Journal of Geomatics Vol. 13, No. 1, April 2019

Study of discontinuity adaptive MRF models with kernel-based noise classifier

Ishuita SenGupta[1*], Anil Kumar[2] and Rakesh Kumar Dwivedi[1]
[1]College of Computing Sciences and Information Technology, TeerthankerMahaveer University, Moradabad, India
[2]Indian Institute of Remote Sensing, Dehradun, India
*Email: ishuitasengupta8@gmail.com;

(Received: Jan 12, 2019; in final form: June 14, 2019)

Abstract: The paper aims to determine the effect of adding contextual models and kernel functions with fuzzy based noise classifier with remote sensing data. The non-linearity between class boundaries can be handled through the kernel functions and contextual models eliminates the probability of isolated pixels. Nine different Kernel functions have been combined with conventional Noise Clustering without Entropy classification method (KNC), to classify data obtained from Landsat-8 and Formosat-2 satellites. For contextual support Markov Random Field (MRF) models were introduced with KNC. Standard regularization model (smoothness prior) and four Discontinuity Adaptive (DA) models (edge preserving priors) have been studied with KNC and abbreviated as KNC-S-MRF, KNC-DA1-MRF, KNC-DA2-MRF, KNC-DA3-MRF, KNC-DA4-MRF, respectively. An increase in overall accuracy has been observed when a comparative analysis has been done with the established Noise classifier.

Keywords: Kernel functions, Kernel based Noise Clustering without Entropy (KNC), Markov Random Field (MRF) models, Regularization Model, Discontinuity Model.

1. Introduction

Expansion of Remote Sensing applications have directed to availability of colossal quantity of data. Challenges to sustain the quality of such data also have increased, requiring more robust framework for processing and analysis of these data. Traditional classification techniques designate each pixel to a single land cover class resulting in a hard (or 'crisp') partitioning (Zhang and Foody, 2001). Due to coarse spatial resolution, more than one land-cover type may exist within a pixel, such a pixel is termed as mixed pixel (Foody, 1996) and ignorance of it resulted in a reduction in classification accuracy. Incorporation of mixed pixel has been facilitated in all stages of a classification process to produce accurate and meaningful land cover classifications from remote sensing images (Ibrahim et al., 2005).

The extensive use of fuzzy logic (Zadeh, 1978) for classification leads into soft classifiers. Among the most prominent fuzzy classifiers, Fuzzy *c*-mean (FCM) had been successfully used for estimation and mapping of sub-pixel level land cover composition (Foody, 2000, Fisher and Pathirana, 1990), although it failed to handle noise. Possibilistic *c*-Means (PCM) was developed to overcome the drawback of FCM as PCM was able to surpass the effect of hyperline constant found in FCM (Chawla, 2010). The challenging problem of noise removal was considered from different perspectives (Jolion and Rosenfeld, 1989, Krishnapuram and Freg, 1992) and among them Noise clustering was found to give the best performance (Dave, 1991; Dave 1993). Lately, it was proven that the Noise clustering algorithm is a generalization, where PCM and FCM are its special cases (Dave and Sen, 1997).

Studies related to spatial contextual information in the classification process illustrates improvement in the classifiers robustness against noise when compared to purely spectral based classification algorithm. (Krishnapuram and Keller, 1996; Foody, 2000). Inclusion of contextual information while classifying helps in removing isolated pixel problem. MRF based contextual methods were used for classification and fusion of multi-source data and it was proven that the classification accuracy has improved and is more reliable over other contextual methods (Solberg et al., 1996; Binaghi et al, 1997). A Robust Fuzzy *c*-Means (RFCM) algorithm was developed by adding contextual information to the objective function of FCM using MRF, while performing image segmentation of Magnetic Resonance Images of brain (Pham, 2001). An Adaptive Bayesian Contextual classifier, which combines the advantages of Adaptive classifier and Bayesian Contextual classifier demonstrated, using MRF modeling of joint probabilities of classes of each pixel and its neighborhood could improve the classification accuracy by mitigating the effect of Speckle error (Jackson and Landgrebe, 2002). Providing contextual support to Noise classifier was proposed earlier with the aim to overcome sensitivity of noise and outliers on the classification result using S-MRF or DA-MRF models (Harikumar, 2014). Integration of contextual information onto support vector machines classifier using MRF model was achieved by reformulating the prior energy function in terms of suitable SVM-like kernel expansion (Moser and Serpico, 2010).

The kernel methods map the input data to a higher dimensional space where the data turn out to be linearly separable (Awan and Sap, 2005). Studies related to kernels and assessment of fusion with fuzzy based classifier has been done earlier also. An unsupervised Kernel Noise clustering algorithm was also proposed (Chotiwattana, 2009) based on distances of kernel method (Gaussian and higher order polynomial) and was found to be relatively more resistant against noise. PCM (Possibilistic*c*-Mean) has been modified with KPCM by replacing Euclidean norm with Gaussian Kernel, resulting to increase in robustness to noise (Ganesan and Rajini, 2010). To deal with the drawback of fuzzy clustering

KFCM was introduced (Ravindraiah and Tejaswini, 2013). Local kernel like KMOD and inverse multiquadratic kernel as well as the global kernels were studied and incorporated to enhance the capability of FCM (Bhatt and Mishra, 2013). (Rhee et. al., 2012) proposed a kernel based possibilistic clustering technique, in which Fuzzy Kernel c-Means (FKCM) algorithm for initialization of PCM was used and PCM was modified using kernel induced metric replacing Euclidean distance measure and showed better results than FCM, PCM, and FKCM. Incorporation of eight kernels with Fuzzy *c*-Means classification to handle the nonlinearity among classes has shown improved accuracy (Byju, 2015). Entailing Kernels with fuzzy based classifier have shown effective results than the conventional ones.

The objective of present paper is to develop a novel method that combines the positives of spectral classification with the contextual spatial information. Supervised Noise Clustering has been opted as the base classifier, and adding nine different kernel functions as the distance functions with it lead to derive a kernel based classifier, termed as, KNC (Sengupta *et.al,* 2019). For contextual support Markov Random Field (MRF) models have been incorporated with KNC. Standard regularization model (smoothness prior) and four Discontinuity Adaptive (DA) models (edge preserving priors) have been studied with KNC and abbreviated as KNC-S-MRF, KNC-DA1-MRF, KNC-DA2-MRF, KNC-DA3-MRF, KNC-DA4-MRF, respectively. Image to image accuracy assessment has been formulated jointly with computation of overall accuracy using Fuzzy error matrix (FERM) of every kernel specified.

2. Study area and dataset used

The datasets used have been acquired from Landsat-8 and Formosat-2 satellites. Landsat8provides moderate-resolution imagery, from 15 meters to 100 meters, of Earth's land surface and operates in the visible, near-infrared, short wave infrared and thermal infrared spectrums. Formosat2 captures panchromatic and multispectral data simultaneously with 2meters and 8meters resolution respectively. The sensors' spectral wavebands specifications are enlisted in table 1a and b. The site for the study work is situated in Haridwar district in the state of Uttarakhand, India. Area extends from 29°52'49" N to 29°54'2" N and 78°9'43" E to 78°11'25" E. The site is identified with five land cover classes i.e. Water, Wheat, Forest, Riverine Sand, Fallow Land.

