International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Image Retrieval using Feature Extraction Techniques in Machine Learning

D.Shunmuga Kumari[1]
Assistant Professor, Department of Computer Science and Information Technology,
Vels Institute of science technology and advanced studies Chennai[1]

D. Shalinigayathri[2]
Assistant Professor, Department of Computer Science,
Agurchand Manmull Jain College,Meenambakkam, Chennai, Tamilnadu, India[2]

Abstract — Nowadays remote sensors satellite images are used in vast areas and play an essential role in graphical reference for urban, agricultural areas, disaster mitigation, and tracking systems. Machine learning algorithms give an intelligent approach for Image retrieval depending on pattern recognition and classification. An input query image is retrieved from an image search engine by image pattern recognition and classification through an evaluation of different Machine learning algorithms. The classification algorithms of SVM Non-Linear and Random Forest are evaluated here for image retrieval. Random Forest averages the results of several decision trees performed to distinct subsets of a dataset to boost the dataset's forecast accuracy. Support vector machine produces great accuracy while using less computing power for both linear and Non-linear. Taking SVM Nonlinear and Random Forest as examples, this paper compares and analyzes the image classification algorithms for retrieval. This study proves that the accuracy of SVM Non-Linear is 77%, and the accuracy of Random Forest is 66% using the UC Merced dataset.

Keywords— Image Retrieval, Feature Extraction, Machine Learning, Content-Based Image Retrieval (CBIR), Image Processing, Feature Representation, Deep Learning, Convolutional Neural Networks (CNN), Image Classification, Visual Similarity, Texture Analysis, Color Histogram, Edge Detection, Image Recognition, Computer Vision, Image Matching, Neural Networks, Similarity Measurement, Image Database, Pattern Recognition, Artificial Intelligence.

I. INTRODUCTION

Satellite Image Processing is a type of remote sensing that collects coherent information about the earth's surface using pixel resolutions. Spatial Resolution has a high resolving capacity or ability to distinguish objects. Satellite imaginary is a great challenge while images are not high resolution always, more complex for comparison, clouds block view is another distortion, double-checking for ground-truthing which causes a lot of time-wasting. Feature extraction and similarity assessments are used to retrieve images. Image Processing on, Therefore, the classification algorithm gives better results on retrieving remote sensor satellite images with ineffective time and accuracy. The machine-learning algorithm has its significant advantage of image classifications results in higher dimensions. Comparing and analyzing image retrieval using three classification algorithms have greater significance to classify the image pattern recognition. There are many classifiers developed by many researchers like naïve Bayes, k-nearest neighbor, support vector machine, decision tree, and random forest. In this paper, we will be comparing two different classification methods. Experimental evaluations were conducted on the land cover satellite image dataset. Then we will elucidate the four different classification performances for best retrieval of these methods.

II. RELATED WORK

According to Kun Fang [1] realize the image classification based on the Naïve Bayes method with multiple features gives a fast and good prediction. In literature four feature extractions of gray histogram features, SIFT features SURF features, and reducing the dimension of the dataset. Compare the SURF and SIFT features which are excellent in scale-invariant features. SURF gives the strategies of easy detection and description for the targeted image in an efficient way. SIFT Feature is also high and has some limitations to prominent points. The result shows that the SURF feature proves the best for image processing [1].
H. Sebai and A. Kourgli [2], improving high-resolution satellite image retrieval with a Linear SVM classifier and data augmentation was achieved, that compares the results different classes of land use land cover area through descriptor of Dense SIFT, SIFT(RGB), SIFT + Cross Adaptive Illuminance gives better results for complex images. This classifier related to

© 2026 IJSRET

color representation in SIFT gives better results if the objects are different shapes and sizes of HRS images [2].

Ansari, Mohd. Aquib and M. Dixit [3] proposed a method of CBIR system which involved HSV color histogram, discrete wavelet transforms, global as well as local histogram descriptor and SVM Classifier based on RBF kernel function, used for classifying the data in an efficient way than the other methods, here the classification done by more than two classes when compared to linear and polynomial function. The RBF-SVM classifier helps to classify the image dataset effectively and the method produced better results than the existing scheme of image retrieval [3].

Bhosle, Nilesh & Kokare, Manesh[4], bring in found of the classification-based on Random Forest approach is seen as a developing way to overcome the semantic gap, it takes an input vector and categorizing according to each tree in the forest, finally, the majority of votes are used to conclude the input vector class. Imbalanced training set problem is overcome by dynamic learning approach. [4].

