Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 1, pp. 519–534, DOI 10.12694/scpe.v25i1.2316

CROP FIELD BOUNDARY DETECTION AND CLASSIFICATION USING MACHINE LEARNING

D.BHAVANA*AND MYLAPALLI JAYARAJU†

Abstract. Crop classification and detection of crop field boundaries empower farmers which helps the agricultural businesses to estimate crop field dimensions and yields accurately. Our research focuses on estimating crucial agricultural inputs such as seeds, pesticides, insecticides, and fertilizers to enhance overall production. The conventional method of manually identifying field boundaries is both time-consuming with labour-intensive. In contrast, our study harnesses data from diverse satellites such as Sentinel, Landsat, and MODIS, encompassing valuable land usage information. By integrating this data with machine learning algorithms, we achieve real-time monitoring of crop fields through effective classification and boundary identification. For the classification of crop fields within our study area, we recommend employing the Classification and Regression Tree (CART) algorithm. Additionally, we leverage normalized difference indices, such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI), as features for classification. We compare these features with Support Vector Machine (SVM) and Random Forest (RF) algorithms. Subsequently, we utilize the Canny edge detection technique to identify boundaries within the classified crop areas. Notably, our approach utilizes the Google Earth Engine (GEE) as a primary platform for extracting features, conducting data training, and visualizing information. The proposed algorithm yields impressive results with a high level of accuracy. Notably, the CART algorithm achieves a remarkable accuracy rate of 96.1%. Furthermore, we incorporate NDWI-based Canny edge detection into our methodology. The outcomes convincingly underscore the practicality and applicability of our research in real-world scenarios.

Key words: Crop field boundary detection,Normalized difference indices, Canny edge detection

1. Introduction. To satisfy the expected rise in food demand, agricultural production must be raised while reducing its negative environmental effects. There are many powerful tools and technologies for optimizing agriculture. Machine learning is one such technology. Utilizing satellite data and applying machine learning algorithms provides farmers with a healthy environment for farming, thus after detection of the boundaries right amount of water, fertilizers and pesticides can be supplied to the farmers and this helps them reach optimal yield while using less amount of renewable resources.

Land-use details are very important in modern farming, hence field boundary detection has been opted in order to give accurate and up-to-date results [1]. Boundary detection helps the suppliers, government, and policymakers know about the areas under crops and their yield. Usually, existing administrative maps are considered and field boundaries are detected manually based on the surveyed data. But it involves a huge amount of manual labor as many number of maps are to be updated [2]. Due to this, the yield predictions produced will not be accurate. Thus, Automated Detection is way better and more helpful it easily identifies the boundaries of different crop fields across the country with minimal human involvement. This will be beneficial in the countries like India, where digital records are not available excessively. The exact detection of boundaries helps in obtaining more precise information about the crop yield. This is where Earth Observational satellite data comes into play.

Lately, Satellite imagery is abundantly available which is cost-effective and frequently updated. Boundary detection is typically seen as a mid-level method for determining the borders of (and between) objects in scenes, with tight linkages to both grouping/segmentation and object form.However, the satellite imagery have their limitations. They are usually available at low image resolution [3]. The properties of the images also change depending upon the land area covered. Hence, there is little research interest on field boundary detection

*Professor, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur District, AP 522502, India (Corresponding author: bhavanaece@kluniversity.in)

†PG student, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur District, AP

when compared to other applications of satellite imagery. Therefore, the robust solutions for automatic field boundary detection are rare.

Many recent studies on field boundary detection, make use of several machine learning algorithms. There are several other studies like, Turker and Kok used Gestalt laws along with perceptual grouping in order to detect the boundaries in known fields. Rahman et al used an approach which makes use of statistical data on crop rotation patterns. Tiwari et al. used fuzzy logic rules along with color and texture information of the images and finally after identifying the boundaries they refined them using snakes. Yan, L.; Roy, D.P used web enabled Landsat data (WELD) along with watershed algorithm in identifying the automated crop field boundaries [4]. Recently, Watkins and van Niekerk compared various edge detection kernels along with watershed, multi-threshold, and multi-resolution segmentations to identify their potential in field boundary detection. Their results have shown that Canny edge detection and watershed have produced best results when compared to other algorithms [4].

In this study we propose a method which is NDWI based canny edge detection which helps in automatic detection of crop field boundaries. Normally, traditional edge detection techniques make use of spatial information of an image but this approach makes use of both spatial and spectral information [5-7]. In this study the identification of crop field boundaries is performed in unknown fields with minimal amount of prior information. A normal human cannot know the crop areas in a particular region; they must perform the study manually in order to identify the crop areas in a location. Hence, in order to reduce this effort the classification is proposed in this research. Crop classification helps in finding out the areas where the crops are present in a particular location. Our approach use Sentinel-2 and Landsat-8 datasets from Google earth engine where a classification algorithm CART is applied and proposed in order to classify and identify the crop area and furthermore NDWI based Canny Edge algorithm is applied to the detected agricultural area in order to obtain the boundaries for crop fields. This data is used as training data and further after classification, all the algorithms were analysed and compared in order to propose a better algorithm through this research.

2. Literature Survey. In the Corresponding approach 3 different locations around Southern India are selected as the region of interests. The Vijayawada region situated in Krishna basin in Andhra Pradesh state is assumed as ROI-1, Mydukur area which is situated in Rayalaseema region of Andhra Pradesh state was assumed as ROI-2 and the Alappuzha region of the Kerala state was considered as the ROI-3 [5]. The 3 regions were selected in order to visualize and identify the performance of the algorithm for any region when it is applied. The goal of this research is to identify crop lands by using classification and detect the crop field boundaries at the identified crop lands.

