International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125

e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

Performance Evaluation of SVM and Random Forest Algorithm for Estimation of Land Use and Land Cover: A Case Study of Pulicat Lake, India

Mule Abhi Roop[1], Dr. D. Gowri Sankar Reddy[2]

[1]Research Scholar, Department of Electronics and Communications Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India.

[2]Associate Professor, Department of Electronics and Communications Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupati, Andhra Pradesh, India.

Email ID: *abhiroop.mule@gmail.com[1], gowri.durgam@gmail.com [2]*

Abstract

Pulicat Lake is the second largest lake of India. It measures around 759 square kilometres. The lake has rich diversity in flora and fauna. It supports various ecosystems with fisheries and birds. The lakes are more prone to spatial and temporal variations. The observations of spatial and temporal changes of natural resources in lakes is vital in monitoring the ecosystem and diversity. Landsat 8 data provides abundant information for the analysis of surface water, vegetation and soil. Several algorithms evolved for the study of lakes using remote sensing data. The present study aims in the estimation of water bodies, vegetation and soil in and around the lake and to provide information for the conservation of the Pulicat lake. The Pulicat lake is considered as region of interest and the study is carried out using Support Vector Machine (SVM) classifier and Random Forest Algorithm. The performance of the Support Vector Machine (SVM) classifier and Random Forest Algorithm is evaluated using the metrics user accuracy, producer accuracy, overall accuracy and kappa coefficient. The Random Forest Algorithm gives better accuracy in classification of Pulicat lake when compared with the Support Vector Machine (SVM) classifier.

Keywords: *Landsat – 8 OLI Images, Overall Accuracy, Pulicat Lake, Random Forest Algorithm, Support Vector Machine (SVM).*

1. Introduction

Lakes and ponds are a diverse set of inland fresh water habitats that exist across the globe and provide essential resources for both terrestrial and aquatic organisms. India known for its diverse natural resources, among which Lakes are one of the major sources of natural habitat in India. India has numerous lakes like Chilika, Dal, Shivajisagar, Kolleru, Pulicat etc. The lakes can be classified in to freshwater, salt water, natural and artificial lakes. The lakes are highly affected by the global warming and hazardous human activities. [1] Pulicat is a vast coastal shallow, brackish water lagoon along the coast of Bay of Bengal. The lake has wide spread of 96% in the state of Andhra Pradesh and 3% spread with mouth in the state of Tamil Nadu. The lake is separated from the Bay of Bengal by Sriharikota Island, home to the Satish Dhawan Space Centre. The lake supports a colossal number of flora and fauna adapted to this brackish water ecosystem [19]. It is a unique Ecotone that supports rich biodiversity, from aquatic life such as mudskippers, seagrass beds, and oyster reefs to more than 200 avian species(birds), including migratory birds such as Eurasian curlews, bar-tailed godwits, sand plovers, and flamingos. Major inflows to the lake include Arani, Kalangi, and Swarnamukhi rivers. [1-2] Despite its ecological significance, Pulicat Lake is facing several threats such as pollution, both from industrial activities and domestic sewage. [3] The rapid development of aquaculture and fishing activities in the lake is also contributing to the degradation of its ecosystem. This lake has experienced an accelerated decline in water quality. [9-10] Remote sensing data provides significant information for mapping and managing of natural resources on earth. Satellite image interpretation and GIS can be used to detect and analyze spatial changes and quantify the water area in lakes. A series of Landsat satellites evolved to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

