International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Machine Learning for Cloud Cover Detection Using Multispectral Satellite Images

Dr. Preeti Verma

Assistant Professor, Computer Science & Engineering, Rewa Engineering College,(M.P.)

Abstract
Reliable cloud detection in satellite based remote sensing applications is a vital pre-processing step. It can be viewed as a classification technique by partitioning objective pixels into cloud or non-cloud shadow classes. However, certain cloud pixels, particularly narrow pixels in the cloud, can be regarded as a blend of cloud reflection and land-based objects. Hence, it is difficult to classify them using a single classifier. In the present study, a technique based on an ensemble classifier is proposed to identify these pixels. The proposed ensemble classifier uses the Extreme Learning Machine (ELM), Naïve Byes (NB), and K-Nearest Neighbour (KNN) classifier. The model is trained and evaluated using the Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) datasets. The findings of the analysis indicate that the proposed classifier provides better classification accuracy. The performance of the model other than RGB band is also analyze and results shows that adding images from different bands significantly improves the results.

Keywords: Machine learning, cloud cover detection, remote sensing, satellite images.

1. Introduction

For numerous scientific investigations, such as the studies of water supply, climate change, the earth energy budget, etc., the development of cloud masks connected with remote sensing photographs is crucial. Remote sensing plays a key role in determining the physical parameters associated with clouds, with a considerable impact on planet earth's radiation budget [1]. Cloud detection involves labelling each pixel of a scene with a binary variable to determine whether or not this pixel is a cloud. The etiquette map is known as the cloud mask [2]. The open data policy of NASA's landsat satellites in 2008 drew enormous attention to remote sensing. The Copernicus programme of the European Space Agency (ESA) began providing free satellite data, including optical imaging and radar measurements of the earth's surface, as well as chemical composition studies of the troposphere, in 2011. Innovators and firms will benefit from combining these data with modern data analytics techniques [3]. The fundamental idea of this cloud detection approach is to employ an Extreme Learning Machine (ELM) to compute a label probability for each pixel. In supervised mode, the ELM uses fully-labeled images to build a cloud index for each raw image pixel.

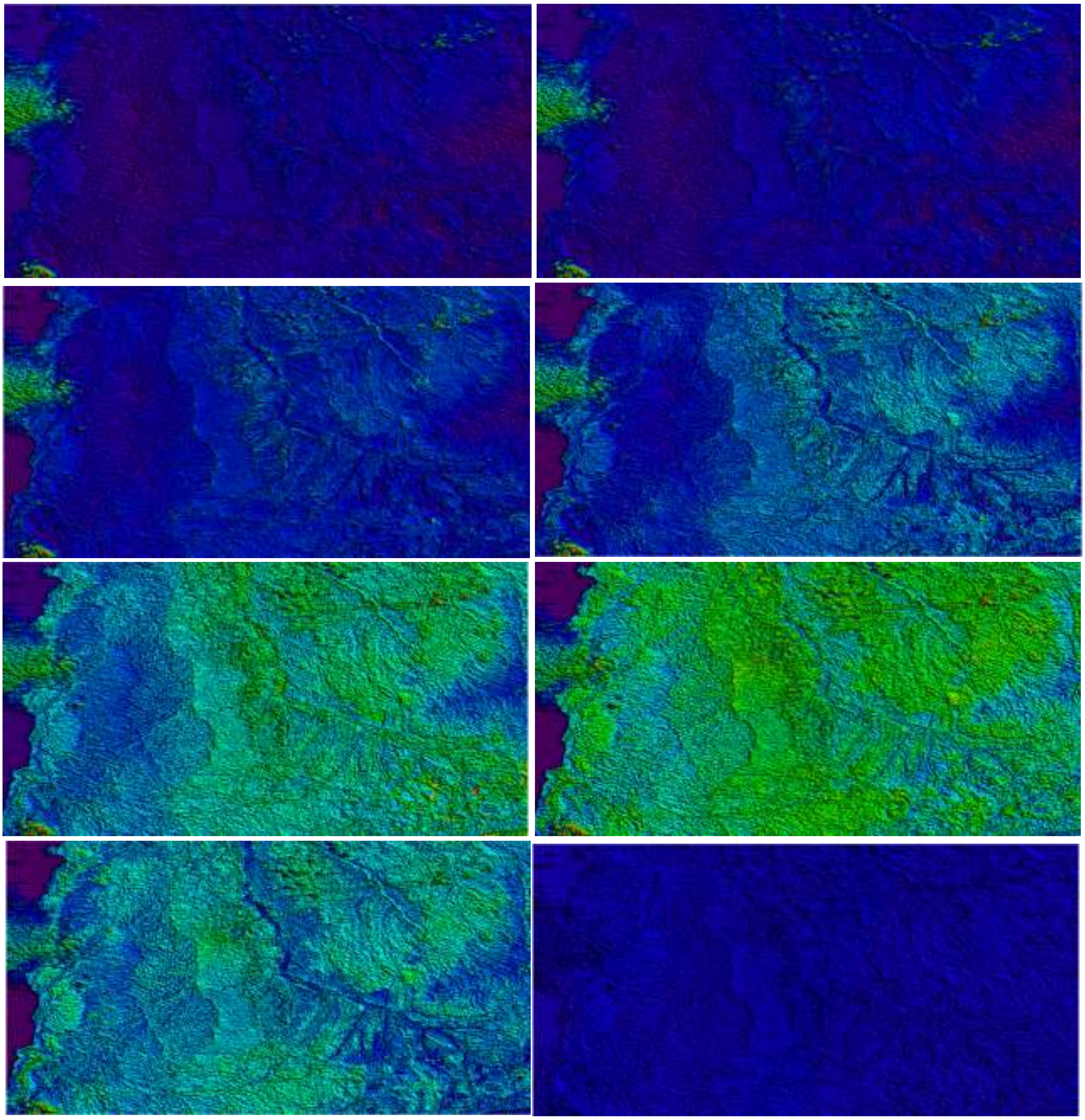
ELM is a fast convergent training approach for single hidden layer feed forward neural network (SLFN) [4]. The ELN learning speed is thousands of times quicker than classic feed forward network learning techniques like back-propagation (BP) [5].

