# Automated Classification of Galaxies Using Transformed Domain Features

*M.A. Mohamed[1] and M.M. Atta[2]*
*[1]Faculty of Engineering – Mansoura University – Egypt:*
*[2] MET Institute of Engineering and Technology-Mansoura-Egypt:*

**Summary**
Classifying galaxy information is one of the most important challenges for astronomers as it can provide important clues about the origin and evolution of the universe. In this paper; the performance of ten artificial neural networks (ANNs) based classifiers was evaluated and tested, based on a selected set of features. These features were extracted in frequency and wavelet domain; and then divided into three categories: (i) Fourier transform based; (ii) Cosine transform based, and (iii) wavelet transform based. The number of features in each category was determined empirically. The results showed that; (i) the support vector machine provided the best results with Fourier based features; (ii) the Jordan/Elman network (JEN) provided the best results in cosine and wavelet based features. In general, the cosine transform based features with JEN classifier provided the best results among all transformed based classifiers; about 0.45718% error in classification. The data set was taken from standardized catalogue from Zsolt Frei website.

**Key words:**
*Hubble Sequence, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Artificial Neural Networks (ANNs)*

## 1. Introduction

Galaxies have a wide variety of appearances. Some are smooth and some are lumpy. Another some have a well-ordered symmetrical spiral pattern, while others are diffuse, patchy, and irregular. For many decades, astronomers have tried to find the underlying order in the diverse properties of galaxies. This desire has led to the development of the Hubble sequence; a system of classification that groups galaxies into a series of types, based only on images; Fig.1 is the famous Hubble tuning fork diagram. Although the Hubble sequence is used to classify galaxies solely on the basis of images seen through a single filter (i.e. images that would appear to be black and white if printed), astronomers have found that many other properties of the galaxies, such as their colours and spectra, also correlate with their Hubble classification. In other words, galaxies that differ structurally or morphologically also tend to differ in their colours and spectral properties. The origin of this remarkable correlation is still not understood [1, 2].

### 1.1The Hubble Classification Scheme

Morphological galaxy classification is a system used by astronomers to classify galaxies based on their structure and appearance. The most common classification scheme is the system devised by Sir Edwin Hubble in 1936; he proposed the following classifications: (i) Elliptical: E0, E3, E5, and E7; (ii) Spiral: S0, Sa, Sb, Sc, and Sd; (iii) Barred spiral: SBa, SBb, and SBc; and (iv) Irregular: Im, and Ibm. Galaxies which fall early in this sequence are sometimes called early-type and galaxies near the end of the sequence are called late-type. A galaxy's Hubble type is based on the prominence of many different features, some of which are labelled in Fig.1. The most important of these features are: (i) the size of the bulge compared to the disk. Galaxies with a strong bulge but no detectable disk are classified as elliptical; denoted as 'E'. Galaxies with disks but no detectable bulges are classified as late-type spirals 'Sd' or irregulars 'Irr'. In between these two extremes are galaxies with both disks and bulges. These are classified as 'S0', 'Sa', 'Sb', and 'Sc', with galaxies of smaller bulges appear later in this sequence; (ii) the presence or absence of spiral arms. Galaxies with spiral arms are classified as 'Sa', 'Sb', 'Sc', and 'Sd'. Galaxies with smooth disks, but no spiral arms, are classified as 'S0'. Galaxies with lumpy irregular disks, but without spiral arms, are classified as irregulars 'Irr'; (iii) the tightness of the spiral arms: in early-type galaxies 'Sa', 'Sb', the spiral arms are very tightly wound around the galaxy. In many cases the arms are so tightly wound that one could not count how many spiral arms are present. In "late-type" galaxies 'Sc', 'Sd', the spiral arms are open and very loosely wound, making it easy to trace a spiral arm outwards from the center of a galaxy; (iv) the lumpiness of the spiral arms: the spiral arms of early-type galaxies tend to be very smooth, with few lumps. The spiral arms of late-type galaxies 'Sc', 'Sd' tends to be very lumpy. The lumps are actually compact regions of star formation known as HII regions, and (v) the presence or absence of bars: many spiral galaxies have linear arrangements of stars in their central regions that are distinct from the central bulge. These features are known as 'bars'. Spiral galaxies with bars are denoted with a capital 'B' after the initial 'S', so that a Sa galaxy with a bar is classified as a SBa, 'B' is always capitalized for barred galaxies [3].

