IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.8, Issue.10, October 2020 E-ISSN: 2347-2693

Water Body Extraction from Multispectral Image Based on Spectral and Spatial Data

Devashri Jalwadi

Dept. of Computer Science and Engineering, Punyashloka Ahilyadevi Holkar Solpaur University, Solapur India

Author's Mail Id: jdevashri@rediff.com, Mobile no.+91-9881542994.

DOI: https://doi.org/10.26438/ijcse/v8i10.1116 | Available online at: www.ijcseonline.org

Received: 26/Sept/2020, Accepted: 07/Oct/2020, Published: 31/Oct/2020

***Abstract*—** Industrialization ,and urbanization lead to a change in land-use patterns and an increase in the utilization of water resources. In biogeochemical cycles, it requires good estimates of the areal extent and shape of water bodies. So timely monitoring of surface water and delivering data on the dynamics of surface water are essential for policy and decision-making processes. Change detection based on multispectral and multi-temporal remote sensing data is one of the most acceptable and ever-growing surface water change detection mechanisms in recent years. In this paper, a study has been conducted and we present an automated procedure that allows extraction of water body from a multispectral image based on its spectral data and spatial information.

***Keywords*—**Multispectral Image, Spectral Data, Spatial Data, Remote Sensing Data, Water Body Extraction, Bio Geochemical cycles.

I. INTRODUCTION

This paper is a study of a multispectral image for the extraction of water bodies based on spectral and spatial data. Inland waters play an essential role in global biogeochemical cycles. Lakes and reservoirs are important sinks of reactive nitrogen. The global annual emission of carbon dioxide from inland waters to the atmosphere is similar in magnitude to the carbon dioxide uptake by the oceans and the global burial of organic carbon in inland water sediments exceeds organic carbon sequestration on the ocean floor. These results have been obtained by extrapolating information on areal carbon processing in lakes based on indirect estimates of the total number and area of lakes on Earth.

Water body extraction is an important task in different disciplines, such as lake coastal zone management, coastline change, and erosion monitoring, flood prediction, and evaluation of water resources. Timely monitoring of surface water and delivering data on the dynamics of surface water are essential for policy and decision-making processes. In recent years, integration of remote sensing data with Geographic Information Systems (GIS) has been used in automatic or semiautomatic water body extraction and mapping Automatically extracted shorelines from Landsat TM and ETM+ multi-temporal images with sub-pixel precision techniques.

Developed an approach called the Automatic Water bodies Extraction Method that combines remote sensing and GIS to extract water bodies and study their abundance and morphometry. However, automatic coastline extraction is a complex process due to water-saturated land transition zones at the land-water boundary. To determine the spatially accurate coastline position, two methods have been explored: image classification and spectral water indexing. Multi-class support vector machine (SVM) classification for water body extraction and coastline detection has been commonly used by many researchers because it successfully minimizes errors and maximizes the geometric characteristics of edge areas. Additionally, it has shown considerable potential in the supervised classification of remotely sensed data, requiring very limited training.

However, several water-indexing methods for the extraction of water bodies from remotely sensed data have been introduced by researchers. Introduced the Normalized Difference Water Index (NDWI) to extract water features from Landsat TM using band 2 and band 4. Introduced another NDWI for water extraction from Landsat TM using bands 3 and 5. introduced the Modified Normalized Difference Water (MNDWI) for Landsat TM using bands 2 and 5. Introduced the Automated Water Extraction Index (AWEI) to improve water extraction accuracy in areas that include shadows and dark surfaces. Introduced a simple Enhanced Water Index (EWI) based on the Modified Normalized Difference Water Index (MNDWI). It can effectively distinguish water surfaces from background information such as desert, soil, and vegetation. Investigated NDWI, MNDWI, NDMI, WRI, NDVI, and AWEI for the extraction of surface water from Landsat data and used a novel surface water change detection process based on the principal components of multi-temporal NDWI.

In this study, water body extraction techniques were applied to Lake Burdur to determine decreasing trends in the lake surface area in specified time intervals. The study focuses on the performance of each satellite-derived index and SVM classification. The spectral and spatial performances of the applied satellite-derived indexes and SVM were evaluated with Pearson's r and the Structural Similarity Index Measure (SSIM). Until now, there has been no spatial performance analysis applied to satellite-derived indexes based on SSIM. Our study contributes to the effectiveness of the SSIM-based quality evaluation of satellite-derived indexes. The SSIM analysis provides a simple quantitative interpretation by comparing the correlations of luminance, contrast, and structure locally between images and averaging these quantities over the entire image.

