JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-20, No.-8, August (2025) pp 166 - 182 ISSN (Print) 0973-8975

CONSENSUS CLUSTERING USING WEIGHT OF CLUSTERS AND CLUSTERINGS: A DUAL-WEIGHTED APPROACH

Sunandana Banerjee[1], Deepti Bala Mishra[2]

[1]Sammilani Mahavidyalaya, Kolkata, West Bengal, India.

[1] GITA Autonomous College, Bhubaneswar and BPUT, Odisha, India.

[2] GITA Autonomous College, Bhubaneswar

Email: [1]sunandana.banerjee@gmail.com, [2]mishradeeptibala@gmail.com

Corresponding Author: **Sunandana Banerjee**

https://doi.org/10.26782/jmcms.2025.08.00012

(Received: May 10, 2025; Revised: July 25, 2025; Accepted: August 07, 2025)

Abstract

This paper presents a novel consensus clustering framework that integrates both cluster-level and clustering-level weighting strategies. Traditional consensus clustering methods either weight the clusters or the base clusterings, but often fail to optimally combine these two strategies. We propose a dual-weighting scheme where weights are assigned to clusters based on internal and external consistency, and to the base clusterings based on their agreement with the ensemble. By applying a combined weight, we ensure that both high-quality clusters and consistent clusterings contribute more to the final consensus. Experimental results on several benchmark datasets demonstrate the superiority of the proposed method over existing clustering ensemble techniques.

Keywords: Consensus Clustering, Clustering Ensemble, Clustering Techniques, Dual-Weighted Approach,

I. Introduction

Clustering is a fundamental task in unsupervised learning, widely applied in areas such as data mining, pattern recognition, and image processing. Its objective is to partition a dataset into mutually exclusive and exhaustive groups, called clusters, where data points within the same cluster exhibit higher similarity to each other compared to those in different clusters. Clustering methods differ significantly in how they define and optimize the notion of similarity. As a result, no single clustering algorithm consistently performs well across different types of datasets. This observation has led to the development of clustering ensembles, which combine the outputs of multiple clustering algorithms or multiple runs of the same algorithm to produce a more reliable and robust consensus clustering.

Consensus clustering, also known as clustering ensemble, aims to find a median clustering from an ensemble of base clusterings generated by different algorithms. Empirical studies suggest that consensus clustering generally yields better results compared to the worst base clustering and, in some cases, approaches the performance of the best one. By leveraging the diversity among the base clusterings, consensus cluster- ing improves the robustness, accuracy, and stability of the final clustering. Despite its promise, the effectiveness of consensus methods largely depends on how the consensus function weighs the contributions of each base clustering and the clusters within them. A key challenge is ensuring that both the base clusterings and the individual clusters that capture significant structure in the data are emphasized in the final consensus.

Several approaches have been proposed to improve the quality of consensus clustering. These can be broadly categorized into two main strategies: cluster-level weighting and clustering-level weighting [XI]. Cluster-level weighting assigns weights to individual clusters based on their external agreement with other clusters in the ensemble. Clustering-level weighting, on the other hand, assigns weights to entire base clusterings based on their internal quality or overall consistency with the ensemble.

It has been observed that a consistent clustering may not necessarily contain high-quality clusters, and conversely, a high-quality cluster may not belong to a consistent clustering. Therefore, there is a need to combine these two complementary strategies. This paper proposes a novel method, Dual-Weighted Consensus Clustering (DWCC), which integrates both cluster-level and clustering-level weighting into a unified framework. The core idea is to assign weights to each cluster based on the internal quality of its corresponding base clustering and agreement with other clusters (as in cluster-level weighting), while also assigning weights to each base clustering based on its consistency with the ensemble (as in clustering-level weighting). By combining these two weighting schemes, we aim to build a more robust and accurate consensus clustering that reflects both the reliability of the base clusterings and the quality of the clusters within them. The proposed method differs from existing ensemble clustering techniques in that it balances the contributions of both individual clusters and base clusterings. This dual-weighting approach allows us to filter out unreliable base clusterings while still capturing valuable information from high-quality clusters.

The rest of this paper is organized as follows: Section 2 reviews related work in cluster and clustering-level weighting strategies. Section 3 presents the problem formulation. Section 4 describes our proposed method. Section 5 presents empirical evaluations of the method on several benchmark datasets. Finally, Section 6 concludes the paper and suggests avenues for future research.

II. Literature Review

Consensus clustering has evolved as a robust approach for stabilizing clustering solutions by aggregating multiple clusterings. Fred *et al.* [I] pioneered the evidence accumulation clustering method, which revealed stable cluster structures by integrating multiple clustering solutions, establishing consensus clustering as a means to enhance clustering accuracy and stability. Strehl *et al.* [II] further proposed clustering ensemble as a framework for knowledge reuse, demonstrating that combining diverse solutions leverages the strengths of each algorithm through techniques like co-association matrices and graph-based consensus functions.

