Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 1, 2026, 61-69.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

K-means-CRBM: An Efficient Unsupervised Tool for Feature Learning

Reza Kharghanian* and Zeynab Mohammadpoory

Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran.

Article Info

Article History:
Received 26 June 2025
Revised 23 August 2025
Accepted 30 September 2025

DOI:10.22044/jadm.2025.16416.2767

Keywords:
Convolutional Restricted Boltzmann Machine, Feature Learning, Representation Learning, K-means Clustering.

**Corresponding author: reza.kharghanian@shahroodut.ac.ir (R. Kharaghanian).*

Abstract

The Convolutional Restricted Boltzmann Machine (CRBM) is a generative model that extracts representations from unlabeled data, achieving success in various applications. However, its unsupervised nature may yield suboptimal representations for specific classification tasks. This paper proposes adapting k-means clustering to enhance CRBM parameters, aligning features with informative cluster centers. A novel criterion combining generative and soft-K-Means objectives optimizes both cluster centers and CRBM parameters, allowing for continued unsupervised feature learning.
Experiments on MNIST, CIFAR10, and three facial expression datasets (JAFFE, KANADE, BU) show that the proposed method enhances the learning process and offers a more informative representation compared to standard and classification CRBM.

1. Introduction

The use of good data representations as inputs to machine learning algorithms is a crucial factor that directly affects the performance of subsequent classification or prediction tasks. Feature engineering and feature learning can be classified as two major categories of feature extraction. Feature engineering uses a set of predefined handcrafted feature descriptors, for example, Gabor filters in different scales and orientations [1]. Nonetheless, there is no guarantee that extracted features are optimal for the classification or recognition task at hand. Thus, they can prove to be insufficient to encode information required to achieve good performance. On the other hand, feature learning or representation learning approaches [2] are designed to automatically learn features and select relevant and the most predictive ones by using training data in either a supervised or an unsupervised manner. Autoencoders [3] and Deep Belief Networks (DBNs) [3] are examples of deep learning models for unsupervised feature learning. The DBN was first proposed by Hinton [3] to encode the input data to a lower-dimensional space and be used as an initialization process for deep neural networks. Restricted Boltzmann Machines (RBMs) as building blocks of DBNs are probabilistic models optimized to model the probability distribution of input data. RBMs can also be used as generative models of many different types of labeled or unlabeled data.
In the latent space, the model learned a representation from the input data, which can be regarded as compressed features of the input. Therefore, this feature space can be classified using a classifier such as a Support Vector Machine (SVM). In this research, a new objective function is introduced to adapt k-means clustering to pre-cluster features before feeding them to the classifier.
The remainder of this paper is organized as follows. Section 2 provides related research on CRBM. Section 3.1 describes the standard CRBM and its extension to use supervised information. The proposed K-Means-CRBM is presented in Section 3.2. In Section 4, an experimental evaluation of the proposed method and a comparison with previous methods are provided. Discussion constitutes Section 5. Finally, the paper is concluded in Section 6.

2. Related works

The ability to represent dependency structure between random variables makes RBMs a powerful tool in many applications. DBNs are

structurally composed of stacked pre-trained RBMs [4]. RBMs have been successfully used for dimensionality reduction, document classification, object tracking, and speech recognition. Along with these applications, RBMs have also undergone vast theoretical research studies, which have led to the rise of numerous varieties of RBM, such as Temporal and Recurrent RBM, Continuous RBM, Gaussian-Binary/Bernoulli RBM, self-connected RBM, Convolutional (Classification) RBM, 3D Convolutional RBM, and Centered CRBM, etc. Interested readers may refer to [5-7] for a comprehensive review of theoretical research studies and applications.

Over the past few years, (C)RBMs have sustained their attractiveness as versatile models, being investigated in diverse applications, such as Seizure Detection [8, 9], image recognition [10], image feature extraction [11], facial pain detection [12], Recommender Systems [13], Sentimental Analysis [14], and a hybrid combination of autoencoder and RBM for multiclass classification problems [15]. Moreover, theoretical investigations have also attracted significant attention, as demonstrated in studies such as [12-14, 16].

The standard RBM is an unsupervised generative model that can be adapted to enrich features by utilizing labeling information. A discriminative learning objective function aimed at optimizing classification performance was introduced in [17]. Tulder et al. [18] updated the RBM classification model by adding label units. Yin et al. [19] found that ClassRBM has higher reconstruction errors than standard RBM due to label information. To enhance ClassRBM's classification performance, Yin proposed a model focused on minimizing reconstruction error. A two-stage training process was employed: one for ClassRBM and another for the adjustment model. Incorporating label information introduces additional learnable parameters. As reconstruction error in ClassRBM rises, it indicates that label information helps the network focus on relevant features.

