Hybridized Clustering Algorithm and Ensemble Learning for Monitoring Paddy Crop Growth Analysis

Sathiya Priya.R
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, TamilNadu, India
sathiyapriya.r2020@vitstudent.ac.in

Rahamathunnisa.U
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, TamilNadu, India
rahamathu.u@vit.ac.in

Abstract: *In food security control and agricultural management, crop monitoring is important to improve the crop yield rate and ensure the paddy growth. The paddy crop is monitored via remote sensing satellite images that cover the entire cultivating region. The satellite images are affected due to satellite movement and human errors, which causes over fitting issues. In addition, the remote images are facing difficulties while managing system robustness. The research issue is overcome by applying the Hybridized Clustering Algorithm-based Ensemble Learning Hybridized Clustering Algorithm with Ensemble Learning (HCA-EL) approach. Initially, the satellite images are handled to remove geometric, radiometric, and atmospheric corrections using pre-processing techniques. This process uses the Ground Control Points (GCP) and histogram process to remove unwanted information. Then vegetation index value data points are obtained to progress the overall crop monitoring. The data points are analyzed using spectral clustering based on the eigenvector and eigenvalue, minimizing the dimensionality issues. Finally, the crop monitoring criteria, such as soil quality and water levels are computed to improve paddy growth. The monitoring process is done with the help of the Ensemble Learning Extreme Networks (ELEN), which uses the voting criteria to identify the quality parameters. The HCA-EL approach-based crop monitoring process was implemented using the Python tool with respective performance metrics like Accuracy, Error Rate, MCC and F1-Score.*

Keywords: *Crop monitoring, satellite images, vegetation index, eigenvector, eigenvalue, hybridized clustering algorithm, ensemble learning.*

Received March 28, 2024; accepted July 15, 2024
https://doi.org/10.34028/iajit/21/4/10

1. Introduction

The agricultural management system's crop progress and monitoring process [24] was important because it used to improve the yield rate and reduce the paddy disease in earlier stage. The automatic crop monitoring process widely supports farmers in increasing their agricultural activities and seasonal yields [20, 22]. The monitoring process covers crop age, growing phases, and yield during the growing season. The automatic systems are developed with additional and customized settings to progress the crop growth prediction rate [47]. The automatic monitoring process uses information technology to access the paddy remotely. During this study, crop spatial information is continuously assessed to handle effective decisions [35]. The Remote sensing techniques monitor the crop health, cultivation areas, production, damage, and loss of information. Frequent crop assessment reduces plant disease and maximizes the production rate. The satellite sensing images track the crop's growth and flaws, minimizing human effort. The paddy images [15] are captured at particular intervals with specific conditions to evaluate the growth cycle. Several researchers use Sentinel 2 [21, 27] imagery images to analyse crop growth in different seasons. The sentinel satellite images, developed seven years with 10 meters of spatial resolution and 60-meter resolution [9, 25] of three spectral and four spectral bands, are highly employed to record the vegetation details. The images are corrupted because of satellite movement and other human-made errors. Therefore, image processing and intelligent techniques are widely applied to investigate crop growth. The noise images are handled using noise-removal techniques to address the atmospheric [40], radiometric [29] and geometric correction processes [45]. The pre-processing techniques remove noise from images and enhance the image quality, which helps to maximize the overall image analysis efficiency. Then researchers practice the segmentation procedures to recognize the vegetation area from noise-removed images. Generally, clustering techniques [23] are utilized to group the vegetation and non-vegetation regions. The clustered details are used to identify the paddy ages and growing stages. However, the existing systems consume high computational challenges while handing a large volume of crop images. In addition, the similarity between vegetation and non-vegetation is difficult to predict due to the green region [48].

The research issues are addressed with the help of the Hybridized Clustering Algorithm with Ensemble Learning (HCA-EL). The approach uses a hybrid clustering algorithms that identify the vegetation and non-vegetation region using a base clustering approach. The spectral cluster is the base clustering approach that identifies the affinity matrix, eigenvector, and eigenvalue that minimize the over fitting issues. In addition, each data point is allocated to a particular group depending on the voting measure. The voting process minimizes the decision deviations, maximizing the overall crop monitoring process. Then soil quality and water levels are accessed depending on the ensemble learning extreme networks. The network process each data, and the weighting parameters are updated depending on the classification accuracy. The clustering and classification algorithm resolves the research issues. Then the overall objective of this study is listed as follows:

1. To progress the crop growth monitoring rate by identifying the vegetation and non-vegetation region from sentinel-2 images.
2. To minimize the deviation between the clustering process using the combination of spectral clustering and ensemble learning
3. To evaluate the efficiency of the crop monitoring using the experimental analysis.

Structure of the proposed work has been organizes as follow:

Section 1 describe about the Introduction part, section 2 describes about the related work for the proposed model in detailed manner, section 3 describes about Proposed Model Hybridized Clustering Algorithm Based on Ensemble Learning HCA-EL for paddy crop monitoring. Section 4 describes about Hybridized Clustering Algorithm (HCA) based paddy region segmentation analysis, section 5 describes about ensemble learning for crop growth analysis, section 6 describes about results and discussions and section 7 describes about conclusion of the proposed work.

2. Related Work

Sethy *et al*. [38] applied the Internet of Things (IoT) and Deep Learning (DL) to monitor the paddy field. The VGG-16 pre-trained neural model classifies the paddy lead disease and estimates the nitrogen status. The DL extracts the deep features, and transfer learning is applied to classify the images. The extracted deep features are processed with the help of a Support Vector Machine (SVM) that classifies the images. Then VGG-16 learning function predicts the nitrogen status and four leaf diseases with 79.86% accuracy. The introduced IoT with the DL model controls the humidity and temperatures effectively. Sharma *et al*. [39] evaluated multi-temporal sentinel-1 SAR information using Artificial Neural Network (ANN) to forecast the paddy yield. The study area is investigated to the Normalized Difference Water Index and Normalized Difference Vegetation (NDVI) based on smart sampling. Then, peak stage signature backscattering coefficients are extracted to compute the yield. The extracted coefficients are processed using the ANN approach that recognizes paddy yield with a 0.72 error rate. Pozzobon de Bem *et al*. [30] recommended Convolution Neural Networks (CNN) to identify the rice crop in Southern Brazil Sentinel-1 time series images. The temporal images are analyzed using the CNN approach to extract the features from rice plantations. This study uses the Vertical-Horizontal (VH) polarized bands, Verticle-Verticle (VV) polarized bands, and VV and VH datasets to monitor the crop. This study attains 0.91% of intersection units and 0.93% of F1-Score while monitoring rice crop plantations.

Devi *et al*. [11] the plant features are derived by hybridized methods such as Grey Scale Co-occurrence Matrix (GLCM), scale invariant feature transform, and wavelet transform. The plant features are processed by various classifiers such as support vector machine, naive bayesian, Multi-class Support Vector Machine, Back Propagation Neural Networks (BPNN), and K-Nearest Neighbouring (KNN). Among the various classifiers, Multi-class SVM attains high recognition accuracy. Ramesh *et al*. [32] identified and classified paddy leaf disease using the Jaya Optimized Deep Neural Networks (JODNN). The captured paddy images are converted into HSV images which are processed with the help of the clustering approach that segments the normal and disease-affected regions. Then features are extracted that features are analyzed with the help of an optimized neural model. The neural model recognizes the plant diseases with 92% of sheath rota and 90.57% of normal leaf images.

Rammohan *et al*. [33] utilized Recurrent Neural Networks (RNN) and Deep Convolution Neural Networks (CNN) to forecast the crop yield. This research forecasts the crop yield depending on the crop genotypes, weather, yield output activities, and soil. According to the factors, genetic development span and environmental factor dependency are designed to minimize the challenges in the untested setting environment. Then the back propagation neural model is applied for the learning process that maximizes the crop yield prediction rate up to 95.5%. Geetha *et al*. [14] predicted the Cauvery delta region paddy yield using Stacked Long-Short Memory Neural Networks (SLSTM). The paddy images are collected from the Cauvery region according to environmental factors. The collected images are processed using LSTM, identifying the paddy crop yield with minimum validation and training loss. Divakar *et al*. [12] designed Deep Feature Learning (DFL) model to map the rice yield. The DFL uses the convolutional LSTM to forecast the crop yield from Satellite Imagery. The convolution layer analyses the temporal and spatial features to predict the output. This analysis uses the Principal Component Analysis (PCA) to reduce the feature dimensionality, minimizing the computation cost. The selected features are classified using DFL that recognizes the yield with maximum accuracy. Joshua *et al*. [18] explored different machine learning techniques to predict the paddy yield in the Tamil Nadu Eastern part. This study uses the Back Propagation Neural Networks (BPNN), Support Vector Regression (SVR), Radial Basis Functions (RBFNN), and General Regression Neural Networks (GRNN) to predict the crop yield at the Cauvery delta zone. These methods recognize the yield rate with minimum error rates and high recognition accuracy.

