Proceedings of the International Conference on Intelligent Computing and Communication for Smart World (I2C2SW-2018)
14-15 December 2018

Classification of Cloud/sky Images based on kNN and Modified Genetic Algorithm

S.Akila Rajini
Department of CSE
Anna University
Chennai, India
akila.cse.psncet@gmail.com

G.Tamilpavai
Department of CSE
Government College of Engineering
Tirunelveli, India
tamilpavai@gcetly.ac.in

***Abstract*—The main objective of this paper is to identify the types of clouds from the high-dimensional feature set extracted from sky/cloud images. The identification of types of clouds finds its scope in prediction of weather, natural disasters like storms, thunder storms etc. The feature set of sky/cloud images are fuzzy, and incomplete and highly spatial and temporal. Hence, it is necessary to optimize the feature set to for further analysis. Neighborhood Component Analysis (NCA) is used for optimizing the feature set extracted from sky/cloud images and it also improves the speed of classification. K-Nearest Neighbor (kNN) classifier is combined with Genetic Algorithm that uses Tabu search is used for classification. Finally, this paper identifies the following types of clouds. (i) Pattered cloud (ii) Thick dark cloud (iii) Thick white cloud and (iv)Veil cloud. kNN algorithm is modified to include condensed training set representation to improve the classification performance. The experimental results are verified with Singapore Whole sky imaging categories (SWIMCAT) database and achieves better performance in classifying the types of clouds.**

Keywords—Neighborhood Component Analysis, k-Nearest Neighbor Classifier, Genetic Algorithm, Tabu Search, Condensed Training set, Pattered cloud, Thick dark cloud, Thick white cloud, Veil cloud, SWIMCAT database

I. INTRODUCTION

The appearance of cloud keeps on changing and to work with such images is very critical. Such cloud images are difficult to capture and monitor. Whole Sky Imagers (WSIs) is a sky imaging device used to capture upper hemisphere of the sky images. The observed sky images are in hyper-spectral band containing high spatio-temporal features. From the captured sky images, high-dimensional features like cloud fractional coverage, clouds' base height, sky polarization, texture of cloud, shape, and radiance of clouds over the entire sky. The data associated with such sky images are incomplete, vague, and unclear. Hence, such high-dimensional feature set need to be optimized for selecting the discriminating features from the observed fuzzy feature set. Feature set help the image processing to work with great computational efficiency and enhanced performance. As sky imaging contains high dimensionality of the data, learning from such data is highly inefficient. Moreover, it is obvious that noise in the images reduces the accuracy and performance of algorithms. Therefore, it is necessary to select discriminative features from noise-free input imagery.

The captured sky images have issues in extracting features such as altitude related information of cloud because of poor spatial resolution. Additionally, there exist different views in cloud features associated with satellite and sensors. Images captured using downward-pointing satellite sensors have solar obstruction. In order to overcome these issues, the input imagery need to undergo basic image preprocessing works.

WSI produces images with dynamically changed data of cloud formations at regular intervals of time. There are many algorithms for computing the fraction of the sky covered by clouds, identifying cloud types, finding cloud base height etc. It is easy to identify the type of cloud based on blue/red ratio. In this paper the statistical features obtained from red and blue channels are used for classification of cloud types.

Hence, an analysis on color channels is essential to deal with the hyper spectral data of sky images by representing them in low dimensional subspace and thereby proceed for classification. Principal component analysis (PCA) is used to determine the discriminating features from the observed low dimensional subspace. However, PCA has several drawbacks with classification and hence in our work Neighborhood component analysis is preferred.

Neighborhood Component Analysis (NCA) is most suitable for k-Nearest Neighbor (k-NN) classifier. NCA uses the class labels again in classification process to obtain maximum reparability among the interclasses. And such dimensionality reduction procedure enhances the performance of classification.

NCA is used for dimensionality reduction of captured rough set for further processing in weather prediction, and by meteorological department. NCA applies linear transformation on the input feature space in order to reduce its dimensionality by learning. The learning process finds the neighbor to reduce the dimension of feature vector. And stochastically selected neighbor is assigned with a vote that can be used to speed up the classification process. Such dimensionality reduction is essential as it identifies the discriminating features that are essential for classification and thereby reduces the complexity in computation. The neighborhood is selected based on the distance measure. NCA offers superior performance for the classification of hyper-spectral images of cloud/sky.

