



SUPPORT VECTOR MACHINE METHOD TO IDENTIFY AND CLASSIFY ASTRONOMICAL OBJECTS

Sunitha P .
Assistant Professor,
Department of Computer
Science, Malnad College of
Engineering, Hassan, Karnataka,
India

Naveen Venkatesh,
Student,
Department of Computer
Science, Malnad College of
Engineering, Hassan, Karnataka,
India

Atreya B. M,
Student,
Department of Computer
Science, Malnad College of
Engineering, Hassan, Karnataka,
India

Abstract - Astronomy is the exploration of celestial objects, for instance, the moon, planets, stars, universes and for the most part everything beyond the Earth's atmosphere. In the past, astronomy has been used to measure time, mark the seasons, and navigate the vast oceans. It inspires us with attractive images and promises answers to the big questions. It acts as a window into the immense size and complexity of space, putting Earth into perspective and promoting global citizenship and pride in our home planet. Astronomical images hold an abundance of knowledge about the universe and its origin. Distinguishing objects in astronomical pictures is not an effortless task even for experienced astronomers. They are at distance measured at light-years, so it is likely that they will show up as faint bright points or mixed with different objects. This work aims at training a computer to identify and classify astronomical objects with satisfying efficiency.

Keywords— Astronomical Objects, Image Processing, Segmentation, Extraction,

I. INTRODUCTION

Astronomical images hold plenty of knowledge about the space and its genesis. The archives of such information have been expanding at a tremendous rate with changes in the innovation of the hardware. However, due to the large amount of data and the fact that many objects can be almost undetectable, a search by humans is inefficient, very slow, and inaccurate, if not almost impossible. Hence, it is necessary to develop highly robust, fast, efficient, and computer automated algorithms to detect the astronomical objects by means of image processing and computer vision.

The pictures give data about a huge variety of celestial objects (source) existing in the universe. It is even possible that some spots in these images may be considered as objects when actually they are not. Nevertheless, by considering the above few challenges connected with the astronomical images, identifying spatial objects has been a complicated task as

mentioned by Marc Masias(2014). Astronomical identification is the initial phase in the procedure of building astronomical objects. After identification, two different procedures are also performed:

- Classification: It categorizes the items into various types like stars, galaxy, clusters, extended objects etc.
- Photometry: To represent the flux, magnitude, and intensity of the object.

Machine Learning is a subfield of soft computing within computer science which has evolved by the investigation of pattern recognition and computational learning theory in artificial intelligence. It investigates the study and development of calculations which extract and forecasts on information. The algorithms work by constructing a model from illustration inputs to conclude data-driven expectation, besides following the strict static system guidelines. Machine Learning is kind of Artificial Intelligence (AI) that furnishes the computers with the ability to learn without modification. It concentrates on the enhancement of computer programs that can prepare them to build and differ when given with new data. Hongwen, et. al. (2007) suggests that ML is a strategy for data analysis which mechanizes investigative model building. Accessing algorithms which are repeatedly extracted by data, Machine Learning allows computers to find hidden insights without being explicitly programmed.

In computer vision and image processing, OTSU's method, named after Nobuyuki Otsu is used to automatically perform clustering-based image thresholding, or, the reduction of a gray level image to a binary image. The algorithm assumes that the image contains two classes of pixels following bimodal histogram. It then calculates the optimum threshold separating the two classes so that their combined spread is minimal, or equivalently, so that their inter-class variance is maximal. Consequently, OTSU's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis. OTSU's method is also directly related to the Jenks optimization method. OTSU is utilized to consequently



perform clustering-based picture thresholding reduction of a gray level picture to a parallel picture. The algorithm accepts that the picture includes two classes of pixels following foreground and background pixels. Immediately after it computes ideal limit to distinguish the two classes by that their joined spread is insignificant, so their inter-class change is maximum. Florin, et. al. (2011) adds that Otsu's technique comprehensively scans for the threshold that minimizes the intra- change characterized as a weighted aggregate.

Further, classification of the objects is experimented using Multi-Class Support Vector Machine (MSVM) method. The project aims at training a computer to identification and classification of astronomical objects with good efficiency.

