2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

A novel robust kernelized FCM based multi-objective simultaneous learning framework for clustering and classification

1st Saketh Innani
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
saketh20ucse163@mahindrauniversity.edu.in

2nd Pawan Chinnari
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
pawan20ucse119@mahindrauniversity.edu.in

3rd Soumen Sinha
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
soumen20ucse179@mahindrauniversity.edu.in

4th Mehek Khan
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
mehek20uari094@mahindrauniversity.edu.in

5th Neha Bharill
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
neha.bharill@mahindrauniversity.edu.in

6th Om Prakash Patel
Department of Computer Science and Engineering
Ecole Centrale School of Engineering, Mahindra University
Hyderabad, India
omprakash.patel@mahindrauniversity.edu.in

***Abstract*—Clustering and classification are the two important tasks involved in pattern recognition. Both tasks are interrelated with each other. The generalization ability of classification learning can be enhanced with clustering results. On the contrary, the class information helps in improving the accuracy of clustering learning. Thus, both learning strategy complements each other. To amalgamate the benefits of both learning strategies, therefore in this paper, we proposed a novel robust kernelized Fuzzy c-Means based multi-objective simultaneous learning framework (RKFCM-MSCC) for both clustering and classification. RKFCM-MSCC employs multiple objective functions to compose the clustering and classification problem, respectively. Both the formulated objective functions are simultaneously optimized using the particle swarm optimization approach. Moreover RKFCM-MSCC uses Bayesian theory that make these multiple objective functions dependent on the single parameter i.e., cluster centers that connect both the clustering and classification learning. The Pareto-optimal solution attained with the RKFCM-MSCC approach complements the clustering and the classification learning process. The effectiveness of the proposed RKFCM-MSCC is empirically investigated on four benchmark datasets and the results are compared with the state-of-the-art approaches.**

***Index Terms*—Clustering, Classification, Kernelized Fuzzy c-Means, Multiobjective Optimization, Pareto Optimal Solution**

I. Introduction

Pattern recognition consists of two parts, i.e., clustering and classification learning. Clustering and classification are the two fundamental problems in machine learning. Clustering learning is a technique used to group a set of data points into formed clusters based on their similarities. The formed groups or clusters are analyzed to explore the underlying structure of the data and to gain a better understanding of the nature of the data. On the other hand, classification learning is a technique used to construct a discriminant function that can distinguish between samples with different class labels. The goal is to learn a function that can map a set of input features to an output class label, which can then be used to classify newly encountered samples. Both clustering and classification learning played an important role in various fields, such as data mining, image processing, and pattern recognition. It has been seen that clustering results analyze the structure of the data that helps in improving the ability of classification learning, thus extract prior knowledge as much as possible for a given problem to enhance the generalization capability of a classifier. Our proposed algorithm also gives a positive result on the above assertion. Alternatively, incorporating class information can enhance the performance of clustering learning. Some supervised clustering and semi-supervised clustering methods have been developed that uses class information to direct the clustering process. The experimental outcomes of these approaches have shown that class information can considerably enhance the effectiveness of clustering results. Therefore, we can conclude that both clustering and classification learning are complementary to each other.

Clustering and classification learning are typically established using various models or criteria, making it challenging to combine both into a single framework. In the past, many algorithms were proposed by researchers [1], [3]–[8] to handle clustering and classification learning sequentially or independently to show the combined benefits of both learning strategies. These algorithms first optimize the clustering process using the clustering criterion, allowing the data structures to be explicitly exposed. Then the algorithms optimize the classification criterion connected with the obtained structural information based on the clustering result to assign the class label to new samples. Such algorithms sequentially optimize the clustering and classification criterion but fail to simultaneously optimize such two criteria.