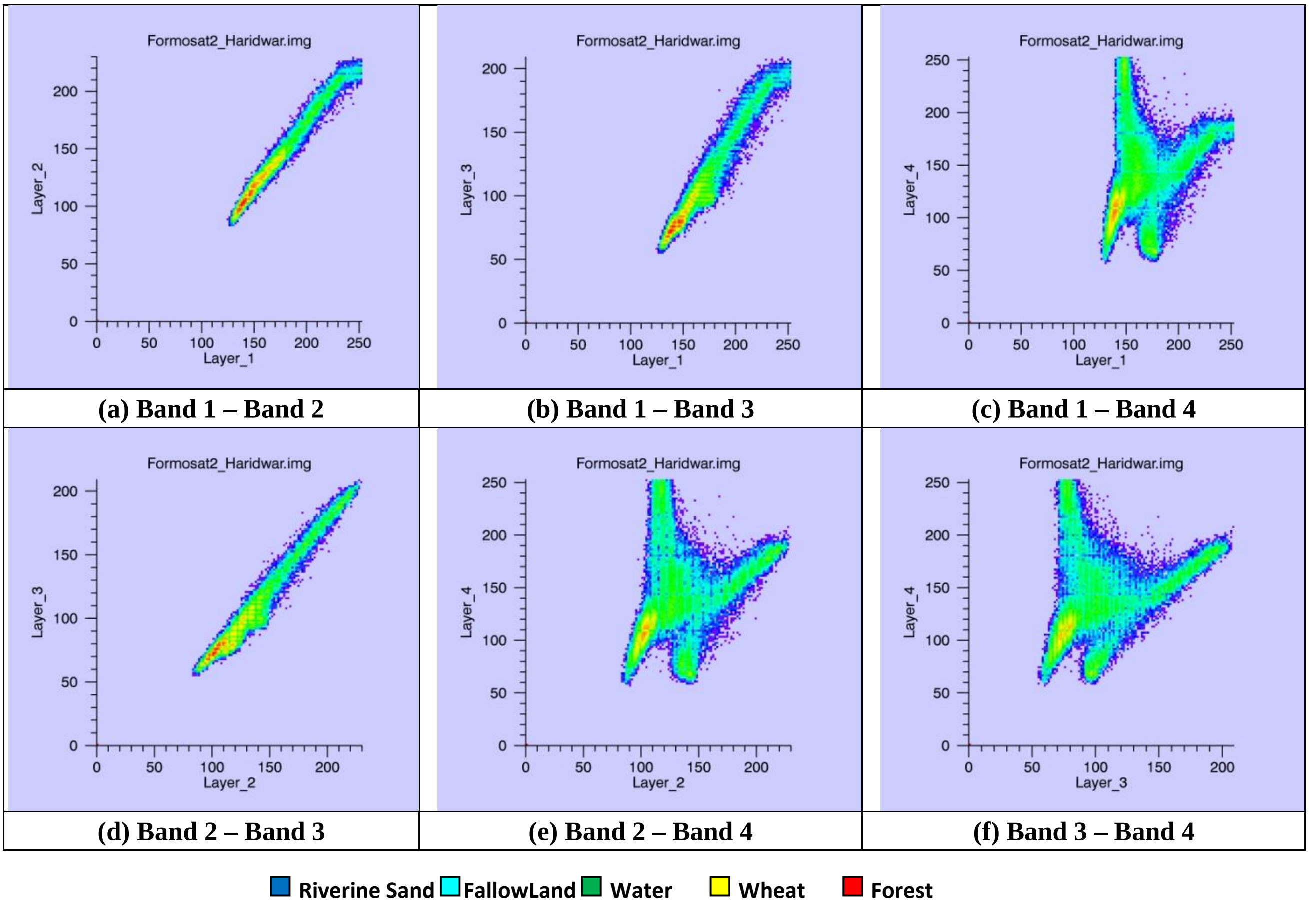


Figure 1: Non-linearity in different classes as 2D scatter plots for Formosat2 for all classes identified. (Generated using ENVI 5.0) (Sengupta *et.al,* 2019)

Table 1: (a) Data Details of Landsat8

Spectral Band	Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.43 - 0.45	30
Band 2 - Blue	0.45 - 0.51	30
Band 3 - Green	0.53 - 0.59	30
Band 4 - Red	0.64 - 0.67	30
Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
Band 6 – Short Wavelength Infrared 1	1.57 - 1.65	30
Band 7 –Short Wavelength Infrared 2	2.11 - 2.29	30
Band 8 - Panchromatic	0.50 - 0.68	15
Band 9 – Cirrus	1.36 - 1.38	30

Table 1: (b) Data Details of Formosat2

Spectral Band	Wavelength (μm)	Resolution (m)
Band 1 - Blue	0.45 - 0.52	8
Band 2 - Green	0.52 - 0.60	8
Band 3 - Red	0.63 - 0.69	8
Band 4 - Near Infrared (NIR)	0.76 - 0.90	8
Band 5 - Panchromatic	0.45 – 0.90	2

The 2D scatter plot in figure 1 shows presence of non-linear data in the specified dataset. Samples taken from the site cannot easily identify individual classes indicating the presence of non-linearity or cannot be separated linearly. While using 4thband as NIR (Near Infrared), there is drastic change of reflectance energy, either increasing for vegetation case or decreasing in water case that is why while using 4th band scatter plot is non-linear.

3. Hybrid classification – Kernel based noise classification with MRF Models

Incorporating contextual features with fuzzy based classifiers have quantified the classification. Markov Random Fields (MRF) used for modelling spatial contextual information and integrated into the objective function of the noise classifier and have shown positive impact in the classification accuracy (Harikumar, 2014). The novelty of the present work is to incorporate kernel methods with supervised Noise Clustering, and to integrate contextual MRF models with it.

3.1 Kernel methods used

The aim of kernel method is to identify a linearly separating hyperplane that separates the classes in higher dimensional feature space (Hofmann *et al*., 2008). The feature map (φ), given in Eq. (3.1), is the mapping function that non-linearly maps the data to a higher dimensional feature space and the kernel function (K), mentioned in Eq. (3.2), implicitly computes the dot product between two vectors $\boldsymbol{x}$ and $\boldsymbol{x}_i$ in higher dimensional feature space without explicitly transforming $\boldsymbol{x}$ and $\boldsymbol{x}_i$ to that higher dimensional feature space.

$$\Phi : R^p \to R^q \text{, where p<q} \quad (3.1)$$

$$K\left(\vec{x}, \vec{x}_i\right) = \phi(x).\phi(x_i) \quad (3.2)$$

A total of nine kernels functions have been considered in this study categorized as: four local kernels, three global kernels, spectral kernel, hypertangent kernel.

3.1.1 Local kernels

They are based on evaluation of the quadratic distance between training samples and the mean vector of the class. Only feature vectors that are close or in proximity of each other have an influence on the kernel value (Kumar, 2007). In this research, the value of the input vector was normalized between [0, 1] and thus acceptable result can be produced at "σ" equals 1. The different local kernels were defined as follows:

Radial Basis Function (RBF)

The RBF kernel is defined by exponential function as shown in equation (3.3). Here, $\boldsymbol{x}_i$ is the feature vector in the data and $\boldsymbol{v_j}$ is the mean vector of class j. σ determines the width of the kernel; a and b are the constants. By replacing a and b by 1 the Gaussian kernel can be obtained. In this study the value of a and b were taken to be 2 and 3 respectively (Kandpal, 2016).

$$K\left(\vec{x}_i, \vec{v}_j\right) = e^{\left(\frac{\left\|x_i^{\,a} - v_j^{\,b}\right\|^2}{2\sigma^2}\right)}\text{, where } \sigma, \text{a, b} > 0 \quad (3.3)$$

KMOD- (Kernel with Moderate Decreasing)

KMOD is the distance based kernel function (Ayat, *et al*., 2001) as shown in equation (3.4). It shows better result in classifying closely related datasets (highly correlated) and has shown better accuracy than Radial Basis Function (RBF) and polynomial kernel.

$$K\left(\vec{x}_i, \vec{v}_j\right) = e^{\left(\frac{\gamma}{\sigma^2 + \left\|x_i - v_j\right\|^2}\right)^{-1}}\text{, where } \sigma, \gamma > 0 \quad (3.4)$$

The parameter γ and σ controls the decreasing speed of the kernel function and the width of the kernel respectively. In this study the value of γ was taken to be one.