A primary challenge in consensus clustering is weighing the contributions of each base clustering or cluster to optimize the final consensus. Weighted consensus clustering addresses this by assigning weights based on the quality or reliability of base clusterings or individual clusters [XVIII][XIX]. Initial efforts focused on global weighting, where a single weight is applied to each base clustering. For example, Zhou *et al.* [XX] and Berikov *et al.* [XII] used metrics such as normalized mutual information (NMI) and algorithm stability to evaluate base clustering quality, allowing more reliable clusterings to exert greater influence. Similarly, Vega-Pons *et al.* [XIII] utilized intra- and inter-cluster distances to weight base clusterings. However, global methods often overlook quality variations within clusters in the same base clustering, leading to potentially suboptimal results. To address this, local weighting strategies were developed to assign weights to individual clusters. Ali *et al.* [XVI] introduced a cluster-specific weighting approach based on cluster validity indices, while Huang *et al.* [XXI] and Nazari *et al.* [XIV] focused on cluster-level reliability measures, assigning higher weights to stable clusters to enhance their role in the final consensus. Yang *et al.* [XV] utilized cluster weights derived from internal edge stability for community detection, and Banerjee *et al.* [IV] proposed a cluster-level weight calculation based on overlap with clusters from other base clusterings.

One recent approach, the Two-Level Weighted Ensemble Clustering (TWEC) framework by Gu *et al.* [VIII], integrates global and local weighting by combining entropy-based uncertainty measures with base clustering reliability metrics. This two-tiered approach assigns weights to clusters based on uncertainty and adjusts them according to base clustering reliability, improving robustness and stability in consensus results. TWEC includes two consensus functions, Two-Level Weighted Evidence Accumulation (TWEA) and Two-Level Weighted Graph Partitioning (TWGP), which demonstrated superior performance compared to ten state-of-the-art algorithms across multiple datasets.

Another significant advancement is the Dual-Level Clustering Ensemble Algorithm (DCEA) by Shan *et al.* [X], which introduces an adaptive selection framework to eliminate low-quality or redundant base clusterings, enhancing both efficiency and accuracy. DCEA also reconstructs relation matrices using spatial location and co-occurrence frequency to improve data structure capture and uses Dempster-Shafer (DS) evidence theory to adjust reliability for conflicting clustering results.

In cases where access to large and diverse base clustering libraries is limited, traditional consensus methods often struggle to achieve high-quality results. To address this, Sedghi *et al.* [9] proposed the Clustering Ensemble Extraction (CEE) framework, which generates new clusters from existing clusterings by using Jaccard similarity to group similar clusters and extract new ones, ensuring high diversity and quality in the final consensus. CEE is particularly useful when base clusterings have low diversity or small size, and it introduces two consensus functions that enhance clustering quality without requiring access to original dataset features.

Autoencoder-based methods have also been impactful. Geddes *et al.* [XI] introduced an ensemble method for single-cell RNA-seq data using autoencoders, offering robust clustering for complex biological datasets. Li *et al.* [XII] proposed a point-cluster-clustering architecture, which balances representativeness by incorporating weighted clusterings. Spectral clustering has also contributed significantly. Huang *et al.* [XXII] developed scalable spectral clustering methods for ensemble clustering that efficiently handle large datasets, and Jia *et al.* [XXIII] demonstrated bagging-based spectral clustering for ensemble selection, underscoring the importance of ensemble diversity.

Overall, consensus clustering has evolved from basic aggregation techniques to advanced frameworks incorporating weighting, autoencoder-based methods, and uncertainty measures. These advancements make consensus clustering adaptable for various fields, such as bioinformatics and large-scale data analysis, providing robust solutions for data with high dimensionality, heterogeneity, and complexity.

III. Research Gap and Motivation for the Proposed Method

Despite the progress in consensus clustering, there remains a significant gap in methods that can effectively combine the strengths of both cluster-level and clustering-level weighting. Most existing approaches (except Gu et al. [XIII]) prioritize either individual clusters or base clusterings, but fail to integrate these two strategies in a unified framework. As a result, existing consensus clustering methods may either underutilize high-quality clusters from unreliable base clusterings or give too much weight to consistent but poorly partitioned base clusterings.

The proposed method, **Dual-Weighted Consensus Clustering (DWCC)**, aims to fill this gap by combining both cluster-level and clustering-level weighting into a single framework. By assigning weights to clusters based on internal and external quality measures, and assigning weights to base clusterings based on their entropy, DWCC balances the contributions of high-quality clusters and reliable base clusterings. This approach ensures that the final consensus clustering is robust, accurate, and aligned with the true underlying structure of the data. In the next section, we present the problem formulation.

III.i. Problem Formulation

The goal of this section is to formulate the problem of consensus clustering using both cluster-level and clustering-level weighting strategies.

Dataset and Clustering Ensemble

Let $D = \{s_1, s_2, ..., s_n\}$ represent a dataset of n data points, where each $s_i \in D$ is a data instance characterized by a set of features. A clustering algorithm partitions this dataset into clusters, grouping similar data points.

