Modularized Filter-Wrapper Feature Selection Scheme for Finding Optimal Image Moments in Maps of Radio Galaxies

Seyed Alireza Bashiri Mosavi[1]*   Mohsen Javaherian[2]
Mohammad Sadeghi[3]   Halime Miraghaei[2]

*[1]Imam Khomeini International University- Buin Zahra Higher Education Center of Engineering and Technology,
Iran*
*[2]Research Institute for Astronomy and Astrophysics of Maragha (RIAAM),
University of Maragheh, 55136-553, Maragheh, Iran*
[3]Department of Physics, University of Kurdistan, Pasdaran Street, P.O. Box 66177-15175, Sanandaj, Iran
* Corresponding author's Email: a.bashiri@bzeng.ikiu.ac.ir

Abstract: The increases in releasing high-dimensional radio galaxy images necessitate addressing automatic methods for compacting the curse of dimensionality. Hence, in this paper, the modularized filter-wrapper feature selection scheme (MFWFS) is exploited to select the most discriminative features (MDFs) of the galaxy images. First, we employed the Fanaroff–Riley (FR) radio galaxy images in the data-gathering phase. Next, we applied the MFWFS scheme to 528 moments of radio images for selecting MDFs in dual-phases. The preliminary optimal features (POFs) are selected in the filter phase concerning the triple information theory criteria. In the wrapper phase, the obtained POFs are fed to the twin support vector machine (TWSVM) classifier to extract MDFs. Finally, the quad-final MDFs are introduced to the experimental comparison strategy through the cross-validation technique. Our results show that the best-laid MDFs in a subset of 96 features have accuracy close to 80% with a dimension reduction rate of 5.5.

Keywords: Radio galaxies, Catalog, Twin support vector machine, Feature subset selection.

1. Introduction

Observing galaxies in radio bands provides valuable insights into active galactic nuclei (AGN) studies. It significantly helps to understand the properties of AGN, their environment, the supermassive black holes, as well dark matter halos, and energy feedback into the intergalactic medium [1]. It also can describe evolving steps of galaxy transition from blue star-forming to red quenched ones [2]. Since radio waves are not effectively scattered by galactic dust, investigating radio galaxies helps to understand the underlying process of the high redshift universe and gives an appropriate explanation about their future behavior. One of the significant physical parameters of radio galaxies is their shapes consisting of an extended range of sizes from compact to mega-parsec scale sources [3].

Fanaroff and Riley (FR) [4] proposed a seminal classification for two populations of extended radio galaxies: type I and type II [4]. In the FRI galaxies (called edge-darkened), the emission profile decreases from the intensity centroid of the galaxy to the edge. Gradual diffusing of galaxy brightness with approaching the edge reveals the gentle downward slope in the tail of the profile. In the FRII galaxies (called edge-brightened), the emission profile increases by moving away from the intensity centroid of the galaxy so that the brightness of the galaxy reaches its peak near the edge. With the emergence of high-resolution instruments and accurate techniques for detecting detailed information of galaxies, the new types of populations were added to the classification list of

galaxies [5]. One of the new type of categories is hybrid morphology radio sources (HyMoRS) [6].

In the age of data, receiving a vast amount of data with different dimensions in each branch of science necessitates automatic techniques for feature subset selection (FSS). In astronomy, so many classification methods were developed to predict the class of received data sets (for radio galaxy classification, refer to [7] and references therein). But, to reduce the curse of dimensionality leading to increase both the efficiency of the algorithm and the flexibility of the classification method, it seems that the FSS algorithm must be developed and used in the interface of received data and classification approaches (See Section 2). To do so, we propose an efficient method based on the TWSVM classifier to reduce the length of moments series (the curse of dimensionality) and categorize FRI and FRII radio images. The series features are selected in the procedure of the modularized filter-wrapper feature selection (MFWFS) scheme. Then, the most discriminative features (MDFs) with different lengths are grouped in the iterative process so that the TWSVM finds the optimal range for moments with higher accuracy and more reserved memory in the shortest possible computational time.

The outline of this paper is as follows: the problem definition is presented in Section 2. The employed data sets are described in Section 3. A comprehensive review of methods is described in Section 4. The obtained results are presented and discussed in Section 5. Also, the comparison results between our method and the other FSS techniques are depicted in Section 6. Concluding remarks are interpreted in Section 7.

2. Defining a problem statement

The FSS has been developed to use in many fields of science for reducing the curse of dimensionality. For astronomical targets, extracting relevant features against high-dimensional radio galaxy data based on FSS approaches has not been addressed in articles. Moreover, reviewing the FSS literature shows that FSS focused on a whole-manner strategy on feature space based on filter methods for selecting discriminative features. For example, in [8], the most relevant features based on filter FSS (Relief) have been concerned with selecting optimal features. The ReliefF FSS to select optimal features has been explained in [9]. The min-redundancy and max-relevance (mRMR) FSS were applied on the set of high-dimensional features [10]. In [11], the FCBF as feature pre-screening has been employed for finding relevant features. However, in

the previous studies in the field of FSS [8-11], statistical and machine learning-based techniques (filter and wrapper method) are applied cohesively (on the whole of feature space) caused to loss of many useful features. In fact, selecting the most relevant features by conducting the partial-manner policy based on the coupling filter and wrapper approach leads to the survival of the optimal-blurred features. For the first time, we proposed an automatic supervised method based on modularized hybrid FSS for applying to the moments series of radio galaxies. This efficient method can be adjusted to use as a data reduction procedure step before classifying any type of image data.

3. Description of the radio galaxy sample

We have used the latest and the largest catalog of FR radio galaxies identified based on the cross-matching the seventh data release (DR7) of the Sloan Digital Sky Survey (SDSS) with the National Radio Astronomy Observatory (NRAO) Very Large Array Sky Survey (NVSS) and the Faint Images of the Radio Sky at Twenty centimeters (FIRST) survey published in [12]. The catalog is an outcome of series of works which first introduced by Reference [13]. They present an algorithm to find out single- and multi-component radio sources with their optical counterparts. The technique has been applied to the SDSS DR7 resulted in a catalog of 18286 galaxies having radio emission by Reference [14]. Among them, galaxies with multi-component radio structures have been visually inspected to construct a catalog of 1329 morphologically classified radio galaxies [12]. The morphologies include FRI, FRII, and FR hybrid as well a few extended sources remain unclassified. The sample of FR hybrid is excluded from the analysis due to the small sample size. In this study, we used $5' \times 5'$ fits images of radio galaxies from the FIRST cutout server. A higher resolution of the FIRST survey ($\theta \sim 5''$) compared to the NVSS ($\theta \sim 45''$) enables us to extract more details of the radio images. The pixel size of each image is $\sim 1''.8$ with a typical noise level of about 0.15 mJy.