Gaussian

The Gaussian kernel is a special case of radial basis function kernel (Byju, 2015), shown in equation (3.5). Here, $\boldsymbol{x_i}$ is the feature vector in the image and $\boldsymbol{v_j}$ is the mean vector of the class.

$$K\left(\overrightarrow{x_i}, \overrightarrow{v_j}\right) = e^{\left(-\frac{\left\|x_i{}^a - v_j{}^b\right\|^2}{2\sigma^2}\right)}, \text{ where } \sigma > 0 \quad (3.5)$$

Inverse Multi-Quadratic (IMQ)

The inverse multi-quadratic kernel is defined as in equation (3.6) (Byju 2015; Kandpal, 2016). Here the value of c was taken to be one.

$$K\left(\overrightarrow{x_i}, \overrightarrow{v_j}\right) = \frac{1}{\sqrt{\left(\left\|x_i - v_j\right\|^2 + c\right)}}, \text{ where } c > 0 \quad (3.6)$$

3.1.2 Global kernels

In global kernels, the samples that are far away from each other have an influence on the kernel value. All the kernels which are based on the dot-product are global (Kumar, 2007). The different global kernels are as follows:

Linear kernel

Linear kernel is one of the simplest kernel functions. It is defined as the inner product of the input feature vectors, as shown in equation (3.7).

$$K\left(\overrightarrow{x_i}, \overrightarrow{v_j}\right) = x_i . v_j \quad (3.7)$$

Polynomial

The polynomial kernel is a positive definite kernel i.e. each element of the kernel matrix (a kernel matrix is a $n \times n$ matrix of feature vector) is positive, shown in equation (3.8). P defines the degree of the polynomial function and c is the constant (Kandpal, 2016). In this work value of P has been taken from 1 to 4. The value of c has been taken to be zero.

$$K\left(\overrightarrow{x_i}, \overrightarrow{v_j}\right) = \left(x_i . v_j + c\right)^P, \text{ where } c \geq 0 \quad (3.8)$$

Sigmoid

Sigmoid kernel is a hyperbolic tangent function, as shown in equation (3.9). The parameter α work as scaling parameter for the kernel function and defines width of the kernel. The best possible value for α and c were when $\alpha >$ 0and c< 0 (Byju, 2015).

$$K\left(\overrightarrow{x_i}, \overrightarrow{v_j}\right) = \tanh\left(\alpha . x_i . v_j + c\right) \quad (3.9)$$

3.1.3 Spectral kernel

The spectral kernel takes into consideration the spectral signature concept (Kandpal, 2016), as shown in equation (3.10). These kernels are based on the use of spectral angle ($\boldsymbol{x}$,) to measures the distance between the feature vector $\boldsymbol{x}$ and the mean vector of the class $\boldsymbol{v}_i$. It is expressed as follows**:**

$$\alpha(x, v_i) = \arccos\left(\frac{(x.v_i)}{\|x\|\|v_i\|}\right) \quad (3.10)$$

3.1.4 Hyper tangent kernel

The hyper tangent kernel is a hyperbolic tangent function, as shown in equation (3.11). The adjustable parameter σ defines the width or the scale of the kernel. Here x and v_i are the feature vectors in the data. It has been seen that the hyper tangent kernel outperforms other kernels when applied to a large data set (Kandpal, 2016).

$$K(x, v_i) = 1 - \tanh\left(-\frac{\| x - v_i \|^2}{\sigma^2}\right) \quad (3.11)$$

3.2 Markov Random Field (MRF) models

Contextual information refers to the relationship of an entity with its neighbourhood and in context of an image pixel; it refers to the information obtained from the neighbourhood pixels. Proper use of context can improve the classification accuracy (Jackson and D. A. Landgrebe, 2002; Solberg.*et.al,* 1996; Tso and Mather, 2009; Magnussen *et.al*., 2004). Markov Random Field (MRF) is a useful tool for modelling the contextual information and widely used to image segmentation and restoration problem (Besag, 1974; Li, 2009).

Study over MRF Models have been accomplished and propagated stating the relevance of neighbourhood pixel with local interaction (Harikumar, 2014). A prior in an image context, refers to the information about the image data available beforehand. Analytical regularizers are used for representing the prior energy. The general form of the regularizer is given in equation 3.12.

$$U(f) = \sum_{n=1}^{N} U_n(f) = \sum_{n=1}^{N} \lambda_n \int_n^b g\left(f^{(n)}(x)\right) dx \quad (3.12)$$

Here $U(f)$ is the prior energy represented using the nth order regularizers, $g\left(f^{(n)}(x)\right)$ is the Potential function that in turn is dependent on the irregularity in $f^{(n-1)}(x)$, N is the highest order considered and λ_n is the weighting factor and is always greater than or equal to 0. Over smoothening of the boundaries can lead to blurred image boundary, therefore, to control smoothening the Adaptive Potential Function (APF) placed within the regularizers and hence four different APFs have been used and hence four DA models. Table 2 demonstrates the mathematical models of the MRF Models to study in the present work.

Table 2: Mathematical Design of the priors (Li 1995; Li, 2009; Harikumar,2014).

Contextual Model	Mathematical Design of Priors
Smoothness Prior – S	$g\left(f^{(n)}(x)\right)=g(\eta)=\eta^2$
Discontinuity Adaptive Prior (Type 1)- DA1	$g_{1\gamma}(\eta)=(-\gamma e)^{-\frac{\eta^2}{\gamma}}$
Discontinuity Adaptive Prior (Type 2) - DA2	$g_{2\gamma}(\eta)=\frac{-\gamma}{1+\frac{\eta^2}{\gamma}}$
Discontinuity Adaptive Prior (Type 3)- DA3	$g_{3\gamma}(\eta)=\gamma\ln\left(1+\frac{\eta^2}{\gamma}\right)$
Discontinuity Adaptive Prior (Type 4)- DA4	$g_{4\gamma}(\eta)=\gamma\lvert\eta\rvert-\gamma^2\ln\left(1+\frac{\eta^2}{\gamma}\right)$

3.3 Kernel Based Noise Clustering without entropy classification (KNC)

The KNC classifier has been derived by using kernel methods with Noise clustering without entropy classifier (NC). The objective function of the NC in fuzzy mode (Dave, 1991; Hathaway *et.al*, 1996; Harikumar, 2014) expressed as shown in equation (3.13):

$$J_{NC}(U,V)=\sum_{i=1}^{N}\sum_{j=1}^{C}(\mu_{ij})^m D\left(\vec{x}_i,\vec{v}_j\right)+\sum_{1}^{N}(\mu_{i,c+1})^m\delta \quad (3.13)$$

.

Where $U=N\times C+1$ matrix, $V=(v_1\dots v_C)$, C is the number of classes, N is the total number of pixels in the image, m is the fuzzification factor and is normally positive (Sengar *et.al*, 2012), μ_{ij} represent the membership value of i^{th} pixel in the j^{th} class, $\mu_{i,c+1}$ represents the membership values of the noise class, v_j is the mean value (cluster center) of the j^{th} class, x_i is the vector value of the i^{th} pixel, D is the Euclidean distance between $\vec{x}_i$ and $\vec{v}_j$ and δ is a positive constant called the Noise distance.