This paper focuses on Convolutional RBM [20], an enhanced version of RBM that utilizes weight sharing to efficiently learn from full-size images while reducing model parameters. A novel optimization problem is introduced that combines the Convolutional RBM criterion with k-means clustering to enhance data representation compactness by reducing the distance between features and cluster centers, thereby improving inter-cluster separability. The proposed method enhances performance without using data labels, unlike classification RBM [17] and convolutional classification RBM [18]. It employs sparsity regularization [21] in the overcomplete CRBM model to maintain mean activation of feature maps near a small constant, preventing trivial feature learning. Additionally, a further sparsity constraint is introduced to encourage different feature learning across feature maps. The proposed method outperforms CRBM and convolutional classification RBM in feature extraction, as demonstrated by evaluations on MNIST, CIFAR10, and facial expression datasets (JAFFE, KANADE, and BU). It achieves lower reconstruction error and produces more informative features, making it effective for data compression.

3. Proposed method

This section details the K-Means-CRBM and its formulation, beginning with a brief overview of the standard CRBM (sCRBM) and its supervised extension.

3.1. Convolutional RBM

sCRBM is a convolutional variant of the restricted Boltzmann machine (RBM) that shares weights across image locations, allowing it to scale to full-sized images [20]. It is a generative energy-based model, with parameters trained using contrastive divergence (CD) learning [4, 22], as an effective alternative to the intractable stochastic gradient descent used in traditional RBMs.

Following the definitions and notations in [20], sCRBM consists of two layers: input layer V and hidden layer H. The visible layer consists of an $N_v \times N_v$ array of either binary or real-valued units. Likewise, the hidden layer is composed of K "groups", where each group, also called a feature map, is an $N_H \times N_H$ array of binary units. The hidden layer is connected to the visible layer through a set of shared weights and each feature map is computed by convolving a feature detector of size $N_W \times N_W$ with the previous layer, $N_W = N_v - N_H + 1$. In addition, all units in each group share a bias b_k .and all visible units share a single bias c.

The visible layer comprises L channels, with the number of filters determined by the feature maps and input channels. For $L+1$, filters equal hidden groups, while for RGB input $(L>1)$, filters equal $L \times K$. Here, we assume $L=1$.

The sCRBM's joint probability distribution function is defined by its energy function.

$$P(v,h)=\frac{1}{Z}\exp(-E(v,h)) \tag{1}$$

where Z is the partition function or normalization constant. $E(v,h)$ is the energy of joint configuration (v,h) of the visible and hidden units. The energy function for a sCRBM with real-valued input units is also defined as follows:

$$E(v,h)=\frac{1}{2}\sum_{p,q=1}^{N}v_{v,q}^{2}-\sum_{k=1}^{K}\sum_{i,j=1}^{N_H}h_{i,j}^{k}\left(\tilde{w}^{k}*v\right)_{i,j} -\sum_{k=1}^{K}b_k\sum_{i,j=1}^{N_H}h_{i,j}^{k}-c\sum_{p,q=1}^{N}v_{p,q} \tag{2}$$

where $*$ stands for convolution, $\tilde{w}^{k}$, $h_{i,j}^{k}$, and $v_{p,q}$ denote flipping feature detectors horizontally and vertically, hidden unit on location (i,j) in group k, and visible unit on location (p,q) respectively. The joint and conditional probability distributions are defined by the sCRBM input type and its energy function.

The conditional independence of hidden and visible units allows for straightforward derivation of their conditional distribution, as demonstrated in Equations (3) and (4).

$$p\left(h_{i,j}^{k}=1|v\right)=sigmoid\left(\left(\tilde{w}^{k}*v\right)_{i,j}+b_k\right) \tag{3}$$

$$p\left(v_{p,q}|h\right)=\mathrm{N}\left(c+\sum_{k=1}^{K}\left(w^{k}*h^{k}\right)_{p,q},1\right) \tag{4}$$

where $sigmoid(.)$ is the logistic sigmoid function and $N(.)$ is a Gaussian density.

CRBM is an unsupervised generative model that can be adapted to a supervised model by altering its energy function and integrating label units into the network architecture [18]. Using class information aids the model in learning features pertinent to specific classification tasks. This paper employs the supervised generative CRBM (sgCRBM) model from [18], with β set to 1. The unit activation probabilities for this model are:

$$p\left(h_{i,j}^{k}=1|v,h\right)= sigmoid\left(\left(\tilde{w}^{k}*v\right)_{i,j}+\sum_{m}U_{m,k}y_m+b_k\right) \tag{5}$$

$$p\left(y_m=1|h\right)= sigmoid\left(\sum_{m}U_{m,k}\sum_{k=1}^{N_H}h_{i,j}^{k}+d_m\right) \tag{6}$$

where $U_{m,k}$ .connects the label unit y_m .to the hidden group h^{k} and d_m is the bias for the label units. Labels are represented in one-hot coding, where each class corresponds to a unit that is activated while others remain off. The conditional probability for the visible nodes is the same as that stated for the standard CRBM in Equation (4).

For detailed information on convolutional RBMs and CCRBMs, refer to sources [20] or [18].