4. Method

We attempted to employ different techniques in multi-levels with a powerful mathematical foundation to have robust results for supervised classification. Our proposed algorithm involves the procedures of converting radio images to the Zernike moments, reducing the moments series in the wrapper phase, and both selecting the most discriminative features and grouping the survived

features into two classes by using the twin SVM classifier with the kernel RBF. We explained all these procedures and their steps in detail as follows.

4.1 Zernike moments (ZMs)

ZMs are described in the polar coordinates that the basic sets of these moments are complex and orthogonal. Orthogonality implies that there is no excess information, and the moments do not overlap with each other. Thus, each moment is unique according to its order [15]. Moreover, the image reconstruction is computationally simple because of the orthogonality property of the Zernike polynomials [16]. This method is one of the most common ways of obtaining image moments. If we have an ordered pair that represents the polynomial order and the phase angle coefficient, we can define the Zernike function as follows:

$$V_{pq}(r,\theta) = R_{pq}(r)\, exp(im\theta) \qquad (1)$$

$$R_{pq}(r) = \sum_{s=0}^{(p-q)/2} (-1)^s \frac{(p-s)!}{s!\,(\frac{p+|q|}{2}-s)!\,(\frac{(p-|q|)}{2}-s)!} r^{p-2s}$$

Where s is a numerator, p is the positive integer order number, and q is the repetition integer number. These two parameters have the constraints: $|q| \leq n$, $p-|q|$ is even [16]. Since the Zernike function is defined in the polar coordinates, it is described by two components r and θ. So, we must map pixels of the image 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 [15].

$$I(r,\theta) = \sum_{p=0}^{p_{up}} \sum_{q=0}^{p} Z_{pq} V_{pq}(r,\theta) \qquad (2)$$

Thus, we can obtain the ZMs as follows:

$$Z_{pq} = \frac{p+1}{\pi} \int \int_{unit\ circle} V_{pq}^*(r,\theta) I(r,\theta) r dr d\theta \quad (3)$$

For a regular pair (p, q), the number of ZMs $[\sum_{p=0}^{p=p_{up}}(p+1)]$ is 528 features. It was shown that the best order of the Zernike is $p_{up}= 31$, which has the least reconstruction error [16]. The reconstructed image is very similar with the original one [7, 16]. ZMs give general and detailed properties of a shape from an image. The low-order moments are suitable for obtaining general shape properties, and the

higher-order moments provide partial shape properties [17]. Selecting higher-order images may lead to better reconstructed images but will cause noise in the curves obtained from ZMs [18].

The nature of the exponential Fourier term [exp(imθ)] in the complex Zernike polynomials makes it rotation invariant when the magnitude of moments series $|Z_{pq}|$ are computed. It means that the $|Z_{pq}|$ is invariant for image rotated in different angles. Two approaches were proposed to make an image scale and translation invariant. In the first proposed approach, the image moments are normalized to the geometrical moments. It was verified that the reconstructed image does not loss the quality of original image [16]. In the second approach, the image is resized to surround the object in a square. Then, the scale invariant process is fulfilled by mapping square into a unit circle. For making the image translation invariant, the center of the circle must be corresponded with the centroid of the object [19]. Fig. 1 and Fig. 2 show the magnitude values of the ZMs with order $p_{up} = 31$ extracted from both three original FRI and FRII radio images, respectively. As it is seen in Fig. 1 and Fig. 2, the $|Z_{pq}|$ with different characteristics of building blocks such as amplitudes and shape parameters can make each one unique to segregate by the classifier.

4.2 Filter phase in MFWFS scheme

Based on information theory, the triple criteria of relevancy, interdependency, and redundancy were concerned for selecting preliminary optimal features (POFs) of ZMs in the MFWFS scheme. The mutual information (*MI*) and entropy are the fundamental factors defined in the following form:

$$MI(X;Y) = H(X) - H(X|Y) \qquad (4)$$

Let *X* be a discrete random variable and probability density function $p(x) = \Pr\{X = x\}$.

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \qquad (5)$$

According to Eqs. (4) and (5), the triple analysis indexes of symmetric uncertainty (*SU*), interdependence (*I*), and redundancy (*R*) were exploited in the filter phase calculations for selecting POFs. The *SU* index evaluates the amount of information shared by the two variables:

$$SU_{i,c}(F_i, C) = 2 \frac{MI(F_i;C)}{H(F_i)+H(C)} \qquad (6)$$

Where F_i is the ith feature and C is representative of the class. Indexes I and R can be formulated in a single form called IR:

$$IR(i,j) = 2\frac{MI(F_i; C|F_j) - MI(F_i; C)}{H(F_i) + H(C)};$$
$$-1 \leq IR(i,j) \leq 1 \qquad (7)$$

Where F_j is the jth fold. Actually, i and j are numerators set in a loop. The detailed description of the filter phase in the MFWFS scheme is articulated in the following five steps:

Step1) The value of SU is computed for each feature (Eq. (6)).

Step2) The feature with the highest SU is selected as the first POFs (*f*-POFs) and then removed from the features set.

Step3) The IR value (Eq. (7)) of the rest of the features is calculated the first iteration regarding the *f*-POFs called $^1IR_{f-POFs}^{rest}$. The $^1IR_{f-POFs}^{rest}$ is multiplied to SU^{rest}. Then, we select the maximum value of the obtained results as the second POFs (*s*-POFs).

Step4) In the second iteration, the *s*-POFs are removed from the features set. The value of $^2IR_{s-POFs}^{rest}$ is calculated and then multiplied to SU^{rest}.