### 1.2 Understanding Galaxy images

Two images of the same spiral galaxy are shown in Fig.2. The left panel shows the Familiar image that you would see if you took a picture of the galaxy through a single filter. The image on the right shows the "negative" (or

Manuscript received February 5, 2010
Manuscript revised February 20, 2010

"inverse") of the image on the left. The negative image has been altered so that dim parts of the original image are now white, and bright parts are now dark. Astronomers use these "negative" images to classify galaxies, because they typically show more detail than the more traditional "white-on-black" view.

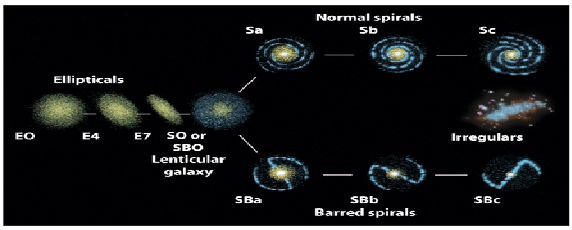


Fig.1 Hubble's Tuning Fork Diagram

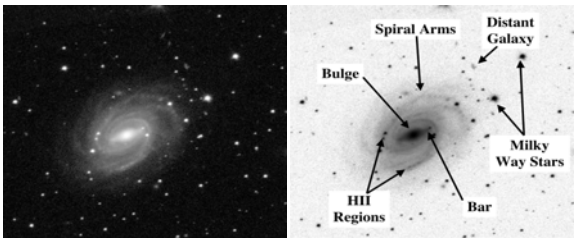


Fig. 2 Two Images of the Same Spiral Galaxy

### 1.3 System Architecture

The architecture of the system; shown in Fig. 3, is divided into four main phases: (i) image pre-processing; (ii) feature extraction; (iii) machine learning, and (iv) classification. In the image pre-processing phase, each galaxy is individually scaled, rotated, cropped, and centred to appear uniform for more accurate feature extraction. Then, in feature extraction, the dimension of the data has been reduced and thus retaining the most salient features using many techniques. In the machine learning and classification phase; we have used ANNs based classifiers to make learning process from these features.

## 2. Data Collection

The images used in this paper were obtained from the Zsolt Frei Catalogue provided by the department of Astrophysical Sciences at Princeton University [4]. This catalogue contains approximately 113 different galaxy images and is often used as a benchmark for astronomical study as the images are carefully calibrated. Taking into account the occurrence of galaxies and charged coupled device (CCD) camera multiple images stacking and supported by Sloan Digital Sky Survey (SDSS) [5].

These 113 galaxies were divided into two categories: (i) 31 galaxies were taken with the 1.5m telescope of the Palomar observatory in 1991, and (ii) 82 galaxies were taken with the 1.1m telescope of the Lowell observatory in 1989. At Palomar the acquisition process was accomplished using a camera with a resolution of 800×800 TI CCD, on the other hand, at Lowell the camera had a resolution of RCA 512×320 CCD. Palomar images are available in three passbands of the Thuan-Gunn system: g, r, and i. Lowell images are in two passbands (J and R) of the filter system developed by Gullixon et al. [6]. 134 images were randomly chosen from the catalogue to evaluate the performance of used techniques taking in account 30% of the selected images for the testing process.

The selected dataset of the Zsolt Frei catalogue has the following features: (i) high resolution; (ii) good quality, and (iii) careful calibration. Due to these characteristics, there will be no need for the preprocessing phase; image enhancement, histogram equalization, gamma correction, and debluring, in the proposed system. Hence, the following processing algorithms are performed to the raw dataset.

