Journal of Integrated
SCIENCE & TECHNOLOGY

Land cover change detection using TIMESAT software and machine learning algorithms near Ujani Dam: A case study
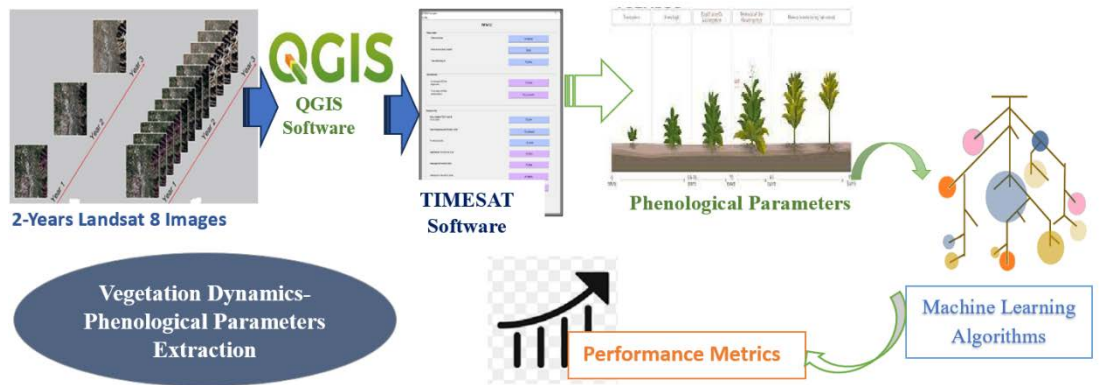
Manisha Kumawat,* Arti Khaparde

Dr. Vishwanath Karad MIT World Peace University, Pune. India.

ABSTRACT

This work focuses on the case study of vegetation dynamics using Machine Learning algorithms on satellite time series data. Vegetation covers a substantial portion of the Earth's land surface, having an important role in the Earth's energy balance. Several factors, such as climate, abiotic environments or biotic interactions, and the dynamics of population and their activities have a great impact on the vegetation processes. Estimation of phenological and seasonal parameters is mainly used for land cover change detection. Estimation of seasonal parameters requires high spatial and temporal resolution data. Time-sequence of vegetation index data from Landsat 8 is used for monitoring vegetation seasons and for investigating and obtaining seasonality-related factors. This research focuses on obtaining seasonal parameters using the TIMESAT program package available in MATLAB. Preprocessing of Satellite images is done to remove the effect of noise and cloud. The comparative analysis is done for four different time sequence smoothing techniques like asymmetric Gaussian, Savitzky-Golay, and Seasonal Trend Decomposition using LOESS (STL), and Double-logistic functions. It is seen with an output of TIEMSAT that Double logistics gives the best result verified in situ. Further, to avoid human interference, and to get better accuracy in the prediction of phenological parameters, different machine learning techniques are applied to TIMESAT output, and performance parameters like f1-score, accuracy, and precision are compared. Random Forest, Decision Tree, and Naïve Bayes methods are used where the suggested, Naïve Bayes algorithms along with a Double logistic filter give enhanced 94% accuracy with a precision of 96 % as compared to the Decision tree and random Forest algorithm.

Keywords: NDVI, TIMESAT, Seasonality Parameters, Decision Tree, Random Forest, Naïve Bayes

INTRODUCTION

The study of vegetation dynamics has been an active research topic since the late 19th century. Mapping vegetation is important and useful for natural resources management, land planning, and/or environmental policy-making decision support. With the increasing availability of remotely sensed images, and the evolution of sensor technology (spectral, spatial, radiometric, and temporal resolutions), vegetation mapping has become possible. The temporal resolution has a crucial impact in the field of study, as it provides data on different stages of vegetation growth. Vegetation has unique spectral signatures, which evolve with the plant vegetative cycle. This advantage along with the use of long time-series periods leads to a more detailed and rigorous monitoring of the vegetation dynamics changes. Thus, vegetation mapping through quantifying vegetation cover presents valuable insight for understanding the natural and man-made environments, from local to global scales.

It is crucial to have high-quality land data sets to promote sustainable practices and gather information regarding the unsustainable usage of natural resources. The deployment of space missions with the goal of providing high spatial resolution coverage of Earth every few days has enhanced the accessibility of satellite

*Corresponding Author: Manisha Kumawat
Tel: +91 8237196799
Email: manisha.kumawat@mitwpu.edu.in

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   1

image time series and will continue to do so in the future years.[2] The majority of experts examined phenological tendencies and linked them to changes in environmental factors like temperature and precipitation to forecast the start of drought in dry climates.[3,4]

Enhanced Vegetation Index (EVI) and Normalize difference vegetation index (NDVI) are two examples of time-series remote sensed information indices that are used to indirectly validate phenology. These indexes were used by several researchers to group vegetation into phenological clusters. A categorized map based on agro-climatological area is produced by grouping plants with the same phenological relevance. The majority of phenological cycles in nature may be seen with unaided eyes when vegetation undergoes seasonal change, and they also provide seasonal parameters like curve, profile, and signal when doing NDVI validation on data collected over a year. Plant seasonality in time series can be effectively replaced by NDVI data spanning a year's length[5]. These types of data are created using trustworthy spatial extents that are displayed in the order of NVDI rasterized layers. The entire layer represents a unique time-based aggregation or snapshot. The time series will have twenty-four unique layers of pixels in each zone, for instance, if data is gathered every day and formed twice a month. Then, time series data collected over many years are subjected to an inter-annual comparison and evaluated for internal consistency[6]. Vegetation Index data are frequently used as input information in classification methods for creating data divisions in time series generation. In this case, NDVI is pre-classifying. Each input pixel associated with a comparable spatial location over the entire image is represented as a time series. Input for classification methods is thus supplied as vector information with "n" dimensions, where n is the total pictures in a time sequence. Every image region displays a seasonal image curve and the entire NDVI vector over time.

CONTRIBUTIONS

In this study, TIMESAT has been used for the extraction of seasonality parameters. High-resolution satellite sensors have produced a lot of data recently. This work processes several satellite vegetation index time series datasets and monitors crop phases using the software TIMESAT. The beginning and harvesting dates of the growth period, the total periods, integrated data, and other seasonality factors are extracted with the aid of this program. Different fitting functions, such as the double logistic fitting method and the asymmetric Gaussian, are used to analyze the original signals. The data is further smoothed using a modified Savitzky-Golay filter. Then to find performance parameters machine learning algorithms are used. Specifically, the important contributions of the research are as stated below:

The author has extracted seasonal parameters using TIMESAT software in MATLAB & comparative analysis is done with different filter & fitting techniques combinations. Extracted phenological parameters using different fitting functions & filters and comparison of results has been done. Validation of results is done with in situ data received from farmers after discussion with them. Compared the results of LANDSAT 8 & AWiFS satellite data. The authors have used the application of algorithms of machine learning like the Random Forest, Decision Tree, and Naïve

Bayes algorithm. These algorithms were used on extracted phenological parameters from TIMESAT. Performance parameters are calculated with these algorithms.

LITERATURE SURVEY

RELATED WORK

Based on the histogram analysis of a time sequence dataset, [7] has created a Spatiotemporal categorization of vegetation indices. RENDVI, NDVI, and SAVI, three different vegetation indexes, have been used for the classification. Phenological variables were taken from TIMESAT software and applied the Random Forest (RF) technique to data mining used by H. Bendini et.al.[8] The three smoothing procedures, asymmetric Gaussian function (AG), Savitzky Golay (SG), and double logistic function (DL), were developed using the TIMESAT software package. The classification of satellite images has been done using supervised learning classifiers like SVM [10]. Different benchmark datasets were used for the comparative analysis of the developed SVM-based time sequence vegetation categorization method in accordance with performance metrics like accuracy and kappa coefficient.[10] In order to predict the risk of fire, Y. Michael et.al.[11] created a machine learning-based dynamic data-related satellite vegetation index categorization method. For the classification of mean data, three machine learning algorithms - XGBoost, RF, and logistic regression - were used.[12] The time-series study for the time-sequence classification of vegetation Adaptive Coyote Crow Search Optimization (ACCSO), a new hybrid metaheuristic method, is used to improve the classification performance of the TWDTW concept. According to M. Kumawat et.al.[13] the moth flame-based bird swarm optimization (MF-BSA), a new hybrid meta-heuristic method is well suited to optimize the TWDTW idea for improving classification performance. The software program PhenoSat, which can extract phenological data from a time sequence of satellite-based vegetation index data, was utilized.[14] TIMESTAT, Enhanced TIMESAT, and SPIRITS are examples of software tools for satellite time-series analysis. PhenoSat offers just seven seasonal activities. Only straightforward methods for removing noise and anomalies from the input data, such as outliers and missing data, were provided.[15] TimeStats. Outliers and missing values are replaced with time series that have been influenced, which are obtained by averaging the data or by extrapolating from nearby data.

A time-weighted DTW model for navigating the terrestrial usage and cover was created using "remote sensing pictures time series."[16] The selection of a seasonally relevant window, which is necessary nowadays for evaluating crop systems and vegetation types with several yearly growth cycles, is not allowed. In another way, the suggested technique required additional computational time to process huge applications. The suggested strategy also fared poorly in terms of categorization consistency and border impacts because it neglected the surrounding communities. As per D. Ashourloo,[17] an automatic classification of vegetation-covered areas was created utilizing the Landsat-8 OLI time series data and Alfalfa mapping. With the aid of the annual time series Landsat data, J. Xiao et.al.[18] established a feature-generated land cover classification model.