In the context of a clustering ensemble, we apply multiple clustering algorithms or repeated runs of the same algorithm with different initializations to form a collection of base clusterings. Let $E = \{C_1, C_2, ..., C_M\}$ be an ensemble of M base clusterings, where each base clustering C_a partitions the dataset D into k_a clusters:

$$C_a = \{C_{a1}, C_{a2}, ..., C_{ak_a}\}, \qquad \forall a \in \{1, 2, ..., M\}.$$

Here, C_{ak} represents the k-th cluster in the base clustering C_a, and k_a is the number of clusters in C_a. Each cluster $C_{ak} \subseteq D$ and the clusters are mutually exclusive in the same base clustering:

$$C_{ak} \cap C_{al} = \varnothing \quad \text{for} \quad k \neq l, \quad \forall a.$$

The task is to compute a consensus clustering $C^* = \{C^*_1, C^*_2, ..., C^*_k\}$, where k is the true number of clusters, which represents the best compromise between the different base clusterings, i.e., a median clustering in E.

IV. Weighted Consensus Clustering Formulation

To improve the accuracy and robustness of the consensus clustering, both clustering-level and cluster-level weights can be incorporated.

IV.i. Clustering-Level Weights:

Each base clustering $C_a \in E$ is assigned a weight w_a based on its quality or contribution to the overall ensemble. The clustering-level weights can be derived from internal validation metrics such as the silhouette index [XXVII], Dunn's index [XXVIII], or external validation measures [XXIX] that assess the clustering performance with respect to the entire ensemble:

$$W = \{w_1, w_2, ..., w_M\}, \text{ with } w_a \geq 0 \text{ and}$$

$$\sum_{a=1}^{M} w_a = 1$$

These weights are used to prioritize higher-quality clusterings in the ensemble, ensuring that more reliable clusterings have a stronger influence on the consensus.

IV.ii. Cluster-Level Weights:

Within each base clustering C_a, individual clusters C_{ak} are assigned weights μ_{ak} to reflect their stability or consistency across the ensemble. The cluster-level weights are determined based on external measurements compared to other clusters in different clusterings:

$$w_{ak} \geq 0 \quad \text{with} \quad \sum_{k=1}^{k_a} w_{ak} = 1 \quad \forall a$$

This allows the consensus process to prioritize stable, high-quality clusters, even if the base clustering as a whole may not perform well.

V. Weighted Consensus Objective Function

The consensus clustering C^* is computed by maximizing the weighted similarity between the consensus clusters and the base clusters. Let $\phi(C, C_{ak})$ be a similarity function that measures the agreement between a clustering C in the space of all possible clusterings and clusters in a base clustering C_a. Then, the objective function can be written as:

$$C^* = \underset{C \in \Omega}{argmax} \sum_{a=1}^{M} w_a \cdot \sum_{k=1}^{k_a} w_{ak} \cdot \phi(C, C_{ak}) \tag{1}$$

Where Ω is the space of all possible clusterings. By incorporating clustering-level and cluster-level weighting strategies, the final consensus clustering C^* reflects the strengths of both individual clusters and all clusterings within the ensemble. In recent literature, several heuristics have been proposed that use the objective function defined in Equation (1) to find the consensus clustering. Two of the most prominent works in this area are by Huang et al. [III] and Banerjee et al. [IV]. In the following section, we propose our method for weighted consensus and demonstrate its superiority over the aforementioned approaches.

VI. Proposed Method: Dual-Weighted Consensus Clustering (DWCC)

The Dual-Weighted Consensus Clustering (DWCC) method incorporates both *cluster-level* and *clustering-level* weighting for consensus clustering from an ensemble of base clusterings. This section explains the method step-by-step using mathematical notations. We propose the probability measure of clusters C_{ak} and C_{bl} occurring simultaneously as

$$p(C_{ak}, C_{bl}) = \frac{|C_{ak} \cap C_{bl}|}{|C_{ak} \cup C_{bl}|} \tag{2}$$

We explain the advantages of the new probability measure:

i. **Better Handling of Differing Cluster Sizes**:

ii. The original measures by Huang et al. [III] and Banerjee et al. [IV] are influenced by the size of a single cluster. In contrast, the new measure uses the size of the union of clusters, meaning that it takes into account both the overlap and the total size of the combined clusters. This normalization reduces bias towards either very small or very large clusters.

iii. **Reduced Sensitivity to Small Overlaps and Outliers**: Clusters that have a small overlap but a large union will have lower probabilities, reflecting weaker association. This helps prevent small, coincidental overlaps from having too much influence, especially when clusters have different sizes. By using the union in the denominator, the measure is also less sensitive to small outliers that might create misleading overlaps. If one cluster contains only a few points overlapping with another large cluster, the union will be large, and the probability will be small, preventing disproportionate influence from small overlaps.

iv. **Smoother Treatment of Intermediate Cases**: The new measure effectively interpolates between the extremes of total disjointness and complete overlap, assigning intermediate probabilities based on how much the two clusters overlap. This makes the clustering process smoother and potentially more robust in handling diverse datasets.