The k-nearest neighbor (kNN) method is a predictive classifier that can be efficiently used to classify sky images into the variety of cloud types based on some similarity measures. Euclidean distance is used to find the similarity or closeness of two learnt images. Similar images are assigned with class labels as per the type of the cloud. The K-NN is not appropriate for noisy, or redundant training dataset. Hence, in this paper for cloud/sky images, NCA is applied first and for the reduced feature space, kNN is applied. kNN analyzes the training data for noise and outliers. Training data is condensed by removing the outliers and including the prototype and absorbed points that are essential for classification. kNN then works with condensed training images and evaluates its performance on the test images.

II. Related Work

Many distance metrics are available in order to estimate closeness of feature set that can be further used for classification of clouds in sky images. Among the available metrics, Kullback-Leibler (KL) distance measure plays a high scope in band selection in hyperspectral imaging.

For sky/cloud segmentation using color channels of hyperspectral images bimodality [2] and Principal component analysis [3] are used to select the favorable channel. Yet, there are certain issues like correlation among the color channels during cloud segment results in poor accuracy.

Recently, hybrid thresholding algorithm [4] uses the color channels especially red-blue channel for identifying the types of cloud from ground-based sky images.

Color plays a vital feature in cloud segmentation of hyperspectral imaging of sky images. Since color features are discriminating features that identifies the cloud in the sky. RGB color space model is used in segmentation by creating binary masks[5].There is another method that uses atmospheric pressure [6] for identification and classification of clouds using red and blue channels. Superpixel-based segmentation [7] also uses red and blue channels for the process of identification of clouds using sky images. Saturation(S) [8] channel is used for calculating cloud coverage and thereafter segmentation is applied over the color channel for extracting the discriminating color channel of sky images.

There exists a model that used the locus [9] of cloud pixels in RGB color channel extracted from the several color channels of sky images used to detect the existence of cloud. An adaptive threshold technique [10] identifies the presence of cloud with discriminating features extracted from red and blue channels of sky images. From the extracted feature set of sky images, superpixel based classification [11] techniques provided good accuracy of classification of types of clouds. For the process of cloud segmentation, several color models are constructed. Using the color models, several color channels are observed. To select red blue channel in particular, Rayleigh method [12][13][14] is used. But in this method, manually-defined parameters and thereafter decision-making process based on the parameters shows the efficiency of cloud segmentation well. But these methods are error-prone. After segmentation, labels are assigned for each classification and these algorithm shows as the best, flexible and robust.

There exists a method that encodes the textural features of input aerial imagery [15] and used in landscape image segmentation [16]. The accuracy produced in such segmentation is high. Additionally, Schmid filters are also applied to identify the types of clouds in sky/cloud images [17].When Red-Blue ratio of the color channel is used for segmentation, it had resulted in good accuracy and it is being tested with other statistical tools [18][19][20] and techniques on 16 different color channels for the task of sky/cloud image segmentation. The image database that is publicly available for sky/cloud segmentation is the HYTA database [21]. It consists of 32 images captured under various scenarios by varying illumination conditions. It also includes binary segmentation ground-truth images.

According to the literature review, most of the methods used for cloud classification are pixel-based. [23] [24] [25] [26] It uses the color channel and estimates the red-to-blue ratio (RBR) of each pixel. Next a threshold is applied to RBR which in turn determines cloud pixels. The pixels with low RBRs are classified as clear sky and higher are labelled as clouds. The neural network [27] is also used as the best classifier to classify the types of clouds in sky images. Prototype selection chooses a representative pattern from the training data. This is NP-hard problem [28], and hence, existing algorithms produces acceptable solutions.

A Support Vector Machine (SVM) [29] has been used in cloud classification model with high temporal and spatial resolution sky images. This method considered shape, radiance features of input images. It has identified four different classes of clouds. To improve further, the spectral and textural features are extracted by the statistical tonal analysis and gray level co-occurrence matrix (GLCM).Then support vector classification model with radial basis function (RBF) kernel function is built to classify the different clouds in the sky images. But this method is not accurate with regard to cloud dynamics. Extreme Learning Machine (ELM) [30], a soft computing technique uses a single hidden-layer feed-forward network. It randomly selects the number of hidden layer neurons in the network, the input weights and hidden layer. The output layer weights has been calculated by the least squares method. ELM has fast learning speed. It used gradient-based learning algorithms. To have high classification accuracy and reduce the computational complexity, more than one hidden layers can also be included. ELM works better only when more features are combined for classification.

