Anomaly Detection across Multiple Farms Using Remote Sensing

Deepak Kumar [1] Harsh Raj [2] Hrishikesh Bodkhe [3] Nitesh Wadhavinde [4]

Department of Computer Science, Indian Institute of Technology Ropar

Springer CCIS ICA 23

Motivation

An anomaly across multiple farms can be defined as a farm whose growth is not the same as its neighbouring farms, with nearby sowing dates. Anomaly detection across multiple farms can help detect early signs of crop diseases, enabling corrective actions before the disease spreads leading to improved crop yields. Remote sensing technologies, such as satellites and drones, can capture images and data on crops and the surrounding environment from a distance, providing farmers with real-time information on crop growth. By making more informed decisions based on anomaly detection through satellite images, farmers can achieve better crop yields, increased profits, and more sustainable agriculture.

Dataset

The data is taken from EOS Data Analytics which gives NDVI data of selected farms over a period of time using SENTINEL-2 with 10m spatial resolution. The dataset contains the NDVI statistical data for 13 farms near the IIT ROPAR region recorded 2-3 times a month for the past 3 years. The data recorded is for the wheat farms. The sowing dates for each farm are added to the dataset and the day of the crop is calculated with respect to its sowing date. The columns' day number and NDVI average values are taken to constitute a training dataset for training the model and it is normalized using Min-Max Scaler. The original dataset contains over 412 rows and 10 columns and the training dataset contains 412 rows and 2 columns.

Table 1. Sample farm data

Farm name	Sowing date	Date	Avg NDVI	Day
farm1	2020-11-20	2020-11-20	0.1942	0
farm1	2020-11-20	2020-12-15	0.4961	25
farm1	2020-11-20	2021-01-27	0.7439	68
farm13	2022-11-18	2023-02-13	0.7753	87
farm13	2022-11-18	2023-02-26	0.6886	100

Data collection and analysis

Figure 1. Visual Wheat growth timeline seen from satellite

Figure 1 shows different images of the same wheat farms near IIT Ropar, taken from the Sentinel-2 satellite. Growth changes can be observed over the whole period from sowing to harvesting. Sowing dates of farms can be different. To balance out the data for training we have used sowing date as a baseline for measuring the growth stages of wheat for individual farms.

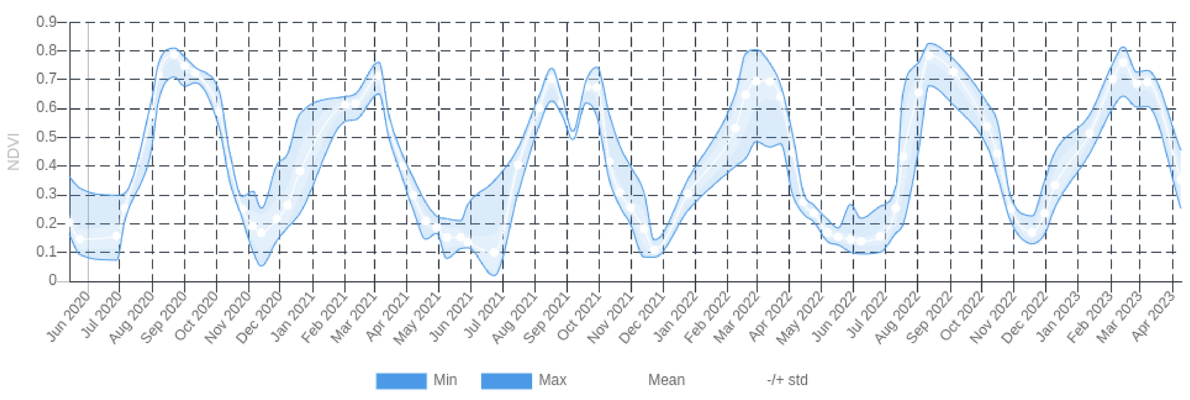


Figure 2. Wheat NDVI timeline of a farm over 3 years

In Punjab, generally, the wheat is sown in November and harvested in April. From Figure 2 it can be observed that the NDVI values start to rise from November because wheat is in its vegetative state. Once wheat reaches maturity, the leaves turn yellow and the plant begins to dry out, known as the ripening or senescence stage then the NDVI values start to fall from March to April. This rise and fall in NDVI values can be used to measure the growth stages of wheat for farms.

