International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 12 (2018) pp. 10220-10226

Dimensionality Reduction of Image Features using an Autoencoder

Mercy Rajaselvi Beaulah P.* and Manjula D.**

**Department of Computer science & Engineering, Easwari Engineering College, Chennai-600089, Tamilnadu, India*
Email: mercyeec@gmail.com
***Department of Computer science & Engineering, Anna University, Chennai-600025, Tamilnadu, India.*

Abstract

"Curse of Dimensionality" is one of the important problems that image analysis confronts. High-dimensional image analysis is a challenge for researchers during image annotation, classification, and retrieval. Dimensionality reduction provides an effective way to solve this problem, which can improve the learning accuracy, reducing the computation time, and facilitate a better understanding of the learning models. This paper proposes a novel approach for dimensionality reduction using an Autoencoder. An Autoencoder is an unsupervised learning algorithm that applies back propagation, setting the target values to be equal to the inputs. The reduced dimensions computed through the autoencoder are used to train the various classifiers and their performances are evaluated. From the performance of the classifiers, it is proven that the reduced dimensions obtained using the autoencoder have good pattern recognition ability than the dimensions obtained through other approaches.

Keywords: Dimensionality reduction, Image features, Autoencoder, Classifiers.

INTRODUCTION

With the rapid development of modern technology and the internet, many computer applications have generated large amounts of image data at an unprecedented speed. These image data often have the characteristics of high dimensions, which pose a great challenge for image data analysis and decision-making [1-6]. Dimensionality reduction provides an effective way to solve this problem. Feature selection and Feature extraction are two ways of dimensionality reduction. Feature selection has been proven in both theory and practice in effectively processing high-dimensional data and in enhancing the learning efficiency. It removes the redundant and irrelevant features. Unlike feature selection, feature extraction transforms the original data into features with strong pattern recognition ability, where the original data can be regarded as features with weak recognition ability. Feature extraction can improve the learning accuracy, reducing the learning time, and simplify the learning results.

This paper presents a novel approach for dimensionality reduction, an Autoencoder. The autoencoder is categorized as a feature extraction method which transforms the original features into a new reduced feature set. The main contribution of this paper is the design of the autoencoder for reducing the dimensions of the training image data set and comparing the accuracy of the different classifiers which are trained with the reduced feature set of the training image data set.

The rest of the paper is organized as follows: The related works are reviewed in section 2. In section 3, the proposed system is presented. Experimental results are discussed in section 4. Finally, section 5 concludes the paper.

RELATED WORK

Dimensionality reduction methods play a great role in improving the accuracy and reducing the training time of the learning models. This section provides a brief summary of the literature on the dimensionality reduction methods applied on image features.

Min-Ling Z. et. al. [7] designed a system for performing dimensionality reduction for image data sets. Initially, the Principal component analysis is used to remove the irrelevant and redundant features. Then, the genetic algorithm is used to select the subset of the features which improves the classification accuracy. The correlation between the labels is addressed by the genetic algorithm function. The system is able to achieve only 72% average precision.

Bolun C. et.al [8] proposed a feature selection algorithm based on ant colony optimization. The existing ACO-based feature selection algorithm traversed $O(n^2)$ edges for '*n'* feature space. Bolun C. et. al. proposed the ACOFS algorithm, in which artificial ants traverse the graph with $O(2n)$ edges. The classification performance and the feature set size are taken as a heuristic guidance to select the small-sized feature set with the higher classification accuracy. The algorithm is tested for 19 features and at the 50th iteration, 8 features are selected. When the number of classes increases, the precision and the recall start decreasing.

Jiye L. et. al. [9] proposed a feature selection methodology based on the rough set theory and it is stimulated from multi-granulation. Initially, the algorithm selects subtables which are considered as small granularity. The reduct of the original data is estimated on these small granularities. An approximate reduct is obtained by fusing all the estimates on small granularities. In this system, the computation is more and the accuracy is not remarkable.

Amir S. et. al. [10] proposed a system to address the curse of dimensionality. They have combined the filter methods and the wrapper methods for feature subset selection. Through the filter method, the irrelevant and redundant features are removed. Through the wrapper method, an Incremental Wrapper Subset Selection with Replacement algorithm and the Fuzzy Imperialist Competitive algorithm are used to select the optimal subset of the features. The performance of feature selection is evaluated by Support vector machine (SVM) and Artificial neural network (ANN). The SVM performance is remarkable, but the performance of ANN is declined as the training set size is reduced.

