The Astronomical Journal, 161:94 (9pp), 2021 February https://doi.org/10.3847/1538-3881/abd314


# Morphological-based Classifications of Radio Galaxies Using Supervised Machine-learning Methods Associated with Image Moments

Mohammad Sadeghi[1], Mohsen Javaherian[2], and Halime Miraghaei[2]
[1] Department of Physics, University of Kurdistan, Pasdaran Street, P.O. Box 66177-15175, Sanandaj, Iran
[2] Research Institute for Astronomy and Astrophysics of Maragha (RIAAM), University of Maragheh, 55136-553, Maragheh, Iran; javaherian@maragheh.ac.ir
*Received 2020 September 21; revised 2020 December 5; accepted 2020 December 9; published 2021 January 29*

## Abstract

With the advent of new high-resolution instruments for detecting and studying radio galaxies with different morphologies, the need for the use of automatic classification methods is undeniable. Here, we focused on the morphological-based classification of radio galaxies known as Fanaroff–Riley (FR) type I and type II via supervised machine-learning approaches. Galaxy images with a resolution of 5″ at 1.4 GHz provided by the Faint Images of the Radio Sky at Twenty centimeters (FIRST) survey are employed. The radial Zernike polynomials are exploited to extract image moments. Then, the rotation, translation, and scale-invariant moments of images are used to form a training set (65% of the radio galaxy sample) and a test set (the remaining 35%). The classes of the test set are determined by two classifiers: a support vector machine and a twin support vector machine (TWSVM). In addition the genetic algorithm is employed to optimize the length of moment series and to find the optimum values of the parameters of the classifiers. The labels of outputs are compared to identify the best performance classifier. To do this the confidence level of classifications is estimated by four different metrics: precision, recall, F1 score, and accuracy. All tests show that implementing TWSVM with the radial basis function as a kernel achieves a confidence level of more than 95% in grouping galaxies.

*Unified Astronomy Thesaurus concepts:* Radio galaxies (1343); Support vector machine (1936); Classification (1907); Catalogs (205)

## 1. Introduction

Radio galaxies identified by their luminous nonthermal and polarized emissions over the radio passband of the electromagnetic spectrum are considered as an important branch of active galactic nuclei (AGNs). The energy injected by AGNs into the interstellar and intergalactic medium via radiation, outflows, and radio jets is one of the main candidates to quench star formation in galaxies, describing the transition of galaxies from blue star-forming to red and almost dead (Baldry et al. 2004; Croton et al. 2006). Feedback from AGNs can also explain why massive galaxies stop further growth as observed in the bright end of the local galaxy luminosity function (Benson et al. 2003). Therefore, they are generally believed to have a significant impact on the formation and evolution of galaxies. Studying these objects has developed our understanding of the physical processes around the supermassive black holes and their energy feedback to the intergalactic medium (McNamara & Nulsen 2007). They are also very effective tools for investigating the high redshift universe as radio waves are unaffected by the galactic dust, unlike other electromagnetic wavelengths. Thus, detecting and identifying radio-loud AGNs are some of the main goals of the new generation of radio telescopes.

Radio galaxies display a wide range of morphologies from compact sources such as compact symmetric objects, gigahertz-peaked spectrum sources, and compact steep spectrum radio sources (O'Dea & Baum 1997; O'Dea 1998) to very extended mega-parsec scale sources such as giant radio galaxies (e.g., Saripalli et al. 1986; Dabhade et al. 2020). According to the seminal classification of the extended radio galaxies, which was developed by Fanaroff & Riley (FR; 1974), there are two morphologically different populations of radio galaxies. In the FR type I population (FRI), the distances between the peak of the emission on the opposite sides of the radio source was smaller than half of the total size of the radio source so that the brightness decreases from the core of the emission to the edge (also called edge-darkened). They mostly show a tail of diffuse emission the faded away into the intergalactic medium. In the FR type II population (FRII), the peak of the emission is located in the second half of the distance between the core and the edge of the emission (also called edge-brightened). Recently, compact radio sources have been added to this classification as a third class of radio galaxies as FR0 by Baldi et al. (2015). The three types of radio galaxies seem to be different in either host galaxy properties or the environments they inhabit (Miraghaei & Best 2017). Investigation of these differences helps to shed light on the jet formation and its interaction with the environment.

While there are the main classifications, the morphology is not discrete. A more complex extended source population has also been observed including hybrid morphology radio sources (HyMoRS; Gopal-Krishna & Wiita 2000; Harwood et al. 2020) that are observationally FRI on one side and FRII on the other side, FRI lobed or tailed (e.g., Laing et al. 2011; Mingo et al. 2019), bent-tailed FRI such as narrow-angle tailed (head–tail) and wide-angle tailed radio sources (Miley et al. 1972; Burns 1981; Smolčić et al. 2007), and double–double or restarting FRII (Miraghaei & Best 2017; Mingo et al. 2019).

Due to the limitations on radio surveys in either the sensitivity or resolution in the past four decades, our knowledge about radio galaxies is limited. The largest FRI/FRII radio galaxy samples of an order of a few hundred galaxies provide poor statistics on the FR radio galaxy studies. Fortunately, progress is expected to be rapid with the more recent radio surveys such as the LOFAR Two-Metre Sky Survey (LoTSS; Shimwell et al. 2017), MeerKAT International GHz Tiered Extragalactic Exploration survey (MIGHTEE;

Jarvis et al. 2016), GaLactic and Extragalactic All-sky MWA survey (GLEAM; Wayth et al. 2015), and eventually the Square Kilometer Array (SKA) surveys. As the statistics improve in the current and future radio surveys, new challenges for identifying and classifying the extensive radio structures will arise. Automated classification of radio galaxies is one of the main issues. Therefore, it is necessary to propose new methods and improve the algorithms to handle a vast amount of data. To do this, machine-learning techniques are widely used in different data-rich areas of science. In the field of astronomy, there are outstanding works which use supervised and unsupervised methods to accurately identify and classify targets such as exoplanets (Shallue & Vanderburg 2018), solar features (Javaherian et al. 2014; Arish et al. 2016), unresolved binary stars (Kuntzer & Courbin 2017), Laser Interferometer Gravitational-Wave Observatory (LIGO) findings (Zevin et al. 2017), and strong lenses (Jacobs et al. 2017).

