Ship Detection in Satellite Imagery by Multiple Classifier Network

Ahsan Raza Siyal[1], Zuhaibuddin Bhutto[2], Kashif Saleem[3], Abdul Sattar Chan[4], Mudasar Latif Memon[5], Murtaza Hussain Shaikh[6], and Saleem Ahmed[7]

[1]Electronics Engineering Department, Dawood University of Engineering & Technology, Karachi, Pakistan
[2]Department of Computer Systems Engineering, Balochistan University of Engineering & Technology, Pakistan
[3]Telecommunication Engineering Department, Dawood University of Engineering & Technology, Karachi, Pakistan
[4]Eletrical Engineering Dept. Sukkur IBA University, Sukkur
[5]IBA Community College Naushehro Feroze, Sukkur IBA University, Pakistan
[6]Department of Computer Systems Engineering, Kyungsung University, Busan, South Korea
[7]Computer System Engineering Department, Dawood University of Engineering & Technology, Karachi, Pakistan

Summary
Automatic ships detection in low-resolution satellite imagery is a challenging task. In this paper, we compare state of the art classifiers like support vector machine, Random Forest, Linear discriminant analysis, K nearest neighbor and deep Convolutional neural network for ship detection and classification in satellite imagery. We proposed a novel method with Convolutional neural network which improves robustness of the system, accuracy for ship detection and reduces noise due to weather conditions and high waves. The open-source dataset of planet ship is used to test the results of the proposed scheme, containing preprocessed 2800 images. The dataset has two main classes “ship” and “no-ship”, within ship class 700 images of various type and sizes of ships having diverse atmospheric conditions. The class having 2100 images of no-ship with random land covers with the ocean, earth and cities land cover with ocean featuring no ship in them. The simulation results presented here show that the proposed scheme gives improved detection accuracy and robustness in low-resolution satellite images having bad weather conditions.

Keywords
satellite, state-of-the-art, convolutional neural network, robustness.

1. Introduction

There is a large number of satellite imagery is being captured on a daily basis that has outgrown the capability of many to manually extract and analyze required information from satellite images, this leads recent researchers to employ artificial intelligence (AI) and machine learning (ML) algorithms to automate the process. Raising need for implementing machine learning on a larger number of images requires more computational power and complexity of the model, to be trained would be increased as the data increases with passage of time and that creates a space for modern researchers to focus on methods that generate acceptable results with trade on complexity and computational power requirements. Automating the process of analyzes and information extraction would be useful to many matters such as observe, track supply chain and port activity. Previously, the research related to ship detection was centered on detecting ships in synthetic aperture radar images [1] (SAR). Recently researcher has shown more interest in optical satellite imagery to detect ships because it has more spatial contents and high resolution compared with SAR images. The state-of-the-art machine learning approaches provide techniques to learn patterns and classify labeled datasets, the support vector machine has proven well-established classifier. The capability of SVM depends on the chosen kernel function for the problem.

2. State-of-the-Art Classifier

2.1 Support Vector Machine

SVM is a classifier that creates an optimum linear boundary to decide classes regions and the line that splits decision boundary is known as hyper plane. The decision boundary is generated by the elements of training sets with weighted combination and the elements are termed as support vectors and generate the decision boundary line between classes, equation 1 represents N number of sets.

$$(x_1, y_1), \ldots (x_i, y_i) \ \ldots \ (x_N, y_N) \ x_i \in R^N, y_i \in \{-1,1\} \quad (1)$$

If the data can be separated linearly, maximum margin classification separates two classes by hyper-plane which maximizes the distance of support vectors. Generally hyper-plane is termed as optimal separating hyper-plane (OSH) and representation shown by equation 2.

$$f(x) = \sum_{i=1}^{N} \propto_i y_i \left(x_i^T x \right) + b \quad (2)$$

The subsets contained by training samples that are support vectors whose $\propto_i$ outlines the solution and has non zero value.

