Applied Soft Computing 11 (2011) 5662–5673

Rough-wavelet granular space and classification of multispectral remote sensing image

Saroj K. Meher[a,*,1], Sankar K. Pal[b,2]

[a] Systems Science and Informatics Unit, Indian Statistical Institute, Bangalore Center, Bangalore 560059, India
[b] Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 22 July 2010
Received in revised form 6 February 2011
Accepted 23 March 2011
Available online 2 April 2011

Keywords:
Wavelet information granulation
Rough neighborhood sets
Rough-wavelet granular computing
Pattern recognition
Remote sensing

ABSTRACT

A new rough-wavelet granular space based model for land cover classification of multispectral remote sensing image, is described in the present article. In this model, we propose the formulation of class-dependent (CD) granules in wavelet domain using shift-invariant wavelet transform (WT). Shift-invariant WT is carried out with properly selected wavelet base and decomposition level(s). The transform is used to characterize the feature-wise belonging of granules to different classes, thereby producing wavelet granulation of the feature space. The wavelet granules thus generated possess better class discriminatory information. The granulated feature space not only analyzes the contextual information in time or frequency domain individually, but also looks into the combined time–frequency domain. These characteristics of the generated CD wavelet granules are very useful in the pattern classification with overlapping classes. Neighborhood rough sets (NRS) are employed in the selection of a subset of granulated features that further explore the local/contextual information from neighbor granules. The model thus explores mutually the advantages of shift-invariant wavelet granulation and NRS. The superiority of the proposed model to other similar methods is established both visually and quantitatively for land cover classification of multispectral remote sensing images. With experimental results, it is found that the proposed model is superior with biorthogonal3.3 wavelet, and when integrated with NRS, it performs the best.

1. Introduction

Granular computing (GrC) refers to that where computation and operations are performed on information granules (clumps of similar objects or points). Its nature and applicability has been changed rapidly from a label to conceptual and computational paradigm of study that deals with information and knowledge processing [1]. Many researchers [2,3] have used GrC models to build efficient computational algorithms that can handle huge amount of data, information and knowledge. These models mainly deal with the efficiency, effectiveness and robustness of using granules, such as classes, clusters, subsets, groups and intervals, in problem solving [4–6].

GrC can be studied based on its notions of representation and process. However, the main task to be focused is to construct and describe information granules, a process, called information granulation [7–9] on which GrC is oriented. Several attempts have been made to construct information granules specifically in spatial domain and applied to various areas, including automatic target recognition [10], color image segmentation [11], and remote sensing image classification [12]. Multispectral remote sensing images contain information over a large range of variation of frequencies, and this too changes over regions. These data have both spectral features with correlated bands and spatial features correlated in the same band. Simultaneous utilization of these spectral and spatial (contextual) information in an effective manner can significantly improve its analysis. There have been several attempts in utilizing the merits of local information [13] in a band for the classification of remote sensing images. For example, texture features [14] extracted from angular second moments, contrast, correlation, entropy and variance based on the grey level co-occurrence matrices have found wide applications. However, these methods are computationally expensive. Later on, Gaussian Markov random fields and Gibbs random fields were proposed to characterize textures [13]. The aforesaid statistical approaches to texture analysis are mostly restricted to the analysis of spatial interactions relatively over small neighborhoods in a single scale/band.

One efficient way to deal with the problems of simultaneous utilization of spectral and spatial information is to analyze the image by a number of subsampled approximations of it at different resolutions, called "multiresolution analysis" [15]. In this regard, wavelet transform (WT) has received tremendous attention as a

* Corresponding author. Tel.: +91 80 28483002x534; fax: +91 80 28484265.
E-mail address: saroj.meher@gmail.com (S.K. Meher).
[1] Member, IEEE.
[2] Fellow, IEEE.

promising tool for analyzing texture regions of image, in both spatial (time) and spectral (frequency) domains. This characteristic of the WT thus encourages one to use it for the extraction of contextual information of pixels in remote sensing images by wavelet granulation (i.e., clump of similar information in WT domain) of feature space. Many investigations on texture classification using WT have already been reported [13]. The WT, in general, is categorized as shift/time variant and shift invariant. In shift variant WT [15], the filtered versions of each (sub)image are downsampled by a factor of two and it results in a non-redundant analysis process. Although shift variant WT is quite attractive for various applications, it does not maintain the indispensable property of textural analysis, like time invariance and makes it insufficient for dealing with such problem. The shift-invariant WT, on the other hand, does not perform the downsampling operation of shift variant WT and thus provides a shift invariant analysis in the exploration of local/contextual information of pixels in an image.

In general, the process of wavelet granulation can be broadly categorized as class-dependent (CD) and class-independent (CI). WT is used in both cases for time–frequency representation of patterns and generation of wavelet granulation of the feature space. WT decomposes the original frequency band of an image into four equal areas subbands with one-level of decomposition, thereby producing wavelet granulation of feature space and characterizing four wavelet granules along the axis. With this process of granulation, each feature of the pattern is described by four wavelet granules over the whole space for one-level of WT decomposition, and hence called CI method. However, this process of granulation does not take care of the class belonging information of features to different classes. This may lead to a degradation of performance in pattern classification, particularly for data sets with highly overlapping classes. On the other hand, in CD granulation, each feature explores its class belonging information to different classes. In this process, features are described by the wavelet granules equal to $4 \times L$(number of classes) for one-level of WT decomposition, and individual class information is restored by the generated wavelet granules.

Rough set theory, as proposed by Pawlak [8] (henceforth it will be abbreviated as PaRS), has been proven to be an effective tool for feature selection, uncertainty handling, knowledge discovery and rule extraction from categorical data [16]. The theory enables the discovery of data dependencies and performs the reduction/selection of attributes contained in a data set using the data alone, requiring no additional information. PaRS can be used as an effective tool to deal with both vagueness and uncertainty in data sets and to perform granular computation. PaRS based feature selection not only retains the representational power of the data, but also maintains its minimum redundancy [16]. However for the numerical data, PaRS theory can be used with the discretisation of data that results in the loss of information and introduction of noise. To deal with this, neighborhood rough set (NRS) [17,18] is found to be suitable that can deal with both numerical and categorical data sets without discretisation. The advantage of NRS is that it facilitates to gather the possible local information through neighbor granules that is useful for a better discrimination of patterns, particularly in class overlapping environment. Various synergistic integrations of rough sets and other soft computing tools like fuzzy sets and artificial neural networks with application specific merits are described in [19,20].

In this article, we describe a rough-wavelet granular space using CD wavelet granulation and NRS based feature selection. The model provides a synergistic integration of the merits of both CD wavelet granulation and the feature selection capability of the theory of NRS using neighborhood information. The resulting output of this judicious integration can be used as an input to any classifier for pixel classification. To demonstrate the effectiveness of the proposed rough-wavelet granular space based model, we have used here different classifiers, such as k-nearest neighbor (k-NN) ($k = 1$, 2 and 3) classifier, maximum likelihood (ML) classifier [21] and multi-layered perceptron (MLP) [22]. However, other classifiers may also be used. We have demonstrated the potentiality of the model with two real and one synthetic multispectral remote sensing images having their spectral (band) values as input features. The superiority of the proposed model to others is validated both visually and quantitatively. Performance measures such as β index [23], Davies–Bouldin (DB) index [24] and computation time are considered for real life remote sensing images. For the synthetic noisy remote sensing images, percentage of overall classification accuracy is computed.

Apart from demonstrating a way of integrating the merits of rough sets and wavelet transform for handling overlapping classes, the significance of the present work lies with the following two operations: First, based on class dependency knowledge, wavelet granulated feature space is generated in time–frequency plane using the shift-invariant WT. Second, the neighborhood rough sets are applied on these wavelet granulated features for computing the approximate reducts that select a subset of features. The experimental results with both synthetic and real life multispectral remote sensing images revealed that the proposed model preserved the homogeneity and structure of various regions of remote sensing images and improved the classification accuracy in terms of various quantitative measures. Different wavelets are used for the present study. Comparison of results showed that the performance of the proposed model is further improved with the biorthogonal3.3 (bior3.3) wavelet.

The organization of the article is as follows. A brief description of shift-invariant WT and image feature representation is made in Section 2. Section 3 describes the proposed model for classification with its characteristic features. Different indexes for performance measurement are discussed in Section 4. Results and discussion are included in Section 5. Finally, the concluding remarks are given in Section 6.

