Effective Feature Extraction Using Residual Attention and Local Context Aware Classifier for Crop Yield Prediction

Vinaykumar Vajjanakurike Nagaraju[1], Ananda Babu Jayachandra[2], Balaji Prabhu Baluvaneralu Veeranna[3], Ravi Prakash Madenur Lingaraju[4]

Department of Information Science and Engineering-Malnad College of Engineering, Hassan, Visvesvaraya Technological University, Belagavi, Karnataka, India[1, 2]

Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)-Malnad College of Engineering, Hassan, Visvesvaraya Technological University, Belagavi, Karnataka, India[3]

Department of Artificial Intelligence and Machine Learning-Kalpataru Institute of Technology, Tiptur, Visvesvaraya Technological University, Belagavi, Karnataka, India[4]

***Abstract*—Crop yield forecasting plays a key role in agricultural management and planning which is highly essential for food security and production in regional to global scales. However, a prediction of crop yield is considered a challenging task due to the difficulty in extracting spatial context and local semantic features, and difficulty in handling spatiotemporal relations. In order to address these issues, a comprehensive feature extraction is developed along with an effective deep-learning classifier. In this paper, the Residual Attention and Local Context Aware Classifier (RALCAC) is developed for obtaining appropriate features from the remote sensing crop yield images. The developed RALCAC helps to obtain the spatial context using Residual Attention (RA) module and local semantic information that are beneficial in understanding the detailed depiction of the crop. Further, the Convolutional Long Short Term Memory (ConvLSTM) is used to obtain the prediction of crop yield using the comprehensive features from the RALCAC. The RALCAC is analysed by means of Root Mean Squared Error (RMSE) and coefficient of determination. The existing research such as DeepYield, SSTNN and 3DCNN are used to compare the RALCAC method. The RMSE of RALCAC for the MODIS dataset is 3.257, and it is lesser when compared to the DeepYield.**

***Keywords*—Convolutional long short term memory; crop yield prediction; residual attention and local context-aware network; root mean squared error; spatial context data**

I. Introduction

Agriculture is an enriching field which clears the way out of economic pressure and has a believable macro-economic part in various economies. Crop production is a complicated phenomenon which is influenced by parameters of agro-climatic information. An improvement in crop yield quality and production while minimizing the costs and environmental pollution is a key objective in the precision agriculture [1-3]. Crop yield is referred as a key representation of sustainable development in agricultural field. An appropriate management practices are required to be adopted for stable organisation of land for crop production [4]. The different climatic situations that influence the crop yield are landscapes, soil quality, climatic situations, water quality and availability, genotype, harvest activity planning, pest infestations and so on. Further, the processes and strategies of crop yield are altered along with time and is non-linear and complex, because of an extensive development combination of interrelated factors, categorised and influenced by external and non-arbitrate run aspects [5].

A precise and timely estimation of crop yield before harvesting in a large scale is challenging for administrative planning and food security, specifically in frequently varying global and international situations. Simultaneously, an earlier prediction of yield is frequently needed to accomplish the decision making in transportation, storage, processing, harvest and marketing of agricultural merchandises [6-9]. The crop monitoring is obtained via interviewing farmers, field visits and manual data collection in regional level before informing the local statistical officers. But this manual process is expensive, inconsistent and time-consuming, the data is available only after harvesting [10, 11]. Remote sensing data is primarily confined to perform crop identification and classification for an extended period [12]. Remote sensing technology is discovered as well appropriate to gather the information over agricultural areas in recurrent intervals with lesser amount of time. Thus, the remote sensing offers an important contribution to provide a rapid comprehensive image. These remote sensing images display the crop development circumstances in chronological and geographical way which denotes their own extraordinary ability [13-15].

The crop yield prediction can be applicable in the following applications: 1) Precision agriculture: The predictive insights is used to plan the field operations such as planting, irrigation and schedule for harvesting for enhancing the yield; 2) Agricultural planning and decision-making: Precise yield prediction helps the farmers and investors for managing the risks related to the climate and diseases. The following issues such as inappropriate feature extraction, restriction against the generalization and failure in handling dynamics among the data. The aforementioned issued are taken as motivation for this research. Therefore, the comprehensive feature extraction using RALCAC is developed along with the ConvLSTM for an effective prediction.

The contributions of this research are concise as follows:

- The RALCAC uses the architecture of dual-encoder which improves the feature extraction capacity. The integration of spatial context data is achieved by using the residual attention, while local semantics also obtained that confirms the representation of local features and variations within the crop fields. Therefore, the combination of spatial context and local semantics is spatial features which used for detailed depiction of crop.
- The ConvLSTM based classifier is used to enhance the prediction of crop yield based on the spatial context and local semantics extracted from remote sensing images. The capacity of handling spatiotemporal dependencies of ConvLSTM is used to achieve an effective generalization during prediction.

The remaining paper is sorted as follows: The existing researches related to the crop yield estimation is given in Section II. The detailed information about RALCAC based feature extraction and ConvLSTM based prediction is provided in Section III. The outcomes of RALCAC are provided in Section IV, discussion is given in Section V and the paper is concluded in Section VI.

II. Related Work

The existing researches related to the crop yield estimation is given in this section.

