IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311067|

Automated Detection of Water Bodies in Agricultural Soils using Deep Learning

Jayashree .D. Bhoj[1], Anmol S. Budhewar[2], Harshal Chavan[3], Sarthak Bukane[4], Sagar Pawshe[5], Kunal Kokate[6]

Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

Associate Professor, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

ABSTRACT: Water body detection plays a critical role in agricultural management, influencing irrigation practices, water resource distribution, and overall crop productivity. Traditional methods for water body identification, which rely heavily on manual interpretation or basic spectral indices such as the Normalized Difference Water Index (NDWI), face significant challenges in complex agricultural landscapes due to the presence of mixed land cover types, including wet soils and vegetation. This paper presents an automated approach for detecting water bodies in agricultural regions using remote sensing data and advanced machine learning techniques. The study leverages multi-spectral satellite imagery from sources such as Sentinel-2 and Landsat, which offer a balance between spatial and temporal resolution, making them suitable for monitoring water dynamics in agricultural settings. The proposed methodology involves a comprehensive image preprocessing pipeline, including atmospheric correction, cloud masking, and enhancement of spectral features to improve the visibility of water bodies. Various spectral indices are employed to extract relevant features, while deep learning-based models, specifically Convolutional Neural Networks (CNNs), are utilized for automated feature extraction and classification. The study compares the performance of different machine learning algorithms, including Support Vector Machines (SVM), Random Forest, and CNNs, to classify water and non-water regions, with a particular focus on distinguishing water bodies from wet soils and irrigation channels that frequently occur in agricultural areas. Experimental results demonstrate the effectiveness of the machine learning-based approach in improving the accuracy and automation of water body detection compared to traditional techniques. The proposed method achieves high precision and recall rates, demonstrating its potential for real-time monitoring and large-scale assessment of water resources in agricultural landscapes. Validation is conducted using ground truth data and high-resolution imagery, ensuring the robustness of the results across different seasons and regions. Additionally, this study addresses key challenges such as cloud cover and mixed land cover by integrating multi-source data and employing region-specific model fine-tuning. The findings of this research contribute to the advancement of precision agriculture by providing a reliable tool for monitoring water bodies, thereby enabling more efficient water management and sustainable agricultural practices. The paper also discusses future directions, including the integration of radar-based remote sensing for cloud-affected regions, the application of advanced deep learning models for pixel-level segmentation, and the potential for near real-time implementation using cloud computing platforms.

KEYWORDS: Waterbody detection , Remote sensing, Agricultural soils

I. INTRODUCTION

Water is an important component in agricultural activities affecting the health of soil and crop growth and in turn, agricultural productivity. Appropriate management of water resources is key to sustainable agriculture in areas that face water scarcity or seasonal water availability on agricultural land. It is very important to monitor water bodies like rivers, lakes, irrigation canals, and even the temporary accumulation of water, for managing irrigation practices, the planning of cropping cycles, and assessing impacts from drought or heavy rainfall. All traditional methods of identification and monitoring of water bodies depend on field surveys and manual interpretation of aerial images, which are timeconsuming as well as labor-intensive, and thus often full of inaccuracies. Such complexity must be captured with more efficient, automated techniques. Remote sensing could also be employed as a tool for detecting water bodies and monitoring the environment in large spatial extents. It employs high spatial resolution satellites such as Landsat, Sentinel-2, and MODIS, which provide valuable multi-spectral data to use for categorizing water bodies according to their spectral properties. As water reflects differently from the soil and vegetation, a specific spectral

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311067|

indices such as Normalized Difference Water Index (NDWI) and Modified Normalized Difference Water Index (MNDWI) have been used for automatic detection of water presence within the satellite images. This, however, makes it challenging to detect water bodies in agricultural areas because wet soils, vegetation, irrigation channels, and other landscape features generally produce spectral signatures similar to that of water. This may cause misclassification or inaccuracies in the traditional techniques of water detection. Machine learning is highly promising for the advancement of applications of remote sensing because it offers the classifying and pattern recognition processes in complex datasets without human intervention. Machine learning algorithms can process large sets of satellite imagery, detect the minute differences that distinguish water from other types of surface features, and enhance the accuracy of the detection of water bodies. Techniques such as SVM, Random Forest, and CNNs can be applied in building classification models that differentiate water areas from the rest of the agricultural landscape. Mainly, CNNs have gained wide popularity because they can learn to extract relevant features autonomously from the raw data and can be applied in complex tasks like image classification and segmentation