In the past, several approaches were proposed by researchers inheriting the advantages of both clustering and classification learning. Here, we presented a review of some of the work carried out in this field. Maglogiannis et al. [5] proposed a radial basis function neural network (RBFNN) which is a feed-forward multi-layer network. This approach first executes unsupervised clustering learning to find out the parameters of the basis function [5] then, it optimizes the connection weights between the hidden and output layers using the mean squared error (MSE) classification criterion between the target and actual outputs. However, RBFNN cannot truly combine the benefits of clustering and classification learning in a single method. Cai et al. [1] proposed a robust fuzzy relational classifier (RFRC) to enhance the robustness of the FRC classifier. In this classifier to enhance the robustness firstly the kernelized fuzzy c-means (KFCM) [2] is used and then the soft class labels are used to replace the hard class labels. In this way, incorporating the soft class labels and KFCM makes the RFRC classifier reflect a better relationship between classes and clusters and hence significantly boosts the robustness and accuracy of the FRC classifier. Kim and Oommen [4] proposed the VQ + LVQ3 algorithm. It first employs learning vector quantization (LVQ) to optimize the locations and class labels of the cluster centers and then employs the 1NN classifier to perform classification on top of the obtained cluster centers. Kuo et al. [9], proposed a sequential clustering and classification approach using a deep learning technique and a multi-objective sine-cosine algorithm. In this study, they introduced a novel data analytics-based sequential clustering and classification (SCC) approach.

All of the methods discussed above first optimize the clustering criterion and then optimizes the classification criterion associated with the clustering result, resulting in a two-step learning paradigm that fails to achieve simultaneous optimization for both criteria. To overcome this problem, Cai et al. [10] proposed a multi-objective simultaneous learning framework (MSCC) for both clustering and classification learning. In this study, they utilize multiple objective functions and then simultaneously optimized the clustering centers embedded in these functions, this not only improved clustering performance but also simultaneously attain promising classification performance. The problem with the MSCC framework is that it is not robust against outliers and not suitable for the non-spherical data structure. To overcome this problem, we proposed a novel robust kernelized Fuzzy c-Means based multi-objective simultaneous learning framework (RKFCM-MSCC) for both clustering and classification learning. The proposed RKFCM-MSCC framework works toward enhancing the robust- ness against outliers, accuracy, and it is suitable for the non-spherical datasets.

II. The Proposed Method

To achieve robust clustering and classification simultaneously, we proposed a novel robust kernelized Fuzzy c-Means based multi-objective simultaneous learning framework (RKFCM-MSCC) for both clustering and classification. To implement this approach, we first perform the clustering using the kernelized clustering algorithm that uses the radial basis kernel function. Then, we employ the Bayesian theory to establish the connection between both the clustering and classification objective functions by making them only dependent on the same set of cluster centers which is considered a parameter to optimize. We then employ a multi-objective framework to formulate the clustering and classification problems. Finally, we use MOPSO [11] to simultaneously optimize the clustering centers incorporated in both of these objective functions. In the subsequent section, we explained the step-by-step process of the proposed RKFCM-MSCC framework that includes training and testing of the classifier and then the optimization of both the objective functions embedded with cluster centers using the MOPSO approach.

A. Training of the Classifier

The training of the classifier comprised of two steps: Firstly, the kernelized Fuzzy c-Means algorithm is applied to the training data for the exploration of the underlying structure of the data. Then in the second step, a fuzzy relational matrix $\mathbf{P}$ is constructed from the obtained membership matrix and the given class labels to uncover the statistical relationship between the formed clusters and the given classes.

1) Kernel Based Fuzzy c-Means: The kernel based Fuzzy c-Means make use of the kernel trick for projecting the non-linear problem (from original low dimensional input space) into linear problem in high dimensional feature space [12]. The kernel function K for all x_i, v_k in the original input space $\mathbf{X}$ satisfies the inner product operation as follows:

$$K(x_i, v_k) = \phi(x_i)^T \phi(v_k) \tag{1}$$

Where ϕ represents an implicit non-linear map from the input space X to a high dimensional feature space R.

$$\phi : x \rightarrow \phi(x) \in \mathbb{R}^d \tag{2}$$

Where $x_i,\ v_k \in \mathbb{R}^d$ such that $i = 1,..,S$ and $k = 1,...,K$. Thus, through this mapping ϕ, the kernelized version of FCM [1] is determined as follows:

$$J_u(\{v_k\}) = \sum_{i=1}^{S} \sum_{k=1}^{K} m_{ik}^u \left\| \phi\left(x_i\right) - \phi\left(v_k\right) \right\|^2, \quad u > 1 \tag{3}$$

Here, each data point x_i satisfy the constraint $\sum_{k=1}^{K} m_{ik} = 1$. Through kernel substitution, we get the following equations:

$$\|\phi(x_i) - \phi(v_k)\|^2 = K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k) \tag{4}$$

