

Synergistic Use of TanDEM-X and Landsat-8 Data for Crop-Type Classification and Monitoring

Subhadip Dey ¹, Graduate Student Member, IEEE, Ushasi Chaudhuri ², Graduate Student Member, IEEE, Narayanarao Bhogapurapu ³, Graduate Student Member, IEEE, Juan M. Lopez-Sanchez ⁴, Senior Member, IEEE, Biplab Banerjee ⁵, Senior Member, IEEE, Avik Bhattacharya ⁶, Senior Member, IEEE, Dipankar Mandal ⁷, Member, IEEE, and Yalamanchili Subrahmanyeswara Rao, Member, IEEE

Abstract—Classification of crop types using Earth Observation (EO) data is a challenging task. The challenge increases many folds when we have diverse crops within a resolution cell. In this regard, optical and Synthetic Aperture Radar (SAR) data provide complementary information to characterize a target. Therefore, we propose to leverage the synergy between multispectral and Synthetic Aperture Radar (SAR) data for crop classification. We aim to use the newly developed model-free three-component scattering power components to quantify changes in scattering mechanisms at different phenological stages. By incorporating interferometric coherence information, we consider the morphological characteristics of the crops that are not available with only polarimetric information. We also utilize the reflectance values from Landsat-8 spectral bands as complementary biochemical information of crops. The classification accuracy is enhanced by using these two pieces of information combined using a neural network-based architecture with an attention mechanism. We utilize the time series dual co-polarimetric (i.e., HH-VV) TanDEM-X SAR data and the multispectral Landsat-8 data acquired over an agricultural area in Seville, Spain. The use of the proposed attention mechanism for fusing SAR and optical data shows a significant improvement in classification accuracy by 6.0% to 9.0% as compared to the sole use of either the optical or SAR data. Besides, we also demonstrate that the utilization of single-pass interferometric coherence maps in the fusion framework enhances the overall classification accuracy by $\approx 3.0\%$. Therefore, the proposed synergistic approach will facilitate accurate and robust crop mapping with high-resolution EO data at larger scales.

Index Terms—Agriculture, classification, crop-type mapping, Landsat-8, phenology, TanDEM-X.

I. INTRODUCTION

THE rapid evolution of multiple sensors allows us to combine information from different images providing more meaningful and necessary information. Images captured by various sensors offer specific details. Therefore, the integration of information is more valuable than individual information obtained from a single sensor. This fused information is critical for planning and decision-making. Various earth-orbiting satellites capture data in diverse parts of the electromagnetic spectrum. Several remote sensing data products are available for Earth observation (EO), viz., multispectral (MS), hyperspectral (HS), and synthetic aperture radar (SAR). These data products provide specific information about the Earth's surface under observation.

On the one hand, SAR imagery contains target geometrical and dielectric information while capturing images in all weather conditions. However, it fails to provide geochemical information of targets. On the other hand, optical images provide biochemical properties of targets perceived in terms of spectral signature. Thus, SAR and optical imagery offer complementary information about targets, and hence a combination of these images would benefit enhance spatial and spectral information [1], [2].

In this context, the synergy between SAR and optical images to enhance crop classification accuracy is presented by [3]. Their study used single-channel SAR data (HH) and Landsat thematic mapper (TM) bands to classify six different crop types in the Saskatchewan region. They reported that the sole utilization of HH polarized SAR data produced 31% to 45% accuracy using the maximum likelihood classifier. In contrast, the combination of SAR and TM data produced 77% classification accuracy. Similarly, Sandholt [4] utilized the SMAP algorithm to combine EMISAR and SPOT images to classify six different crop types. The reported a classification accuracy of 95% which was due to the high information content in the multipolarized SAR data. In another study, Qi *et al.* [5] utilized Landsat and ERS data to extract both soil and plant information. Thus, optical and SAR data synergy has proven to be helpful in mapping and monitoring crops.

Later, Blaes *et al.* [6] proposed a hierarchical parcel-based classification strategy to classify many crop types using SAR

Manuscript received May 9, 2021; revised June 24, 2021 and July 20, 2021; accepted August 3, 2021. Date of publication August 10, 2021; date of current version September 15, 2021. This work was supported in part by the the German Aerospace Center (DLR) which provided all the TanDEM-X data under project POLI6736, in part by the State Research Agency (AEI), in part by the Spanish Ministry of Science and Innovation, and in part by the EU EFDR funds under Project TEC2017-85244-C2-1-P. The work of N. Bhogapurapu and S. Dey was supported by the Ministry of Education (formerly Ministry of Human Resource and Development-MHRD), Government of India. (Corresponding author: Avik Bhattacharya.)

