Adaptive Granulation-Based Convolutional Neural Networks With Single Pass Learning for Remote Sensing Image Classification

Sankar K. Pal, *Life Fellow, IEEE*, and Dasari Arun Kumar, *Senior Member, IEEE*

***Abstract*—Convolutional neural networks (CNNs) with the characteristics like spatial filtering, feed-forward mechanism, and back propagation-based learning are being widely used recently for remote sensing (RS) image classification. The fixed architecture of CNN with a large number of network parameters is managed by learning through a number of iterations, and, thereby increasing the computational burden. To deal with this issue, an adaptive granulation-based CNN (AGCNN) model is proposed in the present study. AGCNN works in the framework of fuzzy set theoretic data granulation and adaptive learning by upgrading the network architecture to accommodate the information of new samples, and avoids iterative training, unlike conventional CNN. Here, granulation is done both on the 2-D input image and its 1-D representative feature vector output, as obtained after a series of convolution and pooling layers. While the class-dependent fuzzy granulation on input image space exploits more domain knowledge for uncertainty modeling, rough set theoretic reducts computed on them select only the relevant features for input to CNN. During classification of unknown patterns, a new principle of roughness-minimization with weighted membership is adopted on overlapping granules to deal with the ambiguous cases. All these together improve the classification accuracy of AGCNN, while reducing the computational time significantly. The superiority of AGCNN over some state-of-the-art models in terms of different performance metrics is demonstrated for hyperspectral and multispectral images both quantitatively and visually.**

***Index Terms*—Deep adaptive granulation, fuzzy rough feature selection, image classification, pixel uncertainty, remote sensing, roughness measure, weighted class-membership.**

I. Introduction

CONVOLUTIONAL neural network (CNN) is one of the most extensively used deep neural network (DNN) models for remote sensing image classification and it can process multiple arrays of data, such as multiband remote sensing images with regularly arranged pixels. CNN consists of mainly three specific hierarchically connected structures like convolutional layer, pooling layer, and fully connected neural network. The series of convolutional and pooling layers in the CNN operate on the input image to obtain the informative features of the objects in the input image. These informative features are used to classify the objects using multiple fully connected neural networks. CNN can extract informative features of ill-defined objects in scene-level/pixel-level remote sensing images [1]. This characteristic of CNN motivated researchers to use it in various remote sensing image analysis like image registration [2], image restoration [3], image fusion [4], image segmentation [5], change detection [6], and land use/land cover classification [7], [8]. Hong et al. [9] used CNN with multimodality learning (MML) to classify the objects in complex scene images. The model used pixelwise classification and spatial information modeling in the said task. Hong et al. [10], presented a mini graph-based CNN to classify the objects in the hyperspectral remote sensing images. The model is computationally faster than the conventional graph-based CNN and it can classify unknown samples without retraining the network and improve the classification performance. Kumar [11] proposed a knowledge encoded CNN model for multispectral remote sensing image classification. Morphological operators were used in the convolutional layers of this model to obtain informative features of the objects. Wu et al. [12] presented a cross-channel reconstruction module-based CNN to classify the objects in hyperspectral and synthetic aperture radar (SAR) images. The model used reconstruction strategy across modalities to learn the features of the objects in the image. Hong et al. [13] developed a shared and specific feature learning-based CNN model to classify the objects in multispectral, hyperspectral, and SAR images. The model is used to decompose multimodal RS data to enable the information blending for heterogeneous data sources. Li et al. [14], proposed a transfer learning-based deep CNN for object classification in high resolution remote sensing images. The model works on the principle of transferring the weight parameters of convolution layers during the learning process to attain better performance. Li et al. [15] used 3D-CNN to classify the objects in high resolution hyperspectral remote sensing images. The concepts like principal component analysis and autoencoder are used in 3D-CNN to deal with the higher dimensionality of the data.

While the performance of the CNN in remote sensing image classification is appreciated, it has limitations like, the model

Manuscript received 26 July 2022; revised 28 October 2022; accepted 15 November 2022. Date of publication 18 November 2022; date of current version 7 December 2022. This work was supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India, through "National Science Chair" Professor Award to Sankar K. Pal. *(Corresponding author: Dasari Arun Kumar.)*

Sankar K. Pal is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata, West Bengal 700108, India (e-mail: sankar@isical.ac.in).

Dasari Arun Kumar is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata, West Bengal 700108, India, and also with the Department of Electronics and Communication Engineering, K.S.R.M. College of Engineering, Kadapa, Andhra Pradesh 516005, India (e-mail: arunkumar.mtech09@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3223180

directly operates on the raw images with uncertainty in the pixels that may lead to misclassification of the object. Further, a large number of network parameters in convolutional layers, and a fully connected neural network, and multiple training iterations may lead to overfitting and an increase in computational complexity/time [16].

A. Related Work

1) Uncertainty in the Pixels: The pixels in the raw images have uncertainty and it arises due to the beloningness of a pixel to more than one classes in the region of interest. The uncertainty in the pixels of image occurs due to its acquisition in different spatial, spectral, radiometric, and temporal resolutions [17]. Classifying these raw images without addressing the uncertainty in the pixels would lead to misclassification of the objects and it affects the overall performance of the CNN. This specifies the necessity of addressing the uncertainty in the pixels of remote sensing images. The concepts of fuzzy sets are used to address the uncertainty in the pixels of images. Pal and Mitra [18], suggested class independent granulation (CI) to represent the uncertainty in the pixels using π-type membership function and these membership values are further processed using a fully connected NN. This method is used in many remote sensing image analysis approaches like [19], [20], [21], [22], [23], [24], and [25]. Bastin et al. [26] used fuzzy spectral signature to represent the uncertainty in the pixels of Landsat images. This method used multilayered stacks of membership images for easy visualization of uncertainty in the pixels. Dungan et al. [27] used probability distribution to represent the uncertainty in the pixels of the input image and it was visualized using interactive maps of first-, second-, and third-order statistics. While Zhong et al. [28] used adaptive mimetic fuzzy clustering algorithm and Wu et al. [29] used fuzzy local C-means clustering-based method to deal with the uncertainty of pixels in remote sensing images, Chen et al. [30] applied fuzzy rough set-based adaptive genetic algorithmic method to address this task. These fuzzy-set-based methods use class independent approaches to handle uncertainty in the pixels of input image. These methods do not consider the class-belonging information of the pixels and can result in misclassification of objects in the image. The methods in these studies are used to represent the uncertainty in the pixels of input images in pixel-level classification of multispectral remote sensing images.

In the recent years, advancement in the imaging technology produced images with finer spatial resolution. The objects in these images are classified using patch-based classification methodologies, unlike pixel-level methods. Also, the uncertainty in the pixels of hyperspectral image patches is complex to address due to the finer spatial resolution of images. Recently, a few methods are suggested to deal with the uncertainty in the pixels of images in the patch-based hyperspectral remote sensing image classification. Zhao et al. [31] used autoencoder-based spectral unmixing methods to handle uncertainty in the pixels of hyperspectral remote sensing images. Further, a CNN-based autoencoder network is used to classify these hyperspectral remote sensing images. Li et al. [32] used pixel-by-pixel clustering framework to represent the uncertainty in the pixels of hyperspectral remote sensing images. The model used deep CNN architecture with large number of weight parameters to classify hyperspectral remote sensing images. In the present study, we use class-dependent granulation (CD) [33] for the first time to deal with the uncertainty in the pixels of image patches for the patch-level hyperspectral/multispectral remote sensing image classification. It is used in pixel-based classification of multispectral remote sensing images in the studies [34], [35], [24], and [25]. In CD method, a pixel in the input image is represented with its fuzzy membership to the classes in the dataset [33]. This increases the dimensionality of input image and, thereby increases the computational complexity of the model. In such situations, rough set-based theoretic approach [36] is used to select the informative features out of them by eliminating redundant ones.

2) Adaptive Granulation, Single-Pass Learning, and Roughness of Granules: The image patches in the dataset has diverse features of objects within the class. To learn these features of objects in the image patches, a large number of weight parameters/filter coefficients are used in CNN and this increases its computational time. Recently, the concept of data granulation is used to group the features of the objects (named as granules) in the image patches. These granules are processed using CNN and it is named as granular computing. Granular computing is inspired by human computing in which the computations are performed on the information granules rather than individual samples/patterns. A granule is a group of elements/objects with similar characteristics [37], [38]. The concepts of granulation are applied in data analysis [39], [40], [41], [42], [43]. Zeng et al. [44] suggested a multilevel data granulation-based CNN model for higher resolution remote sensing patch images. The model used granulation to capture the deep spatial and spectral information of objects in the image patches and this improved its performance. Jean et al. [45] used data granulation in the hyperspectral RS image patches. The model used shape/size features of the objects in the image patch to select the less number of filter coefficients in the convolutional layer of CNN and it could generate better performance in less number of iterations. These models used the concept of granulation to consider the information of diverse features of objects in the image patches and select the less number of filter coefficients in convolutional layers of CNN. The granulation-based CNN models produced better performance within less time, unlike conventional CNN.

Recently, Leite et al. [46], [47], introduced data granulation and adaptive learning in a neural network in classifying real-time data, and named as adaptive granular neural network. The work was motivated by adaptive neural networks, suggested by Alexandridis et al. [48] and Palnitkar et al. [49]. The adaptive granular neural network learns the diversified features of the objects in the image patches by updating the fuzzy membership of granules in a single training iteration called single pass. These characteristics of adaptive granular neural network make easy it to learn the features within a few seconds and result in better performance, unlike conventional neural network. This concept is used for remote sensing image classification in the studies like [50], [51], [52], and [53]. While the adaptive granular neural

network model is credited for learning the features in the images within a single pass, it has two limitations to be addressed. 1) In the testing stage of adaptive granular neural network, it assigns the unknown sample to a class to which the sample has maximum membership. This process works well with the datasets having equal number of classes. In real life situations, a dataset is likely to have unequal number of samples in its classes. In such situations, considering only the maximum membership of a sample to classify it, will not improve the overall accuracy of model. 2) In case of dataset having overlapping classes, an unknown feature vector may have maximum-membership value to more than one class. This makes the decision-making indecisive in adaptive granular neural network. In our present study, these two limitations are addressed by considering the concepts of weighted class-membership and roughness of overlapping granules in decision-making. We introduce a CNN-based deep architecture with eleven layers involving fuzzy adaptive granulation for hyperspectral/multispectral patch-based image classification. The model possesses the advantages of CD granulation of input image with reduced features, and dynamic granulation that evolves automatically to adapt input features of objects in the image patch. The system achieves better classification accuracy in much less time. We name it an adaptive granulation-based convolutional neural network (AGCNN).

