

# Automated morphological classification of galaxies based on projection gradient nonnegative matrix factorization algorithm

I. M. Selim<sup>1,2</sup> · Mohamed Abd El Aziz<sup>3</sup>

Received: 12 May 2016 / Accepted: 16 January 2017 / Published online: 16 February 2017  
© Springer Science+Business Media Dordrecht 2017

**Abstract** The development of automated morphological classification schemes can successfully distinguish between morphological types of galaxies and can be used for studies of the formation and subsequent evolution of galaxies in our universe. In this paper, we present a new automated machine supervised learning astronomical classification scheme based on the Nonnegative Matrix Factorization algorithm. This scheme is making distinctions between all types roughly corresponding to Hubble types such as elliptical, lenticulars, spiral, and irregular galaxies. The proposed algorithm is performed on two examples with different number of image (small dataset contains 110 image and large dataset contains 700 images). The experimental results show that galaxy images from FIGI catalog can be classified automatically with an accuracy of ~93% for small and ~92% for large number. These results are in good agreement when compared with the visual classifications.

**Keywords** Galaxies: classification · Nonnegative matrix factorization (NMF): image processing

## 1 Introduction

Galaxies are gravitationally bound celestial object composed of gas, dust, and billions of stars. Shape and general visual appearance of galaxies gives astronomers much

✉ I. M. Selim  
i\_selim@yahoo.com

Mohamed Abd El Aziz  
abd\_el\_aziz\_m@yahoo.com

<sup>1</sup> Astronomy Department, National Research Institute of Astronomy and Geophysics, Cairo, Egypt

<sup>2</sup> Higher Technological Institute (HTI), Cairo, Egypt

<sup>3</sup> Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt

information about their composition and their evolution. The morphology of galaxies is a significant role in the large scale to understanding observed phenomena and study the universe. This is the first step towards a greater understanding of the origin and formation process of galaxies. Many authors conclude that the galaxies morphological analysis is a problem of interest in astronomy because it is often considered as a convenient way to distinguish between galaxies that have different physical properties [1–4].

Recently, increasing size of telescopes and the CCD camera have allowed extremely large datasets of images to be produced. In modern sky surveys containing millions of galaxies, there are too much data to feasibly analyze manually, one of the most prominent examples of this is the Sloan Digital Sky Survey. This has reinforced the need for robust methods that can perform morphological analysis of large galaxy image databases. Artificial neural networks (ANNs) have recently been utilized in astronomy for a wide range of problems, e.g., from adaptive optics to galaxy classification. Recently, the advancements in computational tools and algorithms have started to allow automatic analysis of galaxy morphology according to the different Hubble types. These algorithms are a widely studied problem in the field of astronomy and include model-driven methods such as [5, 6] and galaxy morphologies [7, 8]. GALFIT [9], GIM2D and refineance their in, CAS [10], Gini [11], Ganalyzer [12], SpArcFiRe [13] and [14] and references therein. Automated classification of galaxies not only using CCD images, but also using spectra, [15, 16]. Not only classifiers can differentiate between broad galaxy morphological types of elliptical and spiral galaxies ([17]; and refinances their in [18]) but also can differentiate between four basic objects [19] and references their in. Clearly, automated classification algorithms will prove invaluable for the analysis of such datasets, but these algorithms are yet to be applied on such scales.

The classification of galaxies based on images and spectra has been a long-term goal in astronomy. Artificial neural networks were first applied to astronomical data sets in order to classify stellar spectra [5]. Classification of galaxies based on spectra in supervised [16] and unsupervised [15]. A common problem facing many of the astronomical sciences today is an increasing amount of data, this will leads to make the classification is a hard problem. Therefore, the dimension reduction methods is needed before apply classification method. Since galaxies images are represented by its light intensity and it is measured by a nonnegative value.

In recently years, the nonnegative matrix factorization (NMF) has become a popular dimension reduction method. The NMF refers to the problem of approximating a nonnegative matrix by a product of two nonnegative matrices.

