# ML-based Vegetative Drought Prediction employing Satellite Remote Sensing and Precipitation Datasets

Jyoti S. Shukla
*Dept. of Electrical Engineering*
*Indian Institute of Technology*
Dharwad, India
jyoti.shukla.21@iitdh.ac.in

Rahul Jashvantbhai Pandya
*Dept. of Electrical Engineering*
*Indian Institute of Technology*
Dharwad, India
rpandya@iitdh.ac.in

2023 IEEE 13th International Conference on Pattern Recognition Systems (ICPRS) | 979-8-3503-3337-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICPRS58416.2023.10179057

***Abstract*—In an agriculturally-dominant economy such as India, gross crop production and adequate land use constitute essential variables to contemplate, when assessing the socioeconomic well-being of all societal segments. Indubitably, agriculture growth relies on climatic factors and the availability of natural resources to suffice the demand-supply chain. Drought, a natural hazard, grievously impairs land conditions and water availability determinants, bringing in a crisis in agriculture productivity and several other interlinked spheres. In this paper, we demonstrate the effective implementation of Linear Regression, a widely favored Machine Learning (ML) - based model, to capture the short-term drought severity, incorporating two datasets comprising NOAA/AVHRR satellite-data-derived Vegetation Health Index (VHI) and VHI+Rainfall readings as the features. The Karnataka state of India is selected as the Region of Interest (ROI), and the compiled dataset spans the period 1981 through 2021. A thorough comparison of the proposed model with several anointed reference models is proffered. The model is further evaluated on datasets for various timescales, and the outcomes are documented. The proposed ML model outperforms the other conventional models as validated through the performance metrics, inclusive of coefficient of determination (R2), mean squared error, and mean absolute error, obtaining values of 0.98, 0.0076, and 0.048 for the VHI dataset and 0.97, 0.0148, and 0.067 for VHI+Rainfall datasets respectively on the selected ROI.**

***Index Terms*—Linear Regression (LR), Vegetation Health Index (VHI), Rainfall, Drought prediction, Machine Learning (ML), Satellite Remote Sensing (RS) data**

## I. Introduction

The increase in global warming, exploitation, and excessive use-misuse of scarce nonrenewable natural resources have stemmed in catastrophic harm and a steep decline of the resources, which are vital for a sustainable civilization. The repercussions are witnessed in factors such as haphazard weather changes, seasonal fluctuations, and depleting natural resources leading to unprecedented natural disasters such as flooding, crop failure, droughts, famine, etc. Identically, drought is one of the most devastating natural disasters, with long-term consequences for agriculture and the economy [1], [2]. Owing to global digitalization and technological advancements ensuing, a more sophisticated solution for drought prediction and monitoring is desired. The use of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) is a potential opportunity for drought monitoring solutions.

ML-based drought monitoring is a burgeoning field of research in which multiple research initiatives have been performed leveraging diverse Remote Sensing (RS) data-derived vegetative indices such as the Vegetation Condition Index (VCI), Temperature Condition Index (TCI), Vegetation Health Index (VHI), Normalized Difference Vegetation Index (NDVI), and several others [1]–[5]. The RS indices are contrasted with certain meteorological indices in multiple research endeavors, including the Standardized Precipitation Index (SPI), Standardized Streamflow Index (SSI), and Standardized Precipitation - Evapotranspiration Index (SPEI) [1], [5], [6]. For Spatio-Temporal (ST) vegetative drought analysis and associated disciplines, sophisticated ML models, notably CART, SVM, RF, KNN, SVR, and Multiple Linear Regression (MLR) are explored [1], [2], [4]. Several other ML models, such as XGBoost and GBoost, forecast agricultural production by synthesizing meteorological data, including average temperature, air quality index, and precipitation. Furthermore, the combined indices supplied to the models execute a trend extraction mechanism [7]. In [8]–[10], trend extraction and analysis utilizing VHI and rainfall characteristics are conducted employing the Mann-Kendall (MK) test, and the trend magnitudes are further recouped through the Theil-slope Sen's estimator.

Meanwhile, several DL models have proven to perform satisfactorily in this discipline; in [5], a DL-based strategy is utilized where the deep belief networks are trained using SSI across multiple durations to forecast drought accurately. Analogously, a multitude of DL models, encompassing MLP and Radial basis function, ARIMA, ARIMAX, LSTM, Convolutional Neural Network (CNN)-YOLO and ConvLSTM models engaged various vegetative indices for effective drought prediction [5], [11]–[14]. Also, an ensemble learning approach is considered in [15] wherein the DL-based CNN model and Barnacles mating optimizer are linked together for the task of drought prediction with various satellite data-derived vegetative indices given as inputs to the model.

In this study, we concentrate on effective vegetative drought forecasting utilizing two datasets, the NOAA/AVHRR satellite data-derived VHI dataset coupled with precipitation data and

979-8-3503-3337-4/23/$31.00 ©2023 IEEE

the other being a standalone VHI dataset, for the period 1981 to 2021. The development of the ML-based model, Linear Regression (LR), for this heterogeneous feature set of rainfall and satellite data (VHI data) integrated together, is a novel combination in this domain to the best of the author's knowledge. Furthermore, the proposed model outcomes are contrasted to several ML/DL-based reference models to ensure its competitive advantage, along with presenting the robustness of the model by testing it on different timescales in the range of 1 to 4 timeshifts. By virtue of the metric scores, enveloping, Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R2), the model performances are further elucidated. The remainder of this paper is oriented as specified: Section II provides background about the datasets and processing. The proposed model's approach is described in Section III. The results obtained by the model are detailed in Section IV, and the paper is concluded in Section V.