Manuscript received August 5, 2019
Manuscript revised August 20, 2019

Fig.1 elaborates the graphical illustration of hyper-plane, where x_1, x_2, and x_3 represents the support vectors.

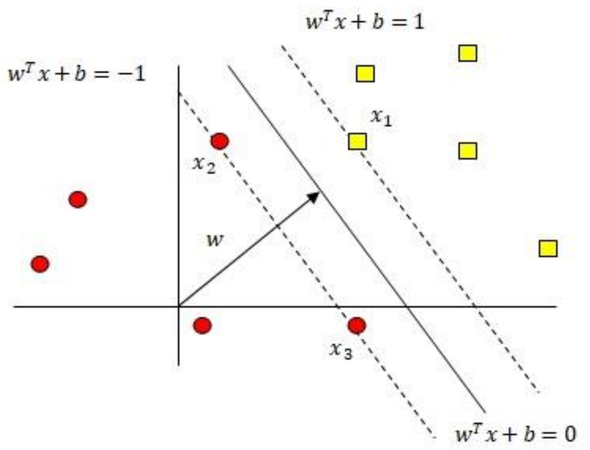

Fig. 1 Graphical illustration of optimal hyper-plane

The classes are represented by the subsets $y_i = +1$ and $y_i = -1$. In case if data can be separate linearly then equation 3 represents hyper-plane.

$$D_i(x) = w^T x + b = 0 \tag{3}$$

where an input vector is represented by x, weight vector w and bias b. support vector machine provides w and b parameters such as distance will be greater than 1 between the nearest sample point and hyper-plane.

If data cannot be separated linearly, input vectors can be mapped on the dot product of space F having higher dimensions which are termed as feature space. Hyper-plane would be now created from feature space as a linear function of vectors having high dimension. In order to resolve this problem kernel method is implemented.

In equation 2, substituting $x_i^T x$ by $\varphi^T(x_i)\varphi(x)$ gives equation 4:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} y_i \propto_i \varphi^T(x_i)\varphi(x) + b\right) \tag{4}$$

$\varphi^T(x_i)\varphi(x)$ is substituted by $k(x_i, x)$ in equation 4, simply kernel method applied to support vector machine (SVM) supported by Mercer's theorem. Non-linear SVM formulated as below:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} y_i \propto_i k(x_i, x) + b\right) \tag{5}$$

2.2 Random Forest

The organization of several small decision trees generates random forest classifier; independently selected random input vector creates each tree which can only vote for a most popular class within the tree. Classifying input vector(x), $C_{rf}^{B} = majority\ vote\{C_b(x)\}_1^B$, where $C_b(x)$ is predicted the class of b^th tree. The characteristics of random forest that make it distinguish from conventional classifiers are the combination of many classifiers. The allowance of the random forest to oblige increase the versatility of trees by creating them from a subset of the training data set [9]. The bagging method is employed to generate training dataset drawing substitute N examples, where 'N' represents the size of the dataset. The classifier during the training may use some of the data within dataset more than once while other data might not be used at all, and it is the reason random forest achieve greater constancy. The use of variant input data makes it more robust and increases classification accuracy [10]. The random forest classifiers are immune to overtraining due to employing bagging-based methods [15], unlike other classifiers that use boosting method [11].

The selection of suitable characteristics that optimizes variation within the classes is the key requirement for designing the tree. There are several probabilities for setting parameters utilized in decision trees, some of them are Chi-square and Gini-index. The random forest usually utilizes Gini-index which may be written as:

$$\sum\sum_{i\neq j}\left(f\,\frac{(c_i,T)}{|T|}\right)\left(f\,\frac{(c_j,T)}{|T|}\right), \tag{6}$$

where $(f(C_i,T)/|T|)$ illustrates the probability of the selected case which is related to C_i by utilizing known feature combinations the decision tree is stretched to its maximum depth. The methods that work on tree principle their performance can be affected by pruning approaches, so it expands without pruning which enhances its performance.