Subhadip Dey, Ushasi Chaudhuri, Narayanarao Bhogapurapu, Biplab Banerjee, Avik Bhattacharya, Dipankar Mandal, and Yalamanchili Subrahmanyeswara Rao are with the Microwave Remote Sensing Lab, Center of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, Maharashtra 400076, India (e-mail: subhadipdey23071994@gmail.com; ushasi2cool@gmail.com; narayanarao.bhogapurapu@gmail.com; getbiplab@gmail.com; avikb@csre.iitb.ac.in; dipankar.agrileng@gmail.com; ysrao@csre.iitb.ac.in).

Juan M. Lopez-Sanchez is with the Institute for Computer Research (IUII), University of Alicante, E-03080 Alicante, Spain (e-mail: juanma-lopez@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2021.3103911

and optical data. They implemented this hierarchical classification strategy to consider the variability of the spectral signatures within each crop type. Their study used fifteen ERS and RADARSAT and three optical images to discriminate agricultural crop types. This particular combination increased the classification accuracy by $\approx 5\%$. McNairn *et al.* [7] produced an operational crop inventory map using the Envisat ASAR and optical images. They showed significant classification accuracy over the diverse agricultural landscape in Canada. Within the overall classification accuracy of 85%, they noted that the SAR-optical combination was able to identify major crop types.

In another study, Torbick *et al.* [8] utilized a decision tree framework to combine multitemporal and multiscale PALSAR, MODIS, and Landsat images for mapping the extent of rice, hydroperiod, crop calendar, and cropping intensity. This multi-season crop data produced a classification accuracy of 89% over the operational rice extent. In parallel, Hong *et al.* [9] attempted to find the optimal approach to fuse SAR and optical data using the wavelet-IHS (intensity, hue, and saturation) technique. A multiresolution image segmentation technique followed this. The overall classification of the fused product was 72%, which was higher than the sole use of SacnSAR and MODIS images.

Alongside this, Seoran and Haack [10] utilized the L-band ALOS PALSAR and TM data to classify four different crops. The stack-based fusion method produced an overall accuracy of 94.1%. Skakun *et al.* [11] employed a multilayer perceptron technique to combine multitemporal and multipolarized RADARSAT-2 and Landsat-8 data. They observed that the combination of SAR and optical data improved the overall classification accuracy by 0.09% and 1.94%, as compared to the use of SAR and optical data individually. They also reported that the use of multitemporal dataset helped gain $\approx 25\%$ accuracy. Furthermore, the VV–VH combination performed better for winter wheat and spring barley, while HH–HV for sugarbeet and soybeans. Several other studies also pointed out the advantages in the synergetic use of optical and SAR data in the domain of agricultural crop mapping [12]–[14].

The time-series of dual-pol SAR images acquired by TerraSAR-X and TanDEM-X have been successfully applied for crop classification by exploiting either the backscattering coefficient at the two copolarization channels [15]–[17] or extracted sets of polarimetric features [18]–[20] as inputs to the classifier. However, the best reported overall accuracy is $\approx 70\text{--}80\%$ when only backscattering coefficient features are utilized. However, polarimetric scattering power components were shown to be more effective in capturing the changes in the scattering mechanism with the advancement of crop phenological stages.

Hence, this study utilizes the newly developed three-component model-free scattering power component for dual copolarimetric (HH–VV) SAR data. The advantage of this model-free decomposition technique is its three roll-invariant scattering power components. Moreover, we resolved several limitations of state-of-the-art model-based decomposition techniques using this novel method. For example, the frequently observed overestimation of the volume scattering power component and the occurrence of negative power components in several model-based decompositions are fixed in this model-free

technique. Moreover, the method is adaptive for any scattering scenarios with high stability in the power components. These scattering power components were shown to be effective in monitoring and mapping different morphological characteristics of crops [21]. This work also explored the added value of the interferometric products derived over the agricultural crops using TanDEM-X data along with the scattering power components. We utilized the single-pass interferometry parameters that combine two simultaneous images acquired over the same scene to evaluate the performance of the interferometric products. Besides, we also use the Landsat-8 reflectance features to capture the biochemical changes inside the crop canopy. One of the most critical approaches to effectively combining this information is calculating a weighted map incorporating information from different source images. These weighted combinations are estimated by specific predefined or hand-crafted techniques devised by the users in the existing literature. These methods are often based on a nonlearning paradigm, and therefore, the estimates are nonadaptive for each resolution cell [22]. Hence, we propose a neural network-based adaptive learnable framework for calculating the weight map. The neural network architecture works as a self-adaptive global function estimator. Also, we use a bilevel attention module for self and cross-attention. The attention mechanism enables the overall framework to focus more on specific essential elements of the input feature space. Effectively, it guides the network to find out the crucial constructs. Moreover, the network architecture along with the attention mechanism approximates complex functions more intuitively [23].