A data science platform that boosts productivity. Create and test better machine learning models. Increase business flexibility by swiftly deploying enterprise-trusted data and supporting data-driven business objectives with ML models. Data augmentation refers to the idea of modifying the training data in order to provide more useful training instances. Technology has resulted in massive increases in raw data volumes. There is a need for more effective handling of sensitive data as well as secure storage and rapid access owing to the ever-evolving area of data science and the tools it employs (algorithms, processes, and soft computing).There are many fields in which the use and development of data science techniques may be found. The goal of data science is to help people make better decisions by organising and collecting large volumes of data so that it can be reviewed, analysed, and chosen from. More and more researchers are paying attention to and using ANNs, SVMs and EVMs because of their increasing accuracy in a variety of areas when they are optimised using a big ground reference dataset. This study paper employs the second strategy.

2. Literature Survey

Researchers have employed a variety of soft computing approaches and algorithms, as well as data mining and artificial intelligence applications, to try and solve a variety of challenges [6-8]. Many automatic cloud detection systems have been developed and routinely deployed to satellite photos in the previous two decades. Based on the features employed, existing cloud identification algorithms can be divided into two classes: spectral and texture-based [9]. This method is successful but fails to distinguish between materials like ice and clouds that have similar spectral characteristics. Sophisticated detector noise or ambient factors can affect spectral characteristics. Texture-based features are less affected by these impacts [10-12]. To distinguish clouds from clear-sky pixels, spectral analysis approaches have been frequently used. For cloud cover evaluation of Landsat-7 imagery, the automatic cloud cover assessment (ACCA) algorithm has been officially used [13]. To distinguish clouds and snow from clear observations, the ACCA method uses two unique tests and over 30 parameters. The authors proposed a polarised picture characteristic cloud detection technique. They made use of features like multi-spectrum and polarization [14]. The dynamic threshold is calculated using various atmospheric and surface models at various times and locations [15]. The paper focuses on several colour models that use multi-spectral thresholds and pixel-to-pixel processing. In terms of the threshold value, the model recognizes cloud and clear cloud, but it has problems distinguishing between snow/Ice and cloud because the reflectance of snow/Ice and cloud is the same in the visible region [16]. The authors suggested a cloud identification approach for Landsat 8 multi-spectral remote-sensing pictures. They employed the Simple Linear Iterative Cluster (SLIC) approach, followed by a two-step super pixel classification strategy to determine if each pixel is cloudy or not. To extract high-level information from the cloud, they employed a double branch Principal Component Analysis (PCA) Network (PCANet) and a support vector machine classifier [17]. Texture approaches are generally based on the spatial distribution of numeric counts, which can be calculated using statistical measures like the grey level cooccurrence matrix or recognised filters like Gabor or discrete cosines. Inventors employs an automatic cloud identification system based on spatial texture analysis and a neural network. On the spectrum analysis for each pixel cloud detection, it used the multi-spectral synthesis approach and the cloud detection index. Super pixel segmentation, spectral and texture feature extraction as well as super pixel information extraction from feature histogram activities were all performed by the authors in the correct sequence [18]. For cloud mask extraction, an optimum threshold and support vector machine (SVM) are used. By extracting grey feature vector and frequency feature vector from remote-sensing photos, the authors created a texture feature vector [19]. They rated the cloud region by comparing actual feature values and thresholds obtained through the training procedure. The approach successfully distinguishes between cloud and ground objects. Researchers suggested a cloud identification approach for high-resolution satellite photos based on multi-features of ground objects such as colour, texture, and shape. The textures are then extracted using the multi-scale decomposition of the domain transform. Although this approach has a high overall accuracy rate, it is readily mistaken if the non-cloud area is as reflective as the cloud area [20].

Following the preceding research, narrow cloud pixels can be viewed as a mix of cloud reflection and land-based objects. So classifying them with a single classifier is tough. To identify these pixels, an ensemble classifier is developed. The proposed ensemble classifier uses ELM, NB, and KNN classifiers. The SPARCS datasets are used to train and test the model. The research shows that the proposed classifier is more accurate. The results suggest that combining photos from other bands enhances the results greatly.