Here, a 2 stage mapping procedure has been utilized to extract the pertinent data from the acquired Landsat pictures. The classification of vegetation using a land cover mapping method due to the study of altered patterns in acquired satellite images has been reported.[19] The conversion of land cover data has been successfully correlated with changes and modifications in the likelihood of the time series dataset through the use of a temporal segmentation algorithm to find trends. The F. Petitjean[20] utilized time warping to examine the vegetation land cover data from satellite images. Different heuristic algorithms and time series classification models were used to test the method's efficacy, and the findings showed that it behaved well compared to other methods.

A novel approach due to NDVI time series has been used to evaluate the recovery of the vegetation following an earthquake utilizing lower bound distance and DTW.[21] The suggested method performs better at distinguishing between natural fluctuations and perturbations to the vegetation. On the other hand, the suggested approach performed less well and had weakly resilient performance when applied to the cloud filter since When processing the original NDVI time series, it was more sensitive. Y. Burstyn et.al.[22] reported a strategy for leveraging the hierarchical DTW to cluster the various geological time series. The suggested approach was used to cluster the hierarchies as well as to set up a single path matrix for handling the real inputs. To choose the best match among all signals, The emphasis of the proposed model is on the non-local similarities between "emphasizing input signals and non-local outliers." However, it falls short in executing the pertinent geoscience applications at different scales. H. Nemmour et.al. integrated the neural networks along with a fuzzy mechanism for detecting the robust changes in the environment.[33] This developed model has studied the potentiality of the detection of variation in the land cover. Initially, the developed framework has conducted diverse validations for determining the desired separate networks. Further, through the utilization of approaches according to the fuzzy integral notation, different neural networks were fused. T.K. behera et.al.[34] introduced a deep learning-based vegetation analysis method with the utilization of UAV-based aerial images. This deep learning-related model has been considered as the dense module and it helped to prevent the networks' feed-forward nature to eliminate the gradient vanishing problem in the real-time images. Here, the deep structure CNN has been utilized to expand and contract the symmetric paths for retrieving the global features of the image very effectively. These extractions of global features helped segment the vegetation class more precisely from the aerial images. Finally, the obtained implementation results were compared with the two different benchmark datasets like NITRDrone datasets and UDD datasets. The experimentally obtained outcome showed that the construction of these dense related connections was useful for significantly decreasing the requirement of trainable parameters and increasing the efficacy of the time series vegetation and categorization methods. A severe natural calamity, drought frequently causes a decline in the health of the vegetation due to a lack of water. Unfortunately, it is challenging to achieve high-resolution vegetation impact that is consistent over time and space. Although some of this information can be provided by remotely sensed goods, these products frequently have data gaps and low spatial or temporal resolutions. Trade-offs between spatial resolution and revisiting times, where high temporal resolution is matched by coarse spatial resolution and vice versa, are a recurring aspect of remote sensing products. V. Jaarsveld et.al.[35] have successfully used machine learning techniques for drought-affected areas. Vegetation dynamics must be modeled in order to anticipate the future because satellites can only provide data for understanding past and present vegetation conditions. Even though many other aspects are also connected, Y. Sun et.al.[36] has made an effort to simulate NDVI using only meteorological data and a deep learning technique (bidirectional long short-term memory model, BiLSTM) toZ anticipate the vegetation activities and stresses. By establishing the connection between climatic factors and vegetation activity, the BiLSTM sequence processing model can forecast NDVI. Additionally, since open ecosystems are vulnerable to natural disturbances, long-term trends, and transient events, they can have dramatically fluctuating vegetation states in contrast to forests or other relatively stable ecosystems. As a result, it is difficult to anticipate the vegetation state in these kinds of ecosystems. Y. Ma et.al.[37] examined how to estimate vegetation changes in an open ecosystem using deep learning-based methods.

All of these works used different methods to study land cover. Techniques for time series categorization are required, as are the most popular methods, for the classification of vegetation using time series, which are discussed. There has been a description of the performance indicators used to assess the success of the vegetation categorization model. Additionally, the necessity of supervised learning methodologies and deep learning-based classification techniques was emphasized. The performance of time-series data-based machine learning algorithms like SVM, RF, and logistic regression approaches to vegetation categorization was discussed. Additionally, the effectiveness of the deep learning-related vegetation classification models like CNN, LSTM, and DNN is examined in comparison to the various baseline techniques.

DATA ACQUIRING

SATELLITE DATA

Landsat 8 data from the website have been downloaded https://earthexplorer.usgs.gov/. Landsat 8's Time-Series data is collected with a 16-day frequency. The most recent Landsat satellite, Landsat 8, supports the operational land imager (OLI) and thermal infrared sensor (TIRS) payloads. Based on the Worldwide Reference System-2 (WRS-2) path/row system, Landsat 8 takes 740 images every day. The image has dimensions of 185 km by 180 km. For this investigation, the surface-reflected photos from 1 January 2017 to 31 December 2018 are employed.[23]

Additionally, Resourcesat-2 AWiFs sensor images are used for research. Over an orbital swath of 740 kilometers, AWiFS has a 5-day repeat cycle with a resolution of 56 meters. The same reference system as Resourcesat-1 is used by Resourcesat-2, an ISRO (Indian Space Research Organization) data continuity mission with improved spectral bands of the IRS-P6/ResourceSat-1. The three electro-optical cameras that make up each ResourceSat satellite's payload are LISS-3, LISS-4, and AWiFS. In comparison to the on-board WiFS capturing device, the Advanced Wide Field Sensor

(AWiFS)[24] camera provides a spatial resolution of (188 m vs. 56 m), radiometric resolution (7 bits vs. 10 bits), and spectral bands (2 vs. 4).

LOCATION/ STUDY AREA

The study location is agricultural land near Ujani dam, Solapur district, Maharashtra, India. One location using the geographic coordinate system is considered from the same farm with latitude and longitude coordinates, 18.142369N, 75.219999E. The Solapur district has a poor natural environment and is primarily located in arid climates.

SUGGESTED ALGORITHMS & DESCRIPTION

Downloading Landsat 8 L2 level surface reflected pictures is the initial step. To create a time series and provide data to TIMESAT, follow the instructions as per Figure 1. Time series are used to study the seasonal growth of vegetation. Bio-climatic zones' spatial distribution, morphing large-scale circulation patterns, or modifications to land usage can all be shown by huge temporal sequences of vegetation index data.[27] To track the seasons of vegetation, there are a few frequently used approaches for analyzing and extracting seasonality elements from dataset series. To assess TIMESAT performance characteristics, many filtering techniques are utilized. Machine learning methods like Decision trees and Random Forests[28] are applied to the output of TIMESAT. In the assessment, four-time sequence smoothing techniques such as Savitzky-Golay, asymmetric Gaussian, Seasonal and trend Decomposition using LOESS(STL), and double-logistic functions are used, filtering approaches like median, STL, etc. Supervised machine learning methodologies include Ensemble Decision Trees and Random Forests. Another created using supervised learning and the Bayes theorem is the Naive Bayes algorithm. Being a categorization algorithm, due to the likelihood that an object would appear, it offers predictions. It's a well-known classification method that has proven to be efficient and user-friendly in a variety of settings.[29]

PREPROCESSING/PREPARATION OF DATA

MULTIBAND GEOTIFF CREATION FOR TIME SERIES FORMATION

After downloading single-band image files containing satellite image data are received.

The following are the steps for creating multiband images:

1. Set the multiple satellite band layers in QGIS. To create a multi-band file with the desired outcome, the band layers in the Layer browser must be graded in ascending order.

2. If the processing toolbox is not displayed, choose PROCESSING --> TOOLBOX.

3. In the processing toolbar's search field, as shown in Fig, type Merge to locate and launch the GDAL\OGR --> Merge tool.

4. Click on the levels you want to be a part of the multi-layer band, then press the button [...]

5. Layer stack Activate.

6. Pick an output raster type that matches the data storage format of the single input bands.

7. The name of the output file should end in .tif to distinguish it as a GeoTIFF file.
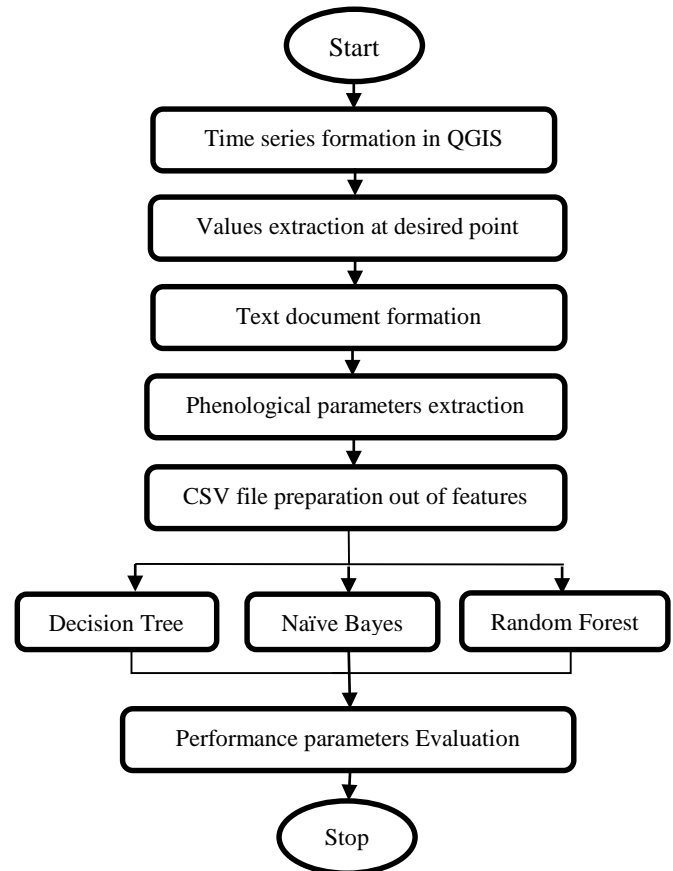