Gavahi et al. [16] presented the DeepYield architecture for forecasting of crop yield, whereas the DeepYield was the combination of ConvLSTM and 3-Dimensional Convolutional Neural Networks (3DCNN). The intrinsic spatiotemporal patterns were considered in ConvLSTM to ensure the crop yield forecasting process. Further, the DeepYield was used to perform precise and robust crop yield forecasting, whereas the end-to-end learning was utilized for an automatic process of input. The local semantics were required to be highlighted for enhancing the feature extraction process.

Qiao et al. [17] developed Spatial-Spectral-Temporal Neural Network (SSTNN) to predict the crop yield which was the integration of 3D convolutional (Conv) and recurrent neural networks. The joint spatial-spectral-temporal representation was recognized by incorporating a spatial-spectral learning and temporal dependency, capturing modules in SSTNN. An effect of imbalanced dissemination of crop yield labels was eliminated by using a loss function. The crop yield prediction was high, when the SSTNN was processed with a huge amount of temporal information.

Fernandez-Beltran et al. [18] presented large-scale rice crop dataset (RicePAL) which has the multi-temporal S2 and climate/soil information from Terai districts of Nepal. The inherent data restraints were adapted 3DCNN for precise estimation of rice crop yield. The developed Convolutional Neural Networks (CNN) was developed for controlling the amount of layers while the fixing 3D Conv blocks were used to minimize the over-fitting. Nonetheless, an extra temporal dimension increased the amount of network parameters that made it possible for the 3DCNN to operate well only for larger patches.

Oikonomidis et al. [19] developed the hybrid deep learning approaches for predicting the crop yield. The developed models were XGBoost, XGBoost with scaling, integrated XGBoost with scaling and feature selection, hybrid CNN-XGBoost, CNN-Recurrent Neural Networks (RNN), CNN- Deep Neural Networks (DNN) and CNN-Long Short Term Memory (LSTM). The XGBoost was utilized as estimator to accomplish the feature selection. Here, the data dependencies and information were obtained by using the CNN. Next, the predictions were done by using the DNN as feed forward propagation approach. The developed XGBoost resulted in higher RMSE while performing the crop yield prediction.

Abbaszadeh et al. [20] presented a framework for combining the deterministic outputs from two DNN for creating the probabilistic simulation. The developed framework was Copula-Embedded Bayesian Model Averaging (COP-BMA) that combined the set of multivariate Copula operations into BMA. This COP-BMA reduced any consideration over the shape of conditional probability distribution function which used to offer precise and consistent predictive distributions. However, the contextual information was required for further improving the prediction.

Mohan, A et al. [21] developed the Temporal Convolutional network (TCN) with a customized dilated convolution unit for forecasting the rice crop yield. The correlation among the temporal and spatial parameters was analyzed using the TCN and it minimized the prediction error. The TCN's causal property and dilated convolution were resulted in the multivariate time-based evaluation and provided the enhanced prediction. The local features and changes within the crop were required to be extracted for further enhancing the prediction.

Qiao, M et al. [22] presented the knowledge-guided temporal multi-head attention approach that combined the prior information and scores of multi-head self-attention for combining the dynamical temporal correlation. Specifically, the prior attention distribution was introduced in self-attention learning based on the dynamic temporal graph convolution transformer. The features of spatially nearby places were aggregated based on geospatial relations for enhancing the capacity of prediction. The temporal dynamics of the features was required to be considered during the prediction for handling the dependencies between the data.

Boppudi, S. and Jayachandran, S [23] developed the hybrid mode according to the improved feature ranking fusion that fused the features from Relief, Recursive Feature Elimination (RFE) and Chi-Square method. The imbalanced data was handled by using an improved synthetic minority oversampling technique. Finally, the prediction was accomplished by integrating the LSTM with Deep Belief Network (DBN) classifiers. The selection of appropriate features was used to enhance the prediction by using the LSTM-DBN. However, the generalization with different datasets was required to be considered for an effective analysis.

Kolipaka, V.R.R. and Namburu, A [24] presented the deep learning-based system for predicting the agricultural production. This research considered a Two-stage classifiers where stage 1 performed pre-prediction and stage 2 performed the final classification for predicting the yield. The pre-prediction stage was incorporated the LSTM, Recurrent Neural Network (RNN) and LSTM for pre-prediction while the improved Convolutional Neural Network (CNN) was used in classification stage. In improved CNN, the Dingo Optimized Sand Piper (DOSP) was used for fine tuning the CNN to improve the prediction. The spatial features were required to be considered for further enhancing the prediction performances.

The limitations from the related works are specified as follows: inappropriate feature extraction, failed to obtain the generalization and ineffective in handling the spatiotemporal dependencies in prediction. In order to address these issues, the RALCAC is developed along with the ConvLSTM for an effective crop yield prediction. The encoders used in the RALCAC extracts the features of spatial context and local semantic information for effectively depicting the crop. Further, the capacity of handling the spatiotemporal dynamics of extracted features by ConvLSTM offers an effective prediction with generalization capacity.