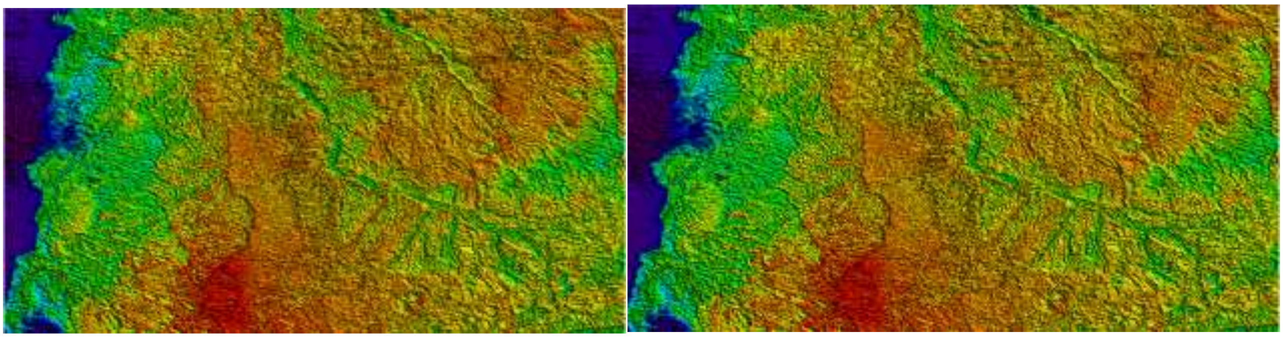


Figure 1: Landsat 8 images for bands from 1 to 10.

3. Dataset Description

NASA Landsat 8 is an 11-band multi-spectral satellite with a 15m resolution and a 16-day revisit duration (Figure 1). It consists of two instruments: the OLI and the TIR (TIRS). A broad panchromatic band gives 15m resolution compared to the other bands (https://landsat.usgs.gov/what-are-banddesignations-landsat-satellites). Band 9 is designed to identify cirrus clouds, which commonly compromise satellite images. The last two bands of TIRS provide surface temperatures from 10.60 m to 12.51 m. The high geographic resolution and multi-spectral capabilities of satellites have proven invaluable in monitoring our environment, and NASA's open data policy has boosted the use of satellite data in academics and business.

Table 1: Landsat-8 bands description

Band Type	Wavelength	Resolution
Band 1 - Coastal aerosol	0.43-0.45µm	30m
Band 2 - Blue	0.45-0.51µm	30m
Band 3 - Green	0.53-0.59µm	30m
Band 4 - Red	0.64-0.67µm	30m
Band 5 - Near Infrared (NIR)	0.85-0.88µm	30m
Band 6 - SWIR 1	1.57-1.65µm	30m
Band 7 - SWIR 2	2.11-2.29µm	30m
Band 8 - Panchromatic	0.50-0.68µm	15m
Band 9 - Cirrus	1.36-1.38µm	30m
Band 10 - Thermal Infrared (TIRS) 1	10.6-11.19µm	100m

Landsat 8 Biome dataset is composed of 96 Landsat 8 sceneries collected and annotated by the authors. It was meant to be readily divided in two for machine learning model validation and training. It has eight biomes: barren, woodland, grass/crops, shrub land, urban, water, and wetlands, and four classes: cloud, thin cloud, cloud shadow, and clear. However, only 30 of the 96 scenes include shadow annotations. The settings are also separated into groups of photos with little or no clouds (35%), mid-overcast (35%) to cloudy (> 65%). There are 80 10001000 pixel scenes in the 1.6 GB Landsat 8 SPARCS dataset. "Cloud shadow over water", "water", "snow", "land", "cloud", and "flooded". An extra analyst annotated 6 sub-scenes of the SPARCS dataset to explore the agreement amongst analysts on annotating these datasets [21, 22].

The potential information contained in different wavelengths group of imaging spectrum can be concise as follows [23, 24]:

Table 2: Landsat-8 bands description with wavelength

Type of band	Wavelength	Information
Visible	0.6-0.8 µm	These are the visible channels, which are essential for cloud detection, cloud tracking, scene identification, aerosol monitoring, and land surface and vegetation monitoring.
Near Infrared	1.6 µm	This bands can distinguish between snow and cloud, ice and water clouds, and aerosols, among other types of particles in the atmosphere.
Infrared	3.9 µm	In low clouds and fog. Cloud monitoring improves low-level wind coverage and helps calculate land and sea surface temperatures at night. The MSG spectral band has been expanded to longer wavelengths to boost SNR.
Infrared	6.2-7.3 µm	This channel observes water vapour and wind. Supports designating semitransparent cloud height as well.
Infrared	8.7 µm	Provides detailed information on thin cirrus clouds and helps distinguish ice from water.
Infrared	9.7 µm	Ozone radiances may be used in numerical weather prediction. The development of the entire ozone field may be tracked.
Infrared	10.8-12 µm	Well-known split-window channels. Important for determining cloud-top and sea-surface temperatures.
Infrared	13.4 µm	The CO_2 absorption receptor If there are no clouds, it can give temperature data from the lower troposphere to determine static instability.