B. Novelty

The novelty of the proposed model and contribution is described as follows.

1) Developing a fuzzy adaptive granulation-based deep network architecture requiring a smaller number of weight parameters and only single pass learning.
2) Incorporating the weighted class-membership,and roughness measure on overlapping granules for handling ambiguous patterns during the decision-making process, where the concept of using roughness measure is unique.
3) Using CD-granulation on input images, followed by computation of rough reducts, to consider only the relevant class-based information of pixels as input features.
4) Demonstrating the superiority of the proposed model over many state-of-the art deep models both quantitatively and visually.

C. Problems Addressed

The concepts like CD-granulation, adaptive granulation, and roughness of overlapping granules used in AGCNN model, address three major issues in remote sensing image classification.

1) Granulation is useful in modeling the indiscernibility in pixels for further processing. In our work, CD-granulation based on class label information handles better the uncertainty in indiscernibility in the pixels of hyperspectral image patches.
2) Adaptive granulation is used to derive salient information from the hyperspectral image patches. Use of adaptive granulation on the image patches, results in fewer weight parameters in the network architecture and enables single pass learning of the network. These improve the performance with less computational time.

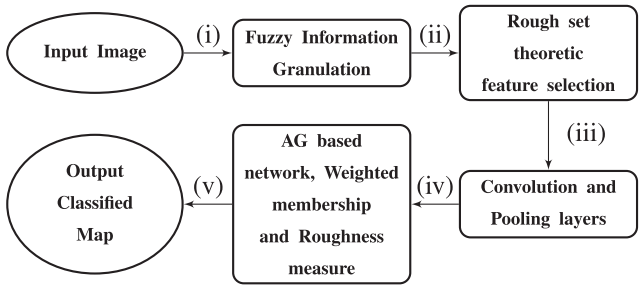

Fig. 1. Functional block diagram of AGCNN model.

3) The hyperspectral images have the pixels of classes with overlapping characteristic and these pixels often possess maximum membership to more than one class; this makes the model indecisive during classification of such pixels. To reduce this problem, the roughness of overlapping granules is used in the decision-making process.

All these factors make the proposed AGCNN superior in terms of performance and computational time.

II. Proposed Adaptive Granulation-Based CNN for Remote Sensing Image Classification

The functional mechanism of the proposed AGCNN model is divided into five modules, as shown in Fig. 1. 1) In fuzzy granulation layer, the pixels of input image are represented with their membership to the granules (Section II-B). 2) In feature selection layer, informative features are selected from granulated features using rough set theoretic approach (Section II-C). 3) The series of convolution and pooling layers (CLs and PLs) are used to obtain 1-D representative feature vector of objects in the image. 4) The feature vectors are used to build the architecture of adaptive granulation-based network. 5) Weighted membership of feature vector and roughness measure on granules are used to label the unknown test images. The layers of AGCNN are shown in Fig. 2.

In Fig. 2, eight layers (L_1, L_2, L_3,..., and L_8) of AGCNN are given with corresponding output at each layer. L_1 is the input image, L_2 is the fuzzy granulation layer with an output (J), $L3$ is the feature selection layer with an output (K), L_4 is convolution layer with an output (L), L_5 is pooling layer with output (M), and L_6 and L_7 are consecutive convolutional and pooling layers. L_8 is ($1D$) representative feature vector of the input image. The remaining three layers (L_9, L_{10}, and L_{11}) of AGCNN model are given in Fig. 5.

The description of the output at each layer of AGCNN is given in Table I. In Table I, the size of input image (I) is considered as (m, n, p) [Example : (9,9,4)], where m is number of rows, n is number of columns, and p is number of bands in the image. In the fuzzy granulation layer, a pixel in I is represented with its membership to c classes, where c is number of classes. The output of granulation layer (J) has a size (m, n, q), where $q = p \times c$ and q is number of granulated features. The output of feature selection layer has a size (m, n, r), where $r < q$, r is the number of selected features out of q. A series of convolutional and pooling layers are applied to informative features (m, n, r) to

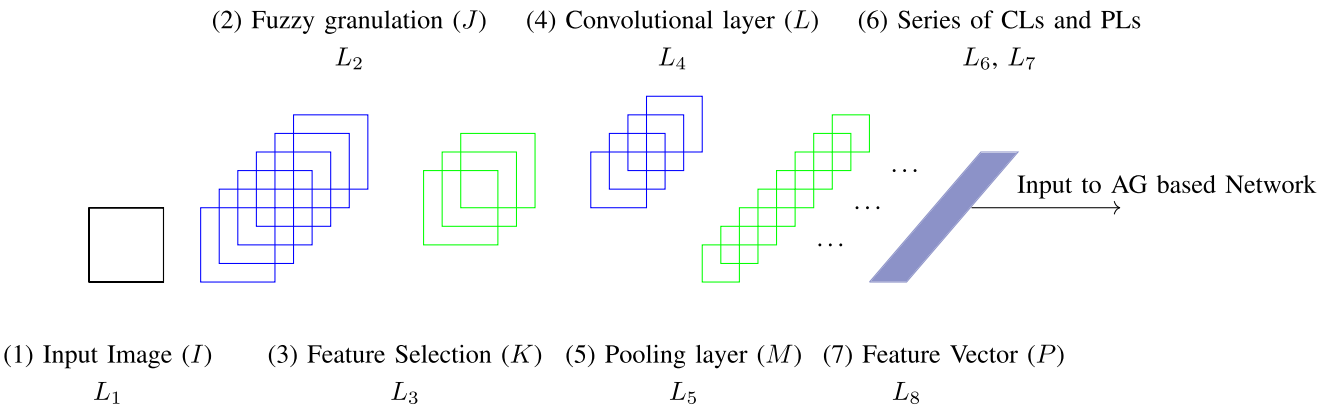

Fig. 2. Layers in AGCNN for remote sensing image classification.

TABLE I
DESCRIPTION OF LAYERS IN AGCNN

Layer Number	Layer name	Shape	Example	Description
L_1	Input Image (I)	(m,n,p)	(9,9,4)	$m = n = 9$, $p = 4$
L_2	Fuzzy granulation (J)	(m,n,q)	(9,9,20)	$q > p$, $q = p \times c$, $c = 5$
L_3	Feature selection (K)	(m,n,r)	(9,9,10)	$r \leq q$
L_4	Convolution layer (L)	(m,n,r)	(9,9,10)	filter size $a \times a$, $a = 3$
L_5	Pooling layer (M)	($\frac{m}{a}$,$\frac{n}{a}$,r)	(3,3,10)	max pooling
L_6	Convolution layer	($\frac{m}{a}$,$\frac{n}{a}$,r)	(3,3,10)	filter size $a \times a$
L_7	Pooling layer	(1,1,r)	(1,1,10)	max pooling
L_8	Feature vector (P)	(1,r)	(1,10)	input to AG based network

obtain the 1-D representative feature vector (1,1,r) of an image. The 1-D feature vectors generated from the images are used to build the architecture of AG-based network (Fig. 5).

A. Input Image

The input image patch (I) is feed forwarded through the series of layers in AGCNN. The size of I is considered based on the spatial resolution of the image. In case of hyperspectral images, the size of I is considered as 9,9,110 because the image is acquired in finer spatial resolution. In case of multispectral images, the size of I is considered as 20,20,4 and this is because the image is acquired in coarser spatial resolution. The size of input image is selected based on the criteria that finer the spatial resolution of the image is, smaller is the size of input image to CNN and the coarser the resolution of the image is, the larger is the size of input image to the CNN.

B. Granulation Layer

In conventional CNN model, the convolution filters are operated on raw images. In the present study, the convolution filters are operated on the membership values of pixels to the granules. In fuzzy granulation layer, the pixels of input image are represented with their memberships to the granules. There exist two types of pixel granulation methods namely, 1) CI granulation and 2) CD granulation.

1) Class Independent: In CI granulation, a pixel (Px) with p features is represented with its membership to three fuzzy granules, namely, *low(l)*, *medium(m)*, and *high(h)* [18]. The membership of a pixel is computed using the π-type function. Let, $Px = [f_1, f_2, \ldots, f_p]$ is a pixel with p features. The fuzzy membership of a feature f_j to a granule is

$$\mu(f_j; i, \alpha) = \begin{cases} 2\left(1 - \frac{\|f_j - i\|}{\alpha}\right)^2, & \text{for } \frac{\alpha}{2} \leq \|f_j - i\| \leq \alpha \\ 1 - 2\left(\frac{\|f_j - i\|}{\alpha}\right)^2, & \text{for } 0 \leq \|f_j - i\| \leq \frac{\alpha}{2} \\ 0, & \text{or else} \end{cases} \tag{1}$$

α and i are the radius and center point of fuzzy granule, respectively. The membership of a feature is maximum (1) at the mid point of granule. The membership is 0.5 at the cross-over points and decreases to 0 on either side of the center of granule. The membership of Px to the granules is

$$Px = \begin{bmatrix} \mu_l(f_1), & \mu_m(f_1), & \mu_h(f_1), \\ \ldots, & \ldots, & \ldots, \\ \mu_l(f_p), & \mu_m(f_p), & \mu_h(f_p) \end{bmatrix} \tag{2}$$

where $\mu_l(f_p)$, $\mu_m(f_p)$, and $\mu_h(f_p)$ are the memberships of Px to the granules l, m, and h for the feature value f_p. The membership of each pixel to three granules is computed using the π-type function.