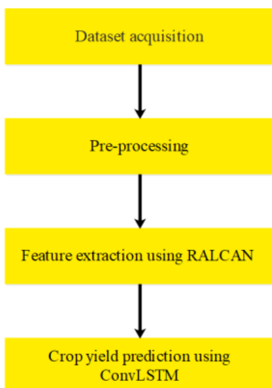

Fig. 1. Block diagram of proposed method.

III. PROPOSED METHOD

In this proposed method, the crop yield prediction using remote sensing images is achieved by using the RALCAC and ConvLSTM classifier. The main processes existing in the proposed method are dataset acquisition, data pre-processing, feature extraction using RALCAC and prediction using ConvLSTM. The residual attention unit existing in the RALCAC integrates residual linking and attention operation for retaining whole edge data, highlighting crucial semantics and improves the generalization capacity used to enhance the prediction. Fig. 1 shows the crop yield forecasting using RALCAC and ConvLSTM.

A. Dataset Acquisition

This research considers three different dataset such as MODIS dataset, MOD09A1 dataset and RicePAL dataset for evaluation.

1) MODIS dataset

a) Yield data: The measurements of soybean related to country are gathered from USDA NASS Quick Stat tool. For performing the model training, the yield information [25] between 2003 and 2019 is utilized as labels of ground truth.

b) MODIS surface reflectance: A surface spectral reflectance with seven bands is obtained by MODIS/Terra Surface Reflectance (SR) [26] which is acquired at spatial resolution of 500m for every eight days. A finest possible SR observation exists in each pixel, but this SR observation is chosen from all the observations of the 8-day window.

c) MODIS land cover: The MODIS Land Cover (LC) [27] type is combined by the Terra and Aqua which offers the yearly LC categories formulated from six recognition schemes. The cropland areas masking is done by annual University of Maryland (UMD).

d) MODIS land surface temperature (LST): A time surface temperature of average of 8 day per pixel, day and night is provided by MODIS of Version 6 LST. A 7 thermal infrared bands are employed by LST approach for collecting the temperature data.

2) MOD09A1 dataset: This MOD09A1 dataset [28] has seven spectral bands and a 500m spatial resolution is utilized for obtaining the required reflectance data. For wheat yield, this dataset includes a time series of 32 images obtained among October to July while time series of 20 images are obtained among May to October.

3) RicePAL dataset: The RicePAL dataset [18] has 3-year multi-temporal S2 imagery acquired from Terai area of Nepal along with its ground truth. Moreover, a climate and soil information are incorporated in the data for supporting the yield forecasting.

B. Pre-Processing

The datasets considered in this proposed method comprised of SR, MODIS LST and Land Use LC have 7, 2 and 1 band. The latter is utilized for masking the cropland zones through each county. The tiles are mosaiced into a single image which encloses the degree of the CONUS. A clipping is done for mosaiced raster through each country and the images for the chosen time intervals are combined by generating 3D tensors. The MODIS SR and LC has a spatial resolution of 500m which is dissimilar from the MODIS LST. Therefore, the 500m images are scaled up to 1km resolution by employing linear interpolation. Further, the 4D tensors with the measurement of $Time \times Height \times Width \times band$ is developed by concatenating each product's band to 3D tensors. The input image size is increased by including the rows and columns of zero in the zero padding process which is used to make the images in similar sizes before giving them to the RALCAC.

C. Feature Extraction using RALCAC

The RALCAC used in feature extraction utilizes the architecture of encoder and decoder for constructing the model. In that, the Residual Attention (RA) module is incorporated in encoders for obtaining the higher level semantic data from the pre-processed image, multi-scale spatial data is obtained by Multi-Scale Dilated Convolution (MSDV) and abstracted feature data is amplified by sing decoders that obtains the pixel-by-pixel semantic segmentation. The spatiotemporal features of pre-processed image are extracted by using RA and MSDV.

The developed RALCAC receives two different inputs such as pre-processed image and multi-feature information. The multi-feature information includes the features of color, texture and shape that made the complete utilization of rich feature information. Here, the color moments are chosen as color features and texture features are extracted by Gray Level Co-occurrence Matrix (GLCM), and the detection of edge according to contours are chosen as shape features. The information is extracted pixel by pixel while processing the color and texture features.

1) Architecture of encoder and decoder: The encoder and decoder extracts comprehensive feature data of the input image. The adjusted RsNet-50 is considered as dual encoder baseline architecture that has a multi-layer residual mapping block. This mapping blocks are additionally separated as two main modules such as identity blocks and conv blocks, and each Conv block has 3×3 Conv layer and two 1×1 Conv layers. Further, the identity block contains additional 1×1 Conv layer when compared to the Conv block over the shortcut that is employed to modify the channel's dimension. The architecture of decoder has Conv and up-sampling blocks. The feature map's spatial size is increased by up-sampling, while the local feature extraction is accomplished by Conv layer in the amplified feature map. The RALCAC receives two different inputs and the allocation of two symmetrical encoders with different weight values improve the feature extraction ability. The given input is processed over various conv and identity blocks followed by the feature maps being obtained layer by layer. Accordingly, the multiple dimensionality reduction causes losses in the spatial and spectral data of input. Hence, the underlying feature data with in-depth features are combined based on the skipping connections among encoder and the decoder. This skipping connections are used for an effective extraction of crop data in complex situations.