2.3 Linear Discriminant Analysis (LDA)

Consider $X_1 = \{x_1^1, \dots, x_{l1}^1\}$, $X_2 = \{x_1^2, \dots, x_{l2}^2\}$ are distinct classes, where $X = X_1 \cup X_2 = \{x_1, \dots, x_l\}$. LDA method is also called Fisher's linear discriminant which is found by the vector w which maximizes

$$j(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_W \omega} \tag{7}$$

where,

$$S_B := (m_1 - m_2)(m_1 - m_2)^T$$

and

$$S_W := \sum_{i-1,2}\sum_{x\in x_i} (x - m_i)(x - m_i)^T$$

where S_B and S_W are the class scatter matrix and scatter matrix inside the class, respectively. The m_i maybe given as:

$$m_i := \frac{i}{l_i} \sum_{j=1}^{li} x_j^i \tag{8}$$

$J(\omega)$ to be optimized in a way that numerator maximized which projected the class and its denominator minimizes.

Optimal Bayes classifiers are compared by Posteriori probabilities of every class and it generates a model to the class with optimal probability. The distribution step is an important one in most of the classes and produces close estimation. However, if the distribution of all classes tried to be standard, may reach quadratic discriminant analysis which measures the Mahalanobis distance of the pattern towards the center of the class. In order to reduce complexity in the problem, it is assumed that all classes have equal covariance, it turns quadratic discriminant analysis into linear. Maximizing the vector 'ω' in the same direction by Bayes discriminant optimal classifier making it easy to demonstrate for the two-class problem. Fisher's linear discriminant has recognized to be very dominant because linear models prevent overfitting and robust against noise.

2.4 K Nearest Neighbor

Suppose $X = [x_1, \dots, x_N]$ is the data to be trained having 'N' elements and dimensionality D and $X_i = [x_{i1}, \dots, x_{ik}]$ are the nearest elements of X_i . Data to be tested is X_t with N_t elements. $X_0 = [x_{01}, \dots, x_{0k}]$ are random samples from the test data set having K number of nearest elements from training data set which also consist labels $[l_1, \dots, l_k]$. Let suppose there are 'C' number of classes in the data and may be denoted as Ω = [Ω1, . . ., ΩC]. Locating the nearest points of the test data from training data.

The distance metric obtained from the training data enhances classification and given as,

$$dis\left(x_i, x_j\right) = \left\| T\left(x_i - x_j\right) \right\|^2, \tag{9}$$

where T represents a linear transformation. It has assumed that points which have been tested have equal weights. If the different weights would be assigned to the nearest neighbors, then the classes with the largest values of its nearest neighbor's sum would classify the point being tested. The weighted k nearest neighbors is given as,

$$j^* = \arg\ \max_{j=1,\dots,C}\ \sum_{i=1}^{k} \omega_i \delta\left(l_i, j\right) \tag{10}$$

where δ represents the Kronecker delta and ω_i represents the weight for x_{0i}. The distance weighted k NN (DS-WkNN) is found using:

$$\omega_i = \frac{dis\left(x_{0k}, x_0\right) - \left(x_{0i}, x_0\right)}{dis\left(x_{0k}, x_0\right) - \left(x_{01}, x_0\right)}, \tag{11}$$

where $[x_{01}, \dots, x_{0k}]$ are in ascending order in accordance to their distance from X_0.

3. Edge Detection Methods

Edge detection methods play a vital role in many modern technologies having image processing capabilities to extract desired information. The edge detection methods used to identify object outlines to distinguish a specific object from the background in an image. These methods also employ to get better visibility and appearance of a blurred image. In this paper, for above-mentioned reasons, it is used to extract features to detect ships from satellite imagery. Four popular methods are been used in the paper however, many methods have been introduced for the same purpose.