Replacing the distance function D with (3.14), KNC objective function derives, as stated in equation (3.15) (Sengupta *et.al.* 2019).

$$D\left(\vec{x}_k,\vec{v}_i\right)=\left\|\varphi(x_k)-\varphi(v_i)\right\|=K(x_k,v_i) \quad (3.14)$$

$$J_{KNC}(U,V)=\sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}{}^m K(x_i,v_j)+\sum_{i=1}^{N}\mu_{i,c+1}{}^m\delta \quad (3.15)$$

Furthermore, KNC has been modelled with the smoothness prior, and discontinuity adaptive (DA) .Thus, the hybrid classifier mentioned from equation (3.16) to (3.20) will be referred as KNC S-MRF,KNC DA1-MRF, KNC DA2-MRF, KNC DA3-MRF and KNC DA4-MRF classifiers respectively. $U\left(\frac{u_{ij}}{d}\right)$, denotes the posterior probability, β is the weight factor associated with a pixel's neighbors and Nj represents the neighborhood window around pixel i

$$U\left(\frac{u_{ij}}{d}\right)=(1-\lambda)\left[\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^m K\left(\vec{x}_i,\vec{v}_j\right)+\sum_{i=1}^{N}(u_{i,C+1})^m\delta\right]+\lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{C}\sum_{j'\in N_j}\beta(u_{ij}-u_{ij'})^2\right] \quad (3.16)$$

$$U\left(\frac{u_{ij}}{d}\right)=(1-\lambda)\left[\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^m K\left(\vec{x}_i,\vec{v}_j\right)+\sum_{i=1}^{N}(u_{i,C+1})^m\delta\right]+\lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{C}\sum_{j'\in N_j}(-\gamma e)^{-\frac{\eta^2}{\gamma}}\right] \quad (3.17)$$

$$U\left(\frac{u_{ij}}{d}\right)=(1-\lambda)\left[\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^m K\left(\vec{x}_i,\vec{v}_j\right)+\sum_{i=1}^{N}(u_{i,C+1})^m\delta\right]+\lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{C}\sum_{j'\in N_j}\frac{-\gamma}{1+\frac{\eta^2}{\gamma}}\right] \quad (3.18)$$

$$U\left(\frac{u_{ij}}{d}\right)=(1-\lambda)\left[\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^m K\left(\vec{x}_i,\vec{v}_j\right)+\sum_{i=1}^{N}(u_{i,C+1})^m\delta\right]+\lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{C}\sum_{j'\in N_j}\left(\gamma\ln\left(1+\frac{\eta^2}{\gamma}\right)\right)\right] \quad (3.19)$$

$$U\left(\frac{u_{ij}}{d}\right)=(1-\lambda)\left[\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^m K\left(\vec{x}_i,\vec{v}_j\right)+\sum_{i=1}^{N}(u_{i,C+1})^m\delta\right]+\lambda\left[\sum_{i=1}^{N}\sum_{j=1}^{C}\sum_{j'\in N_j}\left(\gamma\lvert\eta\rvert-\gamma^2\ln\left(1+\frac{\eta^2}{\gamma}\right)\right)\right] \quad (3.20)$$

3.4 Mean-variance method

Verifying the edge preservation is significant; a standard method to analyze separately the distributions of grey levels of the two regions on either side of the edge, where the difference between the averages values within the two regions indicates the steepness of the edge (Wen and Xia, 1999).

The membership value of a unit in a fraction image is high if the pixel exists at a location of a known class (Class A) and for the unknown class it is low (Non-Class A), elaborated in figure 2. Consequently, the mean of the membership value will be high and the variance will be low in case of a homogeneous area for a known class location in a fraction image, leading to edge preservation. This concept has been used here to verify the edge preservation, to optimize contextual parameters.

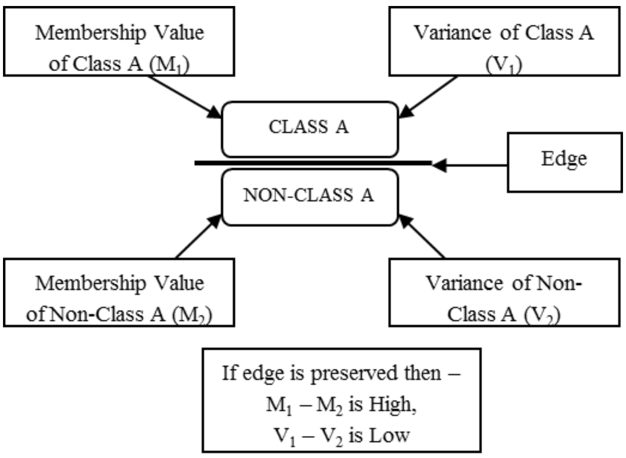


Figure 2: Method to verify edge preservation

4. Accuracy assessment techniques

To assess the discussed soft classifier, simulated image technique has been opted as well as FERM (Fuzzy error matrix) has been opted for computing the accuracy of KNC-S-MRF, KNC-DA-MRF models and NC S-MRF, NC-DA-MRF models. Accuracy assessment of sub-pixel classified output has been done with Java based tool (*Kumar et.al, 2006*).

4.1 Simulated image technique

The simulated image technique has been introduced to evaluate the fuzzy based classifier behaviour. The concept is to assign membership values to feature vectors from mean vector of the classes on the basis of distance measure. It is generated on the sample data for each class with desired number of bands. The technique facilitates to compare the classifier output with known input over defined location and also makes easy to identify the behaviour of classifier with the mixed pixels. Distribution is such that, the membership values of pure pixel in the classified output of a class must be maximized (close to 1). The mixed pixels were simulated with two variations, one with composition of 50:50 in between two different classes and other with composition of 30:30:40, the target membership value of 0.50, 0.40 and 0.30 is expected from the pixel with 50%, 40% and 30% belongingness for a class respectively (Figure 3).

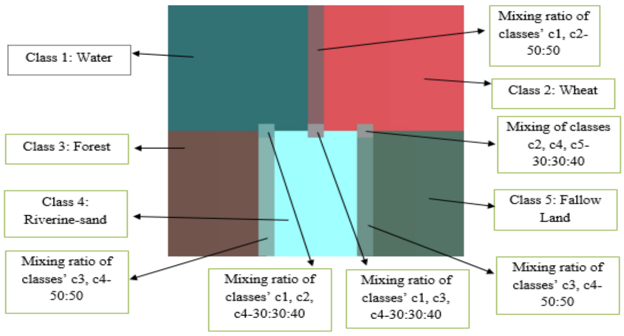


Figure 3: Simulated Image of Formosat 2 (Class Distribution)

4.2 FERM (Fuzzy Error Matrix)

It is a square array of positive fractional value varying between [0, 1]. The column R_N usually represent the sample elements assigned to the reference class *n* while the rows indicate the sample elements assigned to the classified class *m* (Binaghi *et al.*, 1999). The element in fuzzy error matrix (***M***) at row *m* and column *n* for a feature vector *x* is computed as shown in equation (4.1).

$$\mathrm{M}(m,n) = \sum_{x \in X} \min\left(\mu_{C_m}(x), \mu_{R_n}(x)\right) \qquad (4.1)$$

In equation (4.1), *x* is the overall sampled data set. μ_{C_m} and μ_{R_n} are the membership values for the classified and referenced data. The "min" operator is the traditional fuzzy set operator, it returns the minimum membership value between the classified and referenced data set for a class.

5. Results and analysis

5.1 Parameter estimation

The objective function of all KNC S-MRF and KNC DA-MRF classifiers involves certain parameters, which need to be initialized before the optimization of the membership values, thus, implementation of this hybrid classifier has been done in Java. Base classifier estimation has been done by a series of classification upon simulated image, using different kernels for every combination of m ranging between [1.1, 5.0] and the resolution parameter, δ, taken in the range of 10 to 10^6. Hybrid parameters have been estimated through simulated annealing (Bertsimas and Tsitsiklis, 1993) and mean variance method, initial T_0 has been set to 3 where optimized final temperature has been taken to be 0.90; λ has been defined in the range between 0 and 1, range of β to be 1 to 100, and that of Υ in between 0 and 1.