For classifying different types of galaxies, some automated identification processes were proposed. Baron & Poznanski (2017) developed a method based on the Random Forests algorithm for detecting "weird" galaxies. Bastien et al. (2017) employed a set of machine-learning algorithms to discriminate bent radio galaxies in a mixture of point-like/extended images. In recent works, deep machine-learning techniques were used to classify radio galaxies (Aniyan & Thorat 2017; Lukic et al. 2018). One of the novel unsupervised methods that works based on neighborhood similarity and K-means clustering was proposed by Ralph et al. (2019) to segregate features observed in radio-astronomical images. In terms of accurate classification of radio galaxies, a study by Lukic et al. (2019) revealed that convolutional neural networks have a better performance than capsule networks. The convolutional neural network which exploits both unsupervised and supervised methods was also used in Ma et al. (2019) to classify radio AGNs. Using the transfer-learning method, Tang et al. (2019) provide a 13-layer deep convolutional neural network to segregate radio galaxies based on their morphologies. Finally, Mingo et al. (2019) carried out the morphological classification of radio galaxies and host galaxy properties using an automated classification method that works based on thresholding and masking.

In this paper, a sample of morphologically extended radio galaxies is classified into FRI/FRII types using image moments and two different classifiers: a support vector machine (SVM) and a twin support vector machine (TWSVM). Our analysis is based on the main classification of radio galaxies to examine the success of the SVM and TWSVM methods. Sources with more complex structures such as HyMorS, which are neither FRI nor FRII, are excluded from this analysis.

The paper is organized as follows. The radio galaxy sample and image data set are described in Section 2. Different steps of the proposed detection method consisting of extracting the Zernike moments, SVM-based classifiers, and the genetic algorithm (GA) are presented in Section 3. In Section 4, we present the results and discuss the accuracy of our classification methods. Finally, the summary and concluding remarks are given in Section 5.

## 2. Description of the Data Set

The FR radio galaxy sample has been drawn from Miraghaei & Best (2017). They have published a catalog of 1329 morphologically classified extended radio sources with their optical counterparts following the study published by Best (2009). The catalog contains extended radio galaxies classified as FRI, FRII, and FR hybrid, which contains a few sources left unclassified. The classification is performed based on the visual inspection of the multicomponent radio sources identified from the radio source catalog of Best & Heckman (2012). Best & Heckman (2012) cross-matched the seventh data release (DR7; Abazajian et al. 2009) of the Sloan Digital Sky Survey (SDSS; York et al. 2000) with the National Radio Astronomy Observatory (NRAO) Very Large Array Sky Survey (NVSS; Condon et al. 1998) and the Faint Images of the Radio Sky at Twenty centimeters (FIRST) survey (Becker et al. 1995), following the techniques of Best et al. (2005).

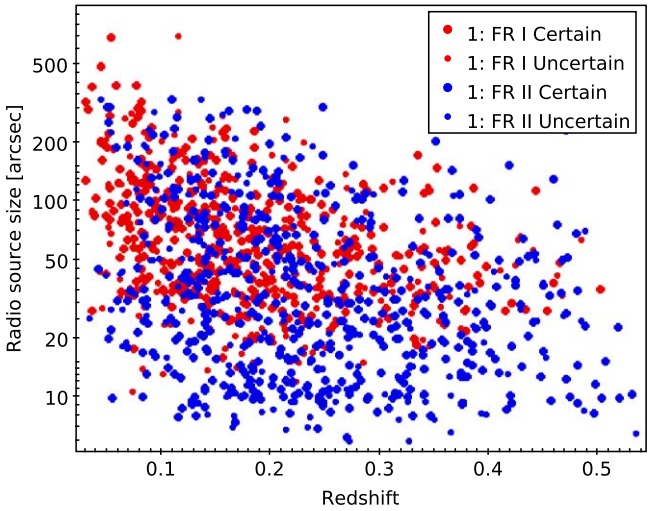


**Figure 1.** The size of FRI (red) and FRII (blue) radio galaxies with respect to the redshift of the host galaxies. The certain and less secure sources are displayed by the large and small circles, respectively.

Given the NVSS/FIRST sensitivity and resolution limitations and based on the intensity and extent of the radio sources, the classifications were labeled as certain or uncertain. In total 410 galaxies were classified as FRI certain, 198 as FRI uncertain, 446 as FRII certain, and 200 as FRII uncertain. The catalog also includes the identification of more complex structures such as wide-angle tailed, head–tail, diffuse, and double–double radio galaxies. The FR uncertain sources, FR hybrids, and sources labeled in the catalog with complex structures were totally excluded from this analysis. The number of FR hybrids and the other sources with complex structures in the catalog was not enough to provide appropriate statistics for each type and were therefore excluded from the analysis.

The fits images have been extracted from the FIRST cutout server. The resolution of FIRST is 5″ with a typical rms of 0.15 mJy. The pixel size of the images is 1.″8. We have used $5' \times 5'$ images in this analysis. The images are then $167 \times 167$ pixels. The radio sizes of the FR radio galaxies concerning the redshifts of their host galaxies are shown in Figure 1. Examples of FRI and FRII radio galaxy images are shown in Figures 2 and 3.

## 3. Description of Identification Methods

Here, we aim to present a brief description of the employed algorithms for detecting galaxies automatically. We first exploit the Zernike function to extract image moments, and then, the GA is employed to optimize the length of the moments series. Next, we use both nonlinear SVM and TWSVM classifiers to classify galaxies and compare the classifiers' performance.