3.1 Sobel Operator

Sobel operator/filters are used in machine vision to detect edges in a given image. It approximates the smoothing and gradient. The kernel which has dimensionality 3x3 would be convoluted with the given image to find gradients in both horizontal and vertical axis. There are two convolution kernels G_x to mask horizontal and G_y for vertical orientation.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

These two kernels are intended to create eminence in horizontal and vertical edges within the image. Both kernels may be convolved separately with the image in order to stress gradient. Consider $\boldsymbol{I_m}$ as an image matrix then both vertical derivative and horizontal derivative are found using:

$$H_d = G_x * I_m \tag{12}$$

and

$$V_d = G_y * I_m. \tag{13}$$

The horizontal and vertical gradient can be combined to obtain absolute gradient magnitude for any point in either orientations [12], The gradient magnitude is given by:

$$|G| = \sqrt{{G_x}^2 + {G_y}^2}. \tag{14}$$

The approximate magnitude is computed using:

$$|G| = \left|G_x\right| + \left|G_y\right|. \tag{15}$$

The orientation angle of the edge given by:

$$q = \arctan\left(\frac{G_y}{G_x}\right). \tag{16}$$

3.2 Prewitt's Operator

The Prewitt filter is also 3x3 dimensional and works on principle alike Sobel kernel [13]. This operator has also two kernels to determine gradients for both orientations. Prewitt filter is known for quick response method for edge detection. The only factor that makes it distinguish from Sobel kernel is the Spectral response which makes it suitable for well-contrasted and noiseless images.

$$G_x(prewitt) = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix},$$

$$G_y(prewitt) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

3.3 Roberts Cross Operator (ROC)

The ROC finds the spatial gradient in both dimensions in the image. It makes edges prominent using high spatial frequencies that usually specify edges, therefore, it usually stresses these regions. The input image to the kernel is the grayscale image and as same for output. This operator uses two 2x2 convolution kernels for vertical gradient and the horizontal gradient is:

$$G_x(Robert\ Cross) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$G_y(Robert\ Cross) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

The operator makes edges prominent which are 45° to the pixel grid. The maximum response would be on these edges, after convoluting two kernels separately with the image vertical and horizontal derivative may be used together to extract absolute gradient magnitude for all pixels of a particular gradient. The absolute gradient can be computed similar to the Sobel operator.

The orientation angle that increases the spatial gradient of the edge is found using:

$$q = \arctan\left(\frac{G_{y2}}{G_{x2}}\right) - \frac{3\pi}{4}. \tag{17}$$

3.4 Laplacian of Gaussian (LOC)

The LOG operator utilizes the signal-to-noise ratio of the image to find an optimal filter to emphasize edges. Firstly, a low pass filter is used to smoothing the image and then the high pass filter of the Laplacian operator. Gaussian filter function [14] may be given as,

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left(x^2+y^2\right)\right), \tag{18}$$

where σ^2 is the standard deviation. The low pass filtered image f(x,y) multiplied with Gaussian filter function gives us:

$$\begin{aligned} g(x,y) &= \nabla^2\left[f(x,y) * G(x,y,\sigma)\right] \\ &= f(x,y) * \nabla^2 G(x,y,\sigma) \end{aligned} \tag{19}$$

where $\nabla^2 G$ is the LOG operator.

4. Proposed Scheme

The selection of the classifier for the given problem depends on the nature of data and it is always the matter to concern before going further to the learning process, because it is time-consuming and computational heavy task. Any algorithm cannot produce similar results for different data sets. The algorithms available in current times have different specialties and disadvantages too for specific cases. The major issue in the recent days is to not have a system that performs with high accuracy of object detection and classification in satellite imagery that is taken under different environments and noise interferences due to rain and high waves, for this reason, models suffer reduced accuracy and robustness. Feature extraction is also a vital part of the classification process before model to be trained the data must be extracted with desired information which we intended to learn by the model. It reduces the training data, time and computational power requirements. The feature extraction technique directly affects the classifier accuracy so, the selection of the feature extraction method is also very crucial in the creation of the model.

In the proposed method, features are extracted from the input image by Sobel kernel which emphasizes edges in it and the classification is performed by multiple classifier systems. Block diagram of the proposed system is shown in figure 2.