II. RELATED WORK

Lake Burdur, which is located in SW Turkey, has shrunk abruptly in recent decades. Therefore, regular and reliable measurements of the lake area are necessary to monitor the dynamic changes of lake water area for water resource balance analysis. Previous studies of the lake area were based on visual interpretation and manual digitization of satellite data. In this study, the spatiotemporal changes of Lake Burdur from 1987 to 2011 are investigated based on SVM classification and satellite-derived water body extraction indexes, including NDWI, MNDWI, and AWEI using Landsat TM and ETM+ data. The performances of the applied indexes were tested using Pearson's r, the SSIM and the Root Mean Square Error (RMSE). Overall, the SVM and NDWI were found superior to other indexes. The approach is highly significant for time-series analyses of extracted shorelines using any number of Landsat satellite images taken in different time intervals, and it provides an important comparison that can be used to investigate shoreline changes.

Inland waters play an essential role in global biogeochemical cycles. Lakes and reservoirs are important sinks of reactive nitrogen. The global annual emission of carbon dioxide from inland waters to the atmosphere is similar in magnitude to the carbon dioxide uptake by the oceans, and the global burial of organic carbon in inland water sediments exceeds organic carbon sequestration on the ocean floor. These results have been obtained by extrapolating information on areal carbon processing in lakes based on indirect estimates of the total number and area of lakes on Earth. The abundance of lakes in large regions is difficult to estimate due to poor or incomplete lake inventories in many parts of the world. Early estimates on the global abundance of lakes suggested that 1.8% of the non- oceanic area is covered by lakes. Different approaches have been proposed to achieve more accurate estimates. Based on the Pareto distribution, Downing et al. (2006) showed that the global extent of lakes is twice as high as previously thought (304 million lakes, 4.2 million km2, covering > 3% of continents), and that small lakes represent a substantial lake area previously not accounted for. However, uncertainty in these estimates. These uncertainties call for methods that allow the direct mapping of lake abundance with greater accuracy.

Satellite remote sensing is the only practical way to determine the spatial and temporal patterns of inland water globally. Because the size distribution of lakes is globally dominated by small lakes, and the greatest uncertainties in current statistical methods applicable to them, high spatial resolution imagery is required. Several satellites provide data with a spatial resolution of 2 m or better. Unfortunately, these satellites do not provide full global coverage. The best data currently available for mapping lakes globally is the GeoCoverTM mosaics of the Landsat imagery covering all continents with 14.25 m spatial resolution and minimal cloud coverage. Several algorithms and techniques have been proposed for retrieving water bodies from remote sensing data, but the scope has so far been restricted to a few lakes or at most the regional scale.

These methods include digitizing through visual investigation, thresholding, edge detection using a single or a combination of multiple bands and algebraic operations (e.g. band ratio, spectral water indexes), classification techniques, spectral transformation, and texture analysis. The threshold method is considered popular for delineating water bodies because it is easy to use and less computationally time- consuming than alternatives approaches. This method is based on threshold values of the band intensity, which spatially corresponds to the land-water interface. Usually, the threshold values derive from histogram analysis of the image for one single band. However, separating water bodies from some other land cover types based on a single threshold in a single unique channel is frequently problematic. Moreover, since the optical properties of water are highly variable in space and time, the analysis cannot be restricted to one standard cut-off value. Consequently, threshold computations should use a range of threshold values and employ multiple band analysis. Identification of inland water can be improved by using specific spectral indices or rationing, where for each pixel the Digital Number (DN) value of one band is divided by the value of another band. This can be useful to reduce or eliminate the effect of shadowing, particularly dark shadows.

Unfortunately, in our context, most of the common spectral indices cannot be applied because Landsat 7 ETM + does not include the relevant bands (Table). Both supervised and unsupervised classification procedures are frequently applied for identifying and classifying water features in images. A supervised maximum-likelihood classification was used to map wetlands on Landsat MSS imagery.

The number of classes and the spectral signature attributed to each class in the scene. Unsupervised classification based on “iso data” (Iterative Self Organizing Data Analysis Technique) or “k-mean” clustering is often used to generate spectral signatures of each class. Water pixels can also be separated from land pixels by morphological segmentation. More advanced approaches use both morphological segmentation and spectral thresholding to identify water bodies. As described above, several techniques are available for lake extraction. Some of the methods cannot be automated, and some cannot be applied to large regions. Some approaches are complicated due to the nature of the data used. However, there is no previous method that allows automated lake detection over large geographic regions.