4. Mathematical background

4.1 Extreme Learning Machine (ELM)

The ELM, a single hidden layer feed forward-neural network (SLFN), which can be trained quickly and accurately for vector machines are a growing technique in machine learning. ELM was initially developed in 2006 and has three layers of conventional feed forward architecture. Its first layer is an input layer and the second layer (or the hidden) is used to project the input layer to greater dimensions using a large number of nonlinear sigmoid neurons, and the last layer is used as the output and is made up of linear output neurons [25-27]. The ELM structure is shown in Figure 2.

$$y(p) = \sum_{j=1}^{m} \beta_j g\left(\sum_{i=1}^{n} w_{i,j} x_i + b_j\right) \tag{1}$$

Where, β_i is the weights between the input layer and the hidden layer, and β_j is the weights between the output layer and the hidden layer. b_j is the hidden layer neuron threshold value, $g(.)$ activation function. Weights of the same input layer ($w_{i,j}$) and bias (b_j) are assigned randomly. At the beginning of the input layer neuron number (n) and hidden layer neuron number (m), the activation function ($g(.)$) is allocated. Now, based on this information, by combining and rearranging the parameters known in equilibrium, the output layer becomes as in equation 3 [28].

$$H(w_{i,j}, b_j, x_i) = \begin{bmatrix} g(w_{1,1}x_1 + b_1) & \cdots & g(w_{1,m}x_m + b_m) \\ \vdots & \ddots & \vdots \\ g(w_{n,1}x_n + b_1) & \cdots & g(w_{n,m}x_m + b_m) \end{bmatrix} \quad (2)$$

$$y = H\beta \quad (3)$$

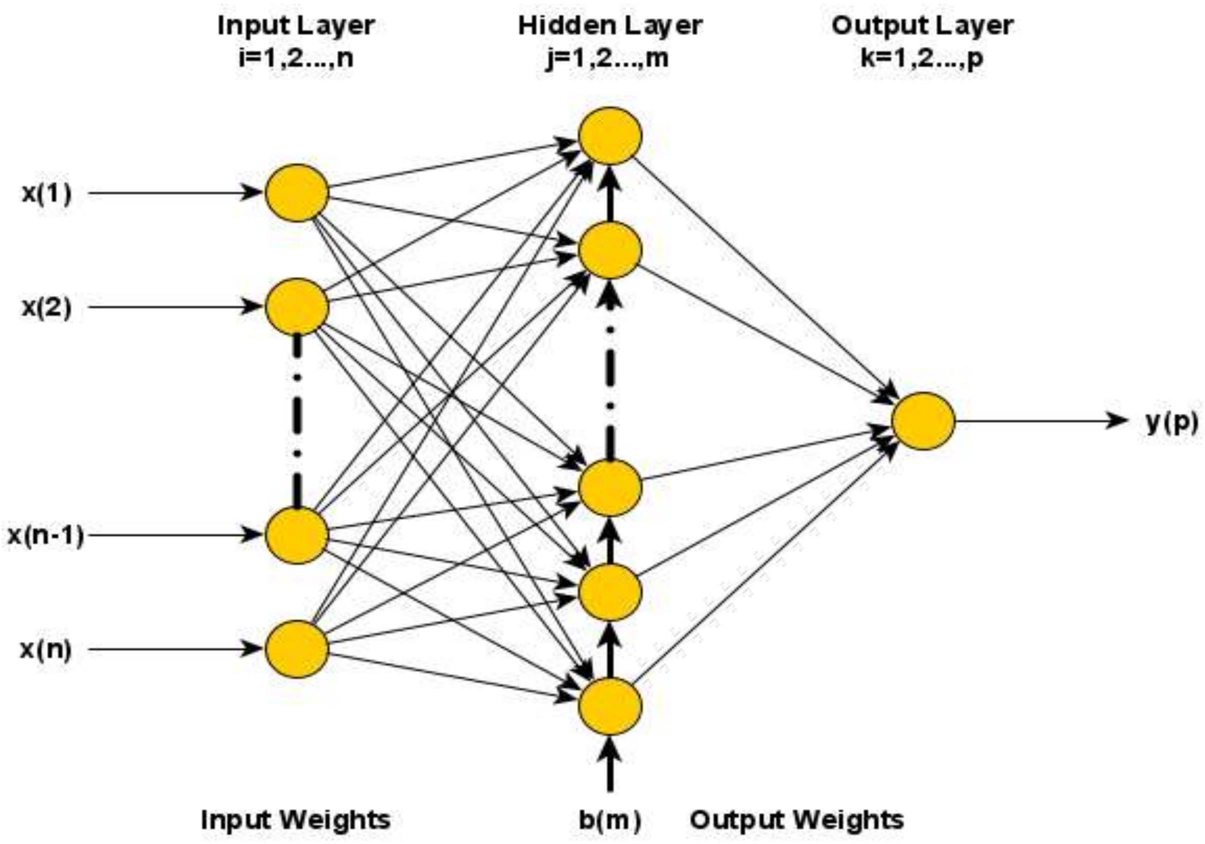
The goal is to minimize the error as much as possible in all training algorithm models. The output y_p error function obtained from the real output value y_o in ELM is $\sum_k^s(y_o - y_p)$ (with "s": training data number) and $\left\|\sum_k^s(y_o - y_p)^2\right\|$ can be minimized.

The output y_p obtained from the actual output value y_o must be equal to y_p for both of these functions. When this happens the unknown parameter in equation (3) the H matrix can be a very low probability matrix, which means that it is impossible that the number of data in the training set will be equal to the number of features each data includes. Therefore, it will be a challenge to take the inverse of H and to find weights (b). To overcome this situation, the pseudo-inverse of the matrix H can be taken by using Moore-Penrose Inverse. The output weights can therefore be found through $\beta = y \times MPI(H)$.