3.2. K-means-CRBM

Let us denote a set of cluster centers by $M=\{m_d|1<d<D\}$, where m_d is a representation of cluster d, and D is the number of clusters. Given the hidden layer outputs, feature maps, $p\left(h^{k}|v\right),k=1,2,...,K$ denoted as Z The feature maps Z would be of size $N_H\times N_H\times K$, where N_H is the spatial size of feature maps, and K is the depth of feature maps. For the representation $Z_{i,j}\in R^{1\times K}$ extracted from the spatial location $i,j\in[1,2,...,N_H]$, the Euclidean distance between $Z_{i,j}$, and each cluster center m_d is computed. Soft-assignment k-means clustering is used to determine distances between $Z_{i,j}$ and cluster centers. Each data point is assigned to multiple clusters with varying weights calculated as follows:

$$Q\left(Z_{i,j},m_d\right)=\frac{e^{-\alpha\left\|Z_{i,j}-m_d\right\|_2}}{\sum_{k=1}^{D}e^{-\alpha\left\|Z_{i,j}-m_k\right\|_2}} \tag{7}$$

where $Q\left(Z_{i,j},m_d\right)$ shows how much $Z_{i,j}$ belongs to cluster m_d, and α is a hyperparameter. Also, note that $\sum_{d=1}^{D}Q\left(Z_{i,j},m_d\right)=1$.

K-means-CRBM optimizes network parameters through a cost function comprising two components: a generative cost function and a k-means clustering cost function, which minimizes the weighted distance of image features to cluster centers.

The system learns cluster centers and network parameters simultaneously, thereby enhancing separability between cluster centers and simplifying the classifier's decision surface. This can be achieved by minimizing the following cost function over training data:

$$\min_{\{W,b,c\}}\{-\sum_{l=1}\log\left(\sum_{h}P\left(v^{l},h^{l}\right)\right)+ \lambda_{sparsity}\sum_{k=1}^{K}\left|p-\frac{1}{N_H^2}\sum_{i,j=1}^{N_H}P\left(h_{i,j}^{k}=1|v^{l}\right)\right|^2+ \lambda_{kmeans}\sum_{i,j=1}^{N_H}\sum_{d=1}^{D}\frac{e^{-\alpha\left\|Z_{i,j}-m_d\right\|_2}}{\sum_{k=1}^{D}e^{-\alpha\left\|Z_{i,j}-m_k\right\|_2}}\left\|Z_{i,j}-m_d\right\|_2^2\} \tag{8}$$

Equation (8) includes the generative objective function of sCRBM with sparsity regularization, ensuring mean activation in hidden groups aligns with target sparsity p, with a large enough sparsity weight $\lambda_{sparsity}$ [20]. The final term represents the soft-assignment k-means cost function, with a weighting parameter, λ_{kmeans}, that adjusts the k-means contribution, can be optimized using a validation set during training.

3.2.1. Optimization of K-means-CRBM

The objective function in Equation (8) is optimized using a gradient-based method.

The k-means term gradient in the objective function is calculated in parallel with CD training after hidden-layer states have been inferred. Training a K-means-CRBM involves merging the k-means gradient with the stochastic gradient estimator from the generative component for each sample or mini-batch.

The cluster centers are initialized by first training the generative part of the objective function to establish a meaningful representation for some epochs. Next, feature maps for the training data are computed using the pre-trained network, followed by calculating cluster centers through k-means. After that, network simultaneously learns parameters and cluster centers using the complete objective function in Equation (8).

The partial derivative of the k-means term with respect to the model's parameters can be derived directly from Equation (8). For simplicity, let us define I to be the k-means part of the objective function in Equation (8), i.e.:

$$I = \sum_{i,j=1}^{N_H} \sum_{d=1}^{D} \frac{e^{-\alpha \left\| Z_{i,j} - m_d \right\|_2}}{\sum_{k=1}^{D} e^{-\alpha \left\| Z_{i,j} - m_k \right\|_2}} \left\| Z_{i,j} - m_d \right\|_2^2 \tag{9}$$

Then, the gradients of q are given by the following:

$$\frac{\partial I}{\partial M} = -\alpha \frac{Z - m_d}{\left\| Z - m_d \right\|_2} S_d + \alpha \sum_{k=1}^{D} \frac{Z - m_k}{\left\| Z - m_d \right\|_2} S_d S_k \tag{10}$$

$$\frac{\partial I}{\partial c_d} = -\alpha \frac{Z - m_d}{\left\| Z - m_d \right\|_2} S_d \left((1 - S_d) \left\| Z - m_d \right\|_2^2 + \sum_{k=1, k \neq d}^{D} S_k \left\| Z - m_k \right\|_2^2 \right) + 2 \left(Z - m_d \right) S_d \tag{11}$$

where $S_d = \frac{e^{-\alpha \left\| Z - m_d \right\|_2}}{\sum_{k=1}^{D} e^{-\alpha \left\| Z - m_k \right\|_2}}$.

The gradient update equations for Dis-CRBM are as follows:

$$\Delta W = \Delta W_{gen} + \lambda_{kmeans} \left(\frac{\partial I}{\partial W} \right) \tag{12}$$

$$\Delta b = \Delta b_{gen} + \lambda_{kmeans} \left(\frac{\partial I}{\partial b} \right) \tag{13}$$

$$\Delta c = \Delta c_{gen} \tag{14}$$

$$\Delta M = \left(\frac{\partial I}{\partial M} \right) \tag{15}$$

The first term in Equations (12) to (14) is evaluated in the same way as in sCRBM. The visible bias update equation remains unchanged from the sCRBM due to a zero gradient with respect to the visible bias. Thus, our method is applicable to both Gaussian and Binary CRBMs. The pseudocode is demonstrated in algorithm 1. What stands out in this algorithm is that following the posterior Q^0 is inferred, the exact gradients of the generative and K-Means components can be computed efficiently in parallel.