5.1.1 Base classifier parameter estimation

A series of kernel-based classification has been applied upon simulated image with every combination of defined*m* and δ. For brief demonstration, figure 4, displays the kernel wise membership values of wheat class, with varying δ, where it stops increasing at δ=10^4, here, *n* is representing the degree of resolution parameter and is related as $\delta = 10^n$. Remaining classes have also shown similar behaviour. Similarly, membership values have been computed for both pure and mixed pixel variations to optimize *m*. Table 3 demonstrates kernel wise membership values across varying *m* from 1.5 to 5.0. The tabular representation consists of three variations of pure pixel and mixed pixels as elaborated in section 4.1. Variation in optimized value of *m* have been seen from kernel to kernel therefore a specific range has been defined as an optimal range of *m* that is found to be [2.7, 5.0], leading to stability of membership values. Polynomial Kernels with degree 1 to 4 have shown poor performance leading to minimal membership values close to 0. Membership values computed, as shown in figure 4 and table 3 supports integer based values, hence, small fractional values rounded off to zero.

Table 3: Kernel wise membership value representation of wheat class with varying fuzzification factor (*m*) (a) Pure Pixel Composition (b) Mixed Pixel Composition (50:50) (c) Mixed Pixel Composition(30:30:40)

(a)

Kernels → m↓	Linear	Hypertangent	Gaussian	Sigmoid	KMOD	IMQ	Radial	Spectral	P1	P2	P3	P4
1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
2.5	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.10	0.00	0.00
2.7	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.89	0.98	0.12	0.00	0.00
3	0.94	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.94	0.16	0.00	0.00
3.5	0.86	0.86	0.86	0.86	0.85	0.91	0.86	0.86	0.86	0.20	0.00	0.00
4	0.78	0.79	0.79	0.79	0.77	0.78	0.78	0.75	0.78	0.24	0.00	0.00
4.5	0.71	0.72	0.72	0.72	0.70	0.71	0.71	0.69	0.71	0.25	0.01	0.00
5	0.64	0.65	0.66	0.66	0.64	0.65	0.65	0.64	0.64	0.26	0.02	0.00

***P1 – Polynomial (Degree 1), P2 – Polynomial (Degree 2), P3 – Polynomial (Degree 3), P4 – Polynomial (Degree 4)**

(b)

Kernels → m↓	Linear	Hypertangent	Gaussian	Sigmoid	KMOD	IMQ	Radial	Spectral	P1	P2	P3	P4
1.5	0.23	0.36	0.36	0.04	0.21	0.35	0.35	0.05	0.23	0.00	0.00	0.00
2	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.00	0.00	0.00
2.5	0.20	0.25	0.26	0.15	0.25	0.21	0.25	0.16	0.20	0.00	0.00	0.00
2.7	0.19	0.21	0.22	0.15	0.22	0.20	0.20	0.16	0.19	0.00	0.00	0.00
3	0.19	0.24	0.24	0.17	0.22	0.23	0.23	0.18	0.19	0.00	0.00	0.00
3.5	0.19	0.24	0.24	0.17	0.22	0.23	0.23	0.18	0.19	0.00	0.00	0.00
4	0.18	0.22	0.23	0.18	0.22	0.22	0.22	0.18	0.18	0.01	0.00	0.00
4.5	0.18	0.22	0.22	0.18	0.21	0.22	0.22	0.18	0.18	0.02	0.00	0.00
5	0.16	0.22	0.23	0.18	0.22	0.22	0.22	0.18	0.16	0.04	0.00	0.00

***P1 – Polynomial (Degree 1), P2 – Polynomial (Degree 2), P3 – Polynomial (Degree 3), P4 – Polynomial (Degree 4)**

(c)

Kernels → m↓	Linear	Hypertangent	Gaussian	Sigmoid	KMOD	IMQ	Radial	Spectral	P1	P2	P3	P4
1.5	0.01	0.02	0.02	0.02	0.04	0.03	0.03	0.01	0.01	0.00	0.00	0.00
2	0.05	0.07	0.07	0.07	0.07	0.07	0.07	0.05	0.05	0.00	0.00	0.00
2.5	0.11	0.13	0.14	0.14	0.15	0.14	0.14	0.11	0.11	0.00	0.00	0.00
2.7	0.11	0.15	0.14	0.16	0.15	0.15	0.15	0.14	0.11	0.00	0.00	0.00
3	0.13	0.15	0.16	0.15	0.16	0.16	0.16	0.13	0.13	0.00	0.00	0.00
3.5	0.14	0.16	0.16	0.16	0.17	0.16	0.16	0.15	0.14	0.01	0.00	0.00
4	0.14	0.17	0.17	0.16	0.17	0.17	0.17	0.15	0.14	0.02	0.00	0.00
4.5	0.15	0.17	0.18	0.17	0.18	0.18	0.18	0.16	0.15	0.03	0.00	0.00
5	0.15	0.18	0.18	0.17	0.18	0.18	0.18	0.16	0.15	0.04	0.00	0.00

***P1 – Polynomial (Degree 1), P2 – Polynomial (Degree 2), P3 – Polynomial (Degree 3), P4 – Polynomial (Degree 4)**

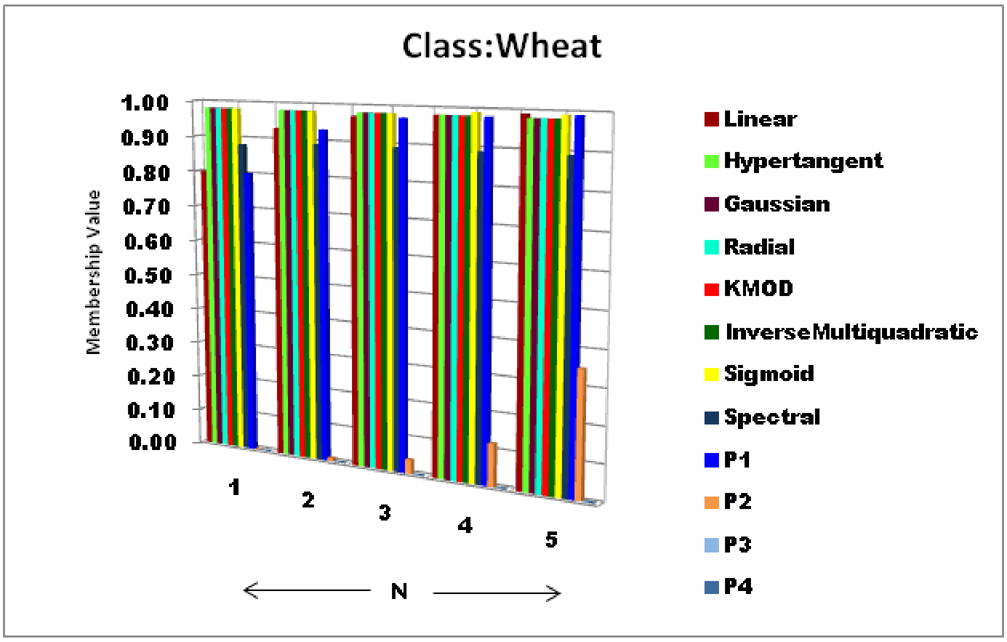


Figure 4: Kernel wise membership value representation of wheat class in accord with Noise Distance ($\delta = 10^N$). *P1 – Polynomial (Degree 1), P2 – Polynomial (Degree 2), P3 – Polynomial (Degree 3), P4 – Polynomial (Degree 4)

5.1.2 Contextual parameter estimation

Contextual parameters include, the weight factor which controls the spatial and spectral component (λ), neighborhood weight in the case of in case of S-MRF models (β) and constant involved in the case DA model (γ). Estimation has been done upon the fractional images of KNC-S-MRF and KNC DA-MRF classification for Landsat8. Table 4 displays the optimal range of hybrid parameters in case of KNC-S-MRF found to be λ lying between 0.7to 0.9, β=7 to 20, similarly in KNC DA1-MRF λ=0.1with Υ=0.1, KNC DA2-MRF λ=0.7 with Υ=0.7 to 0.9, KNC DA3-MRF λ=0.7, 0.8 with Υ=0.6 to 0.9 and KNC DA4-MRF λ=0.9 with Υ=0.8, 0.9.