II. STUDY AREAS

The selection of study area is an integral part of deciding the test performance and generality of an automated method of water body detection in agricultural landscapes. A few agricultural regions characterised by diverse geographic, climatic, and land use conditions were selected to represent different scenarios where the need for monitoring water resources arises for sustainable agriculture. These areas cover the region of a mix of irrigated and rain-fed lands with widely contrasting regional seasonal variations in water availability and 1 6 distribution. It covers areas with diverse landscape characteristics, such as fields with different crops at times of varying soil moistures and water-related infrastructure such as irrigation canals and reservoirs. The diversity allows testing whether the model can distinguish between real water bodies and other land cover types like moist soils or recently irrigated fields that can have a spectral appearance similar to that of real water. A combination of geographical diversity is ensured by choosing the study area in two ways: areas with different environmental conditions, such as arid and semi-arid areas, and areas with distinct climatic, geological, or soil conditions. Geographical variation offers the possibility of studying applicability under various climatic settings of the proposed water detection approach in the context of agriculture. In addition, the areas studied cover various agricultural uses such as intensive irrigation, mixed cropping, and dryland farming in order to capture the variability in water use and management. There is also seasonal variability, an important factor when these areas are selected to reflect the different agricultural calendar stages, like growing season, dry season, and rainy season. The approach of this methodology will be tested under different conditions of water availability such as during drought or after the rainfalls that affect the visibility in satellite imagery. Access to ground truth data or high-resolution imagery is used, and availability of this is only considered feasible while seeking regions for this research when accurate validation of water body detection results is sought. Additional validation, perhaps with ground truth data in the form of field surveys, existing water maps, or other reference data, would be useful for validating the machine learning models as to whether they are accurate enough and needing to be adjusted in some instances. Research could be conducted over regions like the Nile Delta in Egypt, the Indo-Gangetic Plains in India, the Great Plains in the United States, or the Murray-Darling Basin in Australia. Each of these regions presents special challenges for water-body detection: intense irrigation, seasonal flooding, extremely variable soil moisture, and complex landcover patterns. This research seeks to develop a robust and flexible model for the detection of water bodies in agricultural settings by exploiting the diversity of representative study areas, which could be applied more broadly in precision agriculture and waterresource management over diverse landscapes.

III.DATASET DESCRIPTION

The dataset was created using Landsat-8 and SMAP global soil moisture images. These images are in the form of .tif files which are high resolution images consisting of 11 different bands and each band has its individual .tif file. The pixels of each band depict some useful value which has been used to calculate the required input variables and to create the dataset. Data of three different SMI (Soil moisture index) areas is extracted from Landsat-8. The resulting facts were then converted into 12 input variables by using various formulae mentioned within the section IV.B. NDVI, SAVI, NDMI, MNDWI, MSI, MSAVI, MSI, NIR, TIRS1, TIRS2, SWIR1, SWIR2 are calculated using each pixel of the required area for which the .tif images are taken from the Landsat-8 satellite. These indices act as input variables for the model. Input variables along with Soil moisture index which acts as an output variable, The output variable i.e., SMI was extracted from SMAP global soil moisture images for the same area and same time frame. The SMI taken from SMAP combined with the other 12 input variables combined formed the dataset. These data are taken from three