Step5) The filter phase steps (1-4) continue k iterations. Among the initial features set, the survived POFs in the kth iteration are selected as inputs to participate in the wrapper phase procedure equipped by the TWSVM-RBF.
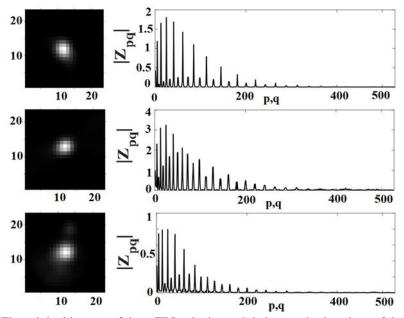

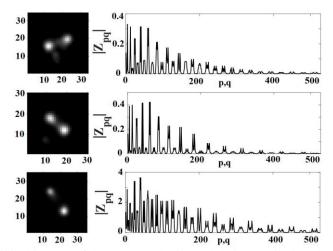Figure. 1 The original images of three FRI galaxies and their magnitude values of the ZMs, Z_{pq}

Figure. 2 The original images of three FRII galaxies and their magnitude values of the ZMs, Z_{pq}

4.3 TWSVM: wrapper phase and classification method

Vapnik (1995) proposed a supervised method to classify data sets by maximizing the margin between different classes [20]. In this basically binary classifier, the parallel decision boundaries are determined by training samples called support vectors. The support vector machine (SVM) computes the maximal margin hyperplane between two various groups of data using Lagrange multiplier optimization. In the updated version of the SVM which twin support vector machine (TWSVM) [21], the classifier utilizes the concept of SVM to find nonparallel decision boundaries (instead of parallel hyperplane) for each class of data sets. In this approach, a pair of equations for each hyperplane is solved to maximize the distance between two decision boundaries. We consider the general case that the data sets are nonlinearly separable. For training set with N numbers, we have $T = \{x_i, y_i | x_i \in R^n, y_i \in \{-1, +1\}\}_{i=1}^N$, wherein x_i is input arrays and y_i is corresponding class labels (outputs). Suppose that we have two classes of data in n-dimensional real space R^n: class 1 denoted by matrix $x_1 (\in R^{r \times n})$ includes r samples with label +1, and class 2 denoted by $x_2 (\in R^{s \times n})$ includes s samples with label -1. The TWSVM searches the solutions for two nonparallel boundaries that each one has the nearest distance to the data points of one class while farthest from the other class. The equations of two hyperplanes with their normal vectors ω_1 and ω_2 are expressed in the matrix form as follows:

$$\omega_1^\top \Theta(x_1, x_2) + b_1 = 0, \quad \omega_2^T \Theta(x_1, x_2) + b_2 = 0 \quad (8)$$

Where b_1 and b_2 are the bias terms in R space. The function $\Theta(x_1, x_2)$ maps data samples onto space with a higher dimension to make data sets separable. The upper index $\top$ denotes the transpose of the matrix. To find the optimum solution for the sets of $\{\omega_1, \omega_2\}$ and $\{b_1, b_2\}$ in the nonlinear TWSVM, the following formulation is introduced:

$$min_{v_1, b_1, \alpha} \frac{1}{2} ||v_1^\top K(x_1^\top, u) + e_1 b_1||^2 + c_1 e_2^\top \alpha \quad (9)$$
$$s.t. \quad -(v_1^\top K(x_2^\top, u) + e_2 b_1) \geq e_2 - \alpha, \quad \alpha \geq 0$$

and

$$min_{v_2, b_2, \beta} \frac{1}{2} ||v_2^\top K(x_2^\top, u) + e_2 b_2||^2 + c_2 e_1^\top \beta \quad (10)$$
$$s.t. \quad -(v_2^\top K(x_1^\top, u) + e_1 b_2) \geq e_1 - \beta, \quad \beta \geq 0$$

Where

$$u = [x_1, x_2]^\top, \quad v_1 = \begin{bmatrix} \omega_1 \\ b_1 \end{bmatrix}, \text{ and } v_2 = \begin{bmatrix} \omega_2 \\ b_2 \end{bmatrix}.$$

The sets $\{c_1, c_2\}$ and $\{\alpha, \beta\}$ are regularization parameters and slack variables, respectively. The all-ones vectors e_1 and e_2 are the same sizes as data numbers in the positive and negative classes, respectively. Instead of using the function Θ, the nonlinear TWSVM uses the Kernel function, K to separate data points in higher-dimensional space. We employed the Gaussian radial basis (RBF) function kernel K (x, x') as the following form:

$$K(x, x') = \exp(-\frac{||x - x'||^2}{2\sigma^2}) \quad (11)$$

Where $||x - x'||^2$ is squared Euclidean distance between the two data objects. The parameter σ is the standard deviation of data points. Using the standard Lagrange multiplier method is the common way to solve this kind of optimization problem (The detailed explanation about the solution can be found in [22]).

4.4 MFWFS scheme

Here, the proposed methodology for selecting the MDFs of the ZMs was concerned in the form of the MFWFS scheme. As it can be seen in Fig. 3, the MFWFS consists of the triple parts: (i) dividing moments series into segments (i.e., 22 subsets), (ii) applying filter phase to each segment via 4-fold cross-validation technique for extracting POFs of the ZMs, and (iii) using the TWSVM-RBF classifier for evaluating the performance of the selected POFs for each fold to obtain the final optimal features of ZMs. The last step of the algorithm is the wrapper phase which gives the arrays of MDFs. Generally, this procedure can be expressed in the following steps:

Step1) The radio galaxy data sets (array with 726×528 elements) were modularized in both dimensions into 22 subsets (called $G - segment^i$). The size of each $G - segment^i$ is obtained by:

$$G - segment_{[33 \times 24]}^i = Modularization(D) \quad (12)$$
$$D = [ZMs]_{samples(726) \times features(528)},$$
$$i = 1, 2, \dots, 22$$

Step2) The filter phase was applied to i^{th} $G - segment_{[33 \times 24]}$. For precise mining on the ZMs feature space, each $G - segment_{[33 \times 24]}^i$ was folded
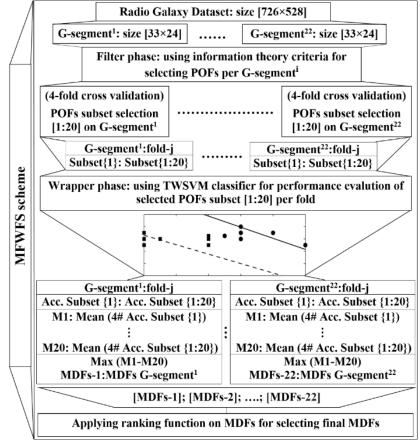
Figure. 3 MFWFS scheme for selecting MDFs of ZMs

based on the cross-validation technique (4-fold cross-validation). To determine the maximum number of the selected POFs in each segment, we define the *MaxSelFea* set to be 20. In fact, after conducting the filter phase, for 4 folds in each segment, we have sets of 1 to 20 POFs, namely POFs subset {1}, subset {1, 2}, …, subset {1:20}:

$$[POFs\{1\#\}, POFs\{2\#\}, …, POFs\{20\#\}]_j^i = \\ filter\ phase\left({}^{fold^j}segment_{[33\times24]}^i\right) \quad (13) \\ i = 1,2,…,22\ and\ j = 1,…,4.$$