2. Shift-invariant discrete wavelet transform and image feature representation

The wavelet transform (WT) is primarily developed for the analysis of non-stationary signals. The transform works on a dual plane instead of working on a single plane (time or frequency). The transform performs the decomposition of signal into a number of scales and each scale represents a particular coarseness of that signal. The discrete WT (DWT) has become largely popular because of its computationally efficient implementation using the Mallat algorithm [15]. Broadly, the DWT can be categorized as shift/time/translation-variant (non-redundant) and shift-invariant (redundant). Two-dimensional (2D) shift-variant DWT (SV-DWT) (extension of one-dimensional SV-DWT) [15] is a separable filter bank in row and column directions and it performs the one-level decomposition of an image into four subimages in four equal areas subbands, as shown in Fig. 1a. H and L in Fig. 1a denote the highpass and lowpass filters, respectively. $\downarrow 2$ denotes the downsampling operation by a factor of 2. The approximate image LL is the low-frequency component obtained by lowpass filtering of the input in both row and column directions. The detail images LH, HL and HH are the high-frequency components including horizontal, vertical and diagonal information, respectively. For more levels of DWT decomposition, the lower frequency component (LL) is recursively processed. With this process, the SV-DWT with Q-level of decomposition generates a total of $3Q + 1$ subbands.

The SV-DWT is well-liked for several reasons. Among them, the compression ability of the transform is better explored with no

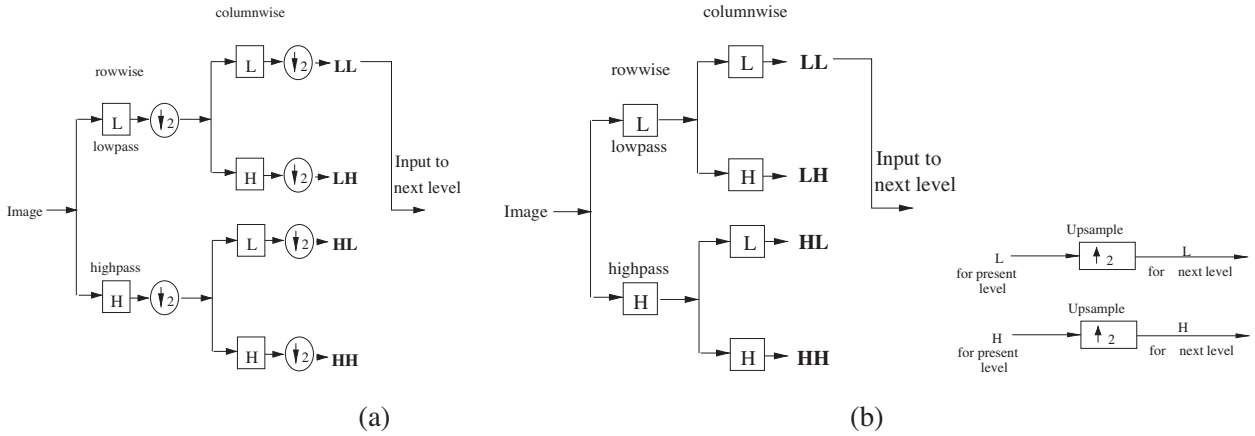

Fig. 1. Two-dimensional wavelet transform for one-level decomposition: (a) SV-DWT and (b) SI-DWT.

loss or redundancy of information between the levels. However, the major drawback of SV-DWT is its variation in time (i.e., the coefficients of a delayed information are not a time shifted version of those of the original), which is particularly important in texture analysis, e.g., land cover regions in remote sensing image. An appropriate solution to this problem is the shift-invariant DWT (SI-DWT), where the decomposition is performed without downsampling operation, and the filter coefficients (L and H) are upsampled (↑2) by a factor of 2 for using them at next level of decomposition, as shown in Fig. 1b. As a result, the SI-DWT provides a shift invariant representation of the input. Similar to 2D SV-DWT, 2D SI-DWT decomposes the original frequency band into four equal areas subbands with one-level of decomposition and the corresponding frequency partition is shown in Fig. 2. The sizes of the subimages obtained by SV-DWT decrease with the increase of decomposition levels, whereas their sizes remain same as the original using SI-DWT. This redundant representation of SI-DWT is more demanding in terms of both memory and time because the subimages obtained by SV-DWT require inverse transformation to retain the size of the input image, that is required for pixel-wise processing, as performed in the present study of land cover classification of remote sensing image.

2.1. Feature representation of multispectral image using SI-DWT

For multispectral remote sensing images we have used the spectral (band) values as features. For example, in a four-band remote sensing image, a pixel **F** in (x, y) coordinate is represented by four numeric features and can be expressed as $\mathbf{F}(x, y) = [F_1(x, y), F_2(x, y), F_3(x, y), F_4(x, y)]$, where each of the features $(F_1, \ldots, F_4)$ in (x, y) coordinate represents the spectral values of four-band of images. Thus **F** is visualized as a point in *four*-dimensional vector space. We perform the DWT decomposition of these images up to the desired level (s) and the corresponding subimages are obtained. Since the pixels of the subimages at different levels represent the information of the original pixels, we have used these pixel values to construct the pattern vector. The subimages are then cascaded so that the extracted features of the original multispectral image can be obtained. Fig. 3 shows the cascading of subimages of a single band image obtained by Q-level of DWT decomposition. The cascading process can be extended for the subimages of multiband images.

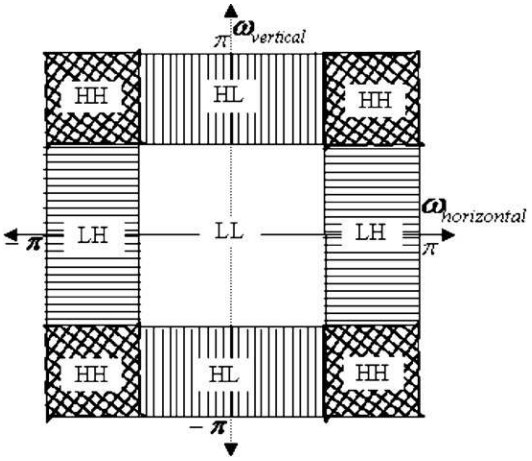

Fig. 2. Two-dimensional WT and its spectral subspaces for one-level decomposition.

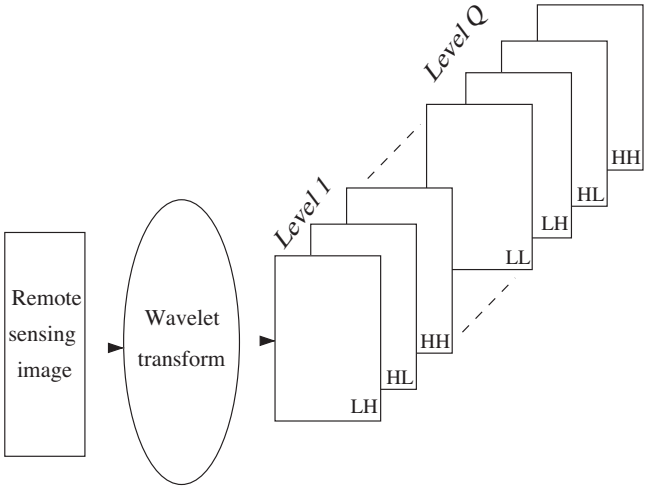

Fig. 3. Feature elements generation with one spectral band of image using WT decomposition.

3. Proposed model for classification

The model has three steps of operation as illustrated in Fig. 4, namely, wavelet based granule generation, rough set based feature selection using reducts, and classification based on the selected features. These are described in Fig. 4.

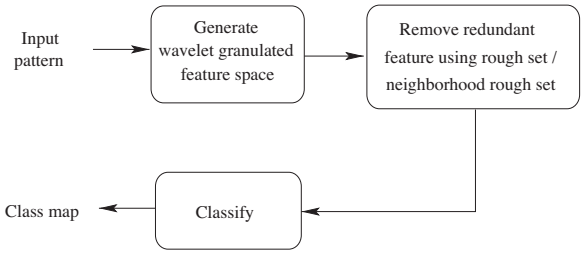

Fig. 4. Schematic flow diagram of the proposed model for classification.

3.1. Class-dependent (CD) granule generation

For class-dependent (CD) wavelet granulation of the input pattern of a multispectral remote sensing image, we have used the shift-invariant discrete wavelet transform (SI-DWT) to characterize the feature values. With CD wavelet granulation, $L \times G$ number of granules are used to characterize the feature values of each pattern, where L = total number of classes and $G = (3Q + 1)$ number of frequency planes characterizing G number of granules, obtained from Q level of WT decomposition. Each feature is thus represented by $L \times G$ number of frequency planes or characterizing $L \times G$ wavelet granules along the axis. The CD granulation explores the class dependency of a pattern into different classes based on individual features and the granules thus provide an improved class-wise representation of input patterns. The granules preserve the interrelated class information to build an informative granular space which is potentially useful for improved classification for the data sets with overlapping classes.