## 3. Performance Indices

Four different measures have been used to evaluate the performance of the ANN-based classification techniques: (i) mean-square-error (MSE); (ii) normalized mean-square-error (NMSE); (iii) correlation coefficient (r), and (iv) error percentage (%error). The mean squared error of an individual case (i) is evaluated by the equation:

$$MSE = \sum_{j=1}^{n} (P_{ij} - T_j)^2 / n \qquad (1)$$

where P(ij) is the value predicted by the individual case i for fitness case j (out of n fitness cases or sample cases); and Tj is the target value for fitness case j. The normalized mean square error is an estimator of the overall deviations between predicted and measured values. It is defined as:

$$NMSE = \sum_{j=1}^{n} (P_{ij} - T_j)^2 / (n \times P \times T) \qquad (2)$$

where:

$$P = \sum_{j=1}^{n} P_{ij} / n \text{ and } P = \sum_{j=1}^{n} T_j / n$$

The correlation coefficient(r) is a quantity that gives the quality of a least squares fitting to the original image. For two data sets x, y; the auto correlation is given by:

$$r = \mathrm{cov}(x, y) / (\sigma_x \times \sigma_y) \qquad (3)$$

where $\sigma_x$ and $\sigma_y$ are the standard deviation of image x and y. Finally; the error percentage is calculated as the percentage difference between the measured value and the accepted value.

## 4. FEATURE EXTRACTION

Feature extraction phase is generally used to reduce galaxy data dimensionality. In this paper three basic feature extraction techniques were extracted for the selected database of galaxy images [6].

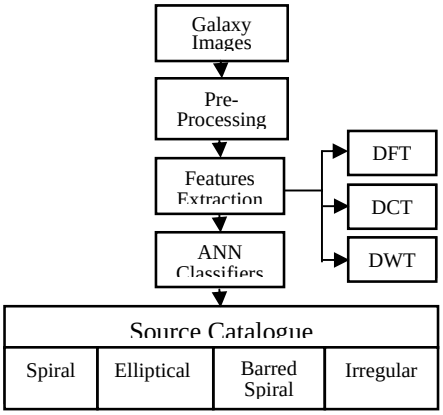


Fig.3. system architecture

### 4.1 Discrete Fourier Transform (DFT)

A DFT decomposes a sequence of values into components of different frequencies. This operation is useful in many fields but computing it directly from the definition is often too slow to be practical. Fast Fourier Transform (FFT) is a way to compute the same result more quickly; so FFT is an efficient algorithm to compute DFT, and it's inverse [7]. The sequence of N spatial complex coefficients $x_0$, $x_1 \ldots x_{N-1}$ is transformed into the sequence of *N* frequency complex coefficients $X_0$, $X_1 \ldots X_{N-1}$ by DFT according to the formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i 2\pi k \frac{n}{N}} \quad \text{where } k = 0,1,\cdots N-1 \qquad (4)$$

### 4.2 Discrete Cosine Transform (DCT)

A DCT expresses a sequence of finitely data points in terms of a sum of cosine functions oscillating at different frequencies however the use of cosine rather than sine functions turns out that cosine functions are much more efficient; due to: (i) energy compaction; (ii) decorrelation; (iii) separability; (iv) symmetry; and (v) orthognality. The DCT, and in particular the DCT-II, is often used in signal and image processing. The DCT is given according to the formula:

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] \quad \text{where } k = 0,1,\cdots,N-1 \qquad (5)$$

The DCT-II implies the boundary conditions: $x_n$ is even around $n$=-1/2 and even around $n$=$N$-1/2; $X_k$ is even around $k$=0 and odd around $k$=$N$ [8].

### 4.3 Discrete Wavelet Transform (DWT)

The wavelet transform has gained widespread acceptance in signal and image processing; because of their inherent multiresolution nature. Wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important.

The most commonly used wavelets were formulated by the Belgian mathematician Ingrid Daubechies in 1988. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale [9]. The DWT of a signal (x) is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response (g) resulting an output that given by:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]\, g[n-k] \qquad (6)$$

The signal is also decomposed simultaneously using a high-pass filter; h. The outputs are then divided into: (i) detail coefficients; from the high-pass filter, and (ii) approximation coefficients; from the low-pass. It should be noted that, the two filters are related to each other and they are known as a quadrature mirror filter. However, half the frequencies of the signal have now been removed and half the samples can be discarded according to Nyquist's rule. The filter outputs are then sub sampled by 2 as follow:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]\, g[2n-k] \qquad (7)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]\, h[2n+1-k] \qquad (8)$$

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled, Fig.4. This decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled. This is represented as a binary tree with nodes representing a sub-space with different time-frequency localization. The tree is known as a filter bank as shown in Fig.5. For the galaxy images the proposed coefficients are extracted using the technique showed in Fig.6.