3. Data Acquisition. Satellite imagery and Ground truth data are explained in in detail within this section.

Satellite Data. Satellites are referred as the 'eye of the sky'. Satellite imagery also known as Satellite Data is a collection of group of images which consists information about the Earth which is gathered by man-made satellites. Satellite data is generated by using remote sensing technologies. Satellite Data has authentic information about Earth surface, weather and others. This data helps us understand long term changes and act accordingly. Benefits of using satellite data are: We can monitor large areas at a time. We can demonstrate movements on large areas especially sea. Satellites can deliver the data irrespective of the conditions on Earth atmosphere like light and weather conditions. The satellite images show the Earth surface conditions extremely well.

Sentinel – 2. Sentinel-2 is mission for observing Earth by Copernicus programme which is a European union's earth observation programme [6]. The Sentinel 2 is maintained and operated by the European space agency. Sentinel captures optical imagery at high spatial resolution around about 10m to 60m over the land and coastal regions.

This Sentinel is a constellation with 2 satellites Sentinel-2A which was launched on 23 June 2015 and Sentinel-2B which was launched on 7 March 2017; and also a third satellite Sentinel-2C is under testing and it is preparing to launch in 2024. This mission specializes in broad range applications like agriculture monitoring, land cover classification and water use and it's quality. These satellites were manufactured by consortium led by Airbus defence and space [7].

Sentinel-2 has multispectral data with 13 bands with a swath width of 290 Km. It provides visible to NIR spectral bands and SWIR spectral bands. Sentinel 2 revisits the same place with same viewing angles after every 10 days. It captures the images at spatial resolution 10m/pixel, 20m/pixel and 60m/pixel.

Landsat - 8. Landsat 8 is developed collaboration between NASA and U.S. Geological Survey (USGS). It was called as the Landsat Data Continuity Mission. Landsat 8 was launched on February 11 in the year 2013 from Vandenberg Air Force Base. The design, Construction, Launch and on-orbit revolution was led and taken care by NASA. Landsat 8 captures images at high resolution which is from 15m to 100m per pixel. Landsat 8 consists of multispectral data with 11 bands. Landsat 8 consists of two sensors – the Operational Land Imager (OLI) and Thermal Infrared Sensor (TRIS). These two sensors provide the images at spatial resolution via bands covering Visible, NIR, SWIR, Thermal and Panchromatic [8-10].

OLI collects the data for visible region, NIR, SWIR and Panchromatic bands. TIRS collects the data for two more bands in Thermal Region. Landsat 8 is regularly providing with around 725 Scenes per day to USGS data archive. Landsat 8 takes approx. 99 minutes per revolution and it completes 14.5 orbits per day. It repeats the coverage which means revisits the same geographical location for every 16 days. The swath width of Landsat 8 is 185 Km. In this Study, the Landsat 8 collection 1 Tier 1 TOA reflectance raster collection from GEE was considered assuming that it is best suited data for the application [11-13].

The Ground Truth Data. The Ground data truth was obtained from Google Earth.This is a user-friendly resource which is helpful for beginner and intermediary learners who are interested in learning more about GIS and wants to perform some analysis and operations in GIS [8]. The data obtained was mapped to the three region of interests as shown in the given below Fig. 4.2.

This research selected the region of interests and mapped the data in sentinel dataset. The image collection were selected in such a way that they possess least cloud cover during the corresponding dates i.e., 01-01-2020 to 01-02-2021. Furthermore, the classification and boundary detection was applied using the ground truth data obtained through Google Earth [9].

4. Methodology. As discussed earlier in this research we have identified agricultural fields in the study site using CART algorithm and then applied NDWI based canny edge detection in order to identify the boundaries of the agricultural fields. The workflow and methodology is explained in the following steps in detail.

The procedure has been done in five steps: 1) feature definition, 2) feature extraction, 3) dataset, 4) classification 5) boundary detection. For any machine learning approach the data cleaning and preprocessing are considered key steps which shown in detail in the below sections.

4.1. Feature definition. To remotely detect and sense any kind of land cover, the basic mechanism to be carried out is to acquire the electromagnetic wave reflectance information from sensors onboard the satellite. This information is then processed and analyzed for variations that could help detecting the targeted land cover type. The reflectance of spectra from every area differs with regard to the vegetation, water that area possess [10]. The information collected through the spectra of the visible, near-infrared, and mid-infrared, as well as the ultraviolet, is what remote sensing of vegetation is principally dependent on. However, as was previously said, the satellite data that we are employing in this study only offers high resolution in the blue, green, red, and near-infrared bands. Therefore, this work could only use certain spectral areas. Further, indices like NDVI and NDWI were computed with the formulae which include coefficients calibrated specifically for satellites like MODIS, Landsat-8 and Sentinel-2 As a result, we employed appropriate indicators in the form of normalised differences relevant to the detection of vegetation, as well as standardised equations that had been carefully considered and calculated [11]. NDVI and NDWI are used for feature extraction, where these two metrics are able to extract vegetation and water content in the region of interest shown in Fig. 4.1.

NDWI. While McFeeters NDWI index is frequently used to define water bodies, it can also be used to track changes in the water content of plant leaves. Given the variety in moisture levels among the crops used in this study, this is particularly helpful. As indicated in equation 2, it may be calculated [12].

$$NDWI = \frac{G - N}{G + N} \tag{4.1}$$

where N is the near-infrared band's surface reflectance and G is the green band's surface reflectance. The NDWI exhibits good results for the boundary identification. As it measures the water content of the plants it

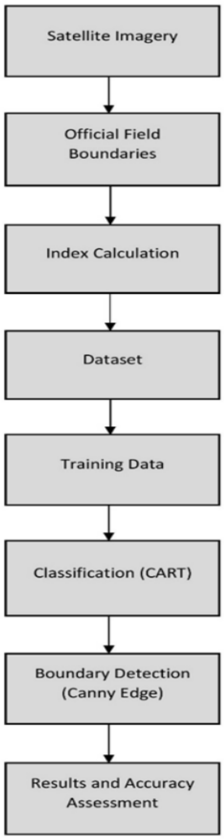

Fig. 4.1: Flow chart of Methodology

was chosen so it can reduce the number of infield edges and helps in finding the edges perfectly.