Prakash *et al*. [31] applied Synthetic Aperture Radar (SAR) to estimate the paddy area in the Cauvery Delta Region. The research study uses the Cauvery region's Samba Season 2020 to 2021. The collected SAR information is processed to extract the multi-temporal features. The extracted features are accessed with the help of the software that identifies the paddy areas with maximum recognition accuracy (93.4%) and Kappa Index (0.87%).

Mandal *et al*. [26] 2020 monitor crop progress from dual polarimetric radar VI using SAR data. This study uses the Canada region information to identify and monitor paddy growth. The region is frequently examined regarding vegetation water content, dry biomass, plant area index and other biophysical variables. The extracted features are processed, and correlations are examined to predict crop yield. Ali *et al*. [4] monitored rice cultivation from the Leaf area index and sentinel-2 data. This research uses the 3240 acre to analyse the cultivation efficiency. The collected sentinel-2 imagery information is explored to get the Normalized Difference Vegetation Index (NDVI). The extracted features are processed using the K-Nearest Neighbouring Method, which predicts the cultivation rate with up to 0.93% accuracy.

Land Cover and Land Use is a consequential process when it come through Remote sensing technology. Azedou *et al*. [6] secern the six category of land use by using the Sentinel 2 satellite imagery. By comprising the five spectral indices like Soil Adjusted Vegetation Index (SAVI), Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI), Green Normalized Difference Vegetation Index (GNDVI), and Shadow Index (SI) the LCLU. The DNN model has outperformed well with 94.5% of accuracy level. And a comparison is made between Random Search, Hyper band and Bayesian Optimization methods. Each of these optimization methods have their own pros and cons. In achieving the highest accuracy level with low time consumption Hyper band is the optimal method. According to the various researcher's thoughts, the paddy crop monitoring is performed from Sentinel 2 images using various deep learning and image processing techniques. The traditional methods consume high computation time to analyse the time series and continuous data processing. In addition, traditional approaches attain minimum prediction accuracy while exploring a large volume of data. The research issues are overcome by applying the Hybridized clustering algorithm based on Ensemble Learning HCA-EL to improve the crop monitoring.

3. Hybridized Clustering Algorithm Based on Ensemble Learning (HCA-EL) For Paddy Crop Monitoring

Frequent crop monitoring increases the yield ratio and reduces paddy diseases in agriculture activities. This study mainly intends to maximize the yield and minimize the computation difficulties while analysing a large volume of data. The study objective is achieved with the help of Hybridized Clustering Algorithm based on Ensemble Learning HCA-EL. The HCA-EL analysis focuses on the Manson season growth using the satellite sentinel 2 images. The satellite images cover the land region used to identify and monitor the crops. The Sentinel-2 images cover crop health, crop growth, production, and weather patterns used to analyse crop yields. The captured land area images are analyzed for geometric, atmospheric, and radiometric corrections. These corrections minimize the computational difficulties while processing the noise satellite images. Then clustering is applied to explore each region to predict crop monitoring. Then the working process of the HCA-EL approach is illustrated in Figure 1.

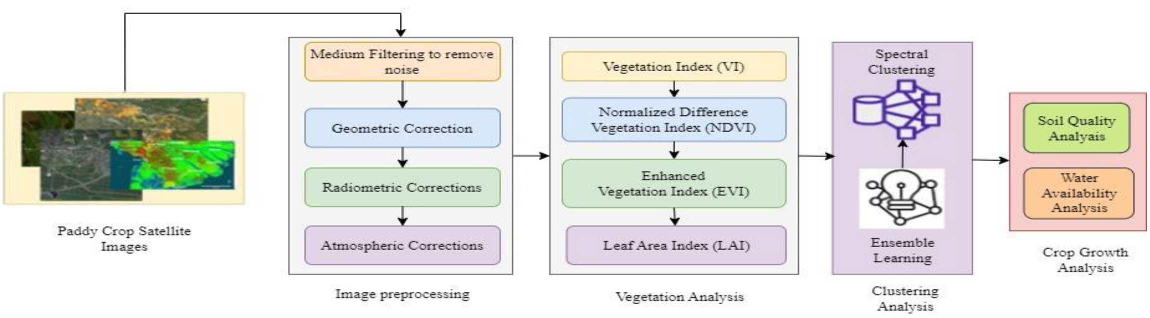


Figure 1. Overall working process of HCA-EL Paddy crop monitoring.

3.1. Data Collection

The research analysis uses real-time paddy information to perform crop monitoring. The paddy satellite images are collected from Tamil Nadu (TN) State, Thanjavur district, located in the Cauvery Delta Region (CDR) [3]. In the entire TN, 3, 39, 657 area covered by Thanjavur district and located in $10°8' \, to \, 11° \, 12'$ in North Latitude and $78°48' \, to \, 79° \, 38'$ in East Longitude. Thanjavur district is located at the TN centre, and it has Pudukkottai in the west, Nagapattinam in the Northeast, River Kollidam in the North, and Tiruvar as neighbouring district in the east. In this study, Thanjavur is choosen due to the best rice cultivating area in Tamil Nadu and the large volume of food grains cultivated in this region. Southwest Monsoon Thanjavur receives 938mm of annual mean rainfall from June to September, while the Northeast monsoon has 462mm. The region has higher rainfall in the northeast monsoon than in the southwest. This region cultivates different paddies such as Kuruvai (May to August), Samba (August to February), and Thaladi (February to March and September to October). Therefore, the Thanjavur region is chosen to examine paddy growth to maximize yield. During the analysis, Sentinel-2 images are captured at 12 days intervals from May to October (2019). The paddy images are collected during the Rabi Season at numerous growth stages. Nearly 250 rice and non-crop points are captured and processed by a fusion technique called hybridized clustering and ensemble learning process. Then the study area utilized [41, 43] in this work is illustrated in Figure 2.

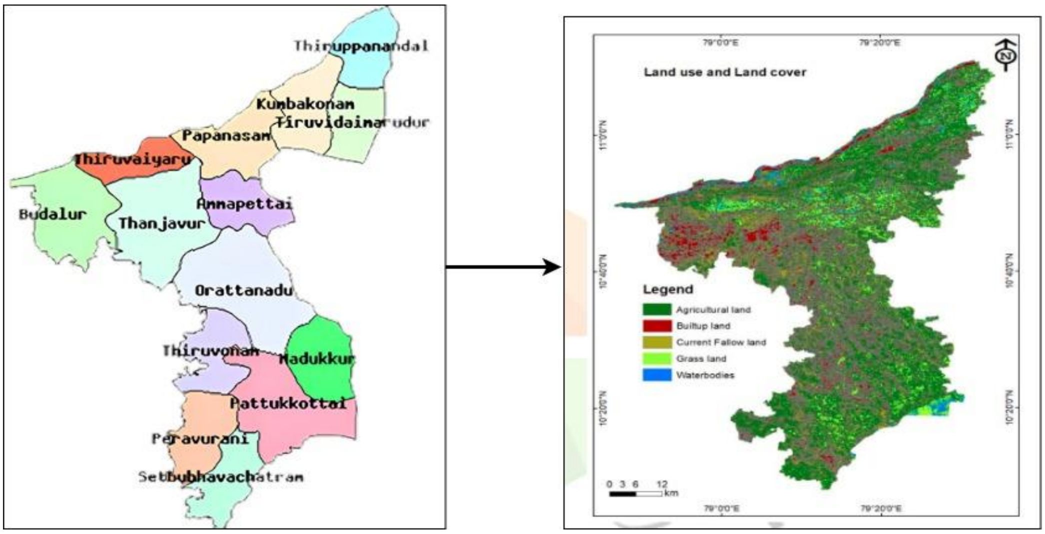


Figure 2. Thanjavur district land use and land cover image.

The study area is examined frequently, and sentinel images are captured, as shown in Figure 2 which is analysed using image processing and clustering algorithm to monitor crop growth.

3.2. Pre-Processing

Pre-processing steps process the gathered sentinel 2 images to remove the irrelevant information such as corrupted pixels and low quality pixels. The noise removal process simplifies the paddy crop monitoring and improves the yield rate. Initially the collected images processed with the help of median filter [17] to remove the inconsistent information. The median filter analyses each pixel and the corrupted information is replaced with the help of median value. The median filtering process simplifies the image analysis and computation. The noise removal process explores the images to make atmospheric, radiometric, and geometric corrections. When it comes to restore the missing pixels in a damaged images Image Inpainting technique is used [42].