2) Residual attention unit: The attention methodology which replicates the human perception and obtains the features is developed. The developed RA uses various weight values for highlighting essential data while reducing unwanted data. Simultaneously, the RA solves the issues created by correlation among various feature channels, decrement in computational efficiency and the deficiency of abstraction and extraction for essential data in the network. The high and low weights are used in RA for highlighting the essential data and eliminating the unwanted data which in turn improves the generalization capacity and network's robustness for obtaining beneficial information in various situations.

The integration of RA and deep learning improves the deep learning performances. In feature mapping, the network frequently creates various residuals in encoder-decoder architecture. An amount of network layers deepens are maximized by using the residuals. The essential data is highlighted by using the various weights in RA which also offers a definite level of interpretability for the features of black box. Thus, the RA utilizes attention operation for highlighting the essential local data and residual links for integrating local context data, thereby obtaining the requirement of emphasising local contextual information. The developed RA has two portions such as, series Conv and shortcut, wherein the RA architecture is shown in Fig. 2. The convergence speed and generalization capacity are enhanced by using the Batch Normalization (BN) layer and ReLU after every Conv layer. In series Conv structure, an each Conv layer of Conv kernel is $\{2^{(i+5)}, 2^{(i+5)}, 2^{(i+6)}\}$, where RA module is denoted as i. Due to the difference in the amount of input and output channels, the architecture of shortcut includes 1×1 Conv, developed for varying the dimension of channel, while the amount of Conv kernels is $2^{(i+6)}$.

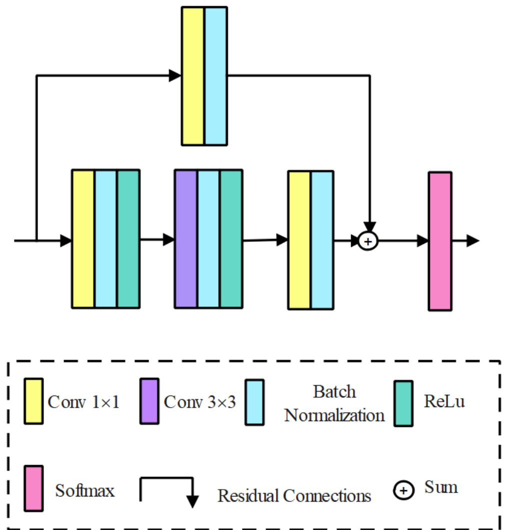

Fig. 2. Architecture of RA.

3) Process of MSDV: The MSDV unit is incorporated among the encoder and decoder in the overall model by using various dilation rates of dilated Conv 1×1 Conv layer for extracting the feature maps from multi-scale. The MSDV with 5 channels is shown in Fig. 3. A 1×1 Conv layer is used in the 1st channel for obtaining feature data and 3×3 Conv layer is incorporated in the 2nd channel. The dilated Conv with dilation rates of $\{1, 2, 3\}$ are appended from the 3rd to 5th channel for increasing the limit of the receptive field without maximizing the model's complexity. Eq. (1) shows the specific computation process of multi-scale dilated Conv.

$$x(l_0) = \sum_{i=1}^{N} m_i(l_0) \tag{1}$$

Where, MSDV of input feature map is denoted as l_0, and multi-scale dilated Conv for layer i is denoted as $m_i()$. Further, the outcomes of each layer is combined and multi scale feature data (x) is achieved from RALCAC.

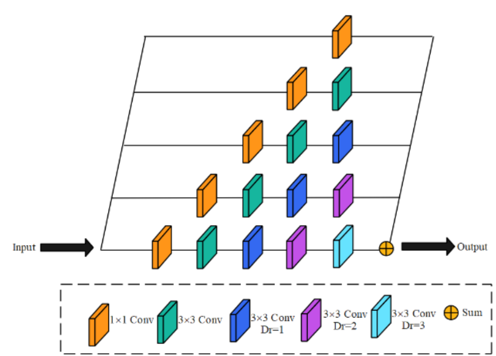

Fig. 3. Architecture of MSDV with five channels.

D. Prediction using ConvLSTM

In this phase, the ConvLSTM which is the integration of Conv filters and LSTM layers is developed for performing the crop prediction based on the features from RALCAC. Generally, the LSTM network has the capacity for maintaining the cell state from the preceding observation's sequence during the unwanted data elimination. The aforementioned principle is ensured by preserving the information over three gates such as input, output and forget gates. The Conv filters are employed to the input to state, and state to state changes of the LSTM. Fig. 4 shows the inner architecture of ConvLSTM. The architecture of ConvLSTM is described in Eq. (2) to Eq. (6).