II. LITERATURE SURVEY

Extensive work is done in the recent past related to the identification and classification of astronomical objects using Machine Learning techniques. Pavla Bromov, et. al. (2014) has proposed the methods in advance, innovations of astronomical that spectra acquisition has brought about a gigantic measure of information accessible in overall telescope archives. It portrays the beginning examinations in the investigation of spectral line profiles of emission line stars. Machine Learning algorithm is applied to consequently recognize Be and B[e] stars spectra in extensive files and order the sorts by the programmed way. Because of the extent of spectra accumulations, the measurements of reduction method depend on wavelet change is considered. G. Gordon (2013) also proposed Background estimation and removal depending on range and color. Utilization of range and color information gives superior outcomes which can accomplish with any of the information origins by itself. The range is an intense sign segmentation which is to a great extent free of color and thus not affected by the classic color segmentation issues. In any case, range itself is also not adequate for the great segmentation: depth estimations are once in a while accessible to every pixel in scene. Color segmentation is correlative in these cases. Segmentation of foreground in a defined frame is to perform through comparison with background statistics in range as well as standardized color.

Giang Hoang Nguyen, et. al. (2009) have discussed Learning Pattern Classification Tasks with Imbalanced Data Sets. A fundamental challenge in imbalance issue is that the small classes are regularly more helpful. However, standard classifiers have a tendency to be weighed down by the huge classes and overlook the small ones. In Machine Learning the imbalanced datasets has turned into a basic issue furthermore generally found in numerous applications, for example, recognition of fraudulent calls, bio-medical, engineering, remote-detecting, computer society and assembling commercial enterprises. To defeat the issues a few methodologies have been proposed.

Zhang, Y. and Zhao, Y (2007), have proposed the methods in SDSS (Sloan Digital Sky Survey) and have made available the data of astronomical objects from the visible region of the

spectrum. This provided a larger number of attributes to experiment with. Since SVMs had shown improved results with more attributes during the classification of these objects, Alternating Decision (AD) tree and Support Vector machines were used for separating quasars. Here the databases from SDSS and 2MASS catalogs were used. Similarly, Saida et. al. have proposed methods based on the success of ADTree algorithm in the identification of quasars. They have used Decision Tables of arranging every single source using the FIRST as well as 2MASS databases. The features and data that they used were largely similar to the BBN, MLP and ADTree models. Point source objects (like stars and quasars) tend to have their magnitude values falling in a particular range making the data more suitable to be classified using decision tree-like structures. In addition, they focused on enhancing the feature selection technique, used the Best first search feature selection. The Decision Table method also had a much higher accuracy (99%) in the case of quasars which proved that this was better than the previous models (BBN, MLP, and ADTree).

Saida Salahova and Azerbaijan(2007), have proposed the methods which are convenient to blend methodologies so as to increment the efficiency to identify and arrange the examined region having complex relief. The methodology of development of the Artificial Neural Networks including Back-propagation Error is proposed that depends on the actual position of spectral and spatial attributes in the picture of Landsat TM satellite. M. P. Lévesque and S. Buteau (2007) have proposed methods that presents an image processing technique to detect satellite streaks. This method is particularly useful in the context of surveillance of space where the positions of active satellites and other orbital debris must be monitored. Their study begins with a detailed analysis of a typical image, which includes several sensor artifacts and signal degradation.

Mohamed Anis Loghmariet et. al. (2006) have proposed methods that manages the issue of visually impaired source partition with remote detecting information depend on Bayesian estimation structure. Considering instance with multispectral pictures having similar zone over various spectral bands. Multispectral pictures comprise various channels where every channel consisting of data collected over the various frequency range. Every item transmits energy on huge spectral bandwidth that actually exists a critical relationship within channels. The objective is to set up new methodology depends on Two-Level Source Separation (TLSS). TLSS comprises with spectral separation with distinct utilized bands as well as spatial separation with neighboring pixels for every picture band. Spectral separation utilized Bayesian approach and depends on second order statistics method which relates overall connection by various spectral bands of the multispectral sensor.

Zhang, Y., Zhao, Y., Gao, D., et. al.(2008,2004), have proposed methods with more data available about astronomy objects from radio surveys and Infrared surveys. Bayesian



Belief Networks (BBN), Multi-Layer Perceptron (MLP) as well as Alternating Decision Trees (ADTrees) were used for separate quasars and non-quasars. Radio and infrared rays are in different ends of the spectrum and can be used to better identify quasars. The radio data was obtained from the FIRST survey and infrared data from the 2MASS. They cross-matched the data from both surveys to obtain one-to-one entries to get the entire set of radio and infrared details on the common objects which were again matched with the Veron Cetty and Veron catalog and the Tycho2 catalog. Y. Zhang and Y. Zhao (2004), have also proposed methods for data mining strategies which will exclusively execute on low-dimensional spaces, feature extraction is fundamental process before managing high dimensional information. Feature extraction is used to finding a minimum subset of actual elements which upgrades more than one aspect, instead of creating a totally new arrangement of dimensions of information. Feature transformation is a preprocessing procedure which changes actual components of data set to precise form, a reduced list of capabilities, by holding much more data as could be expected under the circumstances.

