

Soft Computing in Remote Sensing Applications

A. Senthil Kumar¹ · Anil Kumar¹ · R. Krishnan² · B. Chakravarthi³ · B. L. Deekshatalu⁴

Received: 15 July 2017/Revised: 11 August 2017/Accepted: 24 August 2017/Published online: 7 December 2017
© The National Academy of Sciences, India 2017

Abstract This review paper discusses recent developments in soft computing techniques and applications specific to remote sensing, especially in the last two decades. Even though, the applications have been spread across in many areas of remote sensing, utilization of these techniques has been found to be strongly biased towards remote sensing data classification because of their ability to handle mixed pixels and also varied spatial (ranging 1–250 m) and spectral resolutions (from 1 to 100 of spectral bands). Success with conventional methods for these problems has been only moderate. This has attracted more research in the soft computing data classification. In particular, the soft computing methods outperform conventional methods when they are applied to complex data handling especially for multi-source and very high dimensional data classification with one or more

combination of classifiers. Fairly recent soft computing concepts like deep learning, analytical hierarchical processes and feature extraction from hyperspectral imagery and applications of optimization techniques like particle swarm optimization, genetic algorithms in handling remote sensing problems have been covered to encourage research community for further exploration in such important topics.

Keywords Fuzzy C-means classifier · Support vector machines · Artificial neural networks · Genetic algorithms · Ensemble classifier

1 Introduction

Remote Sensing (RS) community has employed several computing technologies in the areas of satellite data acquisition, correction, registration, analysis, classification and decision support. Various application domains have used different techniques as well. Traditionally, classification of remote sensing has been done at a single pixel level assuming that a single pixel belongs wholly to one class. Also, historically only macro level classes were delineated in remote sensing data. Both these scenarios have changed radically with the advent of high resolution data and also the need to gather very finely differentiated classes. This has necessitated the use of more sophisticated techniques which are capable of separating these fine classes. These methods are essentially data driven, rather than model driven. The set of tools which have been used for this purpose broadly falls into the class of soft computing techniques. This paper will review these techniques with the focus on RS applications.

More specifically, we focus on different soft computing algorithms used in various remote sensing applications.

✉ A. Senthil Kumar
senthil@iirs.gov.in

Anil Kumar
anil@iirs.gov.in

R. Krishnan
drkdrk@gmail.com

B. Chakravarthi
bhagvatc@gmail.com

B. L. Deekshatalu
deekshatulu@hotmail.com

¹ Indian Institute of Remote Sensing, 4, Kalidas Road, Dehradun 248001, India

² Amrita University, Coimbatore, India

³ School of Computer and Information Sciences, University of Hyderabad, Hyderabad 500046, India

⁴ Distinguished Fellow, IDRBT (RBI-Govt of India), Masab Tank, Hyderabad 500057, India

Some of these algorithms include Fuzzy c -means (FCM), Possibilistic c -means (PCM), Artificial Neural Networks (ANN) and Support Vector machines (SVM), Random Sample Consensus (RANSAC), Swarm optimization and their variants. Initially, Soft computing was performed in a wide variety of geostatistical data analysis applications. Bezdek [1] first introduced it where he showed fuzzy partitions and prototypes for any set of numerical numbers. However, soft computing was initially not well-known in the field of geospatial technology as compared to other approaches. Continuous efforts have been made in the last two decades to improve the performance of soft computing algorithms over conventional classifiers, prompting the need of this review to bring out these efforts.

This review deals with publications mostly till 2016 in leading journals. Research papers relevant to different applications of soft computing in thematic applications is also included in this review. This review which is based on selected papers covers the wide range of soft computing in remote sensing: (1) applications from agriculture to urban growth, (2) proven soft computing algorithms to handle satellite imagery having resolution from 10 to 250 m, (3) spectral resolution from single to hundred bands, (4) comparative analysis with the conventional techniques used in the remote sensing and (5) accuracy of soft computing algorithms.

2 Overview of Soft Computing Methods

Soft computing encompasses various paradigms like Fuzzy classifiers, Artificial Neural Networks (ANN), Deep Learning, Compressive Sensing, Bayesian Networks, RANSAC, Swarm optimization and Analytical Hierarchical Process (AHP) like tools for Decision Support. These have been employed by the remote sensing community in the last decade or so. We describe these in the following sections.

2.1 Fuzzy Methods

Fuzzy classifiers capture the natural uncertainty and imprecision of class boundaries in remote sensing data. This is due to the heterogeneous nature of this data and also uncertainties emerging from the definition of classes as well as errors in measurement [2–4].

Fuzzy c-Means (FCM) is one of the earliest clustering algorithms based on Fuzzy set theory; data is characterized by membership values. Zadeh [5] introduced an idea to show the likeness as a point that shares with each class with a membership function which varies from zero to one [1]. The FCM uses the probabilistic constraint [6]. Fuzzy based classifiers uses fuzzy set theory, due to which these

are able to represent vague classes in a natural way. The FCM gives the degree of belongingness in the various clusters i.e., the degree to which a pixel will belong to a clusters [7]. The FCM assigns the membership to the multiple clusters, but it is relative to the total number of clusters, due to constraint conditions imposed on membership values which are also referred as hyper-line constraints [1, 6]. The FCM is an iterative process for the partition of pixels into different class membership values. Each pixel has a membership value which signifies the similarity between the pixel and the cluster [1]. For full description of the algorithm and its formulation, the reader is referred to the article by Krishnapuram and Keller [6]. The membership values generated from FCM and its by products, however, do not always represent to default concept of degree of fitting and closeness. According to earlier studies [6, 8], the FCM is found to be sensitive to noise and outliers in the data. In order to overcome these drawbacks, Krishnapuram and Keller [6] proposed an algorithm based on possibility theory given by Zadeh [9].

The modification in the objective function of the FCM is termed as *Possibilistic c-Means* (PCM). In this method, the probabilistic constraint has been relaxed. It is a clustering algorithm which can be applied as supervised mode while providing class centroid from training data [10]. The membership values generated through this objective function for each class are independent of all others. These membership values to a class indicate by what *degree it belongs* to a class [6, 10]. The degree of belongingness implies degree to which pixel belongs to a class and degree of typicality helps to differentiate a highly typical member of cluster from the moderately typical member of the cluster. The Bandwidth parameter term added in the objective function of FCM defines membership value to be 0.5 at a distance of a class, and is related to the typicality factor. Weight factor m controls the fuzziness in fuzzy based classifiers and its optimal value defers with different fuzzy classifiers [10, 11]. For full mathematical description, the reader is again referred to the research works of Krishnapuram and Keller [6, 11]. As the advancement in this fields proceeds, Wu and Zhou [8] pointed out shortcomings in the PCM approach. They argued that: (1) PCM is profound toward good initialization, (2) it has unwanted leaning to produce equivalent clusters that arise because of columns and rows of typicality matrix being independent of each other, and (3) typicality in PCM not only reduces the effects of noise, but also neglects the membership that makes a class centroid close to the data points. To overcome the drawbacks of the PCM, Zhang and Leung [12] proposed an improved version of PCM which they termed as Improved Possibilistic c -Means (IPCM) and Wu and Zhou [13] proposed a variant model to the PCM and termed it as Modified Possibilistic c -Means (MPCM). Very

little work has been done till now in remote sensing with IPCM and MPCM algorithms. They therefore hold good potential for future research.

2.2 Artificial Neural Networks

Artificial neural networks (ANNs) refers to a whole set of algorithms which have drawn inspiration from the functioning of the human brain. Atkinson and Tatnall [14] brought out an introductory article on neural networks in remote sensing based on research initiated by various groups in early 1990s. Many researchers have designed ANNs to solve problems in pattern recognition, optimization and prediction. Neural networks, approached from the perception of distinguishing analysis, provide a outlive for inevitably selecting the suitable form of edge and locating it [15]. An ANN system contains the set of nodes which are interconnected via set of weights, which allow information travel through the network serially or parallel [15]. ANN can be categorized into two groups: (1) feed-forward networks, in which graph have no loops, (2) recurrent or feedback networks, in which loop is present for the feedback mechanism. ANN is feed forward if while input data in the network moves from input layer to hidden layer to output layers [16].