The main goal of this paper is to developed an algorithm for automated classification of galaxies images based the NMF method. The proposed algorithm is compared with human classifications and other algorithms. The NMF extract a list of features for galaxy images based on some nonnegative constraints are imposed. Therefore, the image can be reconstructed from this list of features.

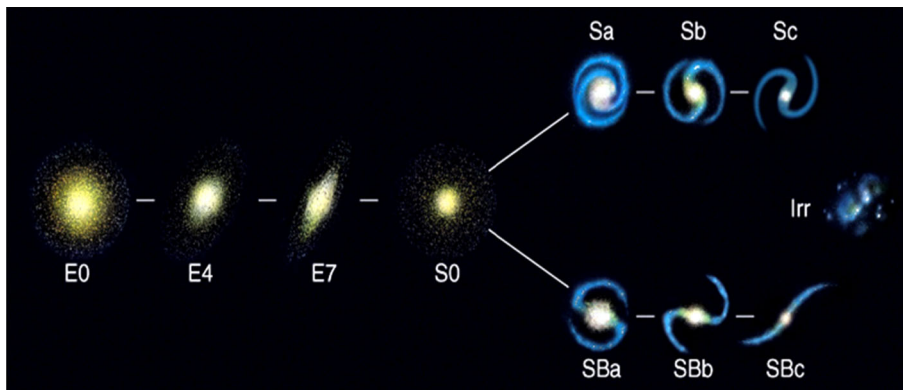
The paper is organized as follows: Galaxy classification was described in Section 2. In section 3, the Nonnegative matrix factorization is introduced. The classification method and algorithm are presented in Section 4. We briefly describe Data and Results in section 5. The conclusion is presented in last section.

## 2 Galaxy classification

Classification of galaxies was established only more than 90 years ago, i.e., during the 1920s. Before that, galaxies were listed in catalogs of nebulae objects that appeared fuzzy in a telescope and were therefore not stars. However, Sir. Edwin Hubble, the American astronomer who showed conclusively for the first time that one of the nebulae (the Andromeda “nebula” M31) is a galaxy in its own right. Hubble was also the first who tried to set out a scheme for classifying the galaxies. His classification scheme which is known as Hubble’s tuning fork diagram as shown in Fig. 1, Hubble [20], including some later additions and modifications, are still in use up to today. Hubble recognized three main types of galaxy: ellipticals, lenticulars and spirals, with a fourth class, that would not fit into any of the other categories called the irregulars. He proposed the following classifications: elliptical: E0, E3, E5, E7; spiral: S0, Sa, Sb, Sc; barred spiral: SBa, SBb, SBc; irregular: Im, IBm. This scheme is commonly referred to as the “Hubble Tuning Fork.” This classification updated by de Vaucouleurs to obtain the Revised Hubble System (RHS) in 1959. Other schemes were proposed by Morgan in 1958 and Van Den Bergh in 1960. A global classification known as the revised morphological types was provided by NASA.

Classifying these galaxies into categories is of great value to astronomers. By studying the structures of galaxies in the same category, we understand the structures processes. In the past, large catalogs of classified galaxies have had many practical applications. Astronomers have used such catalogs to test theories about the universe. The availability of these large datasets has introduced the need for tools that can automatically analyze astronomical images.

There is currently much interest in applying classification algorithms for this problem. The challenge is to design an algorithm which will reproduce classification to the same degree a human expert can do it. Automated classification algorithms will improve the analysis of data sets, but these algorithms are not being applied on all fields of astronomy. The goal of such algorithms is to learn the characteristics of each of the



**Fig. 1** Hubble’s tuning fork diagram of galaxies classification

categories and build a model on this data. This model can then be used to predict the category of futures objects without known classification information. Such an automated procedure usually involves two steps: (i) feature extraction from the digitized image, e.g., the galaxy profile, the extent of spiral arms, the color of the galaxy, or an efficient compression of the image pixels into a smaller number of coefficients (e.g., Fourier or principal component analysis) and (ii) a classification procedure, in which a computer learns from a training set for which a human expert provided his or her classification.