In the case of RBF kernel (also called Gaussian Kernel), $K(x_i, x_i)$ = 1 and $K(v_k, v_k)$ = 1 [12]. Thus, after substituting these values in Eq. (4), it is simplified and represented as follows:

$$\|\phi(x_i) - \phi(v_k)\|^2 = 2(1 - K(x_i, v_k)) \tag{5}$$

In this work, $K(x_i, v_k)$ represents the Radial Basis Function (RBF) kernel, which is defined as follows:

$$K(x_i, v_k) = \exp\left(-\|x_i - v_k\|^2 / \sigma^2\right) \tag{6}$$

Where σ represents the kernel parameter. The selection of kernel parameter play an important role in achieving clustering results. The kernel parameter is defined as follows:

$$\sigma^2 = \frac{\max_{1 \le i \le S} \|\boldsymbol{x}_i - \overline{\boldsymbol{x}}\|^2}{\lambda} \tag{7}$$

Where $\bar{x}$ is the average of all x_i. The scale factor λ is selected from $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 15\}$ according to the trial-and-error approach [13]. To minimize the objective function defined in Eq. (3), the membership matrix m_{ik} and the updated cluster center v_k need to be computed.

$$m_{ik} = \frac{(1 - K(x_i, v_k))^{-1/(u-1)}}{\sum_{k=1}^{K} (1 - K(x_i, v_k))^{-1/(u-1)}} \tag{8}$$

$$v_k = \frac{\sum_{i=1}^{S} m_{ik}^u K(x_i, v_k) x_i}{\sum_{i=1}^{S} m_{ik}^u K(x_i, v_k)} \tag{9}$$

Where, $V = [v_1, v_2......v_K]$. After substituting the value from Eq. (5) in Eq. (3), the objective function defined in Eq. (3) is simplified and denoted as:

$$J_u(\{v_k\}) = 2\sum_{i=1}^{S}\sum_{k=1}^{K} m_{ik}^u (1 - K(x_i, v_k)) \tag{10}$$

After substituting the membership matrix from Eq. (8) in Eq. (10), we now compute the final objective function of the clustering mechanism [10] which is defined as follows:

$$J_u(\{v_k\}) = 2\sum_{i=1}^{S}\sum_{k=1}^{K} \left(\frac{(1-K(\mathbf{x}_i,\mathbf{v}_k))^{-1/(u-1)}}{\sum_{k=1}^{K}(1-K(\mathbf{x}_i,\mathbf{v}_k))^{-1/(u-1)}} \right)^u \times (1 - K(\mathbf{x}_i, \mathbf{v}_k)). \tag{11}$$

After training our data, we get the final membership matrix and a minimized value of the clustering objective function. The Kernel based Fuzzy c-Means clustering algorithm is explained in Algorithm 1.

Once the set of final cluster centers is determined then we compute the $\mathbf{P}$ relation matrix to reveal the statistical relationship between the formed clusters and the given classes.

Algorithm 1 Kernel-Based Fuzzy Clustering Algorithm

1: **Randomly** initialize cluster center $V = [v_1, v_2......v_K]$ for data samples.
2: **Evaluate** the membership matrix by using Eq. (8).
3: **Evaluate** the set of final cluster centers $V = [v_1, v_2......v_K]$ by using Eq. (9).
4: If $(V^{t+1}$ - $V^t) < \epsilon$ then stop, else continue with step 2
5: **Return** M (set of membership matrix), V (set of final cluster centers).

2) $\mathbf{P}$ *Relation Matrix:* The relation matrix $\mathbf{P}$ is computed using the Bayesian theory. First, the cluster posterior probabilities of class membership $p(\omega_l \mid c_k)$ is computed as follows:

$$p(\omega_l \mid c_k) = \frac{p(\omega_l, c_k)}{p(c_k)} \tag{12}$$

In Eq. (12), $p(c_k)$ represents the prior probability, which means the proportion of samples belonging to the kth cluster, i.e., $\mathrm{Num}(x \in c_k)/S$, S denotes the total number of training samples. Likewise, $p(\omega_l, c_k)$ represents the joint distribution which is calculated based on the proportion of samples that lies in the kth cluster and the lth class, i.e., $\mathrm{Num}(\mathrm{x} \in \omega_l$ and $\mathrm{x} \in c_k)/S$. As a result, $p(\omega_l \mid c_k)$ can be expressed as:

$$p(\omega_l \mid c_k) = \frac{\mathrm{Num}(\mathbf{x} \in \omega_l \text{ and } \mathbf{x} \in c_k)}{\mathrm{Num}(\mathbf{x} \in c_k)} \tag{13}$$