An ANN multilayer perceptron (MLP) type showed in Fig. 1, consists of an input layer, hidden layers and a set of output nodes. MLPs have been used with back propagation algorithm, it has two phases: (1) feed forward pass, an input vector is applied to the network and propagated to the output; (2) during back propagation phase, error is

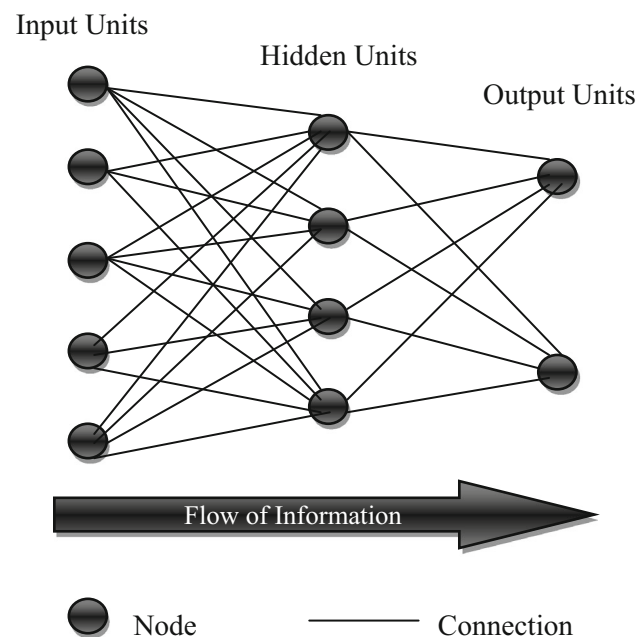


Fig. 1 Schematic diagram of multilayer perceptron type [15]

calculated while comparing target value with output; and weights are then adjusted in accordance with an error-correction rule [17]. Earlier, McCulloch and Pitts [18] had proposed a binary threshold unit as a computational model for an artificial neuron. The mathematical neuron computes a weighted sum of its n input signals and generates an desired output, with minimum target error [19]. Its Architecture and mathematical formulation are well described by [16, 19]. In case of remote sensing, ANN has been used in the areas of change detection, data fusion, land cover classification, spectral unmixing and site suitability for specific vegetation [20]. Due to structure form of ANNs, it is found to be more complex in comparison to conventional classifiers, also having chances to stuck up in loop [17]. For non-linear systems some modification has been implemented in neural network like Fuzzy neural network (FNN) and Probabilistic neural network (PNN). Both these methods employed the back propagation algorithm. In PNN, decision boundaries are modified in real-time using new data as they become available, and can be implemented using artificial hardware “neurons” that operate entirely in parallel [21]. This approach offers tremendous speed for problems in which the incremental time of back propagation is a significant fraction of the total computation time [21]. But the use of the FNN and PNN have still not made much progress in remote sensing. Hence this could be an area for further research.

2.3 Deep Learning

Deep structured learning works with multiple layers of non-linear information processing as applied in the areas such as; classifications, feature extraction and pattern analysis [12, 22, 23]. It is also well known as hierarchical learning because learning is based on several levels in which higher level concepts are defined from lower level ones, and wise versa. Basically the deep learning concept has two key aspects: (1) it consists of multiple layers or stages of non-linear information, (2) for the feature representation method of supervised and unsupervised learning. LeCun et al. [24] proposed in their review paper that deep convolution nets or Convolution neural network (CNN) as a breakthrough in processing of images, audio, videos. Lawrence et al. [25] brought out a study on CNN, in which they represent a hybrid neural network for face recognition. Before applying the CNN, first image sampling by two methods is done as described by Lawrence et al. [25]: (1) Using intensity values create a vector from local window at each point in the window, (2) local sample is represented by forming a vector from intensity using centre pixel and difference in intensity between the centre pixel and all other pixels within the square window. The CNN incorporates multiple feature extraction stages which is a

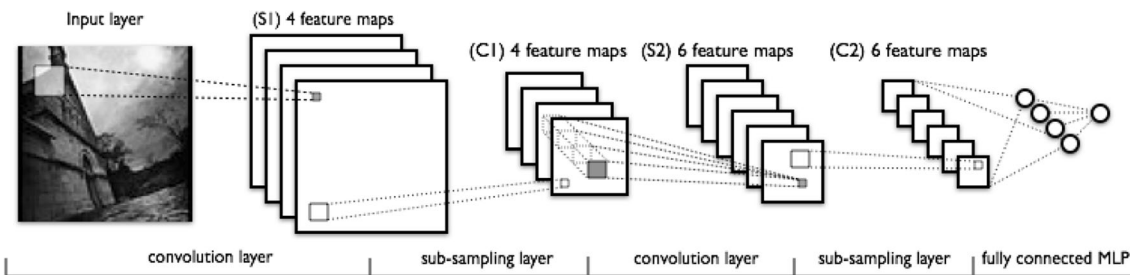


Fig. 2 A typical convolution network (from www.deeplearning.net/tutorial/lenet.html)

multilayer architecture (see Fig. 2). These phases consists of three layers: (1) a *convolution or filter bank layer*, (2) a *non-linearity layer and* (3) a *feature pooling layer*. A three dimensional array with r two dimensional feature maps of size $m \times n$ is given as input to the convolution layer. A three-dimensional array composed of k feature maps of size $m_1 \times n_1$ is output [26]. The Non-linearity layer consists of a pointwise nonlinearity function which is applied to each component of the feature map. The output is a commonly rectified linear unit. In the pooling layer, maximum selection operation is applied within a small spatial region of each feature map. Pooling layer consists of a grid of polling units spaced s pixels apart [26]. The entire network was trained with back propagation with a supervised loss function. The concept of deep learning is still to be implemented widely in the field of remote sensing. Mostly, deep learning or CNN concept were so applied on the test images or for face detection.

2.4 Bayesian Networks

The so-called naïve Bayesian network is one of the most effective classifiers, with its predictive performance is competitive with state-of-the-art classifiers [27]. Conditional probability of each attribute A_i gives class label C while classifier learns from training data. The class was predicted with highest posterior probability while classification was applied using sBayes rule to compute probability of C , using instance of A_1, \dots, A_n , Bayesian network based classifiers have many advantages theoretically, but their overall performance is not as good as the discriminative classifiers like support vector machines [28, 29]. A Naïve Bayesian network (NB) is a simple but effective Bayesian network classifier [30]. When NB is used in real applications, it first partitions the data into various sub datasets by the class label. Each sub dataset is labelled by C_i , the maximum likelihood (ML) estimator given by number of occurrences of the event in the sub dataset C_i . The great advantage of NB is its ability to deal with missing information [30]. The naïve Bayesian classifier is shown in Fig. 3. In NB, first model the joint probability in

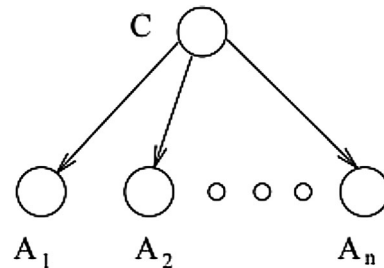


Fig. 3 The structure of the naïve Bayesian network [27]

each subset separately and then apply the Bayes rule for the posterior classification rule. But it discards the discriminative information. It tries to approximate the information in each subset, while not considering the other classes of data [30]. To overcome this problem use of posterior probability model in place of joint probability model is suggested. But this kind of computation for the Bayesian network makes it hard to perform the optimization [30]. Heckerman, Geiger, and Chickering in 1995, describe learning of Bayesian networks in which they combine the user knowledge with the statistical data. Two most important properties were found, namely, equivalence and parameter modularity. They conclude that combining these properties, simplifies the encoding of user's prior knowledge. A user can express prior knowledge for the most part as a single prior Bayesian network for the domain [31]. Cooper and Herskovits [32] have done the work on Bayesian method for construction of probabilistic network from the databases. They have extended the basic method to handle missing data and hidden variables. The different kind of Bayesian models used to classify remote sensing data are: (1) Naïve Bayes (NB), (2) Tree Augmented Naïve Bayes (TAN) and General Bayesian Network (GBN) presented by Solares and Sanz [33]. The implementation of Bayesian model applied in various applications i.e. Medical diagnosis, Clinical decision support, complex genetic support, crime risk factor analysis, and Image classification. But implementation of Bayesian network as a classifier in the field remote sensing is still a biggest challenge.

2.5 Other Soft Computing algorithms

Other well-known soft computing approaches are *Support vector machines* (SVMs) and Analytical Hierarchical Process (AHP). The SVM is by far the most widely explored method in the field of remote sensing in various applications. The SVM is a statistical learning approach with no assumption about data distribution [34]. It was originally proposed by Vapnik [35]. He represented the method with a set of labelled data instances and the SVMs algorithm was developed in such a way that it aims to find a hyperplane that separates the dataset into discrete predefined number of classes [34]. Later some developments in statistical learning were seen in which a modified form of SVMs has been developed, namely, Transductive SVM (TSVMs). TSVMs explores iterative algorithms which gradually search a separating hyperplane using kernels with a transductive process while incorporating both labelled and unlabelled samples in the training phase [36]. Bruzzone et al. [36] describe their proposed technique on TSVMs which is based on a novel transductive procedure, weighing strategy effects of suboptimal model selection and address multiclass cases. Pal and Mather [37] compared the SVM with Maximum likelihood and ANN. He found that SVM can use small training datasets and high-dimensional data. Pixel unmixing of moderate resolution SPOT satellite geometric images was done using pairwise coupling SVM by Li et al. [38, 39]. The SVMs with boosting were used for multiscale classification of RS images [40].