The fundamental contrast between the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) is rooted in the spectral bands employed during calculation and the intended focus of analysis. NDVI is derived from the near-infrared (NIR) and red light bands, primarily indicating the presence and vitality of vegetation. In contrast, NDWI utilizes the green and NIR bands to detect and evaluate the moisture content within various features, such as water bodies and moisture-laden vegetation. While NDVI centers on vegetation, NDWI is tailored for investigations related to water content.

The Normalized Difference Vegetation Index (NDVI) assumes a pivotal role in crop classification due to its capacity to quantify and assess vegetation density and health. By examining NDVI values derived from remote sensing data, distinctions among diverse crop types can be established, enabling the monitoring of growth stages, health conditions, and spatial distributions. This wealth of information contributes to agricultural management, precision farming methodologies, and informed decision-making encompassing aspects like crop yield estimation, irrigation scheduling, and the detection of pests and diseases.

4.2. Feature Extraction. In all classification and border detection algorithms, feature extraction is regarded as the important step. Ground Truth data mapping and identification of the features is considered as the preliminary steps for the Feature extraction. As stated earlier the Ground truth data was obtained from Google Earth. The images used in this research consists of 10m/pixel and 30m/pixel resolution which are provided by Sentinel level 1c dataset and Landsat8 TOA Reflectance dataset. We mapped the ground truth data on to the sentinel level 1c dataset hosted by Google Earth Engine.

4.3. Dataset. The Sentinel-2 Level-1C and Landsat 8 collection 1 Tier 1 TOA reflectance raster collection were used to perform the land cover classification along with the boundary detection of the selected region [13].

4.4. Classification. Using the features extracted above we have performed the proposed machine learning algorithm CART, and also 2 other machine learning algorithms Random Forest and Support vector machine taking the indices and raw bands as the input features to the algorithms.

Classification and Regression Tree. Classification and regression trees are a way of understanding decision tree techniques which are used for classification and regression learning tasks. CART was developed for regression tasks in 1984 by Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. It is also a prediction model that aids in the discovery of a variable reinforced by other labelled variables. To be more specific, the tree topologies anticipate the outcome by asking a series of if-else arguments.

Classification and regression trees (CART), a basic yet effective prediction method. CART, unlike logistic and linear regression, does not create a prediction equation. Instead, data is partitioned along predictor axes into subsets with homogenous dependent variable values—a procedure illustrated by a decision tree that would be used to create predictions from fresh observations. The CART method is a classification technique that is used to construct a decision tree based on Gini's impurity index. It is a simple machine learning method with a wide range of applications. Leo Breiman, a statistician, created the concept to characterise Decision Tree methods that are often used for classification or regression predictive modelling applications.

A decision Tree is a predictive analytic approach used in statistics, data mining, and machine learning. The decision tree is used as the predictive model in this case, and it is used to progress from observations about an item, which are depicted by branches, to the product's predicted values, which is represented by leaves. Decision trees are among the most popular machine learning approaches due to their accessibility and flexibility.

The CART algorithm does this by utilising the Gini Index criteria to find the optimal homogenization for the subnodes. A decision tree's structure is made up of three major components: root nodes, internal nodes, and leaf nodes. The root node is used as the validation set, and it is categorised into two halves based on the best attribute and threshold value. Furthermore, the subsets are divided using the same rationale. This process is repeated until the tree's last pure sub-set is discovered or the maximum number of leaves feasible in that developing tree is reached. This is sometimes referred to as tree pruning.

Gini's Impurity Index is given by the equ.4.2 shown below.

$$Gini = \sum_{i=1}^{c}((pi))^2 \tag{4.2}$$

Gini's impurity index is a measure used in decision tree algorithms to quantify the impurity or disorder of a set of data points. It ranges from 0 to 1, where a value of 0 indicates a completely pure or homogeneous set, and a value of 1 represents maximum impurity or heterogeneity. The index is calculated by summing the squared probabilities of each class within the data set and subtracting the sum from 1.

Fig. 4.2 combines both training and testing data in order to obtain the better results. The CART algorithm works best for the classification of the land cover type.

4.5. The Random Forest. RF classifier is a collection of decision trees in which randomly sampled rows and attributes are supplied to replacement decision trees. The classification is carried out based on the majority vote that is gathered from decision trees, and the prediction is carried out on regression data using the mean of all the decision tree outputs. Simple trees typically have high variation and low bias. The variance is decreased by increasing the number of trees, making it the ideal model to fit the data. The hyperparameters that must be selected in order to train the model are the number of trees and tree depth. It is the best classifier for all kinds of data because it almost never over fits the model and is immune to the curse of dimensionality. Random forest classifier suits well for land cover classification and classifies better than many existing statistical methods.

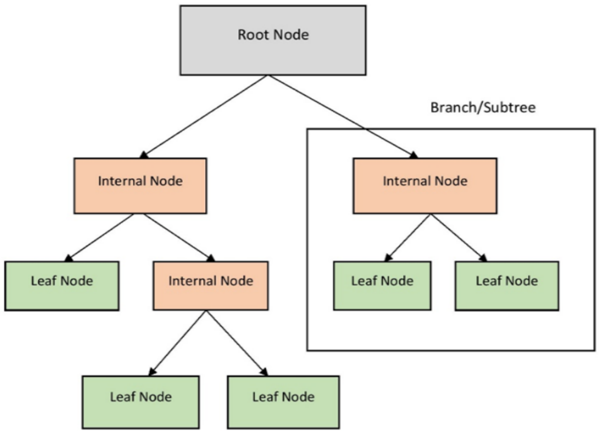

Fig. 4.2: Architecture of CART

4.6. The Support Vector Machine. SVM is considered as one of the best classification models. It has been introduced in the 1960's and later improvised in 1990's. Hyper planes are margins that help in classification of data points. Different classes are constructed by hyper planes that is caused by data points falling on either side of it. The number of features define the dimension of the hyper plane i.e., n features require n-1D hyper plane. Kernel function is a term used for hyper planes. In the current study, a radial-based kernel that classifies in infinite dimensions is applied. Kernel functions assume that the points in space exist in higher dimensions without actually transforming them before calculating the relationship between each pair of points. SVM is used on this data because it works well when the data does not have dimensionality issues and trains more effectively with less samples [14-15].