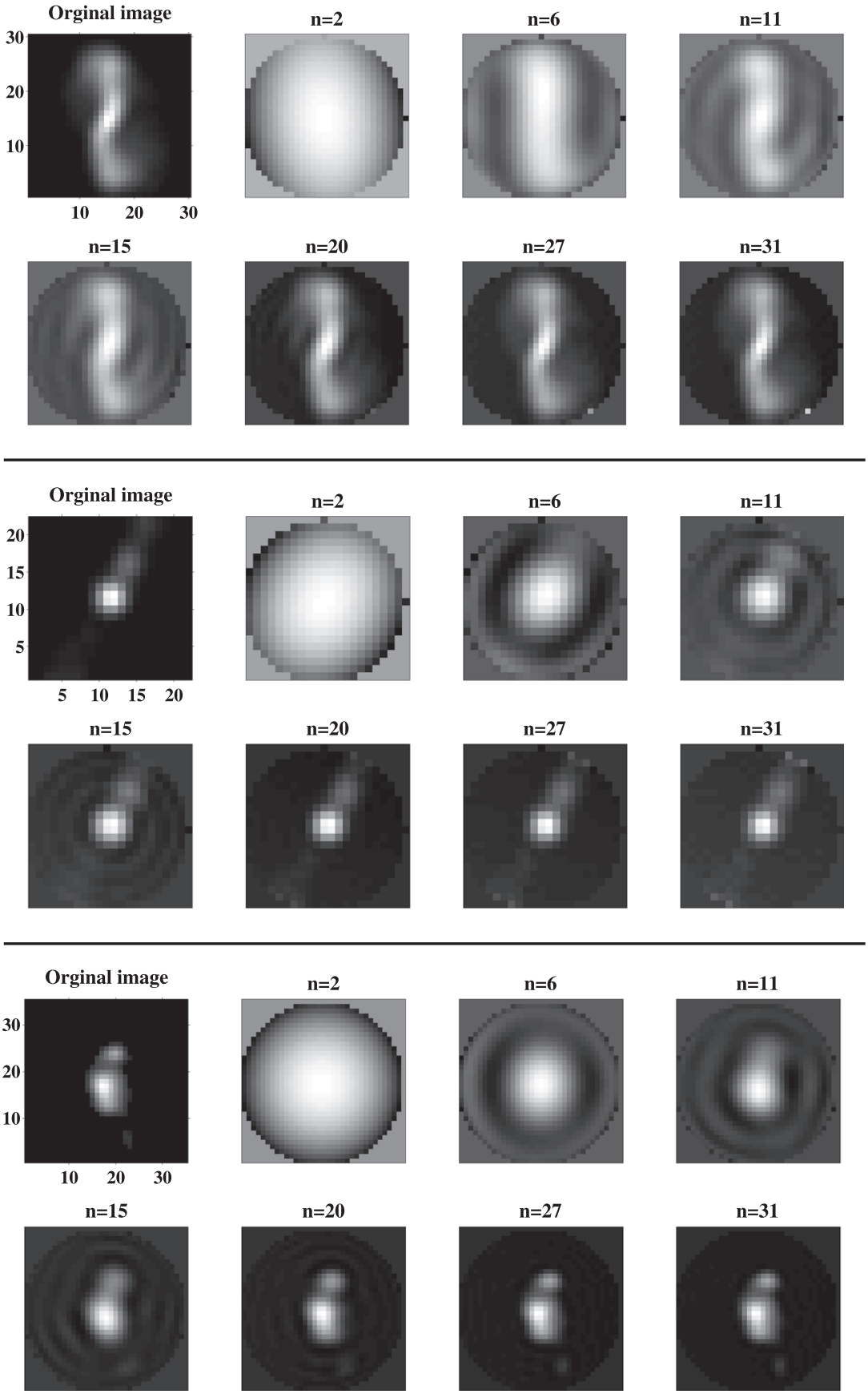


**Figure 2.** Reconstructed image of three FRI galaxies with various order numbers of moments.

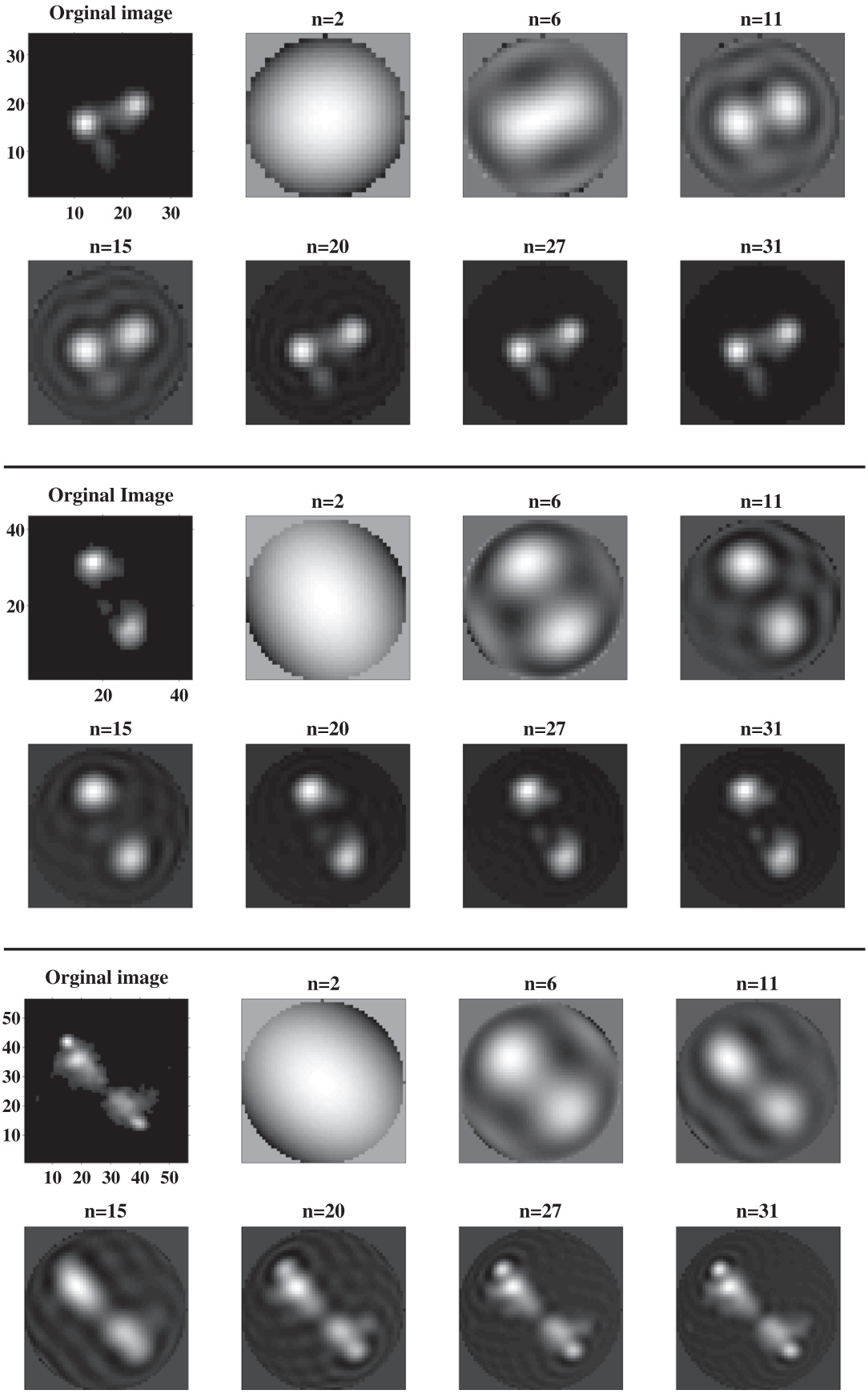


**Figure 3.** Reconstructed image of three FRII galaxies with various order numbers of moments.

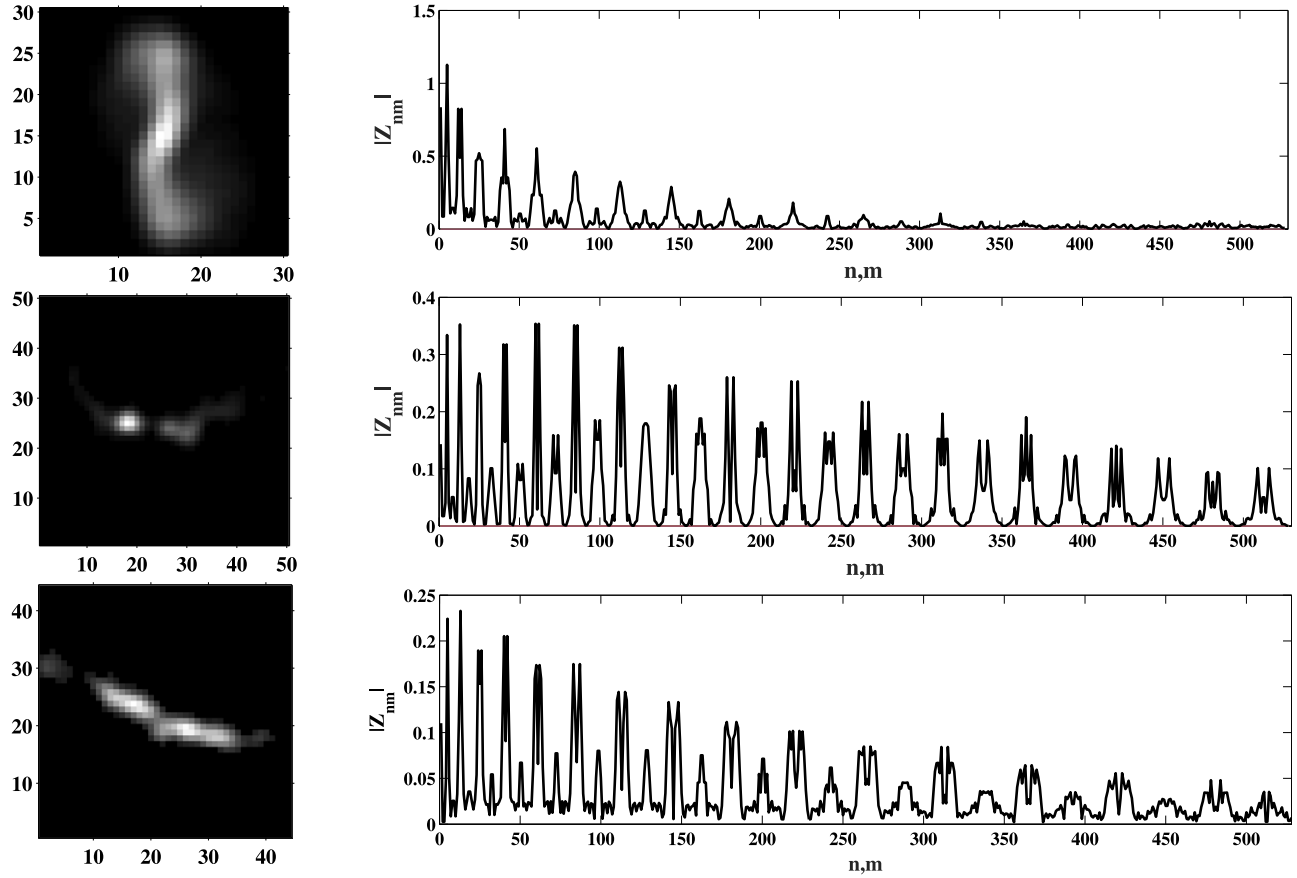


**Figure 4.** Magnitude values of Zernike moments, $Z_{nm}$, of three FRI galaxies.

Again, we use the GA to find the optimum values of parameters defined in classifiers leading to decreases in computational expenses.

### 3.1. Zernike Moments

One of the common ways for obtaining image moments is using moment functions such as Hu, Legendre, and Zernike (e.g., see Ming-Kuei 1962; Khotanzad & Hong 1990). To extract image features in terms of the complete set of Zernike orthogonal polynomials, we used the Zernike function $V(r, \theta)$ as follows

$$V_{nm}(r, \theta) = R_{nm}(r)\exp(im\theta),$$
$$R_{nm}(r) = \Sigma_{s=0}^{(n-m)/2}(-1)^s \times \frac{(n-s)!}{s!\left(\frac{(n+|m|)}{2}-s\right)!\left(\frac{(n-|m|)}{2}-s\right)!} r^{n-2s}, \quad (1)$$

where $s$ is a numerator, $n$ is the positive integer order number, and $m$ is the repetition integer number. These two parameters have the constraints: $|m| \leqslant n$, and $n - |m|$ is even (Abandah & Anssari 2009; Hosny 2010). Since the Zernike function is naturally described in a polar coordinate system denoted by $r$ and $\theta$, each image tile must be mapped to a circle with a unit radius. Then, the image intensity function $I(r, \theta)$ can be expressed as terms of the Zernike polynomials as

$$I(r, \theta) = \Sigma_{n=0}^{n_{up}}\Sigma_{m=0}^{n} Z_{nm} V_{nm}(r, \theta). \quad (2)$$