The SI-DWT identifies both scale and space information of the event simultaneously to build an informative granular space that helps to enhance the classification performance. Based on the number of decomposition level(s), each of the input feature spaces of a particular class is represented by the corresponding number of equal areas frequency planes; thereby producing CD wavelet granules in time–frequency plane. A pictorial view of the generated CD wavelet granules for a two-class data set with one-level WT decomposition in two-dimensional (F_1 and F_2) feature space, is shown in Fig. 5. On the other hand, for class-independent (CI) wavelet granulation, each feature is represented by $G = (3Q + 1)$ number of frequency planes or characterized by G number of granules, obtain from Q level of SI-DWT decomposition. Fig. 5 also shows the CI wavelet granulation for a two-class data set in two-dimensional feature space.

Moreover, the selection of decomposition level is a key factor for the successful application of WT in the analysis of signals or images. The decomposition level depends on the type of requirement and it varies with the problem in hand. To have an objective evaluation, we computed the average entropy, which provides a measure of information of the image for each level. We found that the average entropy value is not changing significantly after a certain level of decomposition. For the present experiment, we stopped the decomposition after second level, as the entropy measure was not changing much after this level and thus we were not getting much extra information, even though the cost of computation kept increasing.

Various distinguishable characteristics like spatio-geometric information and energy at different scales, which are normally called the signature of the land covers in remote-sensing images, are preserved with the DWT decomposition using orthogonal basis [15,25] and further improved using biorthogonal bases [26]. Hence, we have considered biorthogonal group of wavelet bases for the present study. These bases are usually more desirable than orthogonal one because they can maintain linear phase characteristic with finite number of impulse responses and the mother wavelets have

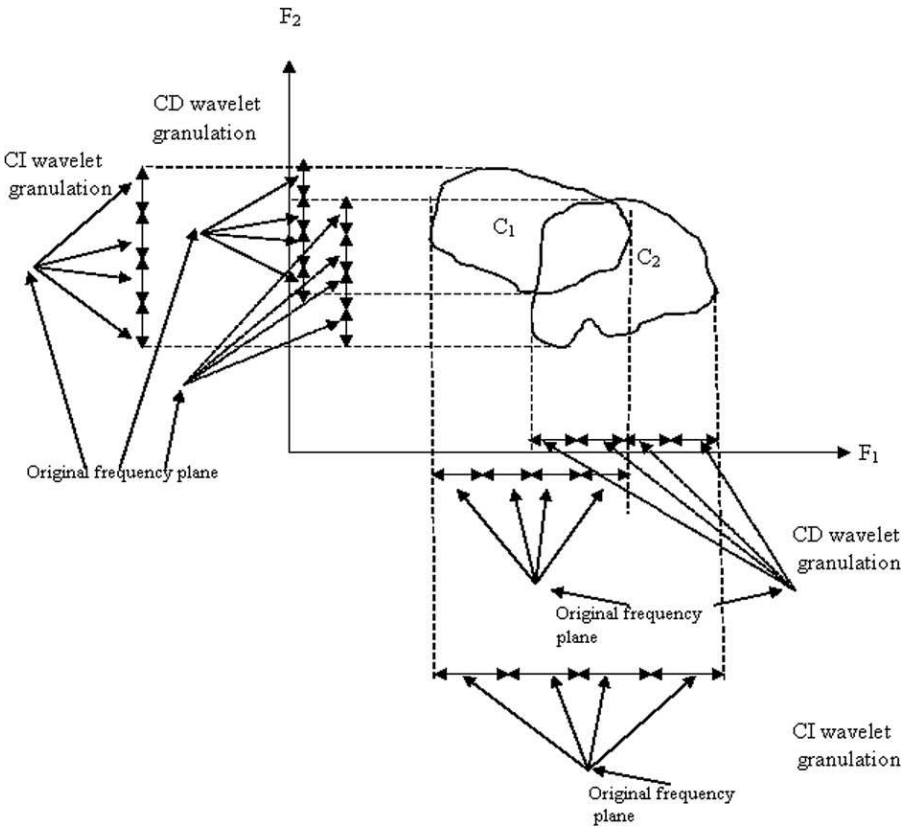

Fig. 5. Wavelet granule generation.

high regularity [26]. It is observed from the experimental results (with the present data sets) that among the biorthogonal (bior) group of wavelets, bior2.2, bior2.4, bior3.1 and bior3.3 provided better performance, and thus we have used these wavelets in the present study.

In the wavelet granulation process, each feature value is represented by large number of subbands characterizing wavelet granules along the axis and it results in the increase of feature dimension. The increased dimension brings great difficulty in solving many tasks of pattern recognition, as in the present case of land cover classification of remote sensing image. This motivates for selecting a subset of relevant and non-redundant features. In this regard, we have used the neighborhood rough set (NRS) [17,18] based feature selection method in the second step of the proposed model (Fig. 4). The advantage in the use of NRS is that it can deal with both numerical and categorical data. NRS does not require any discretisation of numerical data and is suitable for the proposed wavelet granulation of features. Further, the neighboring concept facilitates to gather the possible local information through neighbor granules that provide a better class discrimination information. Thus the combination of these two steps of operation can be a better framework for the classification of patterns in overlapping class environment. The proposed model thus takes the advantage of both CD wavelet granulation using SI-DWT and NRS feature selection methods.

3.2. Feature selection

This section presents some preliminaries relevant to feature selection methods using rough sets (proposed by Pawlak) and neighborhood rough sets (NRS). The details of these theories may be referred to [8,17,18].

3.2.1. Rough sets (PaRS)

Pawlak's rough set (PaRS) theory [8] deals with vague concepts and creates approximate descriptions of objects for data analysis. PaRS is based on the indiscernibility relation that describes indistinguishable objects of the universe. It works with a pair of precise concepts, called as *lower* and *upper* approximations. The *lower* approximation is a description of the domain objects which are known with certainty to belong to the subset of interest, and the *upper* approximation is a description of the objects which possibly as well as definitely belong to the subset. PaRS have been employed to remove redundant conditional features, while retaining their information content. It enables the discovery of data dependencies and the selection of feature subset contained in a data set using the data alone, requiring no additional information. The basic operation involved in PaRS is that it partitions the object space based on a feature set using some equivalence relation. The partition spaces thus generated are also known as *granules*. The generated granules become the elemental building blocks for information granulation process used for data analysis. A measure of significance is then determined by evaluating the change in dependency when a feature is removed from the set. The higher is the change in dependency, the more significant is the feature. Based on this significance a minimum element feature subset (reduct) is searched and located.

Many attempts have been made for finding a reduct of an information system. The simplest solution for locating reducts is to generate all possible subsets and retrieve those with a maximum rough set dependency degree. However, this approach of finding solution is highly expensive for large data sets. For such cases, often one reduct instead of many is required to use for feature reduction. In this regard, the QUICKREDUCT algorithm described by Chouchoulas and Shen [27], is popularly used. The algorithm attempts to calculate a reduct without exhaustively generating all possible feature subsets. It starts with an empty set and adds one feature at a time that results in the increase of rough set dependency. The process goes on until it produces the maximum possible dependency value for a data set. The QUICKREDUCT algorithm is summarised with pseudocode, as shown in Fig. 6.

QUICKREDUCT algorithm:

Input: C, the set of conditional features; D, the set of decision feature
Output: R, the reduct, $R \subseteq C$

a. Initial the feature reduct $R = \phi$ and a temporary variable $T = \phi$
b. **do**
c. {
d. $T = R$
e. For every $a \in (C - R)$
f. {
g. if $\gamma_{R\cup\{a\}}(D) > \gamma_T(D)$, Then $T = R \cup \{a\}$
h. }
i. $R = T$
j. }
k. **until** $\gamma_R(D) == \gamma_C(D)$
l. **return** R

Fig. 6. QUICKREDUCT algorithm for feature selection.

In the present study, we have used QUICKREDUCT algorithm for selecting features generated from the CD wavelet granulation. The selected features are then used in a classifier for classifying the input pattern, as in the third step of Fig. 4.