4.7. Boundary Detection. After the classification is performed we have implemented boundary detection for various crop fields in the different region of interests using canny edge detection algorithm.

4.8. Canny Edge. Canny edge detector is an edge detection algorithm that works in multiple levels in order to identify the edges in given input images. Among all the edge detection algorithms the canny edge detector provides good and reliable detection. This algorithm is adaptable to various environments as it helps in the detection of the boundaries even if the images consists of different characteristics within it. Hence this algorithm was choosen for our crop field boundary detection shown in Fig.4.3.

The Canny Edge Detector was developed by John F Canny in 1986. This technique only extracts the required information from the input and by this it reduces the amount data to be processed. It is being widely applied in computer vision systems. Due to the simplicity of the processing and implementation the canny edge detection is widely used. The canny edge detection algorithm works in 5 different steps those are 1) Applying Gaussian filter in order to remove the noise and smoothen the image, 2) identifying the different intensity gradients in the input image, 3) Applying the gradient magnitude thresholding to get rid of the spurious response to edge detection, 4) Applying double threshold to determine potential edges, 5) Track the edges by hysteresis.

As the first step A Gaussian filter is applied to the input image in order to filter out and reduce the noise in the image. If there is any noise present, it may lead to the false detection. Hence the Gaussian filter is convolved with the input image to smooth it.

The process of utilizing the Gaussian equation to eliminate noise from an input image revolves around convolving the image with a Gaussian kernel. This kernel represents a two-dimensional distribution adhering to the Gaussian probability density function. Its characteristics are defined by both its size and standard deviation (sigma), which dictates the distribution's extent [14-15]. The application of the Gaussian filter involves

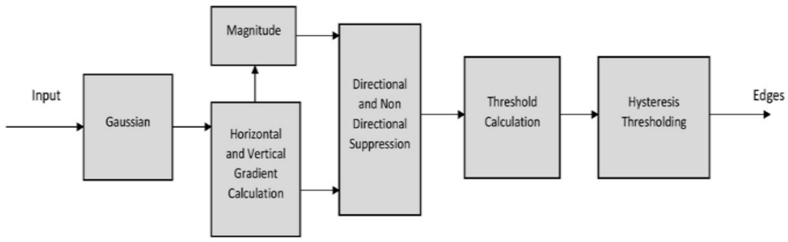

Fig. 4.3: Working of the canny edge detection algorithm

convolving each individual pixel within the image with the Gaussian kernel. This action entails computing a weighted mean of the pixel values found in the vicinity of each pixel. The weights are determined by the values assigned by the Gaussian distribution at those specific positions. Subsequently, the computed weighted average value takes the place of the original pixel value, leading to a notable reduction in noise and the attenuation of high-frequency intricacies present within the image. This iterative procedure is conducted for every single pixel throughout the image, culminating in a filtered image that exhibits significantly diminished noise levels.

In the given input image the edge maybe present in different directions, so in order to overcome this canny edge detector uses four filters to detect vertical, horizontal and diagonal edges. In order to thin out the edges the non-maximum suppression is performed. After performing this step the edges become thinner. But, it can be observed that there is a difference between the intensity of different pixels.

In the process of applying the Gaussian filter to an input image, every pixel within the image undergoes convolution with the Gaussian kernel [14]. This operation entails computing a weighted mean of pixel values surrounding each individual pixel, where the weighting factors are governed by the Gaussian distribution. The computed weighted average is then utilized to substitute the initial pixel value. The outcome is a perceptible reduction in noise and a decrease in high-frequency intricacies present in the image. This iterative procedure is carried out for each pixel within the image, ultimately yielding a filtered image characterized by diminished noise levels.

In the further steps threshold is calculated by identifying the high intensity and low intensity pixels. After getting the threshold results the hysteresis thersholding is performed. This is the final and important step which transforms the weak pixels into strong ones. In this way the Canny edge detection algorithm works. The working flow diagram of the above discussed five steps is shown in Fig. 4.3.

The compactness of the vertical gradient and horizontal gradient depends on the specific characteristics of the proposed work and the image being processed shown in Fig.4.3.

Algorithm:

1. Identification of agricultural fields
2. Partition the datasets into 80% training data and 20% testing data
3. CART classification algorithm is applied on datasets
4. The NDWI based canny edge detection algorithm is applied on the classified data for boundary detection.
5. Qualitative assessment

5. Results and Discussion. As discussed earlier this research is done in two major steps those are classification and boundary detection

5.1. Classification:. The land cover classification for the different region of interests is performed using ML algorithm CART, and two other machine learning algorithms RF and SVM are also performed in order to

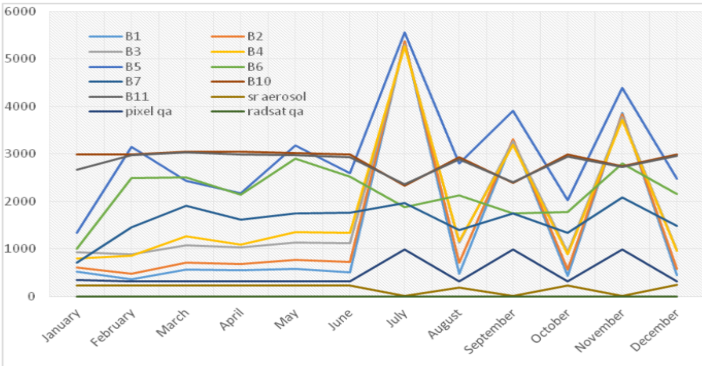

Fig. 5.1: Reflection differences of the all the bands of Landsat 8 in the year of 2021