provide abundant information for remote sensing. The data provided by Landsat has given ample scope for researchers to develop different technologies in the field of remote sensing [21-22]. Landsat aims to monitor the resources on the earth. The data covers complete coverage of the earth through multi-spectral and spatial-resolution satellite images. Spectral signature plot for water, vegetation and soil is shown in figure 6. The Water only reflects in the visible light range. As water has almost no reflection in the near infrared range it is very distinct from other surfaces. The spectral signature for vegetation is very characteristic. The chlorophyll in a growing plant absorbs visible and especially blue and red light to be used in photosynthesis, whereas near infrared light is reflected very effectively as it is of no use to the plant. Therefore the reflection from vegetation in the near infrared and in the visual range of the spectrum varies considerably. The reflection from soil increases slightly from the visible to the infrared range of the spectrum. Image classification is most important part of digital image analysis. The main objective of classification process is to categorize all pixels in a digital image in to several classes. The common image classification approaches are pixel-based and object based methods.[4] The pixel-based classification uses only spectral information such as support vectors, at each pixel location and ignore the remaining spatial information in the image. The object-based approach uses different features of the objects like shape, texture and spectral values are considered for classification. Various learning based algorithms have been developed as an alternative to the traditional pixel-based and object-based approaches. The machine learning based algorithms are developed for improving the performance of satellite image processing. The most widely used machine learning algorithms are Random Forest, Boosting, K-Nearest Neighbor, Artificial Neural Networks, Support Vector Machine (SVM).[9-10] Support Vector Machine (SVM) is initially developed for binary classification which deals with only two classes, Multi class problems are carried out using multiple binary classifiers. The Random Forest algorithm is an alternative method for SVM as this algorithm is a tree based algorithm and can classify many variables and classes without using complex parameters. [6-7]

2. Materials and Methods

2.1. Study Area

The Pulicat located on the east coast is one of the 17 coastal lagoons of India. The lagoon straddles the border of Tamil Nadu and Andhra Pradesh states on the coromandal coast in South India. It is the second largest brackish water lake in India after Chilika Lake. The lake at its southern end, near north of Pulicat town opens into the Bay of Bengal by a narrow pass and this is the opening of the lake into the sea thus functioning as the migratory routes of spawning animals like fish, prawn, and mud crab [22]. The lake is situated in between the coordinates 13°33′57″N, 80°10′29″E. The lake has a length of 60 km and a breadth of 0.2 to 17.5 km. The Landsat 8 data sets with path 142 and row 51 used for the study are obtained from United States Geological Survey Earth Explorer website [25]. (Figure 1)

Table 1 Landsat – 8 OLI dataset

Acquired Image Dates	11-06-2024
Path/Row	142/51
Datum	EPSG:32644
Projection	UTM
Spatial Resolution	30mm
File Format of Acquired Images	Geo – TIFF
Total number of bands	11
Type of sensor	OLI

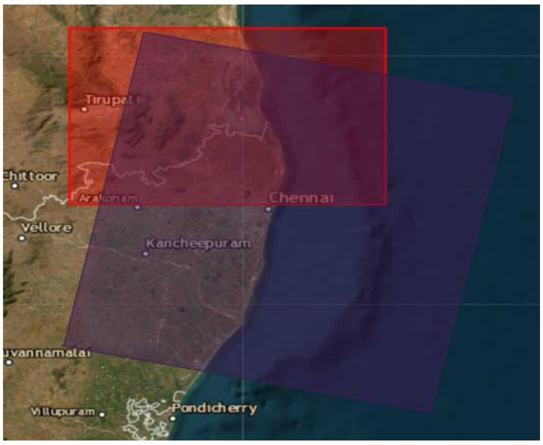

Figure 1 Location of Landsat-8 OLI AOI (11-06 - 2024) Complete Coverage of Path 142 Row 51

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

2.2. Pre- Processing of Data

Preprocessing techniques are performed to create data that is operational for analysis. The metadata file in Landsat – 8 is used for preprocessing The first step involves conversion of DN's to Top of the Atmosphere (TOA) Reflectance. After the conversion raster clipping is performed to the extent of the region of interest. The clipped images of ROI are layer stacked. The stacked image used for the classification is shown in (figure 2)

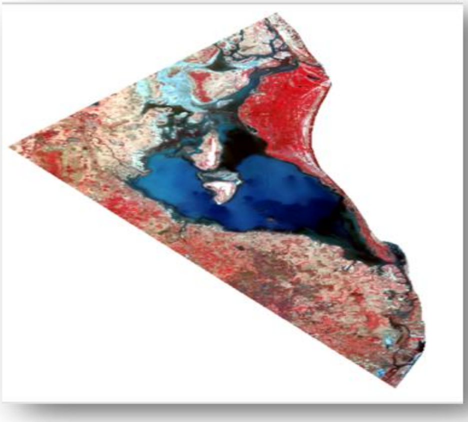

Figure 2 **Layer Stacked Image of Pulicat Lake**

2.3. Conversion of DNS to Top of Atmosphere (TOA) Reflectance

Conversion of satellite image digital number values to TOA reflectance values TOA reflectance is obtained using equation (2).