Generally, the images are affected by satellite movement, which requires geometric corrections to monitor the crops. Then radiometric corrections are done to rectify the digital number deviations to the Atmospheric top radius (ToA). The European Space Agency Sensor Network Asynchronous Processors sensor, is utilized to accomplish the atmospheric corrections. The captured Sentinel 2 images are handled based on the Ground Control Points (GCP) [16] to make the geometric corrections. The GCP has pair points (u, v) denoted as sources and reference coordinates; that have directions (ex: north-south). The source coordinates are mentioned as (c, r); c is column, and r is row. The image distortion is computed using the source and references GCP difference. The GCP represents the image's ground features, which hang on the spatial resolutions of images. Then the coefficient value is estimated to address the polynomial coefficient relation problems. The image coordinate (u, v) and object coordinates are denoted as (x, y, z) computed from the exterior and interior location orientation parameters. Then the sentinel-2 image coordinate is estimated using Equation (1)

$$\left.\begin{aligned} u &= f * \left(\frac{x}{z}\right) + u_0 \\ v &= f * \left(\frac{y}{z}\right) + v_0 \end{aligned}\right\} \tag{1}$$

In Equation (1), image coordinates are denoted as (u, v), object coordinate (x, y, z) is used while capturing the land images, camera focal length is f, principal point image coordinate is (u_0, v_0). According to the computation, the image collinearity equation is framed to identify the relationship between the coordinates. Then polynomial transformation is constructed to perform the geometric correction. Then the polynomic transformation is defined as the same as Equation (2)

$$\left.\begin{aligned} u' = a_0 + a_1 * u + a_2 * v + a_3 * u^2 + a_4 * u * v + a5 * v^2 + \cdots \ldots \ldots \\ v' = b_0 + b_1 * u + b_2 * v + b_3 * u^2 + b_4 * u * v + b5 * v^2 + \cdots \ldots \ldots \end{aligned}\right\} \quad (2)$$

In the Equation (2), corrected image coordinates are denoted as(u', v'), original image coordinates are (u, v), and the polynomial transformation coefficients are denoted as a_0, a_1, a_2..................b_0, b_1, b_2,. After that residual has to be computed to perform the geometric corrections. The residual value (R) is calculated using Equation (3).

$$R = \sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2} \quad (3)$$

In equation (3), the Corrected object GCP coordinate is denoted as (x', y', z') and known ground coordinate GCP value is (x, y, z). The computed R improves the image quality and rectifies the geometric correction. The coordinates of source images are generate the transform matrix called rectification. The rectification process performs the image distortion. After performing the geometric correction, error values are computed to minimize the variations between the reference and source coordinates. The error value is computed for inverse and transforms coordinates using the Root Mean Square Value (RMSE) [8]. The inverse transform value obtains from reference coordinates. The error value is estimated using Equation (4).

$$Root\ Mean\ Square\ Value\ (RMSE) = \sqrt{(x_r - x_o)^2 + (y_r - y_0)^2} \quad (4)$$

In Equation (4), the GCP coordinates to achieve the geometric corrections. During the computation, maximum deviation coordinates are allocated to new grid and the output images are generated as columns and rows. Then radiometric corrections are made to maximize the paddy growth assessment efficiency. During this process, ground source image properties are examined in various areas. The collected images are sent to the sensor for converting the ToA with raster data type and scale. Then dark area images are identified using an image histogram approach. After that, cosine Tz is utilized to compute the atmospheric corrections. The Digital Elevation Model (DEM) to compute the topographic modification of images [5]. Digital Elevation Model plays a vital role plays predicting the vegetation requirement satellite image. Here, image reflectance is computed to perform the atmospheric corrections. The reflection value is computed using Equation (5).

$$Reflectance = \frac{\pi * Radiance * d^2}{ESUN * \sin\theta} \quad (5)$$

In Equation (5), d is denoted as distance, earth-sun is represented as (ESUN), and θ is signifies as the sun elevation angle captured during the image capture. In addition, the image irradiance value is computed from the particular wavelet band ESUN related solar irradiance value. According to (5), the reflectance value is computed from 0 to 1, and the default reflectance value is 32-bit. This process computes the scattering and absorbing point, effectively reducing the atmospheric and outside surface details. Then the pre-processed images are continuously examined to extract the vegetation-related information because it is used to monitor crop growth.

3.3. Description about Sentinel 2 Dataset

The sentinel 2 satellite is Multi Spectral Instrument (MSI) which as High Spatial Resolution ranging from 10m to 60m to monitor the land surface. It has both temporal and Spectral characteristics. The revisit period of this satellite is 5 days with cloud free conditions. The Sentinel 2 satellite provides variety of services like monitoring the land, agriculture, risk mapping and disaster control [34]. The Sentinel 2 satellite consists of 13 spectral bands from B1 to B12 each of these has different resolution, wavelength and specific features The spectral bands meter ranges from 10 to 60-meter pixel size. Sentinel 2 data is used to highlight the different parameter for the plant growth like leaf area index, chlorophyll and water content. Sentinel 2 is used for agriculture application such as monitoring the crop and its growth Sentinel 2 is used to monitor the drought changes and soil degradation problem. The sentinel 2 band details are illustrated in Table 1.

Table 1. Sentinel 2 band details [37].

Band Number	Band Name	Resolution Meter (M)	Central Wavelength (Nm)
B1	ULTRA BLUE	60m	433nm
B2	BLUE	10m	490nm
B3	GREEN	10m	560nm
B4	RED	10m	665nm
B5	VNIR	20 m	705nm
B6	VNIR	20 m	740nm
B7	VNIR	20 m	783nm
B8	VNIR	10 m	842nm
B8a	VNIR	20m	863nm
B9	SWIR	60m	940nm
B10	SWIR	60 m	1375nm
B11	SWIR	20 m	1610nm
B12	SWIR	20 m	2190nm

*NOTE
VNIR: Visible Near Infrared
SWIR: Short Wave Infrared

3.4. Calculation of Vegetation Area

Monitoring the paddy crop through remote sensing technology has a significant role, the author in the paper [36] by using the Vegetation Indices (VI) monitor the paddy crop growth in three different stages, like the first stage of the paddy is known as Tillering, second stage is Flowering and last third stage is ripening. Each of these stage is identified at the time interval like 40 days, 60 days and 100 days. Apart from this to highlight the paddy crop growth health at each of these stage the author utilizes the crop growth. The stages of paddy crop are categorized into Vegetative, Reproductive and Ripening. From the initial stage of cultivation of the crop to the final stage of yield based on the type seasons like Rabi and Kharif the estimated is done. The growth of the crop is known as paddy age. Normalized Difference Red Edge Index (NDRE). A particular area is selected by the geospatial technology which helps to ensure the maximum yield of the crop. Different parameters are included to continuous monitor the growth of the paddy crop like soil moisture, precipitation, and temperature and climate changes. To calculate the vegetation area for the Thanjavur district it is important to know the stages of paddy.

- **Vegetative**: in this stage the seeds are sown in the field, when seedlings occur, they are transferred into the paddy field. Tillering occur after 40 days of planting. In tillering the nodes are completely branched together.
- **Reproductive**: in this stage, Heading will occurs in 60 days. Heading will be depending upon the type of paddy crop sown, in the heading stage a portion of the panicle is found at the rice stem.
- **Ripening**: it is the last stage of the paddy crop where the panicle is cutted and gathered together as stalks. And this stage is also known as Reaping.

When it comes to Summer Autumn season it takes 100 days for Harvest. And in Winter Spring season it takes 90 to 140 days for Harvest. The pre-processed sentinel images are analysed different kinds of spectral indices which are used evaluate growth of the crop. The spectral Indices are

- Vegetation Index (VI)
- Leaf Area Investigation (LAI)
- Normalized Difference Vegetation Index (NDVI)
- Soil Adjusted Vegetation Index (SAVI)
- Normalized Difference Water Index (NDWI).