Several machine learning algorithms have been applied to photometric data extracted from galaxy images and applied to direct galaxies images in order to automatic classification [21–23].

There are many algorithms that classify galaxy images based on ANN such as [8] use artificial neural networks techniques for galaxies classification. The Sloan Digital Sky Survey has already used for morphological classification using automated machine-learning techniques [16]. Also the artificial neural networks were applied to astronomical datasets in order to classify stellar spectra [5, 6] and [24, 25]. This has been one of the most successful attempts at automatic classification.

The matrix factorization is an example of a prevalently used scene-based classification method. There are many types of this factorization such as NMF [26] and principle component analysis (PCA), [27] and independent component analysis (ICA) [28].

However, galaxy images are represented by a set of nonnegative numbers, e.g., numbers of occurrences or light intensities. Because of the nonnegativity is a crucial feature that one needs to maintain during the analysis of objects. Hence, the PCA and ICA are not suitable for images since the results of them contain negative data and some features which nonnegative objects by their nature never or hardly possess, such as zero sum and orthogonality. To solve this, we must add the nonnegativity constraint and this constraint is important in human perceptions. So, the NMF method has been proposed to construct these features not only a good reconstruction of images but also the nonnegativity of the features. Therefore, each feature can be considered again as an image. Together with the participation of each feature in an image, one can establish the composition of every image in a very comprehensible way.

### 3 The nonnegative matrix factorization

The NMF method has been proposed as a novel subspace method in order to obtain a parts based representation of objects by imposing nonnegative constraints. Also, it represents data as a linear combination of basis images by using nonnegativity constraints.

The problem addressed by NMF is as follows: Given a nonnegative matrix  $V \in \mathbb{R}^{n \times m}$  (data matrix, consisting of  $m$  vectors of dimension  $n$ ), it is possible to find nonnegative matrix factors  $W \in \mathbb{R}^{n \times r}$  and  $H \in \mathbb{R}^{r \times m}$  ( $r \ll \min(n, m)$ ) in order to approximate the original matrix:

$$V \approx WH \quad (1)$$

A standard NMF aims to minimize the squared Euclidean distance between  $V$  and  $WH$ :

$$\min_{H,W} f(W, H) = \frac{1}{2} \|V - WH\|_F^2 \tag{2}$$

subject to  $\forall i, j, W_{ij}, H_{ij} \geq 0$

where  $\|\cdot\|_F$  is Frobenius norm, the interpretation of  $W$  and  $H$  is different based on the application. For example, in blind source separation (BSS) [29] and [30], when  $V$  represents the mixed signals,  $W$  is a mixing matrix and  $H$  is a matrix of original sources. In classification  $W$  is a basis matrix and  $H$  is coefficient matrix as in [26]. The multiplicative algorithms proposed by Lee and Seung highly popularized to perform decomposition, which alternatively updates  $W$  and  $H$  as follows:

$$W_{k+1} = W_k \frac{V(H_k)^T}{W_k H_k (H_k)^T}, \quad \text{s.t } W_{ij} \geq 0 \tag{3}$$

$$H_{k+1} = H_k \frac{(W_{k+1})^T V}{(W_{k+1})^T W_{k+1} H_k}, \quad \text{s.t } H_{ij} \geq 0 \tag{4}$$

where  $(\cdot)^T$  represents the transpose of matrix, the multiplicative algorithm is commonly used and simple to implement. However, the multiplicative update algorithm lacks the convergence guarantee for large-scale problem [31, 32].