4. Experimental Results and Analysis

This section details the datasets utilized, describes the CRBM structure as a feature extractor, evaluates the performance of the proposed method alongside sCRBM and sgCRBM, and compares the results with those reported in [23].

4. Datasets

Experiments were conducted to evaluate the proposed method using publicly available datasets: MNIST [24] for digit classification, CIFAR10 [25] for image classification, three facial expression datasets (JAFFE [26], KANADE [27], and BU [28]) for facial expression classification.

The MNIST dataset includes 60K training and 10K test samples of 28×28 handwritten digits. CIFAR10 comprises 60K color images (32×32) across 10 classes, with 50K for training and 10K for testing. Facial expression datasets feature images of six emotions, with JAFFE containing 210 images, KANADE 245, and BU 700. Six basic expressions are happiness, sadness, fear, anger, surprise, disgust, and neutral face. All facial images are gray, centered, and resized to 30×40 pixels.

Based on our experiments, preprocessing input images improves classification performance. For the MNIST dataset, images were normalized to have zero mean and unit standard deviation. For all other datasets, first zero-phase component analysis (ZCA) whitening was applied to whiten the entire dataset as proposed in [29], and then each image was normalized to have zero mean and unit standard deviation. Finally, for the CIFAR10 dataset, color images were used as input, whereas grayscale images were used for the other datasets.

Algorithm 1 A training algorithm for K-MEANS-CRBM with real-valued visible units

Input training data, shared weighs (W), hidden bias (b), visible bias (c), and cluster centers (m)

Initialize = true

repeat {over the training data}

Set $V^0 \leftarrow V$ (e.g.,set the current image as a mini-batch)

Compute the posterior $Q^0 \leftarrow P\left(H \middle| V^0\right)$ (Eq. (3))

// gradient updates for the generative part and for the K-Means part can be computed in parallel //

Sample H^0 from Q^0

for n = 1 to N_{CD} do

sample H^n from $P\left(V \middle| H^{(n-1)}\right)$ (Eq. (4)

Compute the posterior $Q^n \leftarrow P\left(H \middle| V^n\right)$

Sample H^n from Q^n

end for

Compute updates [20]:

ΔW_{gen}

Δb_{gen}

Δc_{gen}

If (initialize)

Initialize cluster centers using kmeans++ after some number of epochs.

Initialize = false.

else

for current training sample (Eq. (7)):

$Z = Q^{0,k}, \left(k = 1, 2, ..., K\right)$.

End if

Compute gradients: $\frac{\partial I}{\partial W}, \frac{\partial I}{\partial b}, \frac{\partial I}{\partial M}$

Update rules for the proposed K-MEANS-CRBM

$$\Delta W = \Delta W_{gen} + \lambda_{kmeans}\left(\frac{\partial I}{\partial W}\right). \qquad \Delta b = \Delta b_{gen} + \lambda_{kmeans}\left(\frac{\partial I}{\partial b}\right)$$

$$\Delta c = \Delta c_{gen}. \qquad \Delta M = \frac{\partial I}{\partial M}.$$

until convergence

Output trained shared weighs (W), hidden bias (b), visible bias (c), and cluster centers (m).

4.2. The CRBM Structure as a Feature Extractor

We used a single-layer CRBM as a feature extractor. The entire zero-padded image is used as input. The number of filters is fixed to 32 while filter sizes are left as hyperparameters to be determined in the model selection procedure. The sparsity regularization in Equation (8) imposes sparsity only in individual feature maps and not across feature map channels. To introduce competition between neighboring feature maps to help filters learn more localized features that explain input data [30], at each spatial location the ρ percent largest activations were kept unchanged while the remaining (1-ρ) percent were zeroed. A pooling layer with Max operation, filter of size 2×2, and stride of 2 have been used following the CRBM to shrink data representation by half. Pooling allows the model to learn features, which are less sensitive to small translations of input while lowering the computational burden in the higher layers. The output of the pooling layer is regarded as the extracted features to create a feature vector.

Model selection involves a grid search across filter size, *l2* regularization, sparsity target, and λ_{kmeans} to optimize parameters for sCRBM, sgCRBM, and K-Means-CRBM (we refer to CRBMs to address all three learning algorithms) based on validation set classification performance. Optimized aforementioned hyperparameters for CIFAR10 are tabulated in Table .

Table 1. Optimized hyperparameters using grid search on fixed number of filters.

Hyper parameter	Optimized value		
	sCRBM	sgCRBM	K-Means-CRBM
Num filters	32	32	32
Filter size	3	3	5
l2 regularization	0.01	0.0001	0.0001
Sparsity target	0.003	0.003	0.005
λ_{kmeans}	-	-	0.01

The same optimization was also employed on other databases.

Initial CRBMs filter weights are drawn from a zero-mean Gaussian distribution with a 0.01 standard deviation. The connection weights U of the supervised generative CRBM were initially sampled from a uniform distribution $\left[-10^{-6}, 10^{-6}\right]$.