These computed index values are utilized to monitor paddy growth. The LAI, EVI, VI, NDVI, SAVI, and NDWI identify the paddy stages and maximize the classification accuracy. The VI estimates the spatial and temporal information that quantifies the vegetation biomass value from the satellite image pixels. The VI value is predicted depending on the plant characteristics like soil background, confounding factors, and atmospheric effects. VI bands are designed to highlight the various vegetation properties and characteristics, it is mostly used for agriculture purpose to monitor the plant health and its growth. The NDVI images are obtained, by using the spectral band reflection of light in near infrared and red band, a colour coded images is obtained for the vegetation covered area. Chlorophyll pigment is the main reason where plants prepare photosynthesis and it absorbs the light in blue and red bands reflects in NIR. In the case of Non vegetation area and bare land the reflectance of NIR will be very low due. The NDVI can highlight the vegetation cover area by subtracting the reflectance value B8 (NIR) and B4 (Red) and dividing the result with the sum of B8 and B4. The obtained result values is mapped to the colour scale where the red represent the low NDVI and the green represent the high NDVI. Which is used for the visual representation of the vegetation covered area The NDVI is used to identify the crop growth, health and its productivity. The NDVI is used the measure the Vegetation greenness for the particular area. Higher the vegetation greenness implies dense vegetation area and lower the vegetation greenness implies sparse vegetation area. And this is indicated by the value of -1 to 1. If the captured images has an NDVI value ranges from 0.036 to 0.240, then it is very minimum vegetation, the paddy is in the beginning stage of vegetation. If the captured image has a 0.0240 to 0.456 NDVI value, then it is low vegetation. If the captured image has a 0.456 to 0.652 NDVI value, then it is medium vegetation, and it is in the Reproductive. If the captured image has a 0.652 to 0.884 NDVI value, then it is high vegetation, the paddy last phase known as Ripening. The NDVI value for the crop growth is illustrated in the Table 2. The NDVI is calculated by using the Equation (6), EVI is almost familiar with NDVI the EVI is used to calculate the vegetation greenness and does the atmospheric correction and the EVI is used to remove the canopy background noise in dense vegetation area calculated by using the Equation (7) and VI is calculated by using the Equation (8).

$$NDVI = \frac{NearInfrard\ Radiation\ band\ (NIR) - Red\ Band\ (R)}{Near\ Infrared\ Band\ (NIR) + Red\ Band\ (R)} \quad (6)$$

$$EVI = G * \left(\frac{(NIR - R)}{(NIR + C1 * R - C2 * B + L)} \right) \quad (7)$$

$$VI = \frac{NearInfrared\ Radiation\ band\ (NIR)}{Red\ Band\ (R)} \quad (8)$$

Table 2. NDVI value.

NDVI Values	Indicates
Below 0	Indicate water or other non –vegetation
Between 0 to 0.3	Barren area or less likely to non-likely vegetation
Between 0.3 to 0.6	Sparse
Between 0.6 to 0.9	Healthy and dense vegetation cover
Above 0.9	Very dense vegetation cover

In addition to this, the LAI value is estimated using Equation (9). The LAI is the ratio of the plant's total area of leaves and plant ground area. The LAI is the representation of the canopy resistance and biomass value.

$$LAI = -\, ln^{\left(\frac{0.069 - SAVI}{0.59}\right)} \Big/ 0.9 \quad (9)$$

In Equation (9), Soil Adjusted vegetation index is defined as SAVI and computed using Equation (10).

$$SAVI = \left(\frac{(NIR-Red)}{(NIR+Red+L)}\right) + (1+L) \quad (10)$$

Soil is the most important factor, for the growth of the crop, soil health and soil quality are considered as a finite and non-renewable resources in agriculture [19], and a healthy soil is identified by growth of the crops and its nutrition's properties which are inherited by the crops from the soil. Alluvial soil is mostly suitable for agriculture purpose, which is fertile and loamy where the pH range of the soil is from 3.5 to 10, the pH ranges from 5 to 7 in higher rainfall area.

The SAVI is used to strive the vegetation index by minimizing the soil brightness and its correct the low vegetation area. SAVI is mostly used in the arid region where vegetation cover is low to identify

In Equation (10), the L implies the green vegetation cover, the value of L depends on following when L=0.5 (Moderate Green Vegetative Cover), L=0 (High Vegetation Cover) and L=1 (No green Vegetation Cover).

Water is the most basic thing to produce a healthy crop. In India 17% of GDP is contributes by agriculture in which the irrigated agriculture serves 20% and the total cultivated land serves 40% [2]. Our Globe is canopied by 71% of water but the availability of fresh water is less than 1%. By efficiently utilizing and managing the water resources can leads to progressive development of our country [1].

For agriculture ground water is the most important criteria to supply water for the field decreases in groundwater leads to depend on the rainfall and other modes of the water for the field. To tackle this issue the author in the paper [13] proposes a new concept known as Groundwater Dependent Vegetation (GDV) and this concept is used to identify the groundwater by using Sentinel2 data. The wet area which remains green in dry period is identified. The hydrogeological parameters are taken into consideration for deriving the accurate result. The vegetation area for agriculture purpose with groundwater is identified 90% using GDV.

The NDWI is used to enhance the features of the water bodies in the satellite image, the primary aspect of using the NDWI is used to detect the water and monitor the changes in the water bodies. The NDWI uses NIR and SWIR spectral bands.

The NDWI can be evaluated by the following ranges from 0.0 to 0.2 indicate Flooding and Humidity, from 0.2 to 1 indicate water surface Area.

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \quad (11)$$

The estimated vegetation area is denoted as γ; (VI, NDVI, EVI, LAI, SAVI, and NDWI) and other areas are relevant to non-vegetation area which is denoted as β.

$$Total\ vegetation\ calculation = \gamma - \beta \quad (12)$$

The above computed vegetation index values, phases, and crop ages are frequently computed. Then, the paddy index value is computed from the paddy region. The paddy region is segmented with the help of the Hybridized Clustering Algorithm (HCA). The detailed description of HCA is discussed as follows.

4. Hybridized Clustering Algorithm (HCA) Based Paddy Region Segmentation Analysis

The next important step is to segment the paddy region from the gathered images. The extracted vegetation index information improves the overall paddy region segmentation. The analysis uses a combination of clustering algorithms like Spectral Clustering [46] and Ensemble Learning [28]. This helps to improve the clustering process. Spectral Clustering exhibits good performances levels in the classification of images. Ensemble Learning are used in multiple domain, but when its used in image classification process the success ratio of this model is higher, in training the base model and combining the methods.

The ensemble learning image clustering process uses the ensemble technique to maximize the clustering efficiency while segmenting the paddy region. The hybridized clustering algorithm ensures the paddy crop monitoring system's accuracy, reliability, and robustness. The pre-processed and vegetation feature extracted images are considered segmentation inputs. Then the clustering process is illustrated in Figure 3. Initially, the base clustering is performed using the spectral clustering. The spectral clustering approach uses the spectral properties of the image to perform the clustering process. Considering the image has a set of data points that are denoted as $X = \{x_1, x_2, x_3, \dots . x_n\}$. The data points are analyzed for constructing the affinity matrix W. The W computes the pairwise dissimilarities and similarities between the data points. The similarity and dissimilarities are estimated based on the Gaussian similarity [44]. Then the W computation is defined as:

$$W_{ij} = sim(x_i, x_j) \quad (13)$$

Where, sim is denoted as similarity measures between the data points.

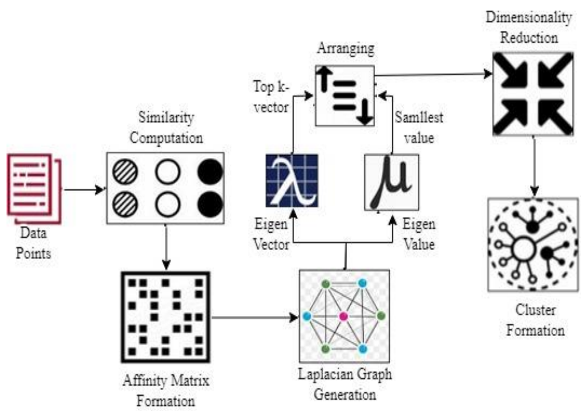


Figure 3. Working process of HCA clustering.