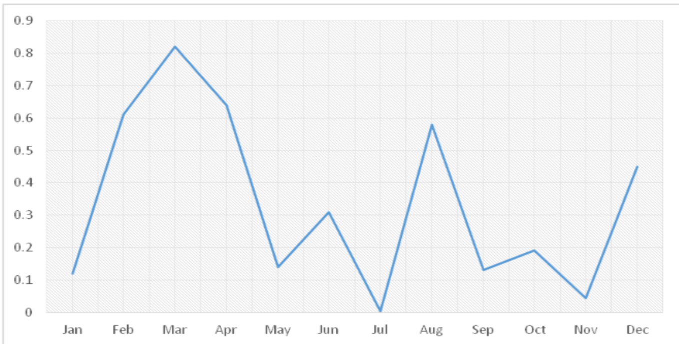

Fig. 5.2: Visualization of NDVI for the year 2021 in the region of interest 1

compare the performance of the proposed algorithm. Our data is trained with vegetation indices NDVI and NDWI and also all the bands in the corresponding year of 2021. 20% of the dataset was used as testing data. The following represents the visualized outputs obtained by using the classifiers.

As observed in the figures of Figs. 5.1-5.4 gives the visualized output differentiates between the land cover type and the colors red, green and blue represents the Urban, Croplands, River respectively. The classified output was very satisfactory with the good amount of accuracy shown in Fig. 5.6.

5.2. Evaluation Matrices. As discussed earlier the data was extracted and cleaning, pre-processing is done. After these steps the data is fed to several machine learning algorithms like CART, RF, SVM and Canny. The effectiveness of a model can be determined by performing evaluation using some standard metrics. The classification models in this study were evaluated using confusion matrix.

Confusion matrix is actually a table which is used to test the performance of a classification model where the true values a known. The table is a combination of Actual values vs Predicted values.

The matrix consists of 2 classes 'Yes' and 'No'. Using this confusion matrix, Accuracy, Recall and Precision are calculated. Let us know some basic terms used in a confusion matrix. These basic terms are used in calculating the Overall accuracy and Kappa Coefficient. Those basic terms are given below:

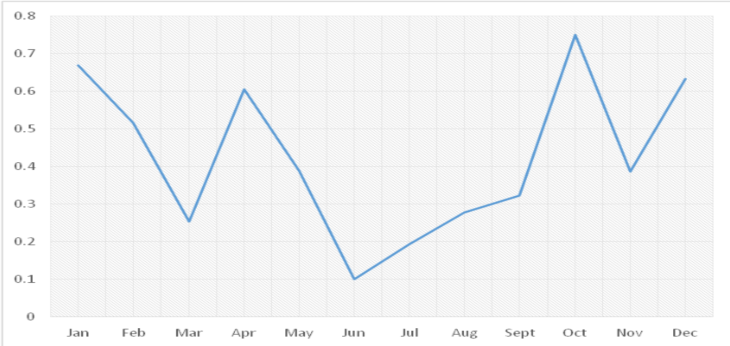

Fig. 5.3: Visualization of NDVI for the year 2021 in the region of interest 2

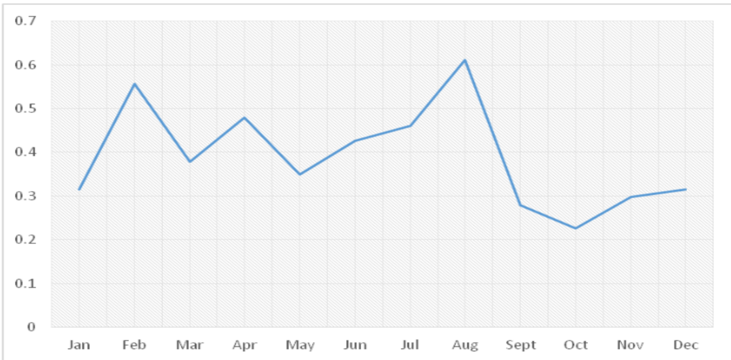

Fig. 5.4: Visualization of NDVIfor the year 2021 in the region of interest 3

True positive (TP). This is a case where the predicted class is Yes and the actual class is also a Yes.

True negative (TN). This is a case where the predicted class is No and the actual class is also a No.

False positive (FP.) This is a case where the predicted class is a Yes and the actual class is No. This is also known as “Type I error”.

False negative (FP). This is a case where the predicted class is a No and the actual class is Yes. This is also known as “Type II error”. In a confusion matrix the rows correspond to actual values and the columns correspond to the predicted values. Using this confusion matrix, Recall, Precision and Accuracy can be calculated.

Recall. Out of all affirmative classes, how many are successfully predicted is calculated. This value must be high for good classifier. Recall is given by following equation.

$$Recall = \frac{TP}{TP + FN} \tag{5.1}$$

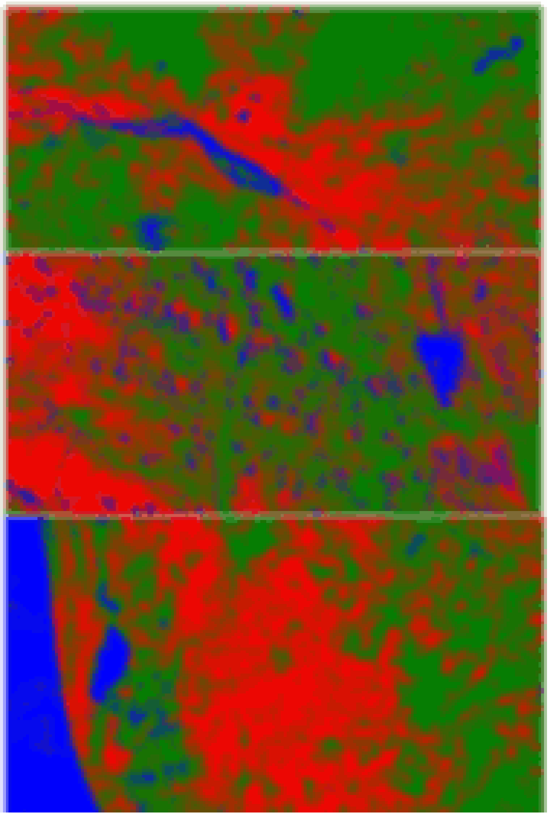

Fig. 5.5: Classified output of CART Classifier for 3 ROIs

Precision. It determines how accurate the model is out of predicted positive, number of actual positive classes. It must be high for a good classification result. It is formulated as shown in equation.