III. METHODOLOGY

In this paper, to examine the performance of several classification algorithms on a variety of data training sample strategies, an area of 256 x 256 pixels of Landcover to examine the performance of several classification algorithms on a variety of data training sample schemes of ten classifications areas.

1. Image Data Set

Dataset has been collected from the data sources of various metropolitan regions around the country, the images were manually extracted from massive imagery from the United States of Geological Survey's (USGS) National Map Urban Area Imagery collection of 10 categories.



Figure 1. Sample Land use Satellite images from the Dataset

Figure1.illustrated with various metropolitan regions around the country, the images were manually extracted from massive imagery from the USGS National Map Urban Area Imagery collection. In Land cover area is divided into 10 categories: agriculture, aircraft, baseball diamond, seashore, buildings, and transportation, chaparral, harbor, intersection, parking lot, storage tanks of each classes taken 100 images.

2. General Flow of Different Classifiers

The process illustrated in Figure 2. is general flow diagram of different classifiers. The procedure is showed by,

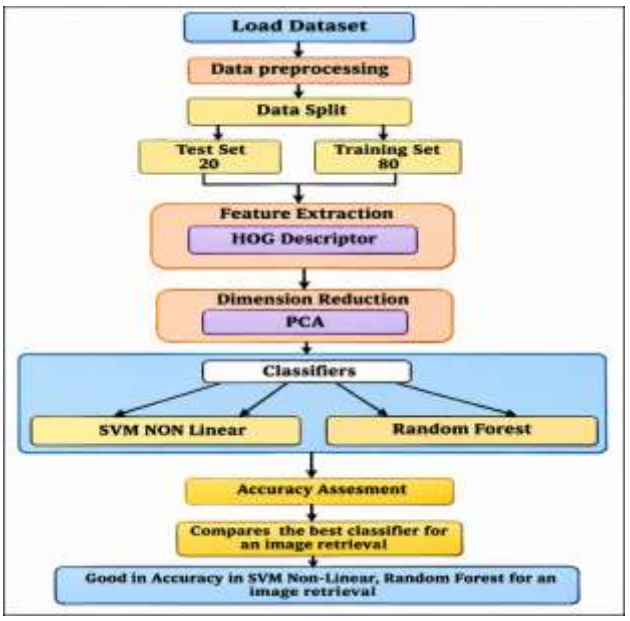


Figure 2. General Flow of different Classifiers

The training samples of each 10 classes having 800 images whereas testing samples of each classes having 200 images. Training samples are trained as model for giving accuracy of good results. From deriving the best classifier which leads to given good prediction result in image retrieval.

IV. CLASSIFICATION TECHNIQUES OF REMOTE SENSOR SATELLITE IMAGES

1. SVM Non-Linear Classifier

To classify data across a high dimensional space, usage of kernel functions gives good results for complex data which are non-linear separable. Conversion of non-linear spaces to linear spaces and turns data into a different dimension, allowing it to be categorized. Usage of two-dimension x and y, for non-linear data adding the third dimension called z=x2+y2, the plane

© 2026 IJSRET

looks three-dimension parallel to the x-axis, for 2 dimensions with z=1 for their circumference radius for non-linear data.

2. Random Forest Classifier

Random forest is one of the supervised learning algorithms which is mainly used for classification with more flexible and easier to use. It creates a decision tree on randomly selected data samples, receives predictions from each tree, and votes on the best option. It also serves as a powerful indicator of the value of the feature. Technically based on the divide and conquer method of a decision tree on the randomly split dataset, the collection of decision tree classifier is a forest. Each decision tree is created by its attributes using an attribute selection indicator such as gain, gain ratio, or Gini index.

V. FEATURE EXTRACTION METHODS

Feature extraction is one of the image compression techniques, which divides and minimizes a large set of raw data into smaller groupings without losing any details. If the dataset is a larger number of variables, it helps to get the best feature from larger datasets by selecting variables into feature vector can reduce the amount of data. This paper focus on histogram-oriented gradients and principal component analysis for this activity [12]. These features are simple to use while still accurately and uniquely describing the real data set.

1. Histogram Oriented Gradients

Image feature extraction process completed by the histogram of oriented gradients (HOG) which gives good result for retrieval accuracy. The method focuses on the shape of an object and counts the number of times a gradient orientation appears in a small segment of an image, focusing on the shape of an object. The process is done by making extraction on gradient and orientation of the edges, then histograms are created by their pixel values [13].