In the scope of this work, we are more interested in exploring the causes for enhanced classification accuracies using both SAR and optical data and their combinations. This study typically does not emphasize the performance comparison of different classification or fusion techniques available in the literature. We have utilized network-based architectures due to their self-adaptive capabilities based on the nature of the data.

- 1) We utilize the scattering power components derived from the novel model-free decomposition technique for dual copolarimetric (i.e., HH–VV) TanDEM-X SAR data [21], [24]. Together, we also use the reflectance of distinct Landsat-8 multispectral bands to analyze crop phenology trends.
- 2) We also demonstrate the contribution of interferometric coherence, in addition to the polarimetric and multispectral features, for the enhancement of classification accuracy. Hence, we utilize all possible information about a target in three dimensions.
- 3) We develop a pixel-wise self and cross attention-based network architecture to effectively fuse SAR and optical time-series data to classify diverse crop types. In this regard, our primary contribution lies in utilizing the attention-based architecture for the pixel-level fusion of polarimetric SAR, interferometric SAR, and optical data to capture the overall characteristics of the agricultural crops throughout the phenological stages.

This study uses the TanDEM-X (HH|VV) SAR data and Landsat-8 multispectral data to classify diverse crop types over Seville, Spain.

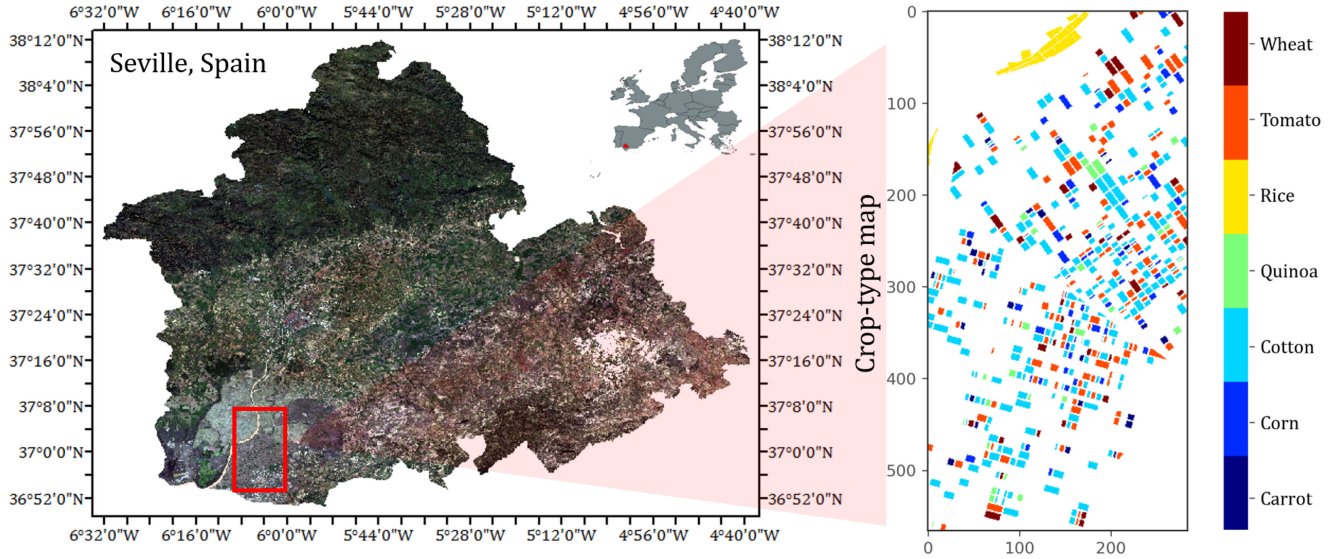


Fig. 1. Landsat-8 image of the Seville test site over Spain obtained from the Google Earth engine platform. We took the samples used for crop-type analysis and classification from the region marked in a red rectangle. The crop types are shown in seven different colors.

The rest of this article is organized as follows: Section II details the study area and dataset. Section III describes methodology. In, Section IV we discuss the results with clear explanations. Finally, Section V concludes this article.