The constraint $\sum_{l=1}^{L} p(\omega_l \mid c_k) = 1$ must be satisfied for each cluster c_k, where L denotes the number of classes. By combining all the $p(\omega_l \mid c_k)$ values, we can create a $K \times L$ matrix, denoted by $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} p(\omega_1 \mid c_1) & p(\omega_2 \mid c_1) & \dots & p(\omega_L \mid c_1) \\ p(\omega_1 \mid c_2) & p(\omega_2 \mid c_2) & \dots & p(\omega_L \mid c_2) \\ \cdots & \cdots & \cdots & \cdots \\ p(\omega_1 \mid c_K) & p(\omega_2 \mid c_K) & \dots & p(\omega_L \mid c_K) \end{bmatrix}$$

The relation matrix $\mathbf{P}$ relies on the partition obtained through clustering and its value is decided by allocating each data point to the closest clustering centers. Next, by using the Bayesian theory, we design a classification mechanism which also relies on the set of final cluster centers obtained through the clustering mechanism.

B. Testing of the Classifier

In the classification phase, we compute the posterior probability $p(\omega_l \mid x_i)$. To introduce the cluster information in the computation of $p(\omega_l \mid x_i)$, we make use of the formed clusters $\{c_k\}$ to reformulate $p(\omega_l \mid x_i)$ through the total probability theorem as

$$p(\omega_l \mid x_i) = \sum_{k=1}^{K} p(\omega_l, c_k \mid x_i) = \sum_{k=1}^{K} p(c_k \mid x_i) p(\omega_l \mid c_k) \tag{14}$$

Where $p(c_k \mid x_i)$ represents the posterior probabilities of the presence of corresponding samples and through its intuitive meaning it can computed using Eq. (8), and $p(\omega_l \mid c_k)$ denotes the cluster posterior probabilities of class membership where ω_l denotes the l^{th} class, c_k represents the k^{th} cluster.

In the classification learning, once the posterior probabilities $p(\omega_l \mid x_i)$ is modeled, then the output class label $f(x_i)$ is computed as follows:

$$f(x_i) = arg \max_{1 \leq l \leq L} p(\omega_l \mid x_i) \tag{15}$$

Once the output class label $f(x_i)$ is determined, then the classification objective is determined as follows:

$$J_u(\{\mathbf{v}_k\}) = \sum_{i=1}^{S} \delta(f(\mathbf{x}_i), y_i) / S \tag{16}$$

where y_i is the class label (ground truth) of $\mathbf{x}_i$ and $y_i \in \{1, 2, ..., L\}$. This objective function is based on the minimization of the misclassification rate.

Algorithm 2 Misclassification Rate

Input: $train_y = \{y_1, y_2, ..., y_S\}$ is the ground truth or class labels corresponding to the data samples present in $train_X = \{x_1, x_2, ..., x_s\}$; M is the membership matrix; $\mathbf{P}$ is the relation matrix.

Output: $J_u(\{\mathbf{v}_k\})$

1: **Compute** the posterior probability $p(\omega_l \mid x_i)$ of data samples present in $train_X$ using Eq. (14).
2: **Compute** the output class labels of each data samples using Eq. (15).
3: **Compare** the output class label $f(x_i)$ with the ground truth y_i and compute the misclassification rate using Eq. (16).
4: **Return** $J_u(\{\mathbf{v}_k\})$

In this work, we have formulated the clustering and classification problem by integrating the two objective functions given in Eq. (11) and Eq. (16).

$$\min J(\{\mathbf{v}_k\}) = [J_1(\{\mathbf{v}_k\}), J_2(\{\mathbf{v}_k\})] \tag{17}$$

The value of $\min J(\{\mathbf{v}_k\})$ only depends on a set of cluster centers. Therefore, by just optimizing the cluster centers embedded in $J(\{\mathbf{v}_k\})$, the clustering $J_1(\{\mathbf{v}_k\})$ and classification $J_2(\{\mathbf{v}_k\})$ criteria can be simultaneously optimized at the same time.