4.2 Feature selection

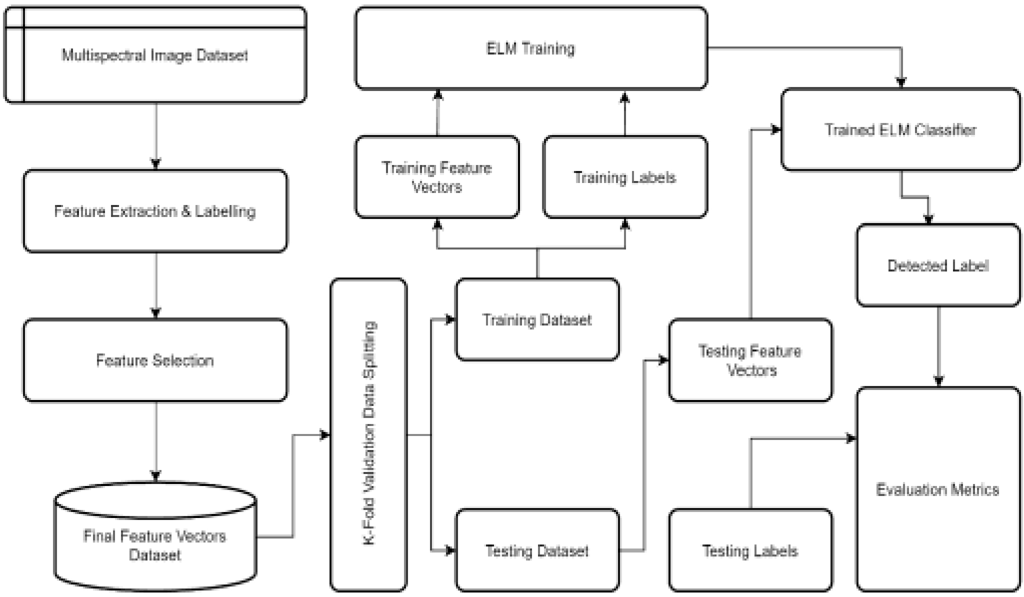
The large dimensionality of the feature vector owing to its non-informational properties is not deemed required as it increases computational complexity and memory usage. It also results in low categorization accuracy. To reduce dimensionality, the most relevant subset of the original feature collection should be

Figure 2: ELM structure with single hidden-layer feed-forward neural network.



International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com



selected. In this paper both the Fisher and minimum Redundancy Maximum Relevance (mRMR) selection methods to pick the most appropriate features. In order to choose the top 10 characteristics, we are using three variations of the mRMR algorithm: Difference in Mutual Information (MID), Mutual Information Quotient

Figure 3: Architecture of the proposed cloud cover detection technique.

(MIQ), and Mutual Information Basis (MIBASE). These properties are provided in Table 3 based on the function selection algorithm [29-32]. Figure 3 shows the suggested architecture for detecting cloud cover.

	Fisher	**MID**	**MIQ**	**MIBASE**
Selected features highest score (top) to lowest score(bottom)	1	4	3	3
	2	3	2	2
	4	1	8	1
	3	7	1	4
	6	2	5	5

Table 3: Top 10 selected features by different algorithms.

5. Proposed Methodology

The proposed cloud cover anomaly detection model is shown in Figure 3. The task in classification is to categorize the each pixel in the multispectral image into pre-defined 7 classes, namely:

1. Cloud Shadow
2. Cloud Shadow over Water
3. Water
4. Snow
5. Land
6. Cloud, and
7. Flooded

Steps of the feature extraction and training vector formation is presented in Figure 4. In this process firstly the images of same geographical area scanned using different frequency bands are collected from Landsat

8 SPARCS dataset [22]. The SPARCS dataset consists of 80 scenes of 1000 × 1000 pixels each. The dataset also contains the 1000 × 1000 pixels annotation image for each scene. These annotation image classify the each pixel in one of the seven predefined classes as mentioned above.

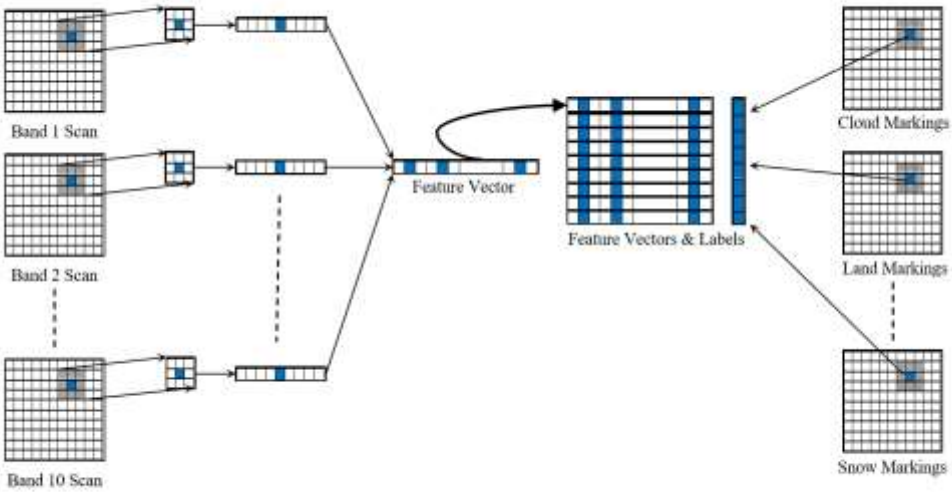


Figure 4. Process of forming feature vectors and their respective labels for supervised learning.