The soft computing tool which is widely in use now days is *Analytical Hierarchical process* (AHP). The AHP structure is well described by Bhushan and Rai [41]. Bhushan and Rai [41] fester the problem into a hierarchy of small problems to easily subjectively evaluate using AHP. Set of pairwise comparisons was used to derive relevant data [42]. Weights play an important role in formulating decision criteria and relative performance measures in term of each individual decision criterion. Beynon [43] has done the mathematical analysis of Dempster-Shafer theory with AHP (DS/AHP), and constructed the functional form of the preferred weights. This method evaluates the range of uncertainty which is expressed by the decision maker. Due to its decision making capabilities, it has been widely used in the field of remote sensing especially in GIS. A new decision making model now popularly known as Fuzzy AHP has emerged. Kahraman et al. [44] have described in their paper the role of fuzzy AHP for multi-criterion objective functions. They have found satisfactory results for the criteria determined. Another AHP-fuzzy approach developed by Hong-feng [45], focuses in the field of remote sensing. For the object level they have taken the regional stability and criterion level defined by crust stability and land surface stability. The has been used to

provided references for regional planning of mid and long range planning, selecting site for important projects and immigrant in Chongqing city [45].

Random Sample Consensus (RANSAC) is yet another interesting soft computing based model. Fischler and Bolles [46] developed a new paradigm of RANSAC for fitting a model, which was capable of smoothing and interpreting data which contain major percentage of gross error. RANSAC uses minimal initial data set as feasible and tries out different randomly selected subsets of data. This enables elimination of outliers. RANSAC implementation is well described in their paper in which they attribute the application of RANSAC to Location Determination Problem. Yaniv [47] worked on RANSAC algorithm for robust parameter estimation which has been applied to wide variety of parametric entities. RANSAC Algorithm developed such a way that its can ignore outlying data elements found in input. Required components were implement to illustrate use of algorithm for estimating parameter values of a hyperplane and hyper sphere [47]. RANSAC has found applications in feature extraction [48–50].

Kennedy and Eberhart [51] explain another soft computing optimization algorithm which is *particle swarm optimization*. Basically it was introduced by them for optimization of non-linear functions using particle swarm optimization technique (PSO). While using primitive mathematical operators, it is inexpensive in terms of memory requirements and speed. They describe its five principles of swarm intelligence which are: (a) proximity principle, (b) quality principle, (c) diverse principle, (d) stability principle, and (e) adaptability principle. Run- kler and Katz [52] worked on minimizing method of Fuzzy *c*-Means model. Particle swarm optimization (PSO) was introduced for minimizing FCM objective function. Alternating optimization (AO) and with ant colony optimization (ACO) were compared by their method developed for PSO. The stochastic methods ACO, PSO–V, and PSO–U are slower than AO [52]. Permana and Hashim [53] propose a method to generate fuzzy membership automatically and used PSO as an optimizer. PSO automatic adjusts membership functions. In their method they found fuzzy system performance after generation showed better results than before generation. Buckley and Feuring [54] worked on evolutionary algorithm which provides solution for fuzzy problems. It has been proven that whole non-dominated set of multi-objective fuzzy linear programming has been explored using fuzzy flexible programming. An evolutionary algorithm designed to solve fuzzy flexible program. Hoffmann [55] developed evolutionary algorithm for fuzzy control system design. Evolutionary algorithm are based on performance index which adjusts the membership functions or scaling factors of predefined fuzzy

controller. Tuning of scaling and membership functions of a fuzzy cart-pole balancing controller was done using evolution strategy, and a genetic algorithm that learns the fuzzy control rules for an obstacle-avoidance behaviour of a mobile robot [55, 56]. Eiben et al. [57] worked on parameter control in evolutionary algorithms, they have found it potential of adjusting the algorithm. Herrera and Lozano [58] adapt a new soft computing algorithm which is genetic algorithm which is based on fuzzy logic controllers. They studied this technique deeply based on fuzzy logic controller. They have developed adaptive real coded genetic algorithm bases on fuzzy logic controllers, and have obtained suitable results.

Genetic Algorithm (GA) is a random optimization technique inspired by the theory of evolution and survival of the fittest. The GAs are iterative stochastic methods that can be used to solve search and optimization problems. There have been many applications reported on the use of GA in remote sensing data processing, image fusion and classifications problems. Garzelli and Nencini [59] applied GA for fusion of very high resolution panchromatic and multispectral images. In image fusion, the MS bands data are to be injected with spatial details of PAN data from representation of the PAN data. The GA was employed to determine the gains that maximize an image quality score index needed to reduce distortions in the output fused image. Yao and Tian [60] have proposed a GA-based selective principal component analysis method for high dimensional data feature extraction from airborne hyperspectral sensor imagery of 60 bands. Singhai and Singhai [61] used GA for optimizing mutual information that measures information redundancy in intensities of floating and reference images.

As can be inferred, soft computing has found numerous applications in remote sensing. For this review, we have consulted more than hundred papers with half of them pertaining to the last 5 years. The fuzzy based soft computing has found a wide range of remote sensing application domains, while the other algorithms of soft computing like ANN, CNN and Bayesian Network still have to find significant applications in remote sensing.

3 Research in Soft Computing on Remote Sensing Data Classification

This section summarizes the soft computation advancement during the past decade. Papers which compare the performance of soft computing as well as incorporating soft computing for a specific application are discussed in the next section.

3.1 Classification Based on Fuzzy Theory

Though fuzzy classification of remotely sensed data has been applied for both supervised and unsupervised discrimination of Earth features, majority of applications has been employed for supervised fuzzy classification. The superiority of fuzzy based classifiers over crisp traditional methods has been reported by many researchers, even with limited quantity of training samples. Wang [4] focused on the problem of conventional classifier as they did not consider the class mixture problem, which led to poor extraction of information. Landsat MSS image were used for the fuzzy supervised classification in which geographical information was represented as fuzzy sets. Two major steps followed by his classifier are: (1) fuzzy parameter from fuzzy training data, and (2) fuzzy partition of spectral space. As a result he found partial membership gives more accurate statistical parameters which in turns provide higher accuracy. Foody and Cox [62] worked on sub-pixel land cover composition using fuzzy membership functions. They focussed on coarser resolution data because the probability of getting mixed pixels was more from the coarser resolution. Two approaches were followed: (1) linear mixture model and (2) regression model based on fuzzy membership functions. Significant correlations were found above a threshold (0.7 in their study) between actual and predicted proportion of land. In another case study presented on tropical forest, Foody et al. [63] found accuracy significantly increased through the use of sub-pixel estimates of land cover.

Fuzzy based classifiers were also employed for both spatial and spectral information for the discrimination between road and building features, because both these features are spectrally near-similar urban land cover classes. Image was segmented using both spectral and spatial heterogeneity to facilitate further object based classification. Shackelford and Davis [64], in their paper represented object-based approach for urban land cover classification using fuzzy based classification approach. They used high resolution MSS image of IKONOS, for which they combined the pixel/object approach. Using this, the authors reported that this technique was suited to extract buildings, impervious surface and roads in dense urban areas with 77, 81 and 99% classification accuracies, respectively. Bárdossy and Samaniego [65] studied fuzzy rule based classification of remotely sensed images. They used Landsat TM scene for the classification. While using simulated annealing as an optimization algorithm, fuzzy classification algorithm can be used with a rule systems. No prior knowledge is required of the rules. In their study, they found significant correlation between the membership values and the percentage of coverage within pixels. Wendling et al. [66] split images into tress of fuzzy

regions. To define the fuzzy regions gradient inverse function was applied with the basic grey level image. Topological features were computed which consist of fuzzy regions. With the help of fuzzy segmentation algorithm a set of sample trees were achieved. Cannon et al. [67] worked on segmentation of thematic mapper using Fuzzy *c*-Means algorithm. They follow a segmentation procedure that utilize a clustering algorithm based on fuzzy set theory. The segmentation uses fuzzy *c*-means in two stages. Large number of clusters resulting from this segmentation process were merged by using similarity measures on the cluster centres. In their study, they found this two stage process was able to separate corn and soybean and several minor classes. Zhang and Foody [68] worked on fully supervised fuzzy approach. They have found fully-fuzzy approach may be deemed more objective and correct than partial-fuzzy approach, when fuzziness is accommodated in one or two of three classification stages [68]. In their paper they focused on two approaches; fuzzy *c*-means algorithm with supervised approach and an ANN approach. Their results confirm the superiority of fully-fuzzy based approach over partial-fuzzy classification, which further gives more relaxation for training samples. Gopal et al. [69] studied global land cover from AVHRR dataset using fuzzy neural network classification approach. They used annual composite normalized difference vegetation index (NDVI) values from AVHRR classified with maximum likelihood classifier and later same data has been classified with fuzzy ARTMAP. Classification accuracy was more than 85% compared to 78% using maximum likelihood classifier, when fuzzy ARTMAP has been trained using 80% of the data and tested on the remaining 20% of the data. This study showed fuzzy neural network as an alternative for global scale land cover classification. Solaiman et al. [70] describe how fuzzy concepts can be used with multisensor data fusion. Land cover classification using ERS-1/JERS-1 SAR data was done using fuzzy based data fusion. Using classes and prior knowledge Fuzzy membership maps of different thematic classes were calculated. FMM's is the iterative process which updates using spatial contextual information. They found three advantages of their classifier viz., (1) due to fuzzy concepts it has flexibility of integrating multi-sensor/contextual and prior information, (2) classification consists of thematic as well as confidence map and (3) confidence map evaluates the classification process complexities. Foody [71, 72] has done vegetation mapping using fuzzy modelling. Number of times vegetation model distribution used in an image classification, however, may not always be appropriate as generates a 'hard' class allocation. He ran the algorithm on airborne thematic mapper (ATM) data. Outputs were generated using three classification techniques: maximum likelihood, ANN and fuzzy sets. The output of hard

classifications such as the maximum likelihood classification and artificial neural network can be "softened" to provide more realistic ground information. Furthermore degree of fuzziness can be modulated by fuzzy sets techniques, such as the fuzzy *c*-means algorithm [71, 72]. Benz et al. [73] did object oriented analysis of remote sensing data for GIS ready information. They explain the principle strategy of object oriented analysis with the combination of fuzzy data.