2) Class Dependent: CI granulation does not consider the class belonging information of a pixel. Because of this, class-dependent (CD) granulation was suggested by Pal et al. [33]. In CD granulation, the features of a pixel is represented in terms of their membership to the classes in dataset. The membership of feature f_j to a class is computed as

$$\mu(f_j; X, R, Y) = \begin{cases} 0, & \text{for } f_j \leq X \\ 2^{W-1}\left(\frac{f_j - X}{R - X}\right)^N, & \text{for } X < f_j \leq P \\ 1 - 2^{W-1}\left(\frac{R - f_j}{R - X}\right)^N, & \text{for } P < f_j \leq R \\ 1 - 2^{W-1}\left(\frac{f_j - R}{Y - R}\right)^N, & \text{for } R < f_j \leq Q \\ 2^{W-1}\left(\frac{Y - f_j}{Y - R}\right)^N, & \text{for } Q < f_j \leq Y \\ 0, & \text{for } f_j \geq Y \end{cases} \tag{3}$$

where the crossover points of membership are P and Q with membership 0.5. R is the center with maximum

membership 1. The values P, Q, and R of a granule are computed as $R = \text{mean}(f)$ (i.e., mean of feature f to a class), $P = R - (\frac{\max(f)-\min(f)}{2})$ and $Q = R + (\frac{\max(f)-\min(f)}{2})$, $\min(f)$ and $\max(f)$ are minimum and maximum values of feature f in class. X and Y are called extreme points of granule, where $X = R - (Q - P)$ and $Y = R + (Q - P)$. In CD granulation, a pixel Px is represented with its membership to c classes

$$Px = \begin{bmatrix} \mu_1(f_1), & \mu_2(f_1), & \mu_c(f_1), \\ \mu_1(f_2), & \mu_2(f_2) & \ldots, \\ \mu_1(f_p), & \mu_2(f_p), & \mu_c(f_p) \end{bmatrix} \tag{4}$$

$\mu_c(f_p)$ is the membership of Px to class c for the feature value f_p.

C. Feature Selection Layer

In feature selection layer, rough set theoretic approach is used to select the informative features from the granulated features. The set of such informative features is called reduct. A reduct is a minimum set of features which can discern the samples in the dataset [36]. Let U is a set of pixels in an image, A is a set of features/bands, $A = (B, C)$, where B is a set of conditional features and C is decision feature. U consists of pixels $X = \{Px_1, Px_2, Px_3, ..., Px_n\}$. The reduct is a minimal set of features in B which can discern the pixels in U based on C. The rough set-based functional dependency ($\gamma_B(C)$) of features in B is

$$\gamma_B(C) = \frac{|\text{POS}_B(C)|}{|U|} \tag{5}$$

where, $\text{POS}_B(C)$ is positive region of set B using C

$$\text{POS}_B(C) = \bigcup_{X \in U/C} \underline{B}X \tag{6}$$

$\underline{B}X$ is lower approximation of set B

$$\underline{B}X = \{Px | [Px]_B \subseteq X\}. \tag{7}$$

The upper approximation of set B

$$\overline{B}X = \{Px | [Px]_B \cup X \neq \emptyset\}. \tag{8}$$

The maximum value of $\gamma_B(C)$ is 1, indicating that the features in B can discern all the pixels in X using C.

D. Adaptive Granulation-Based Network

In conventional CNN, the feature vector is feed forwarded through the fully connected layers of neural networks with fixed architecture. Furthermore, the output error is computed and the weight parameters of CNN are updated by back-propagating the output error for a number of iterations. This increases the time complexity of the CNN. In the present study, we considered an adaptive granulation-based network with less parameters, and the model can learn only in a single pass of training samples.

1) Architecture of Adaptive Granulation-Based Network: The architecture of adaptive granulation-based network is implemented in three steps: a) creating rectangular granules in feature space, b) upgrading the granules to accommodate the new feature vector, and c) building the architecture of adaptive granulation-based network.

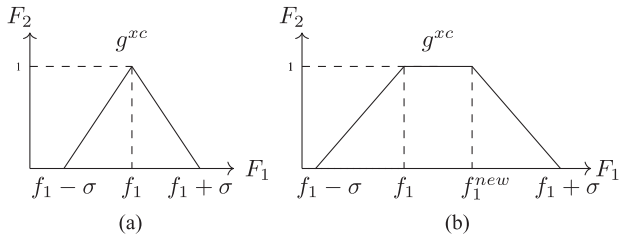

Fig. 3. (a) Generating a granule. (b) Upgrading a granule.

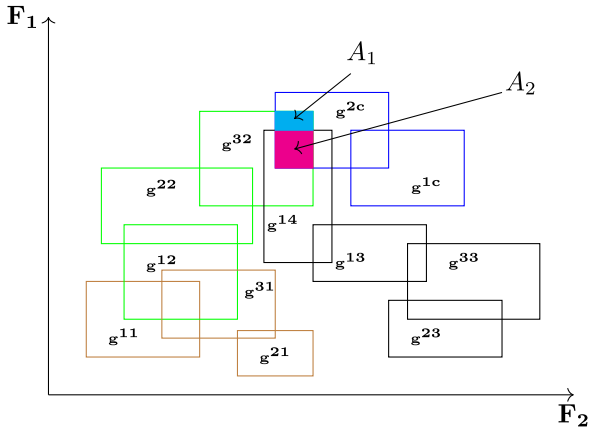

Fig. 4. Granulation of feature space.

The feature vectors generated from the training images by passing through layers L_1, L_2, L_3,..., and L_8 are used as the labeled dataset. During training, a granule is created either with the first feature vector, or with the feature vector of a new class, or if the feature vector does not fit within existing granules. A granule is upgraded if a new feature vector falls within it. The creation/upgradation of the granule is shown in Fig. 3(a) and (b). In Fig. 3(a), xth granule of cth class (g^{xc}) is created along feature axis F_1. The parameter σ is the standard deviation of the class of feature vector and it is used to obtain the extent of granule. In Fig. 3(b), g^{xc} is updated to accommodate new feature value f_1^{new} along feature axis F_1. An example of granules in F_1-F_2 feature space is shown in Fig. 4. In Fig. 4, A_1 and A_2 are overlapping (ambiguous) regions among granules g^{32}, g^{2c}, and g^{14}. Note that a class can have more than one granule.

The architecture of adaptive granulation-based network is built based on the number of granules (shown in Fig. 5). In Fig. 5, the number of input nodes is equal to the number of features in the feature vector. The number of adaptive layer (AL) nodes is equal to the number of granules generated from feature vectors. The number of output layer (OL) nodes is equal to the number of classes in the dataset. The roughness measure layer RML (optional) is used to measure the ambiguity in the overlapping regions of granules when a feature vector has equal membership to more than one class.

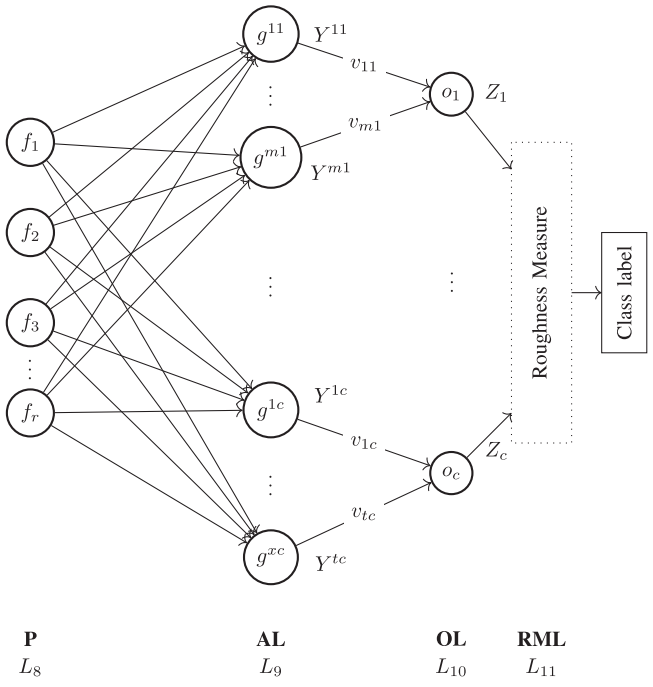

Fig. 5. Architecture of adaptive granulation-based network.

2) Training Adaptive Granulation-Based Network: During training, the membership of a feature vector (P) to a class (c) is obtained by adding its memberships to the granules of class c

$$Y^c = Y^{1c} + Y^{2c} +, \ldots\ldots + Y^{xc} \tag{9}$$

where, Y^c is the membership of P to class c. Y^{xc} is the membership of P to xth granule of class c. The membership of P to a granule is obtained using trapezoidal membership function. Y^{1c}, Y^{2c}....., and Y^{xc} are the outputs at the adaptive layer nodes. The membership of P to a class (Y^c) and the sum of weight parameters (v^c) of class c are feed forwarded to the output layer

$$v^c = v^{1c} + v^{2c} + \cdots . + v^{xc} \tag{10}$$

where, v^{1c}, v^{2c},...., and v^{xc} are the weight parameters between the granules of class c and output node O_c. The output at node O_c is

$$Z_c(Y^c, v^c) = \begin{cases} e\ S\left(\frac{Y^c}{e}, \frac{v^c}{e}\right), & Y^c, v^c \in [0, e] \\ g\ T\ \left(\frac{Y^c - e}{1-e}, \frac{v^c - e}{1-e}\right), & Y^c, v^c \in [e, 1] \\ e, otherwise \end{cases} \tag{11}$$

where, $e \in [0,1]$, $g = e + (1 - e)$, and S and T are the norm and conorm operators.

$$S\left(\frac{Y^c}{e}, \frac{v^c}{e}\right) = \min\left(1, \frac{Y^c}{e} + \frac{v^c}{e}\right), \tag{12}$$

$$T\left(\frac{Y^c}{e}, \frac{v^c}{e}\right) = \max\left(0, \frac{Y^c}{e} + \frac{v^c}{e} - 1\right). \tag{13}$$

A max operator is used to find a node with maximum value in output layer. If the output layer node with maximum value represents the true class of P, then the weight parameters are not updated. Otherwise, the weight parameters are updated as

$$e_{\text{new}} = \gamma(1 - e_{\text{old}}) \tag{14}$$

and

$$v^c_{\text{new}} = v^c_{\text{old}} + \xi(1 - v^c{}_{\text{old}}) \tag{15}$$

where γ, $\xi \in [0,1]$ and these are constants.