Steps for calculating HOG Feature.

Step1. Define the patch dimension on HOG feature (128 x 128) pixel sizes.

Step2. Calculate gradients for x and y direction for computation of overall gradient magnitude.

$$\|\nabla f\| = \sqrt{\left(\frac{\partial \delta}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

Step3. Calculate gradient angle orientation

$$\theta = tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)$$

Step4. Gradient Magnitude and gradient angles for specified patches for the images represented by star for its combination.



Figure 3.a. Airplane Image

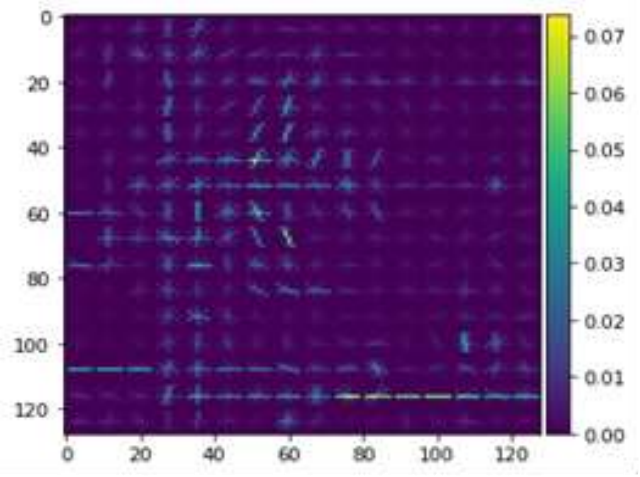


Figure 3.b.HOG of Airplane Image

Airplane images of figure 3.a. is the original image 3.b. is the representations of HOG features of the corresponding image. The star represents gradient magnitude and gradient representation of the images with x and y directions.

5.2. Principal Component Analysis

PCA is the process of image reduction of the dimension on a larger dataset which is compressed into smaller ones without

© 2026 IJSRET

losing any information. Reduction on number of variables for the training samples needs a compromise on the model accuracy, but PCA gives good accuracy also preserve data as much as possible.

Step1: Loading the Image dataset.
Step2: Splitting an image as R, G, B Arrays.
Step3: Apply Principal Components to distinct Arrays.
Step4: Compressing the Images.
Step5: Increasing the number of Principal components, produces the clear output images.

The image is made up of rows of the pixel that are stacked and to produce a single image. Each pixel value reflects the image's intensity value that creates a matrix by treating each row of a pixel as a vector. PCA is employed to compress the multidimensional data into a single column and save the data as much as feasible.

VI. RESULTS AND DISCUSSION

All assessments were done on a Windows 10 computer with an Intel(R) Core (TM) i5 1.8 GHz processor, 1 TB HDD, and 8 GB RAM. Here, Spyder tool is used for our experiment in Machine learning Python Environment.

Table 1. Accuracy Assessment I.

Accuracy For Different Classifiers Using Conventional Methods	
SVM Non-Linear	Random Forest
41%	42%

Table 1. shows the accuracy of different classifiers are assessed and compared for the test dataset by measuring their metrics of precision, recall, and f1 score values, here the accuracy for every classifier used the only conventional method of machine learning with raw images.

Table 2. Accuracy Assessment II.

Accuracy for different classifiers using PCA	
SVM Non-Linear	Random Forest
77%	66%

Table 2. shows the accuracy with PCA for each classifier gives a good result when comparing existing methods. In there the classifier of SVM-Non-Linear gives high accuracy of 77% for 1K image dataset, whereas Random Forest is 66% and Naïve Bayes 68% places their accuracy to the next level of SVM Non-linear.

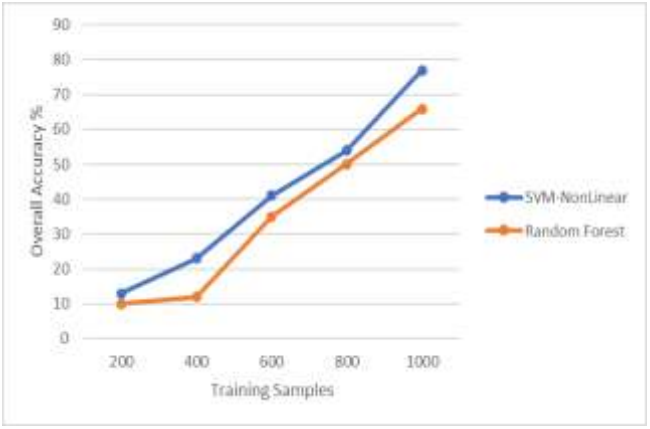


Figure 4. Overall accuracy for SVM Non-Linear and Random Forest.