Many algorithms have been proposed to solve these problems, such as, in [31, 32] the projected gradient update for NMF is introduced, in which the Armijo rule (as in Eq. 7) is used to estimate the learning parameters. In the projected gradient NMF, the two matrices  $W$  and  $H$  are updated as:

$$W_{k+1} = \max(0, W_k - \alpha_k \nabla_W f(W_k, H)) \tag{5}$$

$$H_{k+1} = \max(0, H_k - \alpha_k \nabla_H f(W, H_k)) \tag{6}$$

where  $\alpha_k$  is the fixed step size that selected by an Armijo rule along the projection direction (direction of update variable  $H$  or  $W$ ) to determine the step lengths in the iterative updates. For the computation of  $H$  such a value of  $\alpha_k$  is defined as:

$$\alpha_k = \beta^{t_k} \tag{7}$$

where the parameters  $\beta \in (0, 1)$  and  $t_k$  is the first nonnegative integer  $t$  that satisfies

$$f(W, H_{k+1}) - f(W, H_k) \leq \sigma \nabla f(W, H_k)(H_{k+1} - H_k) \tag{8}$$

where  $\sigma \in (0, 1)$  decide about a convergence speed, here we put  $\sigma = 0.01$  and  $\beta = 0.1$  following [31, 32] as the initial value, and these parameters are updated using (6) and (7).

### 4 The proposed algorithm

The overall process of our algorithm based on projected gradient NMF (PG-NMF algorithm) is given in Fig. 2 and we called it Galaxy-PGNMF Classifier. In which the input to this algorithm is training set and test set (where our algorithm not needs to dimension reduction as preprocessing step since  $r < \min(n, m)$ ). Also, the number of galaxy types (groups) is input (i.e., if it is determined from prior knowledge about catalog). Otherwise, (if it is unknown) we determine the best number of galaxy types by using the consensus matrix, where it reflects the probability that images  $i$  and  $j$  belong to the same type. It is defined as the average of connectivity matrix over runs, where the consensus matrix is defined as:

$$\bar{C} = \frac{\sum_{k=1}^{NR} C_k}{NR} \tag{9}$$

where  $NR$  is the number of run and  $C_k$  is the connectivity matrix of size  $m \times m$  and its element is defined as:

$$c_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ belong to the same type} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

By using the off-diagonal of  $\bar{C}$  the best number of type is determined.

Then, our algorithm consists of two main steps: The first is the training step in which each training sample is normalized using  $l_2$ -norm to make all images have the same scale. Then, the PG-NMF algorithm is used to decompose these sets into basis matrix

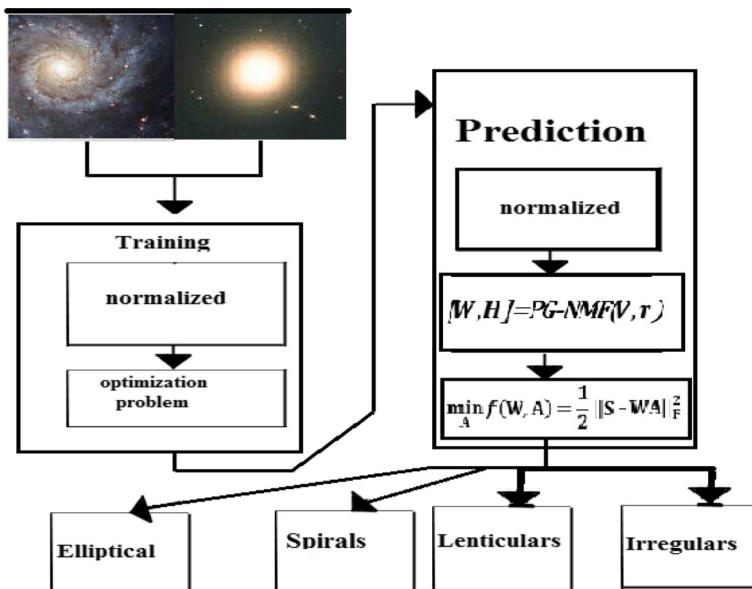


Fig. 2 Nonnegative matrix factorization galaxy classification scheme

$W$  and coefficient matrix  $H$ . The second main step is the test or prediction step, where the test set  $S$  is normalized. The coefficient matrix  $A$  of the  $S$  is computed based on basis matrix  $W$ , where the basis matrix  $W$  contains all information about different galaxy types (i.e., spiral, elliptical, lenticulars, and irregulars) and these basis make the process of separating the images that belong to the same type is easy (like blind source separation).

