline Algorithms

These comparison algorithms are briefly described as follows.

- **Ckm** (Yun et al., 2023), which combines K-means and contrastive learning into a unified model, extracts more accurate cluster centroids by considering both the similarity of a sample to its own centroid and its dissimilarities to other distinct centroids.

Table 5: Clustering performance comparison of CL-QECM with eight baselines on twelve benchmark datasets concerning runtime (seconds)

Datasets	Ckm	PCL(MLP)	IDEC	DEC	DFKM	DTEC	EGMM	ECM	CL-QECM
WIL	123.0905	206.4775	58.9197	46.0837	102.6253	0.3012	0.5228	1.6228	0.9517
PSDAS	861.6387	365.5044	38.4893	36.9889	58.3606	1918.0981	3.9200	1.0276	1.0638
PW	3937.0437	737.8548	41.0000	38.5870	84.9525	23818.7928	3.1891	0.1461	0.4951
Abalone	807.0065	344.7647	34.3406	39.6469	54.4083	1167.1402	1.9447	0.0822	1.3072
Magic	12285.7055	1140.8343	76.7365	58.4970	71.6872	2137087.3959	3.7040	0.4103	0.5858
Australia	39.8053	93.8988	24.3558	7.3456	51.4327	34.7190	0.3705	0.0245	0.0226
Yale32	12.5855	18.7951	28.6325	29.6480	9084.0378	1085.7057	3.3400	0.8684	6.8036
Yale64	19.9905	32.842	51.8053	58.9397	38040.8157	17437.0681	17.2454	1.4686	10.0825
COIL20	223.3878	358.1374	296.8613	272.0154	72917.9969	-	10.7949	3.3860	750.5466
USPS	4210.7908	965.9848	635.6302	555.9249	1942.2497	-	16.7452	7.0039	3663.6692
Isolet	207.7614	391.8234	292.0296	275.7531	46608.7871	-	10.9547	3.0378	41.8002
Yeast	84.8299	258.6052	53.2595	49.2623	82.6570	24.8937	0.9832	0.2637	1.0051

- **ECM** (Masson & Denoeux, 2008) extends fuzzy clustering with Dempster–Shafer evidence theory to generate credal partitions for explicitly representing clustering uncertainty.
- **EGMM** (Jiao et al., 2022) extends the classical Gaussian Mixture Model in the belief function framework. It generates evidential partitions to better characterize cluster-membership uncertainty.
- **DTEC** (Jiao et al., 2023) constructs an unsupervised evidential decision tree to provide interpretable partitions and handle uncertain data via meta-clusters.
- **DFKM** (Zhang et al., 2019) combines autoencoders with fuzzy k-means, employing adaptive loss and entropy regularization for robust simultaneous deep clustering.
- **DEC** (Xie et al., 2016) simultaneously learns feature representations and cluster assignments by iteratively optimizing a KL divergence-based clustering objective.
- **IDEC** (Guo et al., 2017) integrates clustering loss with an autoencoder's reconstruction loss to jointly perform deep clustering and preserve local data structure.
- **PCL** (Li et al., 2020) bridges contrastive learning and clustering via an Expectation-Maximization framework and a ProtoNCE loss to encode semantic structures.
- **UFCER** (Nie et al., 2025) introduces an unconstrained equivalent of entropy-regularized fuzzy c-means, efficiently optimizing reduced variables via gradient descent.

H Additional Experimental Results

The runtime results, averaged over 20 independent trials, are presented in Table 5.

I Statistical Significance Test

In this section, we provide the detailed mathematical formulations and stepwise analytical procedures used to obtain these statistical results. We adopt the Friedman test based on the average rankings over 12 benchmark datasets, with the significance level ϱ set to 0.05. The Friedman statistic is first computed as:

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(K-1)-\tau_{\chi^2}}, \tag{35}$$

Following the Iman–Davenport correction, the corresponding F statistic is obtained as:

$$\tau_{\chi^2} = \frac{12N}{K(K+1)}\left[\sum_{t=1}^{K}\bar{R}_t^2 - \frac{K(K+1)^2}{4}\right], \tag{36}$$

where N denotes the number of benchmark datasets, K denotes the number of algorithms, $\bar{R}_t$ represents the average rankings of the t-th algorithm on all datasets, and τ_F approximately follows a F distribution with $(K-1)$ and $(K-1)(N-1)$ degrees of freedom. The critical value is obtained from the F distribution with $(K-1)$ and $(K-1)(N-1)$ degrees of freedom. For $K=9$ and $N=12$, the critical value at the significance level $\alpha=0.05$ is $F_{0.05}=(8,88)=2.045$. Then, with

Table 6: Ablation study of different components in CL-QECM on clustering metrics.