3.2.2. Neighborhood rough sets (NRS)

As mentioned above the information system is denoted by $I = (U, A)$, where U (the universal set) is a non-empty and finite set of samples $\{x_1, x_2, \ldots, x_n\}$; $A = \{C \cup D\}$, where A is the finite set of features $\{a_1, a_2, \ldots, a_m\}$, C is the set of conditional features and D is the set of decision features. Given an arbitrary $x_i \in U$ and $B \subseteq C$, the neighborhood $\Phi_B(x_i)$ of x_i with given Φ, in feature space B is defined as [18]

$$\Phi_B(x_i) = \{x_i | x_j \in U, \Delta^B(x_i, x_j) \leq \Phi\} \quad (1)$$

where Δ is a distance function.

$\Phi_B(x_i)$ in Eq. (1) is the neighborhood information granule centered with sample x_i. In the present study, we have used three **p**-norm distances in Euclidean space. These are Manhattan distance ($\mathbf{p} = 1$), Euclidean distance ($\mathbf{p} = 2$) and Chebychev distance ($\mathbf{p} = \infty$). The neighborhood granule generation is effected by two key factors such as the used distance function Δ and parameter Φ. The first one determines the shape and second controls the size of neighborhood granule. For example, with Euclidean distance the parameter Φ acts as the radius of the circle region developed by Δ function. Both these factors play important roles in neighborhood rough sets (NRS) and can be considered as to control the granularity of data analysis. The significance of features vary with the granularity levels. Accordingly, the NRS based algorithm selects different feature subsets with the change of Δ function and Φ value. In the present study, we have analyzed the effects of three **p**-norm distances for a variation of Φ values, and selected the best one based on the performance with the present data sets. However, optimal parameters values can be obtained through an optimization technique, e.g., genetic algorithm.

Thus each sample generates granules with a neighborhood relation. For a matric space $\langle U, \Delta \rangle$, the set of neighborhood granules $\{\Phi(x_i) | x_i \in U\}$ forms an elemental granule system, that covers the universal space rather than partitions it as in case of PaRS. A pictorial view of the process of granule generation (as an example) using both PaRS and NRS is shown in Fig. 7.

Let $X = \{a, b, c, d, e, f\}$ be the universal set of five elements (Fig. 7). Partitioning and covering of set X for generating granules are made as $X1 = \{\{a, b\}, \{c, d\}, \{e, f\}\}$ and $X2 = \{\{a, b\}, \{a, c, d\}, \{a,$

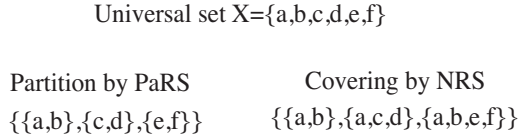

Fig. 7. Example of granule generation using PaRS and NRS.

$b, e, f\}\}$, respectively. A partition of the set X is a division of X into non-overlapping and non-empty "parts" or "blocks" or "cells" that accommodate all the elements of X. Equivalently, a set $X1$ of non-empty sets is a partitions of X if, the intersection of any two distinct elements of $X1$ is empty. On the other hand, a covering of a set X results into overlapping and non-empty "parts" that accommodate all the elements of X. That means a set $X2$ of non-empty sets is a covering of X if, the intersection of any two distinct elements of $X2$ is not necessarily empty. It is noted that the partition of space generated by PaRS can be obtained from NRS with covering principle, while the other way round is not possible. Moreover, a neighborhood granule degrades to an equivalent class for $\Phi = 0$. In this case, the samples in the same neighborhood granule are equivalent to each other and the neighborhood rough set model degenerates to Pawlak's rough set. Thus NRS can be treated as a generalized case of PaRS.

The dependency degree of decision feature D on condition feature set B in a neighborhood information system $\langle U, C \cup D, N\rangle$ with distance function Δ and neighborhood size Φ, is defined as

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|} \tag{2}$$

where $|\bullet|$ denotes the cardinality of a set. $\gamma_B(D)$ is the approximation ability of B to D. For $POS_B(D) \subseteq U$, we have $0 \le \gamma_B(D) \le 1$ and D depends completely on B, and the decision system is consistent in terms of Δ and Φ. For $\gamma_B(D) = 1$, D depends on B in the degree of γ. The dependency function measures the approximation power of a condition feature set. Hence it can be used to determine the significance of a subset of features (normally called as reduct). Significance (SIG) of a subset of features is calculated with the change of dependency, when a feature is removed from the set of considered conditional features.

Based on the significance of a feature(s), the subset of features (reduct) is evaluated. Many sets of reducts can be obtained based on the significance and any of them will work for the feature reduction task. In this regard Hu et al. [18] described a forward greedy search (FGS) algorithm for feature selection using NRS. FGS algorithm begins with an empty reduct. In each step, one feature is added and the change in dependency (significance) is determined, when a feature is removed from the set of considered conditional features. The process is stopped when the significance of reduct is less than a small value ϵ. The algorithm is summarised with pseudocode, as shown in Fig. 8. In the present study, we have used the forward greedy search algorithm for the selection of features in the proposed rough-wavelet granulation based model for classification.

After the features are selected, we use a classifier as in the third step of Fig. 4 to classify the input pattern based on the selected features.

4. Performance measurement indexes

For real life remote sensing image with partially labeled data set, quantitative indexes like β index [23] and Davies–Bouldin index [24], as described next, are used for performance measurement of classifiers. However, for classification of completely labeled synthetic remote sensing images, percentage of accuracy (PA) is used as a measure.

Forward greedy search algorithm:

Input: C, the set of conditional features; D, the set of decision feature
Output: R, the reduct, $R \subseteq C$

a. $R \leftarrow \{\}$
b. **For** each $a_i \in (C - R)$
c. Compute $\gamma_{R\cup a_i}(D) = \dfrac{|POS_{R\cup a_i}(D)|}{|U|}$
d. Compute significance $SIG(a_i, R, D) = \gamma_{R\cup a_i}(D) - \gamma_R(D)$
e. **end**
f. Select the feature a_k satisfying $SIG(a_k, R, D) = \boldsymbol{max}_i(SIG(a_i, R, D))$
g. **if** $SIG(a_k, R, D)) \varepsilon$ // ε, the stopping parameter
h. $R \leftarrow R \cup a_k$
i. go to step 'b'
j. **else**
k. return R
l. **end if**

Fig. 8. Forward greedy search algorithm for feature seection using neighborhood rough set.

4.1. β index

The β index has been defined by Pal et al. in [23], for assessment of image segmentation quality. β is defined as the ratio of the total variation and within-class variation as

$$\beta = \frac{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}})^2}{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^2} \tag{3}$$

where $\bar{\mathbf{x}}$ is the mean grey value of the image pixels (pattern vector), M_i is the number of pixels in the i th ($i = 1, 2, \ldots, C$) class, $\mathbf{x}_{ij}$ is the grey value of the jth pixel ($j = 1, 2, \ldots, M_i$) in class i, and $\bar{\mathbf{x}}_i$ is the mean of M_i grey values of the ith class. Since the numerator is constant for a given image, β value is dependent only on the denominator. The denominator decreases with increase in homogeneity within the class for a fixed number of classes ($\mathbf{C}$). Thus for a given image and given number of classes, the higher the homogeneity within the classes and lower the homogeneity between classes, the higher would be the β value.

Further, in the present work we have evaluated the corresponding percentage of gain of a classifier (b) compared to classifier (a) obtained with respect to β value using the following formula:

$$Gain_b = \frac{\beta \text{ value of classifier } b - \beta \text{ value of classifier } a}{\beta \text{ value of classifier } a} \times 100 \tag{4}$$

4.2. Davies–Bouldin index

Davies–Bouldin (DB) index for cluster validation has been defined in [24]. However, here we are using the index for validating our classification results on partially labeled data sets. The idea behind DB index is that, for a good partition inter-cluster separation as well as intra-cluster homogeneity and compactness should be high. The DB index is based on the evaluation of some measure of dispersion S_i within the ith cluster and the distance between the prototypes of clusters i and j. The DB index is defined as

$$DB = \frac{1}{K}\sum_{i=1}^{K} R_{i,qt} \tag{5}$$

where K is the number of clusters/classes and $R_{i,qt} = \max_{j, j\neq i}[(S_{i,q} + S_{j,q})/d_{ij,t}]$. $S_{i,q}$ is the q th root of q th moment of the points in cluster i with respect to their mean or centroid. $d_{ij,t}$ is the Minkowski distance of order t between the centroids that characterize the extracted classes i and j. The smaller the DB value the better is the partitioning [24]. The corresponding percentage