Thus, we can obtain the Zernike moments as

$$Z_{nm} = \frac{n+1}{\pi}\int\int_{\text{unit cirle}} V_{nm}^{*}(r, \theta) I(r, \theta) r dr d\theta. \quad (3)$$

Since each function can be expanded by the Zernike polynomials, the extracted Zernike moments of image features are unique (Hosny 2010). Empirically, it has been shown that the optimum value for the maximum order number $n_{up}$ is equal to 31, which approximately maintains the quality of the original image after reconstruction and prevents further computational expenses (Alipour et al. 2012; Javaherian et al. 2014). In fact, the normalized reconstruction error as a function of the order number for the image tiles decreases to the smallest value for $n_{up} = 31$. For an $(n, m)$ set, the number of Zernike moments is achievable by $\Sigma_{n=0}^{n=n_{up}}(n+1)$, which equals to 528 features (Javaherian et al. 2014; Alipour et al. 2019). We plot several original image tiles of FRI (Figure 2) and FRII radio galaxies (Figures 3), and their reconstructed images for different order numbers of $n$. As it is seen, for an order number of $n = 31$, the reconstructed image tiles are similar to the original ones.

Because of the exponential Fourier term $[\exp(im\theta)]$ of the Zernike polynomials, which is computable for a shape over different angles, the Zernike polynomials are naturally rotation invariant. There are two popular ways to make the extracted image moments scale and translation invariant: in the first approach, the geometrical moments for normalizing an image are used (Khotanzad & Hong 1990). In the second approach, first, the resizing process of the object is performed to fit the

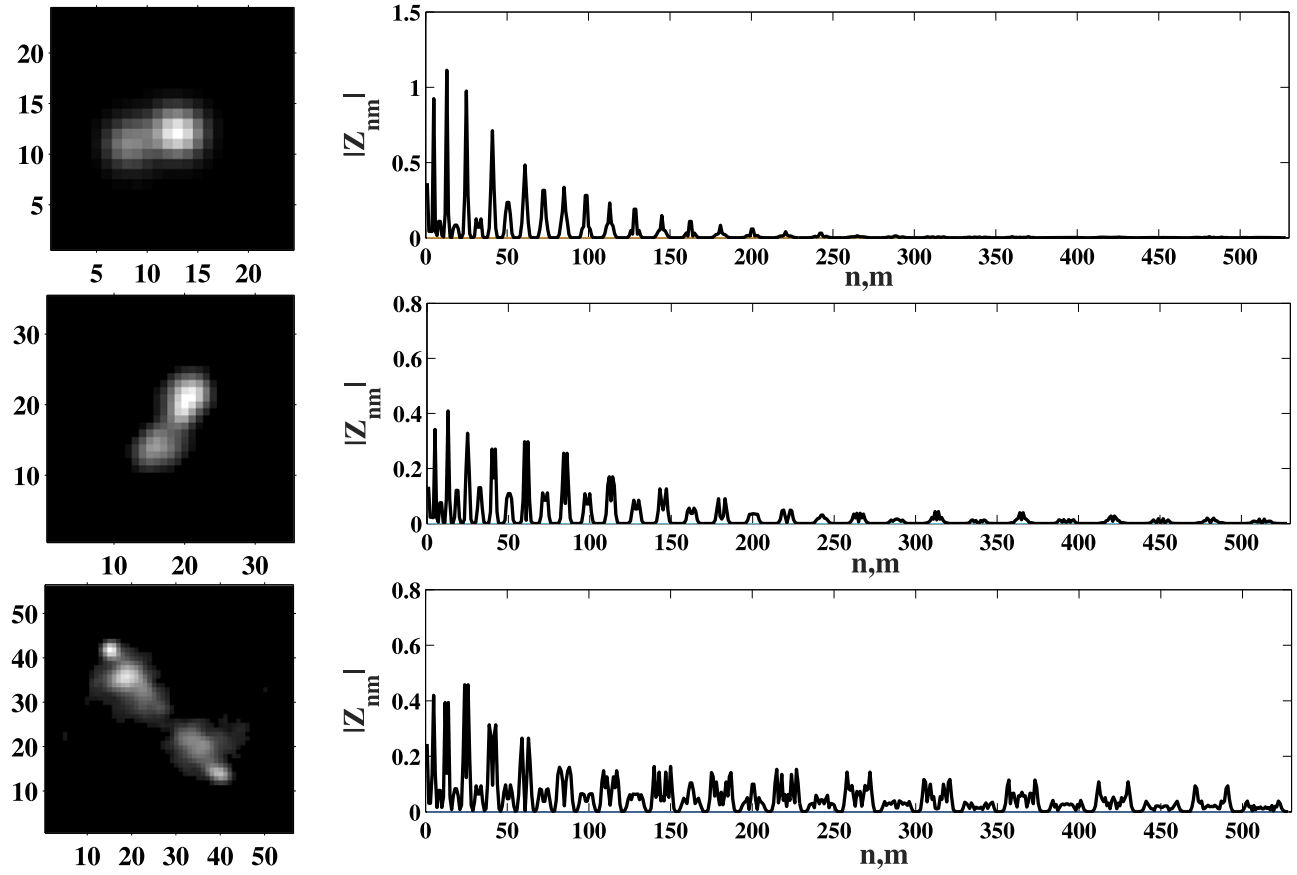


**Figure 5.** Magnitude values of Zernike moments, $Z_{nm}$, of three FRII galaxies.

object inside the unit circle (scale invariant), and then, the centroid of the object is aligned to be set in the center of the unit circle to make the shape translation invariant (Belkasim et al. 2004, 2007). Here, the second approach is automatically applied to the images so that the use of regular moments is not needed. The magnitude values of the Zernike moments of three FRI radio galaxies and three FRII radio galaxies are shown in Figures 4 and 5, respectively. As it is clearly seen, the building blocks of moments (series of different types of Gaussian) are different for the two types of galaxies. These differences make them distinguishable to be easily labeled by classifiers. Choosing a higher-order number of moments may lead to better reconstructed images (Javaherian et al. 2014), but it will be more sensitive to noise (Teh & Chin 1988).