III. METHODOLOGY

We present an approach, GWEM (GeoCoverTM Water Bodies Extraction Method), that combines all the methods described above to eliminate drawbacks of each technique in achieving a robust method that performs well under different circumstances. The method allows automatic extraction of water bodies from GeoCover, and we evaluated the accuracy by comparing the output with accurate lake maps available for Sweden.

A. Water Balance Approach

Water balance approach involves applying the water balance equation to the catchment area of interest over some time T and solving the equation for evapotranspiration, ET as given in Equation,

$$ET = P\ Qin + Gin - Qout - Gout - \Delta\ S$$

where, P is Precipitation, in Q is the inflow of surface water, out Q is the outflow of surface water, in G is groundwater inflow, out G, are groundwater outflow and S changes in the amount of water stored over the time assuming a long-term negligible change in storage. The amount of infiltration of groundwater depends on soils, water table depth, rock layers, surface disturbance, the presence or absence of a liner in the pond, and other factors. The infiltration rate is governed by the Darcy equation. The dimensions of these quantities are L3 or if divided by drainage area, L.

Even as the approach looks simple in concept, it is difficult in practice to measure the true values of the components in Equation. If reasonably accurate information on the balance components is available, the method can provide an accurate estimation of evapotranspiration.

B. Energy Balance Approach

At a land surface, the energy inputs and outputs are balanced according to the energy conservation law. The components of the energy balance can be calculated and the energy available for actual evapotranspiration can be solved by the energy balance equation given below

$$Rn = H + \lambda EE + G$$

where n R is net radiation, H is sensible heat flux from the surface, G is the soil heat flux and E is latent heat flux. ET is estimated as the residual of the land surface energy balance if all other terms are observed and estimated. The units for these terms are commonly W/m2 (1mm of ET per day =28.36 W/m2).

C. Materials and procedures GeoCoverTM Circa 2000

The GeoCoverTM Circa 2000 product is built from the imagery of the Enhanced Thematic Mapper Plus (ETM +) sensor onboard the Landsat 7 satellite. Geo-location information was provided by the National Geospatial-Intelligence Agency (NGA) and the U.S. Geological Survey (USGS). The GeoCover Circa 2000 archive encompasses 8500 scenes (Tucker et al. 2004), which are the basis of the 862 mosaics built from mostly cloud-free images collected in the year 2000 ± 3 years. However, some persistent cloud contamination exists, especially in mountain regions. There are also some radiometric differences in the mosaics because the images were not acquired simultaneously. Stars indicate the three spectral bands (2, 4, 7) in blue, green, and red (BGR) used by GeoCoverTM and sharpened to the panchromatic band spatial resolution (14.25 m).

Table 1. Spectral and spatial characteristics of the Landsat 7 ETM + bands.

Types of Bands	Landsat 7 ETM+		
	Band No	***Wavelength***	***Spatial resolution (m)***
Panchromatic		0.52 – 0.90	14.25
visible	1	0.45 – 0.52	30
	2	0.52 – 0.60	30
	3	0.63 – 0.69	30
NIR	4	0.76 – 0.90	30
SWR	5	1.55 – 1.75	30
	6	2.08 – 2.35	30
TIR	7	10.04 – 12.05	120

The GeoCover mosaic is geo-referenced using Universal Transverse Mercator (UTM) projection and World Geodetic System 1984 (WGS84) datum and ellipsoid. Three (ETM +2, ETM + 4, ETM + 7) of the Landat7 ETM + bands in the visible and near-infrared regions are included (Table 1). Originally, each of the spectral band providing a 30 m spatial resolution was sharpened with the panchromatic band (0.52-0.9 μm) following the cubic-convolution process offering am) following the cubic-convolution process offering a pixel size of 14.25 m, and the Root means square error (RMSE) of the final product is better than 50 m in positional accuracy.

D. The Test Site and Reference Data Set

We validated the lake extraction method against accurately mapped data for all of Sweden. Sweden is a good test area because accurate maps of water bodies are available, and the landscape has a large variety of surfaces, ranging from undeveloped forests and mountains to cities, and also a wide range of optical water properties of lakes. The lakes are both optically deep and shallow and range from oligotrophic to hypertrophic. Lakes with a high concentration of colored dissolved organic matter (CDOM) are abundant, but there are also lakes with high concentrations of phytoplankton and suspended sediment. Seven GeoCover mosaics were put together to cover Sweden (~450,000 km2). The neighboring countries (Norway, Denmark, and Finland) were cut off.