Brad Wallace and et. al (2007), have proposed a Space Surveillance Observatory for DRDC Ottawa in which he has created a small optical sensor to serve an RD instrument for Surveillance of Space (SofS). The sensor is implied as DRDC Ottawa Space Surveillance Observatory (SSO). It depicts the SSO, its reasoning and architecture, the computerized information reduction programming framework and underlying testing execution. In addition, it likewise depicts the way SSO is helping the Canadian Surveillance of Space Project's Concept Demonstrator (CD), NEOS Sat space-based SofS microsatellite as well as DRDC's further RD directions. SSO frequently watches over the star stare mode (SSM) as well as consequently creates metric and photometric information of deep-space RSOs with track-rate mode (TRM).

Summary of Literature survey: An attempt to distinguish satellites streaks in astronomical pictures is entirely new. The first method was proposed by Brad Wallace (2007). Paper begins by describing sensor qualities by considering the end goal to know definitely the principle components of the astronomical picture. The algorithm of streaks identification is proposed. Image-plane artifacts, as well as vast scale background, are reduced. Noisy elements are resolved in every picture pixel as well as items over these elements are distinguished. Picture moments are utilized for ordering items as point-like. Levesque et. al. (2007) have discussed the absolute illustrative presentation. Methodology depends on the progression of procedures, the most vital is being background estimation of star identification, including removal, taken after by iterative coordinated channel utilized for recognizing streaks.

From the above literature survey, referring to the identification and classification of astronomical objects using Machine Learning methods, it is found that numerous

specialists have utilized the accessible dataset containing pictures of celestial bodies. Researchers in the past have carried out their work to distinguish and arrange the objects. In this project, more number of components are being classified and identified. To begin with, the pictures will be pre-processed and then segmented. Further, features will be extracted and these features will be fed as input to the classifier to characterize and recognize the astronomical objects.

III. IMPLEMENTATION

Image processing is the technical analysis of an image by using complex algorithms. Here, image is used as the input, where the useful information is returned as the output. Beforehand image processing only suggested examining the discrete targets in an image. AI and Machine Learning can do real wonders in the realm of image processing. Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. Fig. 1 depicts the process flow of the proposed work.

- Data Acquisition or Input: Dataset is an accumulation of a set of images which are utilized for a specific application. In this project, publicly accessible datasets is considered. The datasets are gathered from Google Search for the training and testing process.

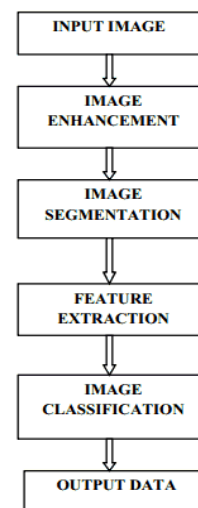


Fig. 1. Process Flow

- Pre-processing : Pre-processing is applied on images at the basic level of abstraction and its aim is to enhance the images which are useful and important for further processing. The Pre-processing steps that carried out are :
 - RGB to Gray Conversion.



- Histogram Equalization for Image Contrast Enhancement.

1) *RGB to Gray Conversion*: It changes over the real natural picture RGB to the grayscale picture using hue and saturation data by holding the luminance. In pre-processing, a grayscale advance picture having an estimation of every pixel leads to the individual example which conveys just valuable data. Pictures of this kind can be called as black and white and are made only gray shade, differing black has the weakest intensity to white having the strongest.

2) *Histogram Equalization for Image Contrast Enhancement*:

Histogram Equalization algorithm upgrades the differentiation of pictures by changing the qualities of an intensity picture so that the histogram of the yielded picture is approximately flat. Chahat et. al. (2013) have inferred that the principle thought of HE-based strategies is to re-locate intensity estimations of pixels to undergo intensity distribution uniform for the most extreme level. To improve the picture, a quality safeguarding Bi-HE (BBHE) technique is utilized to decompose the original picture, and it splits into sub pictures. Dualistic sub-picture histogram equalization (DSIHE) is like BBHE, however, DSIHE utilizes median of partition intensity the quality of the image. The obtained image will be segmented based on Otsu thresholding providing a better representation for further steps.

