Crop Classification with Multi-Temporal Satellite Image Data

Nirbhay Bhuyar[1], Samadrita Acharya[2], Dipti Theng[3]

[1,2]B.E., Department of CSE, G. H. Raisoni College of Engineering, Nagpur

[3]Assistant Professor, Department of CSE, G. H. Raisoni College of Engineering, Nagpur

Abstract:-Agriculture is an important source of livelihood. Crop classification has become important for precision agriculture and helps making many decisions for crop production. However, it's challenging to achieve that precision in this field. Remote Sensing helps to achieve that crop yield assessment, crop health and other parameters. This Paper focusses on how machine learning algorithms can be used for the crop classification with the multitemporal data images from satellite. The models proposed and studied give highest accuracy for crop identification. Detailed analysis with outcome of this is explained further.

Keywords:- Remote sensing, satellite image data, multi-temporal data, crop classification, machine learning, k-means, random forest.

I. INTRODUCTION

County's 15% of GDP is contributed by Agriculture. So, technically it is important to observe and analyses the positive and negative dynamics of the crop development. Therefore, a large amount of development depends on the agriculture and this information is important for planning and resource allocation. Reliable and timely information of the cultivated area is key factor for efficient procurement and India has that good administrative body which can contribute towards significant and rapid development. In this paper we have used the time series and multitemporal properties of satellite images by combining a sequence of images for a particular period of time using machine learning for crop classification. The Remote Sensing is used to obtain the data of particular phenomenon without physical contact. This plays a significant role in analyzing the crop development with its different parameters. Remote sensing data such as satellite images are useful for the surveys.

As Satellite images has its own parameters like wavelengths, bandwidth and spectral bands. These parameters are used to calculate certain indicators called NDVI. The NDVI i.e. Normalized Difference Vegetation Index is exploited for detection of water and vegetation which further includes mapping and monitoring the crop productivity and health. Machine learning refers to algorithm which learns on its own which lead to increase in power of data and computers over past decade making the area extremely popular [2]. Hence this document continues to explore machine learning for classifying crops. We used different supervised machine learning algorithms on different indices for further in detail classification. Hence, we explore an approach to utilize multitemporal vegetation data with other parameters for identification and contribute for compilation of better and advanced method.

The applications for this research include development of a framework that allows the user (analyzers and scientists) to provide their dataset and get the desired output in form of classification. This aims to digitalize the process of field survey records for analysis of crops which was done manually before. This research helps to reduce a lot of human work by reducing paper and manual work interaction and directly sharing of resources. The rapidly evolving automation world now includes digitalization, which increase the productivity. It is the backbone of development, necessary to maintain sustainability of agriculture. This contributes significantly to the economic well-being and agricultural countries around the world.

II. LITERATURE SURVEY

There is a good amount of work done on the Crop Classification using Multitemporal Satellite Imagery with Machine Learning. In relation to crop status, some remote sensing studies have focused on each physical parameters of crop systems, such as nutritional value reference and water availability, as variables for the survey of crop health and yield. Land cover and land use studies are based on maps obtained from the interpretation of images combined with field surveys. The images used represent a frame to cover the study area or are repeated several times on different dates. Discussions often focus on resolution and scale. However, agricultural systems are often more spatial [3].

Traditionally, vegetation monitoring using remote sensing data has been performed using the vegetation index. These are mathematical transformations designed to evaluate the spectral contribution of green plants to multispectral observations. NDVI, obtained by dividing the difference between the infrared and red reflectance measurements by the total, provides an effective measure of photosynthetic active biomass. The normalized difference vegetation index (NDVI), vegetation condition index (VCI), leaf area index (LAI), General Yield Unified Reference Index (GYURI), and temperature crop index (TCI) are all examples of indices that have been used for mapping and monitoring drought and assessment of vegetation health and productivity [5].

In inclusion, the researchers modified the standard NDVI approach by using crop yield masking to improve regional crop forecasts. This technique limits the analysis to pixel subgroups and not to the use of all pixels in a scene. Characteristics of satellite images are defined by four resolution types, namely spatial, spectral, temporal and radiometric resolution [6]

Classification methods in remote sensing mainly consider two aspects, a feature extractor that transforms spatial, spectral, and/or temporal data into discriminative feature vectors, and a

classifier that labels each feature vector to certain types. For crop or vegetation classification, the spatial and spectral features are typically extracted. To aggregate spectral bands into vegetation indices that represent the physical characteristics of vegetation, within which the normalized difference vegetation index (NDVI) is used [7].