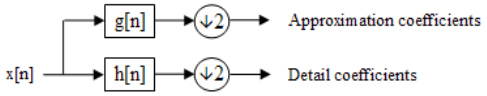


Fig.4. block diagram of filter analysis

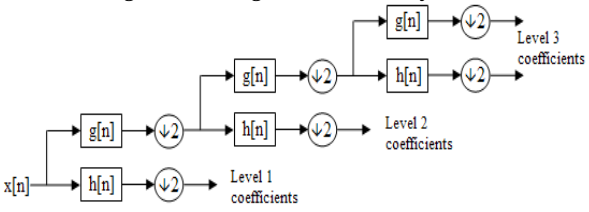


Fig.5. an example for 3 level filter bank

## 5. Galaxy Classification Techniques

An institutive goal of classification is to discriminate between stars and galaxies, while a more ambitious goal may be to classify different galaxy types or to measure structural properties of these galaxies. The end product of

classification is a source catalogue which lists the sources and their properties. There is a large number of ANNs based classifiers however the performance of ten of them will be evaluated and assessed.

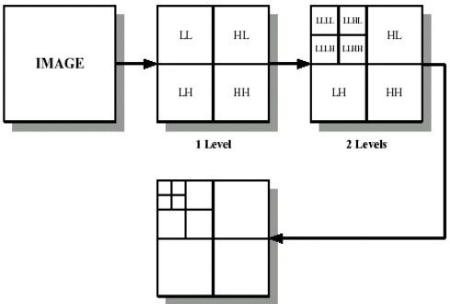


Fig.6. 2D DWT for galaxy image

## 5.1 Multilayer Perceptron (MLP)

MLP is a layered feed forward networks typically trained with static back propagation [10]. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are ease to use, and that they can approximate any input/output map. The key disadvantages are that they train slowly, and require lots of training data (typically three times more training samples than network weights).

## 5.2 Generalized Feed-Forward (GFF)

GFF networks are a generalization of the MLP such that connections can jump over one or more layers. In theory, a MLP can solve any problem that a generalized feed-forward network can solve. In practice, however, GFF networks often solve the problem much more efficiently. It suffices to say that a standard MLP requires hundreds of times more training epochs than the GFF network containing the same number of processing elements [11].

## 5.3 Modular Neural Network (MNN)

MNN networks are a special class of MLP. These networks process inputs using several parallel MLPs, and then recombine the results. This tends to create some structures within the topology, which will foster specialization of function in each sub-module. In contrast to MLP modular networks don't have full interconnectivity between their layers [12].

Therefore, a smaller number of weights are required for the same size network. This tends to speed up training times and reduce the number of required training exemplars. There are many ways to segment a MLP into modules. It is unclear how to best design the modular topology based on the data. There are no guarantees that each module is specializing its training on a unique portion of the data.

## 5.4 Jordan/Elman Network (JEN)

Jordan and Elman networks extend the multilayer perceptron with context units, which are processing elements (PEs) that remember past activity. Context units provide the network with the ability to extract temporal information from the data. In the Elman network, the activities of the first hidden PEs are copied to the context units, while the Jordan network copies the output of the network. Networks which feed the input and the last hidden layer to the context units are also available [13].

## 5.5 Principal Component Analysis (PCA) Network

Principal component analysis networks (PCAs) combine unsupervised and supervised learning in the same topology. Principal component analysis is an unsupervised linear procedure that finds a set of uncorrelated features, principal components, from the input. A MLP is supervised to perform the nonlinear classification from these components [14].