Jason K. G. et. al. [11] incorporated dimensionality reduction on Landsat-8 images used in crop differentiation. They extracted the spectral and textual features from the image data and differentiated the crop using the various machine learning algorithms. They evaluated the system with the full feature set and with the reduced feature set. The Feature selection is done using the Classification and Regression Tree (CART) and the Random Forest (RF) methods. The Principal Component Analysis (PCA) and the Tasselled CAP Transformation (TCT) are used for feature extraction. The overall accuracy and the kappa coefficients are used to evaluate the system. The overall accuracy is reduced when the number of features used for classification is reduced. The Support Vector Machine(SVM) classifier shows a better performance for 75 features.

Liang H. et. al. [12] proposed a dynamic relevance and a joint mutual information maximization methodology to address the feature relevancy and the feature interdependency in the process of feature selection. Based on the joint mutual information, the redundant features are selected. A dynamic weight is assigned to the redundant features which help in eliminating the redundant features and in selecting the interdependent features. The accuracy of the system greatly depends on the number of features selected for classification.

Felipe V. et. al. [13] proposed genetic programming to address the issue of unbalanced data in the process of feature selection. The genetic programming works well with both the balanced and the unbalanced data. Different feature selection metrics are applied to select the different sets of features. These features are combined and given as an input population for genetic programming. The genetic programming produces a new population of the features for further image processing. Here computation complexity is more.

Manizheh G. et. al. [14] proposed an evolutionary methodology for dimensionality reduction. More informative features are selected using the Feature Selection using Forest Optimization Algorithm (FSFOA). The stages of FSFOA are initializing the trees, local seeding, population limiting, global seeding, and updating the best trees. The K-nearest neighbour, Support vector machine, and C4.5 classifiers are used as a fitness function. The system achieved good accuracy for a dataset with less number of features.

From the above study, it is inferred that high-dimensional image processing is a challenge for researchers in the image classification process and it involves more computation in training as well as classifying the test instances. Moreover, the performance metrics used to evaluate the classification process record a low value for a high dimensional image dataset in the literature. So the main issue considered in this paper is handling the high dimensional feature set used for classifying the image instance. It is addressed by the autoencoder methodology which transforms the high dimensional image feature set into a reduced feature set with strong pattern recognition ability. This improves the accuracy of the classifier and also reduces the computation involved in training the classifier.

PROPOSED WORK

From the literature, it is observed that high-dimensional image analysis is a challenge for researchers during image annotation, classification, and retrieval. Dimensionality reduction removes the redundant and irrelevant features and enhances the learning efficiency. Feature selection and Feature extraction are two ways of dimensionality reduction. From the literature, it is proven that feature extraction methods for dimensionality reduction improve the accuracy of the image classification and the image retrieval systems. In this paper a feature extraction method, an Autoencoder is proposed for dimensionality reduction which transforms the high dimensional image feature set into a reduced feature set with strong pattern recognition ability. The main contribution of this paper is the design of the autoencoder for reducing the dimensions of the image data set used for training the image classifier. The effectiveness of the reduced dimensions is evaluated by comparing the accuracy of the different classifiers which are trained using this reduced feature set.

Image feature extraction

In order to build a classifier, image features are extracted from the image data and used to train the classifier. In the proposed system, the micro-structure descriptors are used to build the classifier. The content of the images is composed of many universal micro-structures [15]. These micro-structures are extracted and used for the comparison and analysis of different images. The micro-structures are defined based on an edge orientation similarity, and the micro-structure descriptor (MSD) is built based on the underlying colours in the micro-structures with similar edge orientation. The MSD features simulate human visual processing and effectively integrate the colour, texture, shape, and colour layout information as a whole. The MSD has high indexing performance and low dimensionality.