Now we propose our Cluster-Level Weighting scheme. Let $w(C_{ak})$ denote the weight of cluster C_{ak} in the base clustering C_a. The weight of each cluster is proposed to be based on two factors:

i. **Internal Quality of corresponding base clustering:** The internal quality of clustering C_a is evaluated using the Calinski-Harabasz score [XXIV], a metric that assesses the compactness and separation of the cluster. A high Calinski-Harabasz score indicates that the cluster is well-formed, with data points tightly packed within the cluster and well-separated from other clusters. *Our proposed method, however, does not depend on a specific internal measure like the one mentioned here and can be applied with any internal quality metric found in the literature.*

ii. **External Agreement (Entropy):** The external agreement of cluster C_{ak} with other clusters in different base clusterings is quantified using entropy. Entropy measures the consistency of a cluster relative to other clusters in the ensemble. A lower entropy value indicates a higher level of agreement between the cluster C_{ak} and other clusters across different base clusterings, suggesting that the cluster is more reliable.

The corresponding entropy measure using the proposed probability measure from equation 2 is computed as follows:

$$Ent(C_a) = -\sum_{b=1}^{M}\sum_{l=1}^{k_b} \frac{|C_{bl}|}{n} log_2\left(\frac{|C_{ak} \cap C_{bl}|}{|C_{ak} \cup C_{bl}|}\right) \tag{3}$$

The entropy value reflects the extent of the average agreement of the cluster C_{ak} with other base clusters in the ensemble, where lower entropy suggests a more consistent and reliable cluster.

Cluster Weight Formula

The weight for each cluster C_{ak} is now given by the following formula, which combines both *internal quality* and *external agreement*:

$$w(C_{ak}) = e^{-Ent(C_{ak})} \cdot \frac{Calinski - Harabasz(C_a)}{Max\_Score} \tag{4}$$

Where,

Ent(C_{ak}) measures the entropy of the cluster, indicating its external agreement with other clusters.

i. Calinski-Harabasz(C_a) [XXIV] evaluates the internal quality of the cluster based on its compactness and separation.
ii. Max_Score is the maximum possible Calinski-Harabasz score used for normalization.

Since the Calinski-Harabasz score evaluates cluster compactness and separation, if the clusters become less compact or less well-separated, the Calinski-Harabasz score decreases. This, in turn, lowers the overall weight $w(C_{ak})$ as it is directly proportional to this score.

Higher entropy indicates poor alignment with other clusters in the ensemble. Since the weight $w(C_{ak})$ depends on $e^{-Ent(C_{ak})}$, an increase in entropy reduces the cluster's weight, reflecting its lower consistency.

If both the internal quality (Calinski-Harabasz score) decreases and the external agreement (entropy) worsens, the overall weight $w(C_{ak})$ of the cluster will significantly decrease. The cluster will have less influence in determining the final consensus clustering, as it is both less well-formed internally and less aligned with the ensemble's overall structure.

Clustering Weight Formula

For each base clustering C_a, the clustering-level weight $w(C_a)$ is based on its consistency with the ensemble. We compute the entropy of the clustering C_a, which measures its uncertainty or inconsistency relative to other clusterings in the ensemble. The entropy is calculated using:

$$Ent(C_a) = \sum_{k=1}^{k_a} \frac{|C_{ak}|}{n} \left(-\sum_{b=1}^{M} \sum_{l=1}^{k_b} \frac{|C_{bl}|}{n} log_2 \left(\frac{|C_{ak} \cap C_{bl}|}{|C_{ak} \cup C_{bl}|} \right) \right) \tag{5}$$

A more consistent base clustering will have lower entropy and therefore should receive a higher weight. The clustering-level weight $w(C_a)$ is defined as:

$$(C_a) = 1 - \frac{Ent(C_a)}{Max\_Entropy} \tag{6}$$

where Max Entropy represents the maximum possible entropy in the ensemble.

In this formulation, a clustering-level weight closer to 1 indicates high consistency with the ensemble, while a weight closer to 0 signifies a lack of agreement with other clusterings. Thus, the weight reflects the reliability of the base clustering within the ensemble framework.

Constructing the Similarity Matrix

The next step is to aggregate the information from all the weighted clusters and clusterings into a similarity matrix *S*. The similarity matrix $S = \{s_{ij}\}$ captures the weighted similarity between data points s_i and s_j. Each entry s_{ij} represents the combined effect of all base clusterings and their respective clusters.

For each base clustering $C_a \in E$, and each cluster $C_{ak} \in C_a$, the similarity between data points s_i and s_j is updated as follows:

$$s_{ij} = s_{ij} + w(C_a)\cdot w(C_{ak})\cdot \delta_{ij}(C_{ak}), \tag{7}$$

where,

$$\delta_{ij}(C_{ak}) = \begin{cases} 1 & if\ s_i \in C_{ak}\ and\ s_j \in C_{ak} \\ 0 & otherwise \end{cases}$$

This ensures that data points belonging to the same high-quality cluster from a reliable base clustering receive higher similarity scores.

Consensus Clustering from the Similarity Matrix

Once the similarity matrix *S* has been constructed, the final consensus clustering C^* is obtained by applying a clustering algorithm to the matrix. In this work, we use *agglomerative hierarchical clustering* to produce the consensus clustering. Hierarchical clustering uses the similarity scores in *S* to iteratively group data points into clusters, ensuring that the final consensus reflects both the internal structure of the data and the reliability of the base clusterings. The proposed algorithm is summarised in Algorithm 1.