$$\rho\lambda' = M^{P} * Qcal + A^{P} \quad (1)$$

$$P\lambda = \frac{\rho\lambda'}{\sin(\theta SE)} = \frac{\rho\lambda'}{\cos(\theta SZ)}$$

Where

$\rho\lambda'$ =Top-of-Atmosphere Planetary Spectral Reflectance, without correction for solar angle.

M^{p} = Reflectance multiplicative scaling factor for the band.

A^{p} =Reflectance additive scaling factor for the band.

Qcal =Level-1 pixel value in DN.

$P\lambda$ =Top-of-Atmosphere Planetary Reflectance.

θSZ = Solar Zenith Angle.

θSE =Solar Elevation Angle.

2.4. Support Vector Machine (SVM)

The support vector machine classifier is a supervised learning method which is used for classification and obtaining solutions for regression problems. It uses a subset of training points in the decision function called support vectors. [9-10]. These are the points that are closest to the hyperplane. A separating decision line is defined based on the training data points. The main objective of the SVM is to find an optimal hyperplane that separates the data points from one class to another class. The algorithm ensures maximum margin between the support vectors. The SVM is executed in supervised mode using training data set. Radial Basis Function (RBF), class distributions with non-linear boundaries are widely used functions for obtaining the optimal decision hyperplane. This method is more effective in mapping datasets of high dimensional space to low dimensional space. The SVM training is carried out with a Gaussian RBF. The Gaussian RBF is set with a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error. A small regularization parameter tends to emphasize the margin and ignore the outliers in the training data. A large regularization parameter may over-fit the training data [1]. In this work the SVM is performed using RBF with regularization parameter set to 1.0000 and to classify the surface water, vegetation and soil.

2.5. Random Forest algorithm

Random Forest algorithm is an ensemble learning technique which is performed by constructing an army of Decision trees. It creates a number of Decision trees in the training phase. Each tree is constructed using a random subset of the data set to extract features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance. The decision trees are derived from different subsets of the given data set. This algorithm creates multiple decision trees using a training set, and then combines the predictions of each tree to produce an output. [4-6] Random forest algorithms uses three main hyper parameters, node size, number of trees and the number of features sampled. [7] The Random Forest Classifier with 5 decision trees and node size 2 is used for obtaining the classification of water, vegetation and soil in the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

Pulicat lake. The SVM Classifier and Random Forest algorithms were executed using QGIS (Quantum Geographic Information System) [23], an open source software with Semi - Automatic Classification plugin (SCP). [24] (Figure 3)