To evaluate the accuracy of the developed remote sensing approach, we compared our results to an independent dataset with high spatial resolution, called ViVaN (Virtuellt Vattendrags Nätverk, "virtual watercourse network"). The ViVaN data set was built from several public archives and maps (SMHI 2008; Lantmäteriet 1998). The ViVaN data were imported into a GIS database encompassing a total of 254,111 lakes. According to the ViVaN data set, Swedish lakes cover ~38,465 km2, which represents approximately 9% of the country, and there are 83,059 lakes larger than 1 ha (with a total surface of 37,912 km2). Although there are 171,052 lakes smaller than 1 ha, the latter only represents a total surface area of ~553 km2. We mainly focus on lakes greater than 1 ha because this is the reported lower limit of the database, and since our visual comparison with GeoCover data shows frequent inaccuracies in the ViVaN database for lakes smaller than 1 ha.

IV. RESULTS AND DISCUSSION

The overall methodology (GWEM) can be summarized into these steps: (A) Thresholding and Classification; (B) Texture Analysis (C) Vectorization, and (D) shadow removal.

A. Thresholding and classification

Image processing was performed in ENVI v.4.8 (ITT Vis) software and model developments were generated using procedures written in Interactive Data Language (IDL, Resources Systems). We also used ArcGIS v 10 (ESRI) to create and analyze the lake database made from Landsat mosaics. Although most water bodies are substantially darker than surrounding land, there are cases where the water leaving signal is high, e.g., shallow water areas with bright bottoms, submerged vegetation close to the water surface, aquatic vegetation with floating leaves, strong algal blooms, and high turbidity caused by suspended mineral particles. Statistical analysis of several test sites and validation against aerial photos revealed that simple conventional thresholding method based on the single Digital Number (DN) values of original GeoCover bands (ETM + 2, ETM + 4, ETM + 7), would not allow automatic extraction of water bodies over large areas. Therefore, we developed an automated procedure that employs multiple thresholds, generating various DN magnitudes and spectral shapes of the original bands for each of the resulting classes.

Moreover, it includes additional threshold values deriving directly from Principal Component Analysis (PCA) transformation and from brightness index calculation. Principal Component Analysis (PCA) was performed on the three ETM + bands to enhance water detection. To discern the water bodies from the other land categories, we restricted the thresholding analysis to the first Principal Component (PC1) which explains the major variability in the image and contains the overall scene brightness variation shared by all the input bands.

Fig. 1

The recognition of water pixels was further improved by adding criteria from spectral brightness calculation. Low brightness is associated with wet, dark- colored, rough surfaces whereas higher brightness is associated with dry, light-colored, smooth surfaces. The spectral brightness of lakes depends on water quality. Brightness is also sensitive to the state of the sky and inparticular the type and amount of cloud and the sun angle. Here, the overall brightness is defined as a simple ratio over the entire spectral range. The Modified Brightness Index (MBI) index can be expressed as:

$$\mathrm{MBI} = \frac{(\mathrm{DN}_{\mathrm{ETM+2}})^2+(\mathrm{DN}_{\mathrm{ETM+4}})^2+(\mathrm{DN}_{\mathrm{ETM+7}})^2}{3}$$

where DNETM+2, DNETM+4, DNETM+7 are the Digital Numbers resulting from GeoCover spectral bands. MBI was calculated for each pixel of the mosaic image by applying the equation.

B. Texture Analysis

There are water bodies that are just one or a few Landsat pixels in size (i.e., one or a few units of 14.25 × 14.25 m). However, some of these small objects may be random image noise. Therefore, only objects larger than 10 pixels were considered as water bodies. A simple low pass filter with kernel size 3 × 3 pixels was applied to eliminate small objects. As a result of filtering, any object smaller than 0.1827 ha (0.00187km2) was automatically removed from the data. Segmentation was applied after noise removal, to group connected water pixels into one single water body.