II. STUDY AREA AND DATASET

The study area shown in Fig. 1 is located near Seville in Spain ($37^{\circ}02'29.45''$ N $6^{\circ}02'57.24''$ W). This area has a Mediterranean climatic zone, with varying soil types of cambisol, cromatic luvisol, and rendsic leptosol [25]. The areal coverage of this test site is $\approx 16 \times 8$ km².

The crops cultivated in this area mainly consist of: carrots, corn, cotton, quinoa, rice, tomato, and wheat. The cultivation takes place all around the year. The average size of each field is $\approx 300 \times 300$ m². Information about the crop types was recorded during field campaigns carried out from May to August 2015 by a local institution (FERAGUA). Crop types were assigned, and field borders were confirmed or fixed using a GPS to position the official land-parcel identification system as the initial database. The Landsat-8 acquisition dates are 27-May, 26-Jun, 14-Jul, 30-Jul, and 15-Aug, while the Tandem-X acquisition dates are: 30-May, 02-Jul, 13-Jul, 04-Aug, 15-Aug.

We acquired all TD-X images in descending passes with an incidence angle around 39° with a height of ambiguity of ≈ 5.8 m. All images correspond to the dual-pol mode with the two copolar channels: HH and VV. The original spatial resolution of these images was 6.6 m in azimuth and 3.1 m in the ground range. The pixel spacing (pixel size) was 2.4 m in both coordinates. We radiometrically calibrated these images and applied a boxcar filter to estimate the backscattering coefficients. Finally, we geocoded all products to a grid of 5 m posting. Similarly, we atmospherically corrected all L-8 images and calibrated them to generate surface reflectance values. Later, we resampled the pixels of L-8 to 5 by 5 m resolution and coregistered with the TD-X data.

Fig. 2 shows the crop calendar for wheat, tomato, rice, quinoa, cotton, corn, and carrot. This calendar provides timely information about phenology to promote local crop production. The figure shows the acquisition window of both SAR and optical data with red dotted vertical lines. During this period, one can see that most of the crops are in the growing and harvesting stages. Some of the crop fields were at the postharvest stage. Therefore, specific transitions among the phenological stages for different crop types are evident within the acquisition period. The crop calendar aids the analysis and interpretation of changes in SAR and optical observables corresponding to crop morphological conditions at different phenological stages. Similarly, we also infer the impact of crop phytomorphology in the classification accuracies using this time-series information.

III. METHODOLOGY

In this section, we describe the proposed classification framework. Then, we detail the Landsat-8 and TanDEM-X feature sets that we utilize for fusion and classification purposes. Finally, we describe the network architecture with different modules and cost functions.

A. Architecture

The objective of this work is to perform pixel-based classification of diverse crop-types using TanDEM-X (TD-X) and Landsat-8 (L-8) datasets. For this task, we have considered input dataset, $\mathcal{X} = \{\mathbf{x}_T^i, \mathbf{x}_L^i\}_{i=1}^n$ which are centered around ground-truth pixels, $\mathcal{Y} = \{\mathbf{y}^i\}_{i=1}^n$. Here, $\mathbf{x}_T^i \in \mathbb{R}^{1 \times F_1}$ and $\mathbf{x}_L^i \in \mathbb{R}^{1 \times F_2}$. F_1 and F_2 denote the number of features for TD-X and L-8, respectively, and n denotes the number of sample points. The ground-truth labels are denoted as $\mathbf{y} \in \{1, 2, \dots, 7\}$. These datasets are utilized in the proposed fusion architecture to discriminate various crop types through different modules within the network.