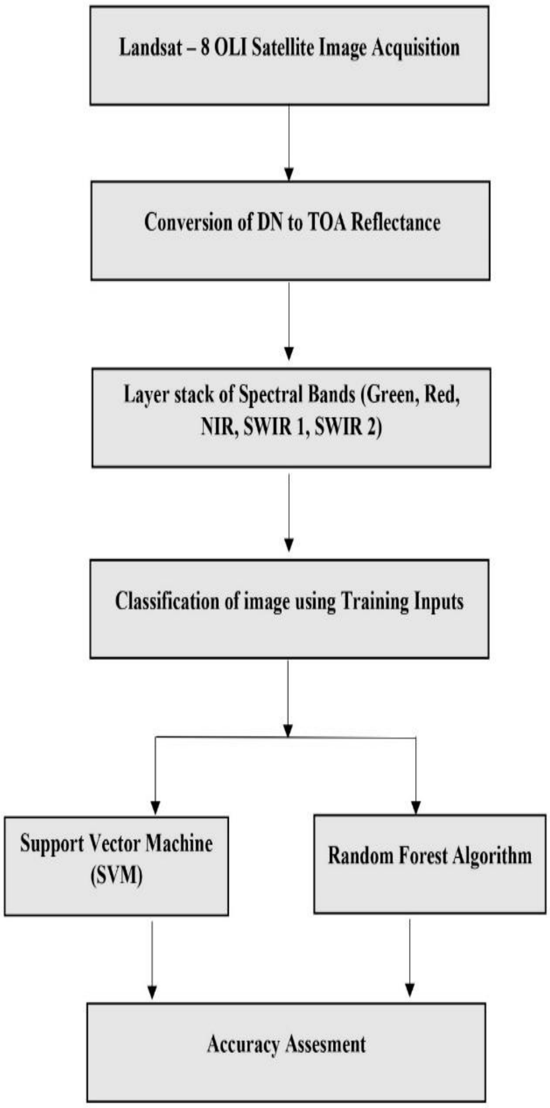

Figure 3 Flowchart for Classification of Landsat – 8 OLI Images Using Support Vector Machine (SVM) and Random Forest Algorithm

2.6. Accuracy Assessment

The performance metrics are performed using the confusion matrix for the reference and classified data as shown below.

Table 3 Confusion Matrix for Reference and Classified Data

Reference Data	**Classified Data**			
	Parameters	**Water**	**Vegetation**	**Soil**
	Water	TP	FN	FN
	Vegetation	FP	TN	FN
	Soil	FP	FN	TN

The User accuracy, Producer accuracy, overall accuracy and kappa coefficient are calculated for both the classification techniques.

2.6.1. Overall Accuracy

The overall accuracy is determined by dividing the sum of the elements along the principal diagonal by the total number of reference pixels in the confusion matrix as shown in equation. (3)

OA% =(TP+TN+TN)/ (TP+FP+FP+FN+TN+FN+FN+FN+TN) x 100%(3)

2.6.2.Kappa Co-efficient

It is a statistical measure of agreement between classification and the reference data. It is a discrete multivariate technique that is used in accuracy assessment. The calculation of kappa co-efficient is shown in equation (4)

$$KC = \frac{[(\mathbf{TS} * \mathbf{TCS}) - \sum(\mathbf{column\ total} * \mathbf{row\ total})]}{(\mathbf{TS} * \mathbf{TS}) - \sum(\mathbf{column\ total} * \mathbf{row\ total})}$$

Where, TCS=Total number of Correct Samples, TS=Total number of Samples.

If KC=1 represents a perfect agreement.
If KC>0.80-0.99 represents a near perfect agreement.
If KC=0.40-0.80 represents a moderate agreement.
If KC<0.40 represents a poor agreement.
If KC=0 represents no agreement.

3. Results and Discussion

In this study, the satellite imagery of Landsat-8 is used for estimation of water, vegetation and soil in Pulicat Lake. The work is carried out in classifying the layer stacked image of five bands (Green, Red, NIR, SWIR 1, SWIR 2). The spectral signatures of water, vegetation and soil are used as training datasets specified by assigning different Macro class ID's. The performance of the SVM and Random

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

Forest algorithms is substantiated with the validation datasets of water, vegetation and soil. The SVM classifier detected more water spread area and soil compared to Random Forest algorithm. The Random forest algorithm detected more vegetation area compared to SVM Classifier. The classified image obtained using the Support Vector Machine (SVM) is shown in Figure 4 and the classified image of Random Forest Algorithm is shown in Figure 5. The spread area of water, vegetation and soil using SVM and Random Forest Algorithm is shown in Table 3. The performance of the SVM classifier and Random forest algorithms is evaluated using the metrics shown in Table 4. The Random Forest algorithm performs better than SVM classifier in estimation of water, vegetation and soil in Pulicat Lake. (Figure 4, 5, 6)