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311067|

different areas whose SMI's are close to 0,0.5 and 1 respectively. These areas are one from a place in Sahara where the SMI is close to 0 and prone to drought, second from a metropolitan area in Maharashtra (India) where the SMI is close to 0.5 with less chances of drought and third from a rainforest in Amazon where the SMI is close to 1 with no chances of drought. This helps to make a better prediction model as there are all types of data present. For water quality, NDWI is considered to detect the change in contents of water bodies and NDCI to calculate the amount of chlorophyll-a present in the water bodies. A. Required bands as input: Band 2: Blue pixels .tif image from Landsat 8 with Wavelength 0.45-0.51 micrometers and Resolution 30 meters. Band 3: Green pixels .tif image from Landsat 8 with Wavelength 0.53-0.59 micrometers and Resolution 30 meters. 7 Band 4: Red pixels .tif image from Landsat 8 with Wavelength 0.64-0.67 micrometers and Resolution 30 meters. Band 5: Near Infrared pixels .tif image from Landsat 8 with Wavelength 0.85-0.88 micrometres and Resolution 30 meters. Band 6: SWIR1 pixels .tif image from Landsat 8 with Wavelength 1.57-1.65 micrometers and Resolution 30 meters. Band 7: SWIR2 pixels .tif image from Landsat 8 with Wavelength 2.11-2.29 micrometers and Resolution 30 meters. Band 10- Thermal Infrared 1 pixels .tif image from Landsat 8 with Wavelength 10.6-11.19 micrometers and Resolution 100 meters. Band 11- Thermal Infrared 2 pixels .tif image from Landsat 8 with Wavelength 11.50-12.51 micrometers and Resolution 100 meters. Red Edge- Red edge bands of Sentinel-2 are used to get the state of the vegetation and lushness with a resolution of 20 meters B. Variables NDVI: The normalized difference vegetation index helps in calculating the presence of green vegetation. NDVI = (NIR – Red) / (NIR + Red) (1) SAVI: Soil-Adjusted Vegetation Index consists of band 4 and band 3 used for land surface reflectance SAVI = ((Band 4 – Band 3) / (Band 4 + Band 3 + 0.5)) * (1.5) (2) EVI: Enhanced vegetation index is used for monitoring changes in atmosphere and components of soil especially in dense forests EVI = 2.5 * ((Band 4 – Band 3) / (Band 4 + 6 * Band 3 – 7.5 * Band 1 + 1)) (3) NDMI: Normalized difference moisture index is used to calculate vegetation of contents in water. The formula of NDMI is the normalized difference between the NIR band and SWIR1 band. NDMI= (Band 5 – Band 6) / (Band 5 + Band 6) (4) MNDWI: MNDWI is Modified normalized difference water index and is used to analyse components present in water. MNDWI = (Green – SWIR) / (Green + SWIR) (5) MSAVI: Modified Soil-Adjusted Vegetation Index consists of red band and Near Infrared Band along with Band 3. It helps in reducing the effects on vegetation through soil MSAVI=(2*Band 4 +1 –sqrt ((2*Band 4+ 1)2 – 8 * (Band 4 – Band 3))) / 2 (6) MSI: The Moisture Stress Index is used to analyse drought impact. MSI=SWIR1/NIR (7) NIR: near infrared band used for healthy and unhealthy vegetation. NIR=Band5 (8) SWIR1: The result is like that of a basic photography with infrared colours but has an excellent clarity. SWIR1=Band6 (9) SWIR2: This band helps in the improvisation of soil moisture content by penetrating thin clouds SWIR2=Band7 (10) TIRS1: TIRS1 or Band 10 provides efficient land temperature TIRS1=Band10 (11) TIRS2: The Thermal Infrared Sensor 2 (TIRS-2) measures thermal radiance emitted from the Earth's land surface. TIRS2=Band11 (12) NDWI: Normalized Difference Water Index detects the changes that take place in water bodies like leaves and aquatic life. NDWI=NIR-SWIR1/NIR+SWIR2 (13) NDCI: Normalized Difference Chlorophyll Index is used to analyse chlorophyll-a concentration in water. NDCI=Red Edge-Red/Red Edge+Red (14) SMI: Soil moisture index is the amount of water content present in the soil. This dataset is taken from Sentinel-1 and SMAP.