## II. Study Area and Dataset Development

### A. Study Area

The study area selected for this work is the entire Karnataka state of India, situated in the southwest, at the confluence of the Western and Eastern Ghats into the Nilgiri Hills. It is divided into three geographical regions: the Coastal Plains, the Sahyadri, and the Deccan Plateau. Geographically, the chosen Region of Interest (ROI) is blessed with a wide assortment of topological variations with an abundance of nature, extending over the western edge of the Deccan peninsular region of India, spanning $191,791\ km^2$. The spatial coordinates for the ROI locate approximately between 11.5°- 18.5° North latitudes and 74°- 78.5° East longitudes. As portrayed in Fig. 1, the Karnataka state's landscape largely comprises agricultural and forested areas, substantially impacting the state's climatic and weather conditions.

The multispectral satellite images obtained and captured by the AVHRR sensor of NOAA satellites are global; furthermore, the derived VHI values are also worldwide data equipped by STAR NESDIS and are subsetted to the Karnataka region. With such diverse terrains dispersed across the state, analyzing drought occurrence and vegetation health is daunting, nevertheless pertinent in assisting with considerable probable socioeconomic hardships.

### B. Data Description

In this study, leveraging the meteorologically-collected rainfall and satellite-data-derived VHI data, two datasets are generated to assess the LR model. One of the datasets is developed utilizing VHI values autonomously, while the second is a combination dataset devised by combining satellite data and rainfall for the designated ROI. Furthermore, the performances of the two datasets are compared.

*1) VHI Dataset:* Satellite data processing has progressed owing to the proliferation of space technology advancements and the accessibility of satellite-based products. With AI/ML, the objective of fetching information through satellite data

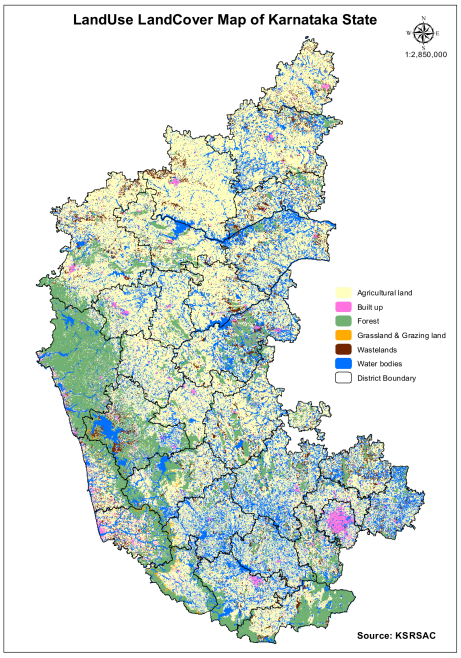


Fig. 1. Land cover of the Karnataka state of India [16].

evolved to be more attainable. In this work, the weekly VHI data is acquired from the NOAA STAR NESDIS open source portal from the $35^{th}$ week of 1981 to the $52^{nd}$ week of 2021. The drought severity levels have been categorized based on the values of VHI (which ranges from 0 to 100) [2]. Table I demonstrates that a VHI value greater than 40 denotes healthy vegetation, and the degrading VHI leads to drought extremes. The calculation of VHI appertains to the computation of NDVI, VCI, and TCI, which are formulated in Eq. (1 - 3) respectively, and further computes the VHI as per Eq. 4.

$$NDVI = \frac{(band_2 - band_1)}{(band_2 + band_1)},\ NDVI \in [-1, 1] \quad (1)$$

$$VCI = \frac{NDVI_i - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} * 100,\ VCI \in [0, 100] \quad (2)$$

$$TCI = \frac{BT_{\max} - BT_i}{BT_{\max} - BT_{\min}} * 100,\ TCI \in [0, 100] \quad (3)$$

$$VHI = a * VCI + (1 - a) * TCI,\ VHI \in [0, 100] \quad (4)$$

In Eq. (1 - 4), $'i'$ denotes the pixel value, $BT$ represents the brightness temperature, $band_2$ and $band_1$ represent the near-infrared $(0.725 - 1.00\ \mu m)$ and red $(0.58 - 0.68\ \mu m)$ spectral bands of NOAA/AVHRR, respectively, and $'a'$ is a parameter used to decide the contribution ratio from both VCI and TCI which in this study is established as $0.5$.

*VHI Dataset Development:* Several missing weeks are encountered in the collected VHI data and are supplanted through the mean refilling approach; following that, the data is further preprocessed using padding, scaling, and cropping operations. From the global VHI data, the ROI is cropped and padded to obtain a structured shape. Post the mean-refilling approach,

TABLE I
DROUGHT SEVERITY CHART BASED ON VHI

| Category | Drought severity | VHI |
|---|---|---|
| – | Normal | $> 40$ |
| L0 | Abnormally dry | $36 - 40$ |
| L1 | Moderate drought | $26 - 35$ |
| L2 | Severe drought | $16 - 25$ |
| L3 | Extreme drought | $6 - 15$ |
| L4 | Exceptional drought | $0 - 5$ |

2105 satellite images are fetched for 1981-2021 and are further employed to generate dataset - I to be utilized in this study:

- Dataset - I (VHI dataset): Post cropping and padding operations, the dimensions obtained for the Karnataka VHI dataset are (192,128,1). The data is further scaled through a min-max scaler to fetch values in the range of [0,1], and the annotated labels are developed as the two-time-shifted compositions of the input images post-scaling. The splitting resulted in 95 samples for testing and 2010 for training sets.