SPATIAL RESOLUTION: Shows the number of details in each pixel of the image. If the spatial resolution of the sensor is 10 meters, each pixel of the image acquired by this sensor covers an area of the ground of 10m x 10m.

SPECTRAL RESOLUTION: Is the number of vibration drivers in the band and is represented by the number of vertical bands in which the sensor collects data. The sensor with high spectral resolution has several thin bands, the sensor with low resolution has wide bands, and each band covers a portion of the appearance.

TEMPORAL RESOLUTION: Represents the satellite revisits for particular phenomenon. The number will represent the resolution.

In [4], states that Hyper spectral remote sensing has helped to enhance more detailed analysis of crop classification. This also focuses the unsupervised classification methods i.e. k-means and ISODATA for the crop identification from the remote sensing image. The models are applied only on agricultural fields, which may be singled out with the prevailing land usage classification models.

Xia Zhang et.al. (2016) [8], introduced a new kind of crop treatments, including the construction itself, and object-oriented nature of the vegetation type groups FBS. Model also accounts for 20 additional sensitive index specters of the features that distinguish the parameters of most vegetation. In order to reduce additional information, resolution of birth will improve the resolution of its kind between the pair. Algorithm is proposed, by the kind of crop has been improved with sufficient accuracy, chlorophyll, execute the office of, the texture of which is associated with the index, reducing the effect of the ends of the face of a better mind, apparitions, to the kind of sensitivity is very much a sign of anthocyanin's, which indicates that you'd like. Therefore, insert the seeds of crops, in order to monitor invasive species is an effective approach to better use in agriculture.

In [8], authors have proposed a new spectral-spatial Multi-temporal image classification method based on the nearest K (KNN). The proposed method consists of a carrier vector machine used to obtain initial probability maps. The resulting pixel probability maps are refined with the proposed KNN filter algorithm based on the correspondence and average of non-local neighborhoods. The model does not require specific segmentation and optimization strategies that use the non-local principle of real-world images using KNN and allow competitive classification with rapid calculation.

Numerous experiments conducted on a set Multi-temporal data sets in which the result shows that the classification results obtained with the proposed method are comparable to the various ordering methods of the recently proposed multi-temporal images.

The vegetation stage (phenology) is influenced by various factors such as available soil moisture, planting date, temperature, photoperiod and soil condition. Therefore, these factors also affect the condition and productivity of the plant.

For example, if the temperature is too high during pollination, maize yields can be adversely affected. Therefore, knowing the temperature at which corn is pollinated will help forecasters to better predict corn yield. This summarized the importance of phenology, "Understanding crop phenology is critical to crop management, and practice management time is increasingly based on the development phase of the crop. the sharing of carbohydrates and nutrients, or of identifying important life cycle events such as flowering or ripening, is also important for crop growth patterns.

Image denoising is also one of the topics of research in this field. As the extraction of features from the satellite image is one of the most important part for the crop classification. Other parameters like process of estimating pixels or unmixing or noise separation, etc. are also responsible for the feature implementation. This manuscript also focuses on the use of machine learning algorithms and satellite data to automatically divide fields. This method uses a partitioning cluster algorithm called Partitioning Around Medoids and takes into account the quality of the groups obtained for each. Evaluate 13 satellite bands for better identification of arable land. The proposed method was tested in a vineyard using the Landsat 8 satellite spectrum and heat band. The results of the experiment indicate the great potential of this method for monitoring field prints from remotely detected multispectral images.

In [1], authors have applied and evaluated a modified Hierarchical grouping method based on Gaussian tests for high definition satellite images. The purpose of the model is to obtain homogeneous groupings within each level of the hierarchy which subsequently

The classification and annotation of image data is done from single scenes to large archives of satellite data. After cutting the image into small patches and extracting the elements from each patch, the k-media are used to divide the extracted pixel sets to form a hierarchical structure. Since pixel vectors typically fall into a high dimensional spatial element, we test different distance metrics to solve the "curse of dimension" problem. Using three different synthetic aperture radars (SAR) and optical image data, the Gabor and Bag-of-Words texture functions are extracted and the clustering results are analyzed by visual and quantitative evaluations.

A data mining approach was used to analyze soil behavior and predict crop yields. Different yield forecasting algorithms used to assist farmers in predicting correct crop selection. Traditional yield forecasts were made taking into account the farmer's experience and harvest in a specific area. The proposed system uses data acquisition technology 15 predicts the category of soil data analyzed. The expected category indicates revenue. The problem of crop prediction is formalized as a rule of classification using the Bayes method and the K-Neighbor method. Therefore, you can use the wavelet algorithm to smear the image by reducing its size. Crop classification includes NDVI and LAI calculations as a means of training existing vegetation rates. If the K position and intensity components are merged, the closest one is important when creating the edges of the graph. Classification is improved by considering the separability of class pairs in object-oriented classification. The above aspects are taken

into account when building the proposed project architecture, thus removing the limitations of the above algorithm.