Table 4: a) KNC estimation over edge verification for Landsat8 data (Class Water)

(Part – 1)

Classifier → Contextual Models ↓	Hybrid Parameters		Linear		Hypertangent		Gaussian		Radial		KMOD		Multiquadratic	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	-1	0	174	103	176	135	169	149	156	200	167	149
DA1-MRF	0.1	0.1	56	467	179	39	181	44	175	48	162	70	173	47
DA2-MRF	0.8	0.8	0	0	188	-25	187	-41	183	-49	169	21	185	-53
DA3-MRF	0.7	0.8	0	0	172	108	177	125	166	134	153	208	167	157
DA4-MRF	0.8	0.9	0	0	170	23	178	-89	170	-122	90	7784	165	55

(Part - 2)

Classifier → Contextual Models ↓	Hybrid Parameters		Sigmoid		Spectral		Polynomial Degree=1		Polynomial Degree=2		Polynomial Degree=3		Polynomial Degree=4		NC (Euclidean)	
	β/ϒ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	148	5752	132	387	-1	0	0	0	0	0	0	0	-1	0
DA1-MRF	0.1	0.1	184	15	133	142	56	467	0	0	0	0	0	0	56	467
DA2-MRF	0.8	0.8	182	39	134	138	0	0	0	0	0	0	0	0	0	0
DA3-MRF	0.7	0.8	177	124	134	44	0	0	0	0	0	0	0	0	0	0
DA4-MRF	0.8	0.9	60	8483	92	5382	0	0	0	0	0	0	0	0	0	0

***MD – Mean Difference, VD – Variance Difference**

Table 4: b) KNC estimation over edge verification for Landsat8 data (Class Wheat)

(Part – 1)

Classifier → Contextual Models ↓	Hybrid Parameters		Linear		Hypertangent		Gaussian		Radial		KMOD		Multiquadratic	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	0	0	155	103	158	140	151	148	137	159	150	150
DA1-MRF	0.1	0.1	25	113	164	-16	167	0	161	-4	147	2	159	3
DA2-MRF	0.8	0.8	-1	0	154	2	153	171	149	112	85	6134	145	194
DA3-MRF	0.7	0.8	-1	0	158	3	158	89	158	106	145	117	160	56
DA4-MRF	0.8	0.9	-1	0	148	77	140	-150	141	181	134	302	140	186

(Part – 2)

Classifier → Contextual Models ↓	Hybrid Parameters		Sigmoid		Spectral		Polynomial Degree=1		Polynomial Degree=2		Polynomial Degree=3		Polynomial Degree=4		NC (Euclidean)	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	165	161	52	3298	0	0	0	0	0	0	0	0	0	0
DA1-MRF	0.1	0.1	178	57	114	126	25	113	0	0	0	0	0	0	25	113
DA2-MRF	0.8	0.8	135	4814	117	165	0	0	0	0	0	0	0	0	-1	0
DA3-MRF	0.7	0.8	174	45	106	197	0	0	0	0	0	0	0	0	-1	0
DA4-MRF	0.8	0.9	163	179	106	195	0	0	0	0	0	0	0	0	-1	0

***MD – Mean Difference, VD – Variance Difference**

Table 4:c) KNC estimation over edge verification for Landsat8 data (Class Forest)

(Part – 1)

Classifier → Contextual Models ↓	Hybrid Parameters		Linear		Hypertangent		Gaussian		Radial		KMOD		Multiquadratic	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	0	0	178	54	182	61	176	76	166	98	176	73
DA1-MRF	0.1	0.1	48	305	178	54	181	61	176	76	166	98	175	72
DA2-MRF	0.8	0.8	-1	0	178	54	181	61	176	76	166	98	175	72
DA3-MRF	0.7	0.8	0	0	182	-4	186	-18	180	8	171	58	174	36
DA4-MRF	0.8	0.9	0	0	190	-106	193	-16	182	73	167	268	185	67

(Part – 2)

Classifier → Contextual Models ↓	Hybrid Parameters		Sigmoid		Spectral		Polynomial Degree=1		Polynomial Degree=2		Polynomial Degree=3		Polynomial Degree=4		NC (Euclidean)	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	181	64	92	98	53	0	0	0	0	0	0	0	0	0
DA1-MRF	0.1	0.1	179	74	92	98	48	305	0	0	0	0	0	0	48	305
DA2-MRF	0.8	0.8	179	74	92	98	-1	0	0	0	0	0	0	0	-1	0
DA3-MRF	0.7	0.8	194	-88	104	37	0	0	0	0	0	0	0	0	0	0
DA4-MRF	0.8	0.9	195	320	129	161	0	0	0	0	0	0	0	0	0	0

***MD – Mean Difference, VD – Variance Difference**

Table 4: d) KNC estimation over edge verification for Landsat8 data (Class Riverine)

(Part – 1)

Classifier → Contextual Models ↓	Hybrid Parameters		Linear		Hypertangent		Gaussian		Radial		KMOD		Multiquadratic	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	-1	0	183	119	195	95	181	141	163	207	181	138
DA1-MRF	0.1	0.1	32	117	189	117	200	114	187	149	168	206	188	148
DA2-MRF	0.8	0.8	0	0	183	122	195	121	182	154	163	200	184	228
DA3-MRF	0.7	0.8	0	0	175	48	186	56	173	97	154	165	174	80
DA4-MRF	0.8	0.9	0	0	144	5863	208	-83	111	10433	126	5108	79	10864

(Part - 2)

Classifier → Contextual Models ↓	Hybrid Parameters		Sigmoid		Spectral		Polynomial Degree=1		Polynomial Degree=2		Polynomial Degree=3		Polynomial Degree=4		NC (Euclidean)	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	201	-451	118	-73	53	0	0	0	0	0	0	0	-1	0
DA1-MRF	0.1	0.1	199	-492	118	-80	32	117	0	0	0	0	0	0	32	117
DA2-MRF	0.8	0.8	218	75	115	-167	0	0	0	0	0	0	0	0	0	0
DA3-MRF	0.7	0.8	206	-180	126	182	0	0	0	0	0	0	0	0	0	0
DA4-MRF	0.8	0.9	218	-184	130	118	0	0	0	0	0	0	0	0	0	0

***MD – Mean Difference, VD – Variance Difference**

Table 4: e) KNC estimation over Edge Verification for Landsat8 data (Class Fallow)

(Part – 1)