Some studies of fuzzy classifiers have been attempted on temporal data for specific crop, post-earthquake identification and identification of moist deciduous forest. Some of these are discussed in this section. Musande et al. [74] worked on Cotton crop discrimination using Fuzzy Classification approach. They have used temporal data because for the mapping of specific crop temporal data was found to be extensively used. Five spectral indices i.e. simple ratio, NDVI, Transformed NDVI (TNDVI), Soil adjusted vegetation index (SAVI) and Transformed vegetation index (TVI) were investigated on AWIFS, LISS-III of Resources at-1 satellite data. The Possibilistic fuzzy classifier was used to handle the mixed pixel or uncertainty in data. In their work, they have found SAVI indices with dataset-2 outperformed than other indices for the discrimination of Cotton crop. Misra et al. [75] worked on mapping specific crop sugarcane using temporal approach. They have discussed the problem of using single date imagery data for the specific crop identification. They have used temporal data of LISS-III and AWiFS sensor data with the Possibilistic fuzzy classifier (PCM). This fuzzy classifier has found to extract the single class sub-pixel information [75]. As a result, they were able to extract specific crop i.e. sugarcane with the accuracy of 92.8%. Upadhyay et al. [76] worked on effect of World View-2 multispectral add-on bands using soft classification approach for specific crop mapping. For this study, world-view-2 multispectral single as well as two date datasets were used. In this study soft Possibilistic fuzzy classification approach, with class based sensor independent spectral band ratio NDVI index was used. From this study it was concluded that existing bands five, seven and new bands four, six, eight in World View-2 are important for identifying and mapping crops mentioned in this study [76]. Kumar et al. [77, 78] have done the work on Automatic Land cover mapping (ALCM). The aim of this study was to extract single land cover class that is water from mixed pixels present in multiple dataset of AWiFS sensor satellite. They have found PCM able to extract single class with 93.7 and 97.1% accuracy [79]. Kumar et al. [80] have done land cover mapping using fuzzy *c*-Means classifier and density estimation. They have described the problem of occurrence of mixed pixel in the data. A fully fuzzy concept has been implemented by them using density estimation using SVM (D-SVM) and FCM approach. A comparison of method found D-SVM function

using a Euclidean norm yields the best accuracy. Kumar and Dadhwal [81] have investigated entropy based fuzzy classification using uncertainty variation across spatial resolutions. In their study, they have used FCM as a base soft classifier in which entropy parameter has been added. For their research they have used Resourcesat-1 (also known as IRS-P6) dataset from AWIFS and LISS-III used for classification, while LISS-III and LISS-IV sensor have been used as a reference data. From this study it has been observed that output from FCM classifier has higher classification accuracy with higher uncertainty but entropy based FCM classifier with optimum value of regularizing parameter generates classified output with minimum uncertainty [81]. Nandan and Kumar [82] worked on wheat crop identification using hyper-tangent kernel based Possibilistic classifier. They introduce kernels with the Possibilistic classifiers to handle the non-linear classes. The Hyper tangent kernel with Possibilistic classifier was applied to temporal dataset of Formosat-2 and Landsat-8. They have found 5 date combinations was sufficient to discriminate between early harvested and late harvested wheat crops. Another study carried out by [83] worked on wheat monitoring using different kernels with Possibilistic classifiers. Temporal Formosat-2 dataset was used to fill the temporal gaps incorporating Landsat-8OLI data. They have found KMOD and polynomial kernel was found to be effective for separating wheat crop data. They have done separability analysis to optimize the temporal date combination while using temporal indices data. From this study, the datasets representing Sowing, Flowering and Maturity phenological stages of wheat crop were found more suitable [84].

Sengar et al. [85, 86] worked on post-Earthquake built-up damage identification using fuzzy approach. They have chosen the study area of Kashmir (Himalayan region) which was shaken by an earthquake of 7.6 magnitude. They have used remote sensing as a tool to study the damage assessment in built-up area. Temporal images of IRS P6 LISS-IV pre and post-earthquake were used with five spectral indices to identify built-up damage using supervised Noise cluster (NC) classifier. Another disaster study done by Sengar et al. [85, 86] shows that in Bhuj earthquake induces soil liquefaction. Using temporal Landsat-7 data, soil liquefaction identification was identified while using class based sensor-independent (CBSI) spectral ratio along with PCM, Noise cluster (NC) and Noise cluster with entropy (NCE) classifier. It was observed that identification of liquefied soil areas while separating with existing water body in that area, CBSI-based temporal indices provided better results [85, 86].

Another work using soft computing on moist deciduous forest identification was done by Upadhyay et al. [87]. They have used seven date temporal MODIS data for the moist deciduous forest and for the reference AWIFS data

were used. Different indices were applied before fuzzy classification namely Simple ratio, NDVI, SAVI and TNDVI. Possibilistic fuzzy classifier was used to handle the mixed pixel. The overall accuracy found after the fuzzy classification was 96.731%. Upadhyay et al. [79] studied the identification of moist deciduous forest using MODIS temporal indices data. They have used fuzzy based noise clustering approach on temporal MODIS data. It has been found that temporal data set representing to three dates, yields the highest overall accuracy assessment from all accuracy assessment techniques.

Some issues in the fuzzy classifiers which deals with the contextual information are also discussed by some of the researchers. Study of hybrid fuzzy approach for remote sensing image classification was conducted by Harikumar et al. [88]. Incorporation of contextual information through MRF into the fuzzy noise classifier (FNC) has been studied by them. The paper concludes that incorporation of spatial contextual information into the fuzzy noise classifier helps in reducing the noise achieving more accurate classification of satellite images [88]. Dutta et al. [89] researched on issues in contextual Fuzzy *c*-Means classification of remotely sensed data for land cover mapping. They have compared the hard and soft contextual classification with Metropolis algorithm for the better performance. They have found soft contextual classification fails to sample random field efficiently due to high dimensionality of soft output. Kumar et al. [77, 78] focused on some of the findings related to Sub-Pixel classification using HYSI sensor data of IMS-1 satellite. The findings of this research demonstrate that uncertainty estimation at accuracy assessment stage while using single and composite operators, overall maximum accuracy was achieved, while using 40 (13–52 bands) band data of HYSI sensor. In another study, Singha et al. [90] focused their work on importance of discontinuity adaptive Markov random field (DA-MRF) model for contextual fuzzy *c*-Means (FCM) classifiers. The results showed that DA-MRF model with FCM found to be better performance than other MRF models which showing an improved overall accuracy. In another recent study, Ghosh et al. [91] combined k-means (KM), partitioning around medoids (PAM), and fuzzy *c*-means (FCM) while using different cluster sizes for classifying land cover types applying GLAS derived parameters. The overall accuracy (89.41%) of all methods were quite significant with classes like; forest, mango orchard and other rest of classes.

3.2 Classification Based on Artificial Neural Network (ANNs)

ANN is one of the popular tools in the analysis of remotely sensed data. Paola and Schowengerdt [92] and later Atkinson and Tatnall [14] published excellent reviews on

the use of back-propagation ANN for classification of remotely sensed multispectral imagery based on research in early 1990s. ANN classification has been used more commonly in land cover, unmixing and retrieval of biophysical parameters of cover. Heermann and Khazenie [93] have done the classification of MSS remotely sensed data using back-propagation neural network. Methodology was developed in such a way so that it can select both training parameter and data sets for the training phase. The results were compared with other three algorithms: (1) a statistical contextual technique, (2) a supervised piecewise linear classifier and (3) unsupervised approach. They found back-propagation neural network was more feasible for classifying satellite images. They also found some drawbacks of using back-propagation neural network i.e. training time to a reasonable level. Foody et al. [94] have done the classification of remotely sensed data by ANNs. Feed forward which used a variant of the back-propagation learning algorithm was used to classify agriculture crops. After classification, it has been found that ANN appears to characterize the classes better than the discriminant analysis with accuracy up to 98%. The result also showed the independency of two classification techniques on representative training samples and normally distributed data. Erbek et al. [95] studied the comparison between maximum likelihood and artificial neural network algorithm for land use activities. They focused on the problem of Istanbul which was in extreme pressure from urban development due to a rapid increase in population. Due to complex pattern of urban land areas it was difficult to map such a city. One technique they come up with was ANNs. They test ANNs and compare it with maximum likelihood classifier on Landsat TM data. They found in their study that although ANNs take longer time than conventional classifiers they give better results. It shows the great potential for change detection. Hong et al. [96] uses the ANN approach for estimation of precipitation. The algorithm described by them were *Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN)* for cloud classification system (CCS) based on local and regional cloud features from geostationary imagery. The algorithm processes satellite cloud images into pixel rain rates by (1) separating cloud images into distinctive cloud patches, (2) extract cloud features, (3) clustering cloud patches into well organised subgroups and (4) calibrating cloud top temperature and rainfall [96]. Gopal et al. [97] detected the forest change using artificial neural network classifier. To analyse the phenomenon of the prolonged drought in the Lake Tahoe basin in California they used multi-temporal data. Rather than the conventional classifiers, they have used ANNs and multilayer feed-forward network architecture. The study estimates the conifer mortality more accurately than the