III.   METHODOLOGY

The Crop Classification is the classification of the crops, which encompasses various elements such as growing cycle, crop species, crop variety, crop season and specific land type. There are various crops such as Cotton, Paddy, Wheat, Gram, Sugarcane, etc. which are being used to classify them based on the multi temporal classification. Image Classification can be performed using ML models. Experimental study results show that our model achieves significantly higher search accuracy than traditional query refinement schemes. The application definitely reduces the time as the machine learning algorithms calculates many of the parameters in less time.
Methodology adopted, aims at predicting the crop yield of a particular area by providing the required dataset from that area. It involves the following steps.

- Collection of the dataset for agricultural crop mapping and monitoring.
- Performing the pre-processing on the satellite image for extraction of data features for further analysis.
- Deciding Crop patterns with the help of vegetation index and map them with help of data.
- Applying the machine learning for accurate results

Application Flow:
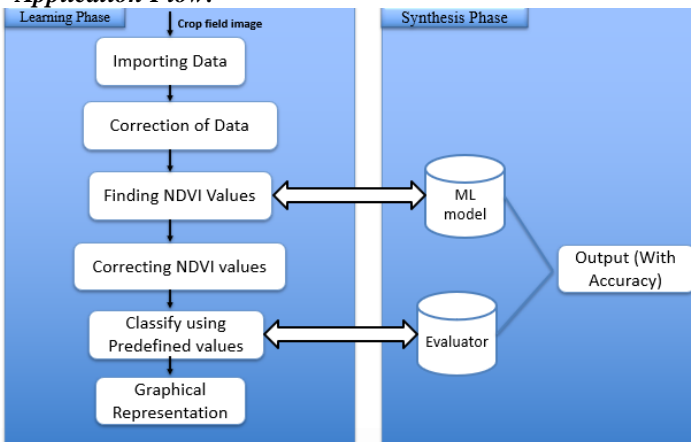
The application flow is bifurcated into two phases:
- Learning phase
- Synthesis phase.

Learning phase:

1. First, we gather the crop field image data, which is a multitemporal data in Tiff format.
2. We import this data for data pre-processing along with importing libraries and functions such as GDAL, NumPy, Pandas, for reading and writing on data.
3. Then we do correction of data; we refine the data by removal of outliers, null entities, cloudiness, Shadow, and Cloud mask.

4. After refining data, we perform feature extraction, which is done on the bases of Pixel value, Greenness, Shape and size of the image data. Next, create a CSV file that contains the value of image in an array format, which will be used for NDVI Value classification.
5. For Finding NDVI values we are using libraries such as TensorFlow, Keras, NumPy, pandas as well as sequential machine learning model. The CSV file obtained by feature extraction is then used for testing and training the data in order to correct NDVI values which hence finally gives the classified NDVI values ranging between -1 to 1.
6. These NDVI values then compared to the predefined crop NDVI FOR crop identification.
7. Now these identified crops are plotted and shown IN graphical representation.

   Synthesis phase:

8. Sequential ml model is used for classification of NDVI values as well as for evaluation for Accuracy.
9. In evaluation phase we are checking the accuracy of the classified NDVI value by further optimizing the error and hence finally obtaining the output with a highest Accuracy of 95.06%.

Dataset Description:
Dataset Name: usgs_cropland_data.
Type: Image from LandSat-5 (TIFF format).
Domain: Agriculture and Rural Development.
The dataset contains 10 images of same land for particular acres but with different time period for more precise result.
The data are images in TIFF format and features like different wavelengths, bandwidths, pixels and raster bands.

We studied and analyzed a variety of supervised and unsupervised machine learning methods. The classification is of two types.
Supervised Classification: This requires a basic knowledge of the study area from external sources or internships. Training camps are used to train a computer to identify several components. The training pitch selection is based on the eye features of each section, observation of familiar areas, Ariel and landscape images. The interpreter enters one pixel and the algorithm incorporates subsequent pixels with pixel values similar to those previously selected.
Unsupervised Classification: This does not require prior knowledge in the study area. Separate pixels from a visual class by examining pixels with similar DLs. Organizations with the same physical characteristics are run. It is based on the choice of space for the classification and sequencing algorithm.