Artificial Neural Network [31] algorithm used spectral, texture, size features and so on. Gray level co-occurrence matrices (GLCM) method is used to reduce the dimension to 15 features with 3 categories. Self-Organizing Map (SOM) technique is used and it geometrically transforms the input two-dimensional discrete map. SOM is simple and unsupervised self-organizing process that is easy to implement.

Table I. Summary of survey

Technique /Method	Features used	Images used
ELM[29]	Texture features, Color features and SIFT features	Ground-based total-sky cloud Imager

		(TCI)
SVM[30]	The spectral and textural features	The total sky imager
ANN – SOM [31]	Spectral features, gray features, assemblage features	Geostationary meteorological satellite FengYun-2C (FY-2C) data.

Table I shows the survey on the three algorithms depicting their implementation with the types of images used along with the features used for classification.

III. Data And Methods

A. Dataset – SWIMCAT (Singapore Whole sky IMagingCATegories Database)

Public benchmark database related to sky/cloud images are rare. There exists a database called Hybrid Thresholding Algorithm (HYTA) containing sky/cloud images. It also contains their segmentation masks.

There exists another database for sky/cloud image classification called as SWIMCAT and this database has been used in this proposed work of classification of cloud types in sky images. The images of SWIMCAT dataset were captured in Singapore using ground-based whole sky imager. The images were captured from January 2013 to May 2014. This database includes 784 images of sky/cloud patches with five categories namely clear sky, patterned clouds, thick dark clouds, thick white clouds, and veil clouds. Each patch has a dimension of 125x125 pixels. The categories of cloud as per SWIMCAT dataset are listed in the table, Table II.

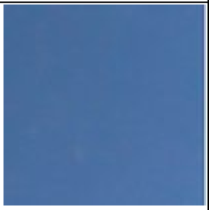
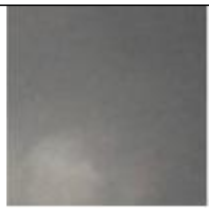
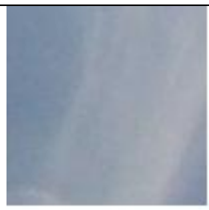
The entire work in this paper is depicted in Figure 3.1. The proposed work accepts the sky/cloud images as input. The input images are highly spectral and dimensional. Extract the feature vectors from such hyperspectral images using Fast Fourier Transformation. The analysis of extracted features shows that the data are time variant and it includes many noisy patterns and hence the data are claimed as rough set. Using Neighborhood Component analysis, the dataset is optimized. As a next step, before passing the optimized data to the kNN classifier, again it is condensed for training dataset to be in reduced dimension. As kNN works efficient with smaller dataset, the condensed dataset with prototype and observed points shows that identifying the types of clouds in sky/cloud images will be efficient.

The kNN algorithm is improved by implementing hybridization of Genetic algorithm and Tabu search producing class labels with good accuracy.

B. Feature Set Optimization using Neighborhood Component Analysis

The input sky/cloud images are hyperspectral and are with high dimensions. And such data are time-variant and contains missing patterns, partially incomplete data, and highly dynamic data. Hence, for further processing of such input images, the entire dataset need to be reduced. Among the various Dimensionality reduction methods, Neighborhood Component Analysis produces accurate result in eliminating the redundancy of dataset and are useful for kNNclassifiers. Being a supervised learning algorithm, Neighborhood component analysis (NCA) accepts hyperspectral images and builds a decision table and then reduces the table. NCA produces higher class separability. For each pixel, Pi identify the probability of its neighborhood, j by applying linear transformation.

TABLE II. FIVE CATEGORIES OF CLOUDS IN SWIMCAT DATABASE

Sl.No.	**Cloud Type**	**Name of Cloud Type**
1		Clear sky
2		Patterned clouds
3		Thick dark clouds
4		Thick white clouds
5		Veil clouds

And hence in this work, totally twelve feature sets are used and hence the dataset is organized into 12-dimensional representation. And the dataset includes the statistical texture features and shape and pattern features. Sky/cloud images are highly spectral and time variant. Hence, working with such data adds more complexity to the classification process. Dataset after dimensionality reduction using NCA is more appropriate for the process of feature extraction and classification.