other approaches. ANN models have been used as a viable alternative for change detection in remote sensing. Dal and Khorram [98] studied change detection in the land cover which were based on ANNs. Neural network algorithm was developed and implemented for an automated land-cover change-detection system USING multispectral image. They developed a four layer trained network which provide categorical information about nature change and detect land cover changes with an overall accuracy of 95.6%. Moody et al. [99] studied ANN on coarser resolution satellite data. They worked on both simulated and real data with feed forward NN model based on MLP structure and trained using back-propagation algorithm. They found overall accuracy increases from 62% to 79% when mislabelled pixels, reclassify using second largest network output. The accuracy increased to 84% if for mislabelled pixel, the second largest sub-pixel class is used as a reference. In their result they conclude that interpretation of complete set of network output can provide information on relative proportion of sub-pixel classes. Warner and Shank [100] have evaluated the potential for fuzzy classification of MSS data using a fuzzy classification that employs back-propagation ANN. They showed two methods to improve NN for fuzzy classification which included modification in functions. The compound linear sigmoid function reduced the overall error and tended to make error more uniform between various proportions of mixed classes.

Ensemble or Meta-Classifiers: A number of papers used combination of multiple classifiers to produce classification outputs from remotely sensed data in the literature. Schematic of a typical ensemble classifier is shown in Fig. 4. This combined classifier (also referred to as an ensemble classifier) has been found to be generally more accurate than any of the individual classifiers [101]. Kumar and Majumder [102] worked on information fusion using tree classifiers. They have worked on three methods which are maximum likelihood (MLC), back propagation neural network and a combination MLC and back propagation

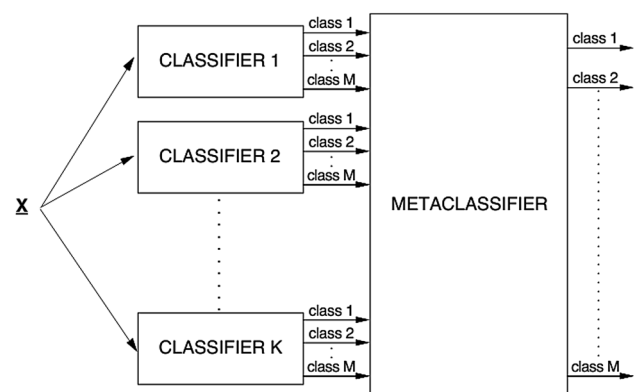


Fig. 4 A basic scheme for an ensemble classifier [105]

neural network. Motivation behind this fusion was to enhance interpretation of a particular pixel having minimum uncertainty in assigning the pixel to one of desired classes [102]. As a result they found that fusion technique was found to be better than the individual classifiers. With fusion using the fuzzy integral their recognizing ability in the presence of additive noise was found to be better [102]. Another research carried out by Kumar et al. [103] using multiple neural networks and fuzzy integral by Sugeno for robust classification of multispectral data [104]. The fuzzy integral was employed for integrating original data with its smoothed version. In both cases, the ensemble ANN combine out performed to that of their individual classification performance [103]. Giacinto and Roli [105] proposed to use a combination of neural network architectures and majority rule logic to select the best classification output, and compared the results with statistical methods. Pal [101] proposed random forest classifier applying combination of tree classifiers with randomly selected features or combination of features at each node to grow the tree. A review on ensemble classifiers can be found in Oza and Turner [106].

Multisource Classifiers: As success with using only satellite imagery based on spectral similarity for land cover classification is limited, the use of ancillary information from other sources with satellite imagery is essential to meet the required accuracy. Many research works on using multi-source and very high dimensional remotely sensed imagery for classification reported in literature. Notably, Benediktsson et al. [107, 108] applied successfully conjugate gradient neural networks along with several thematic and non thematic information sources. Bischof et al. [109] employed textures along with multispectral pixel data with ANN and compared the results with conventional maximum likelihood classification technique. Bischof et al. [110] proposed an approach to optimise the neural networks using minimum description length principle for better land use classification using ANN over the Gaussian maximum likelihood classifier. However, in dealing with multisource and high dimensional data, the determination of source weighting is a difficult issue. Tso and Mather [111] proposed to use the combination of Genetic Algorithm and Markov Random Fields to extract contextual information and also source weighting through energy minimisation approach to classify a combination of six Landsat TM multispectral bands data and microwave polarisation images from SIR-C mission.

Some of other recent applications of ANNs reported in the literature are object based image analysis of high resolution images [112], extraction of temperature, pressure and humidity profiles by analysing radio occultation data from GPS-low-Earth-orbit satellites in the Arctic regions [113], estimation of geometry of seismic source from

interferometric SAR (InSAR) data although the reported performance with nearly 25% errors was not impressive [114], to name a few. As it is difficult to get real In SAR data from earthquakes, the authors trained their ANNs on synthetic interferograms. Auto-associative ANNs were used for dimensionality reduction in pixel unmixing of hyperspectral data by Licciardi and del Frate [115]. The reduced features formed the input to a second ANN for mapping the data to abundance percentages.

An ANN-based simulation of land cover researched by Maithani [116]. Aim was to simulate the process of land cover changes based on different policies which provided on the basis of sustainable development. A predictive neural model was developed to generate future land cover pattern which was based on Baseline, Compacts growth and Hierarchical growth scenarios. The result suggests that there will be regional imbalance due to unabated continuation of the present pattern of land cover transformation [116]. Another urban growth zonation done by Maithani et al. [117] using ANN approach. They have indicated the convention approach which was subjective in nature for mapping urban growth zonation. To reduce the subjectivity ANN were used. The results showed that ANN has potential to map the urban growth zonation which provides valuable input to the urban planners. To reduce subjectivity and calibration time using neural network approach, Maithani [118] developed a Neural Network based urban growth model for a city of India. GIS technique was used to handle attributes while relationship between urban growth and site attributes was establish using ANN. The optimized ANN architecture was applied for future growth simulation, while using various feed forward ANN architectures in this study [118]. An ANN based approach for modelling the urban spatial growth was done by Maithani et al. [119]. In this study a three layered feed forward NN used, to calculate land use transition while training using back propagation algorithm. The model results are evaluated to find out how accurately the model is able to predict the urban morphology using the percent correct match (PCM) metric and Moran spatial autocorrelation index.

The remote sensing community has also explored the use of Active Learning (AL) in classification tasks. In AL, human feedback is dynamically used for improving the performance over a period of time. Munoz-Mari et al. [120] used AL for semi supervised classification of RS image [67] in which a classification map was produced using a supervised classifier. A confidence map was then generated based on the known number of pixels in each class. Relevance feedback (active queries) was then used to update both the confidence and classification maps. Li et al. [38, 39] used a Bayesian approach in conjunction with active learning to segment hyperspectral images. They learn class posterior probability distributions using

multinomial regression in the first step. These are then used for segmenting hyperspectral images in the second step. Performance results are shown on both real and synthetic hyperspectral data. Domain adaptation in supervised classification was attempted by Persello and Bruzzone [121]. In this work, they have adapted supervised classifier trained on one domain to another. They have then iteratively labelled and added to the training set the most informative samples from the new domain while deleting those that are least relevant from the old domain. This work may also be of interest to the computing community as it provides a test bed for the effectiveness of the new transfer learning architectures being developed by the deep learning community [122].

Forest biomass is an important parameter for assessing the status of forest ecosystem. In a recent study, Nandy et al. [123] integrated remotely sensed satellite data and field inventory data using artificial neural network (ANN) technique in Barkot forest region of Uttarakhand state in India. Forest type and density maps were prepared as well as spectral and texture variables were derived from LISS-III of Resourcesat-1. Dhanda et al. [124] reported a robust forest (RF) biomass estimation procedure by combining multisensor data from ICESat/GLAS and high resolution optical data and also using two regression algorithms; random forest and SVM. The study demonstrated that the RF regression algorithm was reported to perform equally well on datasets irrespective of the correlation of underlying variables with the predicted variable, while SVM regression was found to perform well on those datasets which had a subset of underlying variables that were correlated with the predicted variable. Mangla et al. [125] proposed to use a combination of SAR and LIDAR remotely sensed data along with ground based Terrestrial Laser Scanner for forest parameter retrieval using RF regression approach.