In order to extract the MSD features, the given RGB colour space is transformed into an HSV colour space, where H represents hue, S refers to the purity and its value varies from 0 to 1, and V is the amount of the black component mixed with the hue and its value ranges from 0 to 1. The HSV colour space represents the human colour perception. It is quantized into 72 bins. The H component is quantized into 8 bins, and S *and* V are quantized into 3 bins. Edge orientation plays an important role in the human visual system for recognition and interpretation, and contains rich texture and shape information. In order to detect the edge orientation, the cylindrical HSV colour space is transformed into Cartesian coordinate space. Since the Sobel operator is less sensitive to noise, it is used to detect the edge orientation. Let $m(Hx,Sx,Vx)$ and $n(Hy,Sy,Vy)$ be a gradient along the x and y-

directions. The orientation of each pixel is computed by the Eq. (1). The orientation is uniformly quantized into six bins.

$$\Theta = \arccos\left(\frac{mn}{|m|.|n|}\right). \quad (1)$$

where $mn = Hx.Hy + Sx.Sy + Vx.Vy$.

$$|m| = \sqrt{((Hx)^2 + (Sx)^2 + (Vx)^2)}.$$

$$|n| = \sqrt{((Hy)^2 + (Sy)^2 + (Vy)^2)}.$$

The edge orientation image $\theta(x, y)$ is used to define the micro-structure. The 3 X 3 grid is moved from left to right and top to bottom. The pixel value is kept unchanged if its nearest eight pixels are the same as the center pixel. The pixel value is set to empty if its nearest pixels are not the same as the center pixel. If all the nearest pixels are different, then all the 3 X 3 pixels are set to empty. Using the above-mentioned procedure, the micro-structures *MS1, MS2, MS3*, and *MS4* are computed by moving the grid structure from the locations (0,0), (1,0), (0,1) and (1,1) respectively. The final micro-structure is obtained by applying the Eq. (2).

$$MS(x,y) = Max\{MS1(x,y), MS2(x,y), MS3(x,y), MS4(x,y)\} \quad (2)$$

The micro-structure image is obtained by imposing the micro-structure map which preserves the colour of the image. From the micro-structure image, a micro-structure descriptor is extracted. It is represented by 72 dimensions and it expresses how the spatial correlations of the neighbouring underlying colours are distributed in the micro-structure image.

Dimensionality Reduction

The autoencoder is designed to have an encoder with three layers and a decoder with three layers and one code layer. Since the micro-structure descriptor feature of the training image has a 72-dimension feature vector, the input layer and output layer of the autoencoder are designed to have 72 units. There are two hidden layers in both the encoder and the decoder. The hidden layers 1 & 4 have 54 units and the hidden layers 2 & 3 have 36 units. The code layer has 18 units. The architecture of the autoencoder is shown in Figure 1.