C. Feature Extraction

Classification of types of clouds in sky/cloud images are done according to their shape, pattern, and texture features. Fast Fourier Transformation (FFT) is applied to extract statistical features that determine the statistical texture and spectral pattern features of the images using the equation (1).

$$\langle \boldsymbol{X_k} \rangle = \sum_{n=0}^{N-1} \boldsymbol{x_n} \boldsymbol{e}^{-j2k\pi n/N} \quad (1)$$

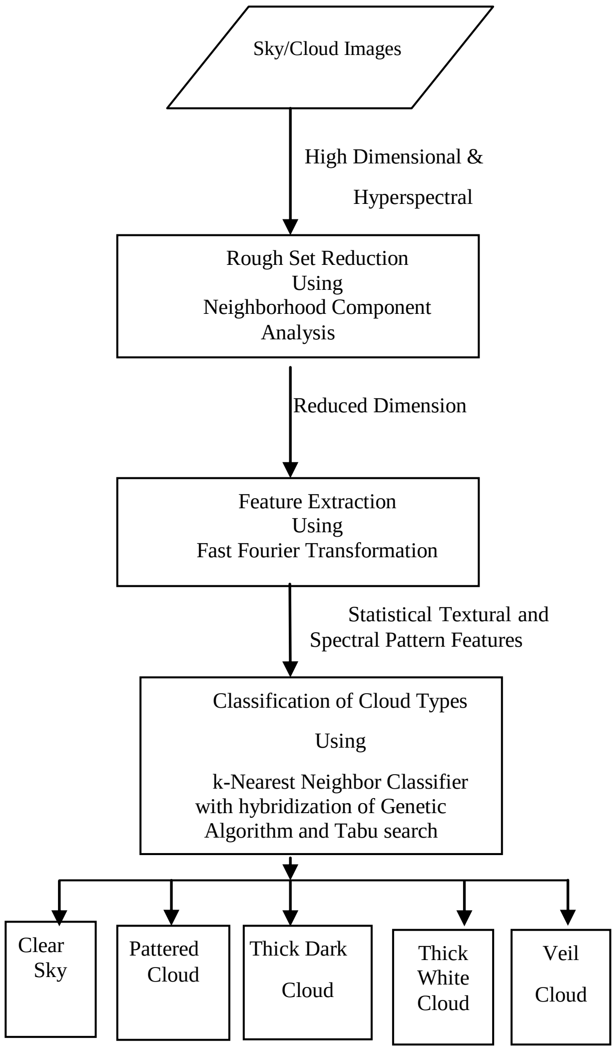


Fig.1. Overall block diagram of the work

The x array is transformed to d-dimensional vector using set of d nested summations.

(i) Statistical Texture Features

Statistical features used in our work includes mean, standard deviation, smoothness, third moment, uniformity and entropy are obtained by applying FFT to determine the uniformity, skewness and transparency of the cloud in sky/cloud images. Equations (1) to (10) shows the mathematical derivations of the features considered for classifications.

The statistical features used in our work are evaluated as follows:

a) Mean (M):

$$M = \sum_{i=0}^{L-1} Zip_{zi} \quad (2)$$

where z is the intensity value and Pz is the frequency distribution of z at various level, L

b) Standard Deviation (SD):

$$SD = \sqrt{\sum_{i=0}^{L-1} (Zi - M)^2 p_{zi}} \quad (3)$$

It measures the contrast in the image.

c) Smoothness (SM):

$$SM = 1 - (1 / (1 - \sigma^2)) \quad (4)$$

where the variance, $\sigma^2 = SD^2/(L-1)^2$.

d) Third moment (TM):

$$TM = \sum_{i=0}^{L-1} (Zi - M)^2 p_{zi} \quad (5)$$

It evaluates the skewness of the histogram of the input imagery.

e) Energy (EN):

$$B = \sum_{i=0}^{L-1} (Pzi)^2 \quad (6)$$

Energy is the gray level differences of the image.

f) Entropy (EY):

$$EY = -\sum_{i=0}^{L-1} (Pzi) \log(Pzi) \quad (7)$$

Entropy measures the product of gray level differences of the image and its logarithmic value.