*2) Precipitation Data:* Precipitation is a crucial parameter for determining changes in land cover and other related climatic aspects, which concerns preserving vegetative health and alerting for any imminent drought crisis [17]. Precipitation data is obtained in the gridded format from the open-source data archive portal of Climate Research & Services group of the Indian meteorological department, Pune, India, for the period 1981-2021 in the NetCDF format with a fine spatial resolution of $0.25^\circ \times 0.25^\circ$ collected from nearly 7000 rain gauge stations [18].

*Precipitation-based Dataset Development:* The daily national gridded precipitation data in NetCDF format is converted into a 2D raster image dataset using GIS-based operations such as interpolation, re-sampling, and re-projecting the gridded data with attributes encompassing latitude, longitude, rain, and time. The resolution disparity between the VHI and rainfall data is resolved by interpolating the ROI, i.e., Karnataka state, which is cropped from the national precipitation data. Following that, the daily rainfall data are averaged to yield a weekly representation corresponding to the temporal resolution of dataset - I, which is then integrated with the VHI data to generate the second dataset of the current study. The purpose of this alliance is to witness how rainfall features influence VHI-based drought analysis. The total number of precipitation image samples in the weekly rainfall dataset enumerated 2105 images for 1981-2021.

- Dataset - II (VHI+Rainfall dataset): The fetched rainfall images are reshaped and padded, acquiring the shape (192,128,1), following which a min-max scaler is applied to the sample, and the data is further scaled, abridging the rainfall values into the range of [0,1]. A train-test data split is performed, entailing 95 in testing and 2010 samples in training set similar to dataset - I. Subsequently, the rainfall data are merged with the exact VHI dataset composition for the given ROI to produce a concatenated input with the shape (192,128,2), with the two time-shifted VHI values as the label having a shape (192,128,1). Such a form of dataset configuration in vegetative drought monitoring applications is novel.

## III. METHODOLOGY

The proposed study aims to forecast the likelihood of vegetative drought by constructing an ML-based LR model on the generated datasets. Certainly, rainfall explicitly impacts vegetative health; hence, substituting it as one of the feature parameters for next time stamp VHI forecasting is a cardinal experiment.

### *A. Proposed Model: Linear Regression*

LR is a profound regression algorithm that describes the linear relationship between a dependent variate $y$ on $'k'$ independent variables $x_1, x_2, ....x_k$. The LR model is straightforward and employs the supervised model technique, which possesses historical values to forecast future outcomes, as demonstrated in Fig. 2. MLR, an extension of LR, refers to the regression problem undertaking more than one independent variable for predicting the output variable, $y$, foreknown to be a continuous entity [19], [20] and is mathematically formulated as denoted in Eq. 5,

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + ..... + b_k x_{ki} + \epsilon_i \tag{5}$$

where, $y_i$ is the predicted output for the $i^{th}$ sample of total $'n'$ number of samples, $b_0$ is the intercept, $b_1, b_2, ..., b_k$ denote the regression coefficients for $k$ variables, and $\epsilon_i$ denotes the residual unexplained error influencing the value of $y_i$. In this study, for dataset - I, the value of $k = 1$, resembling single variate LR (only VHI as a feature), and for dataset - II, as two separate feature sets, VHI+Rainfall, are employed for input, hence $k = 2$. Considering the latter case, Eq. 5 reduces to Eq. 6.

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \epsilon_i \tag{6}$$

The regression coefficients, $b_0$, $b_1$, and $b_2$, are estimated through the model. LR targets to search for the best-fit line which satisfactorily describes the linear association shared by dependent and independent variables, along with minimizing the value of $\epsilon_i$ to the least possible value. The more prominent the linear connection, which also asserts, that the preciseness with which the independent factors explain the variability of the dependent variable, the finer the predictions are derived. To discover the best-fit line, the least-sum-of-squares of error terms (E) method, is the most preferred [19]. Therefore, the estimation accuracy for the regression coefficients is recursively improved, engrossing the error value, E, until a nominal value is achieved; in Eq. 7, $\bar{Y_i}$ is the ground truth value for the $i^{th}$ sample.

$$\mathrm{E} = \sum_{i=1}^{n} \left(\bar{Y_i} - b_0 - b_1 x_{1,i} - b_2 x_{2,i}\right)^2 \tag{7}$$