To predict the class of an unknown sample  $S_i$ , we used a MAX rule that selects the maximum coefficient in  $a_i$  (the coefficient vector of  $A$ ), and then assign the class label of the corresponding training sample to this new sample. The pseudocode of our algorithm is given in algorithm 1.

Algorithm 1. Galaxy-PGNMF algorithm
Input: $V$ : training set $r$ : cluster numbers $S$ : $p$ unknown samples without labels Output: $p$ : predicted class labels of the $p$ unknown samples If $r$ not Determined then using equations (8)- (9) to determine it. Training Step: 1. Normalize training set to have $l_2$ -norm 2. Solve the NMF optimization problem: $[W, H]=PG-NMF(V, r)$ Test Step: 1. Normalize test set to have $l_2$ -norm 2. Solve the NMF optimization problem: $\min_A f(W, A) = \frac{1}{2} \ S - WA\ _F^2$ 3. Predict the class label: $p_i = MAX(a_i)$ 4. return $p$ .

Algorithm 2: PG-NMF algorithm
Input : $V$ input matrix Output: $W$ basis matrix, $H$ coefficients matrix Initial: $W, H, 0 < \beta < 1, 0 < \alpha < 1$ for $k=1, 2, \dots$ , maximum iterations 1. $\alpha_k = \alpha_{k-1}$ 2. If $\alpha_k$ satisfies (7) do $\alpha_k = \alpha_k / \beta$ until $\alpha_k$ does not satisfy (7) or $H_k(\alpha_k / \beta) = H_k(\alpha_k)$ . Else do //decrease $\alpha_k = \alpha_k \cdot \beta$ until $\alpha_k$ satisfies (7). ENDIF 3. $H_{k+1} = \max(0, H_k - \alpha_k \nabla_H f(W, H_k))$ 4. $W_{k+1}$ can be easily obtained by repeating steps similar to (1)-(3). But in (1) consider $V^T$ and $H_{k+1}^T$ as input matrices.

**Table 1** Features of ellipticals, lenticulars and spirals, and the irregulars galaxies

Galaxy type	Feature
Spiral	<ol style="list-style-type: none"> <li>1. Bulge, the bulge is a spherical structure found in the center of the galaxy. The feature mostly contains older stars.</li> <li>2. Disk, the disk is made up of dust, gas, and younger stars. The disk forms arm structures which look like a spiral</li> <li>3. Halo. The halo is a loose, spherical structure around the bulge and some of the disk. The halo contains old stars clusters, as globular clusters.</li> </ol>
Elliptical	Elliptical galaxies have complete 2-dimensional symmetry. They have round but stretch longer a long one axis than along the other.
Lenticular	Lenticular galaxies have a central bulges and disks but no spiral arms. If the disk is faint, it is easy to mistake a lenticular galaxy for an elliptical galaxy- S0 galaxies look very much like E0 galaxies.
Irregular	Irregular galaxies have no regular or symmetrical structure, they have a wide variety of shapes and characteristics. They are frequently the result of collisions between galaxies or gravitational interactions between galaxies.

## 5 Results and discussion

The experiments were implemented in Matlab and run in windows environment with 64 bit support. Performance of the algorithms was evaluated by using different measures of performance such as NPV, sensitivity, accuracy, Specificity, and F-measure. These measures are defined as follows:

- (1) The classification accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \quad (11)$$

where  $TP$ ,  $TN$ ,  $FP$  and  $FN$  are represented the true positive, true negative, false positive and false negative, respectively.