This process is implemented by passing each feature vector in the training set for only one time. Accordingly, the learning process is named as "single pass." The single pass training in adaptive granulation- based network would reduce the computational time drastically. Further, the adaptive architecture of the network has less number of weight parameters (Example : v^{1c}, v^{2c},...., and v^{xc} for class c).

3) Testing of Adaptive Granulation-Based Network: In the testing stage, an unknown feature vector is feed-forwarded through the adaptive granulation-based network. The values at the nodes of output layer are obtained, as explained in Section II-D2. The unknown feature vector is assigned to the class denoted by the output layer node with maximum value.

a) Weighted membership: In real life situations, a dataset is likely to have unequal number of samples in its classes. To take this in account, we multiply the membership of a feature vector to a class with the relative frequency of occurrences of that class (class probability). We denote it as weighted class membership. In a two-class problem, the relative frequency of occurrence of class 1 (ρ_1) is

$$\rho_1 = \frac{n_1}{n_1 + n_2} \tag{16}$$

where n_1 and n_2 are the number of feature vectors belonging to class 1 and class 2, respectively. Similarly, the relative frequency of occurrence of class 2 (ρ_2) is

$$\rho_2 = \frac{n_2}{n_1 + n_2} \tag{17}$$

where $\rho_1 + \rho_2 = 1$. Therefore, if Y^c is the class membership of a feature vector for cth class, then its weighted membership would be $\rho_c Y^c$, where ρ_c is the relative frequency of occurrence of class c.

b) Roughness measure: In case of dataset having overlapping classes, an unknown feature vector may have maximum-membership value to more than one class, i.e., for more than one output node. This makes the decision-making indecisive (tie). The problem arises because the said unknown feature vector has equal membership to the granules of different classes, i.e., overlapping granules. In such situations, we use roughness measure of the overlapping granules to classify P. The overlapping granules are defined as the granules of different classes to which the feature vector P has nonzero membership values. Example, in Fig. 6(a), P has nonzero membership to the granules g^{32} and g^{2c}.

The roughness of an overlapping granule g is

$$R_g = 1 - \frac{|\underline{g}X|}{|\overline{g}X|} \tag{18}$$

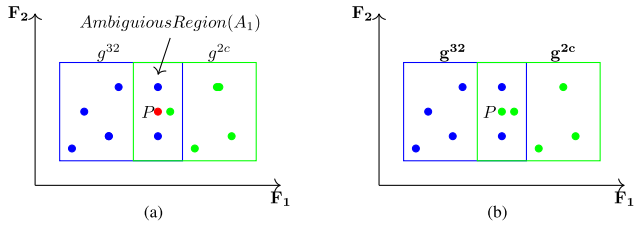

Fig. 6. (a) P in A_1 (Fig. 4) and (b) P is labeled to class c.

where $|\underline{g}X|$ is the cardinality of the lower approximation of g [(7)], $|\overline{g}X|$ is the cardinality of upper approximation of g [(8)], and X is the set of feature vectors in g. Lower approximation of g means the set of feature vectors which definitely belongs to g. The upper approximation of g denotes the set of feature vectors which definitely belongs to g, as well as the feature vectors in ambiguous region/overlapping region, together.

In tie situations, P is assigned to the class, where the roughness of the overlapping granule is minimum after adding P to that granule. For example, initially, the roughness values of granules g^{32} and g^{2c} (Fig. 6) are $R_{g^{32}} = (1\text{-}\frac{4}{6}) = \frac{1}{3}$ and $R_{g^{2c}} = (1\text{-}\frac{3}{4}) = \frac{1}{4}$, respectively. The roughness of the granules g^{32} and g^{2c} is then computed after adding P to g^{32} and g^{2c}, individually. It is found that the changed $R_{g^{32}} = \frac{3}{7} > R_{g^{2c}} = \frac{2}{5}$. This means, the changed-roughness of g^{2c} is less after adding P to g^{2c} as compared to that in class 2. P is therefore assigned to class c [shown in Fig. 6(b)] to resolve the tie. Note that when $|\overline{g}^{32}X| = |\overline{g}^{2c}X|$ and $|\bar{g}^{32}X| = |\bar{g}^{2c}X|$, $R_{g^{32}} = R_{g^{2c}}$ implies P cannot be labeled.

III. Results and Discussion

The performance of AGCNN is demonstrated with eight remote sensing image datasets. The description of these datasets is given in Table II.

1) Hyperspectral Remote Sensing Images: Four hyperspectral RS image datasets were considered in the present study to evaluate the performance of AGCNN model. The datasets are a) Hydice - Washington DC Mall (HWM), b) AVIRIS - Salinas (ASA), c) ROSIS - Pavia University (RPU), and d) AVIRIS - Indian Pines (AIP). The description of these datasets is given in Table II. The datasets consist of densely overlapped pixels in feature space. The distribution of densely overlapped pixels of the HWM dataset in F_1, F_2, and F_3 feature space is given in Fig. 7.

2) Multispectral Remote Sensing Images: In support to hyperspectral RS image datasets, four multispectral RS image datasets were used to test the performance of AGCNN model. These datasets are, a) Landsat OLI - Kolkatta (LOK), b) Sentinel MSI - Visakhapatnam (SMV), c) IRS LISS IV - Hyderabad (ILH), and d) IRS LISS III - Delhi (ILD). The description of these datasets is given in Table II.

A. Model Description

In the present study, six models were considered to demonstrate the performance of AGCNN. The criteria and description of six models are given as follows.

Algorithm 1: AGCNN for Remote sensing Image Classification.

1 During training,
for i = 1 to number of training images,
do,
(i) Perform CD granulation of each pixel (L_2),
(ii) Select efficient features (L_3),
(iii) Apply series of convolution and pooling layers (L_4, L_5, L_6, and L_7),
(iv) Obtain feature vector (L_8), implement 1(a) and 1(b)
(a) Create the granules/update the granules.
(b) Define the architecture of AG-based network.
2 During testing,
for j = 1 to number of test images,
do,
(i) implement steps (i), (ii), (iii) of 1,
(ii) Pass P through AG-based network, obtain values at the nodes of output layer,
(iii) If the values at the nodes are equal, implement 2(a) and 2(b),
(a) Compute roughness of overlapping granules by adding P, individually.
(b) Assign P to the class where roughness of a overlapping granule is minimum.
(iv) Else, assign P to the class of the output node with maximum value.

TABLE II
Description of HSRS and MSRS Image Datasets

Type	*Name*	Image size	Bands (p)	Resolution (meters)	Classes
HSRS	HWM	280 x 308	191	3	6
	ASA	512 x 217	204	3.7	15
	RPU	610 x 340	103	1.3	9
	AIP	145 x 145	200	30	16
MSRS	LOK	300 x 300	9	30	5
	SMV	300 x 300	13	10	5
	ILH	300 x 300	4	23.5	5
	ILD	300 x 300	4	5.6	5

Model framework : Granulation (L_2)+ Feature selection (L_3) + Number of convolutional and pooling layers (L_4 to L_8) + Type of NN + Learning + Membership + Decision-Making.

- *Model 1* (M_1)(Basic CNN): Raw pixels (Un-granulated) + None + 2 + multilayer perceptron + Back propogation + None + None
- *Model 2* (M_2) (New): CI + Reduct + 2 + multilayer perceptron + Back propogation + None + None
- *Model 3* (M_3) (New): CD + Reduct + 2 + multilayer perceptron + Back propogation + None + None
- *Model 4* (M_4) (New): CD + Reduct + 2 + AG-based network + Adaptive + Trapezoidal+ None
- *Model 5* (M_5) (New): CD + Reduct + 2 + AG-based network + Adaptive + Weighted + None
- *Model 6* (M_6)($Proposed$): CD + Reduct + 2 + AG-based network + Adaptive + Weighted + Roughness measure

M_1 is a basic CNN with ungranulated input image. M_2 is a CNN with CI granulated input image. M_3 is CNN with CD

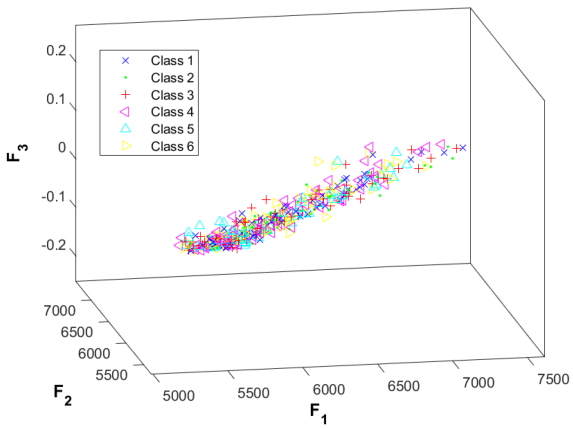

Fig. 7. Samples of HWM dataset in F_1, F_2, and F_3 space.