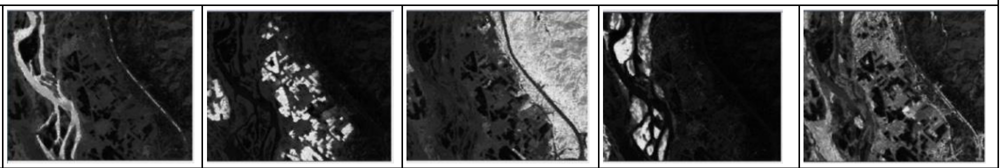
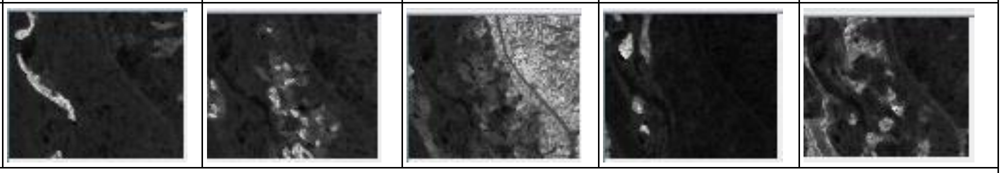
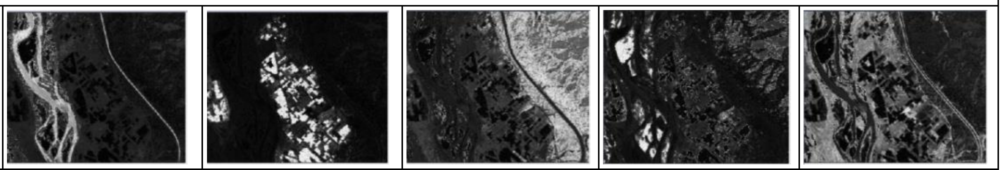
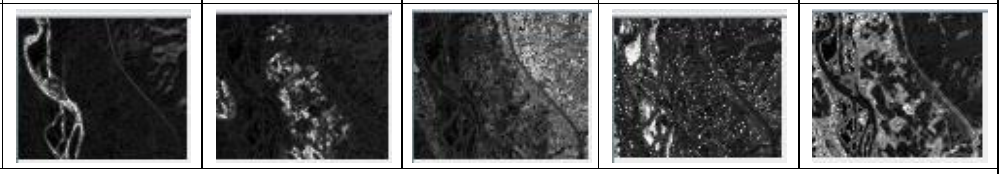
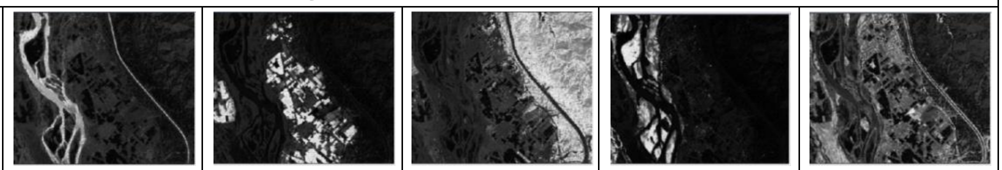
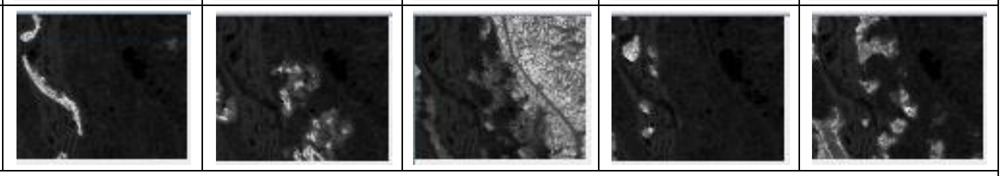
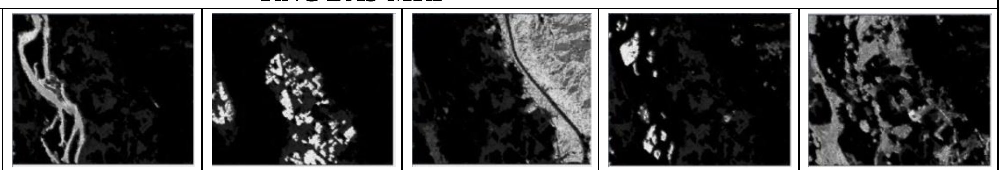
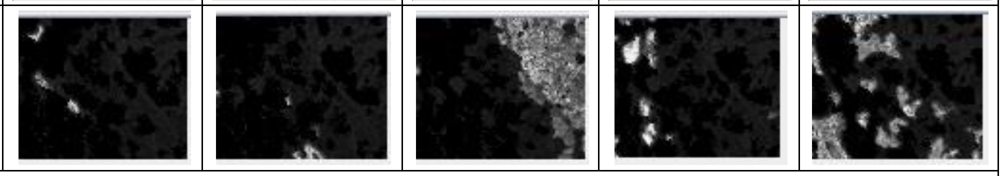
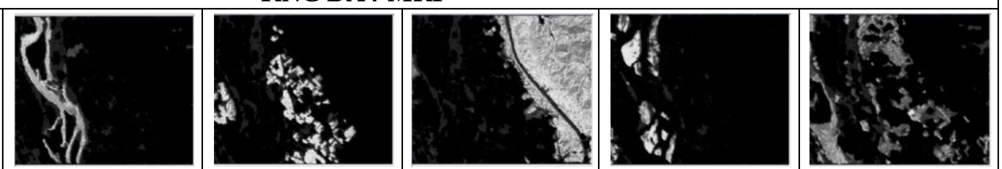
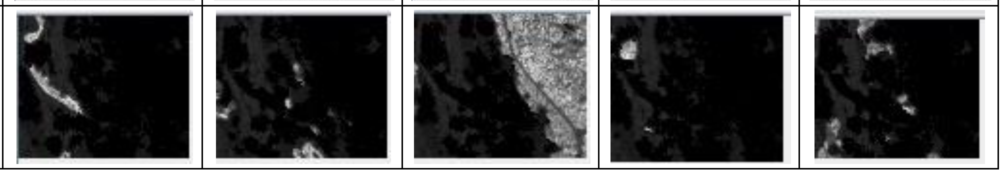
Classifier → Contextual Models ↓	Hybrid Prameters		Linear		Hypertangent		Gaussian		Radial		KMOD		Multiquadratic	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	0	0	132	4315	156	235	152	236	137	276	119	4543
DA1-MRF	0.1	0.1	20	55	162	115	162	138	156	161	142	172	154	141
DA2-MRF	0.8	0.8	0	0	170	75	165	102	167	126	67	6783	164	70
DA3-MRF	0.8	0.8	0	0	168	283	168	367	163	463	150	646	161	450
DA4-MRF	0.8	0.9	0	0	97	10454	129	11975	62	11023	-11	100	56	11030

(Part – 2)

Classifier → Contextual Models ↓	Hybrid Prameters		Sigmoid		Spectral		Polynomial Degree=1		Polynomial Degree=2		Polynomial Degree=3		Polynomial Degree=4		NC (Euclidean)	
	λ	β/ϒ	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD	MD	VD
S-MRF	0.8	20	158	24	19	637	0	0	0	0	0	0	0	0	0	0
DA1-MRF	0.1	0.1	168	49	63	111	20	55	0	0	0	0	0	0	20	55
DA2-MRF	0.8	0.8	170	-281	46	1019	0	0	0	0	0	0	0	0	0	0
DA3-MRF	0.8	0.8	167	94	64	2474	0	0	0	0	0	0	0	0	0	0
DA4-MRF	0.8	0.9	172	145	3	849	0	0	0	0	0	0	0	0	0	0

***MD – Mean Difference, VD – Variance Difference**

Table 5: Fractional images obtained from KNC S-MRF and KNC DA-MRF classifiers on Formosat 2 and Landsat 8 against the optimal parameters

Classifier	Water	Wheat	Forest	Riverine Sand	Fallow
KNC S-MRF					
Gaussian (Formosat2)					
Gaussian(Landsat8)					
KNC DA1-MRF					
Sigmoid (Formosat2)					
Sigmoid (Landsat8)					
KNC DA2-MRF					
Hypertangent (Formosat2)					
Hypertangent (Landsat8)					
KNC DA3-MRF					
Sigmoid (Formosat2)					
Sigmoid(Landsat8)					
KNC DA4-MRF					
Hypertangent (Formosat2)					
Hypertangent(Landsat8)					

5.2 Accuracy assessment

Studied contextual classification has been applied upon Landsat8 and Formosat2 image and fractional images generated with the optimized base, as shown in Table 5. FERM based accuracy assessment has been done using the output of Landsat8 and output of Formosat2 as reference map for KNC S-MRF and for KNC DA1-MRF, KNC D2-MRF, KNC DA3-MRF, KNC DA4-MRF models using all nine kernels. Table 6 shows the overall accuracy of best performing kernels along with the overall accuracy of conventional NC based MRF models. Analyzing the performances of various kernels with these contextual models, KNC DA1-MRF is found to be more promising from the perspective of classification accuracy. Overall accuracy of kernel based contextual models is better than the conventional Euclidean distance based contextual models.