Figure 1. Flow Chart of Methodology

ORIGINAL PICTURES CROPPING

Datasets are regularly downloaded that take up more space than needed. The processing time will be slowed and the map can appear crowded. To fix this, use the QGIS Clipper tool, which may be found in the Raster menu as shown in Figure 2. This will result in a new Raster. Drag Extraction into view, then choose Clipper from the Raster menu.
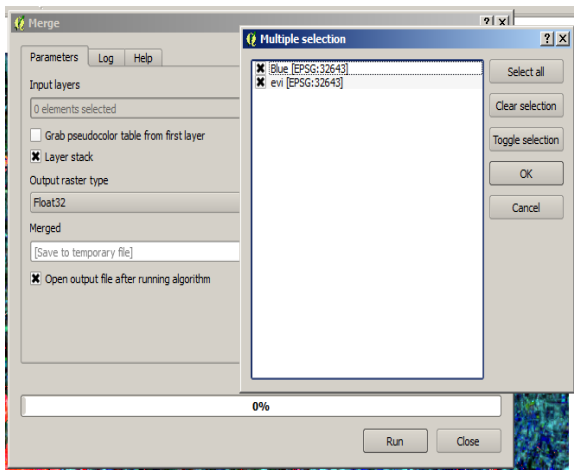

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   4
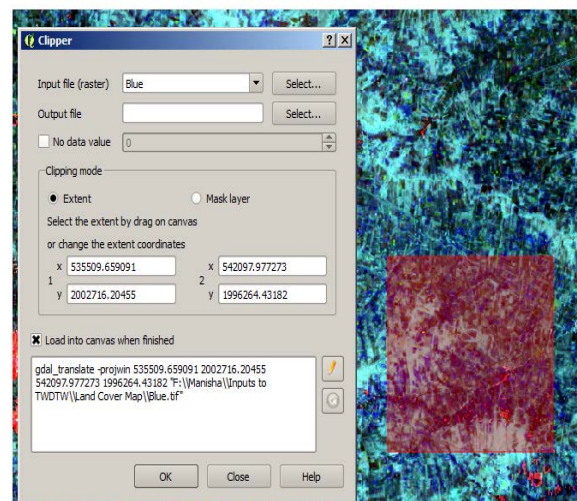
Figure 2. a) Clipping Tool for Raster Images b) The Merge dialogue in QGIS

Once it has been produced, launch QGIS and open the stack image. Hovering the mouse pointer over the chosen raster layer will display the value at that location. As a time series, the set of values will be able to be used to build an ASCII file. Use the check value tool, which is shown in Figure 3, and displays each corresponding value.
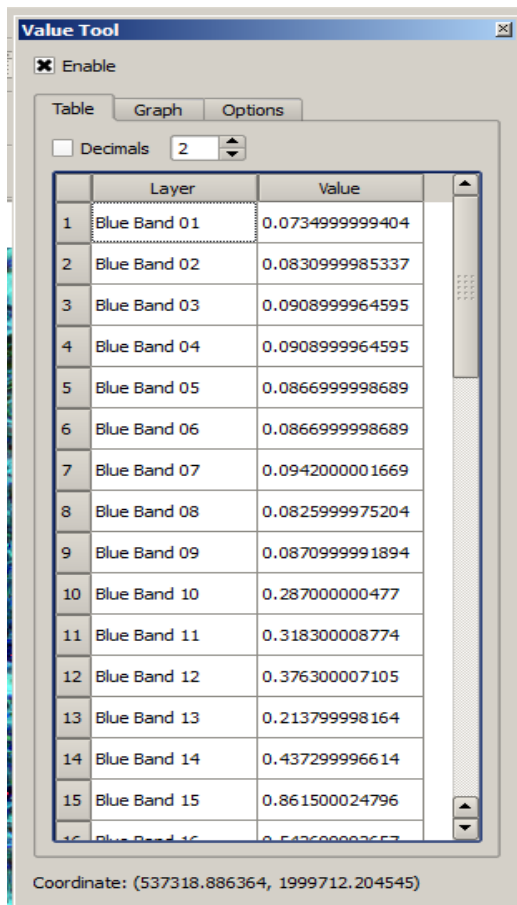

Figure 3. Value Tool Showing Time Series Values

CONSTRUCTION OF TIME SERIES

A formatted image containing NDVI data has been created following the methodology section's earlier instructions. Each NDVI picture I is set up in an array structure for a specific amount of time t. An array of sequential periods (ti, Ii), Se = I1, I2, I3,…IN, is obtained for this place by getting NDVI data from a time series at position (j, k), as per Figure 4.
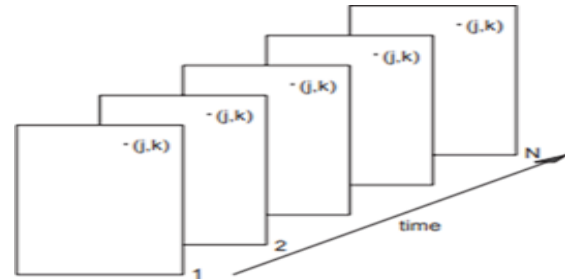

Figure 4. NDVI image time series organization. Each image displays NDVI information at time t. NDVI values over time at a specific spatial point (j, k); a time-series (ti, Ii) with Se = I1, I2, I3, etc. For this site, IN of NDVI data is obtained[26].

Let $Se = I_1, I_2, I_3 \ldots .I_N$ be a set of N pictures with height H and width W i.e., *Se* is a data set generated by combining image time series. Equation 1 provides information about how the time series is constructed from a sequence of photographs.

$$Se = ((I_1 (j, k), I_2 (j, k) \ldots .I_N (j, k)) (j \in (1, W], k \in [1, H]\} \tag{1}$$

This data is further used as input to TIMESAT.

SEQUENTIAL DATA FILE PREPARATION

Time series datasets from the NDVI & EVI are used to build sequential data files, also known as ASCII files[27]. A collection of time series that have undergone TIMESAT processing may be found in these ASCII files.

The header of an ASCII data contains particulars on the total time series that the ASCII file takes into account, nts, the total number of data points that were collected for a year, npyr, and the number of years that have passed through the time sequence, noy. The different time series values for the generalized dataset are shown on a line beneath the header row. The structure of the data in the ASCII file containing the time series is shown in Equation 2 below.

$$noy \quad npyr \quad nts$$

$$nts \begin{cases} y1 \ y2 \ \ldots yn \\ y1 \ y2 \ \ldots yn \\ y1 \ y2 \ \ldots yn \end{cases} \tag{2}$$

Where,

   noy – No of years considered for the study.
   npyr- Total count of every year's samples n
   nts – Total number of time series used

The ASCII file format is seen below. ASCII files are created as text documents.

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   5

Sample input ASCII file of SR NDVI images:

2 23 1

0.3269462 0.22845 0.2136132 0.1971890 0.197864 0.203817 0.2152008 0.2105361 0.197646 0.1001283 0.2722466 0.1200346 0.211185 0.1306333 0.03916074 0.067372 0.562247 0.1840159 0.262488 0.4281277 0.4242928 0.710624 0.6210283 0.6724244 0.5576599 0.557699 0.2793823 0.2134488 0.227110 0.1977861 0.21787709 0.2559774 0 0.166293979 0.2297139 0.2094305 0.0952380 0.6168582 0.0555647 0.7026529 0.6512962 0.7539639 0.0889925 0.709559 0.7142187 0.60872

TIMESAT SOFTWARE

Remotely sensed time series data contains a lot many information to investigate and derive desired results from such data sets; GIS and various other software are used. TIMESAT in MATLAB software is used for obtaining vegetation index time-sequence data derived from satellite spectral measurements, which are applicable in extracting details on seasonal vegetation development.

TIMESAT

A time series data set for the vegetation index is compiled using spectral data taken by space satellites that can be used to calculate the yearly growth of the plant cycle. To investigate and get phenological factors using data time series, the TIMESAT computer package was developed. Numerous time-series data smoothing techniques as well as outlier detection techniques are included in the TIMESAT 3.3 version. To determine the phenological matrices, we can use the 13 repeated parameters that are produced by it.[25,26]

For each image pixel, important phenological metrics, like the beginning and end of the rising season, the duration of a crop cycle, the asymmetry cycle, etc., may be retrieved. The degree to which the fitted functions resemble the original data and input data is what ultimately determines the preferred method. TIMESAT includes a collection of MATLAB files that are used to present both the unprocessed data and the available functions. TIMESAT is a set of MATLAB-based numerical and graphical functions. MATLAB is widely used to perform it. A handful of TIMESAT's characteristics are listed below[27].

1. There are numerous methods for identifying outliers.
2. High-quality data are used to weight the data.
3. High-quality information is used to weight the data.
4. There are two methods for determining the beginning and conclusion of a phenological cycle.
5. Scalable graphical user interface (GUI).
6. Regardless of whether MATLAB is loaded on the computer, it executes.
7. Works with Linux and Windows OS systems as well.
8. Large data sets can be handled using parallel processing.

FITTING METHODS USED

The data is smoothed with filters. Savitzky Golay, Asymmetric Gaussian, and Double Logistic functions are some of the filters that TIMESAT offers, and they are all explained below.[27]

SAVITZKY GOLAY FITTING METHOD

Every point data dpoint i, i = 1;: : : ; n in Savitzky Golay is replaced by a linear combination of surrounding window data, as shown in Equation 3 below:

$$\sum_{j=-n}^{n} fu_j \, dpoint_{i+j} \tag{3}$$

Where, $fu_j = \dfrac{1}{2n+1}$

For all data point *dpoint i* where *i = 1, 2, 3. . . . . . n* keep complete data values. For all 2n + 1 data it can fit in the traveling window, the quadratic polynomial z (t) = fu1 + fu2t + fu3t2 substitutes polynomial values at position ti for the value dpoint i. This process is known as the Savitzky-Golay filter.