(ii) Spectral Pattern Features

Fast Fourier Transformation is applied in order to extract the pattern features of sky images. Initially spectral power images are obtained, then subtract the average of values in the image from all pixels. Next, apply cosine shape at the borders to reduce the edge discontinuity. Then, apply FFT to obtain complex amplitude of harmonics. Finally, apply mod on this amplitude which in turn results in the spectral energy function. Normalize this energy function and obtain the spectral power function. Table III shows the original image, its corresponding RGB image, then its grayscale image, intensity image and the spectral power. And this spectral power function is used to estimate pattern features like (i) Correlation with Clear (CC), (ii) Spectral Intensity (SI) (iii) Fractional Sky Cover (FSC) (iv) Cloud Brokenness (CB) (v) Contrast (CON) and (vi) Homogeneity (HOM).

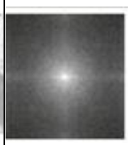
a) Correlation with Clear (CC)

Correlation with Clear (CC) is the linear correlation coefficient between the log of spectral power function of test image and reference image.

b) Spectral Intensity (SI)

Spectral Intensity (SI) measures the distribution of spectral power. That is the Spectral Intensity is the accumulated spectral power two wavenumbers.

TABLE III.ORIGINAL IMAGE, RGB IMAGE, GRAYSCALE, INTENSITY IMAGE AND SPECTRAL POWER

Input Image	RGB	Grayscale	Intensity	Spectral Power
				

c) Fractional Sky Cover (FSC)

FSC is a ratio between number of cloudy pixels to the total number of pixels in the image.

d) Cloud Brokenness (CB)

CB is the ratio of the number of pixels on the perimeter of cloudy areas to the number of cloudy pixels.

e) Contrast (CON)

Contrast measures the gray level differences.

$$CON = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (Zi - M)^2 \, p_{zi} \qquad (8)$$

f) Homogeneity (HOM)

The homogeneity represents the similarity of gray levels.

$$HOM = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{Pzi}{1+Zi-Zj} \qquad (9)$$

D. Hybridization of Genetic Algorithm and Tabu search for k-Nearest Neighbor (kNN) Classifier

The objective function of Genetic algorithm is defined as approximation of posterior probabilities of each output of kNN.The crossover operator determines the relationship between different feature vectors. This crossover minimizes the objective function.The mutation scheme selects the best data from the training sample that fits a particular class.Tabu search creates a dynamic table comprising of data that are with relevant details used for classification.

E. K- Nearest Neighbor Classifier

For the classification of types of clouds in sky/cloud images using the extracted spectral, textural and shape patterns, k-NN classifier is used in this paper. K-NN classifier simple, powerful and requires low computational cost. Even though this classifier is slow and require large memory space, the dataset reduced using NCA will minimize half of the work of k-NN classifier. Hence, in many classification works NCA and k-NN are coherently used.The original input images have the dimension of 125 X 125 pixels. And it has been reduced using NCA to 8-dimensional data containing 8 feature vectors. And in this reduced dataset, some images are considered for training and the other images are used as test images. The steps included in k-NN classifier is illustrated below.

1. Determine parameter k,which is the number of nearest neighbors.
2. Condense the training data.

   (i) Select each point and check whether they are observed point or prototypes to be included in tabu table.

   (ii) Ignore the outliers as they may degrade the performance of classification.
3. Calculate the distance between the testing instance and all training samples for each feature vector by using Euclidean distance measure.
4. Sort the distance and determine the nearest neighbors based on the k^{th} minimum distance of all the images.

   (i) Use Genetic algorithm with crossover and mutation operator to select the closest point.

   (ii) Selected data point is added onto a dynamic table called as Tabu table and it makes the search process easy.
5. Group the r nearest neighbors.
6. Assign the class label to the majority of the category of nearest neighbors.

The extracted feature sets from SWIMCAT dataset are shown in the Table V.

IV. Experimental Evaluation And Results

This work extracted spectral, textural and shape features of the sky/cloud images from the SWIMCAT database. The images of entire database is partitioned into 50% as training set, 25% for testing set and the remaining 25% for validation and is shown in Table IV. The images are used for training, testing and validation in the ratio of 50:25:25 percentage from the dataset.