IV.IMPLEMENTATION

Water Body Detection Model Implementation It is through the use of the Random Forest algorithm in machine learning where satellite imagery data is classified as regions of agricultural land water and nonwater that a water body system is detected. The essential objective is to classify correctly the extent of water bodies to facilitate the effective use of agricultural water. A. Reason for Selecting Random AREA Random AREA has been selected for the analysis due to the fact that there are multiple features which have been extracted from the satellite images which include the spectral band values, the computed water indices, including the NDWI and MNDWI, as well 8 as several texture features. These automatically create a highdimensional dataset. Random Forest can handle complex datasets very well. It is also very efficient computationally and works effectively with large volumes of data with robust results without expensive computational resources. B. Dataset Preparation and Feature Extraction The sources of data collection for this model come from multispectral satellite imagery sources like Sentinel-2 or Landsat-8 covering the agricultural regions with different kinds of water conditions. The nature of the imagery consists of a large amount of pixels, where it works as a data point, while depending on the situation, various features can be extracted that include spectral bands with reflectance values along the different spectral bands ranging from visible to near-infrared. Water Indices: Computed indices in an image like NDWI, Normalized Difference Water Index, and MNDWI, Modified NDWI for the extraction of water features. Texture Features: Obtained metrics from the use of image texture analysis methods in GLCM, GrayLevel Co-occurrence Matrix to allow for distinction of water features from apparently wet soils. The dataset is approximately 1,166,400 samples categorized into 80% training samples and 20% testing samples for excellent

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311067|

generalization by the model. C. Training the Random Forest Model The Random Forest algorithm is used to train the model. This method constructs a collection of decision trees during training. Each tree votes on the classification of each pixel in the image as either water or non-water. The only real hyperparameter to tune is the `max_depth` in case of decision trees. If the depth is larger, then presumably more details can be captured, potentially increasing accuracy. Too large depth limits model complexity, which unfortunately leads to overfitting. Through depth-wise tuning, maximum depth at the value of 20 is utilized as it provides good enough accuracy by trading off model complexity. Beyond that, the accuracy increases very marginally; on the other hand, the training time increases drastically. Other hyperparameters such as the number of trees (set at 100 for `n_estimators`) and the criterion for splitting (`gini` or `entropy`) will be tuned to provide the best performance for the model. D. Model Evaluation Fig.1 Shows the Structur of the drought model Different evaluation metrics will be tested on the trained model: Accuracy: It measures how correctly the model classifies water and non-water pixels. Precision, Recall, and F1 Score: These measures denote how the classification of the model as true positive (water) and not as false positive. RMSE and MAE: These are the measures to determine the error for continuous predictions. (if the task is a regression task). Confusion Matrix,: This is the detailed table representing true positives, false positives, true negatives, and false negatives. Fig 1. Structure of the drought model E. Predicting Water Bodies for New Regions The trained model is then used to classify new satellite images to identify water bodies in various parts of the region. The satellite image undergoes the preprocessing steps, like atmospheric correction and cloud masking, and relevant features are extracted. Finally, feature data is input into the trained Random Forest model for classification of each pixel as water or nonwater. The output is a map of water bodies in which the pixels are labeled according to prediction. Post- 9 processing steps - such as morphological operations, like erosion and dilation - help perfect the detected boundary; small misclassified regions are filtered out by connected component analysis. Fig.2. Structure of the water quality F. Visualization and Comparative Analysis As shown in the figure 3 this shows the dry area in the figure. The generated water body maps are visualized using GIS tools in order to understand the spatial distribution of the detected water bodies. Compared to traditional methods that depend on thresholds of fixed thresholds of water indices used for example, NDWI, this method shows improvements in the accuracy of detection. It applied not only quantitative metrics but also qualitative visual inspection of the maps. Fig3. Illustrate the map of SMI's to figure out the severity and chances of occurrence of dry area. This well-structured implementation gives you the clear description of the process through which this automatic detection of water bodies was attained. This therefore ensures to provide a comprehensive discussion in your research paper. You can also proceed to elaborate on specific details involving the dataset, parameters, or the preprocessing steps you used and how you made it even more specific to your study.