In K-Means-CRBM, cluster centers were initialized with kmeans++ algorithms, and all biases started at zero. Other training settings adhered to the recipes in [31].

4.3. Softmax Regression Model

A Softmax regression model with a modified cost function is employed for object and facial expression recognition. After model selection and validation, CRBM models are trained on the complete training and validation sets to extract features and create feature vectors. Subsequently, a Softmax regression model classifies the images based on these vectors. The flowchart in Figure 1 illustrates the training and testing phases of the feature extraction method and the Softmax classification layer.

4.4. Classification Performance Evaluation

Experiments were conducted on each dataset using three learning algorithms: sCRBM, sgCRBM, and k-means-CRBM. Model selection was performed for each algorithm. Non-Gaussianity measures of weight distribution alongside Kurtosis and classification accuracy on CIFAR10 were used to establish the minimum necessary training iterations for CRBMs [32]. Experiments showed that classification accuracy saturates at around 150 iterations. A similar pattern across other datasets was observed. The CRBM models were trained for up to 400 iterations, with the number of cluster centers matching the number of classes in each dataset.

For the MNIST and CIFAR10 datasets, we randomly excluded $10K$ samples ($1K$ per class) from the training set to create validation set for model selection. We reported classification accuracy, area under ROC curve (AUC), and F1-score averaged over 10 trials, as presented in Table 1 and

Table 3. We employed a Support Vector Machine (SVM) with a linear kernel for a more comprehensive display of the proposed method performance.

Five-fold cross-validation was applied to facial expression datasets. Each fold involved dividing the dataset into training, validation, and test sets, with 20% of images randomly selected for testing. The remaining samples were split into 30% validation and 70% training. Model selection was conducted for each fold, and total accuracy was calculated by averaging the classification accuracy across all folds, as shown in Table 5.

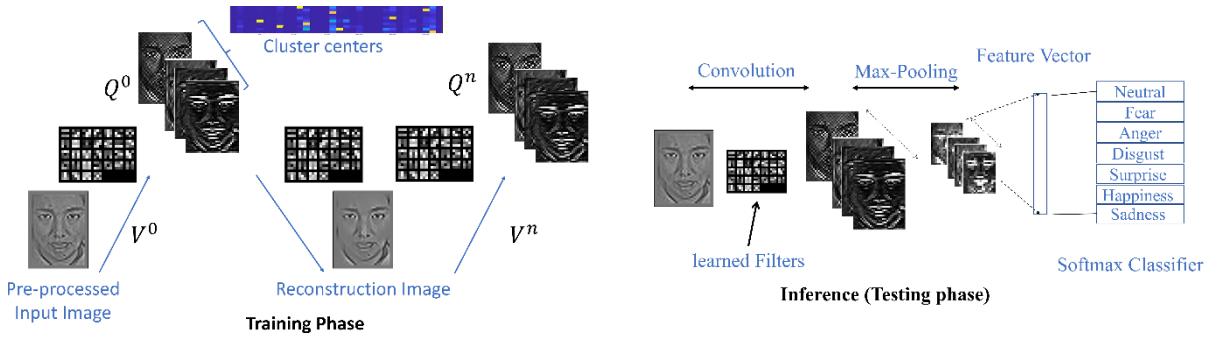


Figure 1. Architecture of the proposed method.

Table 1. Performance of three different CRBM types on MNIST.

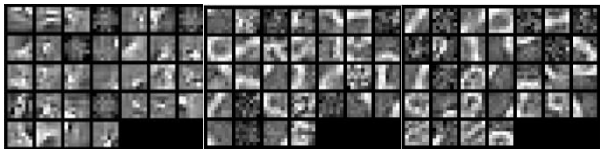
Learning method	Accuracy ± STD%		AUC ± STD%		F1-Score ± STD%	
	SoftMax	SVM	Softmax	SVM	Softmax	SVM
sCRBM	97.62±0.12	97.74±0.11	99.963±0.003	98.97±0.08	97.53±0.07	97.36±0.11
sgCRBM	98.48±0.08	97.96±0.093	99.997±0.0007	99.30±0.05	98.48±0.06	97.95±0.09
K-Means-CRBM	**98.67±0.06**	**98.30±0.08**	**99.998±0.0006**	**99.79±0.02**	**98.66±0.06**	**98.29±0.08**

In all experiments, the proposed K-Means-CRBM outperformed both the standard CRBM and the supervised generative CRBM, demonstrating greater efficiency and robustness. It also reduced the standard deviation of metrics, indicating consistent performance across various weight initializations.

Figure 2 visualizes 32 filters trained on the MNIST dataset. Graphical displays of weights in RBM models help monitor learning, particularly with spatially structured data like images [31]. The K-Means-CRBM model captures more localized features compared to sgCRBM and sCRBM, which exhibit some poorly trained filters.