### 3.2. Support Vector Machine (SVM)

Among supervised statistical machine-learning theories, the SVM has received great attention in classifying data. This approach, which was first proposed by Vapnik (1995), has the capability of determining decision boundaries by use of training samples known as the support vectors. The task of this basically binary classifier is to achieve the maximal margin hyperplane between two different classes based on Lagrange multiplier optimization. Since we frequently deal with the problem of classifying data sets with nonlinear decision boundaries, we focused on the formulation of nonlinear SVM. To solve a classification problem by the use of SVM, suppose that we have the training set with $N$ members as $T = \{x_i, y_i | x_i \in R^n, y_i \in \{-1, +1\}\}_{i=1}^{N}$, wherein $x_i$ and $y_i$ are inputs and corresponding outputs (class labels), respectively. Generally, to separate data belonging to two different classes, the SVM searches the following hyperplane,

$$F(\boldsymbol{x}) = \boldsymbol{\omega}^{\top}\Phi(\boldsymbol{x}) + b = 0, \tag{4}$$

where $\boldsymbol{\omega} \in R^n$ is the weight matrix that is perpendicular to the decision boundaries, and $b$ is the parameter of the model related to the separating margin. The function $\Phi(\boldsymbol{x})$ performs mapping $\mathbf{x}$ onto space with a higher dimension. To optimize the values of $\boldsymbol{\omega}$ and $b$, the following primal problem of SV classification must be solved:

$$\begin{aligned} &\min_{\boldsymbol{\omega},\, b,\, \boldsymbol{\xi}} \quad \frac{1}{2}\,\|\boldsymbol{\omega}\|^2 + c\sum_{i=1}^{N}\xi_i, \\ &\text{s. t.} \quad y_i(\boldsymbol{\omega}^{\top}\boldsymbol{\Phi}(x_i) + b) \geqslant 1 - \xi_i, \quad \boldsymbol{\xi} \geqslant 0, \quad i = 1,\cdots, N, \end{aligned} \tag{5}$$

where the regularization parameter $c > 0$ is determined by the user. The variable $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_N)$, which is known as the slack variable, is the error value of the decision boundary for each training set. By maximizing the regularization term $\frac{1}{2}||\boldsymbol{\omega}||^2$, the margin between the two parallel lines separating two data sets increases. This convex optimization problem can be solved by the standard Lagrange multiplier method (e.g., see Tan et al. 2006; Theodoridis & Koutroumbas 2008).

### 3.3. Twin Support Vector Machine (TWSVM)

The task of a TWSVM is to find a pair of nonparallel hyperplanes that show the maximized distance between these two data sets. Suppose that we have two data sets denoted by

matrices $\boldsymbol{X_1} \in R^{p\times n}$ and $\boldsymbol{X_2} \in R^{q\times n}$ and we represent the data points with $p$ and $q$ samples belonging to classes +1 and −1, respectively, in $n$-dimensional real space $R^n$. Each of the nonparallel hyperplanes describe a boundary proximal to the data points in the related class and has its farthest distance from the other class (Shao et al. 2011) as the following form

$$F_1(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\omega}_1^\top \Phi(\boldsymbol{X}_1, \boldsymbol{X}_2) + b_1 = 0,$$
$$F_2(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\omega}_2^\top \Phi(\boldsymbol{X}_1, \boldsymbol{X}_2) + b_2 = 0. \quad (6)$$

In the same manner as nonlinear SVM, $\{\omega_1, \omega_2\} \in R^n$ are the weight matrices indicating normal vectors to the hyperplanes, and $\{b_1, b_2\} \in R$ are the parameters of the model known as the bias term in the nonlinear TWSVM. To find the solution for such a type of separating hyperplanes, the following useful formulation for the problem of nonlinear TWSVM can be employed (Tomar & Agarwal 2015):

$$\min_{\boldsymbol{u}_1, b_1, \varepsilon} \quad \frac{1}{2}||\boldsymbol{u}_1^\top \boldsymbol{K}(\boldsymbol{X}_1^\top, \boldsymbol{D}) + \boldsymbol{e}_1 b_1||^2 + c_1 \boldsymbol{e}_2^\top \varepsilon,$$
$$\text{s. t.} \quad -(\boldsymbol{u}_1^\top \boldsymbol{K}(\boldsymbol{X}_2^\top, \boldsymbol{D}) + \boldsymbol{e}_2 b_1) \geqslant \boldsymbol{e}_2 - \varepsilon, \quad \varepsilon \geqslant 0, \quad (7)$$

and

$$\min_{\boldsymbol{u}_2, b_2, \eta} \quad \frac{1}{2}||\boldsymbol{u}_2^\top \boldsymbol{K}(\boldsymbol{X}_2^\top, \boldsymbol{D}) + \boldsymbol{e}_2 b_2||^2 + c_2 \boldsymbol{e}_1^\top \boldsymbol{\eta},$$
$$\text{s. t.} \quad -(\boldsymbol{u}_2^\top \boldsymbol{K}(\boldsymbol{X}_1^\top, \boldsymbol{D}) + \boldsymbol{e}_1 b_2) \geqslant \boldsymbol{e}_1 - \boldsymbol{\eta}, \quad \boldsymbol{\eta} \geqslant 0, \quad (8)$$

where $\boldsymbol{D} = [\boldsymbol{X}_1, \boldsymbol{X}_2]^\top$, $\boldsymbol{u}_1 = \begin{bmatrix} \boldsymbol{\omega}_1 \\ b_1 \end{bmatrix}$, and $\boldsymbol{u}_2 = \begin{bmatrix} \boldsymbol{\omega}_2 \\ b_2 \end{bmatrix}$. The parameters $\varepsilon$ and $\eta$ are the slack variables, and the regularization parameters are denoted as $c_1$ and $c_2$. The two vectors $e_1$ and $e_2$ are all-ones vectors with the size of the data number in the positive and negative classes, respectively. The kernel function **K** is a surface that separates data sets in high dimensional space. The famous kernel functions are linear, polynomial, and radial basis functions (RBF). The RBF value depends only on the magnitude of distance (Euclidean or other metrics) between data points and the origin (Buhmann 2003). The above-mentioned minimization problems are solvable by the standard Lagrange multiplier method. For details about the method and solution, the reader can refer to reviews presented by Tian & Qi (2014) and Tomar & Agarwal (2015).