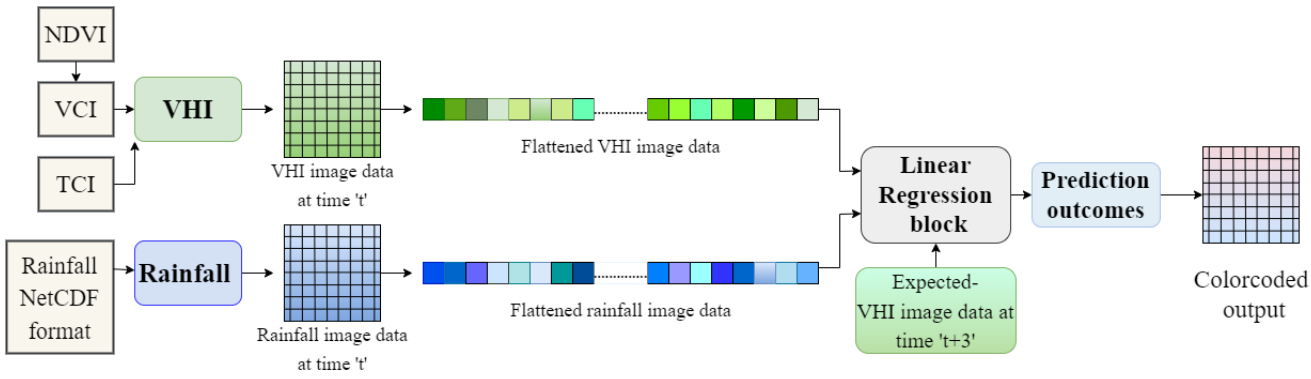


Fig. 2. System model for LR - based drought prediction employing VHI and Rainfall.

The second half part of Eq. 7 represents the predicted value for the $i^{th}$ sample, substituted as $\hat{y_i}$, thus Eq. 7 is rewritten as Eq. 8.

$$\mathrm{E} = \sum_{i=1}^{n} \left(\bar{Y_i} - \hat{y_i}\right)^2 \tag{8}$$

From Eq. 8, the value of residual, $\epsilon_i$, for the $i^{th}$ sample is extracted and is embodied in Eq. 9.

$$\epsilon_i = \bar{Y_i} - \hat{y_i} \tag{9}$$

Subsequently, the sum-of-squares error calculation subsides to Eq. 10.

$$\mathrm{E} = \sum_{i=1}^{n} \left(\epsilon_i^2\right) \tag{10}$$

Consecutively, optimizing the derivative in Eq. 11 yields the optimal value for the regression coefficients.

$$\frac{\partial \mathrm{E}}{\partial b_j} = 2\sum_{i=1}^{n} \epsilon_i \frac{\partial \epsilon_i}{\partial b_j} = 0, \quad j = 0, 1, 2 \tag{11}$$

In Eq. 11, superseding the value of $\epsilon$ from Eq. 9 renders the derivative outcome as demonstrated in Eq. 12.

$$\frac{\partial \mathrm{E}}{\partial b_j} = -2\sum_{i=1}^{n} \epsilon_i \frac{\partial \hat{y_i}}{\partial b_j} = 0, \quad j = 0, 1, 2 \tag{12}$$

The LR model pursues to minimize the error term E and delivers the learned values of the regression coefficients that can precisely account for the relationship between inputs and outputs. The consolidated training dataset is flattened and applied to the LR model as displayed in Fig. 2, following which prediction is made on the unseen test dataset through the trained LR model, and the performance metrics are extracted for further assessment.

### *B. Recommended Models for Performance Estimation*

Assimilating the generated datasets, numerous ML and DL models are evaluated in this study to compare the performance of the preferred time-series ML/DL models to the proposed LR model. Moreover, a comparison is mustered, incorporating quantitative scores and drought maps.

- DL models - the time-based properties in the dataset are addressed by temporally sampling the data operated as input to the DL models. In this research, a two-time-shift temporal sampling of the input to the DL models is adopted.

  1) ConvLSTM+Conv3D model: the ConvLSTM and Conv3D layers are used in conjunction to forecast the short-term VHI value. The model architecture is comprised of 8, 16, 64, 128, 64, and 16 convolutional filters.
  2) ConvLSTM+Conv2D model: this resembles the previous DL model except that the Conv3D layer is swapped out for the Conv2D layer, and there are 8, 32, 128, 64, and 8 filters included.
  3) Conv3D model: the effectiveness of the Conv3D method is assessed in this model architecture, using layers with the following filter values: 8, 8, 16, and 1.
  4) 1DCNN: the 1DCNN architecture is a popular paradigm for time-series processing and is implemented here with filters configured as follows: 64, 50, and 24576; the last layer occupies filters equal to the expected flattened output size, in this case, 24576.

  For all the models, the filter dimension is fixed to three, with normalization performed after every layer, '$SGD$' is adopted as the optimizer, and '$ReLU$' is utilized as the activation function.
- ML models - KNN and Decision Tree (DT) models account for the ML models assessed for comparison. Owing to the limitations of computing resources and dataset enormity, the SVR and MLP models could not be performed, making LR an efficient ML model to process the data in a computationally efficient way.
- Ensemble Models - the ensemble models are developed using two or more models in a congregation; in this work, an ensemble of the DL-based ConvLSTM+Conv3D and ML-based DT, i.e., ConvLSTM+Conv3D+DT model is created to contrast it with the LR performance.

### C. Performance Metrics for Models' Comparison

The proposed and reference models are quantitatively compared by imbibing metrics such as MSE, MAE, R2 score, and computing time.