VII. Discussion

The Dual-Weighted Consensus Clustering (DWCC) algorithm presents a robust approach to achieving high-quality consensus clustering by effectively integrating results from multiple base clusterings. By utilizing a dual weighting mechanism that incorporates both cluster-level and clustering-level weights, DWCC prioritizes high-quality clusters and those that consistently align with others in the ensemble. The use of entropy as a measure of external agreement allows the algorithm to capture the complexity of data structures, while the adaptive similarity matrix enhances accuracy in pairwise comparisons. This combination helps mitigate the effects of noise and outliers, leading to more stable and reliable clustering outcomes.

Since we are using the Calinski-Harabasz internal measurement, which captures the data distribution, the algorithm does not solely rely on the quality of the input ensemble. Therefore, the algorithm's performance is less dependent on the quality of

Algorithm 1: Dual-Weighted Consensus Clustering (DWCC)

Input: Dataset $D = \{s_1, s_2, \ldots, s_n\}$,
Ensemble of base clusterings $E = \{C_1, C_2, \ldots, C_M\}$
Output: Consensus clustering C^*

Initialize similarity matrix $S = \{s_{ij}\}$ of size $n \times n$ to zero;
for *each base clustering* $C_a \in E$ **do**
  Compute clustering-level weight $w(C_a) = 1 - \frac{\text{Ent}(C_a)}{\text{Max Entropy}}$;
  for *each cluster* $C_{ak} \in C_a$ **do**
    Compute cluster-level weight $w(C_{ak}) = e^{-\text{Ent}(C_{ak})} \cdot \frac{\text{Calinski-Harabasz}(C_{ak})}{\text{max score}}$;
for *each pair of data points* $s_i, s_j \in D$ **do**
  if $s_i \in C_{ak}$ *and* $s_j \in C_{ak}$ **then**
    Update similarity matrix $s_{ij} = s_{ij} + w(C_a) \cdot w(C_{ak}) \cdot \delta_{ij}(C_{ak})$;
Perform agglomerative clustering on the similarity matrix S to obtain the consensus clustering C^*;
return Consensus clustering C^*

the initial base clusterings. However, if the base clusterings are very poor or highly inconsistent, the consensus produced may still be suboptimal.

The time complexity for computing cluster-level entropy and clustering-level entropy is $\Theta(n)$ and $\Theta(n^2)$, respectively. Hence, computing the weight of a cluster is $\Theta(n^2)$. Building the similarity matrix takes $\Theta(n^2)$, and since the time complexity of Agglomerative Consensus Clustering is $\Theta(n^2 \log n)$, the overall time complexity of DWCC is $\Theta(n^2 \log n)$. Therefore, DWCC may lead to longer runtimes, particularly with larger datasets.

Vii. i. Experimental Setup

This section presents the experimental setup for evaluating the proposed Dual-Weighted Consensus Clustering (DWCC) method. We detail the datasets used, the comparison methods, the evaluation metrics, and the results obtained from the experiments.

Datasets

The performance of the DWCC method is evaluated on several well-known benchmark datasets [XVII] commonly used in clustering research. The datasets chosen for this study include:

ii. Iris: A classic dataset containing 150 samples of iris flowers, characterized by four features (sepal length, sepal width, petal length, petal width). The dataset includes three classes, with 50 samples each.

iii. Wine: A dataset comprising 178 samples of wine, described by 13 chemical properties. The dataset consists of three classes corresponding to different wine cultivars.

iv. Breast Cancer: This dataset contains 569 samples with 30 features related to breast cancer diagnostic attributes. The samples are classified into two classes: malignant and benign.

Sunandana Banerjee et al.

v. Glass Identification: Comprising 214 instances, this dataset has nine attributes related to different types of glass. The samples are classified into six categories based on their chemical composition.

vi. Ecoli: A dataset of 336 samples, characterized by seven features related to the structure of proteins. It includes eight classes based on protein localization.

vii. Blob: A synthetic dataset created to assess clustering algorithms, containing three distinct clusters. Each cluster has a specific geometric shape and distribution.

Experimental Methodology

To evaluate the performance of DWCC, we compare it with several established consensus clustering methods:

i. Evidence Accumulation Clustering (EAC): It is a foundational consensus clustering approach proposed by Fred et al. [I]. EAC converts base clusterings into a co-association matrix and generates the final consensus clustering through hierarchical agglomerative clustering using the single linkage method.

ii. Locally Weighted Ensemble Accumulation (LWEA): It is proposed by Huang et al. [III] and it represents an early approach to using cluster weights for effective weighted consensus clustering. It applies an entropy-based measure to determine cluster weights and performs average linkage agglomerative clustering on the weighted co-association matrix at the cluster level to obtain the consensus.

iii. Two-Level Weighted Ensemble Clustering (TWEC): Introduced by Gu et al. [VIII], employs the entropy measure from the LWEA method and integrates it with NMI similarity values of the base clusterings to determine the consensus.

iv. Clustering Selected Weighted Clusters (CSWC): The method, introduced by Banerjee et al. [V], is a Cluster Selection approach that employs an entropy-based weighting system to assess cluster consistency in forming the final consensus.

v. Clustering Ensemble Extraction (CEE): developed by Sedghi et al. [IX], utilizes the Jaccard similarity measure to identify new clusters and achieve consensus.