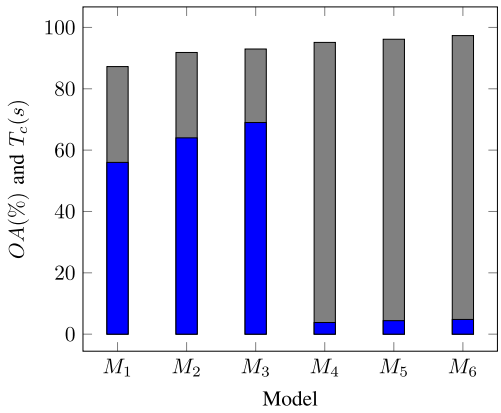

Fig. 8. Performances of M_1 to M_6 for HWM dataset. Gray color bar indicate OA(%)and blue color bar indicates $T_c(s)$ of models.

granulated input image. M_4 is a CNN with CD granulated input image, adaptive granulation-based network and trapezoidal membership. M_5 is similar to M_4 with weighted membership as additional feature. M_1 is basic CNN model and M_2, M_3, M_4, and M_5 are the new models introduced in the present study. M_6 is the proposed AGCNN model.

In the present study, a fourfold comparative analysis is implemented among the six models. In the first facet of comparison, the models M_1, M_2, and M_3 are compared to know the improvement in the performance of model due to CD granulation. In the second facet, comparison between M_3 and M_4 is performed to know the superiority of CNN with adaptive granulation over conventional CNN with CD granulated input. In the third facet, a comparison among M_4, M_5, and M_6 is performed to know the improvement in AGCNN with weighted class-membership and roughness measure in decision-making. In the fourth facet, M_6 is compared with M_1 to know the superiority of AGCNN over conventional CNN with ungranulated input. The models were implemented by the system with configuration Intel(R) Core(TM) i7-2600 CPU 3.40 GHz, 3401 Mhz, 4 Core(s), 8 Logical Processor(s).

B. Training and Testing of Models

Eight datasets are created by considering the image patches of the images given in Table II (In HWM dataset, we considered the size of image patch as $9 \times 9 \times 191$). The dataset is divided into two disjoint sets, namely, training set and test set. The parameters of models from M_1 to M_6 are computed using the training set. The performance of model is evaluated using test set. The division of dataset is done such that 20%, 40%, 60%, and 80% of the image patches are used for training and the remaining 80%, 60%, 40%, and 20% are used for testing the model.

TABLE III
ARCHITECTURE AND NUMBER OF WEIGHT PARAMETERS OF M_1-M_6 FOR HWM DATASET

Model	Input Image I(mxnxp)	Architecture (after L_8)	Number of weight paramaters	NI
M_1	9x9x191	191 : 50 :6	191x50+50x6 = 9850	80
M_2	9x9x191	204 : 50 :6	204x50+50x6 = 10500	80
M_3	9x9x191	216 : 50 :6	216x50+50x6 = 11100	80
M_4	9x9x191	216 : 108 : 6	108	1
M_5	9x9x191	216 : 108 : 6	108	1
M_6	9x9x191	216 : 108 : 6	108	1

C. Performance Metrics

In the present study, the metrics like, overall accuracy (OA), positive predictive (PP), sensitivity (Sn), F-score (F), number of iterations (NI), computational time (T_c), and dispersion measure (Dm) [33] are considered to quantify the performance of six models. Overall accuracy (OA) is the average of accuracies obtained for the models for 20%, 40%, 60%, and 80% of training. Furthermore, the class level performance analysis of the model is implemented using the metrics PP, Sn, F, I, T_c, and Dm. These metrics are derived from the confusion matrix (CM). PP (also called as precision/user's accuracy) is the ratio of predicted true positive divided by the sum of true positive and false positive. More the value of PP and Sn for a model, better the performance of model. Sn is the number of correctly predicted positives divided by the total number of predicted positives. F-score is the harmonic mean of PP and Sn. F-score lies between 0 and 1. The model with F score near to 1, indicates better performance. NI is the number of iterations for which the model is trained. The total time for training and testing the model is called computational time (T_c). Dm quantifies the distribution of classified feature vectors among the classes. Lesser the Dm value, better the performance of the model.

D. Performance of Models with Hyperspectral RS Datasets

The performances of six models were tested with four hyperspectral RS datasets.

1) Performance of Models With HWM Dataset: The architecture of six models after layer (L_8) and the number of weight parameters are given in Table III. In Table III, an input image I $(9\times9\times191)$ is feed-forwarded to all the six models. The MLP in M_1 has 191 input nodes, the number of hidden nodes were considered as 50, and the number of output nodes is equal to the number of classes (6). M_1 is trained for 80 iterations. M_1 does not have a granulation layer, reduction layer, and series of CLs and PLs. In M_2, image (I) is passed through a granulation layer, reduction layer, two CLs and PLs, and MLP. CD granulation of input image has increased the number of features and these

TABLE IV
EXPERIMENTAL RESULTS OF MODELS M_1 TO M_6 FOR *HWM* DATASET

	Training				Performance Metrics			
Model	*20%*	*40%*	*60%*	*80%*	*OA%*	*F*	*NI*	$T_c(s)$
M_1	85.32	86.48	87.38	89.79	87.24	0.52	80	56
M_2	89.83	90.88	92.73	93.93	91.84	0.56	80	64
M_3	90.52	92.78	93.83	94.77	92.97	0.65	80	69
M_4	93.38	94.84	95.07	97.22	95.12	0.71	1	3.8
M_5	94.44	95.67	96.68	97.84	96.15	0.78	1	4.4
M_6	95.87	96.78	97.77	98.94	97.34	0.89	1	4.8

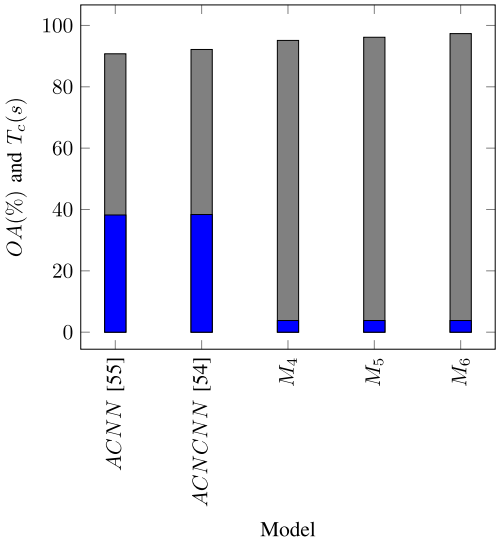

Fig. 9. $OA(\%)$ and $T_c(s)$ of recently published adaptive CNN models ACNN [55] and ACNCNN [54], and M_4, M_5, and M_6 for HWM dataset.

features are reduced in the feature selection layer. M_3 is similar to M_2 with CD granulated input image, the number of features generated using CD granulation is more (1146) compared with CI granulation (576). In the reduction layer, these features are reduced to 216 (for M_3) and 204 (for M_2), respectively. The number of input layer nodes (in M_3, M_4, M_5, and M_6) is equal to reduced features 216. The number of adaptive layer nodes is equal to the number of granules generated during the evolving process (108). In Table III, M_4, M_5, and M_6 have the less number of weight parameters which are updated during the evolving process (in a single iteration). M_1, M_2, and M_3 has more number of weight parameters. The parameters of M_1, M_2, and M_3 are updated by back propagating the output error for 80 iterations.

The performances of six models were tested with the HWM dataset and the results are given in Table IV. In the first facet of comparison, M_3 with CD granulated image obtained 92.97% OA which is 5.73% and 1.13% better than M_1 and M_2, respectively. The F-value of M_3 is 0.65 and it is comparatively better than M_1 and M_2. M_2 is 4.6% better than M_1 and it indicates improvement in the OA due to CI granulated input image. The CD granulation of pixels in the input image improved the performance by 1.13% compared with CI granulation of pixels in the input image. The computational time (T_c) of M_3 is 69 s while the T_c of M_2 and M_1 is 64 and 56 s, respectively. The CD granulation of input image generated more number

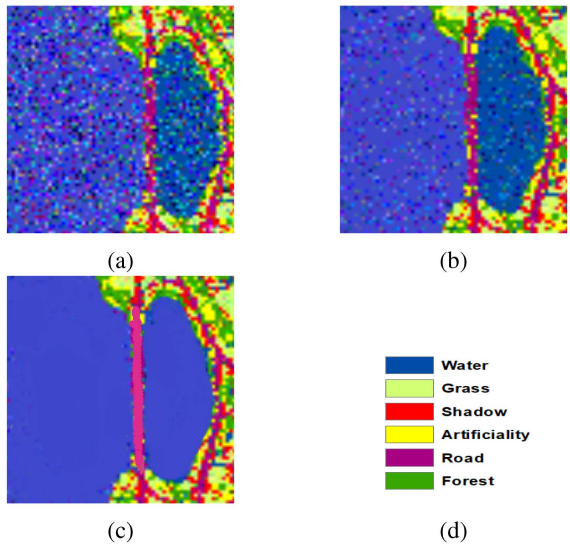

Fig. 10. Part of classified images of HSRS-HWM dataset (a) M_1, (b) ACNCNN [54], and (c) M_6, and (d) Class labels.