Figure 4. Illustrated overall accuracy of different training samples in SVM Non-Linear and Random Forest. Overall accuracy is concerned different categories of images with these two classifiers, whereas both of them give equally same percentage with some minor differences only of greater flexibility. Hence Machine learning algorithms have given prominent support for image retrieval features and give high accuracy among classifications.

VII. CONCLUSION

In this paper, SVM Non-linear, and Random Forest classifiers are applied to remote sensors image datasets to get good retrieval performances. The major goal of this study is to analyze the performances of the classifiers with their accuracy and provide comparison results of best suggestions for future research work. The SVM-Non-Linear method has good accuracy compared to Random Forest. The ten categories of remote sensors images, each of them having 100 images have an accuracy of 77% by using the SVM Non-linear method. The Random Forest approach then yields 66 % accuracy, which is the next degree of accuracy. Since originating the best classifier which leads to giving good prediction results for image retrieval.

REFERENCES

1. Kun Fang, *Naive Bayes Image Classification Based on Multiple Features*, Article ID: 1171 Vol **2**, Issue 1, 2019, Article identifier:12-15., doi: 10.24294/csma.v2i1.1171
2. H. Sebai and A. Kourgli, *Improving high-resolution satellite images retrieval using Linear SVM classifier and*

© 2026 IJSRET

data augmentation, 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), pp. 1-7, doi: 10.1109/PAIS.2018.8598536, 2018.

3. Ansari, Mohd. Aquib and M. Dixit. *A Refined Approach of Image Retrieval Using RBF-SVM Classifier*. International Journal of Signal Processing, Image Processing and Pattern Recognition, **10** (2017): 43-56.
4. Bhosle, Nilesh & Kokare, Manesh., *Random forest-based active learning for content-based image retrieval*. International Journal of Intelligent Information and Database Systems (2020). **13**. 72. 10.1504/IJIIDS.2020.108223.
5. M. R. Kapadia and C. N. Paunwala, *Analysis of SVM kernels for content-based image retrieval system*, 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1409-1414, doi: 10.1109/ICECDS.2017.8389676, 2017.
6. Renukadevi, N. and Dr. P. Thangaraj. *Performance Evaluation of SVM – RBF Kernel for Medical Image Classification*. (2013).
7. Nayak, Janmenjoy & Naik, Bighnaraj & Behera, Dr. H. (2015). *A comprehensive survey on support vector machine in data mining tasks: Applications & challenges*. **8**. 169-186. 10.14257/ijdta.2015.8.1.18.
8. Nurwauziyah, Iva & Sulistyah, Umroh & Gede, I & Putra, I Gede Brawiswa & Firdaus, Muhammad. (2018), *Satellite Image Classification using Decision Tree, SVM and k-Nearest Neighbor*.
9. Ye, Famao & Xiao, Hui & Zhao, Xuqing & Dong, Meng & Luo, Wei & Min, Weidong. (2018). *Remote Sensing Image Retrieval Using Convolutional Neural Network Features and Weighted Distance*, IEEE Geoscience and Remote Sensing Letters. PP. 1-5. 10.1109/LGRS.2018.2847303.
10. Gu, Yating & Wang, Yantian & Li, Yansheng. (2019). *A Survey on Deep Learning-Driven Remote Sensing Image Scene Understanding: Scene Classification, Scene Retrieval and Scene-Guided Object Detection*. Applied Sciences. **9**. 10.3390/app9102110.
11. Shunmuga Kumari D and Arunachalam A S, "*Survey on Intelligent Approach of Content Based Image Retrieval and Classifications*", European Journal of Molecular & Clinical Medicine, **7**, 3, 2020, 5006-5013.
12. Chongtham Rajen Singh and Gobinath R, "*Identify missing countries using GEEBLL iterative method for analyzing tweets opinion*", Materials Today: Proceedings, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2020.10.758, (2020).
13. Britto M John and Gobinath R, "*REHC: Reduction through exclusive homogeneous clusters for imbalance dataset*", Materials Today: Proceedings, (2020).

© 2026 IJSRET