Implementation Details

The experiments are implemented using Python with popular libraries such as scikit-learn for clustering algorithms and NumPy for numerical computations. The ensemble of base clusterings is generated using the K-Means algorithm and its variants (Mini Batch K-Means and Bisecting K-Means) with different random initializations. For each dataset, we create 100 base clusterings to ensure sufficient diversity.

Sunandana Banerjee et al.

For the comparison methods, we adjust their parameters according to their respective requirements. The number of clusters for K-Means is set to the ground truth number of clusters in each dataset.

Evaluation Metrics

To evaluate the performance of the clustering algorithms, we utilize several metrics:

i. Normalized Mutual Information (NMI) [II]: NMI measures the similarity between predicted clusters and ground truth clusters, with values ranging from 0 (no agreement) to 1 (perfect agreement). It is calculated based on mutual information and is normalized by the entropy of each clustering to ensure a consistent scale.

ii. Adjusted Rand Index (ARI)[XXV]: ARI quantifies the similarity between two clustering results by considering the number of pairs of samples that are assigned to the same or different clusters in both the predicted and true clustering. The values range from -1 (no similarity) to 1 (perfect agreement).

iii. Adjusted Mutual Information (AMI)[XXVI]: AMI adjusts the Mutual Information (MI) score between two clustering results for the chance of random agreement. It takes values between 0 and 1, with 0 indicating no correlation between the clusters and 1 representing perfect agreement. Unlike NMI, AMI accounts for the expected similarity between two random clusterings, making it a more reliable measure when the number of clusters varies.

VIII. Results and Discussion

The results of the experiments are presented in Tables 1, 2, 3, 4, 5, and 6.

Table 1: Performance on Iris Dataset

Method	ARI	NMI	AMI
DWCC	0.75619	0.75713	0.75409
EAC	0.71634	0.74191	0.73865
LWEA	0.72822	0.73873	0.73545
TWEC	0.72822	0.73873	0.73545
CSWC	0.74240	0.75182	0.74870
CEE	0.72822	0.73873	0.73545

Table 2: Performance on Wine Dataset

Method	ARI	NMI	AMI
DWCC	0.39434	0.44239	0.43639
EAC	0.39135	0.39559	0.38920
LWEA	0.39135	0.39559	0.38920
TWEC	0.39459	0.39680	0.39042
CSWC	0.39135	0.39559	0.38920
CEE	0.39135	0.39559	0.38920

The experimental results across multiple datasets, detailed in Tables 1-6, demonstrate the efficacy of the proposed DWCC (Dual-Weighted Consensus

Sunandana Banerjee et al.

Clustering) method compared to established clustering ensemble techniques, namely EAC, LWEA, TWEC, CSWC, and CEE. For the Iris dataset, DWCC achieves an ARI of 0.756 and an NMI of 0.757, outperforming all other methods. In the Wine dataset, DWC.

Table 3: Performance on Breast Cancer Dataset

Method	ARI	NMI	AMI
DWCC	0.49142	0.46479	0.46401
EAC	0.28724	0.31908	0.31797
LWEA	0.41060	0.41597	0.41508
TWEC	0.48623	0.46074	0.45994
CSWC	0.41060	0.41597	0.41508
CEE	0.41060	0.41597	0.41508

Table 4: Performance on Ecoli Dataset

Method	ARI	NMI	AMI
DWCC	0.41899	0.59699	0.57951
EAC	0.36958	0.58778	0.57002
LWEA	0.37796	0.57881	0.56067
TWEC	0.31533	0.53243	0.51219
CSWC	0.37878	0.57901	0.56275
CEE	0.40714	0.59245	0.57485

Table 5: Performance on Glass Identification Dataset

Method	ARI	NMI	AMI
DWCC	0.27066	0.40938	0.38098
EAC	0.26585	0.39053	0.36119
LWEA	0.26585	0.39053	0.36119
TWEC	0.26033	0.38569	0.35681
CSWC	0.26706	0.39680	0.36782
CEE	0.26772	0.39909	0.37016

shows strong performance, although TWEC marginally surpasses it in ARI. However, DWCC still achieves notable scores in NMI and AMI. The Breast Cancer dataset results further validate DWCC's strengths, with DWCC achieving an ARI of 0.491 and an NMI of 0.464, outperforming other methods by a substantial margin. For the Ecoli dataset, DWCC's performance is similarly impressive, with an ARI of 0.419 and an NMI of 0.597. On the Glass Identification dataset, DWCC maintains competitive performance with an ARI of 0.271 and an NMI of 0.409, slightly outperforming other methods. Finally, on the Blob dataset, DWCC outperforms other techniques with an ARI of 0.177 and an NMI of 0.180. In summary, the experiments conducted demonstrate the effectiveness of the DWCC method in improving consensus clustering outcomes. The combination of cluster-level and clustering-level weighting strategies results in a robust clustering solution that consistently outperforms existing consensus methods, validating the proposed approach's capability to address the research gaps identified in previous works.