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

Accuracy. It determines the accuracy of the model based on True positive and true negative values. It must be high for good classifier. Which is also termed as overall accuracy. It is given by following equation

$$Accuracy = \frac{TP + TN}{Total} \tag{5.3}$$

Cohen's Kappa. Cohen's Kappa coefficient (k) is a measure of how closely the classified data by a machine learning classifier matches the actual data or the ground truth data. Generally a classifier with kappa statistic value between 0.4 to 0.75 is considered as a good classifier and above 0.75 (>0.75) is considered to be as excellent. It is also calculated by using a confusion matrix. The Cohen's Kappa formula can be given as:

$$Cohen'sKappa = \frac{2 * (TP * TN - FN * FP))}{((TP + FP) + (FP + TN) * (FN + TN))} \tag{5.4}$$

To calculate Cohen's Kappa in the proposed work, the agreement between two raters is measured using the observed agreement and expected agreement. The observed agreement is the proportion of cases where the raters agree, and the expected agreement is the agreement expected by chance. Cohen's Kappa is then

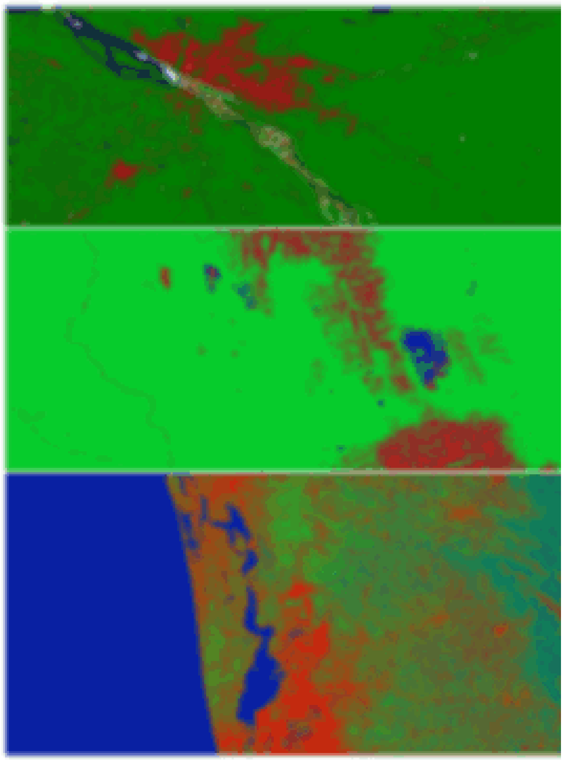

Fig. 5.6: Classified output of the SVM for 3 ROIs

Table 5.1: Confusion Matrix of CART Classifier

+/-	0	1	2
0	66	8	4
1	0	128	30
2	0	5	981

calculated by subtracting the expected agreement from the observed agreement and normalizing it by dividing by 1 minus the expected agreement [16-17].

The overall accuracy in a confusion matrix are calculated by adding all the correctly classified values and further by dividing it with the total number of values in the matrix. To show the performance of classifier models the overall accuracy (OA) is evaluated using the confusion matrix provided by the Google Earth Engine along with Cohen's kappa coefficient for each classifier using the confusion matrix.

It can be observed that some classes are missing in Table 5.1. As these classes don't have enough samples, the model excluded them while being trained and hence, the matrix does not reflect them. The confusion matrix shown above is calculated in Google earth engine and further the evaluation matrices were calculated according to the formulae. The overall accuracy (OA) was obtained as 96.1% and the kappa coefficient (k) is 0.87 for CART which is very good. Furthermore, the classified output was visualised as shown below in Figure 5.7.

It is interesting that other classifiers also have given good results with good amount of accuracies. The outputs were visualized in Google Earth Engine platform for the Random forest classifier also the visualized outputs are shown in below figure 10. The Random Forest Classifier have shown less accuracy when compared to CART classifier because it found difficulty in differentiating between the features given to it as the input.

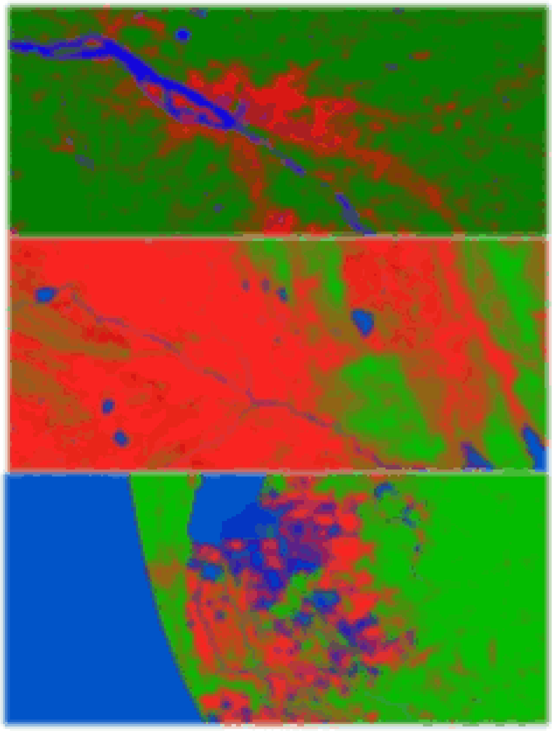

Fig. 5.7: a) Overall accuracy of classifiers b) Kappa coefficient of each classifier

The Random Forest classifier model hosted by Google Earth Engine was applied to the same training data and the results obtained were overall accuracy (OA) is 83.6% and the kappa coefficient (k) is 0.71 which means it was a good statistic. The random forest found difficulties in differentiating among the wetlands, rivers and croplands. The confusion matrix was used to calculate the accuracy and kappa statistic [18]. The support vector machine classifier model hosted by Google Earth Engine was applied to the same training data and the results obtained were overall accuracy (OA) is 85.7% and the kappa coefficient (k) is 0.73 which means it was a good statistic. The confusion matrix was used to calculate the accuracy and kappa statistic. It was applied in order to find an efficient algorithm for the classification of the land around the selected region of interest. But the SVM algorithm was showing accuracy better than RF but lesser than CART shown in Fig. 5.7.