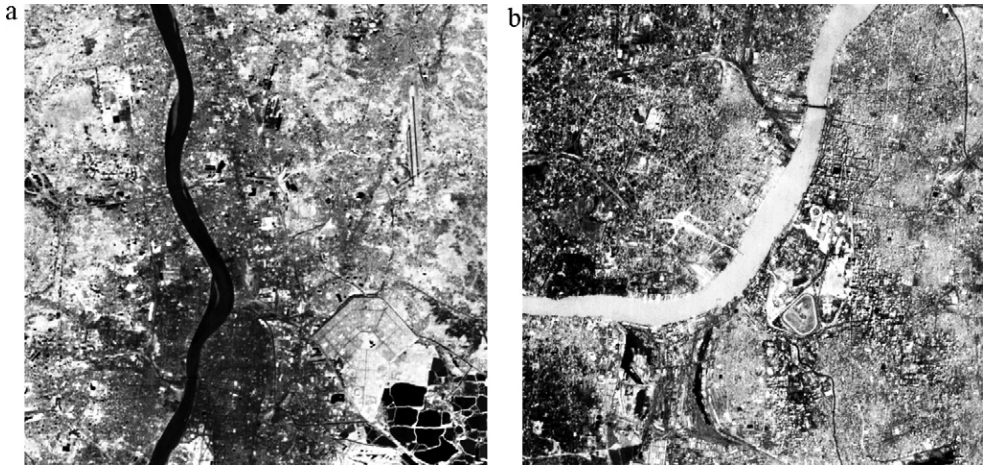

Fig. 9. Original (a) IRS-1A (band-4) enhanced image and (b) SPOT (band-3) enhanced image.

of gain of a classifier over other obtained with respect to *DB* value is also calculated similar to Eq. (4).

5. Results and discussion

For demonstrating the effectiveness of the proposed rough-wavelet granulation based feature selection model, we have used two real life multispectral (four-band) remote sensing images obtained from IRS-1A and SPOT satellites. These images bear different characteristics like spatial resolution, number of bands, and wavelengths, while they have similar land-cover classes. Along with this, a synthetic multispectral (four-band) remote sensing image is also used to validate our model.

5.1. Classification criteria

In the present investigation we have compared the performance of the proposed model with different combinations of wavelet granulation and rough feature selection methods. Five combinations of classification models are considered as mentioned below. Image patterns with its original feature representation are fed as input to these models:

- Model 1: k-nearest neighbor (k-NN with $k = 1$) classifier,
- Model 2: Class-independent (CI) wavelet granulation + k-NN (with $k = 1$) classifier,
- Model 3: Class-dependent (CD) wavelet granulation + k-NN (with $k = 1$) classifier,
- Model 4: CD wavelet granulation + PaRS based feature selection + k-NN (with $k = 1$) classifier,
- Model 5: CD wavelet granulation + NRS based feature selection + k-NN (with $k = 1$) classifier.

The comparative analysis of models is also made with different types of biorthogonal wavelets (e.g., bior2.2, bior2.4, bior3.1 and bior3.3) based granulations. Apart from the performance comparison with different quantitative measures for both real life and synthetic remote sensing images, the efficacy of the proposed model of rough-wavelet granulation and feature selection is justified with the following types of analyses. However, the experimental results with these analyses are provided only for IRS-1A image, because similar trend of comparative performance is observed for the remaining images:

- Variation of classification accuracy with different values of parameter Φ and distances used in NRS based feature selection for optimal value selection,
- Performance comparison of the proposed model with other classifiers such as k-NN with $k = 3$ and 5, maximum likelihood (ML) classifier and multi-layered perceptron (MLP).

5.2. Classification of images

5.2.1. IRS-1A image

In this section, we describe the performance comparison of different models with real life multispectral remote sensing images, namely, IRS-1A and SPOT images. Here the classifiers are initially trained with labeled data of six land cover types and then the said trained classifiers are applied on the unlabeled image data to partition into six regions.

The IRS-1A image (size 512×512) is obtained from Indian Remote Sensing Satellite [5,23,28]. The image has spatial resolution of 36.25 m $\times$ 36.25 m and wavelength range of 0.45–0.86 μm. The whole spectrum range is separated into four spectral bands namely, blue, green, red and near infrared corresponding to band-1, band-2, band-3 and band-4 of wavelengths 0.45–0.52 μm, 0.52–0.59 μm, 0.62–0.68 μm and 0.77–0.86 μm, respectively. Since the image is of poor illumination, we have presented the enhanced image (band-4) in Fig. 9a for the convenience of visualizing the content of the image. However, the algorithms are implemented on the original (poorly illuminated) image. The image in Fig. 9a covers an area around the city of Calcutta, India in the near infrared band having six major land cover classes: *pure water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS).

IRS-1A image is classified with five different models using k-NN classifier ($k = 1$), and the performance comparison in terms of β and DB values, is depicted in Table 1. As expected, the β value is the highest and DB value is the lowest for the training set (Table 1) compared to other models for both the images (IRS-1A and SPOT). In the present experiment, we have compared the performance of models with respect to five criteria, namely, (i) granulated and non-granulated feature space, (ii) different wavelets based granulation, (iii) class-dependent (CD) and class-independent (CI) wavelet granulation, (iv) wavelet and rough-wavelet granulated feature space, and (v) Pawlak's rough sets (PaRS) and neighborhood rough sets (NRS) based feature selection.

As described in Section 3.2.2, performance comparison with the NRS method of feature selection depends on the distance function Δ and parameter Φ of the neighborhood granules. In the present study we analyzed the performance of model 5 for the variation of

Table 1
Performance comparison of models using k-NN classifier ($k=1$) for IRS-1A and SPOT images with different wavelets ($\mathbf{p}=2$, $\Phi=0.30$).

Model	Wavelet	β value		DB value		T_c (s)
		IRS-1A	SPOT	IRS-1A	SPOT	
Training samples	–	9.4434	9.3654	0.5432	1.4656	
1	–	6.9971	6.8960	0.8005	2.8260	385.56
2	bior2.2	7.4001	7.2376	0.7914	2.6316	410.76
3	bior2.2	7.6911	7.5002	0.7631	2.5172	420.37
4	bior2.2	7.9083	7.8563	0.7434	2.3889	391.32
5	bior2.2	8.4113	8.2179	0.6913	2.3781	399.01
2	bior2.4	7.3435	7.1946	0.8001	2.7132	415.76
3	bior2.4	7.6812	7.4904	0.7732	2.6301	422.23
4	bior2.4	7.9176	7.7878	0.7556	2.4013	390.76
5	bior2.4	8.4383	8.2004	0.7040	2.3135	405.35
2	bior3.1	7.3981	7.2172	0.7984	2.6812	409.84
3	bior3.1	7.6733	7.4025	0.7642	2.5971	419.92
4	bior3.1	7.9122	7.8112	0.7485	2.4002	391.81
5	bior3.1	8.4001	8.2034	0.7003	2.3203	402.78
2	bior3.3	7.4026	7.2501	0.7523	2.6242	414.33
3	bior3.3	7.6997	7.5070	0.7225	2.5013	422.20
4	bior3.3	8.1001	7.8711	0.6838	2.3799	390.11
5	bior3.3	8.4567	8.2308	0.6485	2.3011	400.23

both these parameters. We plotted the β values (Fig. 10) of model 5 (using k-NN classifier ($k=1$)) for three $\mathbf{p}$-norm distances for a variation of Φ values ([0,1]) in Euclidean space. These are Manhattan distance ($\mathbf{p}=1$), Euclidean distance ($\mathbf{p}=2$) and Chebychev distance ($\mathbf{p}=\infty$). It is observed from Fig. 10 that the β value varies with Φ for all types of distances. With the increase of Φ value the β value increases at first, reaches to a peak and then decreases. Roughly for all the distances, the highest accuracy is obtained for $\Phi=[0.20, 0.35]$ with maximum for Euclidean distance. Beyond 0.45, the neighborhood rough set based model can not select the relevant features capable of distinguishing patterns. The reason is that with large neighborhood region, the granules accommodate more neighbors, thereby increasing the possibility of possessing irrelevant and contradictory feature information. Further it was seen that the numbers of selected features are different when Φ takes values in the interval [0.20,0.35], although these features are producing similar classification performance. Hence it appears that the value of Φ may be varied in [0.20,0.35] to find the minimal subset of features with similar classification performance. Accordingly, for presenting the results for the remaining data sets, we have taken $\mathbf{p}=2$ (Euclidean distance) and $\Phi=0.30$.

In a comparative analysis from Table 1, it is observed that the classifiers' performance with IRS-1A image, measured in terms of β values is better for the models using granulated feature space. For example, model 1 (without granulation) provides β value of 6.9971, whereas with other models (with granulation) the values are higher. This reflects the ability of better extraction of contextual information in time–frequency plane using wavelet granulation of feature space.