TABLE IV. TRAINING AND TESTING IMAGES OF SWIMCAT DATABASE

Sl.No.	Type of Cloud	No. of		
		Training Images	Testing Images	Validation Images
1.	Clear Sky	150	75	75
2.	Pattern Cloud	49	25	25
3.	Thick dark cloud	176	88	88
4.	Thick white cloud	181	91	91
5.	Veil Cloud	50	25	25
Total Images		606	304	304

TABLE V. FEATURE VECTORS OF SWIMCAT DATABASE

k=3	x	Clear Sky	Pattern Cloud	Thick Dark Cloud	Thick white Cloud	Veil Cloud
M	2	96.1	86.3	92.6	91.5	81.7
	10	93.3	83.6	89.8	88.7	79
SD	2	96.0	92.0	92.5	91.4	87.4
	10	94.5	90.8	91	89.9	86.2
SM	2	95.5	87.0	92	90.9	82.4
	10	93.9	83.1	90.4	89.3	78.5
TM	2	95.5	85.8	92	90.9	81.2
	10	94.8	81.2	91.3	90.2	76.6
EN	2	96.1	88.3	92.6	91.5	83.7
	10	94.6	84.2	91.1	90	79.6
EY	2	95.9	86.7	92.4	91.3	82.1
	10	93.3	83.3	89.8	88.7	78.7
CC	2	1.0	1.0	0.9	0.92	0.88
	10	0.9	0.92	0.88	0.87	0.82
SI	2	0.006	0.013	0.002	0.003	0.001
	10	0.017	0.012	0.015	0.002	0.001
FSC	2	0.00	0.00	0.33	0.58	0.74
	10	0.00	0.17	0.24	0.52	0.71
CB	2	-	-	0.04	0.06	0.05
	10	-	-	0.02	0.01	0.07
CON	2	96.1	86.1	92.6	91.5	81.5
	10	92.3	87.7	88.8	87.7	83.1
HOM	2	96.2	80.4	92.7	91.6	75.8
	10	94.9	85.4	91.4	90.3	80.8

The number of training images are fixed after optimization of input dataset. Optimization determines the relevant feature set that are used for classification. The test images can be chosen in a random way.

The above feature values are adequate to perform the classification. However, to validate the correctness of classification, a contingency matrix is formed for the SWIMCAT dataset as shown in Table VI and for HyTA dataset it is tabulated in Table VII. The overall accuracy is 89.7%.

TABLE VI. CONFUSION MATRIX OF CLASSIFICATION WITH SWIMCAT DATASET

	Clear Sky	Pattern Cloud	Thick Dark Cloud	Thick white Cloud	Veil Cloud	Total	Precision
Clear Sky	127	0	5	6	2	150	90.7
Pattern Cloud	3	35	0	6	5	49	71.4
Thick Dark Cloud	2	2	165	3	4	176	93.7
Thick white cloud	0	3	3	170	5	181	93.9
Veil Cloud	1	5	4	2	38	50	75
Recall	95.4	77.7	93.2	90.9	70.3	**606**	

TABLE VII. CONFUSION MATRIX OF CLASSIFICATION WITH HYTA DATASET

	Clear Sky	Pattern Cloud	Thick dark cloud	Thick white Cloud	Veil Cloud	Total	Precision
Clear Sky	125	0	7	6	2	140	89.2
Pattern Cloud	4	36	0	5	4	49	73.4
Thick Dark Cloud	2	2	160	8	4	176	90.9
Thick white Cloud	0	3	3	165	510	181	91.1
Veil Cloud	1	5	4	0	40	50	80
Recall	94.6	78.2	91.9	89.6	66.6	**596**	88.2

The experimental evaluation of kNN classifier is compared against Extreme learning machine, support vector machine and Self-Organizing Map classifiers with different types of clouds and their accuracy of classification is shown in table VIII.

TABLE VIII. PERFORMANCE OF KNN AGAINST ELM, SVM AND SOM CLASSIFIERS

Technique /Method	Types of Clouds	Accuracy Attained
ELM [29]	Cirrus	**87.67%**
	Cumulus	**90.75%**
	Stratus	**74.50%**
	Clear sky	**93.63%**
SVM [30]	Clear sky	**100%**
	Little mass cloud	**96.67%**

	Large mass cloud	**98.19%**
	Thick cloud	**91.67%**
ANN – SOM [31]	cumulonimbus	**90.74%**
	cirrus	**88.79%**
	clouds in high latitude	**98.51%**
kNN with Genetic	Pattered	**89.7%**
	Thick dark	
	Thick white	
	Veil cloud	