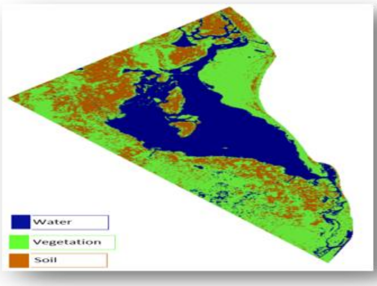

Figure 4 **Classified Image Using Support Vector Machine Classification**

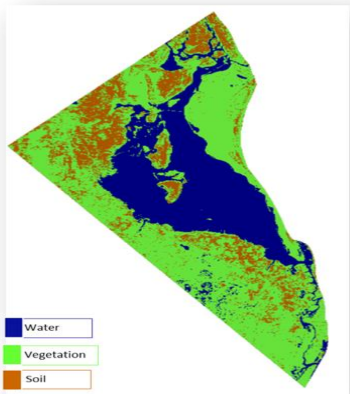

Figure 5 **Classified Image Using Random Forest Algorithm Classification**

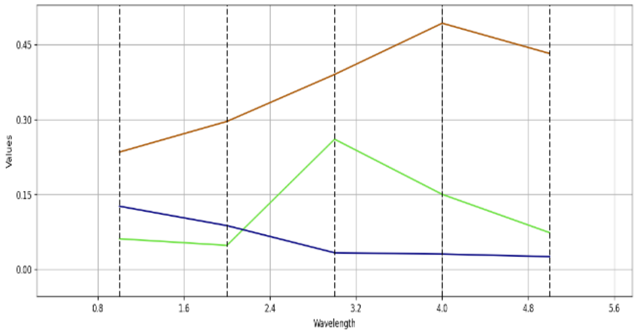

Figure 6: **Spectral Signature Plot for classified data- Water, Vegetation and Soil**

Table 4 **Summary of Classified Area**

Algorithm	Area(km^2)			Total Area (km^2)
	Water	Vegetation	Soil	
Support Vector Machine (SVM)	479.7090	728.1324	340.1739	1548.0153
Random Forest	462.4857	834.2064	251.3232	1548.0153

Table 5 **Accuracy Assessment of Support Vector Machine and Random Forest Algorithm**

Algorithm	UA%	PA%	OA%	KC
Support Vector Machine (SVM)	95.4545	100.00	97.8620	0.9666
Random Forest	97.6744	100.00	98.7468	0.9791

***UA= User's Accuracy, PA= Producer's Accuracy, OA= Overall Accuracy. KC= Kappa Coefficient**

Conclusion

In this study the vegetation, water bodies and soil of Pulicat lake are estimated using SVM classifier and Random Forest Algorithm. Accuracy assessment in terms of overall accuracy and kappa coefficient were computed and compared for SVM classifier and Random forest algorithms. The overall accuracy of Random Forest algorithm classification is 98.74% and for SVM is 97.86%.The kappa coefficient value of Random forest and SVM algorithms are 0.9791

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

and 0.9666 respectively. The result emphasizes that Random forest algorithm gives better accuracy in classification when compared to SVM algorithm.

References

[1]. Rajakumari S,Sundari Sethu, Meenambikai Manickam,Malathi Murugan,Sarunjith Kaladevi Jayadevan, "Study of pulicat lagoon on the basis of deprived vegetation and water area against increased land surface temperature", Research Square Preprint, 2023.

[2]. Gulcan Sarp, Mehmet Ozcelik, "Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey", Journal of Taibah University of Science, 2016; 11:381 – 391.

[3]. Wei.Jiang., Guojin.He.,Tengfei Long, Yuan.Ni, "Detecting Water Bodies in Landsat 8 OLI Image Using Deep Learning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences",Proceedings of ISPRS TC III Mid-term Symposium Developments, Technologies and Applications in Remote Sensing, 2018; XLII-3: 669 – 672.

[4]. Vasavi S, Venkata Kalyan Chintalapudi, Akhila Sree Rajeswari Vuppuluri., "Classification of Water Bodies Using Ensemble of U- Net and Random Forest Algorithm", Journal of Image and Graphics, 2024; 12(1): 76-89.

[5]. Komeil Rokni, Anuar Ahmad, Ali Selamat, Sharifeh Hazini, "Water Feature Extraction and Change Detection Using Multitemporal Landsat Imagery",Remote Sens, 2014; 6: 4173-4189.

[6]. Ozlem Akar, Oguz Gungor, "Classification of Multispectral Images Using Random Forest Algorithm", Journal of Geodesy and Geoinformation, 2012;1(2): 105-112.