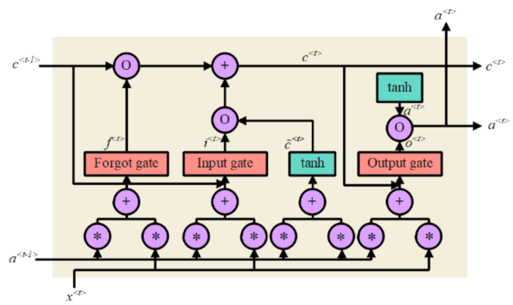

Fig. 4. Inner architecture of ConvLSTM.

$$i^{(t)} = \sigma\left(W_{xi}^{*} x^{(t)} + W_{ai}^{*} a^{(t-1)} + W_{ci}^{\circ} c^{(t-1)} + b_i\right) \quad (2)$$

$$f^{(t)} = \sigma\left(W_{xf}^{*} x^{(t)} + W_{af}^{*} a^{(t-1)} + W_{cf}^{\circ} c^{(t-1)} + b_f\right) \quad (3)$$

$$c^{(t)} = f^{(t)\circ} c^{(t-1)} + i^{(t)\circ} \tanh\left(W_{xc}^{*} x^{(t)} + W_{ac}^{*} a^{(t-1)} + b_c\right) \quad (4)$$

$$o^{(t)} = \sigma\left(W_{xo}^{*} x^{(t)} + W_{ao}^{*} a^{(t-1)} + W_{co}^{\circ} c^{(t-1)} + b_o\right) \quad (5)$$

$$a^{(t)} = o^{(t)\circ} \tanh\left(c^{(t)}\right) \quad (6)$$

Where, $i^{(t)}, f^{(t)}$ and $o^{(t)}$ are the variables returned by input, forget and output gate, respectively, cell output is denoted as $a^{(t)}$, weight matrices are denoted as W, elementwise product is denoted as $(\circ)$, Conv operator is denoted as $(*)$ and sigmoid activation function is denoted as σ.

ConvLSTM is generally used to acquire the intrinsic spatiotemporal patterns of given data. For each required output, eight filters are needed in the architecture of ConvLSTM. The incorporation of Conv filters in LSTM minimizes the model parameters, than the single LSTM which is used to achieve training even deeper that helps to achieve better prediction.

IV. Results and Discussion

The results and discussion of the proposed method are given in this section. The proposed method is analysed by using Python 3.6 software. Here, the Tensorflow 1.14 and Keras library are used for execution of the crop yield prediction. The system is configured with 1 TB memory 128 GB RAM, Windows 10 operating system, 22 GB RAM for RTX 2080 Ti GPU, and i9 processor. The performance measures analysed in this research are RMSE and coefficient of determination (R^2) which are expressed in Eq. (7) and Eq. (8).

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(M_i - O_i)^2}{N}} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(M_i - \bar{O})^2} \quad (8)$$

Where, the model forecast and observed yield value are respectively denoted as M_i and O_i, their respective mean values are represented as $\bar{M}$ and $\bar{O}$, and the amount of predicting data points is denoted as N.

A. Performance Analysis

The primary objective of this research is evaluated using MODIS dataset for soybean forecasting. Further, the proposed method is evaluated in three different datasets such as MOD09A1 dataset for wheat corn yield prediction and RicePAL dataset. The RALCAC is assessed for different features and different classifiers. The different features are color, shape and texture, while the different classifiers are Random Forest (RF), Recurrent Neural Network (RNN) and LSTM.

1) Evaluation of proposed method for MODIS dataset: The MODIS dataset is evaluated for different selection and classifiers as shown in the Table I and II, respectively. Further, the graphs for different features and classifiers are shown in the Fig. 5 and Fig. 6. From the analysis, it is determined that the RALCAC provides better performance than the individual color, texture and shape features. Therefore, the RALCAC uses the multi feature information and pre-processed image for extracting the beneficial data from the images, which further enhance the prediction. On the other hand, the ConvLSTM provides better classification than the RF, RNN and LSTM. The observation of intrinsic spatiotemporal patterns in ConvLSTM is enhances the prediction.

TABLE I. Proposed Method Evaluation with MODIS Dataset for Different Features

Features	*RMSE*	R^2
Color	5.942	0.82
Texture	4.097	0.91
Shape	7.005	0.87
RALCAC	3.257	0.94

TABLE II. Proposed Method Evaluation with MODIS Dataset for different Classifiers

Classifiers	*RMSE*	R^2
RF	4.982	0.82
RNN	5.743	0.77
LSTM	4.226	0.89
ConvLSTM	3.257	0.94

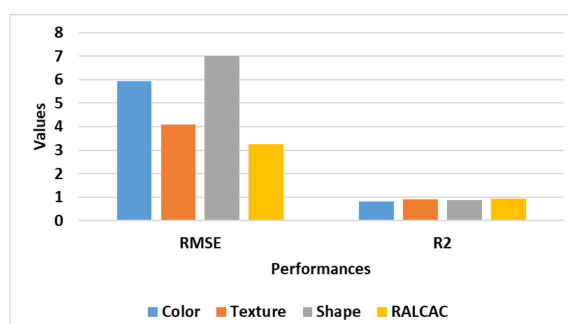

Fig. 5. Proposed method graph of MODIS dataset for different features.