DOUBLE LOGISTIC METHOD

The data is regularly fitted to the local model functions for the highest and lowest data in the time sequence. Equation 4 offers the regional function.

$$z (t) = z (t, f, m) = f1 + f 2 Q(t, m) \tag{4}$$

The linear parameter f = measures the amplitude of a starting point (f1; f2). The non-linear parameters m = (m1, m2,..., mp) and dictate the basis function's shape Q (t, m). The foundation function is provided by Equation 5.

$$Q(t, x1, x2 \dots. x4) = \frac{1}{1+\exp\left(\frac{x1-t}{x2}\right)} - \frac{1}{1+\exp\left(\frac{x3-t}{x4}\right)} \tag{5}$$

The left inflection point's location is determined by x1, while the rate of change is provided by x2. Similar to this, x3 establishes where the right inflection point is, and x4 provides the rate of variation for this location. Additionally, to achieve a smooth shape, the span parameters for this function are restricted.

ASYMMETRIC GAUSSIAN FITTING FUNCTION

Equation 6 provides the basis function for the asymmetric Gaussian.

$$Q(t, m1, m2 \dots. m5) = \begin{cases} \exp\left[-\left(\frac{t-m1}{m2}\right)^{m3}\right] & for \ t < m1 \\ \exp\left[-\left(\frac{m1-t}{m4}\right)^{m5}\right] & for \ t > m1 \end{cases} \tag{6}$$

M1 for this function provides the highest or lowest position relating to the time parameter t, m2, and m3 measures the kurtosis, and breadth of the functional right-hand side. Similar to how m4 and m5 measure the kurtosis and width of the right-hand side.

PERFORMANCE PARAMETER EVALUATION USING MACHINE LEARNING ALGORITHMS

From the Landsat-8 NDVI time sequence, seasonal metrics are assessed. The obtained seasonal parameters have been converted into CSV files. After utilizing all smoothing techniques through the software program TIMESAT, machine learning algorithms are applied to obtain a noticeable improvement in the performance metrics. The Random Forest (RF), Decision Tree, and Nave Bayes algorithms are used to analyze the data on the retrieved seasonal

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   6

measures. The aforementioned CSV files are inputs for the Random Forest, Decision Tree, and Naive Bayes algorithms.

RANDOM FOREST ALGORITHM

For categorization and regression issues, one popular supervised learning paradigm is the Random Forest. It is the preferred algorithm in machine learning. The idea of ensemble learning, which is the act of fusing several categorization techniques, serves as its conceptual foundation to complete a challenging task and improve the methodology. Two methods of randomization are used: (1) random data sampling for bootstrap samples, as in bagging; and (2) random selection of input features for the construction of several base decision trees. Random Forest builds many decision trees[30] as the forest grows, Random Forest produces an internal, unbiased estimate of the generalization error, manages thousands of input variables without variable deletion, provides estimates of significant variables, has an effective method for estimating missing data and maintains accuracy even when a significant portion of the data are missing, and has methods for balancing class error in class population unbalanced data sets. It makes use of various decision trees on a variety of input parameters subgroups, as shown in Figure 5 & Algorithm 1, and aggregates the outcomes to improve the anticipated accuracy of the dataset.
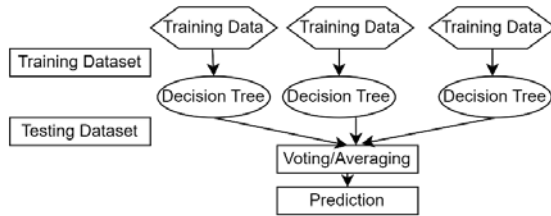

Figure 5. Random Forest using Decision Trees

Algorithm 1: Random Forest as a Regressor or Classifier[31]
For Sa = 1 to Tr: (a) Depending on the training data, generate N-size Btps! bootstrap samples. (b) The following steps must be taken to generate an RF tree Tb to the bootstrapped data: should be carried out iteratively for each node of the tree till the minimal node size n_{min} is achieved. i. From all the input values, pick several variables at random. ii. Take the best split-point variable from a list of candidates. iii. Divide the node into daughter nodes. Ensemble trees Output $\{T_b\}_1^B$. a new point np to make a prediction at: Let $\hat{C}_{Sa}(np)$ be the class estimation of the Sa^{th} RF tree. Predicted output based on majority vote $\hat{C}^B{}_{rf}(np) = majority\ vote\ \{\hat{C}_{Sa}(np)\}_1^B$

SUGGESTED DECISION TREE ALGORITHM

Classification and regression problems can be resolved using the supervised learning technique known as the decision tree. It is a tree-structured extractor, and the features of a dataset are shown inside each node, which also displays the classification conclusion

at each end. The Decision points and End points are the two nodes of a decision tree. Decision points are used to make decisions and have several branches, whereas end points are the outcomes of decisions and do not have any more parts. The Classification and Regression Tree (CART) algorithm is used to construct a tree. As shown in Figure 6[1], a decision tree simply asks a question and divides the tree into subtrees based on the answer (No/Yes).
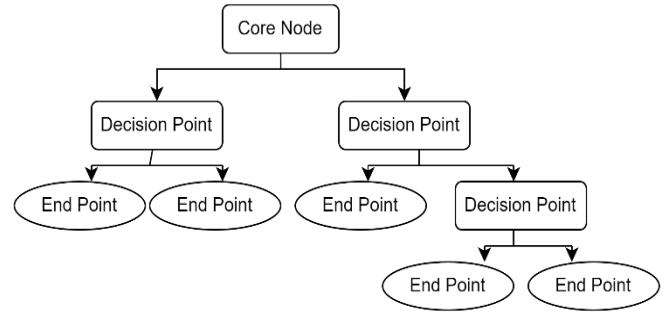

Figure 6. Decision Tree Algorithm

We can choose the ideal attribute for the tree nodes with ease using the Attribute Selection Measure, or ASM. Widely used ASM approaches, which is as follows:

Information Gain: It is the assessment of changes in entropy after a dataset has been segmented based on a characteristic. In a decision tree algorithm, which constantly aims to increases the value of information gain, the node or parameter with the highest information gain is split first. Equation 7 below can be used to compute it[1]:

Information Gain = Entropy(En) − [Weighted Average(Wavg) ∗ Entropy(each feature)]　　　　(7)

An indicator of the impurity in a given attribute is entropy. It describes the data's randomness in accordance with Equation 8.

Entropy(En) = −prb(yes)log2 prb(yes) − prb(no) log2 prb(no)　(8)

Where, En= Entropy, Prb(yes)= yes probability, Prb(no)= no probability

NAÏVE BAYES ALGORITHM

The Naive Bayesian method is a standard classification method that has shown its efficiency and simplicity. It makes the assumption that the classification parameters are independent and do not interact in any way. The Naive Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. Being a probabilistic classifier, it predicts the likelihood that an object will exist.[29]

Bayes Theorem: To determine how plausible a hypothesis is given some prior information, the Bayes theorem, also referred to as Bayes' Rule or law, is utilised. The conditional probability determined using Equation 9.

$$Prb(A|B) = \frac{Prb(B|A)\ Prb(A)}{Prb(B)}　　　　(9)$$

Where, $Prb(A|B)$ is Posterior probability
$Prb(B|A)$ is Likelihood probability, $Prb(A)$ is Prior Probability, $Prb(B)$ is Marginal Probability

RESULTS & DISCUSSION

This section discusses the extraction of seasonal parameters and several agricultural life cycle stages using the TIMESAT software. The TIMESAT software uses a succession of indexed photos from earth observation satellites to provide phenological stages of crops. The suggested machine learning algorithm was used to estimate the performance parameters, and the results were contrasted with those of traditional machine learning techniques.

PRELIMINARY SET-UP

The Quantum Geospatial Information System (QGIS) software is used to pre-process satellite image data. Users can browse, edit, and analyze geographical datasets with the open-source, free QGIS desktop geographic information system programme. Typically, the dataset we download uses far more space than is necessary. For cropping large satellite photos, utilize the clipper tool from the raster menu. Using a merge tool, single band photos are stacked and stored as a multi-band image to create time series. The time series is provided to TIMESAT once it is ready. NDVI time series data produced in MATLAB and Fortran are read using the TIMESAT procedure. The MATLAB 2016 programme was used to access TIMESAT. TIMESAT performs phenological pattern analysis, which yields 13 seasonal parameters.

The recommended Decision Tree and Nave Bayes algorithm were used in the study to gauge performance. The outcomes were compared to those of the conventional Random Forest approach. Positive quantitative metrics were computed. The F1-score index, precision, accuracy, and sensitivity are included in the category of positive values known as Type I. In this test analysis, 45 data points from surrounding farms and agricultural area were taken into account. The NDVI from Landsat 8, the EVI from Landsat 8, and the NDVI from AWiFS were some of the input images used in the proposed model. The outcomes of the random forest method and Naive Bayes are contrasted. After speaking with the farm's owner, information on the agricultural property is gathered on the spot.

EVALUATION OF PHENOLOGICAL STAGES USING TIMESAT SOFTWARE

Some of the 13 main seasonality parameters that are assessed using TIMESAT include the beginning and end of crop cycles, the duration of one phenological cycle, the interval between the beginning and end of a crop cycle, the peak time when NDVI data is maximal, and the maximum value of NDVI.