Methodolgy

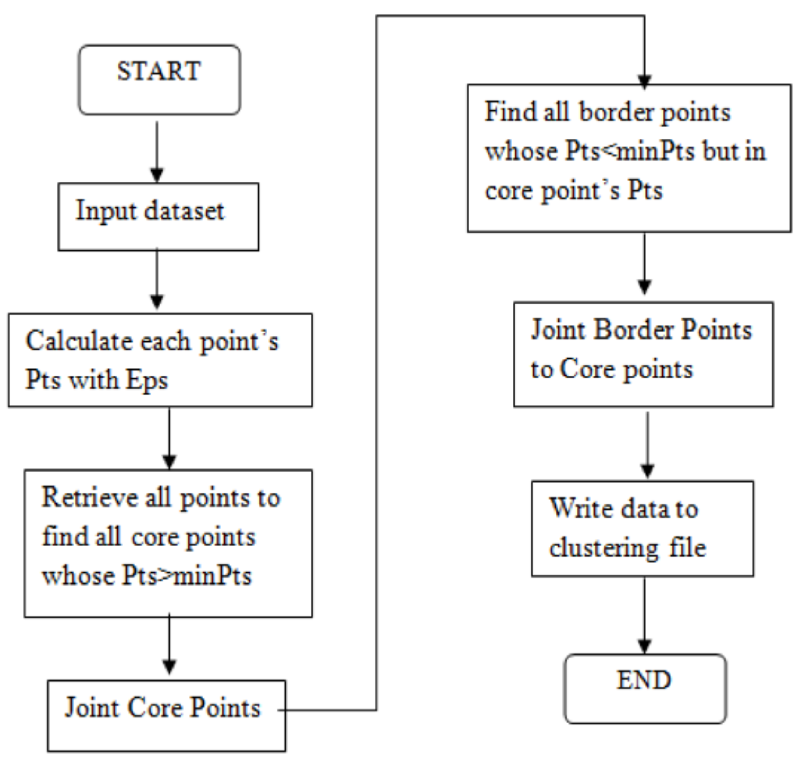


Figure 3. Flowchart of DBSCAN Algorithm [1]

- The training dataset containing crop age (in days) and average NDVI value is created from original dataset.
- The training dataset is standardized using Min-Max Scaler method and the maximum and minimum values for both the columns of the training dataset are recorded.
- The parameters used are: eps = 0.1 and min_samples = 5 and the training data is fit to the model.
- Number of clusters created and labels for each data point assigned by the model are recorded.
- The input data is first normalized using Min-Max Scaler and then put into the predict function which determines the minimum Euclidean distance from any of the core points determined by the DBSCAN model. If the distance is less than the eps used then the input data is not anomalous else it is an anomaly.

Mathematical Formulas Used

The NDVI values from an image are calculated using the given formula: [2]

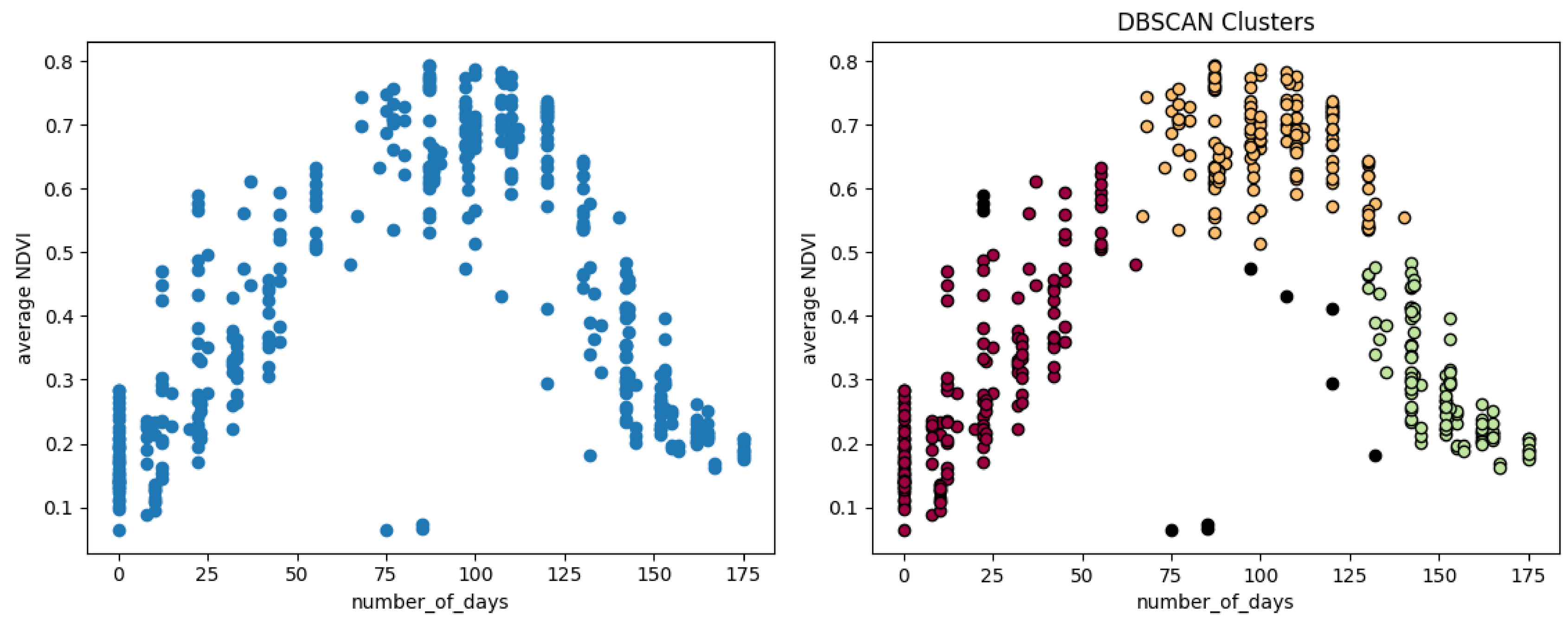
$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \quad (1)$$

where NIR is the near-infrared reflectance and Red is the red reflectance.