Step3) After constructing array of POFs, the TWSVM-RBF is used for evaluating the performance of the selected POFs. First, the subsets {1#}, {2#}, …, {20#} is fed to the TWSVM-RBF for training procedure as follows:

$$TWSVM\ hyperplane_j^i \\ = \{TWSVM^{RBF}[train\ set(POFs\{1\#\})]_j^i, …, \\ TWSVM^{RBF}[train\ set(POFs\{20\#\})]_j^i\} \quad (14)$$

then, the testing procedure can be expressed as:

$$[Acc(POFs\{1\#\}), …, Acc(POFs\{20\#\})]_j^i = \\ TWSVM\ hyperplane[test\ set(POFs\{1\#\})]_j^i, …, \\ TWSVM\ hyperplane[test\ set(POFs\{20\#\})]_j^i\}$$
$$(15)$$

Step4) The MDFs are obtained for each segment:

$$^iMDFs^{Acc\ \&\ subset\{b\#\}} \\ = max_{b\#:[^{Acc}Mean_{1\#}^i||^{Acc}Mean_{2\#}^i||\ …||^{Acc}Mean_{20\#}^i]}^{Acc}(P)$$

$$P = \begin{pmatrix} ^{Acc}Mean_{1\#}^i = \frac{1}{4}\sum_{j=1}^4[Acc(POFs\{1\#\})]_j^i \\ \vdots \\ ^{Acc}Mean_{20\#}^i = \frac{1}{4}\sum_{j=1}^4[Acc(POFs\{20\#\})]_j^i \end{pmatrix}$$
$$(16)$$

Step5) The rank of $^iMDFs^{Acc\ \&\ subset\{b\#\}}$ is computed per segment to select the final MDFs (Fig. 4). In Table 1, the reader can find the maximum accuracy of each segment obtained among subsets {{1#}, …, {20#}}. According to Step 5, there are two types of ranking scenarios to select the optimum
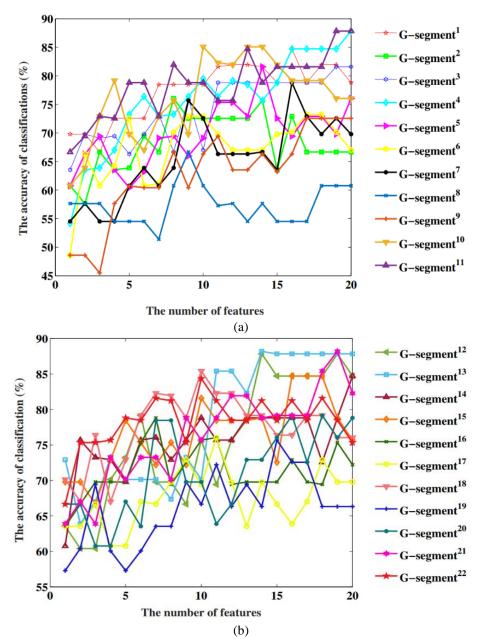
(a)

(b)

Figure. 4 Mean of accuracies vs. different numbers of selected subset (1:20 features) on G-segments: (a) G-segment[1–11] (upper panel) and (b)G-segment[12–22] (lower panel)

set of MDFs rooted in the accuracy level (AL) and the ratio of the accuracy to the number of features (RAN).

• Ranking method (1): the summation of the four highest AL regarding the lower number of features and the three highest RAN regarding AL is higher than 80% (Table 2).

• Ranking method (2): the RAN is higher than 5 (Table 3).

• Ranking method (3): the AL is higher than 80% (Table 4).

• Ranking method (4): the AL is higher than 75% (Table 5).

Using four ranking methods, the quad-final MDFs for the number of 528 moments are obtained to be 96, 154, 201, and 274, respectively, to classify radio galaxy images in the proposed framework.

5. Results and discussion

Receiving a vast amount of radio galaxy images from different telescopes motivated us to provide an optimal approach to categorize data sets. We used the number of 726 cutouts FIRST images with a resolution of 5″. Among the samples of FRI and FRII radio galaxies, the number of 330 FRI images

Table 1. MDFs−#: The Acc ^{max} [Mean^i $_{\{1\#\}}$, …, Mean^i $_{\{20\#\}}$] of each segment^i for subsets [{1#}, …, {20#}]

Segments #	MDF-#: Max. Acc. (%) – subset {#}	Segments #	MDFs-#: Max. Acc. (%) - subset {#}
Segment 1	MDFs-1: 84.72 - subset {16#}	Segment 12	MDFs-12: 87.84 - subset {14#}
Segment 2	MDFs-2: 76.04 - subset {8#}	Segment 13	MDFs-13: 88.19 - subset {14#}
Segment 3	MDFs-3: 81.59 -subset {19#}	Segment 14	MDFs-14: 84.72 - subset {20#}
Segment 4	MDFs-4: 87.84 - subset {20#}	Segment 15	MDFs-15: 84.72 - subset {16#}
Segment 5	MDFs-5: 81.59 - subset {14#}	Segment 16	MDFs-16: 78.71 - subset {7#}
Segment 6	MDFs-6: 73.26 - subset {17#}	Segment 17	MDFs-17: 76.04 - subset {11#}
Segment 7	MDFs-7: 78.81 - subset {16#}	Segment 18	MDFs-18: 85.41 - subset {10#}
Segment 8	MDFs-8: 66.66 - subset {9#}	Segment 19	MDFs-19: 75.69 - subset {15#}
Segment 9	MDFs-9: 72.56 - subset {17#}	Segment 20	MDFs-20: 79.16 - subset {16#}
Segment 10	MDFs-10: 85.06 - subset {10#}	Segment 21	MDFs-21: 88.19 - subset {19#}
Segment 11	MDFs-11: 87.84 - subset {19#}	Segment 22	MDFs-22: 84.37 - subset {10#}