In this work, the seasonality parameters have been evaluated using the Savitzky-Golay, Asymmetric-Gaussian, and Double Logistic fitting functions. According to the overall results, the season started in the first week of October 2017 and ended in the first week of March 2018. The seasonality parameters for TIMESAT have in-situ validation. The harvest took place on or around March 3, 2018, while the crop was planted on or around October 7, 2017. A field investigation and a discussion with the landowner who grew this crop were used to get the information. The findings from the Double Logistic Fitting Function without a filter were superior because the crop was harvested around March 3, 2018. The season of 2017 started on October 8 and finished on October 8, 2017.

PHENOLOGICAL PARAMETER EXTRACTION USING LANDSAT 8 NDVI DATA

Figure 7 displays a filtered graph of NDVI values that was created using Landsat 8 NDVI data and TIMESAT's double Logistics Fitting Function with None Filter. Also used to produce phenological parameters are various combinations of the Savitzky-Golay fitting function, asymmetric gaussian, none filter, and STL filter. Table 1 displays several crop life cycle phases without a filter using a double logistic fitting algorithm.
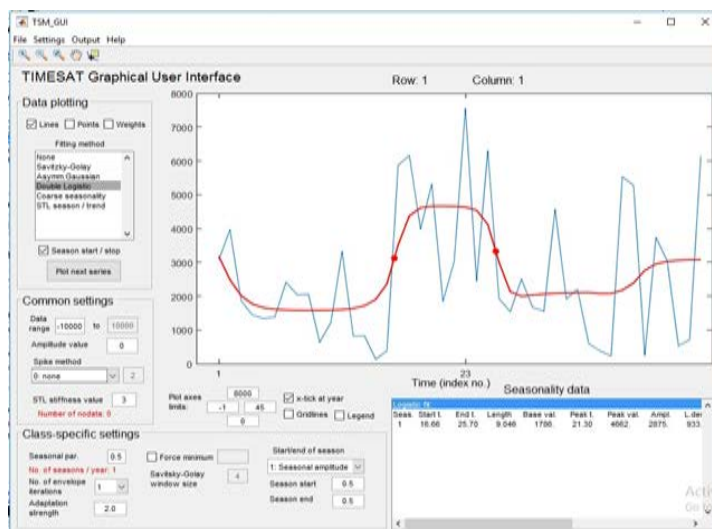

Figure 7. Double Logistic with None Filter GUI with NDVI

Table 1. Seasonality Parameters with Double Logistic with None Filter

Parameter	Phenological Explanation	Value
Starting time	08 October 2017	16.66
Ending time	02 March 2018	25.70
Duration	145 days	9.046
Peak timing	21 December 2017	21.30

PHENOLOGICAL PARAMETER EXTRACTION USING LANDSAT 8 EVI³² DATA

With the use of TIMESAT and a double Logistics Fitting Function with None Filter, and utilizing EVI data from Landsat 8, a filtered graph of EVI values has been displayed in Figure. 8 In addition, phenological parameters are produced using various combinations of the Savitzky-Golay fitting function, asymmetric gaussian, none filter, and STL filter. With no filter and a double logistic fitting procedure, the Table 2 displays several crop life cycle phases.
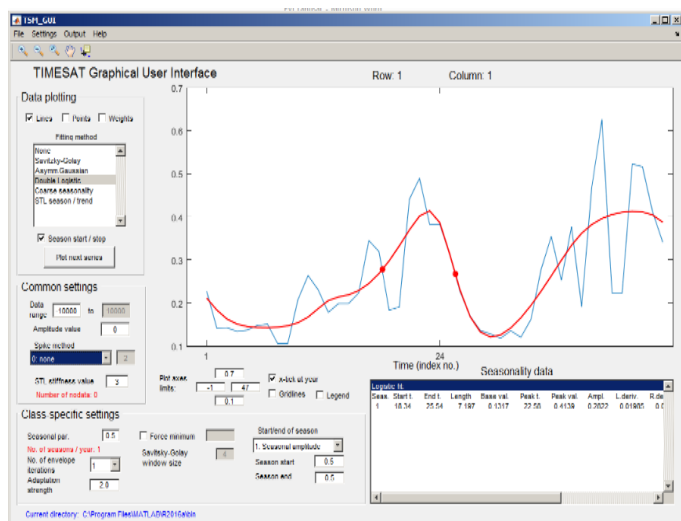
Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   8

Figure 8. Double Logistic with None Filter GUI with EVI Data

EVALUATION OF PHENOLOGICAL MATRICES

Key phenological characteristics are taken into account, as given in Table 3, including Start time, End time, crop cycle duration, and peak time. These phenological characteristics are computed using

Table 2. Seasonality Parameters with Double Logistic Fitting Function with None Filter

Constraints	Phenological data explanation	Seasonality Parameter
Beginning time	02-Oct-17	18.34
Ending time	18-Jan-18	25.54
Duration	107 days	7.197
Peak timing	04-Dec-17	22.58

data from AWiFS, Landsat 8, and Landsat 8 EVI. The Table 3 shows different arrangements of fitting functions, filters, and TIMESAT-retrieved phenological characteristics. Due to the fact that the crop was harvested around 2nd March 2018 and the season began on October 8th, 2017, the Double logistic fitting function produces better results. According to farmer, the crop was planted around October 7 and harvested around March 3. This information was gathered through a field investigation and a conversation with the land owner who grew this crop.

Table 3. Phenological Matrices of Sorghum Crop using NDVI/EVI Data from Landsat 8 and AWiFS

Satellite/Indices	Phenological Parameter	Start Time	End time	Length	Peak time
Landsat 8 EVI	SG with None Filter	5-October-17	17-January-18	142	8-December-17
	SG with STL Replace	5-October-17	17-January-18	104	8-December-17
	Double Logistic with None filter	2-October-17	18-January-18	107	4-December-17
	Double Logistic with STL filter	3-October-17	17-January-18	106	4-December-17
	Asymmetric Gaussian with None filter	23-September-17	19-January-18	118	30-November-17
	Asymmetric Gaussian with STL Filter	24-September-17	19-January-18	117	30-November-17
Landsat 8 NDVI	SG with None	4-October-17	11-March-18	158 days	17-December-17
	SG with STL Replace	4-October-17	12-March-18	159 days	18-December-17
	Double Logistic with None filter	8-Oct-17	2-March-18	145 days	21-December-17
	Double Logistic with STL filter	7-October-17	11-February-18	127 days	15-November-17
	Asymmetric Gaussian with None filter	4-October-17	28-June-18	266 days	18-December-17
	Asymmetric Gaussian with STL Filter	24-September-17	19-January-18	117	30-November-17
AWiFS NDVI	SG with None Filter	26-August-17	25-December-17	121 days	4-November-17
	SG with STL Replace	4-August-17	3-January-18	122 days	27-October-17
	Double Logistic with None	11-August-17	3-January-18	144 days	28-October-17
	Double Logistic with STL	13-August-17	3-January-18	143 days	28-October-17
	Asymmetric Gaussian with None	4-August-17	3-January-18	122 days	27-October-17
	Asymmetric Gaussian with STL	4-August-17	3-January-18	151 days	27-October-17

** Results highlighted in yellow color indicate better results.

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   9

AGRICULTURAL LAND DETAILS RECEIVED FROM FARMER

Figure 9 shows that after visiting the farm, the farmer provided information about the land. Figure 9 shows that the predominant
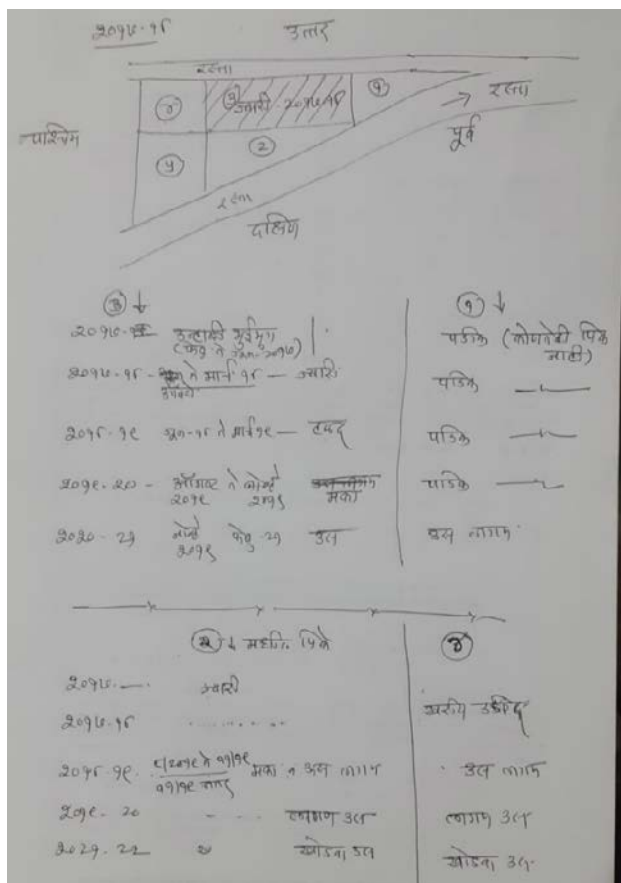

Figure 9. Information Received from Farmer during Agricultural field survey

crop throughout our study period, which spanned from October to March of 2017–2018, was jowar (sorghum). The farmer has written the details in Marathi. Figure 9 gives information provided by a few farmers about the land, after visiting the farm. Figure 9 below shows that the predominant crop throughout our study period, which spanned from October to March of 2017–2018, was jowar (sorghum). The farmer has written the details in Marathi.