- Segmentation: A kind of dimensionality reduction is carried out in this phase. Segmentation is one of the important prerequisites of an image processing framework. It is the capacity to examine pictures and distinguish areas that have particular characteristics. Methods are expected to concentrate irregularities and different distortions from a picture. Image segmentation is frequently used to recognize the foreground from the background. The objective of segmentation is to rearrange the representation of a picture into something that is more important and easier to analyze. It partitions a picture into its constituent regions or items. This is commonly used to recognize objects or other significant data. Segmentation is carried out by Otsu thresholding method
- Feature Extraction: Changes the input information into the arrangement of elements. Feature extraction is used to identify the bit of data which is needed to solve computational task related to the application. The motivation behind the feature transformation is to decrease information by analyzing some of the characteristics that recognize the given patterns. An item is described with estimations, where qualities are fundamentally the same as to items in the existing class and distinct for items in various class. In this project, spatial and texture feature are extracted.

A) *Spatial features extracted*

- Area: Results a spatial that defines the actual number of pixels in region.
- MajorAxisLength: Defines the length (in pixels) of the major axis of the ellipse which has actual standardize second central moments as its regions.
- MinorAxisLength: Defines the length (in pixels) of the minor axis of the ellipse which has actual standardize second central moments as its regions.
- Eccentricity: Defines the eccentricity of the ellipse which has actual second-moment as the region. It is the proportion of the distance between the foci of the ellipse as well as its major axis length. That is around 0 and 1.
- Solidity: Defining the proportion of the pixels in the convex hull that is additionally in the region.
- Equivalent Diameter: Defines the diameter of a spherical particle which will give identical optical behavior to that of the particle being examined.

B) *Texture Features extracted*

Texture is a feature utilized for examination as well as interpretation of pictures. Texture is arrangement of local measurable characteristics of pixel. At some point when the GLCM are created, texture feature can be processed by GLCM.

- Energy: It is average of squared components in GLCM. The range between 0 to 1. For the consistent picture, the energy value is 1.
- Entropy: Entropy is a measure of randomness of intensity image.
- Contrast: It is an estimation of intensity contrast within a pixel as well as its neighborhood pixels in the entire picture.
- Homogeneity: It is an estimation of the components distribution in GLCM. It ranges from 0 to 1.
- Inverse Difference Moment: It is local homogeneity. At the point, while the local grey level is unique and inverse GLCM is high. IDM is high.
- Dissimilarity: Dissimilarity is measured by the certainty associated with a set of variables.
- Classification: Each training point has a place towards one of the various classes. The objective is to build a function which gives new information point accurately defining the place where actually the new point fits in.

The Support Vector Machine has hypothetically ruled Machine Learning strategy that has extraordinary outcomes in pattern recognition. Particularly to coordinate the classification for high-dimensional datasets good Machine Learning algorithms are applied. Mathew et. al. (2013) have experimented SVM to classify galaxy morphologies to spiral, irregular and elliptical galaxies.

The goal of the work is to evaluate SVMs for efficiency as well as cause object-based image that examines as an advanced computational knowledge technique. SVM method

to multi-class classification undergoes in view of actual picture items given by multi-resolution segmentation algorithm. Feature selection process occurs in convenient to the given elements for classification that include spectral, texture as well as shape data. For preparing SVM, test picture objects got by segmentation steps are utilized.

The proposed classification method undergoes outcome of last object classification. Classification outcomes are examined by considering Nearest Neighbor object-based classifier outcomes and are convincible. SVM approach appears to be exceptionally encouraging for Object-Based Image Analysis.

1) SVM Multi-Class Classification (MSVM):

- An SVM is binary classifier which accepts class labels considering only two values either 0 or 1.
- One versus the Rest: Accessing M-class classifiers, build a combination of binary classifiers M, individually prepared to differentiate single class from the other Consolidate them to collect multi-class arrangement as per maximum yield before computing the function. The value could also interpret as a confidence value. Larger the quality the more confident one is having point x belongs to the positive class.
- Train a classifier to each conceivable set of classes.
- Classify unknown point x as count how often point x is assigned toward particular class label. A class label having highest count are accepted to label for the unknown point x.