Sunandana Banerjee et al.

Table 6: Performance on Blob Dataset

Method	ARI	NMI	AMI
DWCC	0.17665	0.18045	0.17892
EAC	0.11393	0.15672	0.15506
LWEA	0.09552	0.11090	0.10920
TWEC	0.13242	0.14890	0.14733
CSWC	0.12032	0.13385	0.13225
CEE	0.09552	0.11090	0.10920

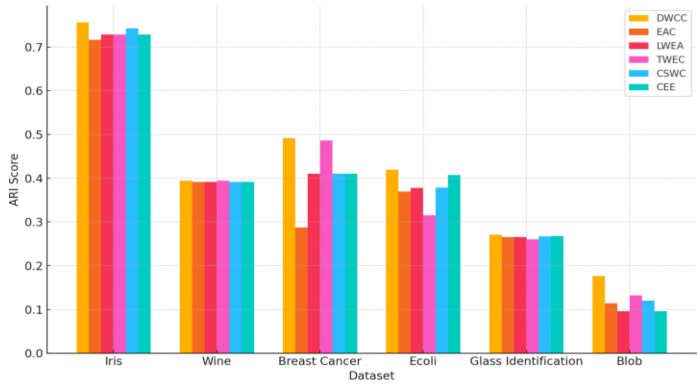

Fig. 1 Adjusted Rand Index (ARI) scores across different datasets for each clustering method.

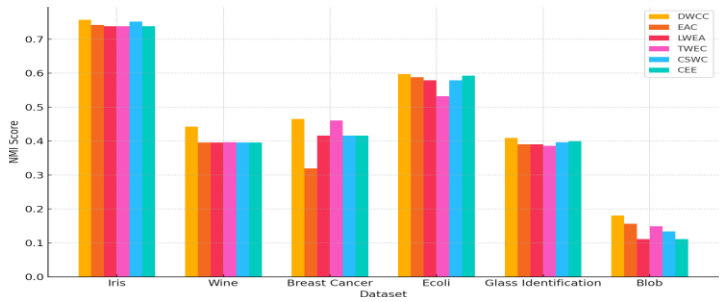

Fig. 2 Normalized Mutual Information (NMI) scores across different datasets for each clustering method.

VIII. Conclusion and Future Work

This paper presents a novel consensus clustering approach, Dual-Weighted Consen- sus Clustering (DWCC), which combines cluster-level and clustering-level weighting strategies to improve the robustness and accuracy of consensus clustering. By weight- ing clusters based on internal quality of corresponding base clustering and consistency with other clusters in the ensemble, and assigning weights to base clusterings based on their reliability, the DWCC method enhances the final clustering outcome. Experimental results on benchmark datasets demonstrate the effectiveness of DWCC over other consensus clustering methods, showing that DWCC better captures the underlying structure of data, even in complex and varied datasets.

Future work will address: (1) conducting an analytical comparison of DWCC with LWEA, TWEC, and CEE through contrasting tables of objective functions/aggregation steps, convergence analysis, and synthetic examples demonstrating behavioral differences; (2) performing sensitivity analyses on weighting parameters (including entropy computation variants, CH normalization bounds, and linear vs. multiplicative combinations) to identify component dominance scenarios; and (3) expanding validation to larger/noisy datasets (e.g., MNIST, 20-Newsgroups) with runtime statistics and additional metrics like homogeneity and silhouette. Additionally, we will explore extending DWCC to dynamic/online clustering scenarios to further assess its utility across diverse applications.

Conflict of Interest:

The authors declare that there is no conflict of interest, the investigation did not involve human participants or animals, and informed consent is not applicable.

References

I. Ali, N., M. S. Uddin, J. M. Kim: "A Cluster Validity Index-Based Weighted Clustering Ensemble Framework." Information Sciences, vol. 478, 2019, pp. 238-55. 10.1016/j.ins.2018.11.053.

II. Banerjee, A.: "Leveraging Frequency and Diversity Based Ensemble Selection to Consensus Clustering." 2014 Seventh International Conference on Contemporary Computing (IC3), 2014, pp. 123-29.

III. Banerjee, A., B. Pati, C. Panigrahi: "SC2: A Selection Based Consensus Clustering Approach." International Conference on Advanced Computing and Intelligent Engineering (ICACIE 2016), Springer, 2016.

IV. Banerjee, A., C. Panigrahi, B. Pati, S. Nayak: "Entropy-Based Cluster Selection." Progress in Advanced Computing and Intelligent Engineering, Springer, 2021, pp. 313-21.

V. Banerjee, A., S. Nayak, C. Panigrahi, B. Pati: "Clustering Ensemble by Clustering Selected Weighted Clusters." International Journal of Computational Science and Engineering, vol. 27, no. 2, 2024, pp. 159-66.

VI. Calinski, T., J. Harabasz: "A Dendrite Method for Cluster Analysis." Communications in Statistics - Theory and Methods, vol. 3, no. 1, 1974, pp. 1-27. 10.1080/03610927408827101.

Sunandana Banerjee et al.