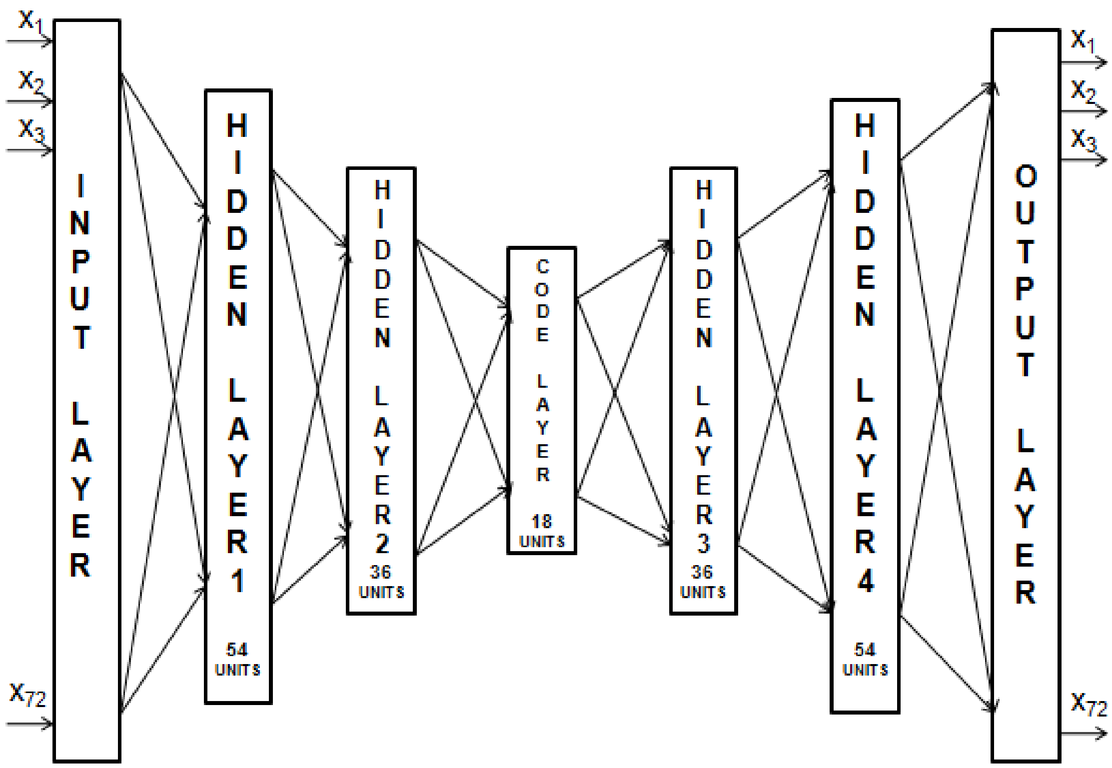

Figure 1. The architecture of an Autoencoder

Let $T = \{T_1, T_2 \ldots\ldots T_n\}$ be the set of training images used for building the classifier, whose dimensions have to be reduced for training the classifier. Let Q be the number of layers. Let $W = (w_1, w_2 \ldots\ldots w_q\}$ be the set of weights of each layer. Let z_q represent the output of the q^{th} layer and let θ_q represent the bias of the q^{th} layer. The input for each layer q is the "output Z_{q-1} of the previous layer along with the weight w_q and bias value θ_q. The logistic function Y is used as the activation function which is nonlinear and differentiable. It maps the large input domain onto the smaller range of 0 to 1. The output of the layer q is computed using the Eq. (3).

$$z_q = Y(z_{q-1} w_q + \theta_q) \quad (3)$$

where $Y = 1/(1 + exp(-T_q))$

[T_q is the input at layer q]

Using the parameters w_q and θ_q of the q^{th} layer, the output z_q of the q^{th} layer is computed. The weights of the network should be learned such that the difference between the input and the output of the autoencoder should be less than the pre-specified threshold *'δ'*. The error at the unit *'j'* in the output layer O is computed using the eq. (4).

$$Err_{oj} = z_{oj} (1- z_{oj})(t_j- z_{oj}) \quad (4)$$

where t_j is the value of the original input features of the training image T_n and z_{oj} is the actual output at the unit 'j' in the output layer O. The error at unit 'i' of the q^{th} hidden layer is computed using the Eq. (5).

$$Err_{qi} = z_{qi} (1- z_{qi}) \sum_j Err_j w_{ij} \quad (5)$$

where w_{ij} is the weight between unit *'i'* of the q^{th} layer and unit *'j'* of the $q+1^{th}$ layer, Err_j is the error of unit *'j'* in the $q+1^{th}$ layer, and z_{qi} is the output at unit *'i'* in the q^{th} layer. The weights and the biases of all the layers are updated to transfer the propagated errors. The weights are updated by using the Eqs. (6) & (7).

$$\Delta w_{ij} = (\gamma) Err_j z_{qi} \quad (6)$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (7)$$

The variable *'γ'* is the learning rate having a value between 0.0 and 1.0. The autoencoder network learns by using the gradient descent method to search for a set of weights to minimize the mean squared distance between the actual output and the original input value. The biases are updated by using the Eqs. (8) & (9).

$$\Delta\theta_j = (\gamma) Err_j \quad (8)$$

$$\theta_j = \theta_j + \Delta\theta_j \quad (9)$$

The weights and the biases are updated after each image's feature is transformed. This process is repeated till the mean squared distance between the original input value and the actual output value is less than the pre-specified threshold value *'δ'*. The time complexity for training the autoencoder network is computed using Eq. (10).

$$\text{Time complexity} = O(ndw) \quad (10)$$

where *'n'* is the number of iterations, *'d'* is the number of images used to train the network, *'w'* is the number of weights to be updated in the network. The output value of the code layer is taken as the transformed 18 dimensions MSD feature vector. This 18 dimension feature vector is used to train the classifier.

Image Classification and accuracy assessment

In order to assess the effectiveness of the reduced dimensions of the image feature vector, the image classifiers are trained using this reduced feature set and the accuracy of the image classifiers is evaluated. The Support Vector Machine (SVM), Artificial Neural Network (ANN), and K-Nearest Neighbour (K-NN) methods are used to train the classifiers.