The computed W value is analyzed to derive the Laplacian graph [7], which helps identify the paddy region. This computation constructs normalized Laplacian to determine the paddy region. According to the spectral clustering variant, the graph Laplacians are selected. Then the Laplacian is represented as:

$$L = D - W \quad (14)$$

D is represented as the diagonal matrix of each data point (summation of respective column or row in the affinity matrix W). The L is analyzed to get the eigenvalue and eigenvectors for performing the eigenvalue decomposition. The eigenvalue decomposition is defined as:

$$L\mu = \lambda\mu \quad (15)$$

μ is denoted as an eigenvector, and the eigenvalue is λ. The extracted eigenvectors are arranged to get the top-k eigenvectors related to the smallest eigenvalue. The selection top-k eigenvector reduces the data dimensionality. Then the dimensionality reduction process is denoted as $X' = [\mu_1, \mu_2, \ldots . \mu_k]$ dimensionality reduced data matrix is denoted as X'. Finally the data points are analyzed, forming clusters in which similar data is assigned to similar groups. Then the clustering process is defined in Equation (16).

$$Cluster = arg\,min \sum \left\| x_i - \alpha_j \right\|^2 \quad (16)$$

In Equation (16), the centroid of the cluster is denoted as α_j and the data point from the dimensionality-reduced matrix is represented as x_i. The spectral selection clustering in ensemble clustering maximizes the overall region identification accuracy. The base spectral clustering process is applied multiple times on data with different variations in initialization strategies, random seeds, and algorithm parameters. The multiple-run process attains the solution and captures the diverse outcomes. The random seeds are nothing but the starting condition of the clustering process, initialization strategies include the data distribution process, and algorithm parameters are utilized to run multiple times to get the paddy region by applying the base algorithm. The obtained cluster solution is combined with the ensemble process to get the ensemble solution. The cluster assignment combination is performed with a majority voting process. Each data point is allocated to the cluster and gets the votes across the ensemble. Then consensus clustering is discovered to obtain a robust and stable clustering solution from the ensemble. The consensus process combines the solution from clustering and gets the optimal solution. The spectral clustering and ensemble classifiers identify the green region from the data dissimilarity or similarity computation. The formed ensemble solution maximizes the clustering efficiency and reduces high-dimensional data analysis difficulties. In addition, vegetation and green region information, water information, and soil level must be analyzed to identify crop growth. Then the overall clustering process is illustrated in Table 3.

Table 3. Hybridized clustering algorithm.

Step 1	Collect all crop region satellite images for thanjavur district
Step 2	Pre-process all the satellite images and correct the Geometric, Radiometric and Atmospheric errors.
Step 3	The Vegetation calculation is done by using the Spectral Indices like NDVI, VI,EVI,LAI,SAVI and NDWI which is given in the Equations (6,7,8,9,10 and 11)
Step 4	Apply Hybridized Clustering Algorithm (Spectral Clustering and Ensemble Learning)
Step 5	Applying Spectral Clustering Process Step 5.1. For all pixel compute the similarity value to form the initial cluster Step 5.2. Compute the affinity matrix from similarity values by using the Equation (13) Step 5.3. Estimate the graph Laplacian matrix by using the Equation (14) Step 5.4. Obtain the Eigen Vector and Eigen Value by doing the Eigen decomposition Step 5.5. Generate multiple set of original input to obtain the Eigen value and Eigen Vector Step 5.6. Dimensionality reduction is done X' to form a cluster Step 5.7. The spectral clustering and ensemble learning uses the voting process to figure out the paddy crop growth.
Step 6	Applying ensemble learning Step 6.1 Ensemble Learning and Elman Networks uses booting technique to identify soil quality. Step 6.2. The training and testing based on voting process, the network parameters are assigned weight and bias, to store the best value by using the Equations (17), (18). Step 6.3. Based on the highest voting the soil and water level are identified.
Step 7	By combining the Spectral clustering and Ensemble Learning, the crop growth monitoring is done.

5. Ensemble Learning for Crop Growth Analysis

The last step of this work is crop growth analysis, which is done by applying the Ensemble Learning Elman Networks (ELEN). The ELEN uses the boosting techniques in the neural model to determine the relationship between green vegetation and soil quality. The network randomly selects the hidden node's weights and bias values. The ensemble learning process is used to train the parameters dynamically based on certain conditions in training, and the testing is performed according to the majority voting. The ELEN uses cross-validations that minimize the over fitting issues, directly increasing prediction accuracy and stability. The clustered input data points are decomposed into subsets R, which has a bias b_i and weight w_i. The R learners are trained using (R-1) subsets to validate the other information. During the validation, network parameters such as $\widehat{w}_i$ and $\widehat{b}_i$ assigned to bias b_i and weight w_i. This validation performs for R trails and output accuracy CA and weight value $\|\hat{\beta}\|$ is computed. The calculated parameters, such as CA and weight value $\|\hat{\beta}\|$ stores the best value in every learning stage, and network parameters are updated to minimize the deviations between the outputs. The data points are frequently analyzed in every iteration to identify the soil quality and water level. Suppose, in K iteration, the network attains the high accuracy or small norm of $CA\ and\ \beta$ causes to assign the network weight and bias value to the $\hat{w}_i$ and $\hat{b}_i$. Then compute the mean classification accuracy on the R subset and stores the value in CA and β^k norms which are computed as follows in Equations (17) and (18).

$$CA^k = \left(\frac{1}{R}\right)\sum_{r=1}^{R} CA_r^k \tag{17}$$

$$\|\beta^k\| = \left(\frac{1}{R}\right)\sum_{r=1}^{R} \beta_r^k \tag{18}$$

The computed CA^k value is compared with CA ($CA^k < CA$) and $\|\beta^k\| > \|\hat{\beta}\|$ then the parameters are updated as $w_i^k = \hat{w}_i$ and $b_i^k = \hat{b}_i$. For every input data point, the output is obtained with the help of the training and network parameters. According to the output value, the decision is taken to identify the soil and water levels. In the decision process, if $\|\beta^k\| > medin\ \{\|\beta\|\}$ then $v_{k,c}^r$ is zero. From the ensemble computation, the total vote received by the model is computed as:

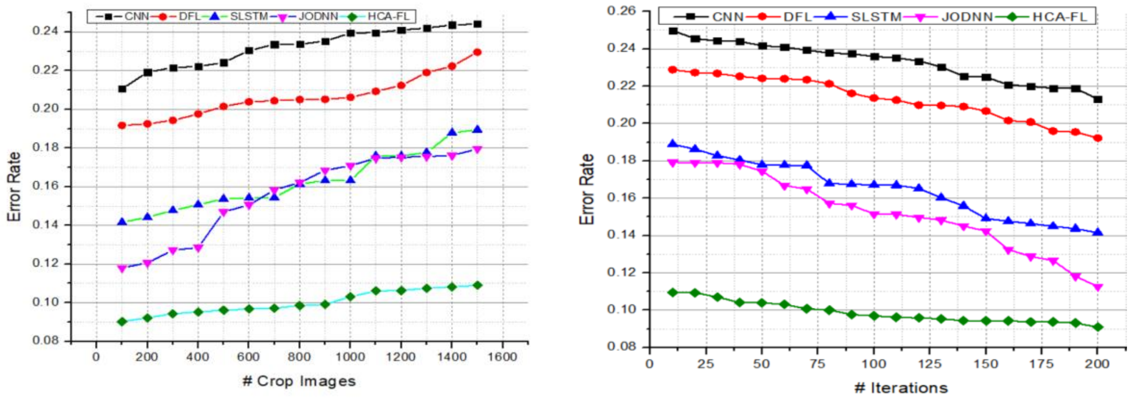
$$V_c = \sum_{r=1}^{R}\sum_{k=1}^{K} v_{k,c}^r \tag{19}$$

The class with the highest vote has the predicted label, which means the data points related to regions are high soil quality and water level. Then the Hybridized clustering approaches are applied using spectral clustering and ensemble learning processes to identify the ground truth value. According to the truth information, the paddy and green vegetation regions are identified. Then the clustered region input data was fed into the ensemble learning extreme network to predict the output value. The ELEN uses the voting concept to identify the high cultivation region soil and water level. The network uses the training and validation for every subset that identifies the relationship amid the reference and ground truth values. At last, crop growth level is predicted to progress the crop monitoring efficiency and harvest crops are predicted.

6. Results and Discussions

This section examines the efficiency of Hybridized Clustering Algorithm based on Ensemble Learning HCA-EL based paddy crop monitoring process. The HCA-EL approach uses the Thanjavur district Sentinel 2 satellite images to explore crop growth. The Sentinel-2 dataset [10] analyzes the system's excellency. The collected satellite images are split into 80% of the training set and 20% for testing purposes. Various vegetation information like EVI, NDVI, LAI, and other information are extracted that are more useful to identify the green region. The vegetation data points are explored using spectral clustering that determines the relationship between each data point. The data similarity, dissimilarities, eigenvector, and eigenvalues are computed to form the cluster with minimum inter and intra-correlations. The above similarity computation clustering is formed to reduce the crop monitoring process. Finally, ensemble learning is applied to attain the ensemble solution for various clusters used to attain the reliability and robustness solution. The cluster region is further examined using ensemble learning Extreme networks that decide according to voting criteria. Here, the decision is taken to identify the soil quality and water level. These two characteristics help improve paddy crop monitoring. Then the efficiency of the system is evaluated using the difference metrics such as accuracy, error rate, Mathew correlation coefficient and F1-score.