- MSE determines the proximity of the regression line to a set of points; therefore, for precise prediction outcomes, the value of MSE is expected to be kept minimal and is calculated through Eq. 13.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\bar{Y}_i - \widehat{y}_i\right)^2 \tag{13}$$

- MAE expresses the definitive aberration amidst the attributes regardless of their direction, computed through Eq. 14; consequently, MAE is aimed to be toward the lower edge for accurate forecasting.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\bar{Y}_i - \widehat{y}_i\right| \tag{14}$$

- The R2 score specifies the ratio between marked variance to the absolute variance accustomed in the data and is formulated as depicted in Eq. 15. R2 ranges in [0,1], with values nearing 1, representing a good regression model fit.

$$R2 = 1 - \frac{\sum_i \left(\bar{Y}_i - \hat{y}_i\right)^2}{\sum_i \left(\bar{Y}_i - \tilde{y}\right)^2} \tag{15}$$

In the Eq. (13 - 15), the notations depict $i$ as the $i^{th}$ input sample, $\widehat{y}_i$ as the expected output value, $\bar{Y}_i$ as the real label value, $\tilde{y}$ denoting the mean of the observed true value and $n$ as the total number of samples. The computing time considered for comparison represents the training time undertaken by the models for each dataset.

## IV. Results and Discussions

The emanated outputs by the models are unscaled into the absolute VHI range of [0,100] and further color-coded in order to disband into vivid degrees of drought severity as per Table I. The simulation experiments are performed employing the graphics cards comprising a 40 Gb NVIDIA (DGX) and 16 Gb NVIDIA RTX (A4000) GPU.

### A. Statistical Comparison among Models

A statistical comparison of the models is exemplified among the datasets. The models' prediction results are applied to derive the performance metrics, MSE, MAE, and R2 scores. Moreover, a metric comparison is established based on the computational time mandated by the datasets for training, as it essentially decides the preferability of the model.

The performance of models on the dataset - I is exhibited in Table II where a single variable LR model is operated with the VHI dataset as the independent variable. The metric scores for LR conceivably transcend the reference models, with mean R2, maximum R2, mean MSE, and MAE being $0.92$, $0.98$, $0.0076$, and $0.048$, respectively. Moreover, LR directs minimal computing time of $2$ minutes compared to the DL models, which demand much more time and memory.

TABLE II
Dataset - I: Quantitative Comparison between the Models

| Model | Epochs | Mean R2 | Max R2 | Mean MSE | Mean MAE | Computational time (in minutes) |
|---|---|---|---|---|---|---|
| LR | – | **0.92** | **0.98** | **0.0076** | **0.048** | **2** |
| KNN regressor | – | 0.88 | 0.92 | 0.0088 | 0.050 | $< 1$ |
| ConvLSTM+Conv3D+DT | – | 0.86 | 0.915 | 0.0100 | 0.0626 | 280 |
| ConvLSTM+Conv3D | 30 | 0.83 | 0.93 | 0.0131 | 0.0771 | 150 |
| DT regressor | – | 0.81 | 0.95 | 0.017 | 0.071 | 280 |
| 1DCNN | 100 | 0.78 | 0.83 | 0.0175 | 0.076 | 4 |
| Conv3D | 10 | 0.29 | 0.35 | 0.060 | 0.19 | 73 |
| ConvLSTM+Conv2D | 40 | 0.22 | 0.37 | 0.075 | 0.179 | 303 |

TABLE III
Dataset - II: Quantitative Comparison between the Models

| Model | Epochs | Mean R2 | Max R2 | Mean MSE | Mean MAE | Computational time (in minutes) |
|---|---|---|---|---|---|---|
| LR | – | **0.85** | **0.97** | **0.0148** | **0.067** | **3** |
| KNN regressor | – | 0.81 | 0.92 | 0.021 | 0.071 | $< 1$ |
| ConvLSTM+Conv3D | 30 | 0.80 | 0.91 | 0.015 | 0.077 | 62 |
| ConvLSTM+Conv3D+DT | – | 0.76 | 0.93 | 0.0200 | 0.070 | 480 |
| ConvLSTM+Conv2D | 40 | 0.73 | 0.91 | 0.0591 | 0.0110 | 81 |
| DT regressor | – | 0.72 | 0.90 | 0.022 | 0.080 | 480 |
| 1DCNN | 100 | 0.70 | 0.80 | 0.0188 | 0.0799 | 6 |
| Conv3D | 10 | 0.63 | 0.70 | 0.0277 | 0.1268 | 26 |

The models' performance metrics on the dataset - II, is demonstrated in Table III; the LR model performance excels when interacting with mixed features also, as documented by the maximum R2 score of $0.97$ and the significantly short computational time of $3$ minutes for training. Nonetheless, a necessary inference asserts that the performance of dataset - I surpasses dataset - II for the specified ROI, revealing that VHI features are more robust to integrated features at learning trends for the given ROI.

Furthermore, as illustrated in Table IV, the proposed model's prediction accuracy is evaluated for multiple timeshifts, and a comparison among the datasets is essayed on predicted outcomes for different timeshifts. The timeshifts correspond to weekly increments; for instance, a timeshift of $1$ predicts the output for the following week given the current input, while a timeshift of $4$ is anticipated to yield a result after $4$ weeks or the next month's VHI condition of the ROI, exploring the model's suitability for temporal comparisons.