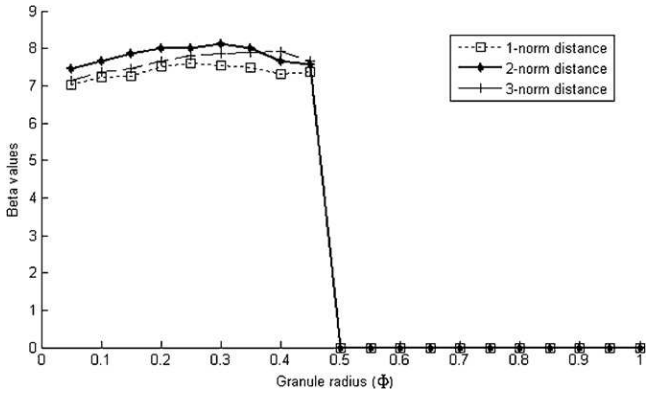

Fig. 10. Variation of β values of model 5 with the parameter Φ for three distances.

Performance comparison among different wavelets, biorthogonal3.3 (bior3.3) is seen to provide improved results compared to bior2.2, bior2.4 and bior3.1, as shown in Table 1. Table 1 reveals that the β value for IRS-1A image, as an example, obtained by model 2 using bior3.3 wavelet is 7.4026, which is higher than the values 7.4001, 7.3435 and 7.3981 using bior2.2, bior2.4 and bior3.1 wavelets, respectively. Similar trend of improvement with other models using bior3.3 wavelet is also observed from Table 1.

Performance analysis between CD and CI wavelet granulation based models (Table 1), the β value for model 3 (CD model) compared to model 2 (CI model) is higher. For example, with bior3.3 wavelet, model 3 provides a β value of 7.6997 whereas it is 7.4026 with model 2. Similarly, for other wavelets based granulation, β values with model 3 are higher than model 2 (Table 1). This clearly indicates that CD granules efficiently explored the class-wise dependency of features to classes and provided an improved class discrimination information responsible for enhanced accuracy.

In another comparison of models with different granular feature space, it is observed that models 4 and 5 (with rough-wavelet granulation) provided higher β values than models 2 and 3 (with wavelet granulation). For example, the β values obtained by models 4 and 5 using bior3.3 are 8.1001 and 8.4567, respectively, are higher than 7.6997 and 7.4016 obtained with models 2 and 3, respectively (Table 1). This justifies the superiority of the rough-wavelet granulation to wavelet granulation and it is true for all types of wavelets used here.

In a comparison of models with NRS and PaRS, it is observed from Table 1 that using bior3.3 wavelet, the β value for the proposed model 5 (8.4567), as an example, compared to model 4 (8.1001) is higher. This is true for all the four wavelet-granulation based models. This signifies that the NRS based feature selection method restores better local information from neighborhood granules that is helpful for improved performance. Thus comparing among the five models of pattern classification with all possible aspects, the proposed model (model 5) that explored and incorporated CD rough-wavelet granular feature space with bior3.3 wavelet and NRS based feature selection methods provided the best performance. As

Table 2
Percentage of gain (Eq. (4)) obtained with respect to β and DB values with IRS-1A and SPOT images (bior3.3 wavelet).

	β gain		DB gain	
	IRS-1A	SPOT	IRS-1A	SPOT
Model 5 over model 1	20.86	19.35	18.98	18.57
Model 5 over model 2	14.23	13.52	13.79	12.31
Model 5 over model 3	9.83	9.64	10.24	8.00
Model 5 over model 4	4.40	4.56	5.16	3.31

Table 3
Performance comparison of models with different classifiers for IRS-1A image ($\mathbf{p}=2$, $\Phi=0.30$, bior3.3 wavelet).

Model	k-NN ($k=3$)		k-NN ($k=5$)		ML		MLP	
	β	DB	β	DB	β	DB	β	DB
1	6.9910	0.8055	7.001	0.8032	7.0121	0.8002	7.1034	0.7984
2	7.4112	0.7514	7.4042	0.7489	7.4501	0.7501	7.5014	0.7300
3	7.6133	0.7200	7.6246	0.7302	7.5987	0.7132	7.7138	0.7001
4	7.9661	0.6788	7.9803	0.6912	7.9110	0.6802	8.0133	0.6619
5	8.4204	0.6411	8.4412	0.6501	8.4212	0.6431	8.5334	0.6305

a whole the gradation of performance of five models with any of the wavelets can be established with the following β relation:

$$\beta_{training} > \beta_{proposed} > \beta_{model4} > \beta_{model3} > \beta_{model2} > \beta_{model1} \quad (6)$$

We also calculate the percentage of gain with respect to β value (Eq. (4)) obtained by the proposed model over others using bior3.3 wavelet, and the results are depicted in Table 2. It is found from Table 2 that the proposed model obtained the gains of 20.86%, 14.23%, 9.83% and 4.40 over models 1, 2, 3 and 4, respectively, which is highly appreciable. The superiority of the proposed model is also validated with the DB index, as shown in Tables 1 and 2.

A comparative analysis with total computational time T_c (given by the sum of the training and testing times), as required by different models using k-NN classifier ($k=1$), is depicted in Table 1. The T_c values for both the images (IRS-1A and SPOT) are same because the number of training samples and pixel sizes (512×512)

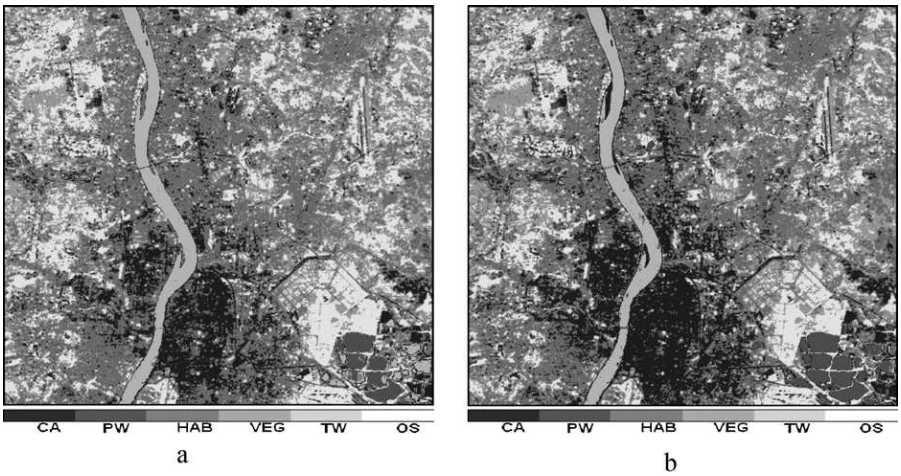

Fig. 11. Classified IRS-1A images with (a) model 1 and (b) model 5 (proposed model).

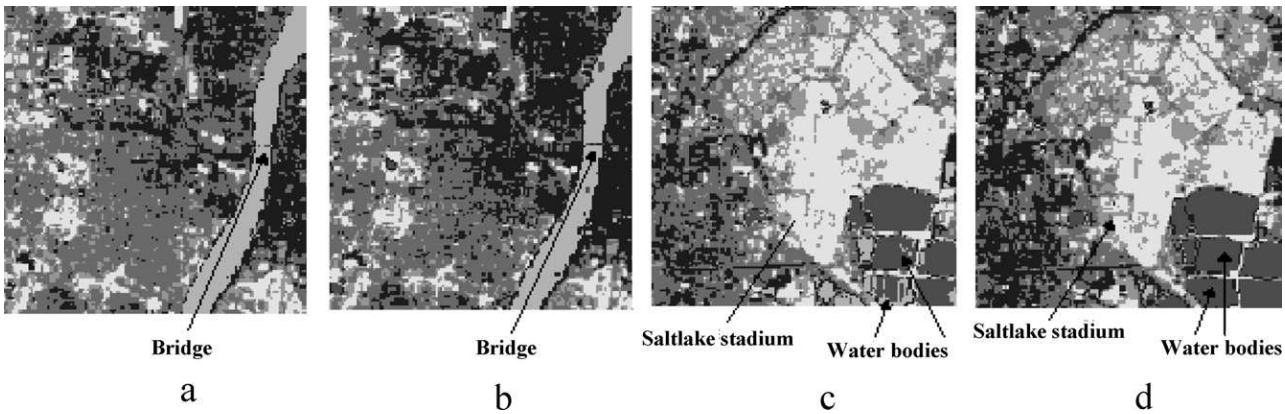

Fig. 12. (Zoomed) Two selected regions of classified IRS-1A image with (a and c) model 1 and (b and d) model 5.