PERFORMANCE ANALYSIS USING MACHINE LEARNING ALGORITHMS

Using smoothing techniques, with and without filtering processes, different time series were evaluated for their influence on the classification of farm land in the Solapur area. The NDVI Landsat-8 picture time series was also used to construct phenological metrics. Prior to retrieving the seasonal measures, the TIMESAT software applied the four smoothing techniques Savitzky-Golay (SG), asymmetric Gaussian function (AG), double-logistic function (DL), and STL. The seasonal measures for each approach were applied to machine learning algorithms for data analysis. Performance metrics (positive measures, or Type I measures), which are Type I measures, were calculated using the supervised machine learning algorithms Decision Tree, Random Forest, and Naive Bayes. Asymmetric Gaussian, Double Logistics, Saviztky-Golay, and STL fitting functions, as well as median and STL filters, were used to fit the time series that were delivered to TIMESAT. For the purpose of evaluating performance indicators, 45 data points from the Solapur districts around the Ujani dam were taken into account. These findings are shown in Table 4. It is clear that the suggested Naive-based and decision tree algorithms outperform the Random Forest method in terms of results. It yields better results with all filtering techniques and fitting functions combined. The parameters that outperform random forest algorithms are the f1-score, accuracy, precision, and recall. To assess the results, the confusion matrix analysis was used. The

Table 4. Performance Metrics Calculated using Decision Tree, Random Forest & Naïve Bayes algorithm.

Fitting Functions	Filters	Decision Tree				Random Forest				Naïve Bayes			
		Accuracy	f1-Score	Precision	recall	Accuracy	f1-Score	Precision	recall	Accuracy	f1-Score	Precision	recall
STL	None	0.93	0.93	0.95	0.93	0.73	0.72	0.78	0.73	0.88	0.89	0.92	0.88
	Median	0.87	0.84	0.9	0.87	0.73	0.68	0.69	0.73	0.87	0.85	0.92	0.87
	STL	0.8	0.78	0.78	0.8	0.87	0.84	0.9	0.87	0.53	0.43	0.38	0.53
Double Logistic	None	0.86	0.8	0.77	0.86	0.71	0.65	0.65	0.71	0.94	0.94	0.96	0.94
	Median	0.87	0.85	0.92	0.87	0.84	0.83	0.86	0.84	0.87	0.86	0.9	0.87
	STL	0.93	0.93	0.95	0.93	0.8	0.78	0.87	0.8	0.87	0.85	0.92	0.87
Savitzky Golay	None	0.94	0.93	0.95	0.93	0.6	0.55	0.55	0.6	0.8	0.79	0.85	0.8
	Median	0.87	0.84	0.9	0.87	0.67	0.58	0.6	0.67	0.8	0.78	0.83	0.8
	STL	0.87	0.85	0.92	0.87	0.6	0.57	0.71	0.6	0.8	0.79	0.85	0.8
Asymmetric gaussian	None	0.87	0.84	0.9	0.87	0.8	0.78	0.87	0.8	0.53	0.49	0.66	0.53
	Median	0.87	0.86	0.88	0.87	0.73	0.71	0.8	0.73	0.8	0.78	0.87	0.8
	STL	0.87	0.85	0.92	0.87	0.85	0.83	0.86	0.85	0.87	0.85	0.92	0.87

** Results highlighted in yellow color indicate better results.

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   10

Double Logistic with no filter applied provides the highest classification accuracy, as shown in Table 4, with a global accuracy of 94%, a 94% f1-score, and 96% precision. The following are Type I positive measures:
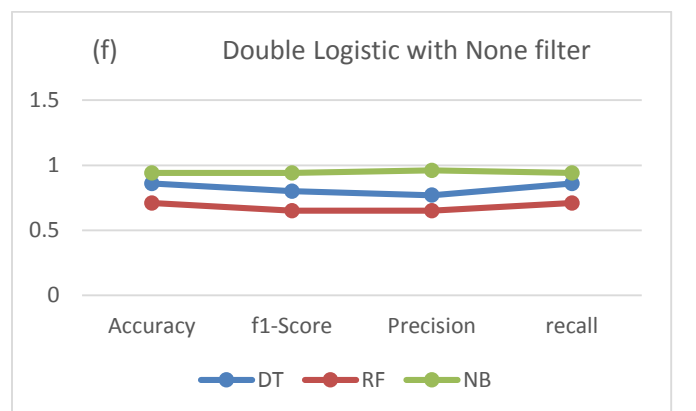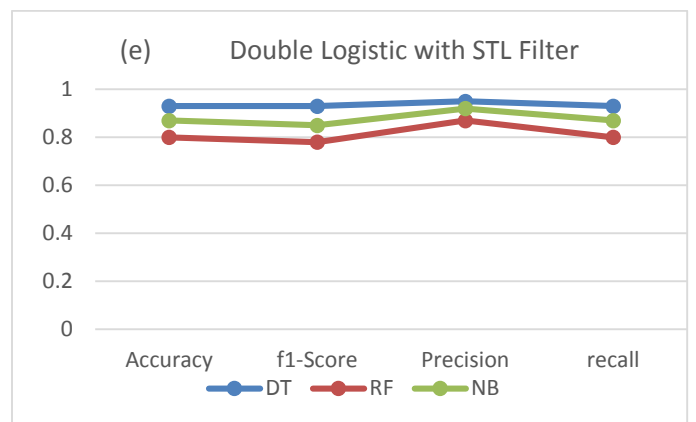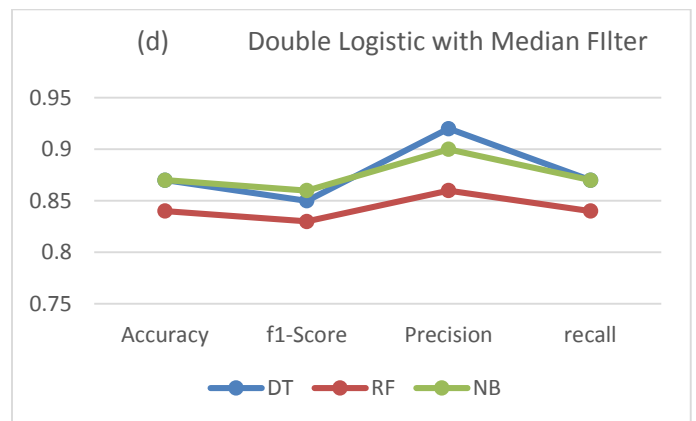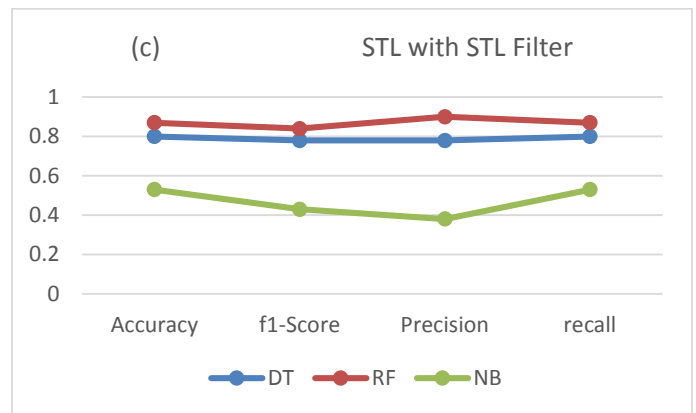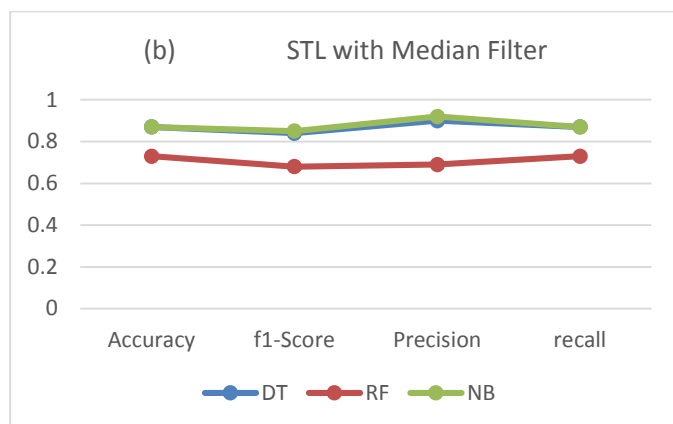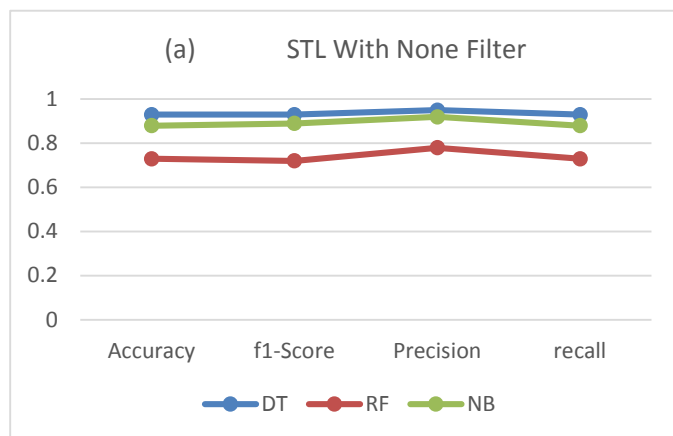
Accuracy: It counts how many of the total number of topics have labels that are accurate.

Precision: Precision is the percentage of all +ve labels that were correctly detected by our algorithm.

Recall: A measurement of the percentage of positive samples among all positive samples that were correctly classified as positive.

F1-Score: Recall and precision are taken into account while computing the F1 Score. It is the harmonic mean of the precision and recall.