C. Optimization of Multiobjective Functions

In this work, we are using Multi-Objective Particle Swarm Optimization (MOPSO) to simultaneously optimize the clustering and classification objective function. MOPSO is an evolutionary technique that combines individual improvement with population cooperation and competition. It has demonstrated excellent performance and fast convergence [11]. Lin et al. [14] proposed an evolutionary algorithm for Many-Objective optimization problems in which they have used more than three objective functions. In our proposed approach, we have used a simplified version of MOPSO to solve the multi-objective optimization problem. MOPSO used in the proposed RKFCM-MSCC algorithm utilizes an external repository to store non-dominated solutions found during the search process. Furthermore, adopting MOPSO in proposed RKFCM-MSCC allow us to obtain the multiple sets of Pareto-optimal clustering centers in the two objective spaces. Since the clustering and classification learning methods complement each other thus, the corresponding the clustering and classification objective function given in Eq. (11) and Eq. (16) can be complementary to certain extent. Consequently, the Pareto-optimal clustering centers that achieve relatively low values on the training data jointly for both clustering compactness and classification error rate can consistently yield the best result for clustering compactness or classification error rate on the test data.

In the proposed RKFCM-MSCC, a particle is represented as $V_i = [v_{i1}, v_{i2}, ..., v_{id}, .., v_{iD}]$ where $D = d \times K$. Each particle flies with a velocity $vel_i = [vel_{i1}, vel_{i2}, .., vel_{id}, ..., vel_{iD}]$. This updated velocity of the particle is calculated based on the experience of particle itself and repository which is defined as follows:

$$\begin{aligned} \text{vel}_{id}^{t+1} = I \times \text{vel}_{id}^{t} + r_1 \times \left(\text{pbset}_{id}^{t} - v_{id}^{t}\right) \\ + r_2 \times \left(\text{Repository}_d(h) - v_{id}^{t}\right) \end{aligned} \tag{18}$$

Where I is inertia weight and set to a value 0.4, r_1 and r_2 are two random numbers which takes the value uniformly distributed in the range $[0, 1]$. pbset_{id}^{t} represents the best position of the i^{th} particle in d^{th} dimension in t^{th} iteration. $\text{Repository}(h) = [\text{Repository}_{h1}, \text{Repository}_{h2}, ..., \text{Repository}_{hD}]$ is the repository consist of nondominated solution. At each iteration, the position of each particle is updated with the given equation as follow:

$$v_{id}^{t+1} = v_{id}^{t} + \text{vel}_{id}^{t+1} \tag{19}$$

The complete process of proposed RKFCM-MSCC Algorithm is presented in Algorithm 3.

Algorithm 3 RKFCM-MSCC Algorithm

Input: The number of particles P is set to 500, the total number of iterations t is set to 100, and the current iteration is set to 1.

1: Initialize all the particles with random positions and velocities.
2: Compute the clustering and classification objective function using Eq. (11) and Eq (16) and **pbest** $_i$ of each particle is set equal to its current position.
3: In the repository, the position of the particle are stored that represent nondominated solutions.
4: While $t > 1$:
   (a) Evaluate the velocity of each particle by using Eq. (18) .
   (b) Evaluate the new position V_i of each particle using Eq. (19).

(c) Compute the value of two objective function using Eq. (11) and Eq (16).

(d) At each iteration determine all the currently nondominated solutions :

```
For i = 1: P
 Non_dominated_flag = 1
 For j = 1: P
      If v_i is dominated by v_j
         Non_dominated_flag = 0
     End
 End
 If Non_dominated_flag = 1
     v_i is the currently
     nondominated location.
End
End
```

(e) Store the nondominated locations into the Repository and remove any dominated locations from the Repository

(f) If the particle current position V_i^t dominates $\mathbf{pbest}_i$

$$\mathbf{pbest}_i = V_i^t$$

else if the $\mathbf{pbest}_i$ dominates V_i^t

$\mathbf{pbest}_i$ is kept

else if no one dominated by other

$\mathbf{pbest}_i$ is updated or kept randomly

End

(g) Update $t = t + 1$

III. Experimental Results

A. Experimental Setup

In this study, all the experimentation is performed on an NVIDIA DGX-1 supercomputer. The supercomputer has the following configuration: Dual 20 Core Intel Xeon E5-2698 V4 clocked at 2.2 GHz, 5120 NVIDIA cores, 512 GB 2.133 GHz DDR4 RDIMM (RAM). All the codes are written in Python version 3.9.13.

B. Dataset Description

In this study, we performed our experimental study on four datasets which were taken from the UCI machine learning repository [15]. The detailed description of the datasets used in this study is presented in Table I.

C. Parameter Specification

In this section, we have presented the specification of the parameters used in the experimental study. Table II shows the values of parameters K, λ, and ϵ used for various datasets. The value of the parameters listed in this table is chosen after exhaustive experimental evaluation.