Results and Analysis

Model estimates

In the DBSCAN model, the label for each data point and the number of clusters is determined and the dataset is plotted with colors depicting their assigned cluster label. The anomalous data points are considered noise, represented in black. DBSCAN model is evaluated for various epsilon distances and min sample values over a total number of clusters and noise points are generated. The eps value of 0.07 and min sample value of 5 gave 3 clusters and 11 noise points and a silhouette score of 0.5369 was obtained for the above DBSCAN model.



(a) Training Dataset plot (b) DBSCAN clustering for training dataset plot

Figure 4

For further study, the Time series forecasting model named ARIMA is trained for the NDVI values of one farm, and its prediction for the same farm is recorded. The plot for the actual NDVI values against the predicted NDVI values for the farm is observed.

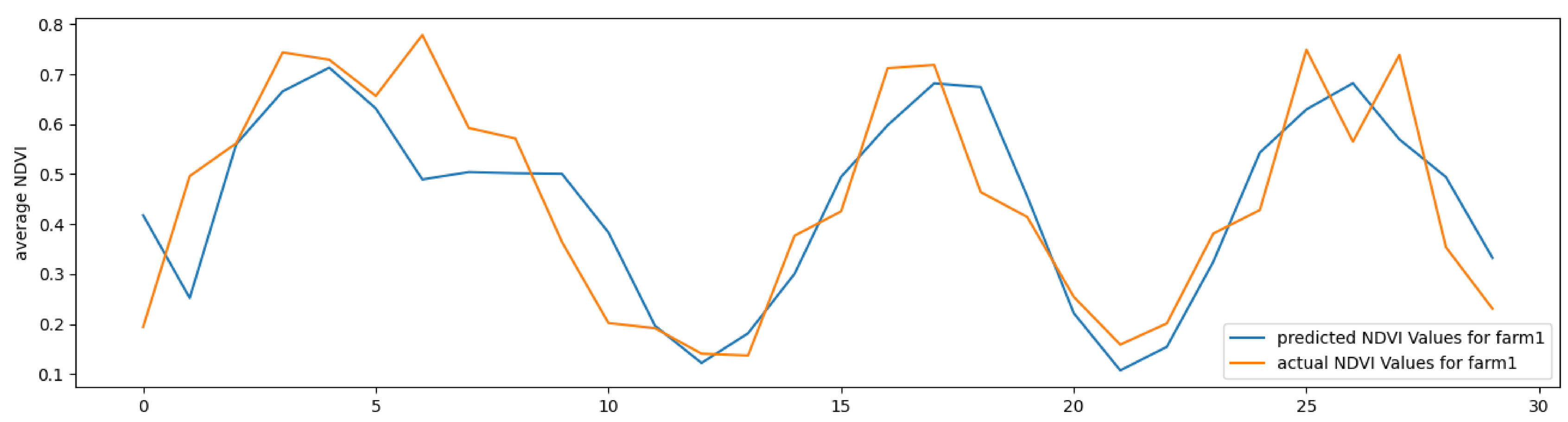


Figure 5. Arima model's test to predict comparison plot

Conclusions

- Anomaly Detection across Multiple Farms through Remote Sensing can help to achieve sustainable agriculture.
- Anomalies can be detected using a clustering algorithm like DBSCAN.
- DBSCAN after proper hyper-parameter tuning, was able to satisfactorily detect the anomalous NDVI value for respective crop age.

What is already known about this subject?

- As per our studies, we have not found any research work that has explored this area of anomaly detection.

What does this study add?

- Opening paths for anomaly detection in farms using satellite images
- Exploring clustering algorithms like DBSCAN, for anomaly detection

Practical implications

- **Precision Agriculture:** By detecting anomalies in crop growth patterns and environmental conditions, farmers can take corrective actions and optimize their farming practices, leading to increased efficiency and productivity.
- **Early detection of Crop Diseases:** Farmers can take immediate corrective action, preventing the spread of the disease and reducing crop losses.
- **Yield Prediction:** By analyzing data on crop growth patterns and environmental conditions which can help farmers make data-driven decisions on resource allocation, pricing, and marketing, thus increasing profitability.

References

[1] Deepak Jain.
Design and implementation of dbscan algorithm for generating clusters in dense region of complex datasets.
10 2017.

[2] Syeda Sultana, Amjed Ali, Ashfaq Ahmad, Muhammad Mubeen, M Zia-Ul-Haq, Shakeel Ahmad, Sezai Ercisli, and Hawa Ze Jaafar.
Normalized difference vegetation index as a tool for wheat yield estimation: A case study from faisalabad, pakistan.
TheScientificWorldJournal, 2014:725326, 06 2014.