   1. K- Means clustering:

   i.   The set of cluster centers is randomly placed throughout the spectral space.
  ii.   Nearest cluster should be assigned with pixel.
 iii.   Each cluster needs mean location which needs to be recalculated.
  iv.   Above two steps should be repeated until movement of cluster centers is below threshold. Assignment of

class types to particular spectral clusters

2. ISO- DATA - The Iterative Self-Organizing Data analysis technique (ISODATA):
Extends k-means. calculate standard deviation for clusters.

After stage iii we can either:

- Combine clusters if centers are close.
- Split clusters with large standard deviation in any four dimensions [4].

3. Random Forrest:

i. Creation of bootstrap data set
ii. Create decision trees
iii. Repeat step 1 and 2
iv. New data point's outcome prediction
v. Model evaluation.

All classification models along with their parameters are studied and analyzed. All the algorithms are performed with the required agriculture dataset which is the multitemporal data from Land Sat 5. The data was of particular time series of the particular area. This methodology shows that out of all evaluated models, Random Forest gives the best score for accuracy with percentile of 95.06%. The detailed analyzed result is explained in result.

IV.   RESULT AND ANALYSIS

The process of crop identification is done in 3 main steps.

1) **Pre-Processing of data**: In this step we first downloaded and analyzed the multitemporal data of agriculture field for the classification. Import the dataset and perform operations to sort the data i.e. removal of all the zero entities and classifying them into classes then process it into a csv file. Now this file contains the NDVI values in form of data and can be used for further classification.
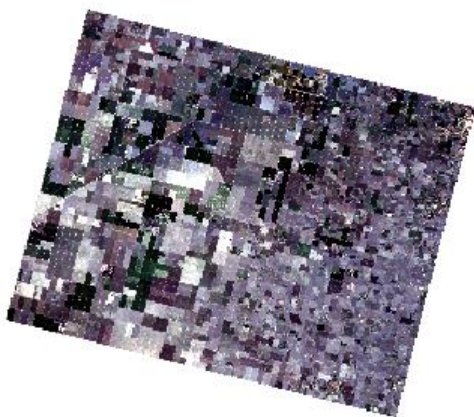

Figure 1: Multi-Temporal data image from satellite before per-processing.

The output we got from pre-processing is the data in csv file format which contains unsorted NDVI values.

2) **NDVI based classification**: This is second step where we studied the NDVI values and classified them according to

raster band using GDAL (Geospatial Data Abstraction Library) and pixel values for classification of the crops.
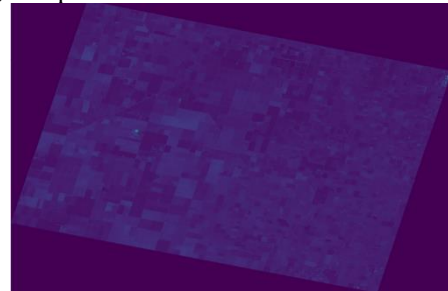

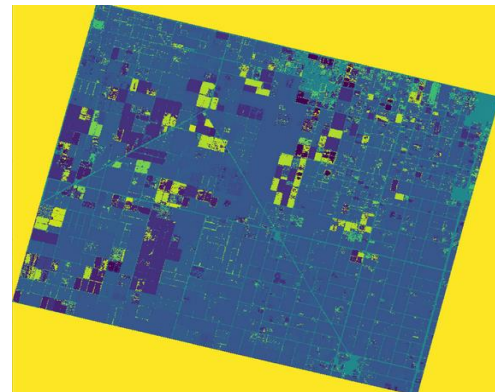Figure 2: Image before NDVI classification.

Figure 3: Image after NDVI classification.

The result of the NDVI classification file contains the perfect NDVI values to identify crop using time series.

The NDVI values obtained is now analyzed to get how man crops are growing in the particular area we selected

3) **Crop Classification Model**: This is the step where we used TensorFlow, Keras and ML models like sequential, Random forest, k-means for training so we can get accurate identification.

As explained before, implementation of these steps with several machine learning algorithm is evaluated with the highest accuracy result given by Random forest method. Below comparison of algorithm shows why the selected model is best.
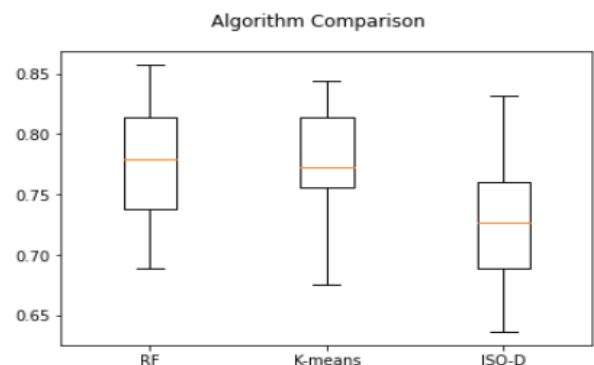

Figure 4: Comparison of all algorithm used.