In the initial stage, the input image is generated as a matrix I_m. Features would be extracted by Sobel kernel which prominent the edges in the image. The Sobel kernel uses two 3x3 convolution kernels to find gradient in both dimensions. The convolution masking is given by using the equations 12 and 13. The gradient is described in equation 14 and the angle of orientation is obtained using equation 15. The training data can be preprocessed with Sobel kernel and this data is used to train multiple classifiers. Each classifier provides its decision and vote for only one class. The voting method would be weighted and provides the most popular decision as to its final.

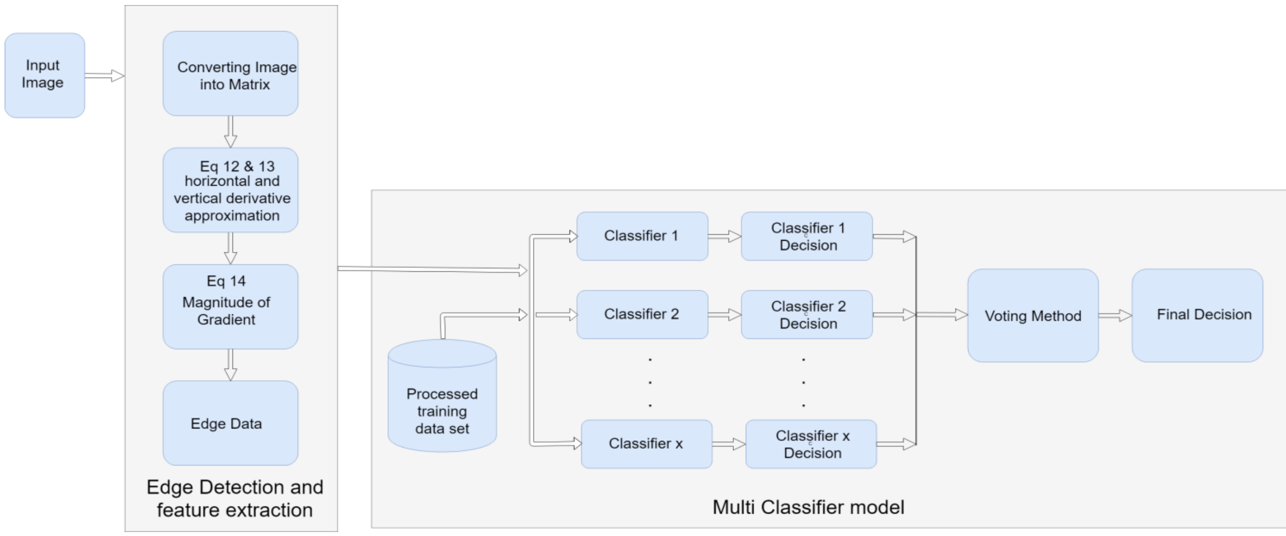

Figure 2: Block diagram representing the proposed model

4.1 Voting Methods

In the paper, the weighted voting system is used to choose the relative significance of the individual classifier vote. The cross-validation accuracy is the factor that determines the weight of the classifier in the model. The voting system depends on three parameters, voters (C1, C2,…, CN), the number of classifiers are denoted by 'N', weights of the classifiers weight (Cw1, Cw2,…, CwN). and quota 'q'. The quota is the threshold value required to decide the winner of the election. The weights and the quota can be given as,

$$C\omega_\sigma = \frac{kc_\sigma * 100}{\sum_{\sigma=1}^{N} kc_\sigma} \tag{20}$$

and

$$[q : C\omega_1, C\omega_2, \ldots, C\omega_N], \tag{21}$$

where, Kc_σ represents cross-validation of individual classifier and $\sum_{\sigma=1}^{N} 1kc_\sigma$ gives the cross-validation sun of 'N' number of classifiers.