IV. RESULTS AND ANALYSIS

In this project, training of objects is carried out manually by considering some of the sample images from the directory and finding the character of each objects considering from every sample pictures. Accurate characteristics of each object from every sample images are identified and loaded with the data in a separate directory for testing. In testing procedure, whenever the function complies with any new picture, each individual object obtained from the given image is compared with the objects stored in the saved directory. In case of any new image show any sign of new instance, they are made to store in the directory for further execution. To detect three objects: satellite, blank, and others, some of the typical features are taken into consideration to identify each single object.

1) GUI Design: An interface for the identification and classification system is developed using which the system can be trained with training images of various datasets and classify the test images using MSVM classifier. Fig. 2 depicts the GUI created by using MATLAB's GUIDE toolbox.

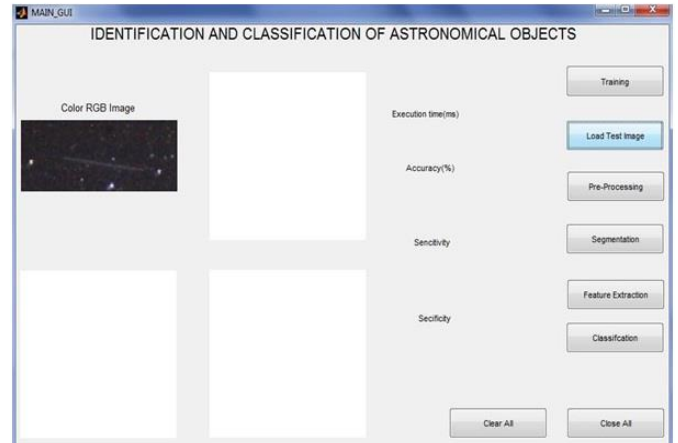


Fig. 2. Graphical User Interface

2) Pre-Processing: The input image is pre-processed by converting them into RGB to gray conversion. The gray image is then enhanced using histogram equalization. Later, extracted image will segment the image based on OTSU thresholding which will provide better representation of the image. Fig. 3 shows the snap shot of pre-processing step.

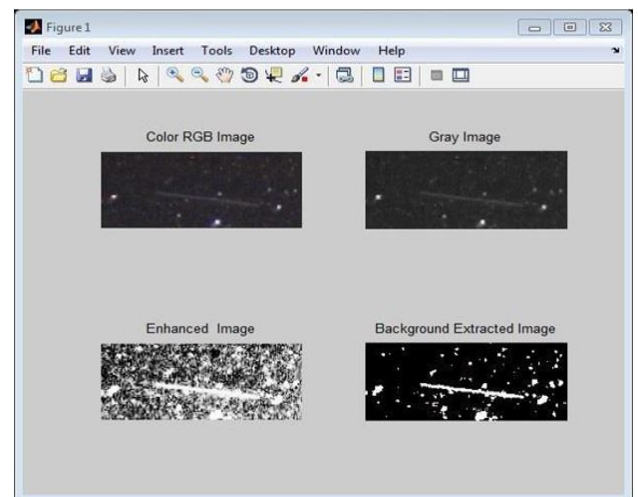


Fig. 3. Pre-Processing

3) Segmentation: Segmentation is carried out by Otsu thresholding method. The process labels the image based on black and white function which helps to label the individual objects, depicted in fig 4. It performs the segmentation on selected image after pre-processing step.

4) Feature Extraction: This is one of the most important phases. Here, spatial and texture features are calculated. GLCM is used to extract texture features. Once the features of the entire image are calculated, feature matrix is constructed whose format is understandable to the classifier. Multiclass Support Vector Machine (SVM) is used as a classifier.