Table 2. Obtained ZMs based on ranking method (1)

Segments #	ZMs	Total ZMs
Segment 13	289:302	
Segment 21	481:499	
Segment 11	241:259	
Segment 12	265:278	96
Segment 18	409:418	
Segment 10	217:226	
Segment 22	505:514	

Table 3. Obtained ZMs based on ranking method (2)

Segments #	ZMs	Total ZMs
Segment 16	361:367	
Segment 2	25:32	
Segment 10	217:226	
Segment 18	409:418	
Segment 22	505:514	
Segment 8	169:177	
Segment 12	265:278	154
Segment 13	289:302	
Segment 17	385:395	
Segment 1	1:16	
Segment 5	97:110	
Segment 15	337:352	
Segment 19	433:447	

Table 4. Obtained ZMs based on ranking method (3)

Segments #	ZMs	Total ZMs
Segment 4	73:92	
Segment 10	217:226	
Segment 11	241:259	
Segment 12	265:278	
Segment 13	289:302	
Segment 18	409:418	
Segment 21	481:499	
Segment 1	1:16	201
Segment 14	313:332	
Segment 15	337:352	
Segment 22	505:514	
Segment 3	49:67	
Segment 5	97:110	

as positive class and the remaining 396 FRII images as a negative class construct our balance data sets. We adopted the complex orthogonal Zernike polynomials to convert the images into a series with 528 moments. The magnitude values of the ZMs as one-dimensional data carry the property of the image as two-dimensional data. We employed the feature selection scheme rooted in the information theory concepts and machine learning techniques to overcome the curse of dimensionality. The MFWFS scheme involved the filter and wrapper phases is applied to the 528-moment series as a hybrid algorithm for optimizing the range of moments. We elaborated on the obtained results in the following paragraph according to subsections 4.2 to 4.4.

In the proposed framework, the performance of obtained quad-final MDFs in 528 features (i.e., 96, 154, 201, 274) (See Section 4.4) are evaluated to predict the class of radio galaxy image. To do this, we conducted a 22-fold cross-validation technique based on the TWSVM-RBF classifier in an iteration manner. In fact, 22-fold training and testing sets were constructed on 726 samples regarding quad-final MDFs of the ZMs, and then the mean of performance for the 22-fold classifier was calculated. The C and σ parameters in TWSVM-RBF are considered to find the optimal pair in the learning procedure. Table 6 shows the evaluated metrics for the performance of the proposed framework. During the procedure for selecting the quad-final MDFs from the 528-moments series, the optimal ranges of features are obtained to be 96, 154, 201, and 274 moments. First, the classification accuracy of ZMs was computed by applying the proposed learning scenario. According to the fine-tuning procedure grouped by fold, the maximum accuracies are given in Table 7. Also, PPV, TPR, TNR, and F-measure were obtained in each fold concerning obtained accuracy.

Table 5. Obtained ZMs based on ranking method (4)

Segments #	ZMs	Total ZMs
Total ZMs based on method (3)	201 ZMs	
Segment 2	25:32	274
Segment 7	145:160	
Segment 16	361:367	
Segment 17	385:395	
Segment 19	433:447	
Segment 20	457:472	

Table 6. The performance metrics

Metrics	Descriptions
Accuracy	Acc=(TP+TN)/(TP+TN+FP+FN)
Sensitivity (Recall)	TPR=TP/(TP+FN)
Specificity	TNR=TN/(TN+FP)
Precision	PPV= TP/(TP+FP)
F-measure	F_1=(2×PPV×Recall)/(PPV+Recall)

Symbols; P: stable sample, N: unstable sample,
T: predicted correctly, F: predicted incorrectly

According to metrics, the obtained performance of the 22-fold classifier for the predicting class of data is about 82%. The performance of the proposed learning framework based on TWSVM-RBF applied to the quad-final ZMs with 274, 201, 154, and 96 features are shown in Tables 8 to 11, respectively.

The classification accuracies of 274, 201, and 154 MDFs are close to 83% that is higher than the classification accuracy of 528 moments (81.39%). Also, for 96 features selected from 528 moments (dimension reduction rate: 5.5), the classification accuracy is close to 80%. In Table 11, we see that the learning procedure on fold 1, fold 6, fold 15, fold 16, and fold 22 with 528 ZMs and 96 ZMs, indicates the dominant of 96 ZMs over 528 ZMs based on the accuracy metric. Also, the classification performance for 96 series moments has equal prediction capacity to 528 ZMs in fold 5, fold 11, fold 12, and fold 17. Generally, the results show the higher efficacy of the proposed method for the different number of selected MDFs in grouping radio galaxy images.

6. Comparison of experimental methods: MFWFS scheme vs. two efficient FSS algorithms

In this section, we aim to compare the performance of the MFWFS with two efficient FSS algorithms for predicting the type of radio galaxy images (type I and type II) is considered in this section. Hence, we focused on the ReliefF algorithm [9] and minimum redundancy maximum relevance (mRMR) [10] which are considered by scholars in the field of

Table 7. The performance of the TWSVM-RBF applied to 528 ZMs per fold in the fine-tuning procedure

No. feature set	22-fold cross validation				
	Max of Acc. index for each fold based on fine-tuning of TWSVM parameters: Accuracy [TPR, TNR, PPV, F-measure]				
	fold 1	**fold 2**	**fold 3**	**fold 4**	**fold 5**
	87.87 [86.67, 88.89, 86.67, 86.67]	75.75 [60.00, 88.89, 81.82, 69.23]	90.90 [86.67, 94.44, 92.86, 89.66]	78.78 [73.33, 83.33, 78.57, 75.86]	75.75[66.67, 83.33, 76.92, 71.43]
	fold 6	**fold 7**	**fold 8**	**fold 9**	**fold 10**
528	81.81 [66.67, 94.44, 90.91, 76.92]	78.78 [73.33, 83.33, 78.57, 75.86]	81.81 [80.00, 83.33, 80.00, 80.00]	75.75 [73.33, 77.78, 73.33, 73.33]	81.81 [60.00, 100, 100, 75.00]
	fold 11	**fold 12**	**fold 13**	**fold 14**	**fold 15**
	72.72 [73.33, 72.22,68.75, 70.97]	81.81 [86.67, 77.78, 76.47, 81.25]	78.78 [93.33, 66.67, 70.00, 80.00]	81.81 [73.33, 88.89, 84.62, 78.57]	78.78 [80.00, 77.78, 75.00, 77.42]
	fold 16	**fold 17**	**fold 18**	**fold 19**	**fold 20**
	75.75 [66.67, 83.33,76.92, 71.43]	81.81 [73.33, 88.89, 84.62, 78.57]	84.84 [86.67, 83.33, 81.25, 83.87]	78.78 [73.33, 83.33, 78.57, 75.86]	96.96 [100, 94.44, 93.75, 96.77]
	fold 21		**fold 22**		
	84.84 [80.00, 88.89, 85.71, 82.76]		84.84 [93.33, 77.78, 77.78, 84.85]		
	Mean of Acc. index of folds: Accuracy [mean]				
	81.39				