In the SVM, the classifier is denoted as $f(q, T+, T--)$, where q denotes the input query image, $T+$ denotes the training images from positive samples and $T--$ denotes the training images from negative samples. There are two approaches in SVM, namely one-against-all and one-against-one. In one-against-all, there is an imbalance between the positive samples and the negative samples and the error transmission rate is also high. In one-against-one, since the number of training samples is restricted, the classification accuracy is good and the error transmission rate is also reduced. So, the one-against-one (pairwise) approach is used for training the classifier. The radial basis function is used as a kernel. If there are C image concepts, then $C(C-1)/2$ classifiers should be trained.

In the ANN, the classifier is designed to have one input and one output layer and two hidden layers. The input layer has 18 units and the output layer has one unit. The first hidden layer has 9 units and the second hidden layer has 4 units. Let $I = \{I_1, I_2 \ldots\ldots I_n\}$ be the set of training images used for building the classifier. Let $W = (w_1, w_2 \ldots\ldots w_k\}$ be the set of weights of each layer. Let h_k represent the output of the k^{th} layer and let θ_k represent the bias of the k^{th} layer. The input for each layer k is the output h_{k-1} of the previous layer along with the weight w_k and the bias value θ_k. The logistic function Y is used as the activation function which is nonlinear and differentiable. It maps the large input domain onto the smaller range of 0 to 1. The output of the layer k is computed using the Eq. (11).

$$h_k = Y(h_{k-1} w_k + \theta_k) \quad (11)$$

where $Y = 1/(1 + exp(-I^k))$

[I^k is the input at layer k]

The weights and the bias are updated using the training images. This process is repeated till the mean squared distance between the expected output value and the actual output value is less than the pre-specified threshold value *'ρ'*.

In the K-nearest neighbor, the parameter K is set to three. The Manhattan distance is used to identify the nearest neighbour of the image to be classified. The class labels of the nearest neighbours are used to classify the query image.

The performance of the Autoencoder for dimensionality reduction is compared with the feature selection methods, namely, the Classification and Regression Trees (CART) and Random Forest (RF) and also it is compared with the

Principal Component Analysis (PCA) method for feature extraction [11]. The PCA transforms the data into a new set of principle components (PCs) that describes the underlying structure of the original dataset. CART is a decision-tree machine learning algorithm used for predictive modelling and data pre-processing. It uses binary recursive partitioning to grow decision trees, while the Twoing and Gini methods explore for significant relationships and patterns, allowing better insight into data. It can be used to create a selected list of predictor variables. RF is a collection of decision trees that form an ensemble learning method for feature selection or classification.

The effectiveness of the reduced feature set is evaluated by computing the classification accuracy of the trained classifiers which are trained by these reduced feature set. The SVM, ANN, and K-NN classifiers are trained using 80% of the training image data set and the performance of the classifiers is evaluated using 20% of the training image data set. Good classification accuracy for a classifier indicates that a reduced dimension does not affect the classifier performance. And also the dimension reduction reduces the computation involved in training the classifier.

EXPERIMENTAL RESULTS AND DISCUSSION

Image data often have the characteristics of high dimensions. Dimensionality reduction provides an effective way to solve this problem. This paper presents a novel approach for dimensionality reduction, an Autoencoder which transforms the original features into a new reduced feature set. The experiment was setup to evaluate the effectiveness of the reduced feature set. In order to demonstrate the efficacy of our proposed approach, we have trained the SVM, ANN and the K-NN classifiers using the original and the reduced feature set and evaluated the accuracy of the classifiers. The ten image classes of animals of the MS-COCO 2017 dataset are used to evaluate the performance of the classifier. In order to measure the performance metrics, the confusion matrix between true positive (TP), true negative (TN), false positive (FP) and false negative (FN) is computed for each classifier. The accuracy is computed using the Eq. (12).

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN). \quad (12)$$

The ANN, SVM, and K-NN classifiers are trained using the features obtained from the CART, RF, PCA [11] and Autoencoder method. The accuracy of the classifiers under different conditions is computed. Table 1 shows the accuracy of the classifiers trained with the original features and with the reduced features obtained through the CART, RF, PCA and Autoencoder methods.