VII. Dunn, J. C.: "Well-Separated Clusters and Optimal Fuzzy Partitions." Journal of Cybernetics, vol. 4, no. 1, 1974, pp. 95-104. 10.1080/01969727408546059.

VIII. Dua, D., C. Graff: UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2019. archive.ics.uci.edu/ml.

IX. Fred, A. L., A. K. Jain: "Finding Consistent Clusters in Data Partitions." Proc. 3rd Int. Workshop on Multiple Classifier Systems, 2002, pp. 309-18.

X. Geddes, T.: "Autoencoder-Based Cluster Ensembles for Single-Cell RNA-seq Data Analysis." BMC Bioinformatics, vol. 20, 2019, p. 660.

XI. Gu, Q., Y. Wang, P. Wang, X. Li, L. Chen, N. N. Xiong, D. Liu: "An Improved Weighted Ensemble Clustering Based on Two-Tier Uncertainty Measurement." Expert Systems with Applications, vol. 238, 2024, art. 121672. 10.1016/j.eswa.2023.121672.

XII. Halkidi, M., Y. Batistakis, M. Vazirgiannis: "On Clustering Validation Techniques." Journal of Intelligent Information Systems, vol. 17, no. 2, 2001, pp. 107-45.

XIII. Huang, D., C. Wang, J. Lai: "LWMC: A Locally Weighted Meta-Clustering Algorithm for Ensemble Clustering." Neural Information Processing: 24th International Conference, Springer, 2017, pp. 167-76.

XIV. Huang, D., C. Wang, J. Lai: "Locally Weighted Ensemble Clustering." IEEE Transactions on Cybernetics, vol. 48, no. 5, 2018, pp. 1460-73.

XV. Huang, D., C. Wang, J. Wu, J. Lai, C. Kwoh: "Ultra-Scalable Spectral Clustering and Ensemble Clustering." IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 6, 2020, pp. 1212-26. 10.1109/TKDE.2019.2903410.

XVI. Jia, J., X. Xiao, B. Liu, L. Jiao: "Bagging-Based Spectral Clustering Ensemble Selection." Pattern Recognition Letters, vol. 32, no. 10, 2011, pp. 1456-67.

XVII. Li, N., S. Xu, H. Xu: "A Point-Cluster-Partition Architecture for Weighted Clustering Ensemble." Neural Processing Letters, vol. 56, 2024, art. 183. 10.1007/s11063-024-11618-9.

XVIII. Nazari, A., M. Kargari, B. Mohammadzadeh Asl, R. Hosseini: "A Comprehensive Study of Clustering Ensemble Weighting Based on Cluster Quality and Diversity." Pattern Analysis and Applications, vol. 22, no. 1, 2019, pp. 133-45.

XIX. Rand, W. M.: "Objective Criteria for the Evaluation of Clustering Methods." Journal of the American Statistical Association, vol. 66, no. 336, 1971, pp. 846-50. 10.1080/01621459.1971.10482238.

Sunandana Banerjee et al.

XX. Rousseeuw, P. J.: "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." Journal of Computational and Applied Mathematics, vol. 20, 1987, pp. 53-65. 10.1016/0377-0427(87)90125-7.

XXI. XXI. Sedghi, M., E. Akbari, H. Motameni, T. Banirostam: "Clustering Ensemble Extraction: A Knowledge Reuse Framework." Advances in Data Analysis and Classification, 2024. 10.1007/s11634-024-00588-4.

XXII. Shan, Y., S. Li, F. Li, Y. Cui, M. Chen: "Dual-Level Clustering Ensemble Algorithm with Three Consensus Strategies." Scientific Reports, vol. 13, 2023. 10.1038/s41598-023-49947-9.

XXIII. Strehl, A., J. Ghosh: "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions." Journal of Machine Learning Research, vol. 3, 2002, pp. 583-617.

XXIV. Tatarnikov, V. V., I. A. Pestunov, V. B. Berikov: "Centroid Averaging Algorithm for a Clustering Ensemble." Computer Optics, vol. 41, no. 5, 2017, pp. 712-18. 10.18287/2412-6179-2017-41-5-712-718.

XXV. Vega-Pons, S., E. Correa-Morris, J. Ruiz-Shulcloper: "Weighted Partition Consensus via Kernels." Pattern Recognition, vol. 41, no. 6, 2008, pp. 1943-53. 10.1016/j.patcog.2007.11.018.

XXVI. Vinh, N. X., J. Epps, J. Bailey: "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance." Journal of Machine Learning Research, vol. 11, 2010, pp. 2837-54.

XXVII. Yang, H., S. Wang, C. Zhou: "A Novel Clustering Ensemble Method Based on Community Detection in Complex Networks." Neurocomputing, vol. 214, 2016, pp. 598-605. 10.1016/j.neucom.2016.06.012.

XXVIII. Zhang, M.: "Weighted Clustering Ensemble: A Review." Pattern Recognition, vol. 124, 2022, art. 108428. 10.1016/j.patcog.2021.108428.

XXIX. Zhou, Z., W. Tang: "Clusterer Ensemble." Knowledge-Based Systems, vol. 19, no. 1, 2006, pp. 77-83.

Sunandana Banerjee et al.