Performance analysis of proposed machine learning algorithms, Decision Tree & Naïve Bayes Algorithm over traditional Random Forest algorithm is shown in Figure 10. The analysis is done in terms of type I measures like Accuracy, F1-Score, Precision, and recall. STL, Double logistic, Savitzky-Golay, and Asymmetric fitting functions along with None, Median, and STL filters are used. It is seen from the graphs that the proposed decision tree and Naïve Bayes give the best results as compared to the random forest algorithm.
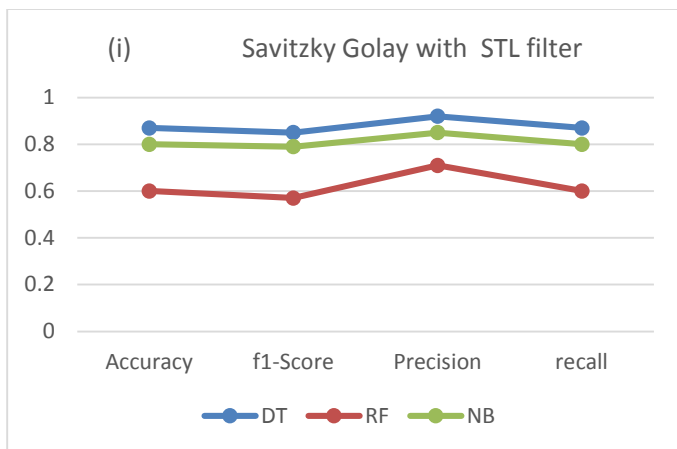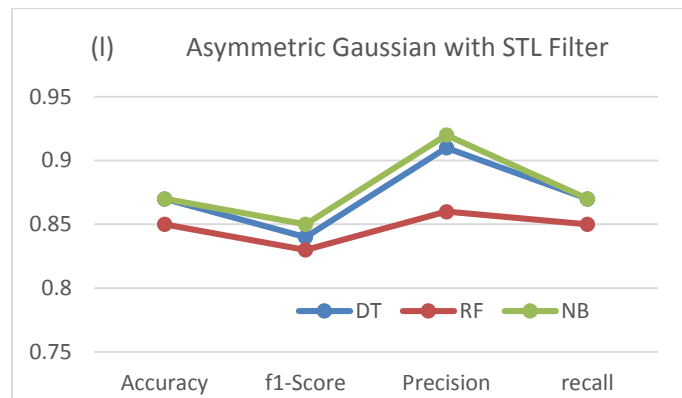

(a) STL With None Filter

(b) STL with Median Filter

(c) STL with STL Filter

(d) Double Logistic with Median FIlter

(e) Double Logistic with STL Filter

(f) Double Logistic with None filter

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   11

(g) Savitzky Golay with Median filter

(K) Asymmetric Gaussian with Median Filter

(h) Savitzky Golay with None Filter
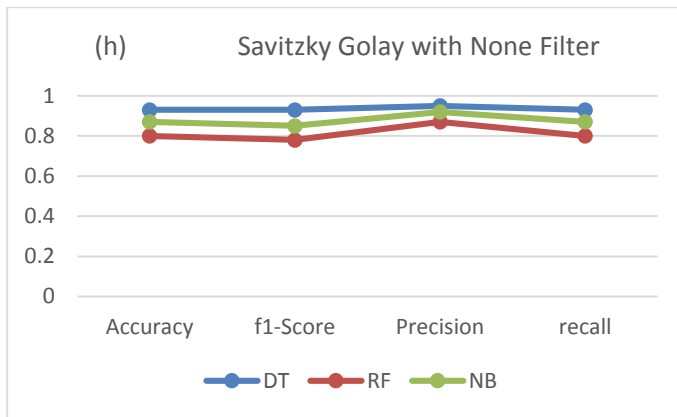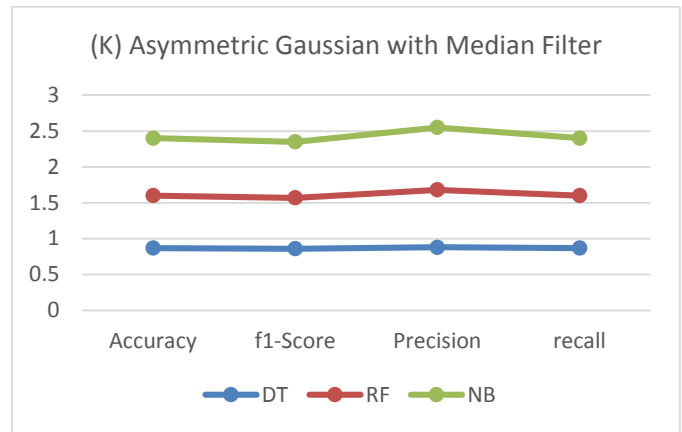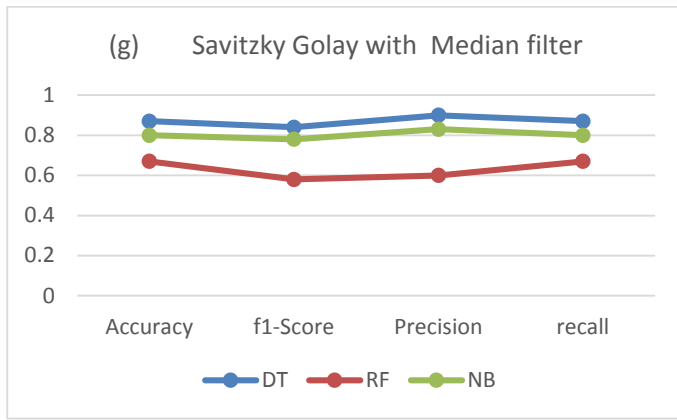

(l) Asymmetric Gaussian with STL Filter

Figure 10. Performance analysis of the suggested ML algorithms a). STL with none b). STL with Median c). STL with STL Filter d). Double Logistic with Median e). Double Logistic with STL f). Double Logistic with None g). Savitzky Golay with Median h) Savitzky-Golay with None i). Savitzky Golay with STL j). Asymmetric Gaussian with None k). Asymmetric Gaussian with Median l). Asymmetric Gaussian with STL.

ANALYSIS OF PERFORMANCE OF SUGGESTED MACHINE LEARNING ALGORITHMS

PERFORMANCE ANALYSIS FOR TYPE I PARAMETERS FOR DOUBLE LOGISTIC FITTING FUNCTION WITH NONE FILTER

Performance Analysis for Type I parameters for Double Logistic Fitting Function with None Filter is shown in Figure 11.
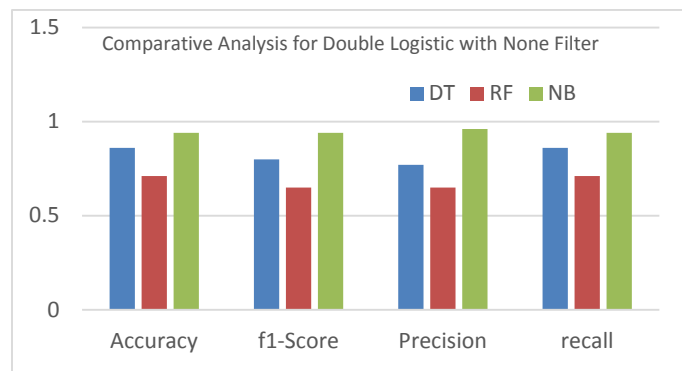

(i) Savitzky Golay with STL filter

Comparative Analysis for Double Logistic with None Filter
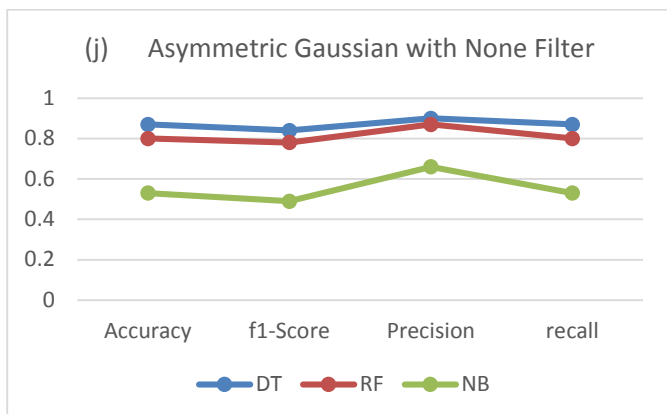
Figure 11. Comparative Analysis for Type I parameters using Double Logistic with None Filter

(j) Asymmetric Gaussian with None Filter

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   12

From the graph it is observed that for the Double logistic fitting function with none filter, Naïve Bayes accuracy is improved by 23%, F1 score improved by 29%, Precision improved by 31% and recall by 23% over Random Forest. Decision Tree algorithm accuracy is improved by 15%, F1-Score improved by 15%, Precision by 12% and recall by 15% over Random Forest.

This study explains phenological parameter extraction from two years of satellite data. Presently there exists a large amount of data from high-resolution satellite sensors. TIMESAT a free software package is used in this work for processing this huge amount of satellite vegetation index time series datasets and monitoring phases of crops. With the assistance of this tool, seasonality parameters are extracted, such as the start and stop dates of the growth period, total duration, integrated data, etc. By using TIMESAT software 13 phenological parameters are extracted. The original signals received from satellite are processed by using different fitting functions like the double logistic fitting method and Asymmetric Gaussian or by applying a modified Savitzky-Golay filter for smoothing of the data. Agricultural land from the Solapur district near the Ujani dam in Maharashtra was considered. On this land, sorghum was taken as the main crop from 1st January 2017 to 31st December 2018. Major 4 categories are considered for classification - Sorghum, Millet, Sugar Cane, and black gram. In this work, the TIMESAT software gives various phases of the crop cycle. For the selected area the suggested machine learning algorithms perform classification of vegetation. These suggested algorithms further avoid the intervention of the farmers once the data has been collected. The naïve bayes algorithm gives 23% improved accuracy over the existing random forest algorithm and the decision tree gives 15% improved accuracy over the random forest algorithm with a double logistic fitting function and none filter.