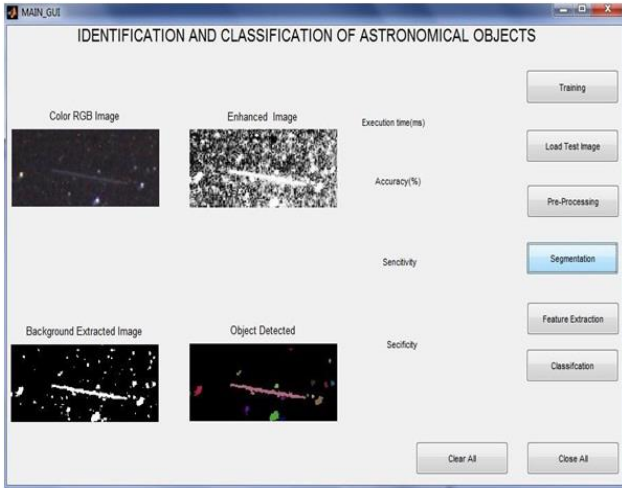


Fig. 4. Segmentation

5) Confusion matrix A confusion matrix is a table that is frequently used to portray the execution of a classification model (or "classifier") on an arrangement of test information for which the genuine qualities are known. In the area of Machine Learning, to particularly concentrate on the issue with measurable grouping, confusion matrix even called as "error matrix" permits representation of the execution of the algorithm. Every section of matrix depicts the instances of a foreseen class where every row depicts occurrences of original class (or vice-versa).

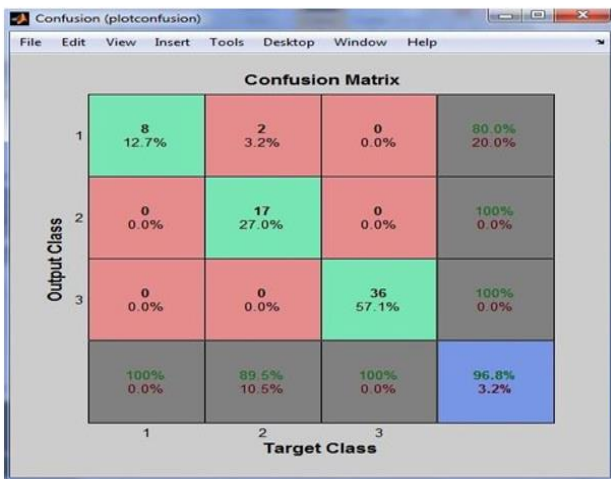


Fig. 5. Confusion Matrix

In fig. 5 columns represent a group of classification where 1 represent satellite, 2 represent a blank space, 3 represent others. Rows define the percentage of all these groups determined. The green diagonal represents the individual groups of classification like first row and column depicts the percentage of satellite-detected which is 8, second row and

column depicts the percentage of others object detected which is 17, third row and column represent the percentage of plane detected which is 36. The last row and column represent the overall accuracy percentage of all objects detected which is 96.8%. First row and second column represent the false representation of object detected which is 2. Performance is measured by determining various parameters like Sensitivity, Specificity, Accuracy, Execution time. The Multiclass SVM classifier is trained and tested for publicly available datasets. The specificity and sensitivity are computed 0.963 and 1 respectively. The system accuracy is figured utilizing which is observed to be 96.82%. Fig. 6 shows the acceptable performance obtained by the proposed work.

V. CONCLUSION AND FUTURE SCOPE

Machine learning techniques are expanding in noticeable quality in all fields and are particularly applicable to modern technology. The proposed approach is a valuable approach, which can significantly support identification and classification of astronomical objects with a little computational effort and has the following benefits:

- Develops a robust system that detects the objects in celestial bodies.
- The system helps to identify the particular objects.
- Potential for significant resource saving.

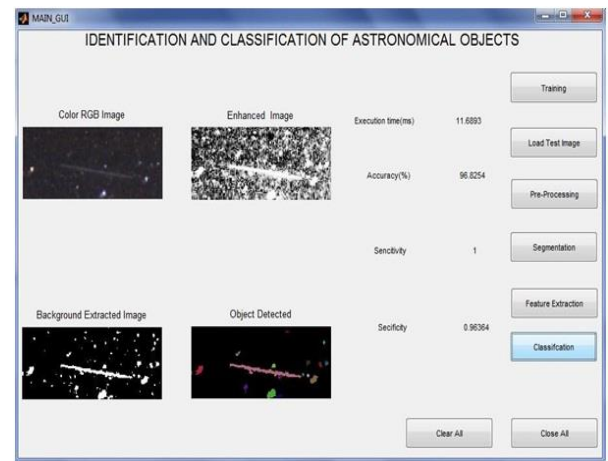


Fig. 6. Performance of the proposed work

- Automated functioning may also lead to quick decision making for scientists.
- Less time consumption

An automated system for identification and classification of astronomical objects using Machine Learning techniques is constructed to identify and classify the objects using publicly available datasets. First, pre-processing is carried out and the specific elements are determined and preceded with the OTSU thresholding methods. Spatial and texture feature is extracted from the segmented region and these extracted values are



given to MSVM classifier to differentiate the objects. Further enhancement can be defined, in view of temporal acceptance of arrangement. It depends on the way that satellites location would be in a similar direction referred in past, therefore miss classifications could be revised. The project is based on publicly available image dataset. It is not efficient for tracking the moving object or in the video set. In future, it can be used for video or real-time satellite images of continues frames and track all objects.