The captured sentinel images are handled according to the GCP that removes the geometric corrections in the images. The GCP is computed for both source and reference images that identify the deviations in the image, maximizing the overall image quality. In addition, atmospheric and radiometric corrections maximize the overall crop monitoring prediction rate. This process uses the histogram, reflectiveness, and elevation process that identifies the relevant and irrelevant details which reduce the image quality. The effective computation of each pixel reduces the changes in the image and increases the crop growth prediction accuracy. The efficiency of the introduced system is compared with existing methods, such as Convolution Neural Networks (CNN) [30], Jaya Optimized Deep Neural Networks (JODNN) [11], Stacked Long-Short Memory Neural Networks (SLSTM) [14], and Deep Feature Learning (DFL) [12]. Then the obtained error rate analysis of various iterations and the number of crops are explored, and the graphical analysis is shown in Figure 4. Figure 4 demonstrates the error rate analysis of the HCA-FL crop paddy monitoring process. The system Excellency is evaluated on various crops and iterations in which HCA-FL ensures the minimum deviations between the output values. The GCP data points of source and reference images are explored frequently, which helps to correct the geometric, atmospheric, and radiometric errors in sentinel images.



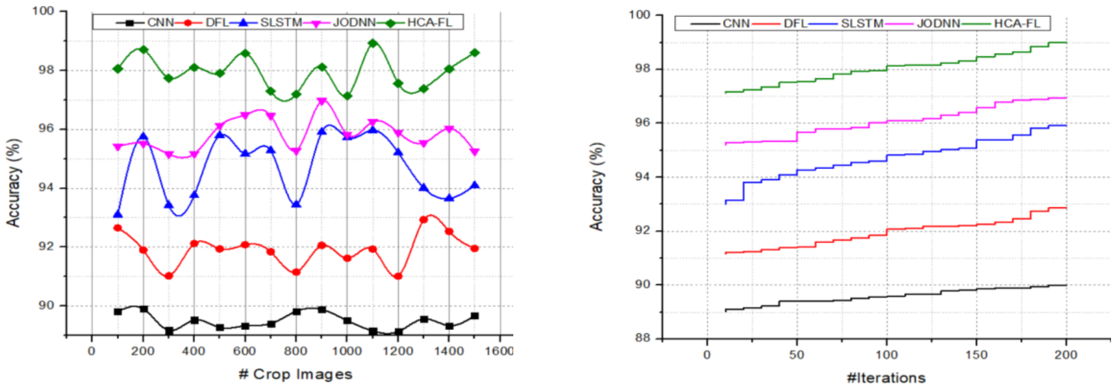
a) Paddy crop.

b) Iterations.

Figure 4. Error Rate Analysis of HCA-FL.

The histogram analysis of the images improves the image representation used to identify the vegetation index with maximum recognition accuracy. In addition, the computed Equations (16) and (17) value attains to the robustness and reliability of the system. According to the above computed CA_r^k and β_r^k values, network parameters weight, and bias values are updated continuously.

The frequent fine-tuning of extreme learning network reduce the deviations between the outputs. Therefore, the HCA-FL attains a minimum error rate value compared to other methods. The low error rate is directly proportional to the prediction accuracy. Then the obtained accuracy value for the paddy crop images and iterations are illustrated in Figure 5.



a) Paddy crop.

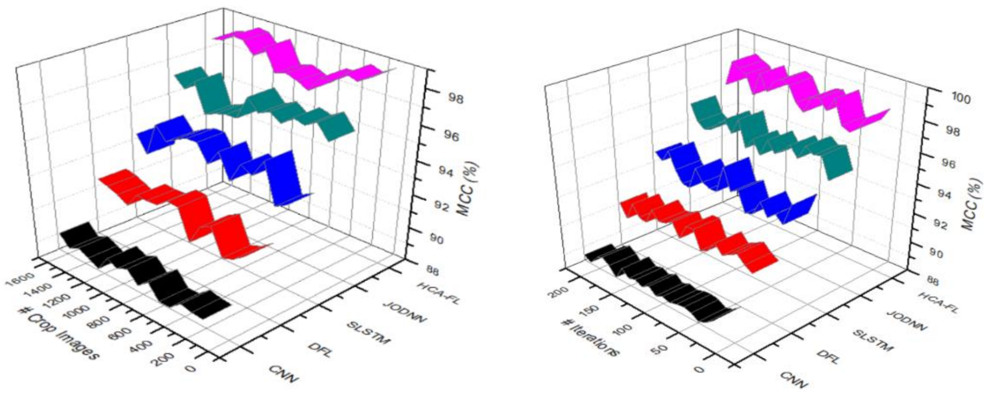
b) Iterations.

Figure 5. Accuracy analysis for HCA-FL.

Figure 5 shows that HCA-FL attains a high accuracy value while monitoring paddy crops in the Thanjavur region. The efficiency of the system is evaluated with various existing methods such as Convolution Neural Networks (CNN) [30], Jaya Optimized Deep Neural Networks (JODNN) [11], Stacked Long-Short Memory Neural Networks (SLSTM) [14], and Deep Feature Learning (DFL) [12] for the various number of crop images and iterations. The HCA-FL approach uses the affinity matrix W while computing the data point features. The W is frequently analysed and derives the (μ, λ) that helps to minimize the feature dimensionality. The top-k eigenvector and eigen values-based data point selection minimize the over fitting issues and improve the overall clustering efficiency. In addition, the computed W's related similarity computation form the cluster and cluster members are decided based on the cluster centroid value. The effective computation of clusters and members helps to improve the overall paddy crop monitoring accuracy (crops-98.13% and iterations-97.98%-in average).

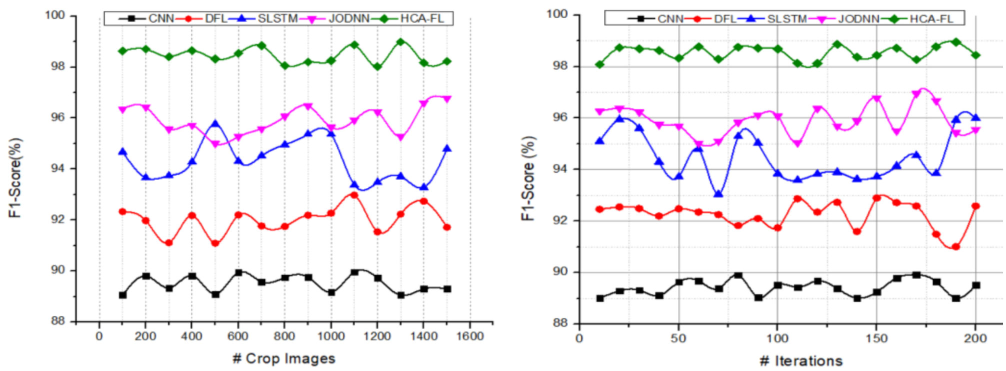
Figure 6 depicts the MCC analysis of the HCA-FL method while monitoring crop growth. The system's efficiency is determined for various numbers of crop images and iterations. Here, the geometrically corrected images are investigated continuously to extract the vegetation information, NDVI, EVI, and LAI details. The extracted vegetation information identifies the paddy ages and growing stages. The paddy stages and ages identify paddy growth. In addition, green regions are further explored using the hybridized clustering approach. The clustering process detects the green region with the help of spectral clustering and ensemble learning. Combining clusters maximizes the overall green region, which helps predict the soil quality and water level. The analysis computes the data similarities form the affinity matrix that reduces the dimensionality and over fitting issues while monitoring the crop. Thus overall, the HCA-FL approach maximizes the crop monitoring efficiency and yielding rate.

The extracted vegetation information identifies the paddy ages and growing stages. The paddy stages and ages identify paddy growth. In addition, green regions are further explored using the hybridized clustering approach. The clustering process detects the green region with the help of spectral clustering and ensemble learning. Combining clusters maximizes the overall green region, which helps predict the soil quality and water level. The analysis computes the data similarities form the affinity matrix that reduces the dimensionality and over fitting issues while monitoring the crop. Thus overall, the HCA-FL approach maximizes the crop monitoring efficiency and yielding rate. Figure 7 depicts the F1-score value of the HCA-FL approach, which is compared with existing methods.



a) Paddy crop. b) Iteration.

Figure 6. MCC Analysis of HCA-FL.



a) Paddy Crop. b) Iterations.

Figure 7. F1-score analysis of HCA-FL.