- (2) Sensitivity measures (also called recall):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \quad (12)$$

**Table 2** The performance of proposed algorithm for small dataset

Average Accuracy	Sensitivity	Error Rate	Spirals		Ellipticals		Lenticulars		Irregulars	
			TP	Accuracy	TP	Accuracy	TP	Accuracy	TP	Accuracy
92.7	93	7.27	28	93	29	90	27	90	18	97



(3) Specificity:

$$\text{Specificity} = \frac{TN}{FP + TN} \times 100 \quad (13)$$

(4) The F-measure or F-score is defined as:

$$F\text{-measure} = \frac{2(\text{precision} + \text{recall})}{\text{precision} \times \text{recall}} \quad (14)$$

## 5.1 Data description

The proposed algorithm is applied to classify an images of the EFIGI catalog [33], a database contains sample all different Hubble types galaxies. The catalog merges data from standard surveys and catalogs (Principal Galaxy Catalogue, Sloan Digital Sky Survey, Value-Added Galaxy Catalogue, HyperLeda, and the NASA Extragalactic Database). The final EFIGI database is a large sub-sample of the local universe which it samples densely. It must be noted that the peculiar, interacting, duple and merging galaxies images are not included in our images. The images that contain an unknown object are not included also. The main feature of galaxies types used in the paper can be explained in Table 1.

## 5.2 Discussion

Since supervised learning is used, the classifier can be biased by the intuition of the person who prepares the standard training data. Therefore, training data for an image classifier that can be used for practical galaxy morphology classification should be selected and reviewed carefully.

### 5.2.1 Small data

In this section, we illustrate the performance of the proposed algorithm by selecting small number of galaxy imgaes. In which we selected 30 images from

**Table 3** Confusion matrix of the classification of spirals, ellipticals, lenticulars, and irregulars for small dataset

	Spirals	Ellipticals	Lenticulars	Irregulars
Spirals	28	0	1	1
Ellipticals	0	29	2	0
Lenticulars	1	1	27	1
Irregulars	1	0	0	18

**Table 4** The performance of proposed algorithm for large dataset

Average accuracy	Specificity	F-Measure	Sensitivity	Error rate	Spirals		Ellipticals		Lenticulars		Irregulars	
					TP	Accuracy	TP	Accuracy	TP	Accuracy	TP	Accuracy
91.9	93	72	72	8	51	92	4	94	86	88	18	92

**Table 5** Confusion matrix of the classification of spirals, ellipticals, lenticulars, and irregulars for small dataset

	Spirals	Ellipticals	Lenticulars	Irregulars
Spirals	51	2	5	2
Ellipticals	1	4	4	1
Lenticulars	2	2	86	4
Irregulars	1	1	4	18

each classes ellipticals, lenticulars and spirals, and 20 images from fourth class the irregulars.

The results are show in Tables 2 and 3, where Table 2 illustrates the accuracy and Table 3 shows the count matrix). From Table 2, we can conclude that the performance of the proposed algorithm for ellipticals, lenticulars and spirals, and the irregulars galaxies is 90, 90, 93, and 97%, respectively. The mean accuracy over all galaxies type is 92.7%. Also, in this table, the measure TP represents the true positive for each type of galaxies.

### 5.3 Large dataset

In this section, we select a large number of galaxy images from catalog as ellipticals, spirals, lenticulars, and irregulars galaxies, 225, 388, 99, and 34 respectively. This dataset is divided into training and testing set which represent nearly 75% (~558 images) and 25% (~188 images) from dataset respectively. The results of the proposed algorithm is given in Tables 4 and 5, where the accuracy for spirals, ellipticals, lenticulars, and the irregulars are 92, 94, 88, and 92% respectively. Also, the average over all types is 91.9%, the F-measure is 72% which indicates that high performance of the proposed algorithm.

In our sample, there are some galaxy images are poorly classified over different morphological classes. A standard problem in automatic classification is the projection of the galaxies on the sky. Not only these poorly classified is results from overlapping or foreground objects and the sky projection but also

**Fig. 3** Images which are not well classified

the most poorly classified galaxies have noise. Our algorithm can work well when SNR is low, in general, all observed images have noise so it would be reasonable to state, e.g., that 90% of the galaxies with a SNR of more than 60 at their effective radius (i.e., half light radius) could be classified correctly. There are some images are not well classification with different morphological type as shown in Fig. 3, with high SNR.