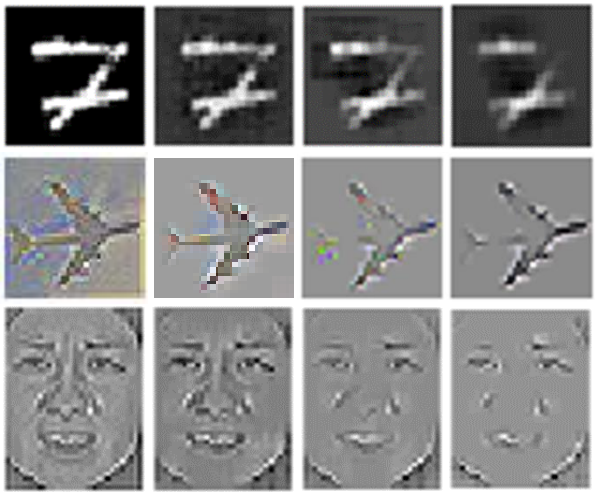
Figure 3 illustrates dataset samples and their reconstructions via CRBMs algorithms. The proposed method enhances detail in reconstructed samples compared to sCRBM and sgCRBM, effectively reconstructing facial features such as

the left eye, nose, and lips, which are crucial for expressing facial emotions.



(a) (b) (c)

Figure 2. Filters learned by (a) K-Means--CRBM, (b) sgCRBM, and (c) sCRBM algorithms on the MNIST dataset.



(a) (b) (c) (d)

Figure 3. (a) original image (b) reconstruction given by K-Means-CRBM (c) reconstruction given by sgCRBM and (d) reconstruction given by sCRBM.

Experiments demonstrated that increasing the number of filters enhances accuracy across all models, as depicted in Figure 4, with the sCRBM model showing poor performance at lower filter counts. As filters increase, sCRBM's efficiency nears that of sgCRBM, with no further gains beyond a certain point, which matches the results reported in [18]. Notably, K-Means-CRBM outperforms others even with fewer filters.

An experiment assessed K-Means classification accuracy on CIFAR10 by varying cluster center numbers, the accuracy is depicted in Table 2. K-Means-CRBM, functioning as the sCRBM without clusters, enhances accuracy by nearly 1% with four cluster centers and by 3.8% when cluster centers match the number of classes.

The proposed K-Means-CRBM method outperformed sCRBM and sgCRBM, as evidenced by Softmax classifier accuracy comparisons. We also evaluated its performance against vector-based neural networks exploiting discriminative information to optimize parameters, as proposed in [23].

Table 2. Accuracy of K-Means-CRBM with different numbers of cluster centers on CIFAR10.

Cluster numbers	0	4	8	10	16	24	32	48	64
Softmax Accuracy	55.93	56.68	60.71	61.22	59.81	60.10	60.20	59.32	58.62

In Table 6, we report the results of the Multilayer Perceptron (MLP) network, which is structurally much closer to the proposed model. In the case of MLP with no hidden layer, the input is the encoded representation extracted by the discriminant denoising Autoencoder. In the case of MLP with one hidden layer, the input is the original image, and the first layer weights are initialized by the learned weights of the discriminant denoising Autoencoder, followed by training the entire MLP network.

As can be seen in Table 6, the K-Means-CRBM method outperforms MLP and MLPh classifiers across all five datasets, as it effectively utilizes local neighborhood information through convolution-based data representation learning.

Table 3. Performance of three different CRBM types on CIFAR10.

Learning method	Accuracy ± STD %		AUC ± STD%		F1-Score ± STD%	
	Softmax	SVM	Softmax	SVM	Softmax	SVM
sCRBM	55.93±0.65	54.71±0.64	90.04±0.28	88.69±0.28	55.68±0.67	54.96±0.65
sgCRBM	57.62±0.46	56.84±0.39	90.93±0.13	89.51±0.15	57.46±0.44	57.05±0.38
K-Means-CRBM	**61.12±0.38**	**61.42±0.27**	**92.34±0.1**	**91.28±0.07**	**61.08±0.38**	**61.66±0.27**

Table 4.Accuracy and the STD of three different CRBM types on facial expression recognition datasets.

Learning method	BU	JAFFE	KANADE
sCRBM	60.95±0.19	58.09±3.1	66.53±2.7
sgCRBM	74.00±0.11	60.95±2.9	78.77±2.2
k-means-CRBM	**74.97±0.10**	**62.74±1.4**	**79.39±1.7**

5.Discussion

Labeling information enhances feature representation for tasks like classification. Label units introduce additional learnable parameters, complicating optimization. The K-Means clustering cost function guides training to make a compact feature space in an unsupervised representation.

The K-Means-CRBM shares the same energy function and learnable parameters as the standard CRBM, which allows for parallel optimization of the K-Means objective after hidden unit activation inference, with simultaneous CRBM parameter updates.

As a result, Time complexity can match that of standard CRBM when employing parallel processing. The proposed optimization method

converges more quickly and yields a detailed feature representation after the same number of epochs. Each of the network parameters plays a significant role in performance especially k_{kmeans}, since determines the amount of participation in updating network weights. Adjusted K-Means parameters enhance non-linear projections, improving classification.