4 Concluding Remarks

In this review, we have brought some important contributions of soft computing in the field of remote sensing. Soft computing itself has a vast collection of tools used in various applications. The algorithms used in soft computing have been developed earlier for other computational purposes and subsequently adapted with necessary modification for their applications in remote sensing. From logical reasoning problems like exclusive 'OR' or traveling salesman problem to handling real world data classifications from medium resolution multispectral sensors like Landsat, SPOT, IRS to very high resolution IKONOS, QuickBird and further to microwave SAR imagery, the soft computing techniques have shown their immense potential

to compete and excel in most cases with conventional statistical methods. From the algorithmic side, there has been a significant discussion on the linear and non-linear classes and feature selection and their consequences to accuracy. But the focus is on classification and accuracy tasks obtained after application of those algorithms discussed in Sect. 4.

For the fuzzy based classification, major contributions were based on how the membership value of the pixels is assigned. There is statistical evidence to support the usage of fuzzy objective function as fuzzy based classification. The most important characteristic of a fuzzy based classifier is its ability to perform with small training data. Compared to other methods, such as back propagation neural networks or maximum likelihood, and the accuracy achieved using these classifiers were more as compared to other soft computing classifiers. The fuzzy classifiers rely on constraints and also on a few boundary points to define its hyperplane. This characteristic has proved to be very useful in many applications, mainly because cutting down the cost of ground truth data collection. The ANN and its variant models have become highly dependable and most widely used classifier. But some limitations of ANN still remain unsolved. As the number of objects to be discriminated increases, it comes with a high computational cost, largely due to its demand for more training data and more weights to be adjusted and also more number of hidden layers as the complexity of network structure increases with number of objects. Deep learning is another soft computing based algorithm. Within deep learning only convolution neural network is extensively used in the field of remote sensing. It has been more popular because of its potential to automatically learn relevant contextual features. But deep knowledge is required for the implementation of this classifier. Its success is based on two steps, (1) generalise the imperfect training data with the manually labelled data and (2) its precise details. But this approach has not been extensively used in remote sensing because it require deep knowledge of convolutional neural network, and also requires accurate training data. But it will be a good topic of upcoming research. Bayesian network combines statistical theory with the ANN architecture which is becoming popular in remote sensing domain. While using several variables, it is possible to characterize phenomena through plausible reasoning inferences using new computer-aided method [126]. Advantages of Bayesian network is that it incorporates expert's knowledge into the process. The most commonly used Bayesian network consists of Naïve Bayes, Tree augmented Naïve Bayes (TAN) and Gaussian Bayesian network. But, not much software is available for this classifier. Hence researchers have developed their own basic tools available with R software and MATLAB. Much research is pending for

software developers to explore optimal and robust algorithmic approaches to handle large volume data typical of very high resolution remotely sensed data.

In summary, this review article has covered major soft computing techniques in use for various remote sensing applications. Special emphasis was given to classification of remotely sensing imagery wherein strong contributions are observed in the literature. It is our hope that the review article will facilitate state-of-art research that has gone in this area which will help the young students, and research community in pursuit of information in this subject.

Acknowledgements Authors sincerely thank Dr. P. S. Roy, Editor of this special issue, for his encouragement to write this paper and constant support throughout the preparation of the manuscript. They sincerely acknowledge Sri. A. S. Kiran Kumar, Chairman ISRO, for his constant encouragement to publish this work. Technical support provided by NRSC scientists is gratefully acknowledged.

References

- Bezdek JC (1984) FCM: The fuzzy c-means clustering algorithm 1. *Comput Geosci* 10(2):191–203
- Ji M (2003) Using fuzzy sets to improve cluster labelling in unsupervised classification. *Int J Remote Sens* 24(4):657–671. <https://doi.org/10.1080/01431160210146226>
- Shalan MA, Arora MK, Ghosh SK (2003) An evaluation of fuzzy classifications from IRS 1C LISS III imagery: a case study. *Int J Remote Sens* 24(15):3179–3186. <https://doi.org/10.1080/0143116031000094791>
- Wang F (1990) Fuzzy supervised classification of remote sensing images. *IEEE Trans Geosci Remote Sens* 28(2):194–201. <https://doi.org/10.1109/36.46698>
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Krishnapuram R, Keller JM (1993) A possibilistic approach to clustering. *IEEE Trans Fuzzy Syst* 1(2):98–110. <https://doi.org/10.1109/91.227387>
- Chawla S (2010) Possibilistic c-Means-Spatial Contextual Information based sub-pixel classification approach for multi-spectral data. *Int Inst Geo Inf Sci Earth Obs, M.Sc Thesis*. <http://www.iirs.gov.in/content/possibilistic-c-means-spatial-contextualinformation-based-sub-pixel-classification-approach>
- Wu X-H, Zhou J-J (2006) An improved possibilistic c-means algorithm based on kernel methods. *Struct Syntactic Stat Pattern Recognit* 4109:783–791. https://doi.org/10.1007/11815921_86
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28
- Foody GM (2000) Estimation of sub-pixel land cover composition in the presence of untrained classes. *Comput Geosci* 26(4):469–478. [https://doi.org/10.1016/S0098-3004\(99\)00125-9](https://doi.org/10.1016/S0098-3004(99)00125-9)
- Krishnapuram R, Keller JM (1996) The possibilistic C-means algorithm: insights and recommendations. *IEEE Trans Fuzzy Syst* 4(3):385–393. <https://doi.org/10.1109/91.531779>
- Zhang JS, Leung YW (2004) Improved possibilistic C-means clustering algorithms. *IEEE Trans Fuzzy Syst* 12(2):209–217. <https://doi.org/10.1109/TFUZZ.2004.825079>
- Wu XH, Zhou JJ (2008) Modified possibilistic clustering model based on kernel methods. *J Shanghai Univ* 12(2):136–140. <https://doi.org/10.1007/s11741-008-0210-2>
- Atkinson PM, Tatnall ARL (1997) Neural networks in remote sensing. *Int J Remote Sens* 18:699–709
- Cross SS, Harrison RF, Kennedy RL (1995) Introduction to neural networks. *Lancet* 346:1075–1079. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000229\)19:4<541::AID-SIM355>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-0258(20000229)19:4<541::AID-SIM355>3.0.CO;2-V)
- Yao XIN, Member S (1999) Evolving artificial neural networks 87(9): 1423–1447
- Mas JF, Flores JJ (2007) The application of artificial neural networks to the analysis of remotely sensed data. *Int J Remote Sens* 29:617–663. <https://doi.org/10.1080/01431160701352154>
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115–133
- Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: A tutorial. *IEEE Computer* 29(3):31–44. <https://doi.org/10.1109/2.485891>
- Benediktsson J, Kneppopoulos I (1999) Information extraction based on multisensor data fusion and neural networks. *Inf Process Remote Sens* 369–395. http://doi.org/10.1142/9789812815705_0017
- Specht D (1990) Probabilistic neural networks. *Neural Netw* 3(1):109–118. [https://doi.org/10.1016/0893-6080\(90\)90049-Q](https://doi.org/10.1016/0893-6080(90)90049-Q)
- Deng L, Yu D (2014) Deep learning: methods and applications. In: *Foundations and trends in signal processing* (Vol 7, pp 197–387). <https://doi.org/10.1561/20000000039>
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Lawrence S, Giles CL, Tsoi Ah Chung, Back AD (1997) Face recognition: a convolutional neural-network approach. *IEEE Trans Neural Netw* 8(1):98–113. <https://doi.org/10.1109/72.554195>
- LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE international symposium on circuits and systems, Paris, France*
- Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. *Mach Learn* 29:131–163
- Huang K, Yang H, Irwin K, Lyu MR (2004) Learning large margin classifiers locally and globally. In: *International conference on machine learning*
- Vapnik VN (2000) *The nature of statistical learning theory*, vol 8. Springer, New York. <https://doi.org/10.1109/TNN.1997.641482>
- Mittal A, Kassim A (2007) Bayesian network technologies: applications and graphical models. <https://doi.org/10.4018/978-1-59904-141-4>
- Heckerman D, Geiger D, Chickering D (1995) Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn* 20:197–243. <https://doi.org/10.1023/A:1022623210503> **243(1992), 197–243**
- Cooper GF, Herskovits E (1992) A Bayesian method for the induction of probabilistic networks from data. *Mach Learn* 9:309–347. <https://doi.org/10.1007/BF00994110>
- Solares C, Sanz A (2007) Different Bayesian network models in the classification of remote sensing images. *Data Eng Autom Learn IDEAL*, 10–16. http://link.springer.com/chapter/10.1007/978-3-540-77226-2_2
- Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. *ISPRS J Photogramm Remote Sens* 66(3):247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Vapnik V (2006) *Estimation of dependences based on empirical data*. Springer, Berlin. <https://doi.org/10.1007/0-387-34239-7>
- Bruzzone L, Chi M, Marconcini M (2006) A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Trans Geosci Remote Sens* 44(11):3363–3372. <https://doi.org/10.1109/TGRS.2006.877950>