## 5.6 Radial Basis Function (RBF) Networks

Radial basis function (RBF) networks are nonlinear hybrid networks typically containing a single hidden layer of processing elements. This layer uses Gaussian transfer functions, rather than the standard sigmoidal functions employed by MLPs. The centres and widths of the Gaussians are set by unsupervised learning rules, and supervised learning is applied to the output layer. These networks tend to learn much faster than MLPs. If a generalized regression (GRNN) probabilistic (PNN) net is chosen, all the weights of the network can be calculated analytically. In this case, the number of cluster centres is by definition equal to the number of exemplars, and they are all set to the same variance [15].

## 5.7 Self Organized Maps (SOMs)

Self-organizing maps (SOMs) transform the input of arbitrary dimension into a one or two dimensional discrete map subject to a topological (neighbourhood preserving) constraint. The feature maps are computed using Kohonen unsupervised learning. The output of the SOMs can be used as input to a supervised classification neural network such as the MLP .This network's key advantage is the clustering produced by the SOM which reduces the input space into representative features using a self-organizing process. Hence the underlying structure of the input space is kept, while the dimensionality of the space is reduced [16].

## 5.8 Time Lagged Recurrent Networks (TLRNs)

Time lagged recurrent networks (TLRNs) are MLPs extended with short term memory structures. Most real-world data contains information in its time structure, i.e. how the data changes with time. Yet, most neural networks are purely static classifiers. TLRNs are the state of the art in nonlinear time series prediction, system identification and temporal pattern classification [17].

### 5.9 Recurrent Networks (RNs)

Actually, there are two types of recurrent networks: (i) fully recurrent networks feedback the hidden layer to itself and (ii) partially recurrent networks start with a fully recurrent net and add a feed forward connection that bypasses the recurrency, effectively treating the recurrent part as a state memory. These recurrent networks can have an infinite memory depth and thus find relationships through time as well as through the instantaneous input space. Most real-world data contains information in its time structure. Recurrent networks are the state of the art in nonlinear time series prediction, system identification, and temporal pattern classification [18].

### 5.10 Support Vector Machine (SVM)

Support Vector Machine (SVM) is implemented using the kernel Adatron algorithm. The kernel Adatron maps inputs to a high-dimensional feature space, and then optimally separates data into their respective classes by isolating those inputs which fall close to the data boundaries. Therefore, the kernel Adatron is especially effective in separating sets of data which share complex boundaries. SVMs can only be used for classification, not for function approximation [19].

## 6. Results

The performances of all tested ANN-based classifiers were evaluated through four performance indices; MSE; NMSE; r, and %Error. Table.1 through Table.10 illustrates the results of the ten discussed ANN classifiers. All these classifiers are trained with the following parameters: (i) 10 processing elements; (ii) one hidden layer, and (iii) 1000 epochs. As a result of the comparative study it has been found that:(i) the DFT features were applied to all ANN-based classifiers, it was found that; the classifier based on SVM provides the best results: MSE=0.00836; NMSE=0.01141; r=0.99846, and %Error=1.13857, (ii) the DCT features were applied to all ANN-based classifiers, it was found that; the classifier based on JEN provides the best results: MSE=0.00311;NMSE=0.00524;r=0.99798, and Error= 0.45718, and (iii) the DWT (based on dB2 and one level decomposition) features were applied to all ANN-based classifiers, it was found that; the classifier based on JEN provides the best results: MSE=0.00442; NMSE=0.00747; r=0.99715, and %Error=0.67527.

On the other hand, the RBF based classifiers provides the worst results for all cases.Fig.7, Fig.8, Fig.9, Fig.10, Fig.11; and Fig.12 show the active cost of the learning and classification curves for these best cases discussed below respectively.