Table 6: Accuracy assessment results for MRF based NC and MRF based KNC using single kernel

	CONTEXTUAL MODELS				
CLASSIFIER	S-MRF	DA1-MRF	DA2-MRF	DA3-MRF	DA4-MRF
NC (Euclidean)	4.00%	6.64%	0.79%	0.91%	0.90%
Gaussian Kernel	76.63%	81.09%	70.38%	59.42%	52.59%
Sigmoid	73.51%	75.38%	66.30%	51.04%	37.97%
Hypertangent	75.95%	78.75%	70.09%	60.91%	50.85%

6. Conclusions

The study focused to realize handling of non-linearity between class boundaries by integrating spatial features with kernel based noise classifier. This model takes account of involving MRF models into supervised kernel based Noise classifier (KNC). KNC S-MRF (smoothness prior) and four different Discontinuity Adaptive KNC DA1-MRF, KNC D2-MRF, KNC DA3-MRF, and KNC DA4-MRF models have been introduced and experimented to characterize both the contextual as well as spectral information. From the experiments performed we found that the KNC DA1-MRF model has performed better than the remaining models. Gaussian Kernel followed by Hypertangent Kernel has shown better output amongst the nine kernels. The study concludes that contextual MRF models when associated with kernel based have shown increase in accuracy as compared to Noise Classifier with Euclidean distance.

References

Ayat, N. E., M. Cheriet, L. Remaki and C. Y. Suen (2001). KMOD-A new support vector machine kernel with moderate decreasing. In Document Analysis and Recognition, (2001). Proceedings, Sixth International Conference, pp.1215–1219.

Awan, A. M., and M. N. M. Sap (2005). Clustering spatial data using a kernel-based algorithm. In Proceedings of the Annual Research Seminar, 306–310

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological), 192-236.

Bhatt, S. R. and P. K. Mishra (2013). Study of local Kernel with fuzzy C mean algorithm. International Journal of Advanced Research in Computer & Software Engineering, 3(12), 636–639.

Binaghi, E., P. Madella, M. Grazia Montesano and A. Rampini (1997). IEEE Transactions on Geoscience and Remote Sensing, 35, 326-340.

Binaghi, E., P. A. Brivio, P. Chessi and A. Rampini (1999). A fuzzy set-based accuracy assessment of soft classification. Pattern Recognition Letters, 20(9), 935–948.

Byju, A. P. (2015). Non-Linear separation of classes using a Kernel based Fuzzy c -Means (KFCM) Approach. MSc. Thesis, ITC, University of Twente, The Netherlands.

Chawla, S. (2010). Possibilistic-c-Means-Spatial contexutal information based sub-pixel classification approach for multi-spectral data. MSc. Thesis, University of Twente, Faculty of Geo-Information and Earth Observation (ITC), Enschede.

Chotiwattana, W. (2009). Noise Clustering algorithm based on Kernel method. Advance Computing Conference (2009). IACC 2009. IEEE, pp. 56-60

Dave, R. N. (1991). Characterization and detection of noise in clustering. Pattern Recognition Letters, 12, 657-664.

Dave, R. N. (1993). Robust fuzzy clustering algorithms. Second IEEE International Conference on Fuzzy Systems 1993,pp. 1281-1286.

Dave, R. and S. Sen (1997). Noise clustering algorithm revisited. Fuzzy Information Processing Society, NAFIPS'97. Annual Meeting of the North American, IEEE, pp.199-204.

Fisher, P. F and S. Pathirana (1990). The evaluation of fuzzy membership of land cover classes in the suburban zone. Remote Sensing of Environment, 34, 121-132.

Foody, G. M. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data, International Journal of Remote Sensing, 17(7), 1317-1340.

Foody, G. M. (2000). Estimation of sub-pixel land cover composition in the presence of untrained classes. Computers & Geosciences, 26(4), 469-478.

Ganesan, P and V. Rajini (2010). A method to segment color images based on modified Fuzzy-Possibilistic-C-Means clustering algorithm. Recent Advances in Space Technology Services and Climate Change 2010 (RSTS and CC-2010) , 157–163, 2010.http://doi.org/ 10.1109/ RSTSCC.2010.5712837.

Harikumar, A. (2014). The effects of discontinuity adaptive MRF models on the Noise classifier. MSc. Thesis, ITC, University of Twente, The Netherlands.

Hathaway, R. J., J. C. Bezdek and W. Pedrycz (1996). A parametric model for fusing heterogeneous fuzzy data, IEEE Transactions on Fuzzy Systems, 4, 1277-1282.

Hofmann, T., B. Scholkopf and A. J. Smola (2008). Kernel methods in machine learning. The Annals of Statistics, 36(3), 1171–1220.

Ibrahim, M. A., M. K. Arora, and S. K. Ghosh (2005). Estimating and Accommodating Uncertainty through the Soft Classification of Remote Sensing Data. International Journal of Remote Sensing, 26(14), 2995–3007.

Jackson Q. and D. A. Landgrebe (2002). Adaptive Bayesian contextual classification based on Markov random fields, IEEE Transactions on Geoscience and Remote Sensing, 40(11), 2454 – 2463.

Jolion, J. M. and A. Rosenfeld (1989). Cluster detection in background noise. Pattern Recognition, 22, 603-607.

Kandpal, N. (2016). Non-Linear Separation of classes using a Kernel based Possibilistic c -Means. MSc. Thesis, ITC, University of Twente, The Netherlands.

Krishnapuram, R. and C. P. Feg (1992). Fitting an unknown number of lines and planes to image data through compatible cluster merging. Pattern Recognition, 25, 385-400.

Krishnapuram, R. and J. M. Keller (1996). The possibilistic c-means algorithm: insights and recommendations. Fuzzy Systems, IEEE Transactions, 4, 385-393.

Kumar, A., S. Ghosh and V. Dadhwal (2006). Sub-pixel land cover mapping: SMIC system. ISPRS International Symposium “Geospatial Databases for Sustainable Development”, Goa, India, September 27-30, 2006.

Kumar, A. (2007). Investigation in sub-pixel classification approaches for land use and land cover mapping. Ph.D. Thesis. IIT Roorkee, India.

Li, S. Z. (1995). On discontinuity-adaptive smoothness priors in computer vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(6), 576-586.

Li, S. Z. (2009). Markov random field modeling in image analysis, Advances in Computer Vision and Pattern Recognition, Springer.

Magnussen, S., P. Boudewyn and M. Wulder (2004). Contextual classification of Landsat TM images to forest inventory cover types. International Journal of Remote Sensing, 25 (12), 2421–2440.

Moser, G. and B. S. Serpico (2010). Contextual remote-sensing image classification by support vector machines and Markov random fields. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3728-3731, 2010

Pham, D (2001). Spatial models for fuzzy clustering. Computer vision and image understanding. Computer Vision and Image Understanding, 84(2), 285-297.

Ravindraiah, R., and K. Tejaswini (2013). A survey of image segmentation algorithms based on Fuzzy clustering. International Journal of Computer Science and Mobile Computing, 2(7), 200–206.

Rhee, F., K. Choi and B. Choi (2012). Kernel approach to possibilistic C-means clustering. International Journal of Intelligent Systems, 24(3), 272–292. http://doi.org/10.1002/int.20336.

SenGupta, I., A. Kumar and R. K. Dwivedi (2019). Performance Evaluation of Kernel-Based supervised noise clustering approach, Journal of Indian Society of Remote Sensing., 47(2), 317–330.https://doi.org/ 10.1007/s12524-019-00938-2

Solberg, A. H., T. Taxt and A.K. Jain (1996). A Markov random field model for classification of multisource satellite imagery. Geoscience and Remote Sensing, IEEE Transactions, 34 (1), 100-113.

Tso, B. and P. M. Mather (2009). Classification of remotely sensed data. Second edition. Boca Raton, CRC, Press.

Wen, W. and A. Xia (1999). Verifying edges for visual inspection purposes, Pattern recognition letters, 20(3), 315–328.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 100, 9–34.

Zhang, J. and G. M. Foody (2001). A fuzzy classification of sub-urban land cover from remotely sensed imagery. International Journal of Remote Sensing, 19(14), 2721-2738.