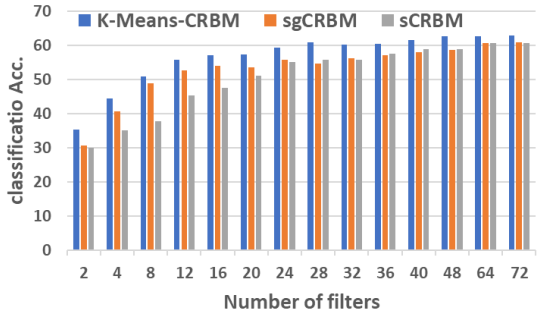


Figure 4.Evaluation of three CRBM models for different number of filters on CIFAR10.

Table 5. Comparison between the proposed method and the MLP classification results reported in [23].

dataset	ACC		
	MLP	MLPh	K-Means-CRBM
MNIST	92.80	96.98	98.66
BU	60.42	72.28	78.00
JAFFE	46.19	60.95	61.90
KANADE	63.67	77.14	81.22
CIFAR10	44.61	56.77	61.6

In addition, hidden units serve as a compressed representation of the input image. The reconstruction process in K-Means-CRBM mirrors that of sCRBM due to the identical energy function. Results indicate that K-Means-CRBM's compressed representation is more informative than sCRBM and even surpasses the sgCRBM model in both quantity and quality. Fine-tuning algorithms can also be applied to the proposed model to improve the quality of the learned features for later use in a classifier. As for the standard CRBM, K-Means-CRBM layers can also be stacked to form a Convolutional Deep Belief Network (CDBN).

The K-Means-CRBM layers maintain the same structure and energy function as standard CRBM layers, resulting in identical inference processes. Therefore, all the new parameters were defined due to the K-Means cost function used only during training and not testing. As a feature learning model, CRBM effectively initializes Convolutional Neural Networks (CNNs), enhancing the learning process and improving performance.

6.Conclusion

This paper introduces a novel CRBM model that enhances feature representation by optimizing feature locations for improved classification. An adapted K-Means algorithm is employed to create a more organized and distinct representation of data in feature space. The proposed K-Means-CRBM effectively improves the feature learning for classification, outperforming sCRBM and sgCRBM. It is applicable in object classification, recognition tasks, and CNN initialization.

Also, K-Means-CRBM is suitable for various unsupervised machine-learning problems as it does not require data labels. Experimental results indicate that K-Means-CRBM provides superior representations compared to sCRBM and sgCRBM across challenging datasets. Tests on MNIST, CIFAR10, JAFFE, KANADE, and BU confirmed that K-Means-CRBM outperforms these alternatives.

The proposed method yielded an improvement of nearly 1% compared to sCRBM while showing only a little improvement compared to sgCRBM on the MNIST dataset. It also achieved over 3.5% and 5.1% enhancements in contrast to sgCRBM and sCRBM, respectively, on the CIFAR10 dataset. Furthermore, in the case of the facial expression datasets, the proposed method improved the accuracy by 14%, 4.7%, and 12.8% compared to sCRBM for BU, JAFFE, and KANADE, respectively. It also enhanced the accuracy by nearly 1%, 2%, and 1% for BU, JAFFE, and KANADE in association with sgCRBM, respectively. In future works, we plan to explore the generalization of our model on face analysis tasks, 3D input, and non-image inputs.

Supplementary information Code will be released at: https://github.com/R-KH02/K-Means-CRBM.

References

[1] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 18, no. 10, pp. 959-971, 1996.

[2] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, no. 8, pp. 1798-1828, 2013.

[3] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[4] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation,* vol. 18, no. 7, pp. 1527-1554, 2006.

[5] M. Joudaki, 2025, doi: 10.20944/preprints202502.1119.v2.

[6] N. Zhang, S. Ding, J. Zhang, and Y. Xue, "An overview on restricted Boltzmann machines," *Neurocomputing,* vol. 275, pp. 1186-1199, 2018.

[7] A. Fischer and C. Igel, "Training restricted Boltzmann machines: An introduction," *Pattern Recognition,* vol. 47, no. 1, pp. 25-39, 2014.

[8] V. V. Narayana and P. Kodali, "Advanced Seizure Detection Framework Using Stacked Convolutional Restricted Boltzmann Machine (SCRBM)," *IEEE Sensors Letters,* vol. 9, no. 3, pp. 1–4, 2025, doi: 10.1109/lsens.2025.3543141.

[9] K. Visalini, S. Alagarsamy, and D. Nagarajan, "Event-Based Epileptic Seizure Detection with Stacked Convolutional Restricted Boltzmann Machine," *IETE Journal of Research,* vol. 70, no. 5, pp. 4880–4889, 2023, doi: 10.1080/03772063.2023.2234854.

[10] S. Sobczak and R. Kapela, "Hybrid Restricted Boltzmann Machine– Convolutional Neural Network Model for Image Recognition," *IEEE Access,* vol. 10, pp. 24985–24994, 2022, doi: 10.1109/access.2022.3155873.

[11] X. Lü, L. Long, R. Deng, and R. Meng, "Image feature extraction based on fuzzy restricted Boltzmann machine," *Measurement,* vol. 204, 2022, doi: 10.1016/j.measurement.2022.112063.