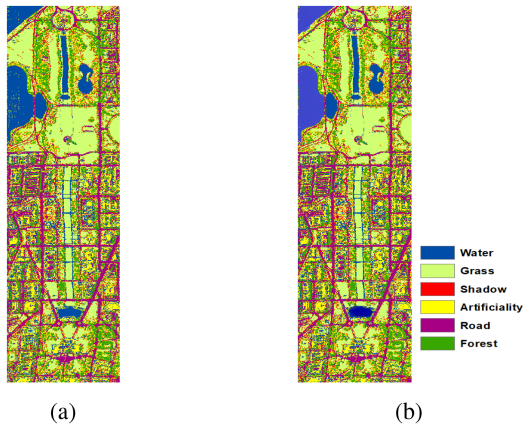

Fig. 11. Panoramic view of classified images obtained (a) HWM-M_1 and (b) HWM-M_6.

of features and this increased the computational time of M_3. The CD granulation of pixels in the input image improved the performance of CNN by 5.73% better than the basic CNN model (M_1). In the second facet of comparison, M_4 is 2.15% better than M_3. M_4 has less T_c (i.e., 3.8 s) and it is ten times better than M_3 in computational time. This significant drop in the computational time is credited to the adaptive granulation-based network with single-pass learning. The superiority of M_4 over M_3 in terms of OA and T_c is credited due to the adaptive granulation-based network architecture that can accommodate the information content of a new sample with a single pass learning. This indicates the superiority of adaptive granulation-based network architecture over fully connected neural network in terms of OA and computational time. The proposed adaptive granulation-based architecture improved the performance by 2.15% within 3.8 s. The superiority of M_4 over M_3 is justified with an F value 0.71. In the third facet, M_6 with weighted membership and roughness measure of overlapping granules in decision-making, produced 97.34% OA with a superiority

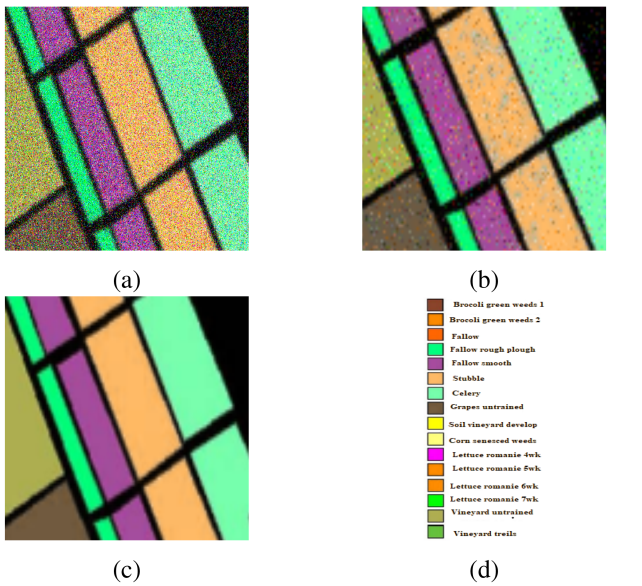

Fig. 12. Classified images of HSRS-ASA dataset. (a) M_1. (b) ACNCNN [54]. (c) M_6. (d) Class labels.

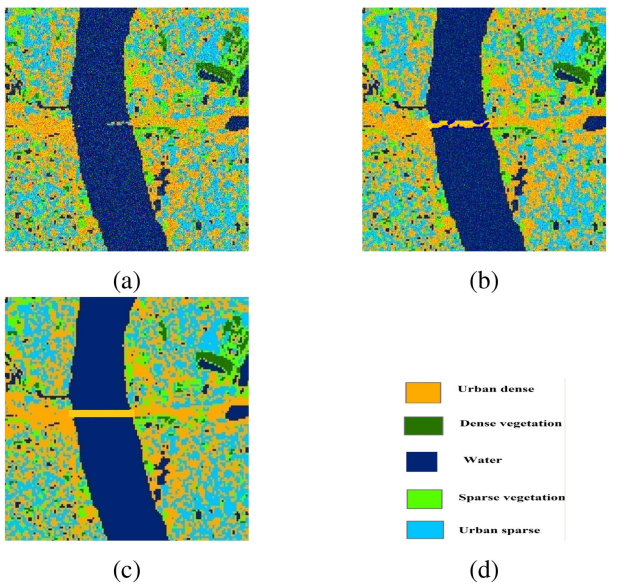

Fig. 13. Classified images of MSRS-LOK dataset. (a) M_1. (b) ACNCNN [54]. (c) M_6. (d) Class labels.

of 1.19% and 2.22% over M_5 and M_4, respectively. This indicates the proposed weighted class-membership and roughness of overlapping granules in decision-making of AGCNN improved the performance by 2.22%. In M_4, M_5, and M_6, the number of granules generated after training is 108. In the models M_4, M_5, and M_6, the number of weight parameters is 108 (equal to number of granules). These weight parameters are updated in a single iteration. Due to this reason, the T_c of M_6, M_5, and M_4 is less and it is within 10 s. The superiority of M_6 over M_5 and M_4 is justified with F value. In the fourth facet, M_6 is 10.1% better and it is 18 times faster than M_1. The superiority of M_6 over M_1 is credited due to CD granulation of input image, feature selection

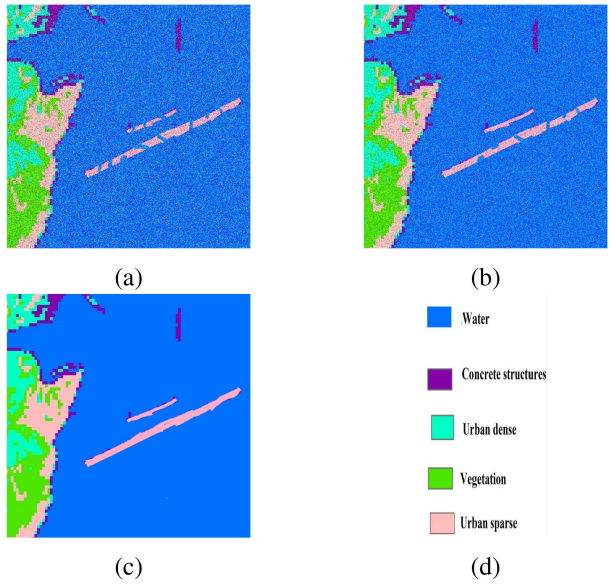

Fig. 14. Classified images of MSRS-SMV dataset. (a) M_1. (b) ACNCNN [54]. (c) M_6. (d) Class labels.

TABLE V
PERFORMANCE OF M_1 AND M_6 IN TERMS OF PP, Sn, AND Dm FOR SIX CLASSES (60% TRAINING)

	Model 1			Model 6		
Class	*PP* (%)	*Sn* (%)	*Dm*	*PP* (%)	*Sn* (%)	*Dm*
1	83.74	90.24	0.693	93.21	95.77	0.245
2	82.12	87.46	0.532	90.34	93.43	0.184
3	81.56	91.67	0.436	92.23	92.86	0.236
4	71.66	80.93	0.578	81.43	90.45	0.198
5	72.86	82.32	0.437	82.26	87.47	0.177
6	73.84	80.44	0.454	83.64	88.85	0.189

using rough set theoretic approach, adaptive granulation-based network with weighted membership and roughness measure in decision-making. AGCNN with the three novel characteristics like CD granulation of pixels in the input image, adaptive granulation-based architecture, and weighted class-membership and roughness measure of overlapping granules in decision-making produced a step-by-step increment in the OA like 5.73%, 2.15%, and 2.22%, respectively, in comparison with basic CNN model (M_1). Pictorial representation of OA(%) (gray bar) and T_c (s) (blue bar) from M_1 to M_6 is shown in Fig. 8. In Fig. 8, the computational time is drastically decreased from 69 s (M_3) to 3.8s (M_4). The decrease in computational time from M_3 to M_4 is due to the adaptive granulation-based network with single-pass learning characteristic in M_4.

The metrics like OA and F provides overall information of the models. In support to OA and F-score, the metrics PP, Sn, and Dm were considered to understand the class-level performance of five models. The performance of M_1 and M_6 in terms PP, Sn, and Dm is given in Table V. In Table V, M_6 has better PP, Sn for six classes compared with M_1. PP and Sn of M_6 for class 1 are 93.21% and 95.77%, respectively. The PP and Sn of M_1 for class 1 are 83.74% and 90.24%, respectively. The

TABLE VI
PERFORMANCE COMPARISON OF AGCNN WITH OTHER ADAPTIVE CNN CLASSIFIERS FOR HWM DATASET

Model	$OA(\%)$	F	T_c	NI
ACNN [55]	90.76	0.49	38.21	80
ANCNN [56]	91.83	0.58	28.36	80
ARCNN [57]	91.58	0.54	41.23	80
OACNN [58]	91.89	0.56	39.36	80
AKCNN [59]	91.24	0.51	38.78	80
AHF-CNN [60]	91.78	0.55	40.24	80
AMKFM-CNN [61]	92.03	0.59	39.24	80
AHA-CNN [62]	91.26	0.53	32.26	80
ACNCNN [54]	92.17	0.61	38.36	80
M_3	92.97	0.65	69	80
M_4	95.12	0.71	3.8	1
M_5	96.15	0.78	3.8	1
M_6	97.34	0.89	3.8	1

values PP and Sn for class 1 increased from M_1 to M_6. The superiority of M_6 over M_1 is justified with less Dm value.

a) Comparison with recently proposed other adaptive CNN models: The performance of AGCNN is compared with some recently developed adaptive CNN models and the results are given in Table VI. In Table VI, our proposed M_6 is seen to outperform all these comparing adaptive CNN models with a maximum OA of 97.34% and minimum T_c of 3.8 s. Among the comparing adaptive CNN models, ACNCNN ([54]) produced highest OA of 92.17% with Tc of 38.36 s and ACNN ([55]) produced lowest OA of 90.76% with T_c of 38.21 s.

One may note the evolution of improvement over the model M_4 to M_6 with additional features, vis-a-vis the best performed comparing method ACNCNN. For example, M_4 with adaptive granulation-based architecture is 2.95% better than ACNCNN interms of OA. It is ten times faster than ACNCNN. This superiority of M_4 over ACNCNN is due to the ability of the granulation-based network architecture to accommodate the information content of new samples in a single-pass learning. M_5 is 3.98% better than ACNCNN in terms of OA. It is also ten times faster than ACNCNN. The further increase in performance here is due to the weighted class-membership-based decision-making. M_6 that adopts the principle of roughness minimization in deciding on ambiguous patterns further enhances the difference to 5.17% with ten time faster in speed. While the comparing adaptive CNN models were trained for 80 iterations, the proposed AGCNN model is trained for a single iteration. The computational merit of AGCNN over other CNN-based models is due to less number of weight parameters in its architecture and single pass learning. The AGCNN has a few hundreds of weight parameters, while the other CNN models have thousands of weight parameters. AGCNN converges to a global minimum in a single iteration, while the comparing CNN models converge to minimum error in multiple number of iterations. These characteristics of AGCNN make it computationally faster than the existing models. Pictorial representation of OA(%) and $T_c(s)$ from M_1 to M_6 is shown in Fig. 8. Fig. 9 shows the same for the comparing deep CNN models with highest and lowest performance, as examples. It clearly depicts the superiority of the proposed model M_6, say, which has OA = 97.34% and T_c = 3.8 s, against OA of 92.17% with T_c = 38.36 s for the best performing recent adaptive model (ACNCNN).