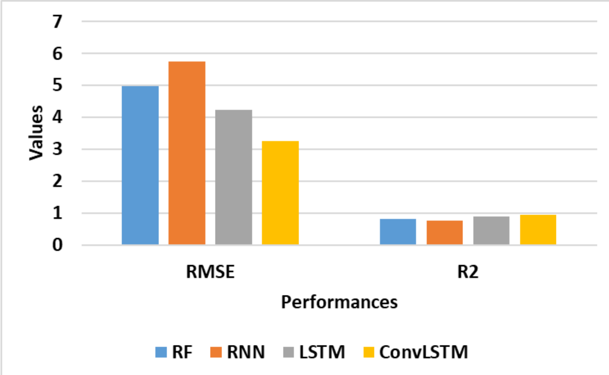

Fig. 6. Proposed method graph of MODIS dataset for different classifiers.

2) Evaluation of proposed method for MOD09A1 dataset: In this section, the time series of 32 images obtained between October to July of the next year in MOD09A1 dataset are used for wheat yield prediction. Table III and IV show the proposed method's evaluation of MOD09A1 dataset for different features and different classifiers. Further, the graph of the proposed method with MOD09A1 dataset for different features and different classifiers is shown in Fig. 7 and Fig. 8. From the analysis, it is found that the RALCAC obtains better performance than the individual features. Moreover, the ConVLSTM provides better performance than the RF, RNN and LSTM. The RALCAC achieves superior prediction because it highlights the required semantics and improves the generalization capacity. Also, the combination of Conv filters and LSTM block in ConvLSTM enhances the prediction.

TABLE III. PROPOSED METHOD EVALUATION WITH MOD09A1 DATASET FOR DIFFERENT FEATURES

Features	*RMSE*	R^2
Color	0.77	0.83
Texture	0.71	0.88
Shape	1.23	0.77
RALCAC	0.63	0.93

TABLE IV. PROPOSED METHOD EVALUATION WITH MOD09A1 DATASET FOR DIFFERENT CLASSIFIERS

Classifiers	*RMSE*	R^2
RF	0.91	0.79
RNN	0.98	0.73
LSTM	0.72	0.81
ConvLSTM	0.63	0.93

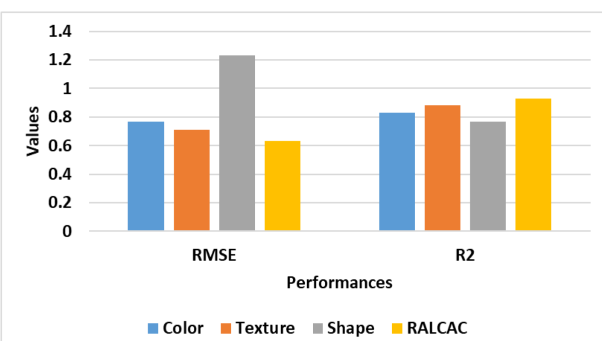

Fig. 7. Proposed method graph of MOD09A1 dataset for different features.

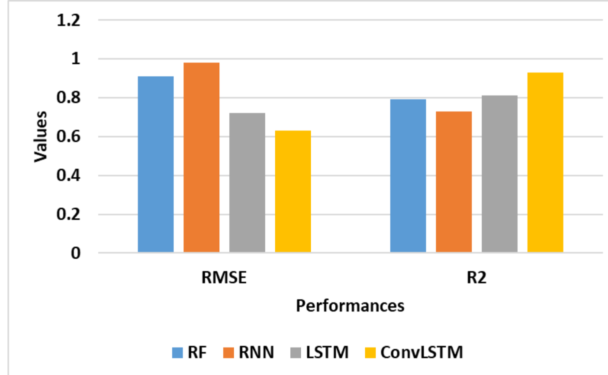

Fig. 8. Proposed method graph of MOD09A1 dataset for different classifiers.

3) Evaluation of proposed method for RicePAL dataset: The RicePAL dataset is evaluated for different selection and classifiers as shown in Tables V and VI, respectively. Further, the graph of the proposed method with RicePAL dataset for different features and classifiers is shown in Fig. 9 and Fig. 10, correspondingly. From the analysis, it is determined that the RALCAC provides better performance than the individual color, texture and shape features. Moreover, the ConvLSTM provides better prediction than the RF, RNN and LSTM.

TABLE V. PROPOSED METHOD EVALUATION WITH RICEPAL DATASET FOR DIFFERENT FEATURES

Features	*RMSE*	R^2
Color	3.986	0.88
Texture	3.227	0.91
Shape	5.025	0.84
RALCAC	2.069	0.94

TABLE VI. PROPOSED METHOD EVALUATION WITH RICEPAL DATASET FOR DIFFERENT CLASSIFIERS

Classifiers	*RMSE*	R^2
RF	3.217	0.84
RNN	4.561	0.79
LSTM	2.844	0.92
ConvLSTM	2.069	0.94

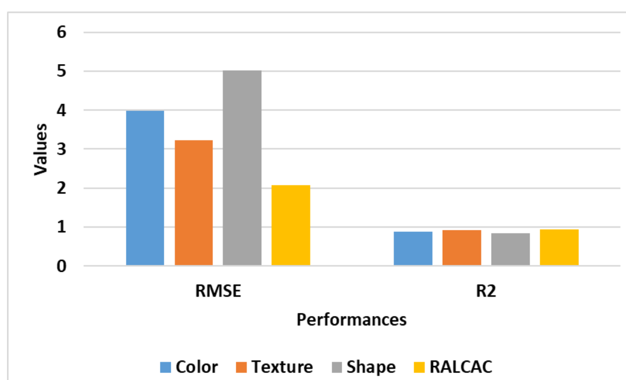

Fig. 9. Proposed method graph of RicePAL dataset for different features.