### 5.3.1 Comparison with other methods

We compare the proposed algorithm with other related works in Table 6. From this table we can conclude that, average of accuracy for most of the different algorithms nearly 91%. However, the number of galaxies type range from two to three types; therefore, the accuracy of the Galaxy-PGNMF algorithm is in

**Table 6** The accuracy of related works

Reference	Aim and accuracy
[34]	Used neural network and a locally weighted regression for classification. This paper compare between three types (E, S and Irr) with accuracy 91% and also compare between two types (E and S) with accuracy 95%
[35]	The author describes an supervised learning algorithm to automatically classify galaxy images. This algorithm used a large set of features is extracted from each image, and then select subset from these features using Fisher scores. Experimental results of classify spiral, elliptical and edge-on galaxies with accuracy of 90% compared to the classification carried out manually by the author.
[19]	The method presented in this paper is divided into three main tasks. The first one consists of locating and matching the objects in the multispectral images. In the second task a new representation for each astronomical object is created using its multispectral images, and also a set of features using principal component analysis are selected. In the last task the classification of the astronomical objects using neural networks, locally weighted linear regression and random forest algorithms. The results show that the method obtains over 93% accuracy classifying stars and galaxies.
[36]	the paper introduced an algorithm to classify images after Enhancement faint regions in images. This algorithm uses the segmentation and Morphological feature descriptors to the segmented images after a preprocessing stage and to extract the galaxy image structure for use in training the support vector machine learning classifier. The proposed algorithm compare between several types of galaxy images for example Irregular/Regular, Elliptical/Not Elliptical, Lenticular/Spiral and Spiral/Barred Spiral and it is accuracy nearly 80%.
[37]	This paper presented a three algorithms to galaxy classification (namely, Naive Bayes classifier, the rule-induction algorithm C4.5 and random forest (RF)). The experiments show that RF obtains the best results considering three, five and seven galaxy types. For all the galaxy classes, i.e., 91.64% accuracy for the three-class case, 54.72% accuracy for the five-class case, and 48.62% accuracy for the seven-class case; Naive Bayes obtained accuracy was of 43.62%. The number of PCs, are 13
[14]	Accuracy 90%

the same range of accuracy (or increase with small difference). Also, the Galaxy-PGNMF algorithm classify four types of galaxy images.

## 6 Conclusions

In this paper, we described an algorithm that can automatically classify images of spiral, elliptical, lenticular and Irregulars galaxies based on Nonnegative Matrix Factorization (NMF) algorithm. The data set used for the described experiments consists of galaxy images manually classified by Baillard and also classified visual by the authors (in this paper four morphological classes are used). In fact, our experiments show that the accuracy for small and large datasets is 92.7 and 91.9% respectively. There are many factors that effect on the accuracy of the proposed algorithm such as, a noise in the image, bias, dust scattering, overlapping or foreground objects, sky projections, and low surface brightness. The correct classification within four Hubble types is different from early type galaxies to late type galaxies as followe, for small (large) example 90% (94%) of ellipticals classifications and 90% (88%) of lenticular classifications 93% (92%) of spiral type classifications, and 97% (92%) for fourth class, the irregulars classifications. These results are compatible with those obtained by the human eye and as learned. We found that early types generally have a high optimum values where the late type low optimum values.

Finally, we conclude from the results that the Galaxy-PGNMF algorithm is a powerful alternative for galaxy morphology classification in CCD images.

Future work to improve classification requires more training data and integrating photometric features measured in different morphological types, repeating the experiments with a larger set of galaxies.