From equation 20, the weights are listed in ascending order. Every individual classifier would vote for only one class and its vote would have weighted significance in the final decision. The sum of weights of all classifiers which has voted for same class must be greater than 'q' which is 50% and it will choose the winner in the election and provides a final decision of the multiple classifier model otherwise the election would be repeated.

5. Experimental Results

The proposed method in the paper is tested on open source data set provided by Planet [16]. The data set contains two classes and 2800 images. "Ship" class contains 700 images of different type and sizes of ships under different weather conditions. The "non-ship" class includes 2100 images of lands, greenery beaches, and oceans with no ships in them. The data set provides images in the form of a vector, each image vector contains 19200 elements within which first 6400 elements represent red pixels. Blue and green pixels also have the same number off elements as red, it forms 80x80 RBG image.

Initially, the data set is already preprocessed, and the four states of the art classifiers are trained on 20 % of the data, that includes 200 images of "ship" class and 360 images of "non-ship" class, in total 560 images are used for training purpose. The model is tested on 80% of the data that includes 500 images of "ship" class and 1740 images of "non-ship" class and total of 2240 images.

Now, the data set is convoluted separately with four edge detection kernels that are Sobel, Robert cross, Prewitt and LOG to evaluate the performance of state-of-the-art classifiers, table 1 shows the accuracy of classifiers without applying edge detection techniques on data set.

Now, the proposed method is evaluated for the performance and accuracy, the data set is convoluted with edge detection kernels, the four classifiers are trained with the 20% of data and tested on 80%.

Table 1. Illustrates the accuracy of conventional schemes without applying edge detection.

S No.	Conventional Scheme	Output
1	Support Vector Machine	80.45
2	Linear Discriminant Analysis	78.21
3	Random Forest	81.76
4	KNN	83.65

Table:2 Proposed scheme results on experimental data

Accuracy of various schemes on data							
S No.	Scheme	Sobel	Robert cross	LOG	Prewitt	10-Flod cross-validation Accuracy	Classifier weight(Cw)
1	Linear Discriminant Analysis	82.61	84.61	71.87	82.86	0.87	23%
2	SVM	89.79	83.49	79.11	89.79	0.97	26%
3	RandomForest	93.24	88.51	80.59	93.19	0.97	26%
4	KNN	93.02	90.50	82.29	93.11	0.95	25%

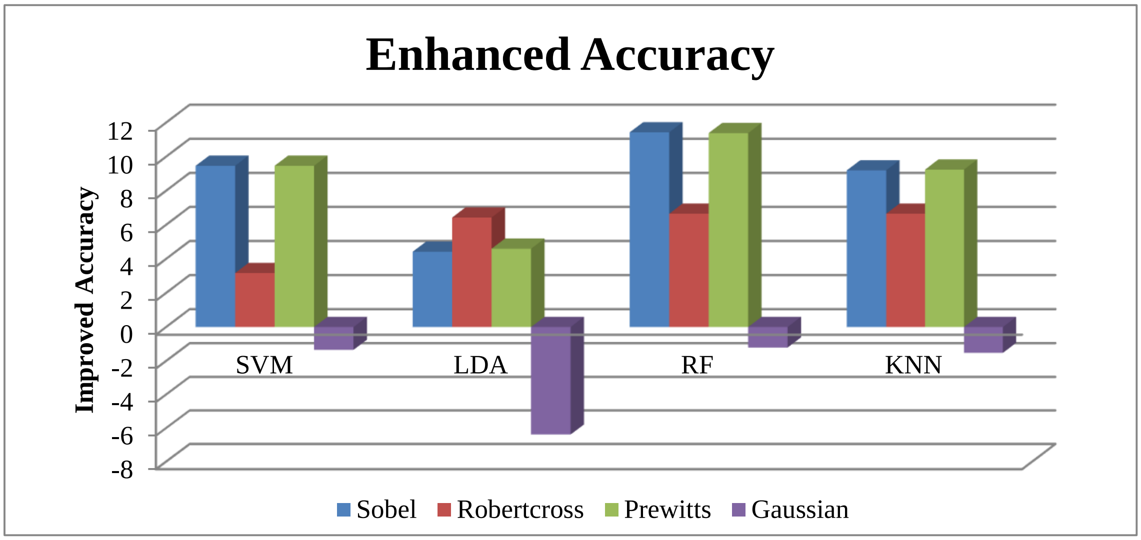

Figure 3: illustrates the enhancement in accuracy of the proposed model after applying the edge detection.