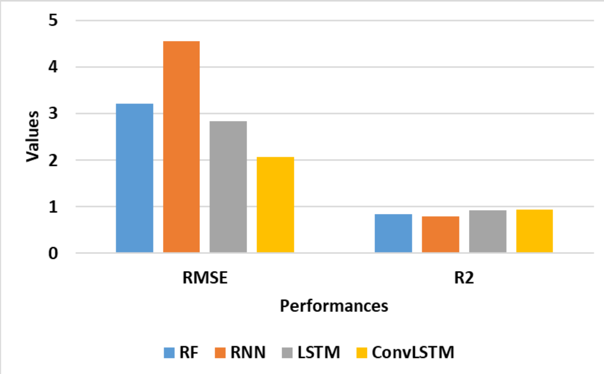

Fig. 10. Proposed method graph of RicePAL dataset for different classifiers.

B. Comparative Analysis

This section shows the comparative analysis of the RALCAC based crop yield prediction. The existing research such as DeepYield [16], SSTNN [17] and 3DCNN [18] are used to compare the RALCAC. Here, the comparison is done for three different datasets such as MODIS, MOD09A1 and RicePAL. Table VII shows the comparative analysis of RALCAC while the graph for MODIS dataset is shown in Fig. 11. From the comparison, it is concluded that the RALCAC provides lesser RMSE than the DeepYield [16], SSTNN [17] and 3DCNN [18] methods. The integrated spatial context information along with the highlighting of local semantics in RALCAC improves the feature extraction which helps to achieve better prediction.

TABLE VII. COMPARATIVE ANALYSIS OF RALCAC

Datasets	Methods	RMSE
MODIS dataset	DeepYield [16]	4.79
	RALCAC	3.257
MOD09A1 dataset	SSTNN [17]	0.67
	RALCAC	0.63
RicePAL dataset	3DCNN [18]	89.03
	RALCAC	2.069

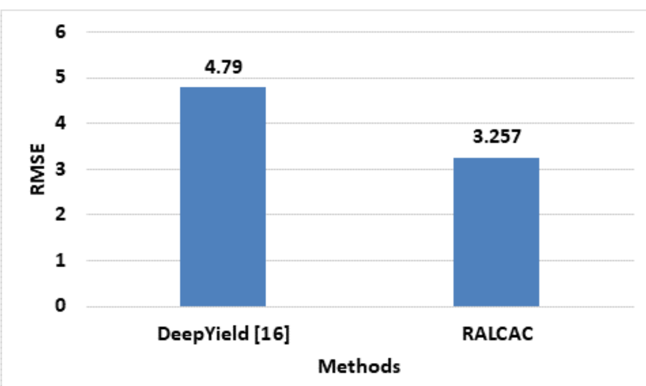

Fig. 11. Comparison graph for MODIS dataset.

V. DISCUSSION

This section provides the discussion about the crop yield prediction performed by the RALCAC and ConvLSTM. The different datasets used for evaluation are MODIS dataset, MOD09A1 dataset and RicePAL dataset. The RALCAC is evaluated with different feature extraction approaches such as Color, Texture and Shape while the ConvLSTM is evaluated with different classifiers such as RF, RNN and LSTM. The evaluation demonstrates that the RALCAC and ConvLSTM has better performance than the aforementioned state of art approaches. Moreover, this RALCAC outperforms well than the DeepYield [16], SSTNN [17] and 3DCNN [18]. The main reason of improved prediction is RALCAC based comprehensive feature extraction and handling of spatiotemporal dynamics using ConvLSTM. The developed RALCAC extracts the spatial context features using RA module and local semantic information during the extraction. Therefore, the RALCAC represents the extensive spatial features and relationships, and local features and changes in the crop fields which effectively depicts the crop. Additionally, the capacity of spatiotemporal handling using ConvLSTM is used for an effective prediction with generalization capacity.

VI. CONCLUSION

In recent times, the evolution of remote sensing offers huge accessibility for performing precise crop yield prediction. In this research, RALCAC based comprehensive feature extraction is developed along with a ConvLSTM classifier. An effective depiction of crop is obtained by extracting the spatial context and local semantic features using the RALCAC which denotes spatial features and its relationships, and local features and changes in the crop fields. Further, the ConvLSTM performs a prediction based on the spatial and local semantic features from the RALCAC. The capacity of handling the spatiotemporal dependencies using ConvLSTM helps to enhance the prediction with effective generalization. From the simulation, it is found that the RALCAC outperforms the DeepYield, SSTNN and 3DCNN. The RMSE of RALCAC for MODIS dataset is 3.257, which is lesser when compared to the DeepYield. In future, a feature selection can be developed for further improving the prediction performances.

REFERENCES

[1] D. Elavarasan and P. M. D. R. Vincent, “A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters,” J. Ambient Intell. Hum. Comput., vol. 12, no. 11, pp. 10009–10022, November 2021.