VI. REFERENCE

- [1] Marc Masias Moyset, (2014), "Automatic source detection in astronomical images".
- [2] Hongwen Zheng and Yanxia Zhang, (2007) "Feature selection for high dimensional data in astronomy", *Advances in Space Research, China*, (Vol. 41(12), pp. 1960-1964).
- [3] Florin Oniga, Melania Miron, Radu Danescu, and Sergiu Nedevschi, "Automatic Recognition of Low Earth Orbit Objects from Image Sequences", (2011), Computer Science Department Technical University of Cluj-Napoca, Romania.
- [4] Pavla Bromov, Petr Skoda, Jaroslav, "Classification of Spectra of Emission-Line Stars Using Machine Learning Techniques", (2014), *International Journal of Automation and Computing*, (Vol. 11, pp. 265-273).
- [5] Martin P. Lévesque, "Automatic Reacquisition of Satellite Positions by Detecting Their Expected Streaks in Astronomical Images", (2009), *Defence R&D Canada*.
- [6] Jean-Luc Starck and Fionn Murtagh, "Handbook of Astronomical Data Analysis".
- [7] G. Gordon, T. Darrell, M. Harville, J. Woodfill, (2013), "Adaptive Background Estimation and Removal based on Range and color", *IEEE*.
- [8] Giang Hoang Nguyen, Abdesselam Bouzerdoum, Son Lam Phung, "Learning Pattern Classification Tasks with Imbalanced Data Sets", (2009). In book chapter in "Pattern Recognition", (pp. 193-208).
- [9] Zhang, Y., Zhao, Y., "Support Vector Machines and Kdtree for separating Quasars from Large Survey Databases", (2008), *Chinese Journal of Astronomy and Astrophysics*, (Vol. 386 (3), pp. 1417-1425).
- [10] Zhang, Y., Zhao, Y., "A Comparison of BBN, ADTree and MLP in separating Quasars from Large Survey Catalogues", (2007), *ChJAA* 7
- [11] Saida Salahova, Azerbaijan National Aerospace Agency, "Remote Sensing and GIS Application for Earth Observation on the base of the Neural Networks in Aerospace Image Classification", (2007), 3rd International Conference on Recent Advances in Space Technologies.
- [12] M. P. Lévesque, S. Buteau, "Image processing technique for automatic detection of satellite streaks", (2007), *DRDC Valcartier*.
- [13] Mohamed Anis Loghmari, Mohamed Saber Naceur, and Mohamed Rached Boussema, "A Spectral and Spatial Source Separation of Multispectral Images", (2006), *IEEE Transactions On Geoscience And Remote Sensing*, (Vol. 44, No. 12).
- [14] Zhang, Y., Zhao, Y., Gao, D., "Decision Table for classifying point sources based on FIRST and 2MASS databases", (2007), *Avances in Space Research*, (Vol. 41(12), pp. 1949-1954).
- [15] Zhang, Y., Zhao, Y., "Automated clustering algorithms for classification of astronomical objects", (2004), *Astronomy & Astrophysics*, (Vol. 422(3)).
- [16] Y. Zhang and Y. Zhao, "Automated clustering algorithms for classification of astronomical objects", (2004), *Astronomy & Astrophysics, Chinese Academy of Sciences, Beijing, PR China*, (pp 1113-1121).
- [17] Brad Wallace, Robert (Lauchie) Scott, Aaron Spaans, "The DRDC Ottawa Space Surveillance Observatory", (2007).
- [18] Chahat Chaudhary, Mahendra Kumar Patil, "Review Of Image Enhancement Techniques Using Histogram Equalization", (2013), *International Journal of Application or Innovation in Engineering & Management*, (Vol. 2, Issue 5).
- [19] Mathew Free, Jeonghwa Lee, "Application of support vector machines to the classification of galaxy morphologies", (2013), *IEEE, International Conference on Computational and Information Sciences*.