## References

1. Buta, R.J.: Vol. 6, Series Editor T. D. Oswalt, Volume editor W. C. Keel, Springer Reference (2011)
2. Bertin, G., Lin, C.: Spiral structure in galaxies. The MIT press (1996)
3. Gray, M.E., Wolf, C., Barden, M., et al.: MNRAS 393:1275 (2009)
4. Seigar, M., Kenefick, D., Kenefick, J., Lacy, C.: Astrophys. J. **678**, L93–L96 (2008)
5. von Hippel, T., Storrie-Lombardi, L.J., Storrie-Lombardi, M.C., Irwin, M.J.: MNRAS.269...97V (1994)
6. Bailer-Jones, C.A.L., Irwin, M., von Hippel, T.: MNRAS 298:1061 (1998)
7. Storrie-Lombardi, M. C., Lahav, O., Sodre, L., Jr., Storrie-Lombardi, L.: AAS...181.6508S (1992)
8. Naim, A., Lahav, O., Sodre, L., Jr, Storrie-Lombardi, C.: MNRA **275** (1995)
9. Peng, C.Y., Ho, L.C., Impey, C.D., Rix, H.-W.: AJ (2002)
10. Conselice, C.J.: yCat..21470001C (2003)
11. Abraham, R.G., van den Bergh, S., Nair, P.: ApJ. **588** (2003)
12. Shamir, L.: ApJ...736..141S (2011)
13. Davis, D.R., Hayes, W.B.: ApJ...790...87D (2014)
14. Schutter, A., Shamir, L.: A&C....12...60S (2015)
15. Almeida, C., Baugh, C., Lacey, C., Frenk, C., Granato, G. L., Silva, L., Bressan, A.: MNRAS. 402...544A (2010)
16. Ball, N.M., Loveday, J.: ASSL..319..771B (2004)
17. Shamir, L.: MNRAS.399.1367S (2009)

18. Banerji, M., Lahav, O., Lintott, C.J., Abdalla, F.B., Schawinski, K., Bamford, S.P., Andreescu, D., Murray, P., Raddick, M.J, Slosar, A., et al.: MNRAS.406..342B (2010)
19. de la Calleja, J., Fuentes, O.: Automated classification of astronomical objects in multispectral wide-field images. FLAIRS Conference. 610–612 (2006)
20. Hubble, E. P.: *Realm of the Nebulae*, ed. Hubble, E. P. (1936)
21. Bamford, S.P., Nichol, R.C., Baldry, I.K., Land, K., Lintott, C.J., Schawinski, K., Slosar, A., Szalay, A.S., Thomas, D., Torki, M., et al.: MNRAS **393** (2009)
22. van den Bergh, S.: *gmc.book.....V* (1998)
23. Land, K., Slosar, A., Lintott, C., Andreescu, D., Bamford, S., Murray, P., Nichol, R., Raddick, M.J., Schawinski, K., Szalay, A., et al.: MNRAS.388.1686L (2008)
24. Shamir, L., Nemiroff, R.J.: ApSoC **8:79S** (2008)
25. Shamir, L., Orlov, N., Eckley, D.M, Macura, T., Johnston, J., Goldberg, I.: ascl.soft12002S (2013)
26. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
27. Jolliffe, I.T.: *Principal components analysis*. Springer- Verlag, New York (2002)
28. Abd El-aziz, Mohamed, O.A., EL-Sayed Waheed, M.: *Math Probl Eng* **2011**:Article ID 150294
29. Khedr, W.M., Abd El-Aziz, M.E., Amer, S.M.: A novel algorithm for multichannel deconvolutive based on  $\alpha\beta$ -divergence. *Int. J. Comput. Appl.* **57**(10), 9–14 (2012)
30. Ozerov, A., Fevotte, C.: *IEEE Trans. Audio Speech Lang. Process.* **18**(3), 550–563 (2010)
31. Lin, C.: *IEEE Trans. Neural Netw.* **18**(6), 1589–1596 (2007)
32. Lin, C.: **19**(10):2756–2779 (2007b)
33. Baillard, A., Bertin, E., de Lapparent, et. al.: *Astron Astrophys* **532** (2011)
34. de la Calleja, J., Fuentes, O.: *Mon. Not. R. Astron. Soc.* **349**, 87–93 (2004)
35. Shamir, L.: *Mon. Not. R. Astron. Soc.* 000:1 (2005)
36. Jenkinson, J, Grigoryan, A.M., Agaian, S.S.: *Proc. SPIE 9399, Image Processing: Algorithms and Systems XIII*, 93990X (2015)
37. de la Calleja, J., Fuentes, O.: *American Association for Artificial Intelligence* (2007)