37. Pal M, Mather PM (2005) Support vector machines for classification in remote sensing. *Int J Remote Sens* 26(5):1007–1011. <https://doi.org/10.1080/01431160512331314083>
38. Li H, Wang Y, Li Y, Wang X (2011) Pixel unmixing in moderate resolution remote sensing imagery using pairwise coupling SVM: a case study. *IEEE Trans Geosci Remote Sens* 49(11):4298–4307
39. Li J, Bioucas-Dias JM, Plaza A (2011) Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Trans Geosci Remote Sens* 49(10):3947–3960
40. Dos Santos JA, Gosselin P-H, Philippe-Foliguet S, Torres RS, Falcao AX (2011) Multiscale classification of remote sensing images. *IEEE Trans Geosci Remote Sens* 49(11):3764–3775
41. Bhushan N, Rai K (2004) Strategic decision making-applying the analytical hierarchy process. <https://doi.org/10.1007/b97668>
42. Engineering I, Triantaphyllou E, Mann SH (1995) Using the analytic hierarchy process for decision making in engineering applications: some challenges. *Int J Ind Eng Theory Appl Pract* 2(1):35–44
43. Beynon M (2002) DS/AHP method: A mathematical analysis, including an understanding of uncertainty. *Eur J Oper Res* 140(1):148–164. [https://doi.org/10.1016/S0377-2217\(01\)00230-2](https://doi.org/10.1016/S0377-2217(01)00230-2)
44. Kahraman C, Cebeci U, Ulukan Z (2003) Multi-criteria supplier selection using fuzzy AHP. *Logist Inf Manag* 16(6):382–394. <https://doi.org/10.1108/09576050310503367>
45. Hong-feng NHN (2001) The application of remote sensing technique and AHP-fuzzy method in comprehensive analysis and assessment for regional stability of Chongqing City, China. In: 2001 International conferences on info-tech and info-net. Proceedings (Cat. No.01EX479), 1(November), pp 5–9
46. Fischler AM, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications ACM* 24(6):381–395
47. Yaniv Z (2010) Random sample consensus (RANSAC) algorithm: a generic implementation, Release 1.0
48. Changjie W, Hua N (2017) Algorithm of remote sensing image matching based on corner-point. In: 2017 International workshop on remote sensing with intelligent processing (RSIP), Shanghai, China, 2017, pp 1–4. <https://doi.org/10.1109/RSIP.2017.7958803>
49. Li L, Yang F, Zhu H, Li D, Li Y, Tang L (2017) An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells. *Remote Sens* 9(5):433. <https://doi.org/10.3390/rs9050433>
50. Yan J, Jiang W, Shan J (2012) Quality analysis on RANSAC-based roof facets extraction from airborne LIDAR data. In: International archives of the photogrammetry, remote sensing and spatial information sciences, Vol XXXIX-B3, 2012 (XXII ISPRS Congress, 25 August–01 September 2012, Melbourne, Australia)
51. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks. <http://doi.org/10.1109/ICNN.1995.488968>
52. Runkler TA, Katz C (2006) Fuzzy clustering by particle swarm optimization. In: IEEE international conference on fuzzy systems, pp 601–608
53. Permana KE, Hashim SZM (2010) Fuzzy membership function generation using particle swarm optimization. *Int J Open Probl Compt Math* 3(1):27–41
54. Buckley PJJ, Feuring T (2000) Evolutionary algorithm solution to fuzzy problems: Fuzzy linear programming. *Fuzzy Sets Syst* 109(1):35–53
55. Hoffmann F (2001) Evolutionary algorithms for fuzzy control system design. *Proc IEEE* 89(9):1318–1333
56. Herrera F, Lozano M, Verdegay JL (1998) A learning process for fuzzy control rules using genetic algorithms. *Fuzzy Sets Syst* 100(1–3):143–158
57. Eiben AE, Hinterding R, Michalewicz Z (1999) Parameter control in evolutionary algorithms. *IEEE Trans Evol Comput* 3(2):128–134
58. Herrera F, Lozano M (1996) Adaptation of genetic algorithm parameters based on fuzzy logic controllers. In: Herrera F, Verdegay JL (eds) Genetic Algorithms and Soft Computing. Physica-Verlag, pp 95–125
59. Garzelli A, Nencini F (2006) PAN-sharpening of very high resolution multispectral images using genetic algorithms. *Int J Remote Sens* 27:3273–3292
60. Yao H, Tian L (2003) A genetic-algorithm-based selective principal component analysis (GA-SPCA) method for high-dimensional data feature extraction. *IEEE Trans Geosci Remote Sens* 41(6):1469–1478
61. Singhai R, Singhai J (2012) Registration of satellite imagery using genetic algorithms. In: Proceedings of the world congress on engineering (WCE 2012), July 4–6, London. U.K. http://www.iaeng.org/publication/WCE2012/WCE2012_pp1201-1206.pdf
62. Foody GM, Cox DP (1994) Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *Int J Remote Sens* 15(3):619–631. <https://doi.org/10.1080/01431169408954100>
63. Foody GM, Lucas RM, Curran PJ, Honzak M (1997) Mapping tropical forest fractional cover from coarse spatial resolution remotely sensed imagery. *Plant Ecol* 131:143–154
64. Shackelford AK, Davis CH (2002) A fuzzy classification approach for high-resolution multispectral data over urban areas. In: IEEE International conference on Geoscience and remote sensing symposium, 2002. IGARSS02 2002, vol 3, no 10, pp 1621–1623
65. Bárdossy A, Samaniego L (2002) Fuzzy rule-based classification of remotely sensed imagery. *IEEE Trans Geosci Remote Sens* 40(2):362–374
66. Wendling L, Desachy J, Paries A (1997) Pattern recognition by splitting images into trees of fuzzy regions. *Intell Data Anal* 1:71–84
67. Cannon RL, Dave JV, Bezdek JC, Trivedi MM (1986) Segmentation of a thematic mapper image using the fuzzy c-means clustering algorithm. *IEEE Trans Geosci Remote Sens*, GE-24(3), 400–408. <https://doi.org/10.1109/TGRS.1986.289598>
68. Zhang J, Foody GM (2001) Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *Int J Remote Sens* 22(4):615–628. <https://doi.org/10.1080/01431160050505883>
69. Gopal S, Woodcock CE, Strahler AH (1999) Fuzzy neural network classification of global land cover from a 1° AVHRR data set. *Remote Sens Environ* 67(2):230–243
70. Solaiman B, Pierce LE, Ulaby FT (1999) Multisensor data fusion using fuzzy concepts: application to land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Trans Geosci Remote Sens* 37(3):1316–1326. <https://doi.org/10.1109/36.763295>
71. Foody GM (1996) Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *Int J Remote Sens* 17(7):1317–1340. <https://doi.org/10.1080/01431169608948706>
72. Foody GM (1996) Fuzzy modelling of vegetation from remotely sensed imagery. *Ecol Model* 85(1):3–12. [https://doi.org/10.1016/0304-3800\(95\)00012-7](https://doi.org/10.1016/0304-3800(95)00012-7)
73. Benz UC, Hofmann P, Willhauck G, Lingenfelder I, Heynen M (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J*