[7]. Gaspar Albert, Seif Ammar, "Application of Random Forest Classification and Remotely Sensed Data in geological mapping on the Jeebel Meloussi area (Tunisia)", Arabian Journal of Geosciences, 2021; 14: 2240.

[8]. Hasti Shwan Abdullah, Mahmoud S. Mahdi, Hekmat M. Ibrahim, "Water Quality Assesment Models for Dokan Lake Using Landsat 8 OLI Satellite Imagery", Journal of Zankoy Sulaimani. 2017; 19-3-4(Part-A):25-42.

[9]. Fatima Hashim, Hayder Dibs, HusseinSahab Jaber, "Applying Support Vector Machine Algorithm on Multispectral Remotely Sensed Satellite Image for Geospatial Analysis", In the Proceedings of 2nd International Conference on Physics and Applied Sciences(ICPAS 2021), 2021.

[10]. Dee Shi, Xiaojun Yang, "Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery, Monitoring and Modeling of Global Changes:A Geomatics Perspective", Springer Remote Sensing/Photogrammetry, 2015; 265-279,.

[11]. S.K. McFeeters, "The use of the normalized difference water index(NDWI) in the delineation of open water features", International Journal of Remote Sensing, 1996; 17: 1425–1432.

[12]. Xu, H., "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery", International Journal of Remote Sensing, 2006; 27: 3025–3033.

[13]. Meera Gandhi.G, S. Parthiban, Nagaraj Thummalu, Christy. A, Ndvi: "Vegetation Change detection using remote sensing and gis – A case study of Vellore District". Procedia Computer Science, 2015; 15: 1199-1210.

[14]. Kim-Anh Nguyen, Yuei-An Liou, Ha-Phuong Tran, Phi-Phung Hoang, Thanh-Hung Nguyen, "Soil salinity assessment by using nearinfrared channel and Vegetation Soil Salinity Index derived from Landsat 8 OLI data: a case study in the Tra Vinh Province, Mekong Delta, Vietnam,Nguyen". Progress in Earth and Planetary Science, 2020; 7(1): 1-16.

[15]. Min, C., Ning, W.,Li, F., "Extraction of water body with different water quality

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0125
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 783-789

types based on Landsat8 image", Journal of Anhui Agricultural Sciences, 2016; 30: 220–222.
[16]. Hajigholizadeh, M. Melesse, A.M. Reddi, L, "A comprehensive review on water quality parameters estimation using remote sensing techniques", Sensors, 2016; 16: 1298.
[17]. Ahmed Mohsen, Mohamed Elshemy, Bakenaz Zeidan,, "Water quality monitoring of Lake Burullus (Egypt) using Landsat satellite imageries", Environmental Science and Pollution Research, 2021; 28:15687–15700.
[18]. Illangovan R., "Restoration of polluted Lakes-A New Approach", In the Proceedings of Tall 2007: The 12th world Lake conference, 2007; (1321-1328).
[19]. [19] Sanjeeva Raj, P.J. "Macrofauna of Pulicat Lake", National Biodiversity Authority Chennai, 2006;1-67.
[20]. Vaithiyanathan Kannan, "Vulnerable ecosystem the Pulicat lake needs government's attention", Earthy worthy, 2022; 1(1).
[21]. Thirunavukkarasu, N., Gokulakrishnan, S., Premjothi, P. V. R., &Inbaraj, R. M., "Need of coastal resource management in Pulicat Lake–challenges ahead", Indian Journal of Science and Technology, 2011; 4(3): 322-326.
[22]. Dr. R. Syamala, E Hemavathy, "Physico-Chemical Parameters And Land Use Patterns Of Pulicat Lake, Tamil Nadu, India", International Journal of Advanced Scientific and Technical Research, 2018; 6(8): 10-37
[23]. QGIS.org(2024) (Quantum Geographic Information System) a free open source software for editing and analysis of geospatial data.
[24]. Congedo, Luca, (2021). Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. Journal of Open Source Software, 6(64), 3172, https://doi.org/10.21105/joss.03172
[25]. United States Geological Survey Earth Explorer website https://earthexplorer.usgs.gov