6. Crop field Boundary Detection. In this study canny edge detection was used for detecting the boundaries of the crop fields identified by the CART classifier. NDWI index is calculated for the time period of Kharif season (The months of June – October in southern India) and Rabi Season (The months of October – February in southern India). NDWI is given as an input to the Canny. Number of images are added during the period of each season this summation of the images improve the results and obtain better output data.The NDWI Index was calculated for each month and fed as the input to the canny edge detection algorithm. Several images added upon each image in order to obtain a better output. The limiting factor was the resolution which is 10m/pixel yet still the data provided good results. As shown in Table. 6.1 the bands with high resolution were used to calculate the NDWI those bands are B3 and B8 these bands are used to calculate the NDWI, usage of the bands with low resolution such as B6, B7 need to be avoided in order to obtain good results. The spectral response was influenced by the dry crop growth and total yield decline in 2020, as illustrated in Fig. 6.1.

In this study we have calculated the NDWI for 2 different seasons and 3 different region of interest in order test the efficiency of the algorithm. The reflectance values and NDWI values are dependent on the weather conditions and the growth progress of the crops in the selected region of interest. Hence, The NDWI is shown in

Table 6.1: Comparison of different algorithms with the proposed algorithm

S.No	Algorithm	Kappa coefficient	Accuracy
1	SVM	0.73	85.6%
2	RF	0.71	83.1%
3	CART	0.87	96.1%

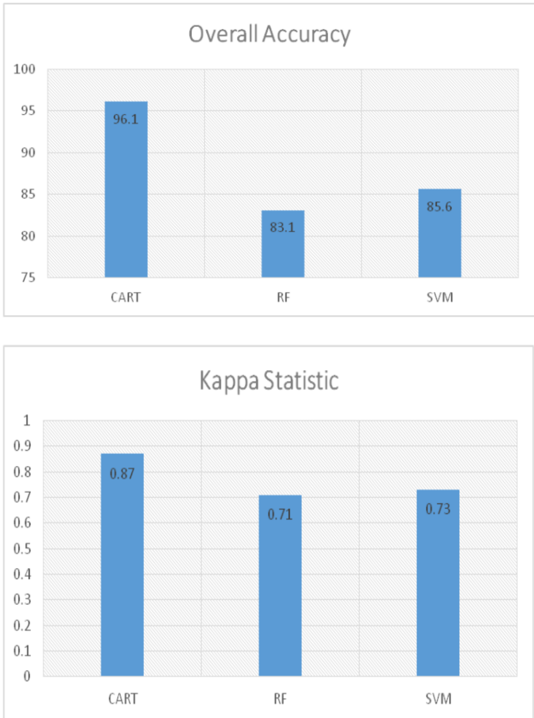

Fig. 6.1: a) Band Reflectance values for ROI-1 in Kharif season b) Rabi season

next figures for each region of interest. The Vijayawada region situated in Krishna basin is assumed as ROI-1, Mydukur area which is situated in rayalaseema region of Andhra Pradesh state was assumed as ROI-2 and the Alappuzha region of the Kerala state was considered as the ROI-3. The visualized output of the canny edge detector which is applied for the Region of interests is shown in Fig. 6.2 and Fig. 6.3.

The trained model, which was previously given, was used for a site close to Alappuzha and for the period of May 2020 to February 2021. This location is not part of the training data set, so we can evaluate how well our model performs in a real-world situation. As seen in the Fig. 6.3 field boundaries are similar in both the Kharif and Rabi seasons by this we can say that our model is efficient and the weather conditions doesn't affect the algorithm.

The results obtained through classification and the crop field boundary detection are satisfactory and the input data show the influence on the performance of the algorithms. Data preprocessing and data cleaning are very important steps that are to be carried before every data science application. As expected, the NDWI-based canny edge detection produced excellent results. The NDVI index and raw bands were used as the features, and the results were excellent and satisfactory. The results are significantly impacted by the use of the raw bands B6, B7, B8, B11, or B12.

Data pre-processing, cleansing, and cloud filtering was another key aspect. The presence of the clouds does

Fig. 6.2: NDWITime series of our ROI-1 during May 2020 – Feb 2021

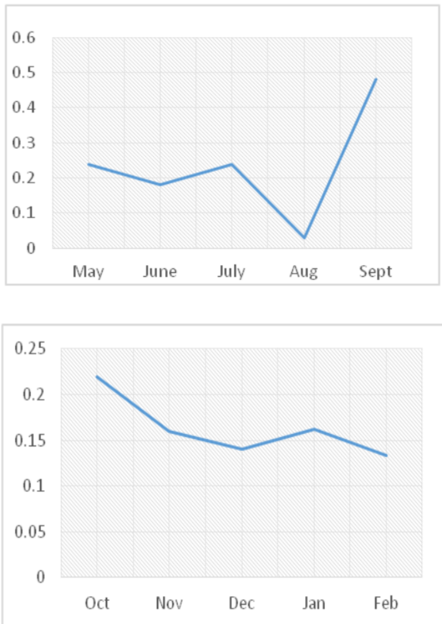

Fig. 6.3: NWDI based Canny output for ROI-2 in Kharif (left) and in Rabi (right) seasons

effect the boundary detection but does not affect much to the classification as compared to the field boundary detection.

The results produced very good classification accuracies and the boundary detection was also very good when compared to the ground truth data. The limiting factor was the resolution which is 10m/pixel yet still the model provided good results. As a result, we contend that our model's performance is excellent and that it is ideally suited for use in real-time scenarios in everyday life. Sentinel 2 and Landsat 8 were preferred as the datasets can be obtained on the regular basis.