TABLE I
Dataset Descriptions

Datasets	Characteristics		
	Instances	No of Classes	No of attributes
Iris	150	3	4
Soyabean	47	4	35
Wine	178	3	13
Thyroid	215	5	3

TABLE II
Parameter Specifications for Datasets

Dataset	Parameters		
	K	λ	ϵ
Iris	12	.001	.001
Soybean	4	1	.001
Wine	6	.1	.001
Thyroid	10	.001	.001

D. Results and Discussion

In this section, all the datasets used are randomly partitioned into two halves: for training the first half of each dataset is used and the other half is used for testing. This process is repeated and independently run ten times. Then averaged accuracies and standard deviations is computed and reported in Table III. Furthermore, the experimental results justifying the effectiveness of the proposed RKFCM-MSCC approach in terms of accuracy when compared with other existing models like RFRC [1], VQ+LVQ3 [4], RBFNN [5], SCC [16], and MSCC [10], respectively. As we can see from the reported experimental results that the proposed RKFCM-MSCC model shows significant improvement in terms of accuracy on all four datasets in comparison to other models like RFRC, VQ+LVQ3, RBFNN, SCC, and MSCC, respectively. The exceptional classification accuracy achieved by the proposed RKFCM-MSCC can be attributed because of its efficient learning mechanism. For Iris Dataset, there is a 1.24 percentage increment in their average accuracy achieved by the proposed RKFCM-MSCC model in comparison to MSCC and achieved much higher accuracy compared to other models. For Wine Dataset, the proposed RKFCM-MSCC model shows an increment of 0.9 percent in terms of average accuracy compared to MSCC and significantly higher compared to other models. In the case of the thyroid dataset, the proposed RKFCM-MSCC model attain an increment of 0.6 percent in terms of average accuracy compared to MSCC and SCC models and significantly higher compared to other models. For the soybean dataset, the proposed RKFCM-MSCC model attains the 100 average accuracy which is equivalent to the MSCC approach but much higher than other compared models. In addition to the average accuracy, we have reported the Pareto optimal solution for both training and test partition of Iris dataset in Table IV. From the reported results, an interesting observation is found that on the test data the best performance of clustering or classification is corresponding to those solutions which attain relatively low values of objective functions for both clustering compactness and classification error rate on the training dataset.

TABLE III
CLASSIFICATION ACCURACY ON REAL-LIFE DATASETS

Datasets (#samples $\times \#dim \times \#class$)	Models					
	RFRC [1]	VQ+LVQ3 [4]	RBFNN [5]	SCC [16]	MSCC [10]	**RKFCM-MSCC**
Iris (150×4 × 3)	95.3 ± 1.1	94.7 ±1.9	96.4 ± 1.6	95.2 ± 1.4	97.1 ± 1.7	98.26 ± 0.64
Wine (178×13 × 3)	96.0 ± 1.7	96.5 ± 1.5	97.3 ± 1.1	97.1 ± 1.8	98.3 ± 1.3	99.2 ± 0.64
Thyroid (215×5 × 3)	91.8 ± 2.0	92.7 ± 2.2	95.3 ± 1.0	96.4 ± 1.5	96.4 ± 1.6	97.00 ± 0.78
Soybean-small (47×35 × 4)	99.1 ± 1.7	96.1 ± 10.4	98.1 ± 1.7	99.6 ± 1.3	100 ± 0.0	100 ± 0.0

TABLE IV
MISCLASSIFICATION RATE AND CLUSTERING COMPACTNESS ON THE TRAINING AND TEST DATA SET OF *IRIS*

Pareto optimal solution	Training Misclassification rate J_1	Training clustering compactness J_2	Test Misclassification rate J_1	Test clustering compactness J_2
S1	0.0061	0.0133	0.0065	0.04
S2	0.0059	0.0133	0.0060	0.04
S3	0.0063	0.0133	0.0056	0.04
S4	0.0063	0.0133	0.0054	0.0266
S5	0.0057	0.0266	0.0063	0.0266
S6	0.0080	0.0133	0.0054	0.0266
S7	0.0061	0.0266	0.0063	0.0266
S8	0.0058	0.0133	0.0064	0.0266
S9	0.0052	0.0133	0.0073	0.0266
S10	0.0060	0.0266	0.0066	0.0266