Table.1: results of MLP classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.43849 | 0.07679 | 0.01364 |
| NMSE | 0.73909 | 0.12944 | 0.02306 |
| r | 0.49311 | 0.93913 | 0.99061 |
| % Error | 13.0751 | 3.24279 | 1.00487 |

Table.2: results of GFF classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.04831 | 0.01396 | 0.02958 |
| NMSE | 0.08142 | 0.02353 | 0.05000 |
| r | 0.95775 | 0.98436 | 0.97582 |
| % Error | 3.00228 | 0.94146 | 1.76802 |

Table.3: results of MNN classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| *MSE* | 0.13445 | 0.04373 | 0.15087 |
| *NMSE* | 0.22663 | 0.07372 | 0.25501 |
| *r* | 0.87432 | 0.95995 | 0.85021 |
| % Error | 5.68324 | 2.45103 | 5.81759 |

Table.4: results of JEN classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.10777 | 0.00311 | 0.00442 |
| NMSE | 0.18164 | 0.00524 | 0.00747 |
| r | 0.88035 | 0.99798 | 0.99715 |
| % Error | 4.24276 | **0.45718** | **0.67527** |

Table.5: results of PCA classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.48388 | 0.31199 | 0.17576 |
| NMSE | 0.81558 | 0.52586 | 0.29707 |
| r | 0.39004 | 0.65896 | 0.81618 |
| % Error | 15.1219 | 10.0772 | 6.07928 |

Table.6: results of RBF classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.48144 | 0.56608 | 0.52590 |
| NMSE | 0.81147 | 0.95414 | 0.88889 |
| r | 0.43377 | 0.42383 | 0.42643 |
| % Error | 15.1285 | **16.2810** | 15.9994 |

Table.7: results of SOM classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.22028 | 0.46530 | 0.38146 |
| NMSE | 0.37129 | 0.78428 | 0.64475 |
| r | 0.80245 | 0.50135 | 0.60854 |
| % Error | 7.44578 | 14.1873 | 11.7651 |

Table.8: results of TLRN classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| *MSE* | 0.13128 | 0.11715 | 0.12257 |
| *NMSE* | 0.22127 | 0.19746 | 0.20717 |
| *r* | 0.89027 | 0.90200 | 0.89609 |
| *% Error* | 2.99721 | 2.6699 | 2.92627 |

Table.9: results of RN classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.24598 | 0.11507 | 0.24618 |
| NMSE | 0.41460 | 0.19396 | 0.41610 |
| r | 0.76844 | 0.90726 | 0.76889 |
| %Error | 7.66501 | 3.7318 | 5.74637 |

Table.10: results of SVM classifier

| | DFT | DCT | DWT |
|---|---|---|---|
| MSE | 0.00836 | 0.00890 | 0.01148 |
| NMSE | 0.01141 | 0.01215 | 0.01572 |
| r | 0.99846 | 0.99729 | 0.99405 |
| % Error | **1.13857** | 1.13740 | 1.13858 |

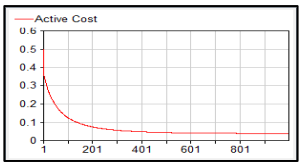


Fig.7. learning curve of DFT/SVM based classifier

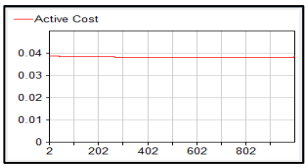


Fig.8. testing curve of DFT/SVM based classifier

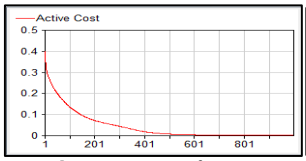


Fig.9. learning curve of DCT/JEN based classifier

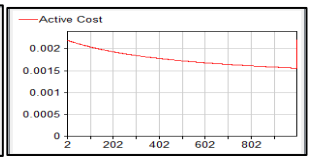


Fig.10. testing curve of DCT/JEN based classifier

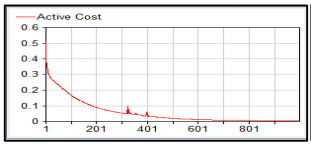


Fig.11. learning curve of DWT/JEN based classifier

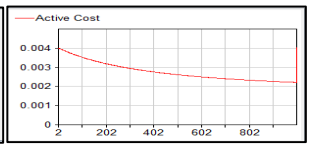


Fig.12 testing curve of JEN based classifier

## 7. Conclusion

The process of classifying galaxy information has gained importance; with the growth of scope development and accurate database capturing. During, the past decade the analysis of galaxy information was concerned with morphological features as well as principal component analysis based features; or different combinations between them. In this thesis, we have developed ANN-based classifiers based on the transformed domain features. As a result of the comparative study it was found that: (i) with DFT-features; the SVM provided the best results; (ii) with DCT-features; JEN provided the best results, and (iii) with DWT-features; JEN provided the best results. As a final conclusion, the DCT-features with JEN-based classifier provided the best results among all discussed classifiers: MSE=0.00311; NMSE=0.00524; r=0.99798, and %Error= 0.45718.