[2] D. Elavarasan and P. M. Durai Raj Vincent, “Fuzzy deep learning-based crop yield prediction model for sustainable agronomical frameworks,” Neural Comput. Appl., vol. 33, no. 20, pp. 13205–13224, October 2021.

[3] A. Sharifi, “Yield prediction with machine learning algorithms and satellite images,” J. Sci. Food Agric., vol. 101, no. 3, pp. 891-896, February 2021.

[4] N. R. Prasad, N. R. Patel, and A. Danodia, “Crop yield prediction in cotton for regional level using random forest approach,” Spatial Inf. Res., vol. 29, no. 2, pp. 195-206, April 2021.

[5] D. Elavarasan and P. M. D. R. Vincent, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications,” IEEE Access, vol. 8, pp. 86886-86901, May 2020.

[6] Z. Ji, Y. Pan, X. Zhu, J. Wang, and Q. Li, “Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index,” Sensors, vol. 21, p. 1406, February 2021.

[7] P. Hara, M. Piekutowska, and G. Niedbała, “Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data,” Land, vol. 10, p. 609, June 2021.

[8] X. Zhu, R. Guo, T. Liu, and K. Xu, “Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data,” Remote Sens., vol. 13, p. 2016, May 2021.

[9] W. Xu, P. Chen, Y. Zhan, S. Chen, L. Zhang, and Y. Lan, "Cotton yield estimation model based on machine learning using time series UAV remote sensing data," Int. J. Appl. Earth Obs. Geoinf., vol. 104, p. 102511, December 2021.

[10] S. J. J. Jui, A. A. M. Ahmed, A. Bose, N. Raj, E. Sharma, J. Soar, and M. W. I. Chowdhury, "Spatiotemporal Hybrid Random Forest Model for Tea Yield Prediction Using Satellite-Derived Variables," Remote Sens., vol. 14, p. 805, February 2022.

[11] H. Tian, P. Wang, K. Tansey, J. Zhang, S. Zhang, and H. Li, "An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China," Agric. For. Meteorol., vol. 310, p. 108629, November 2021.

[12] A. Tripathi, R. K. Tiwari, and S. P. Tiwari, "A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation," Int. J. Appl. Earth Obs. Geoinf., vol. 113, p. 102959, September 2022.

[13] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad, and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," Agric. For. Meteorol., vol. 284, p. 107886, April 2020.

[14] V. Sagan, M. Maimaitijiang, S. Bhadra, M. Maimaitiyiming, D. R. Brown, P. Sidike, and F. B. Fritschi, "Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning," ISPRS J. Photogramm. Remote Sens., vol. 174, pp. 265-281, April 2021.

[15] M. Marshall, M. Belgiu, M. Boschetti, M. Pepe, A. Stein, and A. Nelson, "Field-level crop yield estimation with PRISMA and Sentinel-2," ISPRS J. Photogramm. Remote Sens., vol. 187, pp. 191-210, May 2022.

[16] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting," Expert Syst. Appl., vol. 184, p. 115511, December 2021.

[17] M. Qiao, X. He, X. Cheng, P. Li, H. Luo, L. Zhang, and Z. Tian, "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks," Int. J. Appl. Earth Obs. Geoinf., vol. 102, p. 102436, October 2021.

[18] R. Fernandez-Beltran, T. Baidar, J. Kang, and F. Pla, "Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal," Remote Sens., vol. 13, p. 1391, April 2021.

[19] A. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid deep learning-based models for crop yield prediction," Applied Artificial Intelligence, vol. 36, no. 1, p. 2031822, January 2022.

[20] P. Abbaszadeh, K. Gavahi, A. Alipour, P. Deb, and H. Moradkhani, "Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction," Agric. For. Meteorol., vol. 314, p. 108773, March 2022.

[21] Mohan, A., Venkatesan, M., Prabhavathy, P. and Jayakrishnan, A., 2023. Temporal convolutional network based rice crop yield prediction using multispectral satellite data. Infrared Physics & Technology, 135, p.104960.

[22] Qiao, M., He, X., Cheng, X., Li, P., Zhao, Q., Zhao, C. and Tian, Z., 2023. KSTAGE: A knowledge-guided spatial-temporal attention graph learning network for crop yield prediction. Information Sciences, 619, pp.19-37.

[23] Boppudi, S. and Jayachandran, S., 2024. Improved feature ranking fusion process with Hybrid model for crop yield prediction. Biomedical Signal Processing and Control, 93, p.106121.

[24] Kolipaka, V.R.R. and Namburu, A., 2024. An automatic crop yield prediction framework designed with two-stage classifiers: a meta-heuristic approach. Multimedia Tools and Applications, 83(10), pp.28969-28992.

[25] Yield data: https://www.nass.usda.gov/Quick_Stats/index.php.

[26] MODIS Surface reflectance data: https://lpdaac.usgs.gov/products/mod09a1v006/.

[27] MODIS Land cover data: https://lpdaac.usgs.gov/products/mcd12q1v006/.

[28] E. Vermote, Mod09a1 modis/terra surface reflectance 8-day l3 global 500m singrid v006. NASA EOSDIS Land Processes DAAC, vol. 10, 2015.