TABLE IV
Comparison of LR Model Performance for Different Timescales

| Timeshifts | Mean R2 | | Max R2 | | Mean MSE | | Mean MAE | |
|---|---|---|---|---|---|---|---|---|
| | VHI | VHI+Rainfall | VHI | VHI+Rainfall | VHI | VHI+Rainfall | VHI | VHI+Rainfall |
| 1 | 0.92 | 0.92 | 0.98 | 0.97 | 0.0076 | 0.0076 | 0.048 | 0.048 |
| 2 | 0.92 | 0.85 | 0.98 | 0.97 | 0.0076 | 0.0148 | 0.048 | 0.067 |
| 3 | 0.91 | 0.77 | 0.96 | 0.92 | 0.0081 | 0.023 | 0.050 | 0.085 |
| 4 | 0.90 | 0.5 | 0.95 | 0.91 | 0.0087 | 0.0519 | 0.052 | 0.122 |

As is outlined in Table IV, the LR model consistently delivers satisfactory performance for dataset - I, even for monthly forecasts represented by four timeshifts. However, for dataset - II, a declining performance is observed with increased temporal shifts, suggesting that the mixed features work sufficiently alone for short-term drought predictions for the given ROI.

### B. Drought Maps Comparison among Models

The predicted outcomes are segmented and color-coded to generate drought maps for a qualitative comparison, where a range of color hues represents the different drought severity levels, as per the legend in Fig. 3 and 4. The predicted VHI map is subdivided into a color-coded, segmented map encapsulating values in the [0, 255] range. The drought maps

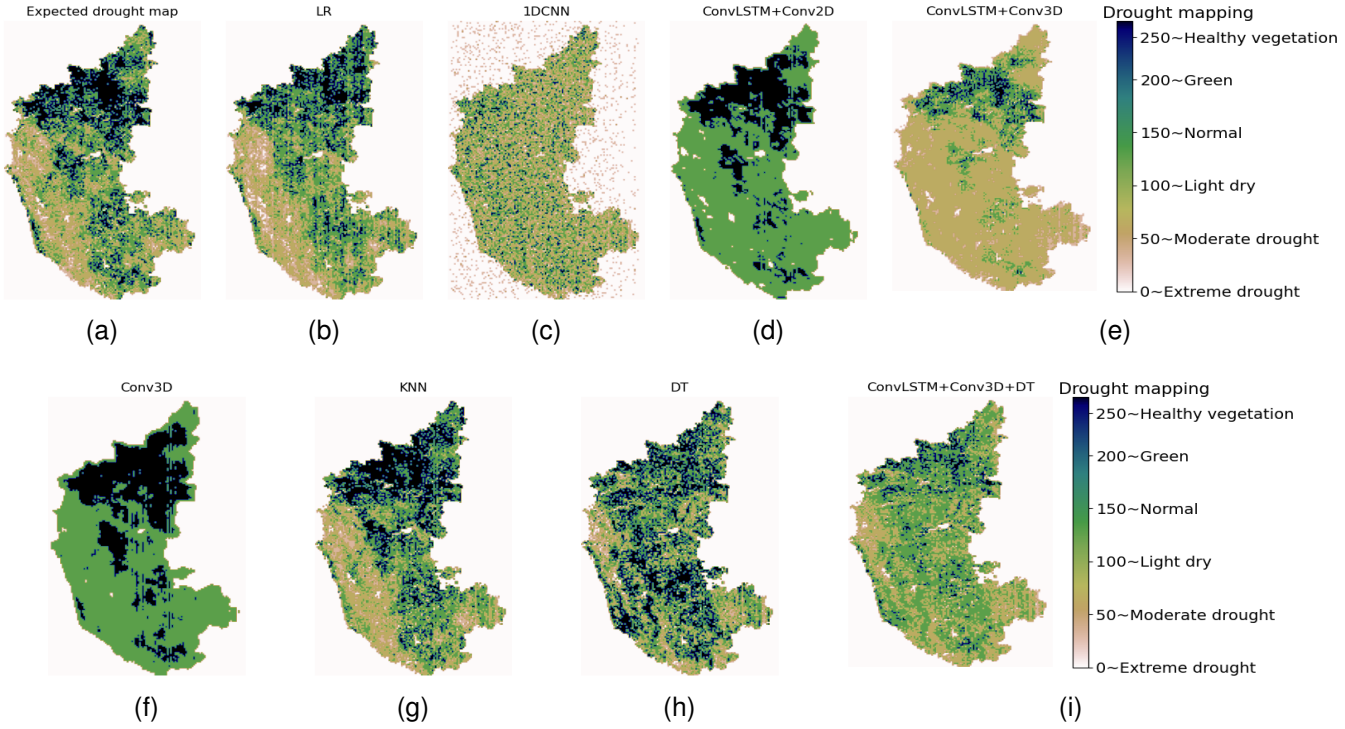


Fig. 3. Drought map comparison for the proposed and recommended models for dataset - I: (a) Expected drought map (b) LR (c) 1DCNN (d) ConvLSTM+Conv2D (e) ConvLSTM+Conv3D (f) Conv3D (g) KNN (h) DT (i) ConvLSTM+Conv3D+DT.

for dataset - I, comprising the VHI dataset, are portrayed in Fig. 3 where Fig. 3 (a) denotes the ground truth or desired drought map, and Fig. 3 (b) indicates the drought map for the suggested model of LR. In the drought maps, the darker hues represent good vegetation cover, whereas the mud shade represents a deteriorating green cover. The generated drought maps for recommended ML/DL models, as displayed in Fig. 3 (c - i), are held ineffective in accurately simulating the land situation for the ROI. On the contrary, the LR model is able to produce a nearly similar outcome which further validates the good metric scores obtained for the LR model.