CONCLUSION & FUTURE SCOPE

This research work has represented a vegetation classification based on time series in two different phases and a comparison was performed on the developed model. This paper has explained how to get different phases of the crop cycle from smoothed data using TIMESAT. Here phenological matrices were extracted from EVI & NDVI data in TIMESAT. Smoothing of data has been done using different fitting functions like Savitzky-Golay, Double Logistic function, STL fitting functions, etc. along with filters. Accuracy assessment of outputs from TIMESAT was performed by utilizing machine learning algorithms. Crop classification can be done based on these phenological parameters as input to machine learning algorithms like Random Forest, Decision Tree, etc. Different performance parameters are evaluated in terms of accuracy etc. It has been observed that suggested machine learning algorithms decision tree & Naive bayes achieved better accuracy over random forest algorithm. Accuracy of Naïve bayes improved by 23% over random forest algorithm & Accuracy of Decision tree is improved 15% over random forest algorithm for Double logistic function. Improvement in precision parameter by suggested algorithms also observed over traditional RF machine algorithm. In future, quantities of image are needed to be enhanced and also different feature vectors are needed to be improved. For study only two

temporal years of data is used with around 15 acres of data. So, it is necessary to utilize complex data for the research to verify the efficiency of the model at the time of analysis. It is important to enhance temporal dimensions as well as spatial integration for occurring effective classification rate. An accurate tool or algorithms are required which deals with the modification attained in the earth when different environmental changes like rainfall, cloudy conditions land slide, wind and so on. Also, TIMESAT provides a time series covering n years, the seasonality characteristics for the n -1 middle most seasons. In theoretically, phenological parameters can be evaluated from just one year of data if the vegetation season peaks in the middle of the time-series. TIMESAT, however, is unable to accurately get phenological characteristics by using data from a single year. Some modifications can be made in TIMESAT MATLAB routine to overcome this problem.

Ethics approval:

The submitted work is original and has not been published elsewhere in any form or language.

Disclosure of potential conflicts of interest:

There is no potential conflict of interest.

Author Contribution Statement: All authors contributed equally.

REFERENCES

1. H.H. Patel, P. Prajapati. Study and Analysis of Decision Tree Based Classification Algorithms. *Int. J. Comput. Sci. Eng.* **2018**, 6 (10), 74–78.
2. F. Petitjean, J. Weber. Efficient satellite image time series analysis under time warping. *IEEE Geosci. Remote Sens. Lett.* **2014**, 11 (6), 1143–1147.
3. P. Rahimzadeh Bajgiran, A.A. Darvishsefat, A. Khalili, M.F. Makhdoum. Using AVHRR-based vegetation indices for drought monitoring in the Northwest of Iran. *J. Arid Environ.* **2008**, 72 (6), 1086–1096.
4. R.M. Trigo, C.M. Gouveia, D. Barriopedro. The intense 2007-2009 drought in the Fertile Crescent: Impacts and associated atmospheric circulation. *Agric. For. Meteorol.* **2010**, 150 (9), 1245–1257.
5. R.A. Geerken. An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change. *ISPRS J. Photogramm. Remote Sens.* **2009**, 64 (4), 422–431.
6. R. DeFries, M. Hansen, J. Townshend. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* **1995**, 54 (3), 209–222.
7. S. Niazmardi, S. Homayouni, A. Safari, et al. Histogram-based spatio-temporal feature classification of vegetation indices time-series for crop mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, 72, 34–41.
8. H. do N. Bendini, L.M.G. Fonseca, T.S. Körting, I.D.A. Sanches, R.F.B. Marujo. Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters. In *Simposio Brasileiro de Sensoriamento Remoto*; **2017**; Vol. 18, pp 4267–4274.
9. L.F.C. Ruiz, L.A. Guasselli, J.P.D. Simioni, T.F. Belloli, P.C. Barros Fernandes. Object-based classification of vegetation species in a subtropical wetland using Sentinel-1 and Sentinel-2A images; Science of Remote Sensing, **2021**; Vol. 3.
10. S.H. Qader, J. Dash, P.M. Atkinson, V. Rodriguez-Galiano. Classification of Vegetation Type in Iraq Using Satellite-Based Phenological Parameters. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, 9 (1), 414–424.
11. Y. Michael, D. Helman, O. Glickman, et al. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Sci. Total Environ.* **2021**, 764, 10,142844.

12. M. Kumawat, A. Khaparde. Time-Variant Satellite Vegetation Classification Enabled by Hybrid Metaheuristic-Based Adaptive Time-Weighted Dynamic Time Warping. *Int. J. Image Graph.* **2022**, 2450016.

13. M. Kumawat, A. Khaparde. Development of Adaptive Time-Weighted Dynamic Time Warping for Time Series Vegetation Classification Using Satellite Images in Solapur District. *Comput. J.* **2022**.

14. A. Rodrigues, A.R.S. Marcal, M. Cunha. PhenoSat – a tool for remote sensing based analysis of vegetation dynamics; Springer, Cham, **2016**; Vol. 20.

15. T. Udelhoven. TimeStats: A Software Tool for the Retrieval of Temporal Patterns From Global Satellite Archives. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, 4 (2), 310–317.

16. V. Maus, G. CĂmara, R. Cartaxo, et al. A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, 9 (8), 3729–3739.

17. D. Ashourloo, H.S. Shahrabi, M. Azadbakht, et al. A Novel Automatic Method for Alfalfa Mapping Using Time Series of Landsat-8 OLI Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, 11 (11), 4478–4487.

18. J. Xiao, H. Wu, C. Wang, H. Xia. Land Cover Classification Using Features Generated from Annual Time-Series Landsat Data. *IEEE Geosci. Remote Sens. Lett.* **2018**, 15 (5), 739–743.

19. H. Yin, D. Pflugmacher, R.E. Kennedy, D. Sulla-Menashe, P. Hostert. Mapping annual land use and land cover changes using MODIS time series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, 7 (8), 3421–3427.

20. F. Petitjean, J. Weber. Efficient satellite image time series analysis under time warping. *IEEE Geosci. Remote Sens. Lett.* **2014**, 11 (6), 1143–1147.

21. X. Zhang, M. Wang, K. Liu, J. Xie, H. Xu. Using NDVI time series to diagnose vegetation recovery after major earthquake based on dynamic time warping and lower bound distance. *Ecol. Indic.* **2018**, 94, 52–61.

22. Y. Burstyn, A. Gazit, O. Dvir. Hierarchical Dynamic Time Warping methodology for aggregating multiple geological time series. *Comput. Geosci.* **2021**, 150.

23. M. E. D. Chaves, M. C. A. Picoli, I. D. Sanches. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sens.* **2020**, 12 (18), 3062.

24. S.N. Goward, G. Chander, M. Pagnutti, et al. Complementarity of ResourceSat-1 AWiFS and Landsat TM/ETM+ sensors. *Remote Sens. Environ.* **2012**, 123, 41–56.

25. L. Eklundh, P. Jönsson. TIMESAT: A Software Package for Time-Series Processing and Assessment of Vegetation Dynamics; Sweden, **2015**; pp 141–158.

26. P. Jönsson, L. Eklundh. TIMESAT - A program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **2004**, 30 (8), 833–845.

27. L. Eklundh, P. Jönsson. TIMESAT 3.3 with seasonal trend decomposition and parallel processing Software Manual; Lund and Malmo University, Sweden, **2017**.

28. H. do N. Bendini, L.M.G. Fonseca, T.S. Körting, I.D.A. Sanches, R.F.B. Marujo. Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters. In *Simposio Brasileiro de Sensoriamento Remoto*; Santos, At, **2017**; Vol. 18, pp 4267–4274.

29. S.K. Apat, J. Mishra, K.S. Raju, N. Padhy. The robust and efficient Machine learning model for smart farming decisions and allied intelligent agriculture decisions. *J. Integr. Sci. Technol.* **2022**, 10 (2), 139–155.

30. V.Y. Kullarni, P.K. Sinha. Random Forest Classifier: A Survey and Future Research Directions. *Int. J. Adv. Comput.* **2013**, 36 (1), 1144–1156.

31. M. Savargiv, B. Masoumi, M.R. Keyvanpour. A new random forest algorithm based on learning automata. *Comput. Intell. Neurosci.* **2021**, 2021.

32. B. Matsushita, W. Yang, J. Chen, Y. Onda, G. Qiu. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: A case study in high-density cypress forest. *Sensors* **2007**, 7 (11), 2636–2651.

33. H. Nemmour, Y. Chibani. Neural network combination by fuzzy integral for robust change detection in remotely sensed imagery. *EURASIP J. Appl. Signal Processing* **2005**, 2005 (14), 2187–2195.

34. T.K. Behera, S. Bakshi, P.K. Sa. Vegetation Extraction from UAV-based Aerial Images through Deep Learning. In *Computers and Electronics in Agriculture*; **2022**; Vol. 198.

35. B. Van Jaarsveld, S.M. Hauswirth, N. Wanders. Machine learning and Global Vegetation : Random Forests for Downscaling and Gapfilling. *Hydrol. Earth Syst. Sci.* **2023**, No. February, 1–24.

36. Y. Sun, D. Lao, Y. Ruan, C. Huang, Q. Xin. A Deep Learning-Based Approach to Predict Large-Scale Dynamics of Normalized Difference Vegetation Index for the Monitoring of Vegetation Activities and Stresses Using Meteorological Data. *Sustain.* **2023**, 15 (8), 6632.

37. Y. Ma, Y. Hu, G.R. Moncrieff, et al. Forecasting vegetation dynamics in an open ecosystem by integrating deep learning and environmental variables. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, 114, 103060.

Journal of Integrated Science and Technology

J. Integr. Sci. Technol., 2024, 12(1), 717   14