C. Texture Analysis

Vectorization and water boundary delineation To enable calculation of surface area, perimeter, and shape of water bodies, the water-pixel groups resulting from the previous processing steps were converted to a polygon vector format with smoothing to remove pixel corner effects in ArcMap. Multi-pixels boundaries (water body) were automatically digitized into shoreline vector geometries. To assess the morphometry of lakes, we employed several previously developed morphometry indices. The Shoreline Development Index is the ratio of water body boundary (the perimeter, P) to the circumference of a circle whose area (A) is equal to that of the given water body Equation,

$$\mathrm{SDI} = \frac{\mathrm{P}}{\sqrt[2]{\pi \mathrm{A}}} \in ¿$$

For a perfectly circular lake SDI = 1, and it increases with the number of inlets, bays, and islands. The circularity can also be calculated by the thickness index Miller Equation.

$$\mathrm{Thickness} = 4\pi \frac{\mathrm{A}}{\mathrm{p}^2} \epsilon ¿ \, 0,1 ¿ ¿$$

To take into account that the perimeter is the scale-dependent quantity (Roche 1963, Kent and Wong 1982), some morphometric indices are based on the length (L) of the equivalent rectangle which encompasses the water body. The compacity index Eq. , expresses the ratio of the square of the water body length (L) to the water body area.

$$\mathrm{Compacity} = \frac{\mathrm{L}^2}{\mathrm{A}} \in ¿ \, 0, \infty ¿$$

The spreading ratio or Morton Index Eq. characterizes how the shape of the water bodies deviates from a circle.

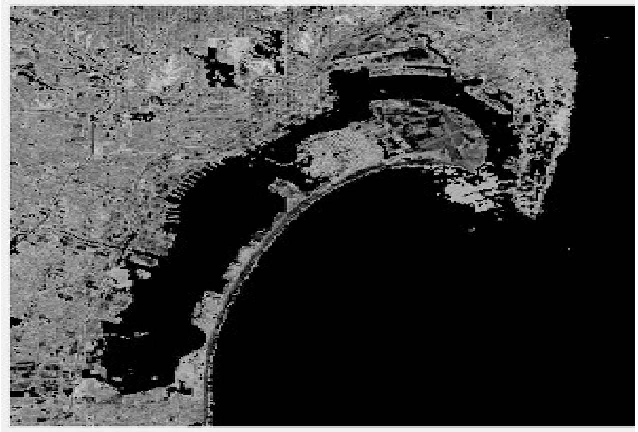
$$\mathrm{Spreading} = \frac{4}{\pi} \frac{4}{\mathrm{L}^2} \in ¿ \, 0,1 ¿ ¿$$

D. Shadow removal

Shadowing of the ground by mountains and clouds can generate misclassification of water bodies. A Digital Elevation Model (DEM) with 50 m cell resolution was merged with the remotely sensed data. Mountain shadow surfaces were modeled from the hill shade algorithm developed in ArcMap. We verified visually that there was a good match between shadow zones from DEM and the shadows observed from the RGB color composite of GeoCover. If mountain shadow zones produced from the DEM overlapped with a detected water body, the water body was deleted from the data set. The DEM was also used to reduce the number of cloud shadows in the GeoCover data. This was done in three steps:

i. Detection of clouds edge pixels.
ii. Identification of isolated water bodies not connected to the river network.
iii. Removal of water bodies that intersect with cloud structures.

The first step consisted of detecting bright cloud boundaries based on their typical high radiance values. The clouds were easy to separate from the rest of the scene by using the high brightness values (DN) in optical band 2 (blue) by visually inspecting the range of cloud types and the digital numbers (DN's). Note that here we only considered the very bright clouds, since clouds may also be misclassified as other landscape surfaces such as ice cover or a certain type of rocks (e.g., calcareous sand). In the second step, based on the digital elevation model (DEM) analysis in the ArcGIS hydrology tools, we designed the drainage pattern of Swedish catchments. Geo-processing was performed to create a depression less DEM and generate data on flow direction, flow accumulation, stream definition, stream segmentation, and watershed delineation. The resulting data were then used to spatially distinguish water bodies, which are connected to the river network from isolated lakes. If cloud zones produced from thresholding overlapped with an isolated lake, it was deleted from the data set. This step was done only on the isolated lake map to avoid true lake removal. Cloud shadow miss classifications may remain in the data if they are connected to the river networks.