- Photogramm Remote Sens 58(3–4):239–258. <https://doi.org/10.1016/j.isprsjprs.2003.10.002>
74. Musande V, Kumar A, Kale K (2012) Cotton crop discrimination using fuzzy classification approach. *Indian Soc Remote Sens*. <https://doi.org/10.1007/s12524-012-0201-z>
 75. Misra G, Kumar A, Patel NR (2012) Mapping a specific crop—a temporal approach for sugarcane ratoon. *J Indian Soc Remote Sens*. <https://doi.org/10.1007/s12524-012-0252-1>
 76. Upadhyay P, Roy PS, Gilbert I (2012) Effect on specific crop mapping using WorldView-2 multispectral add-on bands: soft classification approach. *J Appl Remote Sens*. <https://doi.org/10.1117/1.JRS.6.063524>
 77. Kumar A, Ghosh S, Dadhwal VK (2010) ALCM : automatic land cover mapping. *J Indian Soc Remote Sens* 38:203–210. <https://doi.org/10.1007/s12524-010-0027-5>
 78. Kumar A, Saha A, Dadhwal VK (2010) Some issues related with sub-pixel classification using HYSI data from IMS-1 satellite. *J Indian Soc Remote Sens* 38:203–210. <https://doi.org/10.1007/s12524-010-0027-5>
 79. Upadhyay P, Ghosh SK, Kumar A, Krishna YVN, Raju PL (2014) Moist deciduous forest identification using MODIS temporal indices data Priyadarshi. *Int J Remote Sens*. <https://doi.org/10.1080/01431161.2014.903438>
 80. Kumar A, Ghosh SK, Dadhwal VK (2007) Full fuzzy land cover mapping using remote sensing data based on fuzzy c-means and density estimation. *Can J Remote Sens* 33(2):81–87
 81. Kumar A, Dadhwal VK (2010) Entropy based fuzzy classification parameter optimization using uncertainty variation across spatial resolution. *J Indian Soc Remote Sens* 38:179–192. <https://doi.org/10.1007/s12524-010-0025-7>
 82. Nandan R, Kamboj A, Kumar A, Kumar AS, Reddy KV (2016) Formosat-2 with Landsat-8 temporal—multispectral data for wheat crop identification using Hypertangent Kernel based Possibilistic classifier. *J Geomat* 10(1):89–95
 83. Tripathi RN, Kumar R, Kumar A, Kumar AS (2016) Wheat monitoring by using kernel based possibilistic c-means classifier: Bi-sensor temporal multi-spectral data. *J Indian Soc Remote Sens*. <https://doi.org/10.1007/s12524-016-0651-9>
 84. Upadhyay P, Ghosh SK, Kumar A (2015) Temporal MODIS data for Identification of wheat crop using noise clustering soft classification approach. *Geocarto Int* 37–41. <https://doi.org/10.1080/10106049.2015.1047415>
 85. Sengar SS, Kumar A, Ghosh SK, Wason HR (2013) Earthquake-induced built-up damage identification using IRS-P6 data: a comparative study using fuzzy-based classifiers, (February). <https://doi.org/10.1080/10106049.2012.756939>
 86. Sengar SS, Kumar A, Wason HR, Ghosh SK, Murthy YVNC, Raju PL (2013) Study of soft classification approaches for identification of earthquake- induced liquefied soil. *Geocarto, Nat Hazards Risk* 37–41. <https://doi.org/10.1080/19475705.2013.811444>
 87. Upadhyay P, Kumar A, Ghosh SK (2013) Fuzzy based approach for moist deciduous forest identification using MODIS temporal data. *Indian Soc Remote Sens*. <https://doi.org/10.1007/s12524-013-0267-2>
 88. Harikumar A, Kumar A, Stein A, Raju PLN (2015) International Journal of Remote An effective hybrid approach to remote-sensing image classification, 37–41. <https://doi.org/10.1080/01431161.2015.1047050>
 89. Dutta A, Kumar A, Sarkar S (2010) Some issues in contextual fuzzy c-means classification of remotely sensed data for land cover mapping. *J Indian Soc Remote Sens* 38(1):109–118
 90. Singha M, Kumar A, Raju PNL, Krishna YVN (2015) Importance of DA-MRF models in fuzzy based classifier. *J Indian Soc Remote Sens* 43(1):27–35. <https://doi.org/10.1007/s12524-014-0382-8>
 91. Ghosh S, Nandy S, Patra S, Kushwaha SPS, Kumar AS, Dadhwal VK (2017) Land cover classification using ICESat/GLAS Full waveform data. *J Indian Soc Remote Sens* 45(2):327–335
 92. Paola JD, Schowengerdt RA (1994) A review and analysis of backpropagation neural networks for classification of remotely sensed multi-spectral imagery. *Int J Remote Sens* 16:3033–3058
 93. Heermann PD, Khazenie N (1992) Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Trans Geosci Remote Sens* 30(1):81–88
 94. Foody GM, McCulloch MB, Yates WB (1995) Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. *Am Soc Photogramm Remote Sens* 60:391–401
 95. Erbek FS, Özkan C, Taberner M (2004) Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *Int J Remote Sens* 25(9):1733–1748. <https://doi.org/10.1080/0143116031000150077>
 96. Hong Y, Hsu K-L, Sorooshian S, Gao X (2004) Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J Appl Meteorol* 43(12):1834–1853
 97. Gopal S, Woodcock C, Member A (1996) Remote sensing of forest change using artificial. *Neural Netw* 34(2):398–404
 98. Dal XL, Khorram S (1999) Remotely sensed change detection based on artificial neural networks. *Photogrammetric Eng Remote Sens* 65(10):1187–1194
 99. Moody A, Gopal S, Strahler AH (1996) Artificial neural network response to mixed pixels in coarse-resolution satellite data. *Remote Sens Environ* 58(3):329–343. [https://doi.org/10.1016/S0034-4257\(96\)00107-1](https://doi.org/10.1016/S0034-4257(96)00107-1)
 100. Warner TA, Shank M (1997) An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks. *Photogramm Eng Remote Sens* 63(11):1285–1294
 101. Pal M (2005) Ensemble of support vector machines for land cover classification. *Int J Remote Sens* 29(10):3043–3049
 102. Kumar AS, Majumder KL (2001) Information fusion in tree classifiers. *Int J Remote Sens* 22(5):861–869. <https://doi.org/10.1080/01431160051060327>
 103. Kumar AS, Basu SK, Majumdar KL (1997) Robust classification of multispectral data using multiple neural networks and fuzzy integral. *IEEE Trans Geosci Remote Sens* 35(3):787–790. <https://doi.org/10.1109/36.582004>
 104. Sugeno M (1977) Fuzzy measures and fuzzy integrals: A survey in Fuzzy Automata and Decision Processes. North Holland, Amsterdam, pp 89–102
 105. Giacinto G, Roli F (1997) Ensembles of neural networks for soft classification of remote sensing images. In: European symposium on intelligent techniques (pp 20–21)
 106. Oza CN, Turner K (2008) Classifier ensembles: select real-world applications. *J Inf Fusion* 9:4–20
 107. Benediktsson JA, Swain PH, Ersoy OK (1990) Neural network approaches versus statistical methods in classification of multi-source remote sensing data: *IEEE Trans. Geosci Remote Sens* 28(4):540–551
 108. Benediktsson JA, Swain PH, Ersoy OK (1993) Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. *Int J Remote Sens* 14(15): 2883–2903
 109. Bischof H, Schneider W, Pinz A (1992) Multispectral classification of Landsat images using neural networks. *IEEE Trans Geosci Remote Sens* 30(3):482–490
 110. Bischof H, Leonardis A, Selb A (1999) MDL principle for robust vector quantization. *Pattern Anal Appl* 2:59–72

111. Tso BCK, Mather PM (1999) Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields. *IEEE Trans Geosci Remote Sens* 37:1255–1260
112. Rizvi IA, Mohan BK (2011) Object-based image analysis of high-resolution satellite images using modified cloud basis function neural network and probabilistic relaxation labeling process. *IEEE Trans Geosci Remote Sens* 49(12):4815–4820. <http://doi.org/10.1109/TGRS.2011.2171695>
113. Pelliccia F, Pacifici F, Bonafoni S, Basili P, Pierdicca N, Ciotti P, Emery WJ (2011) Neural networks for arctic atmosphere sounding from radio occultation data. *IEEE Trans Geosci Remote Sens* 49(12):4846–4855
114. Stramondo S, del Frate F, Picchiani M, Schiavon G (2011) Seismic source quantitative parameters retrieval from InSAR data and neural networks. *IEEE Trans Geosci Remote Sens* 49(1):96–103
115. Licciardi GA, del Frate F (2011) Pixel unmixing in hyperspectral data by means of neural networks. *IEEE Trans Geosci Remote Sens* 49(11):4163–4172
116. Maithani S (2015) Neural network based simulation of land cover scenarios in Doon valley, India. *Geocarto Int* 30(2):163–185
117. Maithani S, Arora MK, Jain RK (2010) An artificial neural network based approach for urban growth zonation in Dehradun city, India. *Geocarto Int* 25(8):663–681
118. Maithani S (2009) A neural network based urban growth model of an Indian city. *J Indian Soc Remote Sens* 37(3):363–376
119. Maithani S, Jain RK, Arora MK (2007) An artificial neural network based approach for modelling urban spatial growth. *J Inst Town Plan* 4(2):43–51
120. Munoz-Mari J, Tuia D, Camps-Valls G (2012) Semisupervised classification of remote sensing images with active queries. *IEEE Trans Geosci Remote Sens* 50(10):3751–3763
121. Persello C, Bruzzone L (2012) Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Trans Geosci Remote Sens* 50(11):4468–4483
122. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 27:3320–3328
123. Nandy S, Singh RP, Ghosh S, Watham T, Kushwaha SPS, Kumar, AS Dadhwal VK (2017) Neural network-based modelling for forest biomass assessment. *Carbon Manag (In Press)*
124. Dhanda P, Nandy S, Kushwaha SPS, Ghosh S, Krishna Murthy YVN, Dadhwal VK (2017) Optimising spaceborne LiDAR and very high resolution optical sensor parameters for biomass estimation at ICESat/GLAS footprint level using regression algorithms. *Prog Phys Geogr* 41(3):247–267
125. Mangla R, Kumar S, Nandy S (2016) Random forest regression modelling for forest aboveground biomass estimation using RISAT-1 PolSAR and terrestrial LiDAR data. In: *SPIE asia-pacific remote sensing* (pp 98790Q-98790Q). <https://doi.org/10.1117/12.2227380>
126. Mello MP, Risso J, Atzberger C, Aplin P, Pebesma E, Vieira CAO, Rudorff BFT (2013) Bayesian networks for raster data (BayNeRD): plausible reasoning from observations. *Remote Sens* 5(11):5999–6025. <https://doi.org/10.3390/rs5115999>