(Note: leftmost column label "TWSVM Classifier" appears vertically spanning all data rows.)

Table 8. The performance of the TWSVM-RBF applied to 274 ZMs per fold in the fine-tuning procedure

No. feature set	22-fold cross validation				
	Max of Acc. index for each fold based on fine-tuning of TWSVM parameters: **Accuracy [TPR, TNR, PPV, F-measure]**				
	fold 1	**fold 2**	**fold 3**	**fold 4**	**fold 5**
	87.87 [93.33, 77.78, 77.78, 84.85]	78.78 [86.67, 72.22, 72.22, 78.79]	90.90 [86.67, 94.44, 92.86, 89.66]	78.78 [73.33, 83.33, 78.57, 75.86]	78.78[66.67, 88.89, 83.33, 74.07]
	fold 6	**fold 7**	**fold 8**	**fold 9**	**fold 10**
	84.84 [86.67, 83.33, 81.25, 83.87]	81.81 [80.00, 83.33, 80.00, 80.00]	81.81 [80.00, 83.33, 80.00, 80.00]	78.78 [73.33, 83.33, 78.57, 75.86]	84.84 [86.67, 83.33, 81.25, 83.87]
528 ∝ 274 (1.93)	**fold 11**	**fold 12**	**fold 13**	**fold 14**	**fold 15**
	66.66 [60.00, 72.22, 64.29, 62.07]	78.78 [73.33, 83.33, 78.57, 75.86]	84.84 [93.33, 77.78, 77.78, 84.85]	81.81 [73.33, 88.89, 84.62, 78.57]	81.81 [73.33, 88.89, 84.62, 78.57]
	fold 16	**fold 17**	**fold 18**	**fold 19**	**fold 20**
	75.75 [60.00, 88.89, 81.82, 62.93]	84.84 [86.67, 83.33, 81.25, 83.87]	84.84 [80.00, 88.89, 85.71, 82.76]	81.81 [86.67, 77.78, 76.47, 81.25]	96.96 [100, 94.44, 93.75, 96.77]
	fold 21			**fold 22**	
	87.87 [80.00, 94.44, 92.31, 85.71]			87.87 [86.67, 88.89, 86.67, 86.67]	
	Mean of Acc. index of folds: **Accuracy** [mean]				
	82.77				

Table 9. The performance of the TWSVM-RBF applied to 201 ZMs per fold in the fine-tuning procedure

No. feature set	22-fold cross validation				
	Max of Acc. index for each fold based on fine-tuning of TWSVM parameters: **Accuracy [TPR, TNR, PPV, F-measure]**				
	fold 1	**fold 2**	**fold 3**	**fold 4**	**fold 5**
	87.87 [100, 77.78, 78.95, 88.24]	75.75 [80.00, 72.22, 70.59, 75.00]	90.90 [86.67, 94.44, 92.86, 89.66]	78.78 [73.33, 83.33, 78.57, 75.86]	81.81[80.00, 83.33, 80.00, 80.00]
	fold 6	**fold 7**	**fold 8**	**fold 9**	**fold 10**
	87.87 [93.33, 83.33, 82.35, 87.50]	84.84 [86.67, 83.33, 81.25, 83.87]	81.81 [86.67, 77.78, 76.47, 81.25]	78.78 [73.33, 83.33, 78.57, 75.86]	84.84 [86.67, 83.33, 81.25, 83.87]
528 ∝ 201 (2.63)	**fold 11**	**fold 12**	**fold 13**	**fold 14**	**fold 15**
	69.69 [66.67, 72.22, 66.67, 66.67]	81.81 [86.67, 77.78, 76.47, 81.25]	84.84 [93.33, 77.78, 77.78, 84.85]	81.81 [73.33, 88.89, 84.62, 78.57]	78.78 [73.33, 83.33, 78.57, 75.86]
	fold 16	**fold 17**	**fold 18**	**fold 19**	**fold 20**
	75.75 [60.00, 88.89, 81.82, 69.23]	84.84 [86.67, 83.33, 81.25, 83.87]	84.84 [80.00, 88.89, 85.71, 82.76]	81.81 [80.00, 83.33, 80.00, 80.00]	96.96 [100, 94.44, 93.75, 96.77]
	fold 21			**fold 22**	
	87.87 [80.00, 94.44, 92.31, 85.71]			81.81 [73.33, 88.89, 84.62, 78.57]	
	Mean of Acc. index of folds: **Accuracy** [mean]				
	82.91				

Table 10. The performance of the TWSVM-RBF applied to 154 ZMs per fold in the fine-tuning procedure