V. CONCLUSION

A framework has been developed to classify the types of cloud in sky/cloud images using k Nearest Neighbor (kNN) classifier. Neighborhood Component Analysis (NCA) reduces the rough set containing all details of sky/cloud images and extract the necessity features produces optimized representation of the actual dataset. NCA in this framework has reduced the original rough set to 8-dimensional representation including spectral, shape and textual features. The reduced rough set data improves the speed of classification. To further improve the performance of kNN classifier, condensed data representation is followed, it creates a dynamic table called tabu table used for search. kNN classifier combined with hybrid combination of Genetic algorithm and tabu search assigns the class labels for Pattered cloud, Thick white cloud, Thick dark cloud and Veil clouds. The experimental evaluation on the benchmark dataset SWIMCAT shows good performance of classification on sky/cloud images. This framework automatically classifies the sky/cloud images with accuracy of 89.7% and its computational complexity is reduced as the rough set reduction is applied using Neighborhood component analysis. This automated tool works only with reduced dataset for classification of sky/cloud images. In future, the classification can be done using the hierarchical combination of soft computing techniques to improve the accuracy to further extent.

The pixels of input imagery near the sun may vary greatly due to change in the position of solar disk and lead to misinterpretation. And hence in future to rectify this misinterpretation geometrical features have to be included for the classification process. This framework can be extended for evaluation using the benchmark dataset SEVIRIand in near future it can be evaluated with real dataset and types of clouds can be extended. Eleven new cloud classifications such as the volutus, or roll cloud, as well as the asperitas cloud, the flumen, "cataractagenitus", "flammagenitus", "homogenitus" , "silvagenitus“ etc. discovered in March 2017 can also be identified in future using our framework.

REFERENCES

[1] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, “Clusteringbased hyperspectral band selection using information measures,” IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 12, pp.4158–4171, Dec. 2007.

[2] S. Dev, Y. H. Lee, and S. Winkler, “Color-based segmentation ofsky/cloud images from ground-based cameras,” IEEE Journal of Selected Topics in Applied Earth Observations and RemoteSensing,vol. PP, no. 99, pp. 1–12, 2016.

[3] S. Dev, Y. H. Lee, and S. Winkler, “Systematic study of color spaces and components for the segmentation of sky/cloud images,” in Proc. International Conference on Image Processing (ICIP), pp. 5102–5106, 2014.

[4] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pages, “Retrieving cloud characteristics from ground-based daytime color all-sky images,” Journal of Atmospheric and Oceanic Technology, vol. 23, no. 5, pp. 633–652, May 2006.

[5] A. Heinle, A. Macke, and A. Srivastav, “Automatic cloud classification of whole sky images,” Atmospheric Measurement Techniques, vol. 3, no. 3, pp. 557–567, May 2010.

[6] S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao, “Automatic cloud detection for all-sky images using superpixelsegmentation,” IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 2, pp. 354–358, Feb. 2015.

[7] M. P. Souza-Echer, E. B. Pereira, L. S. Bins, and M. A. R. Andrade,“A simple method for the assessment of the cloud coverstate in high-latitude regions by a ground-based digital camera,” Journal of Atmospheric and Oceanic Technology, vol. 3, no. 5, pp. 437–447, March 2006.

[8] S. L. Mantelli-Neto, A. von Wangenheim, E. B. Pereira, and E. Comunello, “The use of Euclidean Geometric Distance on RGB color space for the classification of sky and cloud patterns,”Journal of Atmospheric and Oceanic Technology, vol. 27, no. 9,pp. 1504–1517, Sept. 2010.

[9] Q. Li, W. Lu, and J. Yang, “A hybrid thresholding algorithm for cloud detection on ground-based color images,” Journal of Atmospheric and Oceanic Technology, vol. 28, no. 10, pp. 1286–1296, Oct. 2011.

[10] Y.Yuan and X. Hu, “Bag-of-words and object-based classification for cloud extraction from satellite imagery,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,vol. 8, no. 8, pp. 4197–4205, Aug. 2015.

[11] J. Calbó and J. Sabburg, “Feature extraction from whole-sky groundbased images for cloud-type recognition,” Journal of Atmospheric and Oceanic Technology, vol. 25, no. 1, pp. 3–14, Jan. 2008.

[12] S. Liu, Z. Zhang, B. Xiao, and X. Cao, “Ground-based cloud detection using automatic graph cut,” IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 6, pp. 1342–1346, June 2015.

[13] J. Shao and W. Foerstner, “Gabor wavelets for texture edge extraction,” in Proc. SPIE ISPRS Commission III Symposium: Spatial Information from Digital Photogrammetry and Computer Vision, vol. 2357, 1994.