TABLE VII
PERFORMANCE OF M_6 (BEST CLASSIFIER) AND M_1 (WORST CLASSIFIER) FOR OTHER HSRS DATASETS

Type	*Dataset*	*Model*	*OA%*	F	NI	$T_c(s)$
HSRS	ASA	M_1	90.16	0.67	80	57
		ACNCNN[54]	91.56	0.71	80	58
		M_6	97.86	0.88	1	4.3
HSRS	AIP	M_1	89.78	0.69	80	62
		ACNCNN[54]	92.86	0.75	80	59
		M_6	98.94	0.89	1	5.1
HSRS	RPU	M_1	88.24	0.65	80	55
		ACNCNN[54]	92.02	0.72	80	56
		M_6	97.62	0.90	1	7.1
MSRS	LOK	M_1	87.54	0.725	80	15
		ACNCNN[54]	90.23	0.72	80	14
		M_6	95.86	0.967	1	1.24
MSRS	SMV	M_1	88.46	0.684	80	16
		ACNCNN[54]	92.08	0.75	80	56
		M_6	96.84	0.974	1	1.32
MSRS	ILH	M_1	90.58	0.702	80	12
		ACNCNN[54]	92.67	0.73	80	61
		M_6	97.08	0.952	1	0.79
MSRS	ILD	M_1	88.76	0.693	80	12
		ACNCNN[54]	92.16	0.71	80	57
		M_6	98.47	0.938	1	0.73

For the purpose of visual inspection and comparison, we include some examples of classified output images of HWM, ASA, LOK, and SMV datasets corresponding to M_1(worst-performer), recently developed adaptive CNN (ACNCNN) [54], and M_6(best-performer). Fig. 10(a) and (d) show a (zoomed) part of such classified images by M_1, ACNCNN, M_6, and class labels for HWM dataset. As seen, the model M_6 could classify the objects in the image much better than M_1 and ACNCNN. M_1 is the worst among the three. For example, the road in Fig. 10(c) is clearly visible compared to those in Fig. 10(a) and (b). For the convenience of readers, as example, the panoramic view (of which the said image is a part) of the classified images by M_1 and M_6 for HWM dataset is shown in Fig. 11(a) and (b), respectively, corresponding to worst and best performance.

Similarly, the agriculture patches in ASA data (Fig. 12), the bridge in LOK data (Fig. 13), and concrete constructions in SMV data (Fig. 14) are seen to be much more distinctly extracted (identified) by M_6, as compared to other two deep models. All these corroborate the findings, as obtained by the aforesaid quantitative indices, and demonstrate further the superiority of the proposed deep model M_6.

b) Performance of models M_1 to M_6 with other datasets: The performances of the deep models M_1 to M_6 were also tested on other datasets. The results of M_1 and M_6 are given in Table VII as examples of best performing and worst performing models in terms of F-score, OA, and computation time. The superiority of M_6 over M_1 is also verified using other metrics.

IV. Conclusion

An adaptive granulation-based CNN model with eleven layers is proposed for remote sensing image classification. The model AGCNN (M_6) considers class belonging information of pixels in image patches through class-dependent granulation, and selection of informative features from the granulated ones in terms of rough set theoretic reducts. The model evolves its architecture automatically to accommodate the information content in new samples during the learning process. During classification, it incorporates the weighted class membership, and roughness measure on overlapping granules representing more than one class. The principle of roughness-minimization on overlapping granules enables appropriate class-assignment to doubtful pixels. The concept of using roughness measure in case of tie is novel. The adaptive granulation characteristics of AGCNN reduces the number of weight parameters and makes it able to learn in a single pass. In our study, the number of weight parameters is seen to be much less (viz., 108) in comparison to thousands of weight parameters in conventional CNN. Further, the single-pass training resulted in much reduced computational time (say, to few seconds), as compared to more than 100 s in conventional CNN.

All these features make the proposed model superior to many state-of-the art adaptive CNN-based models. For example, the AGCNN model M_6 produces 97.34%OA with T_c of 3.8 s for HWM dataset compared to OA of 92.17% with $T_c = 38.36$ s by the best performing adaptive model (ACNCNN) recently developed. The proposed model is ten times faster than the recently proposed adaptive CNN model ACNCNN, as tested with remote sensing datasets. Superiority of AGCNN is also demonstrated visually on the output classified images.

A. Limitations

In AGCNN, the feature vector in the overlapping regions of granules is assigned to a class based on the roughness of the constituting granules. This method can therefore label the sample even in the tie situation, as explained in Section II-D3b. However, the model cannot label the sample in the overlapping region if the granules have equal roughness. Also, during the training of AGCNN, the granules evolve to accommodate the input sample. During the evolving process of AGCNN, the size of the granule is increased by a constant value and may lead to the maximum overlapping of granules and hence to misclassification of the sample.

B. Futurescope

The model can be further improved by addressing the aforesaid two drawbacks. The concept of neighbourhood roughness of the overlapping granules can be incorporated to classify the ambiguous samples in the overlapping region for better class labeling. Further, the extent of granules during adaptive granulation can be controlled by the roughness of overlapping neighboring granules.

Acknowledgment

S. K. Pal acknowledges National Science Chair, SERB-DST, Govt. of India. The authors would like to thank Prof. Landgrebe and Prof. P. Gamba for providing hyperspectral remote sensing datasets.

References

[1] H. Zhang, X. Ma, X. Zhao, and G. R. Arce, "Compressive spectral image classification using 3D coded convolutional neural network," *Opt. Express*, vol. 29, no. 21, pp. 32875–32891, 2021.

[2] N. Girard, G. Charpiat, and Y. Tarabalka, "Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning," in *Proc. Asian Conf. Comput. Vis.*, 2018, pp. 675–690.

[3] Z. Chen, C. Wang, J. Li, N. Xie, Y. Han, and J. Du, "Reconstruction bias U-net for road extraction from optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2284–2294, 2021.

[4] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.

[5] J. Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," *Comput. Res. Repository*, vol. 1, pp. 1–22, 2016.

[6] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

[7] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.

[8] D. Marcos, M. Volpi, B. Kellenberger, and D. Tuia, "Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 96–107, 2018.

[9] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[10] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[11] D. A. Kumar, "Knowledge-based morphological deep transparent neural networks for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2209–2222, 2022.

[12] X. Wu, D. Hong, and J. Chanussot, "Convolutional neural networks for multimodal remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5517010.

[13] D. Hong, J. Hu, J. Yao, J. Chanussot, and X. X. Zhu, "Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 178, pp. 68–80, 2021.

[14] W. Li et al., "Classification of high-spatial-resolution remote sensing scenes method using transfer learning and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1986–1995, 2020.

[15] W. Li, H. Chen, Q. Liu, H. Liu, Y. Wang, and G. Gui, "Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2215.

[16] A. Ma, Y. Wan, Y. Zhong, J. Wang, and L. Zhang, "Scenenet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search," *ISPRS J. Photogrammetry Remote Sens.*, vol. 172, pp. 171–188, 2021.

[17] M. Joshaghani, A. Davari, F. N. Hatamian, A. Maier, and C. Riess, "Bayesian convolutional neural networks for limited data hyperspectral remote sensing image classification," *Comput. Vis. Pattern Recognit.*, vol. 1, pp. 1–9, 2022.

[18] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 683–697, Sep. 1992.

[19] F. Pacifici, F. Del Frate, W. J. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 331–335, Jul. 2008.