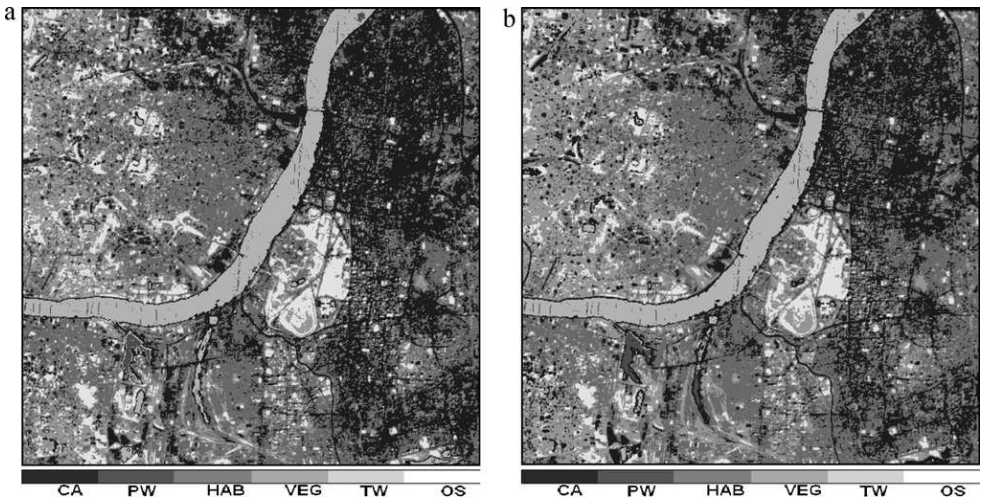

Fig. 13. Classified SPOT images with (a) model 1 and (b) model 5 (proposed model).

of these images are identical. All the simulations are done in MATLAB (Matrix Laboratory) environment in Pentium-IV machine with 3.19 GHz processor speed. It is seen for all the cases that the T_c values for wavelet granulated models (models 2, 3, 4 and 5) are higher than those of the non-granulated model (model 1), with improved performance. Interestingly, models 4 and 5 perform better and at the same time take less computational time than models 2 and 3 for all the cases. That means, the incorporation of the rough set theoretic feature selection step reduces the computation time. Further, the T_c values for model 5 (with NRS based feature selection), as expected, are little higher compared to model 4 (with PaRs based feature selection) at the cost of improved performance. Similar is the case between models 2 and 3, where CD granulation (model 3), as expected, is taking more time than CI granulation (model 2) for improved performance.

So far we have described the effectiveness of the proposed rough-wavelet granulation and feature selection model using k-NN ($k=1$) classifier. The effectiveness of the same model is also described using some other classifiers, e.g., k-NN ($k=3$ and 5), maximum likelihood (ML) classifier and multi-layered perceptron (MLP). The comparative results of all models with these classifiers are depicted in Table 3. The superiority of model 5 to others for different sets of classifiers is evident. Also similar improvement in performance of the models (using different classifiers) with granulated over non-granulated, CD over CI, bior3.3 wavelet granulation over other wavelet granulation and NRS based feature selection over PaRS, is observed, as in the case of k-NN ($k=1$) classifier.

In order to demonstrate the significance of granular computing visually, let us consider Fig. 11a and b depicting the output corresponding to model 1 (without granulation) and model 5 (with CD granulation and NRS feature selection) using bior3.3 wavelet, say. It is clear from the figures that the proposed model 5 performed well in segregating different areas by properly classifying the land covers. For example, the *Bridge* (*Rabindra Setu*) over the south part of the *river* is more prominent in Fig. 11b, whereas it is not so in Fig. 11a. A zoomed version of the said *bridge* region is shown in Fig. 12a and b to have a better visualization. Similarly, the regions such as *Saltlake stadium* and *water bodies* are more distinct and well shaped with model 5, as shown in Fig. 12d (zoomed version). These observations also justify the significance of the β and DB indexes in reflecting the performance of the models automatically without visual intervention.

5.2.2. SPOT image

The SPOT image (size 512×512) shown in Fig. 9b (enhanced image (band-3)) is obtained from SPOT satellite (Systeme Pour d'Observation de la Terre) [23]. The image used here has been acquired in the wavelength range of 0.50–0.89 μm. The whole spec-

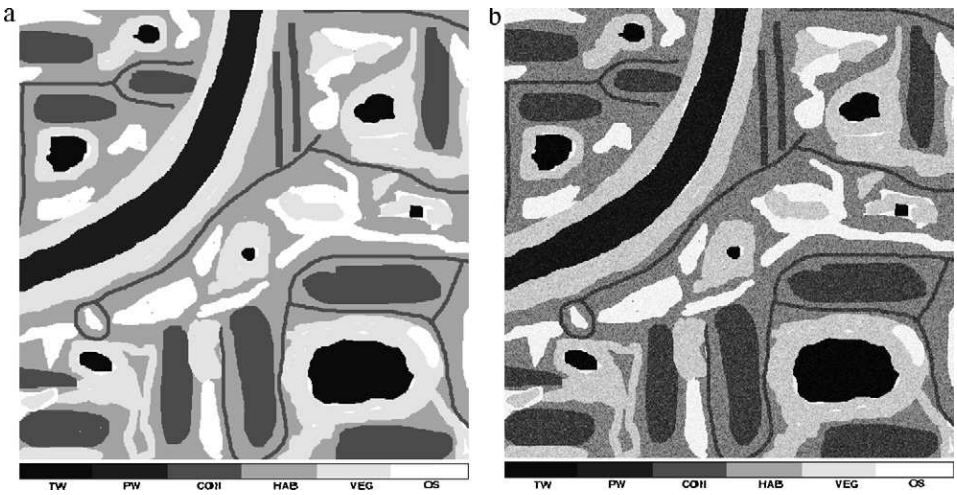

Fig. 14. Synthetic image (band-4): (a) original and (b) noisy ($\sigma=2$).

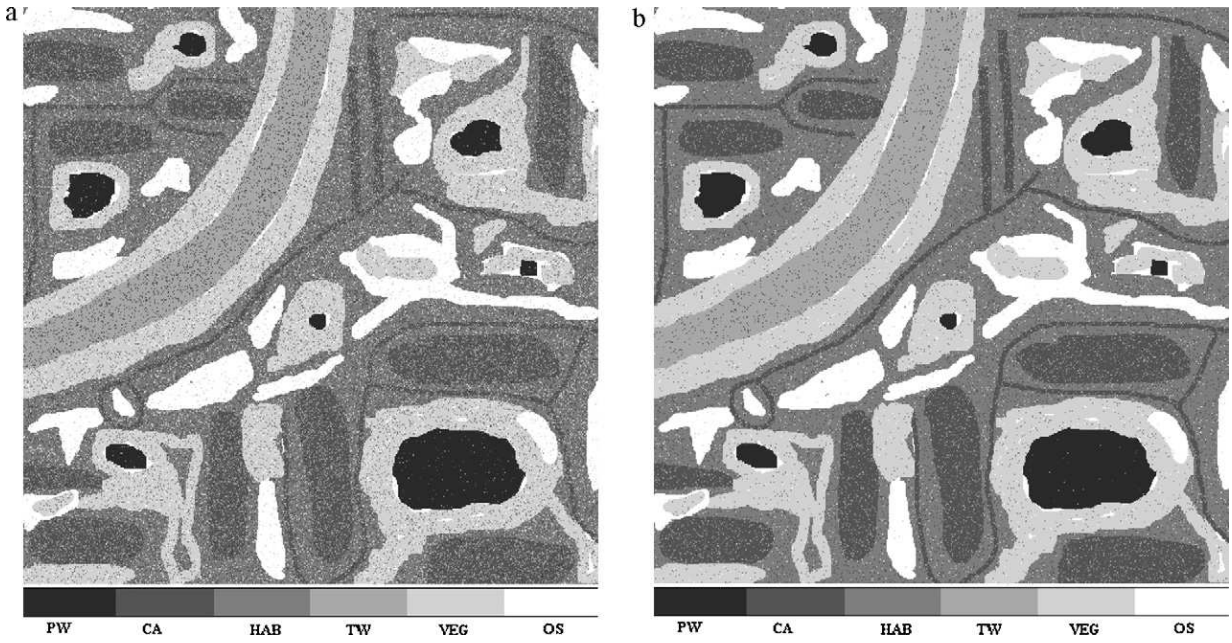

Fig. 15. Classified synthetic image (for $\sigma = 2$) by (a) model 1 and (b) model 5 (proposed model).

trum range is decomposed into three spectral bands namely, green (band-1), red (band-2) and near infrared (band-3) of wavelengths 0.50–0.59 μm, 0.61–0.68 μm, and 0.79–0.89 μm, respectively. This image has a higher spatial resolution of 20 m × 20 m as compared to IRS-1A. We have considered in our experiment the same six classes as in the case of IRS-1A image.