The accuracy of all algorithm for best fit is defined below.

Name	Mean	Standard Deviation
Random Forrest:	0.773428	0.053375
K-Means:	0.713462	0.051592
ISO-Data:	0.726555	0.061821

This Comparison shows that Random forest algorithm gives the highest accuracy with 95.06% for our dataset. Image Classification performed using ML models. Experimental study results show that proposed model achieves significantly higher search accuracy than traditional query refinement schemes.
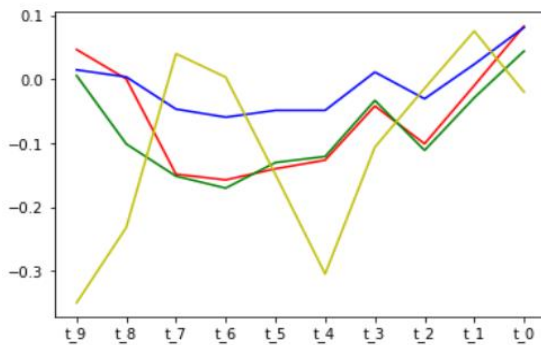

Figure 5: Graphical representation of Identified crops.

Red- Cotton (Ranging: -0.457354 to 0.084001)
Green- Paddy (Ranging: -0.473742 to 0.077552)
Blue- Wheat (Ranging: -0.500975 to -0.159793)
Yellow- Gram (Ranging: -0.489547 to 0.021030)
In this way it is shown that satellite images can be used with sufficient accuracy for crop classification and with minimal image pre-processing. In addition, classifying only the crop types yielded better resolution and improved crop results over the related work.

V. CONCLUSION AND FUTURE SCOPE

Manuscript presented the studied of different agricultural area with usage of the remote sensing. It analyzes multiple papers and documents and the techniques out there for this particular area. The experimental studies are performed with help of machine learning algorithms from multispectral images and it evaluates that random forest gives best result for crop classification. . The image pre-processor is used to remove noise from the image [9], which simplifies work with the image classifier evaluator and allows predicting soil classes more accurately.

For our future work, the proposed work can be extended to a mobile app for users uploading images from farmland. Easy access. The validity of preprocessed data is limited due to unnecessary information. The presence of this unwanted information in the input image, both during training and during classification, makes the performance less efficient because the preprocessor cannot identify the exact contours.

 Image parameters such as climate coefficient, humidity and historical data sets can be used to predict crop yields. Gathering more useful details about soil class, latitude, latitude and appropriate crops will greatly increase your work efficiency. Therefore, we can improve pretreatment equipment and expand other functions so that we can make a significant contribution to the welfare of agriculture around the world.

REFERENCES

[1] Rustowicz, R. M. (2017). Crop Classification with Multi-Temporal Satellite Imagery. Final Reports. Univ. Stanford.
[2] Viskovic, L., Kosovic, I. N., & Mastelic, T. (2019, September). Crop Classification using Multi-spectral and Multitemporal Satellite Imagery with Machine Learning. In 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-5). IEEE.
[3] De Bie, C. A. J. M., Khan, M. R., Toxopeus, A. G., Venus, V., & Skidmore, A. K. (2008, July). Hypertemporal image analysis for crop mapping and change detection. In ISPRS 2008. Proceedings of the XXI congress: Silk road for information from imagery: the International Society for Photogrammetry and Remote Sensing, Beijing, China (pp. 803-812).
[4] Kulkarni, N. M. Crop Identification Using Unsuperviesd ISODATA and K-Means from Multispectral Remote Sensing Imagery.
[5] Doraiswamy, Paul & Moulin, Sophie & Cook, Paul & Stern, Alan. (2003). Crop Yield Assessment from Remote Sensing. Photogrammetric Engineering & Remote Sensing. 69. 665-674. 10.14358/PERS.69.6.665.
[6] Campbell, J. B., & Wynne, R. H. (2011). Introduction to remote sensing. Guilford Press.
[7] Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10(1), 75.
[8] Xia Zhang, Yanli Sun, Kun Shang, Lifu Zhang, Senior Member, IEEE, and Shudong Wang "IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing" Pages 01,Fevruary-2016.
[9] Darve, N. R., & Theng, D. P. (2015). Image processing on eye image using SURF feature extraction. Image, 3(4).