IV. CONCLUSION

This paper proposed the RKFCM-MSCC approach to simultaneously attain better clustering and classification results for the linear and non-linear separable data points. The use of radial basis kernelized fuzzy clustering integrated with the classification is the main highlight of the paper. It yielded better results as compared to existing models. To combine the benefits of classification and clustering learning, several existing algorithms, including RBFNN, RFRC, and VQ+LVQ3 follow a sequential and separate optimization approach for both the clustering and classification criterion. However, this two-step optimization process significantly restricts the effectiveness of both clustering and classification learning. RKFCM-MSCC has a simultaneous optimization process for both clustering and classification. From the experimental results, it can be observed that 1) RKFCM-MSCC attain the promising clustering and classification results at same time; and 2) the Introduction of RBF kernel in Fuzzy c means yielded better results. Undoubtedly, one of the future directions of research is to explore alternative approaches for developing a semi-supervised version of RKFCM-MSCC. Undoubtedly, one of the future directions of research is to explore alternative approaches for developing a semi-supervised version of RKFCM-MSCC.

REFERENCES

[1] W. Cai, S. Chen, D. Zhang, "Robust fuzzy relational classifier incorporating the soft class labels," Pattern Recognition Letters vol. 28, no. 16, pp. 2250–2263, 2007.

[2] D. Q. Zhang, S. C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation ," Artificial intelligence in medicine, vol. 32, no. 1, pp. 37-50, 2004.

[3] C. W. Bong, M. Rajeswari, "Multi-objective nature-inspired clustering and classification techniques for image segmentation," Applied soft computing, vol. 11, no. 4, pp. 3271-3282, 2011.

[4] S. W. Kim, B. J. Oommen, "Enhancing prototype reduction schemes with lvq3-type algorithms," Pattern Recognition, vol. 36, no. 5, pp. 1083–1093, 2003.

[5] I. Maglogiannis, H. Sarimveis, C. T. Kiranoudis, A. A. Chatziioannou, N. Oikonomou, V. Aidinis, "Radial basis function neural networks classification for the recognition of idiopathic pulmonary fibrosis in microscopic images," IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 1, pp. 42–54, 2008.

[6] L. A. Bejarano, H. E. Espitia, C. E. Montenegro, "Clustering analysis for the pareto optimal front in multi-objective optimization," Computation, vol. 10, no. 3, pp. 37, 2022.

[7] N. K. Kamila, L. Jena, H. K. Bhuyan, "Pareto-based multi-objective optimization for classification in data mining," Cluster computing, vol. 19, pp. 1723-1745, 2016.

[8] C. E. Pedreira, "Learning vector quantization with training data selection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol .28, no. 1, pp. 157–162, 2005.

[9] R. Kuo, M. R. Setiawan, T. P. Q. Nguyen, "Sequential clustering and classification using deep learning technique and multiobjective sine-cosine algorithm," Computers Industrial Engineering, vol. 173, pp. 108695, 2022.

[10] W. Cai, S. Chen, D. Zhang, "A multiobjective simultaneous learning framework for clustering and classification," IEEE transactions on neural networks, vol. 21, no. 2, pp. 185– 200, 2009.

[11] Y. Chen, X. Zhao, and J. Hao, "A novel MOPSO-SODE algorithm for solving three-objective SR-ES-TR portfolio optimization problem," Expert Systems with Applications, p.120742, 2023.

[12] T. Li, L. Zhang, W. Lu, H. Hou, X. Liu, W. Pedrycz, and C. Zhong, "Interval kernel fuzzy c-means clustering of incomplete data," Neurocomputing, vol. 237, pp. 316-331, 2017.

[13] S. Abe, "Training of support vector machines with Mahalanobis kernels," In International Conference on Artificial Neural Networks . Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 571-576, 2005.

[14] Q. Lin, S. Liu, K. C. Wong, M. Gong, C. A. Coello Coello, J. Chen, J. Zhang, "A clustering-based evolutionary algorithm for many-objective optimization prob- lems," IEEE Transactions on Evolutionary Computation, vol. 23, no. 3, pp. 391–405, 2019.

[15] D. Newman, S. Hettich, C. Blake, C. Merz, "UCI Repository of machine learning databases," http://www.ics.uci.edu/ mlearn/MLRepository.html, 1998.

[16] W. Cai, S. Chen, D. Zhang, "A simultaneous learning framework for clustering and classification," Pattern Recognition, vol. 42, no. 7, pp. 1248-1259, 2009.