No. feature set	22-fold cross validation				
	Max of Acc. index for each fold based on fine-tuning of TWSVM parameters: **Accuracy [TPR, TNR, PPV, F-measure]**				
	fold 1	**fold 2**	**fold 3**	**fold 4**	**fold 5**
	87.87 [86.67, 88.89, 86.67, 86.67]	75.75 [86.67, 66.67, 68.42, 76.47]	90.90 [86.67, 94.44, 92.86, 89.66]	78.78 [73.33, 83.33, 78.57, 75.86]	75.75 [60.00, 88.89, 81.82, 69.23]
	fold 6	**fold 7**	**fold 8**	**fold 9**	**fold 10**
	84.84 [93.33, 77.78, 77.78, 84.85]	84.84 [86.67, 83.33, 81.25, 83.87]	75.75 [86.67, 66.67, 68.42, 76.47]	78.78 [73.33, 83.33, 78.57, 75.86]	84.84 [86.67, 83.33, 81.25, 83.87]
528 ∝ 154 (3.43)	**fold 11**	**fold 12**	**fold 13**	**fold 14**	**fold 15**
	72.22 [73.33, 72.22, 68.75, 70.97]	81.81 [86.67, 77.78, 76.47, 81.25]	84.84 [93.33, 77.78, 77.78, 84.85]	84.84 [73.33, 94.44, 91.67, 81.48]	81.81 [80.00, 83.33, 80.00, 80.00]
	fold 16	**fold 17**	**fold 18**	**fold 19**	**fold 20**
	72.72 [60.00, 83.33, 75.00, 66.67]	81.81 [80.00, 83.33, 80.00, 80.00]	84.84 [86.67, 83.33, 81.25, 83.87]	81.81 [80.00, 83.33, 80.00, 80.00]	96.96 [100, 94.44, 93.75, 96.77]
	fold 21			**fold 22**	
	81.81 [80.00, 83.33, 80.00, 80.00]			87.87 [86.67, 88.89, 86.67, 86.67]	
	Mean of Acc. index of folds: **Accuracy** [mean]				
	82.36				

(The leftmost column header spans: "TWSVM Classifier")

Table 11. The performance of the TWSVM-RBF applied to 96 ZMs per fold in the fine-tuning procedure

No. feature set	22-fold cross validation				
	Max of Acc. index for each fold based on fine-tuning of TWSVM parameters: **Accuracy [TPR, TNR, PPV, F-measure]**				
	fold 1	**fold 2**	**fold 3**	**fold 4**	**fold 5**
	93.93 [86.67, 100, 100, 92.86]	66.66 [80.00, 55.56, 60.00, 68.57]	87.87 [80.00, 94.44, 92.31, 85.71]	75.75 [80.00, 72.22, 70.59, 75.00]	75.75 [66.67, 83.33, 76.92, 71.43]
	fold 6	**fold 7**	**fold 8**	**fold 9**	**fold 10**
	84.84 [86.67, 83.33, 81.25, 83.87]	72.72 [66.67, 77.78, 71.43, 68.97]	75.75 [60.00, 88.89, 81.82, 69.23]	72.72 [60.00, 83.33, 75.00, 66.67]	78.78 [93.33, 66.67, 70.00, 80.00]
528 ∝ 96 (5.5)	**fold 11**	**fold 12**	**fold 13**	**fold 14**	**fold 15**
	72.72 [73.33, 72.22, 68.75, 70.97]	81.81 [93.33, 72.22, 73.68, 82.35]	75.75 [73.33, 77.78, 73.33, 73.33]	75.75 [73.33, 77.78, 73.33, 73.33]	84.84 [66.67, 100, 100, 80.00]
	fold 16	**fold 17**	**fold 18**	**fold 19**	**fold 20**
	78.78 [73.33, 83.33, 78.57, 75.86]	81.81 [80.00, 83.33, 80.00, 80.00]	72.72 [80.00, 66.67, 66.67, 72.73]	75.75 [86.67, 66.67, 68.42, 76.47]	84.84 [80.00, 88.89, 85.71, 82.76]
	fold 21			**fold 22**	
	81.81 [80.00, 83.33, 80.00, 80.00]			87.87 [86.67, 88.89, 86.67, 86.67]	
	Mean of Acc. index of folds: **Accuracy** [mean]				
	79.05				

(The leftmost column header spans: "TWSVM Classifier")

Table 12. Comparing MFWFS scheme performance with mRMR and ReliefF algorithms

Methods	Prediction accuracy based on best-laid MDFs (quad-final feature set: No. 96, 154, 201, and 274) obtained by FSS schemes			
	best-laid (# 96)	best-laid (# 154)	best-laid (# 201)	best-laid (# 274)
	Mean Acc. of 22-fold (rank)			
MFWFS	79.05 % (2)	82.36 % (1)	82.91 % (1)	82.77 % (1)
	Mean Acc. of 22-fold (rank)			
mRMR	80.98 % (1)	81.25 % (3)	81.81 % (2)	82.22 % (2)
	Mean Acc. of 22-fold (rank)			
ReliefF	78.09 % (3)	81.39 % (2)	79.88 % (3)	79.05 % (3)

FSS. In the first step, two efficient FSS (2eFSS) was applied to the ZMs for selecting the quad-final MDFs including the best-laid feature sets (96, 154, 201, and 274 optimal moment packages). Next, the obtained MDFs by the 2eFSS were fed to the TWSVM-RBF classifier for performance evaluation on categorizing FRI and FRII radio images. As can be seen in Table 12, the MFWFS outperformed the mRMR and ReliefF approaches based on the accuracy metric regarding the 22-fold cross-validation technique (ignoring only 1.93% less Acc. than the mRMR in 96-MDFs). In fact, selecting optimal features in the form of the quad-final MDFs by the MFWFS leads to higher performance than the quad-final features survived by the 2eFSS in the classification process of radio galaxies. A closer look at the results in Table 12 shows that the categorizing FRI and FRII radio images based on the best-laid MDFs (# 96) obtained by the MFWFS has a similar performance with the best-laid MDFs (# 274) selected by the ReliefF. Furthermore, classification results based on the best-laid MDFs (# 154) obtained by the MFWFS show a higher performance than the best-laid MDFs (# 201) and the best-laid MDFs (# 274) survived by the mRMR and ReliefF in classification. Also, employing the best-laid MDFs (# 201) selected by the MFWFS leads to the highest accuracy than the other obtained MDFs packages by the MFWFS, mRMR, and ReliefF.