[12] R. Kharghanian, A. Peiravi, F. Moradi, and A. Iosifidis, "Pain detection using batch normalized discriminant restricted Boltzmann machine layers," *Journal of Visual Communication and Image Representation,* vol. 76, 2021, doi: 10.1016/j.jvcir.2021.103062.

[13] R. Kirubahari and S. M. J. Amali, "An improved restricted Boltzmann Machine using Bayesian Optimization for Recommender Systems," *Evolving Systems,* vol. 15, no. 3, pp. 1099–1111, 2023, doi: 10.1007/s12530-023-09520-1.

[14] E. Karthik and T. Sethukarasi, "A Centered Convolutional Restricted Boltzmann Machine Optimized by Hybrid Atom Search Arithmetic Optimization Algorithm for Sentimental Analysis," *Neural Processing Letters,* vol. 54, no. 5, pp. 4123–4151, 2022, doi: 10.1007/s11063-022-10797-7.

[15] M. Aamir, N. M. Nawi, F. Wahid, M. S. H. Zada, M. Z. Rehman, and M. Zulqarnain, "Hybrid Contractive Auto-encoder with Restricted Boltzmann Machine For Multiclass Classification," *Arabian Journal for Science and Engineering,* vol. 46, no. 9, pp. 9237–9251, 2021, doi: 10.1007/s13369-021-05674-9.

[16] J. Garí, E. Romero, and F. Mazzanti, "Learning restricted Boltzmann machines with pattern induced weights," *Neurocomputing,* vol. 610, 2024, doi: 10.1016/j.neucom.2024.128469.

[17] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted boltzmann machine," *Journal of Machine Learning Research,* vol. 13, no. Mar, pp. 643-669, 2012.

[18] G. van Tulder and M. de Bruijne, "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1262-1272, 2016.

[19] J. Yin, J. Lv, Y. Sang, and J. Guo, "Classification model of restricted Boltzmann machine based on reconstruction error," *Neural Computing and Applications,* vol. 29, no. 11, pp. 1171-1186, 2018.

[20] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," 2009.

[21] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM,* vol. 54, no. 10, pp. 95-103, 2011.

[22] M. A. Carreira-Perpinan and G. E. Hinton, "On contrastive divergence learning.," 2005.

[23] P. Nousi and A. Tefas, "Deep learning algorithms for discriminant autoencoding," *Neurocomputing,* vol. 266, pp. 325—335, 2017.

[24] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, and et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278—2324, 1998.

[25] A. Krizhevsky, G. Hinton, and et al., "Learning multiple layers of features from tiny images," 2009.

[26] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," 1998.

[27] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," 2010.

[28] M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 4, pp. 322-336, 2000.

[29] M. A. Ranzato, A. Krizhevsky, and G. Hinton, "Factored 3-way restricted boltzmann machines for modeling natural images," 2010.

[30] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, "Convolutional sparse autoencoders for image classification," *IEEE transactions on neural networks and learning systems,* vol. 29, no. 7, pp. 3289-3294, 2017.

[31] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Springer, 2012, pp. 599-619.

[32] S. Dieleman and B. Schrauwen, "Accelerating sparse restricted Boltzmann machine training using non-Gaussianity measures," 2012.

K-means-CRBM: ابزاری کارآمد و بدون نظارت برای یادگیری ویژگی

رضا خرقانیان و زینب محمد پوری*

دانشکده مهندسی برق، دانشگاه صنعتی شاهرود، شاهرود، ایران.

ارسال ۲۰۲۵/۰۶/۲۶؛ بازنگری ۲۰۲۵/۰۸/۲۳ پذیرش ۲۰۲۵/۰۹/۳۰

چکیده:

ماشین بولتزمن محدود کانولوشنی (CRBM) یک مدل مولد است که با یادگیری ویژگی‌ها از داده‌های بدون برچسب بازنمایی تولید می‌کند و در کاربردهای گوناگون موفقیت‌هایی کسب کرده است. با این حال، ماهیت یادگیری بدون نظارت این مدل‌ها می‌تواند منجر به تولید بازنمایی‌های غیر بهینه به خصوص در مسائل طبقه‌بندی شوند. در این مقاله، برای بهبود یادگیری پارامترهای CRBM، خوشه‌بندی k-means پیشنهاد شده است، به گونه‌ای که بازنمایی در فضای ویژگی به مراکز خوشه‌ها نزدیک شوند. برای این منظور، تابع هزینه جدید با ترکیب تابع هزینه مدل مولد و تابع هزینه soft-K-Means معرفی شده است تا همچنانکه یادگیری بدون نظارت انجام می‌شود، پارامترهای CRBM و مراکز خوشه‌ها بهینه خواهند شد. آزمایش‌ها بر روی مجموعه داده‌های MNIST، CIFAR10، و سه مجموعه داده حالات چهره (شامل JAFFE، KANADE، و BU) نشان می‌دهد که روش پیشنهادی، فرآیند یادگیری را تقویت کرده و در مقایسه با CRBM استاندارد و CRBM طبقه‌بند، بازنمایی آموزنده‌تری ارائه می‌دهد.

کلمات کلیدی: ماشین بولتزمن محدود کانولوشنی، یادگیری ویژگی، یادگیری بازنمایی، خوشه‌بندی k-means.