V.CONCLUSION

This paper proves that machine learning, particularly by utilizing the Random Forest algorithm, can effectively be used for the automation detection of water bodies in agricultural soils using remote sensing data. It is built upon multi-spectral satellite imagery and advanced feature extraction methods such as spectral indices and texture analysis. The proposed approach will provide a robust and accurate means of identifying water bodies, and given the model's ability to handle large datasets with multiple features, such a model would be ideal for analyzing complex agricultural landscapes where water availability is a very critical factor. The result shows that the automated water body identification system gains higher accuracy and reliability under varied conditions as compared with the traditional threshold-based methods. The system, incorporating machine learning, adjusts dynamically to changes in environmental factors and improves the detection of water bodies under various conditions of seasonality and water management practices. This also verifies good generalization capabilities by the model on testing data, thus opening even broader applications in real scenarios such as drought monitoring, management of irrigation, and water resource planning.The utility of machine learning and remote sensing technologies in environmental monitoring therefore has benefits, as is found in this study. Challenges remain ahead, including further improvements on post-processing techniques towards better definition of the boundary accuracy and more use of sources of data such as further utilization of the highresolution imagery or topographical data for further model improvements in performance. Future applications might also include the integration of temporal analysis for change monitoring on water bodies throughout time, as well as the greater application of more enhanced deep learning techniques towards improved accuracy in detection. 10 Overall, this research provides a promising solution for an automated water body detection in agricultural regions, leading to efficient water management and sustainable agriculture.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311067|

REFERENCES

[1] Haekyung Park, Kyungmin Kim and Dong Kun Lee, "Prediction of Severe Drought Area Based on Random Forest: Using Satellite Image and topography data"- MDPI journal Water 2019, 11(4), 705.
[2] Getachew Berhan Demisse, Shawndra Hill, Tsegaye Tadesse and Solomon Atnafu, "Drought Prediction System for Improved Climate Change Mitigation", Research Gate, July 2014
[3] Zengchao Hao,Vijay P. Singh and Youlong Xia, "Seasonal Drought Prediction: Advances, Challenges, and Future Prospects", AGU, January 2018.
[4] Zengchao Hao, Amir AghaKouchak, Alireza Farahmand and Navid Nakhjiri, "Global integrated drought monitoring and prediction system",,Scientific data, March 2014.
[5] Eric F. Wood,Siegfried D. Schubert,AndrewW. Wood, "Prospects for Advancing Drought Understanding, Monitoring, and Prediction", American Meteorological Society, August 2015.
[6] Sachidananda Mishra,Deepak R. Mishra, "Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters",Science Direct Volume 117, 15 February 2012, Pages 394- 406.
[7] Luke M. Moseley "Drought impacts on water quality of freshwater systems; review and integration"- Science Direct Volume 140, January 2015, Pages 203-214
[8] Amir Sadeghian,Steven C. Chapra,Jeff Hudson,Howard Wheater,Karl,Erich Lindenschmidt " Improving in-lake water quality modeling using variable chlorophyll a/algal biomass ratios"- Science Direct Volume 101, March 2018
[9] Usman Atique,Kwang-Guk An,"Reservoir Water Quality Assessment Based on Chemical Parameters and the Chlorophyll Dynamics in Relation to Nutrient Regime",February 2018.
[10] Anna Cardall, Kaylee Brook Tanner, Gustavious Paul Williams, "Google EarthEngine Toolsfor Long-Term Spatiotemporal Monitoring of Chlorophyll-a Concentrations",May 2021.
[11]Lei Wang,Min Xu,Yang Liu,Hongxing Liu,Richard Beck,Molly Reif,Erich Emery,Jade Young,Qiusheng Wu, "Mapping Freshwater Chlorophyll-a Concentrations at a Regional Scale Integrating MultiSensor Satellite Observations with Google Earth Engine",MDPI journal October 2020.
[12]Stuart Campbell, Cameron Miller, Andrew Steven, Andy Stephens "Photosynthetic responses of two temperate seagrasses across a water quality gradient using chlorophyll fluorescence",Science Direct 24 June 2003.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details