The weights are determined by the factor of cross-validation accuracy during the training of the classifiers and the final decision is provided by the voting method. Table 2 shows the accuracy of the classifiers with their individual weights deciding the significance in the decision.

Figure 3 shows the improvement in the accuracy of classifiers; the highlight of the experiment is Sobel and Prewitt's techniques work very well for Random forest and it improved the accuracy by 11%. The proposed model shows 99% accuracy on the test dataset taking all four classifiers decision in consideration and gives the final decision by the voting method.

6. Conclusion

In the paper, we have proposed ship detection multiple classifier network for satellite images by using Sobel edge detection technique. We tested the accuracy of SVM, Random Forest, LDA and KNN classifiers on open source ship data set, it is found that KNN performs better on preprocessed data set and after, we convoluted the data set with edge detection kernels including Sobel, Robert cross, Prewitt's and Laplace of Gaussian. The improvement in the accuracy of all four classifiers has observed and Random forest beat others by improving 11% in classification accuracy when the Sobel kernel is used for feature extraction. The proposed method takes the decision from all classifiers and the weighted voting system produced 99% of classification accuracy on the data set.

References

[1] D. Crisp, “The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery” Australian Government, Department of Defense, 2004.

[2] F. Wu, Z. Zhou, B. Wang, J. Ma, “Inshore Ship Detection Based on Convolutional Neural Network in Optical Satellite Images”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no.11, pp. 4005-4015, 2018.

[3] F. Yang, Q. Xu, B. Li, “Ship Detection From Optical Satellite Images Based on Saliency Segmentation and Structure-LBP Feature”, IEEE Geoscience and Remote Sensing Letters, Vol. 14, no.5, pp. 602-606, 2017.

[4] M. Yang, M., & Guo, C., “Ship Detection in SAR Images Based on Lognormal ρ-Metric”, IEEE Geoscience and Remote Sensing Letters, vol.15, no.9, pp. 1372-1376, 2018.

[5] H. He, Y. Lin, F. Chen, H. Ming, and Z. Yin, “Inshore Ship Detection in Remote Sensing Images via Weighted Pose Voting”, IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 6, pp. 3091-3107, 2017.

[6] X. Han, Y. Zhong, R. Feng, L. Zhang, “Robust geospatial object detection based on pre-trained faster R-CNN framework for high spatial resolution imagery”, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017.
[7] H. Lin, Z. Shi, and Z. Zou, “Fully Convolutional Network With Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images”, IEEE Geoscience and Remote Sensing Letters, vol. 14, no.10 pp. 1665-1669, 2017.
[8] J. Huang, Z. Jiang, H, Zhang, B. Cai, Y. Yao, “Region proposal for ship detection based on structured forests edge method”, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017.
[9] L. Breiman, Machine Learning (1996) 24: 123. https://doi.org/10.1023/A:1018054314350.
[10] L. Breiman, Machine Learning (2001) 45: 5. https://doi.org/10.1023/A:1010933404324.
[11] M. Pal, and P. M. Mather, “An assessment of the effectiveness of decision tree”, Remote Sensing of Environment, vol. 86, pp. 554-565, 2003.
[12] C. John, "A computational approach to edge detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 679-698, 1989.
[13] X.L Xu, “Application of Matlab in digital image processing”, Modern Computer, pp. 35-37, 2008.
[14] Y.Q. Lv and G. Y Zeng, “Detection algorithm of picture edge”, Taiyuan Science & Technology, vol. 27, no.2, pp. 34-35, 2009.
[15] V. F. .Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp 93-104, 2012.
[16] Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com.