With SPOT image, the comparative results of five models using k-NN classifier ($k = 1$) in terms of β and DB values are shown in Table 1, which revealed the supremacy of the proposed model (model 5) and it is found to be most effective with CD granulation using bior3.3 wavelet. The significance of model 5 is further justified visually from Fig. 13 that illustrates the classified images corresponding to models 1 and 5. It is seen that some of the regions (e.g., *Garden Reach Lake* and *Race Course*) are well-structured and proper-shaped in Fig. 13b compared to Fig. 13a. For example, the shape and boundary of the *Garden Reach Lake* have come up much prominently in Fig. 13b. Similarly, the *Race Course* in Fig. 13b contains large grass region than in Fig. 13a.

5.2.3. Synthetic image

A four-band synthetic image (size 512 × 512) has been generated with six major land cover classes similar to the IRS-1A image. Fig. 14a shows the synthesized image in the near infrared range (band-4). All the five models are tested on the corrupted synthetic image. The synthetic image is corrupted with Gaussian noise (zero mean and standard deviation (σ) = 1, 2, ..., 6) in all four bands. Fig. 14b, as an example, shows the noisy version of the original image with $\sigma = 2$.

50% of the entire data are used as training set and the rest are considered as test set. Training set is selected randomly and an equal percent of samples is collected from each of the classes. We repeat these splitting sets for ten times and the final result is then averaged over them. For performance comparison the percentage of classification accuracy (PA) is calculated with respect to the original image (Fig. 14a).

The performance of five models using k-NN classifier ($k = 1$) in terms of percentage of accuracy (PA) for different σ is shown in Table 4 for 50% training set. The table revealed the superiority of model 5 to others for all the noise levels. Since similar trend of observation, as discussed in the case of IRS-1A image data, is obtained with other measures for the synthetic remote sensing image, we have not put those results here. Fig. 15 shows the resulting classified images obtained by models 1 and 5 for the noisy input image with $\sigma = 2$ (i.e., Fig. 14b). Superiority of model 5 to 1, as indicated in Table 4, is further verified visually from Fig. 15. Here we have shown the classified images obtained from these two models, as an example, because one of them performed the worst and the other performed the best.

Table 4
Classification accuracies (PA) of models using k-NN classifier ($k = 1$) for synthetic image with different σ at 50% training set ($\mathbf{p} = 2$, $\Phi = 0.30$).

Classification model	PA			
	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$
1	95.32	83.51	73.35	62.01
2	96.89	91.33	78.87	64.72
3	97.41	93.15	80.34	67.71
4	98.02	94.02	82.83	70.02
5	98.74	95.11	85.25	72.51

6. Conclusions

In the present article, we described a rough-wavelet model for land cover classification of multispectral remote sensing images. The model formulates a class-dependent (CD) wavelet granulation of input feature space, where the generated granules explore the dependency of features into different classes and make it more suitable for improved class label estimation. For the granulation process, we use shift-invariant wavelet, where the time–frequency plane explores the local/contextual information of pattern. Shift-invariant wavelet granulation provides translation invariant representation of features, which is an indispensable property in textural analysis (e.g., land cover classification of remote sensing images). The advantage of neighborhood rough sets that deal with both numerical and categorical data without any discretisation is also realized in the proposed model. The neighboring concept facilitates to gather the local/contextual information through neighbor granules that provide improved class discrimination information. It may be mentioned here that wavelet

granulation of feature space described in [29] for land cover classification, is similar to the method of class-independent granulation used here.

With extensive experimental results on both types of real life and synthetic multispectral remote sensing images, it is found that the proposed synergistically integrated model performs well with CD wavelet granulation using shift-invariant wavelet transform and neighborhood rough sets. The performance of the models with biorthogonal3.3 wavelet is further encouraging for the data sets with highly overlapping classes. A critical value of the threshold for various distances used in NRS, beyond which classification performance falls drastically, is also determined. Inclusion of rough set theoretic feature selection method not only increases the performance, but also reduces the computational time required for wavelet granules based classification.

Though the model is described here for multispectral remote sensing image classification, it can be used for the analysis of other spatio-temporal patterns wherever wavelet transform is effective.

Acknowledgments

The authors acknowledge the Center for Soft Computing Research: A National Facility, funded by the Department of Science and Technology, Govt. of India. The work was done while Prof. Pal held J.C. Bose Fellowship of the Govt. of India.

References

[1] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, Soft Computing 2 (1998) 23–25.
[2] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers, Boston, USA, 2003.
[3] A. Skowron, J.F. Peters, Rough-Granular Computing, in: W. Pedrycz, A. Skowron, V. Kreinovich (Eds.), Handbook of Granular Computing, John Wiley & Sons Ltd, West Sussex, England, 2008, pp. 285–328.
[4] W. Pedrycz, B.J. Park, S.K. Oh, The design of granular classifiers: a study in the synergy of interval calculus and fuzzy sets in pattern recognition, Pattern Recognition 41 (2008) 3720–3735.
[5] S.K. Pal, P. Mitra, Multispectral image segmentation using rough set initialized em algorithm, IEEE Transactions on Geoscience and Remote Sensing 40 (2002) 2495–2501.
[6] S.K. Pal, Computational theory perception (CTP), rough-fuzzy uncertainty analysis and mining in bioinformatics and web intelligence: a unified framework, LNCS Transactions on Rough Set 5946 (2009) 106–129.
[7] L.A. Zadeh, Fuzzy sets and information granularity, in: M. Gupta, R. Ragade, R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, North-Holland Publishing Co., Amsterdam, 1979, pp. 3–18.
[8] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.
[9] J.F. Peters, Z. Pawlak, A. Skowron, A rough set approach to measuring information granules, in: Proc. of Annual Int. Conf. on Computer Software and Applications, 2002, pp. 1355–1360.
[10] L. Huang, Y.-J. Ma, L. Guo, A rough set-based svm classifier for atr on the basis of invariant moment, in: Proc. of WRI Int. Conf. on Communications and Mobile Computing, 2009, pp. 620–625.
[11] M.M. Mushrif, A.K. Ray, Color image segmentation: rough-set theoretic approach, Pattern Recognition Letters 29 (2008) 483–493.
[12] Y. Wang, X. Liu, Z. Wang, W. Chen, Multispectral remote sensing image classification algorithm based on rough set theory, in: Proc. of IEEE Int. Conf. on Syst., Man Cybern, 2009, pp. 4853–4857.
[13] B. Tso, P.M. Mather, Classification Methods for Remotely Sensed Data, 2nd ed., CRC Press, Boca Raton, FL, 2009.
[14] R.M. Haralick, K.S. Shanmugam, Combined spectral and spatial processing of ERTS imagery data, Remote Sensing of Environment 3 (1974) 3–13.
[15] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed., Academic Press, USA, 1999.
[16] Z. Pawlak, Rough Sets – Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Netherlands, 1991.
[17] T.Y. Lin, Granulation and nearest neighborhoods: rough set approach, in: W. Pedrycz (Ed.), Granular Computing: An Emerging Paradigm, Physica-Verlag, Heidelberg, Germany, 2001, pp. 125–142.
[18] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Information Sciences 178 (2008) 3577–3594.
[19] S.K. Pal, A. Skowron (Eds.), Rough-Fuzzy Hybridization: A New Trend in Decision Making, Springer-Verlag, Singapore, 1999.
[20] S.K. Pal, L. Polkowski, A. Skowron (Eds.), Rough-Neural Computing: Techniques for Computing with Words, Springer-Verlag, Berlin, 2004.
[21] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley Interscience Publications, USA, 2000.
[22] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice Hall, New Jersey, USA, 1998.
[23] S.K. Pal, A. Ghosh, B. Uma Shankar, Segmentation of remotely sensed images with fuzzy thresholding, and quatitative evaluation, International Journal of Remote Sensing 21 (2000) 2269–2300.
[24] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 224–227.
[25] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
[26] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using wavelet transform, IEEE Transactions on Image Processing 1 (1992) 205–220.
[27] A. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorisation, Applied Artificial Intelligence 15 (2001) 843–873.
[28] NRSA, IRS data users hand book, Tech. Rep., 1989, Document No. IRS/NRSA/NDC/HB-02/89.
[29] S.K. Meher, B. Uma Shankar, A. Ghosh, Wavelet-feature-based classifiers for multispectral remote-sensing images, IEEE Transactions on Geoscience and Remote Sensing 45 (2007) 1881–1886.