To generate the feature vectors for training the ELM, a square mask of $m \times m$ pixels is applied over the each pixel of image (Figure 4). The process is repeated for all 10 bands images. Finally these extracted $m \times m$ pixels from all 10 band images are converted into $1 \times (10 \times m^2)$ element row vectors. These vectors are then grouped with the label given in annotation image.
Since it is known (refer Table-1, Table-2) that each frequency band reveal different information about the scanned environment. Taking features from frequency band that may not contain the useful information will cause the high dimensionality of feature vector. The unnecessary high dimensionality may increase the computational complexity and use of memory [29, 33]. It also leads sub-optimal classifier training.
To decrease dimensionality, it is appropriate to choose the most relevant subset of the original set of features. To select the most relevant features, in this work four different feature selection algorithms named as Fisher, and three variants of the mRMR algorithm names as Mutual Information Difference (MID), Mutual Information Quotient (MIQ), and Mutual Information Base (MIBASE), are used to select the top five features. These selected features are presented in Table 3.
Then this dataset is divided into two sets training set and testing set. In the training stage, the training set which consists set of predefined features and respective labels is used to train the extreme learning machine (ELM). The trained ELM is used as a cloud cover detector on the test dataset. In the testing stage, the testing dataset is applied to the trained ELM model to generate classification labels. The last step in the process is to evaluate the performance of the proposed model, this is done by comparing this generated set of labels from the testing step with the original testing set labels.

6. Performance Evaluation Metrics

The classifier is evaluated based on four common measures known as accuracy (Eq. 5), precision (Eq. 6), recall (Eq. 7), and F1 (Eq. 4) to estimate the efficiency of the methods. Accuracy determines the predictive ability of the classifier for normal and anomaly assessments. Accuracy defines the predicted accuracy of the label. Recall determines the completeness of the category. Measurement of F1 is the harmonic mean of precision and recall used to balance accuracy with recall.

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

$$Accuracy = \frac{|TP + TN|}{|TP + TN + FP + FN\ |} \tag{5}$$

$$Precision = \frac{|TP|}{|TP + FP\ |} \tag{6}$$

$$Recall = \frac{|TP|}{|TP + FN|} \tag{7}$$

Where, TP, TN, FP, and FN are representing the true positive, true negative, false positive, and false negative respectively and defined as:

- TP: The number of anomalous instances classifier correctly identified as an anomaly.
- TN: The number of normal instances classifier correctly identified as normal.
- FP: The number of normal instances classifier incorrectly identified as an anomaly.

Table 4: Results comparison

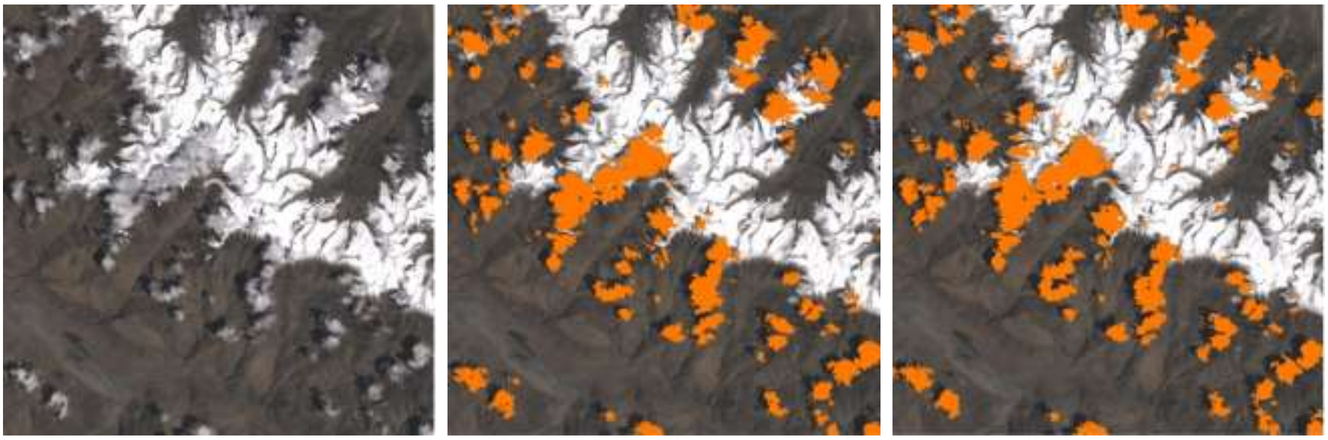
Dataset		Feature Set (Selected Bands)	Accuracy (%)				Precision (%)			
			SVM	HMM	RS-Net [3]	Proposed	SVM	HMM	RS-Net [3]	Proposed
SPARCS		Fisher	84.3	76.3	N/A	**96.7**	74.6	69.3	N/A	**96.3**
SPARCS		MID	82.1	71.3	N/A	**94.3**	56.3	50.5	N/A	**93.6**
SPARCS		MIQ	83.7	79.2	N/A	**93.8**	55.1	48.2	N/A	**95.5**
SPARCS		MIBASE	87.8	79.9	N/A	**92.2**	68.9	67.7	N/A	**93.4**
SPARCS		*ALL	85.5	80.7	94.5	92.5	74.4	75.1	87.6	**89.1**
SPARCS		*ALL-NT	85.7	81.3	95.6	94.2	73.8	74.8	89.5	88.7
SPARCS		*RGBI	88.2	88.2	94.8	**95.6**	71.3	73.3	88.9	83.3
SPARCS		*RGB	89.1	81.1	94.9	**96.1**	75.6	72.8	88.6	**88.7**
SPARCS		*G	90.2	81.4	94.7	**95.6**	75.4	71.5	88.5	**89.1**
Dataset		**Feature Set (Selected Bands)**	**Recall (%)**				**F-1 (%)**			
			SVM	**HMM**	**RS-Net [3]**	**Proposed**	**SVM**	**HMM**	**RS-Net [3]**	**Proposed**
SPARCS		Fisher	70.2	69.3	N/A	**96.7**	72.3	69.3	N/A	**96.5**
SPARCS		MID	66.4	70.5	N/A	**94.3**	60.9	58.8	N/A	**93.9**
SPARCS		MIQ	68.1	69.2	N/A	**93.8**	60.9	56.8	N/A	**94.6**
SPARCS		MIBASE	56.4	70.9	N/A	**92.2**	62.0	69.3	N/A	**92.8**
SPARCS		*ALL	76.1	73.2	83.7	82.1	75.2	74.1	85.6	85.5