[14] M. Galun, E. Sharon, R. Basri, and A. Brandt, “Texture segmentation by multiscale aggregation of filter responses and shape elements,” in Proc. International Conference on Computer Vision (ICCV), pp. 716–723, 2003.

[15] S. Dev, Y. H. Lee, and S. Winkler, “Categorization of cloud image patches using an improved texton-based approach,” in Proc. IEEEInternational Conference on Image Processing (ICIP), 2015.

[16] Q. Li, W. Lu, and J. Yang, “A hybrid thresholding algorithm for cloud detection on ground-based color images,” Journal of Atmospheric and Oceanic Technology, vol. 28, no. 10, pp. 1286–1296, 2011.

[17] S. L. M. Neto, A. von Wangenheim, E. B. Pereira, and E. Comunello, “The use of Euclidean geometric distance on RGB colorspace for the classification of sky and cloud patterns,” Journal of Atmospheric and Oceanic Technology, vol. 27, no. 9,pp. 1504–1517, 2010.

[18] S. Dev, Y. H. Lee, and S. Winkler, “Systematic study of color spaces and components for the segmentation of sky/cloud images,” in Proc. International Conference on Image Processing (ICIP), 2014.

[19] Q. Li, W. Lu, and J. Yang, “A hybrid thresholding algorithm for cloud detection on ground-based color images,” Journal of Atmospheric and Oceanic Technology, vol. 28, no. 10, pp. 1286–1296, 2011.

[20] Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J.,Shields, J., and Washom, B.: Intra-hour forecasting with a totalsky

imager at the UC San Diego solar energy testbed, Sol. Energy,85, 2881–2893, 2011.

[21] Johnson, R., Hering W., and Shields, J.: Automated visibility andcloud cover measurements with a solid-state imaging system.Tech. Rep., Scripps Institution of Oceanography, Marine PhysicalLaboratory, University of California, San Diego, USA, SIORef. 89-7, GL-TR-89-0061, 128, 1989.

[22] Long, C. N., Sabburg, J., Calbó, J., and Pagès, D.: Retrieving cloudcharacteristics from ground-based daytime color all-sky images,J. Atmos. Ocean. Tech., 23, 633–652, 2006.

[23] Shields, J., Karr, M., Burden, A., Johnson, R., and Hodgkiss, W.:Continuing support of cloud free line of sight determination includingwhole sky imaging of clouds, Final Report for ONRContract N00014-01-D-0043DO #13,Marine Physical Laboratory,Scripps Institution of Oceanography, University of California,San Diego, USA, Technical Note 273, 2007.

[24] Roy, G., Hayman, S., and Julian, W.: Sky analysis from CCD images:cloud cover, Lighting Res. Technol., 33, 211–222, 2001.

[25] West, S. R., Rowe, D., Sayeef, S., and Berry, A.: Short-term irradianceforecasting using skycams: Motivation and development,Sol. Energy, 110, 188–207, 2014.

[26] Huo, J. and Lu, D.: Cloud determination of all-sky images underlow visibility conditions, J. Atmos. Ocean. Tech., 26, 2172–2180,2009.

[27] J. E. Shields, T. L. Koehler, and R. W. Johnson, “Whole sky imager,” in Proceedings of the Cloud Impacts on DOD Operationsand Systems, 1989/90 Conference (Science and Technology Corporation, Meetings Division, 101 Research Drive,pp. 123–128,1990.

[28] A. V. Zukhba, “Np-completeness of the problem of prototype selection in the nearest neighbor method”, Pattern Recognition and Image Analysis, vol. 20, pp. 484-494, 2010.

[29] Zhao Zhen, Fei Wang, Yujing Sun, Chun Liu, Bo Wang, Jing Lu, “SVM Based Cloud Classification Model Using Total Sky Imagesfor PV Power Forecasting”, IEEE Power and Energy Society, 2015

[30] Zhengping Wu, Xian Xu, Min Xia, Meifang Ma, Lin Li, “Ground-based Vision Cloud Image Classification based on Extreme Learning Machine”, The Open Cybernetics and Systemics Journal, 2015

[31] Yu Liu, Jun Xia, Chun-Xiang Shi, Yang Hong, “An Improved Cloud Classification Algorithm for China’s FY-2C Multi-channel Images Using Artificial Neural Network”, Sensors, 2009.