Table 1- The Accuracy of the Classifiers under different dimensionality reduction methods

S.No	Dimensionality Reduction method	Number of features	**The accuracy of the Classifiers**		
			SVM	ANN	K-NN
1.	None	72	93.1	84.1	76.2
2.	CART	20	79.6	72.7	62.2
3.	RF	20	86.2	77.8	75.7
4.	PCA	20	93.8	88.2	90.8
5.	Autoencoder	18	94.6	89.8	91.7

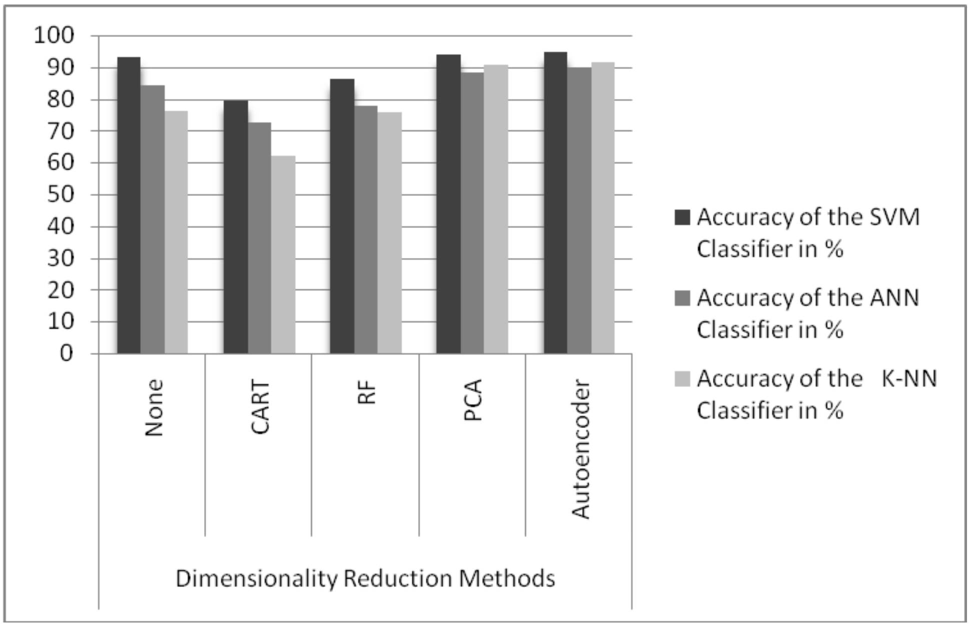

Figure 2. Comparison of the Accuracy of the Classifiers

Figure 2 shows the comparison of the accuracy of the classifiers trained by the reduced feature set obtained through different dimensionality reduction methods. The performance of the SVM classifier is good when compared to that of the other classifiers. The performance of the feature extraction methods, the PCA, and the autoencoder, is better when compared to that of the feature selection methods, the CART and the RF. From the graph, it is shown that the SVM classifier trained by the features obtained from the Autoencoder method produces good accuracy. From the results, it is shown that the accuracy of the classifiers which are trained by the features obtained from the autoencoder method is remarkable when compared to the other methods.

CONCLUSION

The large dimensions of an image data pose a great challenge for image classification, annotation, and retrieval. Dimensionality reduction provides an effective way to solve this problem. Feature selection and Feature extraction are two ways for dimensionality reduction. Feature selection removes the redundant and the irrelevant features. Unlike feature selection, feature extraction transforms the original feature set into a feature set with strong pattern recognition ability. This paper presents a novel approach for dimensionality reduction, an Autoencoder. The Autoencoder is categorized as a feature extraction method which transforms the high dimensional image feature set into a new reduced feature set. The main contribution of this paper is the design of the autoencoder for reducing the dimensions of the image data set and comparing the accuracy of the different classifiers which are trained with this reduced feature set.

In the proposed approach, 72-dimensions micro-structure descriptor features are extracted from the training image dataset. The autoencoder is designed to have three layers in the encoder and three layers in the decoder and one code layer. Since the micro-structure descriptor features of the training images have a 72-dimensions feature vector, the input layer and the output layer of the autoencoder are designed to have 72 units. The hidden layers 1 & 4 have 54 units and the hidden layers 2 & 3 have 36 units. The code layer has 18 units. The weights and the bias are updated after each image feature is transformed. This process is repeated till the mean squared distance between the original input feature and the actual output value is less than the pre-specified threshold value *'δ'*.

The performance of the autoencoder for dimensionality reduction is compared with that of the Classification and Regression Trees (CART) and the Random Forest (RF) methods for feature selection and the Principal Component Analysis (PCA) for feature extraction. The ANN, SVM, and the K-NN classifiers are trained using the original features and the features obtained from the CART, RF, PCA and the autoencoder methods. The accuracy of the classifiers under different conditions is computed. From the results, it is shown that the accuracies of the classifiers which are trained by the features obtained from the autoencoder method are good when compared to the other methods.