The HCA-FL method ensures a high monitoring rate (98.53%-Crop numbers; 98.3% for iterations) while analyzing the large volume of crop images. Effectively utilizing pre-processing procedures, histogram approach, GCP points, and vegetation information effectively recognize the green region. In addition, the ensemble approach uses voting techniques while predicting soil quality and water level. Extreme learning networks update network parameters at every iteration. The frequent computation of vegetation area, paddy ages, crop stage, soil quality, and water level is more useful for identifying crop growth. Then the overall results are shown in Table 4. The HCA-EL approach recognizes the crop growth from highly correlated feature analysis (98.10%). The maximum correlation feature exploration maximizes the crop monitoring rate up to 98.49% F1-score and minimum error rate (0.105) (Table 4). Then the HCA-EL approach recognizes the crop growth from highly correlated feature analysis (98.514%). The maximum correlation feature exploration maximizes the crop monitoring rate up to 97.994% of F1-score and minimum error rate (0.098) (Table 5).

The accuracy, error rate and F1 score is derived by using the Equations (20, (21), and (22).

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(True\ Positive + True\ Negative + False\ Positive + False\ Negative)} \quad (20)$$

$$Error\ Rate = Actual\ Value - Computed\ Value \quad (21)$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (22)$$

From the above equation True Positive implies vegetation area is correctly identified, False Positive implies null value for falsely identified vegetation area, True Negative implies non vegetation area is effectively identified, False Negative implies vegetation area which is wrongly identifies.

Table 4. Overall Efficiency Analysis of HCA-EL for paddy crop.

Metrics	CNN [30]	JODNN [11]	SLSTM[14]	DFL [12]	HCA-FL
Error Rate	0.2273	0.146267	0.168673	0.212087	0.1015
Accuracy	89.68	96.08	94.22	91.93	98.152
MCC	89.502	96.076	94.302	91.96	98.10
F1-Score	89.48	95.908	94.306	91.68	98.49

Table 5. Overall Efficiency Analysis of HCA-EL for Paddy crop iteration.

Metrics	CNN [30]	JODNN [11]	SLSTM[14]	DFL [12]	HCA-FL
Error Rate	0.23452	0.21323	0.16562	0.14381	0.09981
Accuracy	89.4905	91.9	94.545	96.1	98.0375
MCC	89.553	91.92	94.24	95.8365	97.994
F1-Score	89.337	91.774	94.5335	95.8695	98.514

7. Conclusions

Thus the paper analyzes the Hybridized Clustering Algorithm based on Ensemble Learning HCA-EL based paddy crop monitoring process. This work uses Sentinel-2 Satellite images to explore the paddy crop growth. The Thanjavur research area is choosen, and the images are collected, which are processed to improve the radiometric, geometric, and atmospheric corrections. Then different vegetation area information is extracted to maintain the system's robustness and reliability. The extracted vegetation area is further analyzed to derive the green region, and similar data points are grouped to improve the crop monitoring process. Similar dimensionality issues are overcome during clustering by sorting the data points depending on the eigenvector and value. Then ensemble learning extreme machine networks classify the soil quality and water level according to the voting criteria. The discussed system was implemented using the python tool, ensuring a 0.098 error rate and 98.03% accuracy. In the future, an optimized feature selection approach will be incorporated to reduce further computation challenges.

References

[1] Abinaya M., Durgadevi N., Ramya K., Pradeepa I., Balamurugan R., and Nirmal R., "Significance of Water Management and Conservation in Agriculture," *The Pharma Innovation Journal*, vol. 9, pp. 174-175, 2020. https://www.researchgate.net/publication/338447942_Significance_of_water_management_and_conservation_in_agriculture

[2] Agarwal S., Water Usage in Agriculture and How it Can be Managed, https://timesofindia.indiatimes.com/blogs/voices/water-usage-in-agriculture-how-it-can-be-managed/, Last Visited, 2024.

[3] Ajith K., Geethalakshmi V., Ragunath K., Pazhanivelan S., and Panneerselvam S., "Rice Acreage Estimation in Thanjavur, Tamil Nadu Using Lands at 8 OLIIMAGES and GIS Techniques," *International Journal of Current Microbiology and Applied Sciences*, vol. 6, no. 7, pp. 2327-2335, 2017. https://doi.org/10.20546/ijcmas.2017.607.275

[4] Ali A., Savin I., Poddubskiy A., Abouelghar M., Saleh N., "Integrated method for rice cultivation monitoring using Sentinel-2 data and Leaf Area Index Egypt," *Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 431-441, 2021. https://doi.org/10.1016/j.ejrs.2020.06.007

[5] Aqdam K., Rezapour S., Asadzadeh F., and Nouri A., "An Integrated Approach for Estimating Soil Health: Incorporating Digital Elevation Models and Remote Sensing of Vegetation," *Computers and Electronics in Agriculture*, vol. 210, pp. 107922, 2023. https://doi.org/10.1016/j.compag.2023.107922

[6] Azedou A., Amine A., Kisekka I., Lahssini S., and Bouziani Y., "Enhancing Land Cover/Land Use (LCLU) Classification Through a Comparative Analysis of Hyperparameters Optimization Approaches for Deep Neural Network (DNN)," *Ecological Informatics*, vol. 78, pp. 102333, 2023. https://doi.org/10.1016/j.ecoinf.2023.102333

[7] Belkin M. and Niyogi P., "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," *in Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, Vancouver, pp. 585-591, 2001.

[8] Chai T. and Draxler R., "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)," *Geoscientific Model Development Discussions*, vol. 7, no. 3, pp. 1247-1250, 2014. https://doi.org/10.5194/gmd-7-1247-2014

[9] Claverie M., Ju J., Masek J., Dungun J., and Vermote E., "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set," *Remote Sensing of Environment*, vol. 219, pp.145-161, 2018. https://doi.org/10.1016/j.rse.2018.09.002

[10] CopernicusProgramme, https://sentinels.copernicus.eu/web/sentinel/copernicus, Last Visited, 2024.

[11] Devi T. and Neelamegam P., "Image Processing Based Rice Plant Leaves Diseases in Thanjavur, Tamilnadu," *Cluster Computing*, vol. 22, pp. 13415-13428, 2019. https://doi.org/10.1007/s10586-018-1949-x

[12] Divakar M., Sarith M., Elayidom M S., and Rajesh R., "Design and Implementation of an Efficient and Cost Effective Deep Feature Learning Model for Rice Yield Mapping," *International Journal of Computational Science and Engineering*, vol. 25, no. 2, pp. 128-139, 2022. https://doi.org/10.1504/IJCSE.2022.122205

[13] El-Hokayem L., De Vita P., and Conrad C., "Local Identification of Groundwater Dependent Vegetation Using High-Resolution Sentinel-2 Data-A Mediterranean Case Study," *Ecological*

Indicator, vol. 146, pp. 109784, 2023. https://doi.org/10.1016/j.ecolind.2022.109784

[14] Geetha M., Suganthe R., Nivetha S., Anju R., and Anuradha R., "A Time-Series Based Yield Forecasting Model Using Stacked LSTM to Predict the Yield of Paddy in Cauvery Delta Zone in Tamilnadu," *in Proceedings of the 1st International Conference on Electrical, Electronics, Information and Communication Technologies*, Trichy, pp. 1-6, 2022. doi: 10.1109/ICEEICT53079.2022.9768441

[15] Guan K., Li Z., Rao L., Gao F., and Xie D., "Mapping Paddy Rice Area and Yields Over Thai Binh Province in Viet Nam from MODIS, Landsat, and ALOS-2/PALSAR-2," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, pp. 2238-2252, 2018. https://doi.org/10.1109/JSTARS.2018.2834383

[16] Hummel P., "Remotely Sensed Ground Control Points," *The International Archives of the Photogrammetry*, *Remote Sensing and Spatial Information Sciences*, vol. XLI-B7, pp. 797-802, 2016. doi:10.5194/isprsarchives-XLI-B7-797-2016

[17] Jana B., Thotakura H., Baliyan A., Sankararao M., and Deshmukh R., "Pixel Density Based Trimmed Median Filter for Removal of Noise from Surface Image," *Applied Nanoscience*, vol. 13, pp. 1017-1028, 2023. https://doi.org/10.1007/s13204-021-01950-0

[18] Joshua V., Priyadharson S., and Kannadasan R., "Exploration of Machine Learning Approaches for Paddy Yield Prediction in Eastern Part of Tamilnadu," *Agronomy*, vol. 11, no. 10, pp. 2068, 2021. https://doi.org/10.3390/agronomy11102068

[19] Joyalata L., Saxena K., Rakesh M., and Rao K., "Soil Quality and Soil Health: A Review," *International Journal of Ecology and Environmental Sciences*, vol. 38, no. 1, 2012.