The drought maps for dataset - II are displayed in Fig. 4, where Fig. 4 (a) depicts the intended or ground truth drought map, and Fig. 4 (b) depicts the drought map for the proposed model of LR. The LR model's potential to yield a remarkably identical response further bolsters the favorable metric scores earned by the model, despite the different features delivered as input. Fig. 4 (c - i) represent the drought maps for the reference models which, as evident, are subpar qualitatively as compared to the drought map of the proposed model. Furthermore, this underlines LR's prowess in handling multivariable datasets and its ability to support VHI forecasting.

## V. Conclusion

The current research adopts satellite and meteorological data to demonstrate the application of LR for vegetative drought analysis. The proposed model is applied to two datasets (VHI and VHI+Rainfall) with the Karnataka state as ROI from 1981 to 2021 and is compared statistically and qualitatively to several other recommended ML/DL-based models. The results obtained per evaluating models on the two datasets reveal that the LR model accurately anticipates short-term drought conditions for the specified ROIs with the mean R2, maximum R2, mean MSE, and mean MAE scores acquiring values of $0.92$, $0.98$, $0.0076$, and $0.048$ for dataset - I and $0.85$, $0.97$, $0.0148$, and $0.067$ for dataset - II respectively. The model corroborated with promising results on various feature sets akin to converging the contribution of rainfall with satellite data and performed satisfactorily over the Indian state of Karnataka. The computational training time devoured by the LR model on the datasets is optimal, contrary to the other models. Moreover, the excellent performance metrics for the suggested model demonstrate the method's viability for short-term drought forecasting and render it scalable to other applications.

This study's future objectives entail consolidating seasonal drought maps with seasonal trend analysis and further enhancing the current model outcomes by diversifying the feature sets through a greater ROI or increasing the feature temporality.

## Acknowledgment

This research is conducted under the aegis of the ANTRIX Corporation through the aid in developing the Space Data Science Lab at the Indian Institute of Technology Dharwad, Karnataka, India. The GPU facility acquired under SERB grant number 00047 and the DGX GPU equipped by ANTRIX Corporation is partially incorporated to execute the experimentations in this study. The authors are also grateful to Dr.

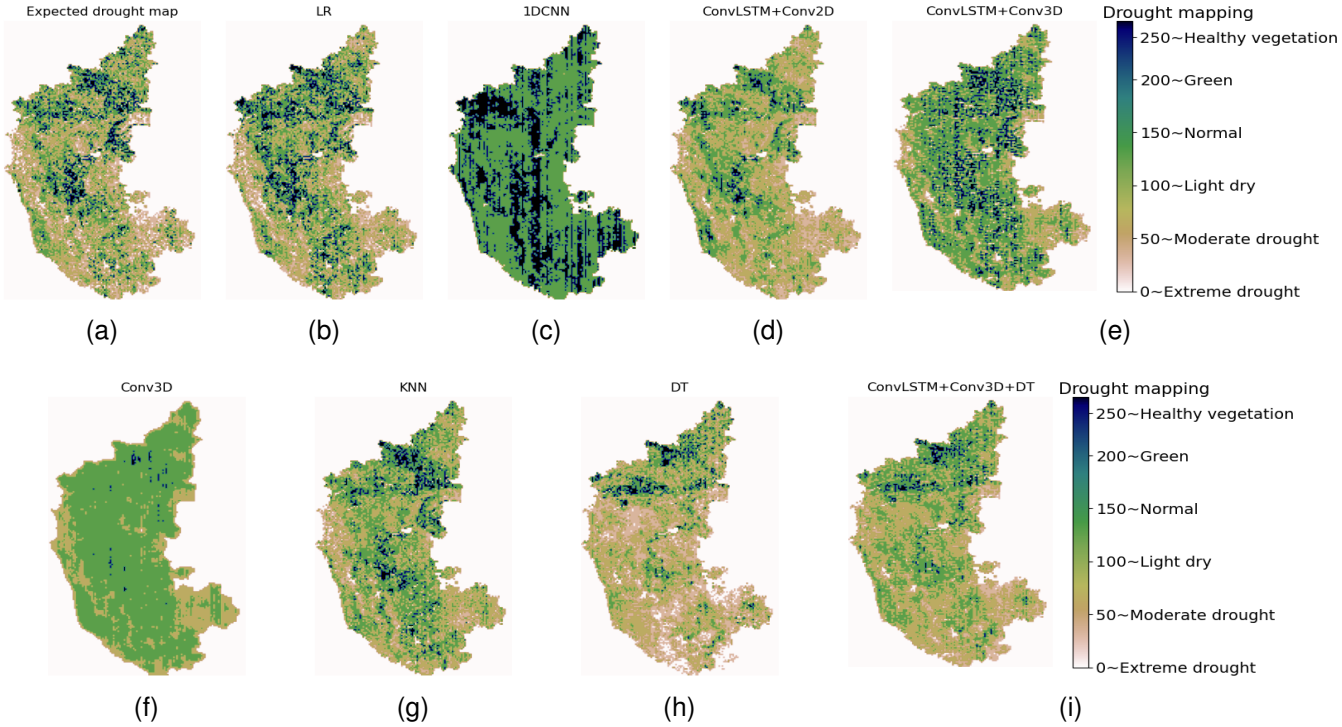


Fig. 4. Drought map comparison for the proposed and recommended models for dataset - II: (a) Expected drought map (b) LR (c) 1DCNN (d) ConvLSTM+Conv2D (e) ConvLSTM+Conv3D (f) Conv3D (g) KNN (h) DT (i) ConvLSTM+Conv3D+DT.