[20] D. Stathakis and A. Vasilakos, "Comparison of computational intelligence based classification techniques for remotely sensed optical image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2305–2318, Aug. 2006.
[21] A. Baraldi, E. Binaghi, P. Blonda, P. A. Brivio, and A. Rampini, "Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 994–1005, May 2001.
[22] Q. Jishuang, W. Chao, and W. Zhengzhi, "Structure-context based fuzzy neural network approach for automatic target detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, 2003, pp. 767–769.
[23] A. Kulkarni and S. McCaslin, "Knowledge discovery from multispectral satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 246–250, Oct. 2004.
[24] S. K. Meher, "Knowledge-encoded granular neural networks for hyperspectral remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2439–2446, Jun. 2015.
[25] S. K. Meher and N. S. Kothari, "Interpretable rule-based fuzzy ELM and domain adaptation for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5907–5919, Jul. 2021.
[26] L. Bastin, J. Wood, and P. Fisher, "Visualising and tracking uncertainty in thematic classifications of satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 5, 1999, pp. 2501–2503.
[27] J. L. Dungan, D. Kao, and A. Pang, "The uncertainty visualization problem in remote sensing analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, 2002, pp. 729–731.
[28] Y. Zhong, A. Ma, and L. Zhang, "An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1235–1248, Apr. 2014.
[29] C. Wu and X. Guo, "A novel single fuzzifier interval type-2 fuzzy c-means clustering with local information for land-cover segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5903–5917, 2021.
[30] Q. Chen, M. Huang, H. Wang, and G. Xu, "A feature discretization method based on fuzzy rough sets for high-resolution remote sensing Big Data under linear spectral model," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 5, pp. 1328–1342, May 2022.
[31] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521914.
[32] Q. Li et al., "A superpixel-by-superpixel clustering framework for hyperspectral change detection," *Remote Sens.*, vol. 14, no. 12, 2022, Art. no. 2838.
[33] S. K. Pal, S. K. Meher, and S. Dutta, "Class-dependent rough-fuzzy granular space, dispersion index and classification," *Pattern Recognit.*, vol. 45, pp. 2690–2707, 2012.
[34] S. K. Meher, "Explicit rough-fuzzy pattern classification model," *Pattern Recognit. Lett.*, vol. 36, pp. 54–61, 2014.
[35] J. Zhou, J. Qin, K. Gao, and H. Leng, "SVM-based soft classification of urban tree species using very high-spatial resolution remote-sensing imagery," *Int. J. Remote Sens.*, vol. 37, no. 11, pp. 2541–2559, 2016.
[36] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991.
[37] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Set Syst.*, vol. 90, pp. 111–127, 1997.
[38] S. K. Pal, B. U. Shankar, and P. Mitra, "Granular computing, rough entropy and object extraction," *Pattern Recognit. Lett.*, vol. 26, no. 16, pp. 2509–2517, 2005.
[39] S. K. Pal, D. Bhoumik, and D. Bhunia Chakraborty, "Granulated deep learning and Z-numbers in motion detection and object recognition," *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16533–16548, 2020.
[40] D. B. Chakraborty and S. K. Pal, *Granular Video Computing: With Rough Sets, Deep Learning and in IoT*. Singapore: World Scientific, 2021.
[41] A. Pramanik, S. K. Pal, J. Maiti, and P. Mitra, "Granulated RCNN and multi-class deep sort for multi-object detection and tracking," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 1, pp. 171–181, Feb. 2022.
[42] S. K. Pal, "Granular mining and Big Data analytics: Rough models and challenges," *Proc. Nat. Acad. Sci., India Sect. A: Phys. Sci.*, vol. 90, no. 2, pp. 193–208, 2020.
[43] S. K. Pal, A. Pramanik, J. Maiti, and P. Mitra, "Deep learning in multi-object detection and tracking: State of the art," *Appl. Intell.*, vol. 51, no. 9, pp. 6400–6429, 2021.
[44] Z. Zeng, X. Chen, Z. Song, and T. Liu, "Remote sensing scene classification by multi-granularity fused CNN," in *Proc. Int. Conf. Intell. Automat. Soft Comput.*, 2021, pp. 879–886.
[45] M. J. Bosco and G. Wang, "Multilevel and multi-granularity of remote sensing imagery application based on deep learning and machine learning algorithm," in *Proc. Int. Conf. Mach. Learn. Mach. Intell.*, 2021, pp. 111–119.
[46] D. Leite, I. Škrjanc, and F. Gomide, "An overview on evolving systems and learning from stream data," *Evol. Syst.*, vol. 11, no. 2, pp. 181–198, 2020.
[47] D. Leite, P. Costa, and F. Gomide, "Evolving granular neural network for fuzzy time series forecasting," in *Proc. Int. Joint Conf. Neural Netw.*, 2012, pp. 1–8.
[48] A. Alexandridis, H. Sarimveis, and G. Bafas, "A new algorithm for online structure and parameter adaptation of RBF networks," *Neural Netw.*, vol. 16, no. 7, pp. 1003–1017, 2003.
[49] R. M. Palnitkar and J. Cannady, "A review of adaptive neural networks," in *Proc. Southeast Conf.*, 2004, pp. 38–47.
[50] D. A. Kumar, S. K. Meher, and K. P. Kumari, "Adaptive granular neural networks for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1848–1857, Jun. 2018.
[51] D. A. Kumar, S. K. Meher, and K. P. Kumari, "Knowledge based progressive granular neural networks for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5201–5212, Dec. 2017.
[52] D. A. Kumar, K. P. Kumari, and S. K. Meher, "Progressive granular neural networks with class based granulation," in *Proc. IEEE Annu. India Conf.*, 2016, pp. 1–6.
[53] D. A. Kumar, S. K. Meher, and K. P. Kumari, "Fusion of progressive granular neural networks for pattern classification," *Soft Comput.*, vol. 23, pp. 4051–4064, 2019.
[54] W. Liang, Y. Wu, M. Li, and Y. Cao, "Adaptive multiple kernel fusion model using spatial-statistical information for high resolution SAR image classification," *Neurocomputing*, vol. 492, pp. 382–395, 2022.
[55] S. M. Jameel, M. A. Hashmani, M. Rehman, and A. Budiman, "Adaptive CNN ensemble for complex multispectral image analysis," *Complexity*, vol. 2020, 2020, Art. no. 8361989.
[56] A. Zhang, X. Yang, L. Jia, J. Ai, and Z. Dong, "SAR image classification using adaptive neighborhood-based convolutional neural network," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 178–193, 2019.
[57] H. Huang, C. Pu, Y. Li, and Y. Duan, "Adaptive residual convolutional neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2520–2531, 2020.
[58] J. Wang, Y. Zheng, M. Wang, Q. Shen, and J. Huang, "Object-scale adaptive convolutional neural networks for high-spatial resolution remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 283–299, 2020.
[59] C. Ding, Y. Li, Y. Xia, W. Wei, L. Zhang, and Y. Zhang, "Convolutional neural networks based hyperspectral image classification method with adaptive kernels," *Remote Sens*, vol. 9, no. 6, 2017, Art. no. 618.
[60] W. Ma et al., "A novel adaptive hybrid fusion network for multiresolution remote sensing images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5400617.
[61] X. Ding et al., "An adaptive capsule network for hyperspectral remote sensing classification," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2445.
[62] S. Pande and B. Banerjee, "Adaptive hybrid attention network for hyperspectral image classification," *Pattern Recognit. Lett.*, vol. 144, pp. 6–12, 2021.

Sankar K. Pal (Life Fellow, IEEE) received the first Ph.D. degree in radio physics and electronics from the University of Calcutta, Kolkata, India, in 1979, and the second Ph.D. degree in electrical engineering along with DIC from Imperial College, University of London, London, U.K., in 1982.

He is currently a National Science Chair, Government of India, and the President, Indian Statistical Institute (ISI). He is also a Distinguished Scientist and former Director of ISI, a former Distinguished Professor of Indian National Science Academy, and a former Chair Professor of Indian National Academy of Engineering. He founded the Machine Intelligence Unit and the Center for Soft Computing Research: a national facility in the institute in Calcutta. In 1975, he joined ISI as a CSIR Senior Research Fellow where he became a Full Professor in 1987, Distinguished Scientist in 1998, the Director in 2005–2010, and the President in 2022–2024. He was the first Computer Scientist and someone outside Statistics and Mathematics to become the Director of the Indian Statistical Institute since its inception in 1931. He is also the first Ex-Employee of ISI to get elected to hold the honorable Chair of President of ISI after its inception in 1931. He was with the University of California, Berkeley, CA, USA, and the University of Maryland, College Park, College Park, MD, USA, in 1986 and 1987, with the NASA Johnson Space Center, Houston, TX, USA, in 1990–1992 and 1994, and with US Naval Research Laboratory, Washington, DC, USA, in 2004. He has coauthored 21 books and about 500 research publications in the areas of pattern recognition and machine learning, image processing, data mining and web intelligence, soft computing, neural nets, genetic algorithms, fuzzy sets, rough sets, cognitive machine and bioinformatics.

Prof. Pal is the Fellow of the Academy of Sciences for the Developing World (TWAS), International Association for Pattern recognition, International Association of Fuzzy Systems, Asia-Pacific Artificial Intelligence Association, and all the four National Academies for Science/Engineering in India. He is a Member of the European Academy of Sciences and Arts. He introduced the Soft Computing concept and research in India. He visited 45 countries as a keynote/ invited speaker or an Academic Visitor. Since 1997, he has been a Distinguished Visitor of IEEE Computer Society (USA) for the Asia-Pacific Region, and held several visiting positions in Italy, Poland, Hong Kong, and Australian universities.

He was the recipient of the 1990 S.S. Bhatnagar Prize (which is the most coveted award for a scientist in India), 2013 Padma Shri (one of the highest civilian awards) by the President of India, and many prestigious awards in India and abroad, including the 1999 G.D. Birla Award, 1998 Om Bhasin Award, 1993 Jawaharlal Nehru Fellowship, 2000 Khwarizmi International Award from the President of Iran, 2000-2001 FICCI Award, 1993 Vikram Sarabhai Research Award, 1993 NASA Tech Brief Award (USA), 1994 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award, 1995 NASA Patent Application Award (USA), 1997 IETE-R.L. Wadhwa Gold Medal, 2001 INSA-S.H. Zaheer Medal, 2005-06 Indian Science Congress-P.C. Mahalanobis Birth Centenary Gold Medalfrom the Prime Minister of India for Lifetime Achievement, 2007 J.C. Bose Fellowship of the Government of India, 2013 INAE Chair Professorship, 2015 DAE Raja Ramanna Fellowship, 2015 INAE-S.N. Mitra Award, 2017 INSA-Jawaharlal Nehru Birth Centenary Lecture Award, 2018 INSA Distinguished Professorial Chair, 2020 National Science Chair, Government of India, and 2021 AICTE Distinguished Chair Professor. He is on the editorial board of 30 internationally well-known scientific journals in computer science and engineering, including several IEEE transactions. His Google Scholar *h*-index is 81 with more than 35 000 total citations.

Dasari Arun Kumar (Senior Member, IEEE) received the Ph.D. degree in spatial information technology from Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India, in 2019.

He was a Postdoctoral Research Associate with the Centre for Soft Computing Research, Indian Statistical Institute, Kolkata, West Bengal, India. He is working as Associate Professor with the Department of Electronics and Communication Engineering and Centre for Research and Innovation, KSRM College of Engineering, Kadapa, Andhra Pradesh, India. He worked as a Project Fellow with Indian Statistical Institute, Bangalore, Karnataka, India. He has authored or coauthored research papers in pattern classification, granular neural networks, deep learning and remote sensing in reputed publishers like Springer, Elsevier, and IEEE.