Metrics	Datasets	Models			
		w/o CL & Phase	w/o Phase	w/o CL	**CL-QECM (Ours)**
Purity	WIL	0.9093±0.07	0.9281±0.04	0.9075±0.07	**0.9460±0.00**
	PSDAS	0.6644±0.01	0.6629±0.01	0.6531±0.03	**0.6834±0.00**
	PW	0.6927±0.00	**0.6928±0.00**	0.6927±0.00	0.6909±0.02
	Abalone	0.5101±0.00	0.5286±0.01	0.5100±0.00	**0.5318±0.01**
	Magic	0.6588±0.02	0.6592±0.02	0.6591±0.02	**0.6763±0.03**
	Australia	0.8507±0.00	0.8507±0.00	0.8507±0.00	**0.8520±0.00**
	Yale32	0.4785±0.03	0.4652±0.03	0.4815±0.03	**0.5152±0.00**
	Yale64	0.6127±0.04	0.6245±0.03	0.6103±0.03	**0.6406±0.04**
	COIL20	0.7073±0.02	0.5378±0.09	0.6343±0.05	**0.7178±0.02**
	USPS	**0.7424±0.00**	0.7404±0.01	0.7370±0.01	0.7413±0.01
	Isolet	0.6334±0.02	0.6329±0.03	0.6330±0.02	**0.6726±0.02**
	Yeast	0.4445±0.03	0.4438±0.03	0.4502±0.04	**0.5470±0.01**
NMI	WIL	0.8025±0.08	0.8244±0.03	0.8021±0.07	**0.8432±0.00**
	PSDAS	**0.1727±0.01**	0.1706±0.01	0.1596±0.04	0.1720±0.01
	PW	0.1080±0.00	**0.1081±0.00**	0.1080±0.00	0.1035±0.00
	Abalone	0.1679±0.00	0.1669±0.01	0.1677±0.00	**0.1718±0.00**
	Magic	0.0290±0.03	0.0236±0.03	0.0322±0.03	**0.0520±0.04**
	Australia	0.3913±0.00	0.3913±0.00	0.3913±0.00	**0.3942±0.00**
	Yale32	0.5428±0.02	0.5380±0.02	0.5431±0.02	**0.5508±0.00**
	Yale64	0.6344±0.02	0.6443±0.02	0.6315±0.03	**0.6515±0.03**
	COIL20	0.7722±0.01	0.6517±0.07	0.7695±0.02	**0.7763±0.01**
	USPS	**0.6170±0.00**	0.6155±0.01	0.6136±0.01	0.6166±0.00
	Isolet	0.7594±0.01	0.7569±0.01	0.7566±0.01	**0.7668±0.01**
	Yeast	0.1725±0.03	0.1743±0.02	0.1734±0.03	**0.2718±0.01**
ARI	WIL	0.8028±0.10	0.8288±0.06	0.7984±0.10	**0.8618±0.00**
	PSDAS	0.2354±0.03	0.2306±0.04	0.2206±0.06	**0.2413±0.03**
	PW	0.1485±0.00	**0.1486±0.00**	0.1484±0.00	0.1455±0.00
	Abalone	0.1445±0.00	0.1812±0.02	0.1442±0.00	**0.1901±0.01**
	Magic	0.0292±0.06	0.0283±0.06	0.0301±0.06	**0.0728±0.09**
	Australia	0.4911±0.00	0.4911±0.00	0.4911±0.00	**0.4946±0.00**
	Yale32	0.2798±0.03	0.2716±0.03	0.2791±0.03	**0.2989±0.00**
	Yale64	0.3789±0.04	0.3986±0.04	0.3735±0.04	**0.4007±0.05**
	COIL20	0.6064±0.02	0.4051±0.11	0.5455±0.05	**0.6156±0.02**
	USPS	**0.5463±0.00**	0.5426±0.02	0.5402±0.02	0.5454±0.01
	Isolet	0.5272±0.02	0.5262±0.02	0.5200±0.02	**0.5591±0.01**
	Yeast	0.0801±0.02	0.0824±0.02	0.0789±0.03	**0.1462±0.01**

$K = 9$ and $N = 12$, the average rankings in terms Purity, NMI and ARI are presented in Tables 1, 2 and 3, respectively, where a smaller average rank indicates better clustering performance. From these we calculate τ_F of Purity, NMI, ARI separately, that is, $\tau_F(Purity) = 13.504$, $\tau_F(NMI) = 9.322$ and $\tau_F(ARI) = 9.874$. It can be seen that τ_F values are greater than the critical value 2.045, indicating that the overall performance of all algorithms is significantly different.

Since there are significant differences in the performance of these algorithms, we conducted a post-hoc test, namely Nemenyi test, to find out which comparison algorithms CL-QECM specifically has statistical differences. The the Nemenyi test calculates the critical difference (CD) for the average ranking values, defined as

$$CD = q_\varrho \sqrt{\frac{K(K+1)}{6N}}, \tag{37}$$

in which q_ϱ is related to K and ϱ. In this work, $K = 9$, $\varrho = 0.05$, so $q_\varrho = 3.102$. Therefore, the critical value domain $CD = 3.468$ is calculated. The test comparison can be intuitively displayed with the Friedman test chart. According to the average ranking results in Tables 1, 2 and 3, Fig. 3 can be drawn. For each algorithm, a dot is used to show its average ranking, and the horizontal line segment centered on the dot represents the size of the critical value domain. If the horizontal line segments of the two algorithms overlap, it means that there is no significant difference between the two algorithms, otherwise it means that there is a significant difference.