### 3.4. Genetic Algorithm (GA)

The GA, which was proposed by Holland (1975) in the early 1970s, was invented to mimic some of the processes observed in the natural evolution of living organisms (Holland 1975). The main idea of the GA is to use the power of the evolutionary process to solve optimization problems. In other words, the GA is an algorithm that uses genetic evolution as a problem-solving model. The GA is very effective when dealing with very large data sets, improving the classification performance over alternative methods. The task of this algorithm is to find efficient solutions (chromosomes) for a problem (population). The GA randomly starts with a population of chromosomes or individuals and then a fitness function is employed. Actually, coding the evolutionary pathways is fulfilled by the fitness function based on a pattern. The fitness function can be a function of statistical parameters (mean and/or variance) of generations, or it can be defined as the objective function that must be optimized. This function evaluates the suitability of each chromosome for survival. In fact, the chromosomes that represent the optimal solution of a problem are reproduced through a repetitive process of some selection operators such as mutation, crossover, recombination, and inversion (Ujjwal et al. 2011).

## 4. Results and Discussions

In this work, we first use the Zernike functions to extract the image invariant moments that can be converted to the original image. The GA is then used to optimize the length of the series of Zernike moments based on the statistical parameters (maximum, minimum, mean, variance, and skewness) of the series of moments. A training set of the selected moments is provided for use in the SVM and TWSVM. The GA is reused to select the optimal parameters of classifiers. The GA minimizes the objective function of differences between the label of outputs (determined by classifiers) and the original label of data. The classifiers are ready to accurately predict the label of each selected Zernike moment in the test set. The SVM and TWSVM codes used in this paper can be found in the SVM-KM MATLAB toolbox and the work done by Jayadeva et al. (2007), respectively.

As mentioned in Section 2, there are 856 FRI and FRII galaxies, some of which have complex structures and so are excluded from the analysis. From the two groups of which 349 galaxies have a label of FRI, 786 galaxies are selected and the remaining 437 are labeled as FRII. About 65% of the total data (i.e. 511 radio galaxies), which was randomly chosen by a random generator algorithm (Press et al. 2007), were introduced as a training data set for the SVM and TWSVM. This includes 232 radio galaxies from class one (FRI) and 279 radio galaxies from class two (FRII). Then, the remaining 35% of the total data consisting of 275 FRI and FRII galaxies were fed to the classifiers as a test data set.

The precision, recall, F1 scores, and accuracy are the most important performance measures that are employed to present the confidence level of classification. By considering the FRI data as the positive class and the FRII data as the negative class, the precision refers to the fraction of FRI data that are correctly classified as the positive class (true positives) to all FRI data (Equation (9)), while the recall is the fraction of true positives to the sum of true positives and the FRII data that are classified incorrectly as the negative class (false negatives). This metric also indicates the sensitivity of the classifier to accurate identification (Equation (10)). The F1 score is the harmonic mean of precision and recall that can be interpreted as the average precision and recall values (Equation (11)). The accuracy is the total proportion of correct predictions (Equation (12)). These accuracy metrics are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (10)$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

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

**Table 1**
Percentage Accuracy of Predicting Test Data

| Method (kernel) | Precision | Recall | F1 score | Accuracy |
|---|---|---|---|---|
| SVM (RBF) | 0.9583 | 0.9746 | 0.9664 | 0.9710 |
| SVM (Poly) | 0.9360 | 0.9915 | 0.9630 | 0.9674 |
| TWSVM (RBF) | 0.9583 | 0.9829 | 0.9705 | 0.9745 |
| TWSVM (Poly) | 0.8468 | 0.8974 | 0.8714 | 0.8872 |
| TWSVM (Linear) | 0.9636 | 0.9060 | 0.9339 | 0.9454 |

where TP refers to the true positives, FP refers to the false positives, and FN refers to false negatives. A true positive is when the prediction matches the label. A false positive is when the positive class is incorrectly predicted. A false negative is when the positive class is predicted to be in another class. In Table 1, the results of applying the SVM and TWSVM with different kernel functions (i.e. linear, polynomial, and RBF) are shown for the test data.

As we see in Table 1, almost all accuracy metrics take values of more than 85%. However, the best performances of classifiers occur when the RBF is used as a kernel in the SVM or TWSVM. Our results suggest that the RBF operates effectively and has more flexibility for discriminating data clusters.

Finally, we intend to make a brief discussion about the strengths of our method when compared with other proposed methods. Our multilevel method involves the following algorithms: (i) applying the Zernike function to each image gives a series of moments that are invariant regarding rotation, scaling, and translation. Therefore, it is not necessary to know the information about the positions of host galaxies in the images. The other characteristic of these series is that they can be fed to every type of supervised classifiers for labeling. This enables us to easily change and/or promote the classifier based on the data if needed. (ii) By using the GA for both optimizing the length of the moments series and estimating the parameters of classifiers over an iteration process, the algorithmic efficiency increases. Moreover, there is no need to manually change these parameters when dealing with another type of image data. Therefore, this approach provides a fully automated method with the least intervention by the user. (iii) Among the supervised classifier introduced in the pattern recognition field, the SVM-type classifiers are accepted as powerful tools for classifying data in so many fields (e.g., see Ma & Guo 2014) because they find the maximum margin hyperplanes between two groups of data sets. We employed both the SVM and TWSVM because of their high accuracy, ability in fast classification, and compatibility with other types of data.

## 5. Summary and Conclusions

This paper explored two supervised machine-learning approaches (SVM and TWSVM) for the morphological classification of radio galaxies presented in Miraghaei & Best (2017) into two Fanaroff–Riley classes: FRI and FRII. The rotation-, translation-, and scale-invariant image moments were extracted by employing the Zernike polynomials, and the lengths of the series were optimized by the GA. The classifiers were trained by the selected Zernike moments of 65% of the total number of radio images in each class. Also, the GA was reused to optimize the parameters of classifiers for finding the maximum margin between two types of data so that they minimized the classification error. It seems that the structure of building blocks of the magnitude values of moments (i.e. amplitudes, shape parameters, the distances between peaks, and the lengths of tails) makes a sufficient difference for the classifiers to discriminate classes. To determine the confidence level of classification, four popular and approved tests were used. The results show that the TWSVM with the RBF as a kernel with a correct recognition (more than 95%) has the best output in radio galaxy classification. Therefore, we provide a novel method with high accuracy for data classification. The method involves more convenient coding steps compared to the previous methods that have used multilayer deep learners. The most significant advantages of this method are the use of fully automated steps with the ability of extension to multicategory classification and its compatibility with other types of image data in various fields. This method may provide a useful tool for classifying radio galaxies in future SKA and LOFAR surveys alongside other proposed approaches.