In this paper, an autoencoder is designed to transform 72-dimensional micro-structure descriptors into 18-dimensional feature vectors for training the classifier as well as classifying the image instance. The autoencoder can be further redesigned and tested for handling a still higher dimensional feature vector and for other types of image features for image classification, annotation, and retrieval.

REFERENCES

[1] Li, Z., Bo, C., and Jing, Z., 2014, "A comparative study of dimensionality reduction methods for large-scale image retrieval", Neurocomputing, 141, pp. 202-210, http://dx.doi.org/10.1016/j.neucom.2014.03.014.

[2] Ahmed, A. K., Heping, P., Huolin, M., Xiurong, C., Yao, Y. Z., Margaret, O., and Nashir, K., 2015, "Application of dimensionality reduction technique to improve geo-physical log data classification performance in crystalline rocks", Journal of Petroleum Science and Engineering, Elsevier, 133, pp. 633–645, http://dx.doi.org/10.1016/j.petrol.2015.06.035.

[3] Jie, F., Licheng, J., Fang, L., Tao, S., and Xiangrong, Z., 2016, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images", Pattern Recognition, 51, pp. 295-309, http://dx.doi.org/10.1016/j.patcog.2015.08.018.

[4] Jinghua, L., Yaojin, L., Shunxiang, W., and Chenxi, W., 2018, "On line multi-label group feature selection", Knowledge-Based Systems, Elsevier, 143, pp. 42-57, https://doi.org/10.1016/j.knosys.2017.12.008.

[5] Annrose, J., and Seldev, C. C., 2017, "An efficient image retrieval system with structured query based feature selection and filtering initial level relevant images using range query", Optik - International Journal for Light and Electron Optics, 157, 1053-1064, https://doi.org/10.1016/j.ijleo.2017.11.179.

[6] Ayşe, D., 2018, "The effect of feature selection on multivariate pattern analysis of structural brain MR images", Physica Medica, 47, pp. 103–111, https://doi.org/10.1016/j.ejmp.2018.03.002

[7] Min-Ling, Z., Jose M. P., and Victor, R., 2009, "Feature selection for multi-label naïve Bayes classification", Information Sciences, 179(19), pp. 3218-3229..

[8] Bolun, C., Ling, C., and Yixin, C., 2013, "Efficient ant colony optimization for image feature selection", Signal Processing, 93(6), 1566-1576. https://doi.org/10.1016/j.sigpro.2012.10.022.

[9] Jiye, L., Feng, W., Chuangyin, D., and Yuhua, Q., 2012, "An efficient rough feature selection algorithm with a multi-granulation view", International journal of approximate reasoning, 53(6), pp. 912-926.

[10] Amir, S., Mansour, Z., Ali, A., Mahdieh, H. Z., Sadegh, K., and Mostafa, M., 2018 "Using a feature subset selection method and support vector machine to address curse of dimensionality and redundancy in hyperion hyperspectral data classification", The Egyptian Journal

of Remote Sensing and Space Sciences, 21(1), pp. 27-36, https://doi.org/10.1016/j.ejrs.2017.02.003.

[11] Jason, K. G., and Adriaan, V. N., 2017, "Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning", Computers and Electronics in Agriculture, 142, pp. 50–58.

[12] Liang, H., Wanfu, G., Kuo, Z., Ping, Z., and Feng, W., 2018, "Feature selection considering two types of feature relevancy and feature interdependency", Expert Systems With Applications, 93, pp. 423-434. https://doi.org/10.1016/j.eswa.2017.10.016.

[13] Felipe, V., Leonardo, R., Marcos, G., Fernando, M., Giovanni, S., Thiago, S., Guilherme, A., and Isac, S., 2018, "A Genetic Programming approach for feature selection in highly dimensional skewed Data", Neurocomputing, 273, pp. 554-569. https://doi.org/ 10.1016/j.neucom.2017.08.050

[14] Manizheh, G., and Mohammad-Reza, F., 2016, "Feature selection using Forest Optimization Algorithm", Pattern Recognition, 60, pp.121-129, https://doi.org/10.1016/j.patcog.2016.05.012.

[15] Liu, G. H., Li, Z. Y., Zhang L., and Xu Y., 2011 "Image retrieval based on micro-structure descriptor", Pattern Recognition, 44(9), pp. 2123-2133, https://doi.org /10.1016/j.patcog.2011.02.003.