[20] Kanjir U., Duric N., and Veljanovski T., "Sentinel-2 Based Temporal Detection of Agricultural Land Use Anomalies in Support of Common Agricultural Policy Monitoring," *ISPRS International Journal of Geo-Information*, vol. 7, pp. 405, 2018. https://doi.org/10.3390/ijgi7100405

[21] Kayad A., Sozzi M., Gatto S., Marinello F., and Pirotti F., "Monitoring Within-Field Variability of Corn Yield Using Sentinel-2 and Machine Learning Techniques," *Remote Sensing*, vol. 11, no. 23, pp. 2873, 2019. https://doi.org/10.3390/rs11232873

[22] Khanal S., Fulton J., and Shearer S., "An Overview of Current and Potential Applications of Thermal Remote Sensing in Precision Agriculture," *Computers and Electronics in Agriculture,* vol. 139, pp. 22-32, 2017. https://doi.org/10.1016/j.compag.2017.05.001

[23] Kowe P., Mutanga O., Odindi J., and Dube T., "A Quantitative Framework for Analysing Long Term Spatial Clustering and Vegetation Fragmentation in an Urban Landscape Using Multi-Temporal Landsat Data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, pp. 102057, 2018. https://doi.org/10.1016/j.jag.2020.102057

[24] Lakhiar I., Gao J., Syed T., Chandio F., and Buttar N., "Monitoring and Control Systems in Agriculture Using Intelligent Sensor Techniques: A Review of the Aeroponic System," *Journal of Sensors,* vol. 2018, no. 672769, pp. 1-18, 2018. https://doi.org/10.1155/2018/8672769

[25] Laso F., Benítez F., Rivas-Torres G., Sampedro C., and Arce-Nazario J., "Land Cover Classification of Complex Agroecosystems in the Non-Protected Highlands of the Galapagos Islands," *Remote Sensing*, vol. 12, no. 1, pp. 65, 2019. https://doi.org/10.3390/rs12010065

[26] Mandal D., Kumar V., Ratha D., Dey S., and Bhattacharya A., "Dual Polarimetric Radar Vegetation Index for Crop Growth Monitoring Using Sentinel-1 SAR Data," *Remote Sensing of Environment*, vol. 247, pp. 111954, 2020. https://doi.org/10.1016/j.rse.2020.111954

[27] Maponya M., Niekerk A., and Mashimby Z., "Pre-Harvest Classification of Crop Types Using a Sentinel-2 Time-Series and Machine Learning," *Computers and Electronics in Agriculture*, vol. 169, pp. 105164, 2020. https://doi.org/10.1016/j.compag.2019.105164

[28] Mohammed A. and Kora R., "A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges," *Computer and Information Sciences*, vol. 35, no. 2, pp. 757-774, 2023. https://doi.org/10.1016/j.jksuci.2023.01.014

[29] Moradi E. and Sharifi A., "A Fast Radiometric Correction Method for Sentinel-2 Satellite Images," *Aircraft Engineering and Aerospace Technology*, vol. 93, no. 10, pp. 1709-1714, 2021. https://doi.org/10.1108/AEAT-11-2020-0262

[30] Pozzobon de Bem P., Abílio de Carvalho Júnior O., Ferreira de Carvalho O., Arnaldo Trancoso Gomes R., and Fontes Guimarães R., "Irrigated Rice Crop Identification in Southern Brazil Using Convolutional Neural Networks and Sentinel-1 Time Series," *Remote Sensing Applications: Society and Environment*, vol. 24, pp. 100627, 2021. https://doi.org/10.1016/j.rsase.2021.100627

[31] Prakash M., Pazhanivelan S., Ragunath K., Muthumanickam D., and Sivamurugan A., "Paddy Area Estimation in Cauvery Delta Region Using Synthetic Aperture Radar," *International Journal of Environment, Ecology and Conservation*, vol.

13, no. 10, pp. S517-S522, 2022. DOI: 10.9734/ijecc/2023/v13i102630

[32] Ramesh S. and Vydeki D., "Recognition and Classification of Paddy Leaf Diseases Using Optimized Deep Neural Network with Jaya Algorithm," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 249-260, 2020. https://doi.org/10.1016/j.inpa.2019.09.002

[33] Rammohan S., Niveditha V., Singh A., and Yuvarani T., "Crop Yield Forecast Using a Hybrid Framework of Deep CNN with RNN Technique," *Computing Technologies and Applications*, 2021.

[34] S2 Mission, https://sentiwiki.copernicus.eu/web/s2-mission, Last Visited, 2024.

[35] Saiz-Rubio V. and Rovira-Mas F., "From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management," *Agronomy*, vol. 10, no. 2 pp. 207, 2020. https://doi.org/10.3390/agronomy10020207

[36] Samsuddin S., Abdul-Maulud K., Sharil S., Karim O., and Pradhan B., "Monitoring of Three Stages of Paddy Growth Using Multispectral Vegetation Index Derived from UAV Images," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 26, no. 4, pp. 989-998. DOI:10.1016/j.ejrs.2023.11.005

[37] SentinelOnline, https://sentinel.esa.int/web/sentinel/home, Last Visited, 2024.

[38] Sethy P., Behera S., Kannan N., Narayanan S., and Pandey C., "Smart Paddy Field Monitoring System Using Deep Learning and IoT," *Concurrent Engineering*, vol. 29, no. 1, pp. 16-24, 2021. https://doi.org/10.1177/1063293X21988944

[39] Sharma P., Kumar P., Srivastava H., and Sivasankar T., "Assessing the Potentials of Multi-Temporal Sentinel-1 SAR Data for Paddy Yield Forecasting Using Artificial Neural Network," *Journal of the Indian Society of Remote Sensing*, vol. 50, pp. 895-907, 2022. https://doi.org/10.1007/s12524-022-01499-7

[40] Su W., Zhang M., Jiang K., Zhu D., and Huang J., "Atmospheric Correction Method for Sentinel-2 Satellite Imagery," *Acta Optica Sinica*, vol. 38, no. 1, pp. 0128001, 2018.

[41] Surega R. and Neelakantan R., "Geoinformatics for Agricultural Land Suitability Assessment for Sustainable Land Management in Thanjavur District, South India," *International Journal of Creative Research Thoughts*, vol. 10, no. 1, pp. 90-101, 2022. https://ijcrt.org/papers/IJCRT2201572.pdf

[42] Susan J. and Subashini P., "Deep Learning Inpainting Model on Digital and Medical Images-A Review," *The International Arab Journal of Information Technology*, vol. 20, no. 6, 2023. https://doi.org/10.34028/iajit/20/6/9

[43] ThanjavurDistrict, https://tnfusa.org/mannvaasanai/thanjavu/, Last Visited, 2024.

[44] Umatani R., Imai T., Kawamoto K., and Kunimasa S., "Time Series Clustering with an EM Algorithm for Mixtures of Linear Gaussian State Space Models," *Pattern Recognition*, vol. 138, pp. 109375, 2023. https://doi.org/10.1016/j.patcog.2023.109375

[45] Venkatesan R. and Prabu S., "Preprocessing of Multimodal Hyperspectral Imaging Using Anisotropic Diffusion Approach," *Journal of Computational and Theoretical Nanoscience*, vol. 15, pp. 2617-2624, 2018. https://doi.org/10.1166/jctn.2018.7509

[46] Wu C. and Zhang J., "Self-Supervised Spectral Clustering with Spectral Embedding for Hyperspectral Image Classification," *International Journal of Remote Sensing*, vol. 45, no. 12, pp. 3913-3936, 2024. https://doi.org/10.1080/01431161.2024.2358547

[47] Yu N., Li L., Schmitz N., Tian L., Greenberg J., and Diers B., "Development of Methods to Improve Soybean Yield Estimation and Predict Plant Maturity with an Unmanned Aerial Vehicle Based Platform," *Remote Sensing of Environment*, vol. 187, pp. 91-101, 2016. https://doi.org/10.1016/j.rse.2016.10.005

[48] Zhao Y., Huang B., and Song H., "A Robust Adaptive Spatial and Temporal Image Fusion Model for Complex Land Surface Changes," *Remote Sensing of Environment*, vol. 208, pp. 4262, 2018. https://doi.org/10.1016/j.rse.2018.02.009

Sathiya Priya.R received her M.Tech Degree from Ganadipathy Tulsi's Jain Engineering College, Vellore, Tamil Nadu affiliated to Anna University, Chennai. She is currently pursuing her Ph.D. in Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India. Her research area of interest include Image Processing, Deep Learning and Cloud Computing.

Rahamathunnisa.U received her Ph.D. from Vellore Institute of Technology, Vellore, Tamil Nadu, India. She is currently working as Associate Professor (Sr) in Vellore Institute of Technology, Vellore, Tamil Nadu, India. Her Research area of interest include Multimedia Networks, Image and Video Processing. Her publication includes Book Chapters, Conferences, Scopus Indexed journals, SCIE and SCI Journals.