SPARCS		*ALL-NT	75.9	75.6	87.8	85.9	74.8	75.2	88.5	87.3
SPARCS		*RGBI	73.8	70.3	83.9	**84.0**	72.5	71.8	86.3	83.6
SPARCS		*RGB	72.8	76.1	84.3	**84.8**	74.2	74.4	86.4	**86.7**
SPARCS		*G	78.7	74.7	83.4	**84.3**	77.0	73.1	85.9	**86.6**

**five different band combinations, ALL = all available bands, ALL-NT = all available bands except thermal (i.e., exclude band 10 and 11), RGBI = the red/green/blue/infrared (RGBI) bands, RGB = the red/green/blue (RGB) bands, and B = the green (G) band alone. Band 8 (the panchromatic band) is not used in this work, and the remaining bands have not been pansharpened.*

FN: The number of anomalous instances classifier incorrectly identified as normal.

These can also be defined by the confusion matrix (Table 5):



(a) (b) (c)

Figure 5: Example of a SPARCS scene (LC81480352013195LGN00_32), (a) RGB scene, (b) Round truth cloud mask, (c) Proposed cloud mask.

Table 5: Confusion matrix

		Known Labels	
		True (Anomaly)	False (Normal)
Classifier's identification result	Positive (Anomaly)	TP	FP
	Negative (Normal)	FN	TN

7. Results analysis

Results of previous methods {SVM (Support Vector Machine), HMM (Hidden Markov Model) and, RS-Net} performances as well as the proposed method are shown in Table-4. For the SPARC dataset when compared to previous methods, the proposed approach delivers better accuracy for Fisher and RGB based features. For MID based features it provides 94.3% accuracy which is 12.2% higher than the 82.1 the previous best provided by HMM. For MIQ based features we get 93.8% which is 10.1% higher than the previous best 83.7% provided by SVM. When comparing In terms of F1 score the proposed classifier works better for all features set. Since the F1 score presents the combined information of Precision and Recall the higher value of it even when the Accuracy measure is lesser shows the better classifier performance.

For the RS-Net [3] and its feature set, the proposed approach delivers better accuracy for RGBI, RGB, and G Band based features and improves the accuracy by 1%. When comparing In terms of F1 score the proposed classifier works better for the RGB and G Band features. Example of a SPARCS scene is shown in Figure 5.

The proposed algorithm gives an average 94.55% accuracy for all feature sets with a standard deviation of 1.57%, whereas RS-Net provides 94.9% accuracy however with an 0.41% standard deviation. When analyzed for the F1 scores the proposed method gives a higher average F1 score of 89.72% with 4.69% of standard deviation in comparison to the RS-Net which provides 86.54% average accuracy with 1.14% standard deviation. This validates that the proposed method gives much uniform performance for all the five features set in terms of accuracy and F1 score.

8. Conclusion

In this paper, an ELM based approach for cloud cover detection using multispectral images is presented. Feature selection algorithms are used during processing phase to select most influential features, while ELM during the classification phase. Finally, the performance of the proposed and previous algorithms is evaluated using SPARCS dataset with four different feature selection algorithms.

The experimental results reveal that the proposed algorithm performs better than SVM, HMM, and RS-Net algorithm and provides much stable performance throughout the datasets and feature selection algorithms. The experimental results show that the proposed algorithm provides significantly improved performance over previous algorithms. The presented model can be taken as the reference for similar works as it contains all the necessary steps required.

Extending RS-Net to predict several classes, including interpretability approaches for classification judgments, and including temporal patterns into the network will be a major focus of future research. Remote sensing places a premium on taking into account the temporal patterns since buildings remain immobile over time, vegetation goes through regular cycles, and clouds move at a much faster and more erratic pace.

References

1. Wang, Z., Du, J., Xia, J., Chen, C., Zeng, Q., Tian, L., ... & Mao, Z. (2020). An Effective Method for Detecting Clouds in GaoFen-4 Images of Coastal Zones. *Remote Sensing*, *12*(18), 3003. https://doi.org/10.3390/rs12183003
2. Latry, C., Panem, C., & Dejean, P. (2007). Cloud detection with SVM technique. In *2007 IEEE International Geoscience and Remote Sensing Symposium* (pp. 448-451). IEEE. https://doi.org/10.1109/IGARSS.2007.4422827
3. Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote sensing of environment*, *229*, 247-259. https://doi.org/10.1016/j.rse.2019.03.039
4. Wang, J., Lu, S., Wang, S. H., & Zhang, Y. D. (2021). A review on extreme learning machine. *Multimedia Tools and Applications*, 1-50. https://doi.org/10.1007/s11042-021-11007-7
5. Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1-3), 489-501. https://doi.org/10.1016/j.neucom.2005.12.126
6. Olson, D. L., Shi, Y., & Shi, Y. (2007). Introduction to business data mining (Vol. 10, pp. 2250-2254). New York: McGraw-Hill/Irwin.