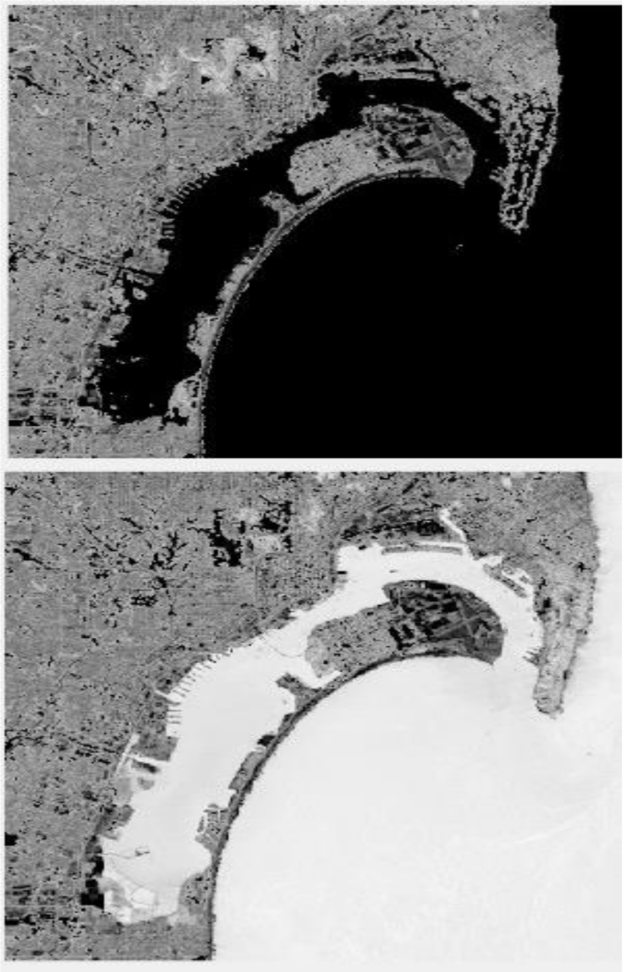


Fig. 2

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have studied many different methods of water body extraction from the multispectral image but we get the best and nearly accurate results with GeoCover Water Bodies Extraction Method (GWEM). A major benefit of GWEM is that it is simple and straightforward, providing a wealth of information for water body retrieval at relatively low cost and independently of previous mapping other than GeoCover.

Our method automatically derives water bodies by a straightforward stepwise procedure built on multi-thresholding decision making, including the original spectral band values, principal component analysis, and MBI calculation. This allowed good separation of water pixels from all other surfaces. Using multiple satellite images for different purposes usually requires atmospheric correction of the imagery.

The GeoCover data are put together from multitude of Landsat images acquired at different times, and it does not contain enough spectral bands to perform atmospheric correction. However, the GeoCover CircaTM 2000 data are relatively homogeneous in the sense that atmospheric correction has a minor effect on classification accuracy.

REFERENCES

[1] Alsdorf, D., D. Lettenmaier, and C. Vörösmarty. 2003.*"The need for global, satellite-based observations of terrestrial surfaces waters".* Eos, Transactions, American Geophysical Union, Vol.**8**, Issue.**12**, pp.**269-277**, **2013.**

[2] Bagli, S., and P. Soille. *"Automatic delineation of shoreline and lake boundaries from Landsat satellite images".* In Proceedings of initial ECO-IMAGINE GI and GIS for Integrated Coastal Management, Seville **13-15 May 2014.**

[3] Battin, T. J., S. Luyssaert, L. A. Kaplan, A. K. Adenkampeuf, A.Richter, and L. J. Tranvik. 2009. *"The boundless carboncycle. Nat. Geosci"* Nature Geosciences', vol.**2** pp.**598-600**, **2009.**

[4] Bennett, M. W. A. *"Rapid monitoring of wetland surface status using density slicing"* In Proceedings of the 4th Australasian Remote Sensing Conference, Adelaide **14-18** Sep **1987.**

[5] Castañeda, C., J. Herrero, and M. A. Casterad. *"Landsat monitoring of playa-lakes in the Spanish Mongers desert"* Journal of Arid Environ. **Vol.63, pp.497-516, 2005.**

[6] Choi, H., and R. Bindschalder. *"Cloud detection in Landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision. Remote Sens. Environ."*vol.**91 Issue.2 pp.237-242 2004.**

[7] Cole, J. J., and others. *"Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. Ecosystems"* vol.**10, Issue.1 pp.171-184, 2007.**

[8] Foody,G.*"Status of land cover classification accuracy assessment. Remote Sens. Environ."* vol.**80 Issue.1 pp.185-201 2002.**

[9] Frazier, P. S., and K. J. Page. *"Water body detection and delineation with Landsat TM data"*Photogram. Engineering and Remote Sensing. Vol.**66 Issue.12 pp.1461-1467 2, 2000.**

AUTHORS PROFILE

Miss.D.D.Jalwadi pursed Bachelor of Engineering in Computer Science . She is currently pursuing M.E in Computer Science . She has attended workshops on Java Technology and Latex .She has teaching experience of 5 years 2014-2019.