7. Summary and conclusions

Here, we proposed an automatic method for classifying radio galaxy data using the supervised framework. We applied the polynomial Zernike function to both FRI and FRII images. Each image is converted to a series with 528 moments involving unique invariant characteristics. Regardless of making an effort for generating well-structured datasets for radio galaxy datasets, the ZMs of images can provide the most significant concern for knowledge discovery from radio galaxies. The great challenge was to deal with long series leading to the curse of dimensionality. To address this concern, we proposed the MFWFS scheme for selecting the MDFs. Based on the MFWFS strategy, we conducted data modularization on two dimensions (sample and features) into the given subset for precise mining on the ZMs space in the dual-phase. In the filter phase, the criteria of relevancy, interdependency, and redundancy were considered in the selected POFs. The MDFs are chosen from the inserted POFs into the wrapper phase using the TWSVM classifier. The experimental comparison was presented for performance evaluation of the proposed strategy according to the cross-validation technique. Based on obtained results, the different number of selected MDFs show that the subsets with 274, 201, and 154 features are subsets with optimal ranges. These subsets have classification accuracies close to 83% greater than the accuracy obtained for the 528-moment series with 81.39%. The value of accuracy approaches 80% in a subset of 96 features with a dimension reduction rate of 5.5. This multilevel method is optimized to reduce the computational time of running a program with more reserved memory. To ensure the efficacy of the MFWFS scheme, experimental comparison in the presence of 2eFSS was considered as the final report in this paper. The results showed that the MFWFS outperformed mRMR and ReliefF algorithms based on accuracy metric regarding cross-validation technique. The optimum fully automated process of this method and its compatibility with image datasets make it applicable and more convenient in all fields that deal with each type of visual data.

For future work, we intend to modify the proposed the proposed method for datasets classified into more than two categories. Also, we aim to make mock data simulating radio galaxy images and extract their Zernike moments. Comparing the ZMs of both simulated data and the original one would

give a better description of the morphology of galaxies. Moreover, this method can be applied to astronomical images such as solar data recorded in various wavelengths to discriminate features that appeared in different layers of the Sun.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, and project administration done by 1st author. Writing—original draft preparation, writing—review and editing, visualization, and supervision done by 2nd. Data curation done by 3rd and 4th authors.

References

[1] D. Eckert, M. Gaspari, F. Gastaldello, A. M. C. L. Brun, and E. O'Sullivan, "Feedback from Active Galactic Nuclei in Galaxy Groups", *Universe*, Vol. 7, No. 5, pp. 142, 2021.

[2] V. Kalinova, D. Colombo, S. F. Sánchez, K. Kodaira, R. G. Benito, R. G. Delgado, E. Rosolowsky, and E. A. D. Lacerda, "Star Formation Quenching Stages of Active and Non-Active Galaxies", *Astronomy & Astrophysics*, Vol. 648, 2021.

[3] P. Dabhade, M. Mahato, J. Bagchi, D. J. Saikia, F. Combes, S. Sankhyayan, H. J. A. Rottgering, L. C. Ho, M. Gaikwad, S. Raychaudhury, B. Vaidya, and B. Guiderdoni, "Search and Analysis of Giant Radio Galaxies with Associated Nuclei (SAGAN). I. New Sample and Multi-Wavelength Studies", *Astronomy & Astrophysics*, Vol. 642, 2020.

[4] B. L. Fanaroff and J. M. Riley, "The Morphology of Extragalactic Radio Sources of High and Low Luminosity", *Monthly Notices of Royal Astronomical Society*, Vol. 167, No. 1, pp. 31-36 1974.

[5] Y. T. Lin, Y. Shen, M. A. Strauss, G. T. Richards, and R. Lunnan, "On the Populations of Radio Galaxies with Extended Morphology at z < 0.3", *The Astrophysical Journal*, Vol. 723, No. 2, 2010.

[6] J. J. Harwood, T. Vernstrom, and A. Stroe, "Unveiling the Cause of Hybrid Morphology Radio Sources (HyMoRS)", *Monthly Notices of the Royal Astronomical Society*, Vol. 491, No. 1, pp. 803-822, 2020.

[7] M. Sadeghi, M. Javaherian, and H. Miraghaei, "Morphological-based Classifications of Radio Galaxies Using Supervised Machine-learning Methods Associated with Image Moments", *The American Astronomical Journal*, Vol. 161, No. 2, 2021.

[8] R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson, and J. H. Moore, "Relief-Based Feature Selection: Introduction and Review", *Journal of Biomedical Informatics*, Vol. 85, pp. 189-203, 2018.

[9] M. R. Sikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF", *Machine Learning*, Vol. 53, pp. 23-69, 2003.

[10] H. C. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp. 1226-1238, 2005.

[11] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", In: *Proc. of the Twentieth International Conference (ICML)*, 2003.

[12] H. Miraghaei and P. N. Best, "The Nuclear Properties and Extended Morphologies of Powerful Radio Galaxies: The Roles of Host Galaxy and Environment", *Monthly Notices of the Royal Astronomical Society*, Vol. 466, No. 4, pp. 4346-4363, 2017.

[13] P. N. Best, G. Kauffmann, T. M. Heckman, and Z. Ivezic, "A Sample of Radio-Loud Active Galactic Nuclei in The Sloan Digital Sky Survey", *Monthly Notices of the Royal Astronomical Society*, Vol. 362, No. 1, pp. 9-24, 2005.

[14] P. N. Best and T. M. Heckman, "On The Fundamental Dichotomy in The Local Radio-AGN Population", *Monthly Notices of the Royal Astronomical Society*, Vol. 421, pp. 1569-1582, 2012.

[15] T. Arif, Z. Shaaban, L. Krekor, and S. Baba, "Object Classification via Geometrical, Zernike and Legendre Moments", *Journal of Theoretical & Applied Information Technology*, Vol. 6, No. 3, 2009.

[16] K. M. Hosny, "A Systematic Method for Efficient Computation of Full and Subsets Zernike Moments", *Information Sciences*, Vol. 180, No. 11, pp. 2299-2313, 2010.

[17] V. G. V. Mahesh and A. N. J. Raj, "Zernike Moments and Machine Learning Based Gender Classification Using Facial Images", In: *Proc.*

of the Eighth International Conference on Soft Computing and Pattern Recognition, 2016.

[18] C. H. Teh and R. T. Chin, "On image Analysis by The Methods of Moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 496-513, 1988.

[19] M. Javaherian, H. Safari, A. Amiri, and S. Ziaei, "Automatic Method for Identifying Photospheric Bright Points and Granules Observed by Sunrise", *Solar Physics*, Vol. 289, pp. 3969-3983, 2014.

[20] C. Cortes and V. Vapnik, "Support Vector Networks", *Springer*, Vol. 20, No. 3, pp. 272-297, 1995.

[21] Jayadeva, R. Khemchandani, and S. Chandra, "Twin Support Vector Machines for Pattern Classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 5, pp. 905-910, 2007.

[22] H. Huang, X. Wei, and Y. Zhou, "Twin support vector machines: A survey", *Neurocomputing*, Vol. 300, pp. 34-43, 2018.