Rajshekhar V. Bhat, Assistant Professor, IIT Dharwad, for his advice and assistance in procuring the necessary resources.

## REFERENCES

[1] S. Y. J. Prasetyo, K. D. Hartomo, M. C. Paseleng, D. W. Candra, and B. H. Simanjuntak, "The machine learning to detect drought risk in central Java using Landsat 8 OLI remote sensing images," in *5th Int'l Conf. on Science and Tech. (ICST)*, vol. 1, Jul. 2019, pp. 1–6.

[2] S. Zhong, L. Di, Z. Sun, Z. Xu, and L. Guo, "Investigating the long-term spatial and temporal characteristics of vegetative drought in the contiguous United States," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 836–848, Mar. 2019.

[3] Z. Li, Y. Han, and T. Hao, "Assessing the consistency of remotely sensed multiple drought indices for monitoring drought phenomena in continental China," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5490–5502, Aug. 2020.

[4] S. Perera, W. Li, E. Linstead, and H. El-Askary, "Forecasting vegetation health in the Mena region by predicting vegetation indicators with machine learning models," in *IEEE Int'l Geosc. and Remote Sens. Symposium (IGARSS)*, Sept. 2020, pp. 4642–4645.

[5] N. A. Agana and A. Homaifar, "A deep learning based approach for long-term drought prediction," in *SoutheastCon*, May 2017, pp. 1–8.

[6] K. S. Dayal, R. C. Deo, and A. A. Apan, "Application of hybrid artificial neural network algorithm for the prediction of standardized precipitation index," in *IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 2962–2966.

[7] A. Shah, R. Agarwal, and B. Baranidharan, "Crop yield prediction using remote sensing and meteorological data," in *Int'l Conf. on Artif. Intellig. and Smart Syst. (ICAIS)*, Apr. 2021, pp. 952–960.

[8] S. Zhong, Z. Xu, Z. Sun, E. Yu, L. Guo, and L. Di, "Global vegetative drought trend and variability analysis from long-term remotely sensed data," in *8th Int'l Conf. on Agro-Geoinfo.*, Sep. 2019, pp. 1–6.

[9] A. Kumar and S. Singh, "A review on Indian summer monsoon rainfall prediction using machine learning techniques," in *2nd Int'l Conf. on Secure Cyber Comp. and Communs. (ICSCCC)*, Jul. 2021, pp. 524–528.

[10] F. Muthoni, "Spatial-temporal trends of rainfall, maximum and minimum temperatures over west Africa," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2960–2973, May 2020.

[11] Y. Dhyani and R. J. Pandya, "Deep learning oriented satellite remote sensing for drought and prediction in agriculture," in *IEEE 18th India Council Int'l Conf. (INDICON)*, Dec. 2021, pp. 1–5.

[12] S. Zhong, Z. Xu, and L. Cao, "Evaluating performance of prediction of vegetative drought using classic and recent sequence-based models," in *9th Int'l Conf. on Agro-Geoinfo.*, Sep. 2021, pp. 1–6.

[13] M. I. Habibie, T. Ahamed, R. Noguchi, and S. Matsushita, "Deep learning algorithms to determine drought prone areas using remote sensing and GIS," in *IEEE Asia-Pacific Conf. on Geosc., Electro. and Remote Sens. Tech.(AGERS)*, Dec. 2020, pp. 69–73.

[14] D. Yang, S. Zhong, X. Mei, F. Niu, and W. Zhu, "Correlation analysis of vegetation drought and influencing factors and a preliminary prediction model based on Convolutional LSTM," in *10th Int'l Conf. on Agro-geoinfo.*, Aug. 2022, pp. 1–6.

[15] V. Sardar, S. Chaudhari, A. Anchalia, A. Kakati, A. Paudel, and B. N. Bhavana, "Ensemble learning with CNN and BMO for drought prediction," in *IEEE 3rd Global Conf. for Advancement in Tech. (GCAT)*, Dec. 2022, pp. 1–6.

[16] G. o. K. Karnataka State Remote Sensing Applications Centre, "Karnataka geographic information system," Accessed Jan. 2023. [Online]. Available: https://kgis.ksrsac.in/kgis

[17] J. Lu, L. Jia, C. Zheng, J. Zhou, M. V. Hoek, and K. Wang, "Characteristics and trends of meteorological drought over China from remote sensing precipitation datasets," in *IEEE Int'l Geosc. and Remote Sens. Symposium (IGARSS)*, Nov. 2016, pp. 7581–7584.

[18] D. Pai, M. Rajeevan, O. Sreejith, B. Mukhopadhyay, and N. Satbha, "Development of a new high spatial resolution (0.25° × 0.25°) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region," *MAUSAM*, vol. 65, no. 1, pp. 1–18, Jan. 2014.

[19] S. Mouatadid, R. C. Deo, and J. F. Adamowski, "Prediction of SPEI using MLR and ANN: A case study for Wilsons Promontory Station in Victoria," in *IEEE 14th Int'l Conf. on Mach. Learn. and Apps. (ICMLA)*, Dec. 2015, pp. 318–324.

[20] E. Sreehari and S. Srivastava, "Prediction of climate variable using multiple linear regression," in *4th Int'l Conf. on Comp. Communi. and Automat. (ICCCA)*, Dec. 2018, pp. 1–4.