## 8. References

[1] J. Fuentes, "Machine learning & image analysis for morphological galaxy classification," Monthly Notices of the Royal Astronomical Society, Vol. 24, 2004.

[2] J.B. Hutchings, "Hubble telescope resolved image & spectra of the z2 QSO 1345+584," The Astronomical Journal, Vol.116; No. 20E25, July, 1998.

[3] A. Jandir, U. de Mendonça, E.C. de Barros, et al., "A mathematical morphology approach to the star/galaxy characterization," J. Braz. Comp. Soc. Vol. 3, No. 3, Campinas Abr. 1997.

[4] Z. Frei, 1999, Zsolt Frei Galaxy Catalog. Retrieved 2002 from Princeton University, Department of Astrophysical Sciences Web site: http://www.astro.princeton.edu/ frei/catalog.htm

[5] Sloan Digital Sky Survey (SDSS), web site: http://www.worldwindcentral.com/wiki/Add-on:SDSS.

[6] G.W. Zack, W.E. Rogers, and S.A. Latt, "Automatic measurement of sister chromatid exchange frequency," Journal of Histochemistry and Cytochemistry Vol. 25, No. 7, pp.741-753, 1997.

[7] T. H. Manjula, H.S.Manjunatha, K.B. Raja, and L.M. Patnaik, "Detecting original image using histogram, DFT and SVM," International Journal of Recent Trends in Engineering Vol. 1, No. 1, May 2009.

[8] A.B. Watson, "Image compression using the DCT," Mathematica Journal, Vol.4, No.1, p. 81-88, 1994.

[9] C. Torrence, & G. Compo, "A practical guide to wavelet analysis," the American Meteorological Society, Vol.79, No.1, January 1998.

[10] D. Ruck, S. Rogers, & M. Kabrishy, "Feature selection using a multilayer Perceptron," Journal of Neural Network Computing, Vol.2, No.2, pp. 40-48, 1990.

[11] T. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward ANN," Neural Networks, Vol. 2, pp. 459-473, 1989.

[12] A. Kehagias and V. Petridis, "Predictive modular neural networks for timeseries classification," Journal of Neural Networks, Vol.10, pp. 31-49, 1997.

[13] V.R. Mankar and A.A. Ghatol, "Design of adaptive filter using Jordan/Elman neural network in a typical EMG signal noise removal," Hindawi Publishing Corporation Vol. 2009, Article ID 942697, 9 pages doi:10.1155/2009/942697.

[14] W.W. Hsieh, "Nonlinear principal component analysis by neural networks," University of British Columbia, Appeared in Tellus, 2001, vol. 53A, pp. 599-615.

[15] J. Park and I.W. Sandberg, "Approximation and radial-basis-function networks," Neural Computation, Vol.5, pp. 305-316, 1993.

[16] T. Honkela, S. Kaski, K. Lagus, and T. Kohonen, "WEBSOM—self-organizing maps of document collections," In Proceedings of WSOM'97, pp. 310–315. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, June-1997.

[17] Y. Liang and A. Kelemen, "Time lagged recurrent neural network for temporal gene expression classification," International Journal of Computational Intelligence in Bioinformatics and Systems Biology Vol. 1, No. 1, 2009.

[18] A. atiya, and A.G. Parlos, "New results on recurrent network training: unifying the algorithms and accelerating convergence," IEEE Transactions on Neural Networks, Vol. 11, No. 3, May 2000.

[19] L.M. Manevitz and M. Yousef, "One-class SVMs for document classification," Journal of machine learning, Vol.2, 2001, pp.139-154.