7. Shi, Y., Tian, Y., Kou, G., Peng, Y., & Li, J. (2011). Optimization based data mining: theory and applications. Springer Science & Business Media.
8. Tien, J. M. (2017). Internet of things, real-time decision making, and artificial intelligence. *Annals of Data Science*, 4(2), 149-178. https://doi.org/10.1007/s40745-017-0112-5
9. AJ, G. C., & Jojy, C. (2015). A Survey of Cloud Detection Techniques For Satellite Images.
10. Jedlovec, G. (2010). Automated detection of clouds in satellite imagery.
11. Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Haar, T. H. V., & Reinke, D. L. (1999). A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, *10*(1), 138-151. https://doi.org/10.1109/72.737500
12. Onsi, M., & ElSaban, H. (2012). Spatial cloud detection and retrieval system for satellite images. *International Journal of Advanced Computer Science and Applications,(IJACSA)*, *3*(12). https://doi.org/10.14569/IJACSA.2012.031235
13. Irish, R. R., Barker, J. L., Goward, S. N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric engineering & remote sensing*, *72*(10), 1179-1188. https://doi.org/10.14358/PERS.72.10.1179
14. Song, X., Liu, Z., & Zhao, Y. (2004, September). Cloud detection and analysis of MODIS image. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 4, pp. 2764-2767). IEEE. https://doi.org/10.1109/IGARSS.2004.1369875
15. Li, C., Ma, J., Yang, P., & Li, Z. (2019). Detection of cloud cover using dynamic thresholds and radiative transfer models from the polarization satellite image. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *222*, 196-214. https://doi.org/10.1016/j.jqsrt.2018.10.026
16. Gupta, R., & Panchal, P. (2015, January). Advancement of cloud detection algorithm in satellite images with application to color models. In *2015 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCCI.2015.7218124
17. Zi, Y., Xie, F., & Jiang, Z. (2018). A cloud detection method for Landsat 8 images based on PCANet. *Remote Sensing*, *10*(6), 877. https://doi.org/10.3390/rs10060877
18. Tan, K., Zhang, Y., & Tong, X. (2016). Cloud extraction from Chinese high resolution satellite imagery by probabilistic latent semantic analysis and object-based machine learning. *Remote Sensing*, *8*(11), 963. https://doi.org/10.3390/rs8110963
19. Changhui, Y., Yuan, Y., Minjing, M., & Menglu, Z. (2013). Cloud detection method based on feature extraction in remote sensing images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*, W1.
20. Zhang, J., Zhou, Q., Shen, X., & Li, Y. (2019). Cloud detection in high-resolution remote sensing images using multi-features of ground objects. *Journal of Geovisualization and Spatial Analysis*, *3*(2), 1-9. https://doi.org/10.1007/s41651-019-0037-y
21. Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley Jr, R. D., Beckmann, T., ... & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote sensing of environment*, *194*, 379-390. https://doi.org/10.1016/j.rse.2017.03.026
22. Hughes, M. J., & Hayes, D. J. (2014). Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sensing*, *6*(6), 4907-4926. https://doi.org/10.3390/rs6064907

23. Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002). An introduction to Meteosat second generation (MSG). *Bulletin of the American Meteorological Society*, *83*(7), 977-992. https://doi.org/10.1175/15200477(2002)083<0977:AITMSG>2.3.CO;2
24. Linares-Rodriguez, A., Ruiz-Arias, J. A., Pozo-Vazquez, D., & Tovar-Pescador, J. (2013). An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. *Energy*, *61*, 636-645. https://doi.org/10.1016/j.energy.2013.09.008
25. Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. *International journal of machine learning and cybernetics*, *2*(2), 107-122. https://doi.org/10.1007/s13042-011-0019-y
26. Tapson, J., & van Schaik, A. (2013). Learning the pseudoinverse solution to network weights. *Neural Networks*, *45*, 94-100. https://doi.org/10.1016/j.neunet.2013.02.008
27. Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1-3), 489-501. https://doi.org/10.1016/j.neucom.2005.12.126
28. Toprak, A. (2018). Extreme learning machine (elm)-based classification of benign and malignant cells in breast cancer. *Medical science monitor: international medical journal of experimental and clinical research*, *24*, 6537. https://doi.org/10.12659/MSM.910520
29. Winter, P., Hermann, E., & Zeilinger, M. (2011, February). Inductive intrusion detection in flow-based network data using one-class support vector machines. In *2011 4th IFIP international conference on new technologies, mobility and security* (pp. 1-5). IEEE. . https://doi.org/10.1109/NTMS.2011.5720582
30. Chen, Y. W., & Lin, C. J. (2006). Combining SVMs with various feature selection strategies. In *Feature extraction* (pp. 315-324). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-35488-8_13
31. Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.
32. Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, *27*(8), 1226-1238. https://doi.org/10.1109/TPAMI.2005.159
33. Verma, P., Patil, S. (2021). A Machine Learning Approach and Methodology for Solar Radiation Assessment Using Multispectral Satellite Images. *Ann. Data. Sci.*. https://doi.org/10.1007/s40745-021-00352-x