7. Conclusion and Future Scope. This study comprehensively tackled aspects of land cover classification and the detection of crop field boundaries, while striving to propose an efficient machine learning algorithm. These applications are particularly vital for analysing crop yields in cases where field information is scarce. The study employed a variety of algorithms, including CART, RF, SVM, and Canny Edge detection. Additionally, indices like NDVI and NDWI were computed, with the incorporation of raw bands as inputs leading to the discovery of supplementary information and increased result accuracy.

Key stages of data pre-processing and cloud filtering were integral to both classification and crop field boundary detection processes. The algorithmic evaluations spanned the time frame of 2020 to 2021. The application of Canny edge detection in conjunction with NDWI yielded notably favourable outcomes. To ensure temporal diversity, this approach was executed on different days during the Kharif and Rabi seasons across 2020 and 2021. Comparative analysis revealed the CART model's superior performance among classifiers. Trained on the specified data, CART exhibited an impressive 96.1% Overall Accuracy and a 0.87 Kappa Coefficient, signifying its efficacy as a classifier. Similarly, RF demonstrated an Overall Accuracy of 83.1% with a Kappa Coefficient of 0.71, while SVM achieved an Overall Accuracy of 85.6% and a Kappa Coefficient of 0.73.

Based on these quantitative metrics, the CART model was advocated for land cover classification due to its robust performance. For precise crop field boundary detection, NDWI-based Canny edge detection was applied during the Kharif and Rabi seasons, producing satisfactory results aligned with ground truth data. Thus, the NDWI-based Canny edge detection method was proposed as a feasible solution. The study's potential extends to predicting crop yields and optimizing pesticide and insecticide usage based on detected boundaries. By leveraging insights from this research, it is possible to make informed decisions that enhance agricultural practices and resource allocation.

REFERENCES

[1] Achanta R and Süsstrunk S. (2017) Superpixels and Polygons Using Simple Non-iterative Clustering. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4895-4904, doi: 10.1109/CVPR.2017.520.

[2] Breunig, Fábio Marcelo, et al. (2020) Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. International Journal of Applied Earth Observation and Geoinformation 85: 102004

[3] Chang L, Peng-Sen S, Shi-Rong L, (2016) A review of plant spectral reflectance response to water physiological changes. Chinese Journal of Plant Ecology, vol. 40, no. 1, pp. 80–91.

[4] Dixon, Dan J, et al. (2021) "Satellite prediction of forest flowering phenology." Remote Sensing of Environment 255): 112197.

[5] NagaJyothi, G., and SriDevi, S. (2017, March). Distributed arithmetic architectures for fir filters-a comparative review. In 2017 International conference on wireless communications, signal processing and networking (WiSPNET) (pp. 2684-2690). IEEE.

[6] Nagajyothi, G., Kumar, G. P., Kumar, B. S., Kumar, B. D., and Damodaram, A. K. (2023, March). High-Speed Low Area 2D FIR Filter Using Vedic Multiplier. In Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022 (pp. 569-578). Singapore: Springer Nature Singapore.

[7] Jyothi, G. N., Anusha, G., Kumar, N. D., and Kundu, D. (2019). Design of FINFET based DRAM cell for low power applications. Computer-Aided Developments: Electronics and Communication, 35-43.

[8] Gamanya, R., P.D. Maeyer, and M.D. Dapper. (2007) An automated satellite image classification design using object- oriented segmentation algorithms: A move towards standardization, Expert Systems with Applications, 32:616– 624.

[9] Kokhan S, Vostokov A. (2020) Application of nanosatellites PlanetScope data to monitor crop growth. In Proceedings of the E3S Web of Conferences, Lublin, Poland, 17–20 September 2020; Volume 171, p. 2014

[10] Kumar, L.,and Mutanga, O. (2018). Google Earth Engine applications since inception: Usage, trends, and potential. Remote Sensing, 10(10), 1509.

[11] Pandey, A., and Jain, K. (2022). An intelligent system for crop identification and classification from UAV images using conjugated dense convolutional neural network. Computers and Electronics in Agriculture, 192, 106543.

[12] Roopashree, S., Anitha, J., Mahesh, T. R., Kumar, V. V., Viriyasitavat, W., & Kaur, A. (2022). An IoT based authentication system for therapeutic herbs measured by local descriptors using machine learning approach. Measurement, 200, 111484.

[13] Kumar, V. V., Raghunath, K. K., Rajesh, N., Venkatesan, M., Joseph, R. B., & Thillaiarasu, N. (2022). Paddy plant disease recognition, risk analysis, and classification using deep convolution neuro-fuzzy network. Journal of Mobile Multimedia, 325-348.

[14] Vinoth Kumar, V., Wang, L., Chen, J. I. Z., Sikdar, B., & Nones, M. (2022). Guest editorial: Trends, perspectives and prospects of sensor technologies in hydrological sciences. Acta Geophysica, 70(6), 2837-2839.

[15] Piao, Shilong, et al. (2019) Plant phenology and global climate change: Current progresses and challenges. Global change biology 25.6: 1922-1940.

[16] Qayyum, Nida, et al. (2020) Glacial lakes mapping using multi satellite PlanetScope imagery and deep learning. ISPRS International Journal of Geo-Information9.10: 560.

[17] Bhavana, D., Likhita, N., Madhumitha, G. V., & Ratnam, D. V. (2023). Machine learning based object-level crop classification of PlanetScope data at South India Basin. Earth Science Informatics, 16(1), 91-104.

[18] Chang L, Peng-Sen S, Shi-Rong L, (2016) A review of plant spectral reflectance response to water physiological changes. Chinese Journal of Plant Ecology, vol. 40, no. 1, pp. 80–91.

Edited by: Polinpapilinho Katina
Special issue on: Scalable Dew Computing for future generation IoT systems
Received: May 26, 2023
Accepted: Sep 25, 2023