The authors thank the anonymous referee for the very helpful comments and suggestions.

## ORCID iDs

Mohammad Sadeghi https://orcid.org/0000-0002-6409-7679
Mohsen Javaherian https://orcid.org/0000-0003-0464-1561
Halime Miraghaei https://orcid.org/0000-0003-1829-9402

## References

Abandah, G., & Anssari, N. 2009, Journal of Computer Science, 5, 226
Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Alipour, N., Mohammadi, F., & Safari, H. 2019, ApJS, 243, 20
Alipour, N., Safari, H., & Innes, D. E. 2012, ApJ, 746, 12
Aniyan, A. K., & Thorat, K. 2017, ApJS, 230, 20
Arish, S., Javaherian, M., Safari, H., & Amiri, A. 2016, SoPh, 291, 1209
Baldi, R. D., Capetti, A., & Giovannini, G. 2015, A&A, 576, A38
Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
Baron, D., & Poznanski, D. 2017, MNRAS, 465, 4530
Bastien, D., Oozeer, N., & Somanah, R. 2017, MS&E, 198, 012013
Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
Belkasim, S., Hassan, E., & Obeidi, T. 2004, in CIT '04 The Fourth Int. Conf. Computer and Information Technology, ed. D. Wei, H. Wang, Z. Peng et al. (Piscataway, NJ: IEEE), 790
Belkasim, S., Hassan, E., & Obeidi, T. 2007, PaReL, 28, 1969
Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38
Best, P. N. 2009, AN, 330, 184
Best, P. N., & Heckman, T. M. 2012, MNRAS, 421, 1569
Best, P. N., Kauffmann, G., Heckman, T. M., & Ivezić, Ž. 2005, MNRAS, 362, 9
Buhmann, M. D. 2003, Radial Basis Functions: Theory and Implementations (Cambridge, New York: Cambridge Univ. Press)
Burns, J. O. 1981, MNRAS, 195, 523
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Dabhade, P., Mahato, M., Bagchi, J., et al. 2020, A&A, 642, A153
Fanaroff, B. L., & Riley, J. M. 1974, MNRAS, 167, 31P
Gopal-Krishna, & Wiita, P. J. 2000, A&A, 363, 507
Harwood, J. J., Vernstrom, T., & Stroe, A. 2020, MNRAS, 491, 803
Holland, J. H. 1975, Adaptation in Natural and Artificial Systems (Ann Arbor, MI: Univ. Michigan Press)
Hosny, K. M. 2010, Information Sciences, 180, 2299
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167
Jarvis, M., Taylor, R., Agudo, I., et al. 2016, arXiv:1709.01901
Javaherian, M., Safari, H., Amiri, A., & Ziaei, S. 2014, SoPh, 289, 3969
Jayadeva, Jayadeva, R., & Chandra, S. 2007, ITPAM, 29, 905
Khotanzad, A., & Hong, Y. H. 1990, ITPAM, 12, 489

Kuntzer, T., & Courbin, F. 2017, A&A, 606, A119
Laing, R. A., Guidetti, D., Bridle, A. H., Parma, P., & Bondi, M. 2011, MNRAS, 417, 2789
Lukic, V., Brüggen, M., Banfield, J. K., et al. 2018, MNRAS, 476, 246
Lukic, V., Brüggen, M., Mingo, B., et al. 2019, MNRAS, 487, 1729
Ma, Y., & Guo, G. 2014, Support Vector Machines Applications (Cham: Springer)
Ma, Z., Xu, H., Zhu, J., et al. 2019, ApJS, 240, 34
McNamara, B. R., & Nulsen, P. E. J. 2007, ARA&A, 45, 117
Miley, G. K., Perola, G. C., van der Kruit, P. C., & van der Laan, H. 1972, Natur, 237, 269
Ming-Kuei, H. 1962, IEEE Trans Inf Theory, 8, 179
Mingo, B., Croston, J. H., Hardcastle, M. J., et al. 2019, MNRAS, 488, 2701
Miraghaei, H., & Best, P. N. 2017, MNRAS, 466, 4346
O'Dea, C. P. 1998, PASP, 110, 493
O'Dea, C. P., & Baum, S. A. 1997, AJ, 113, 148
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (3rd ed.; Cambridge: Cambridge Univ. Press)
Ralph, N. O., Norris, R. P., Fang, G., et al. 2019, PASP, 131, 108011
Saripalli, L., Gopal-Krishna, Reich, W., & Kuehr, H. 1986, A&A, 170, 20
Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94
Shao, Y., Zhang, C., Wang, X., & Deng, N. 2011, IEEE Trans Neural Netw, 22, 962
Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, A104
Smolčić, V., Schinnerer, E., Finoguenov, A., et al. 2007, ApJS, 172, 295
Tan, P., Steinbach, M., & Kumar, V. 2006, Introduction to Data Mining, Always Learning (Boston, MA: Pearson Addison Wesley)
Tang, H., Scaife, A. M. M., & Leahy, J. P. 2019, MNRAS, 488, 3358
Teh, C., & Chin, R. T. 1988, ITPAM, 10, 496
Theodoridis, S., & Koutroumbas, K. 2008, Pattern Recognition (4th ed.; New York: Academic), 119
Tian, Y., & Qi, Z. 2014, Annals of Data Science, 1, 253
Tomar, D., & Agarwal, S. 2015, Egyptian Informatics Journal, 16, 55
Ujjwal, M., Sanghamitra, B., & Anirban, M. 2011, Multiobjective Genetic Algorithms for Clustering (Berlin: Springer)
Vapnik, V. N. 1995, The Nature of Statistical Learning Theory (Berlin: Springer)
Wayth, R. B., Lenc, E., Bell, M. E., et al. 2015, PASA, 32, e025
York, D. G., Adelman, J., John, E. J., et al. 2000, AJ, 120, 1579
Zevin, M., Coughlin, S., Bahaadini, S., et al. 2017, CQGra, 34, 064003