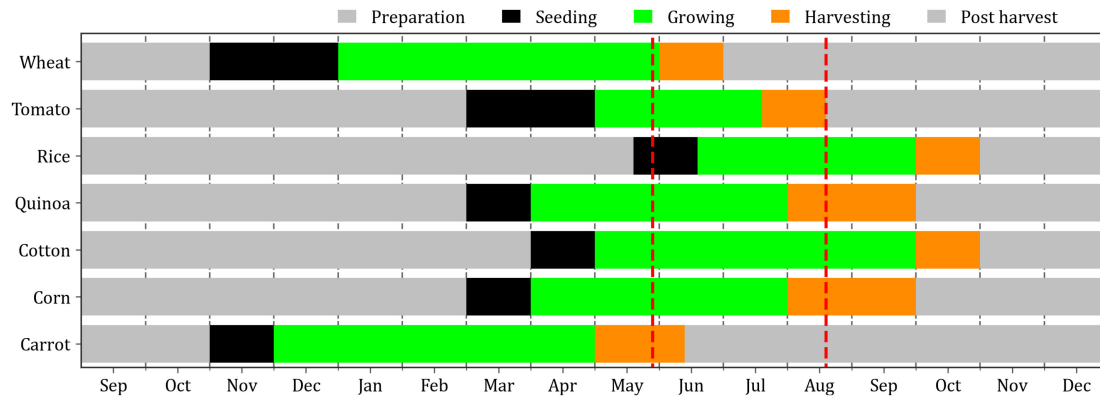


Fig. 2. Crop calendar of carrot, corn, cotton, quinoa, rice, tomato, and wheat crops over Seville, Spain. Different colors indicate the preparation, seeding, growing, harvesting, and post harvest stages. The red dotted lines represent the month of acquisitions of TanDEM-X and Landsat-8 data which are used for crop-type classification.

Feature Set

The feature sets F_1 consists of reflectance values from seven bands, and F_2 consists of polarimetric scattering power components along with interferometric coherence information.

B. Landsat-8 (L-8)

The L-8 band reflectance features (F_1) consists of 7 bands with 30-m resolution. However, as these bands were resampled, the final pixel resolution is 5 m. The wavelength of band-1 is 0.43 to 0.45 μm . This band gives information about the sediments, particles, and organic matter within a resolution cell. Band-2 have a wavelength range of 0.45 to 0.51 μm . This band might help discriminate between dry and moist soil conditions. Band-3 has a wavelength range of 0.53 to 0.59 μm and is useful in identifying greenness within a resolution cell. Band-4 of wavelength 0.64 to 0.67 μm has absorption characteristics depending on the chlorophyll content and health of vegetation. Band-5 has a wavelength range of 0.85 to 0.88 μm . The characteristics of this band directly depend on the chlorophyll content and the spongy mesophyll cells. It is also helpful in distinguishing vegetated surfaces from bare ground. Band-6 and Band-7 of wavelength ranges 1.57 to 1.65 μm and 2.11 to 2.29 μm , respectively, provide information on soil moisture and leaf water content [26]. Therefore, these bands have particular significance in inferring information about crop fields and could be better suited for classification.

C. TanDEM-X (TD-X)

We extract a set of seven features (F_2) from the TD-X data. We obtain the first three features from the model-free three-component scattering power decomposition technique [21], [24], and the following four features as the interferometric coherence information.

Target decomposition using SAR data provides scattering information from a target. The state-of-the-art methods use model-based decomposition techniques to obtain target information. However, the hierarchical process and the branching conditions lead to several stringent limitations. Moreover, the assumptions of *ad hoc* scattering models within a radar resolution cell make the computation of scattering power ambiguous. Common

concerns of these model-based techniques are associated with the overestimation of the volume scattering power, the nonroll invariant scattering power components, and the occurrence of negative scattering power components, and instability.

In addition to this, croplands are usually considered homogeneous, and hence, the scattering from these targets is almost symmetric. Therefore, the existence of the helix power component is negligible. Like the four-component Yamaguchi decomposition, the fourth component in our proposed model-free four-component scattering power decomposition [27] is the insignificant helix power component. Due to this, we have restricted ourselves to the model-free three-component scattering power decomposition technique.

The scattering power components consist of even bounce, odd bounce, and diffused powers [21]. We used the elements of the Kennaugh matrix (k_{11} and k_{44}) and the $n\text{D}$ Barakat degree of polarization [28] to compute the scattering power components. Moreover, these scattering power components are unique and unambiguous and adaptive to the morphological changes of the crops. Also, within the decomposition framework, the target characterization parameters and the scattering power components are roll-invariant.

In addition to this, we have used a set of SAR observables composed of the single-pass interferometric coherence at HH and VV copolar channels and HH + VV (P1), HH - VV (P2) Pauli channels. As the master and slave images of the slant range products were already coregistered, we followed only four steps in the processing chain: 1) Subset of the region of interest, 2) removal of flat Earth and topographic phase components, 3) computation of coherence using a 9×9 boxcar filter, and 4) geocoding.

Following this, we removed the Earth and topographic phase terms from the interferograms. Hence, the leftover phase contains the topographic information about the digital elevation model utilized in the process. One should note that phase removal is necessary for a better estimation of coherence in this context.

After that, the images went through a common-band spectral range filter. Finally, the interferometric coherence at both HH, VV, HH + VV, and HH - VV channels were computed. In this study, the height of ambiguity (HoA) for each TD-X

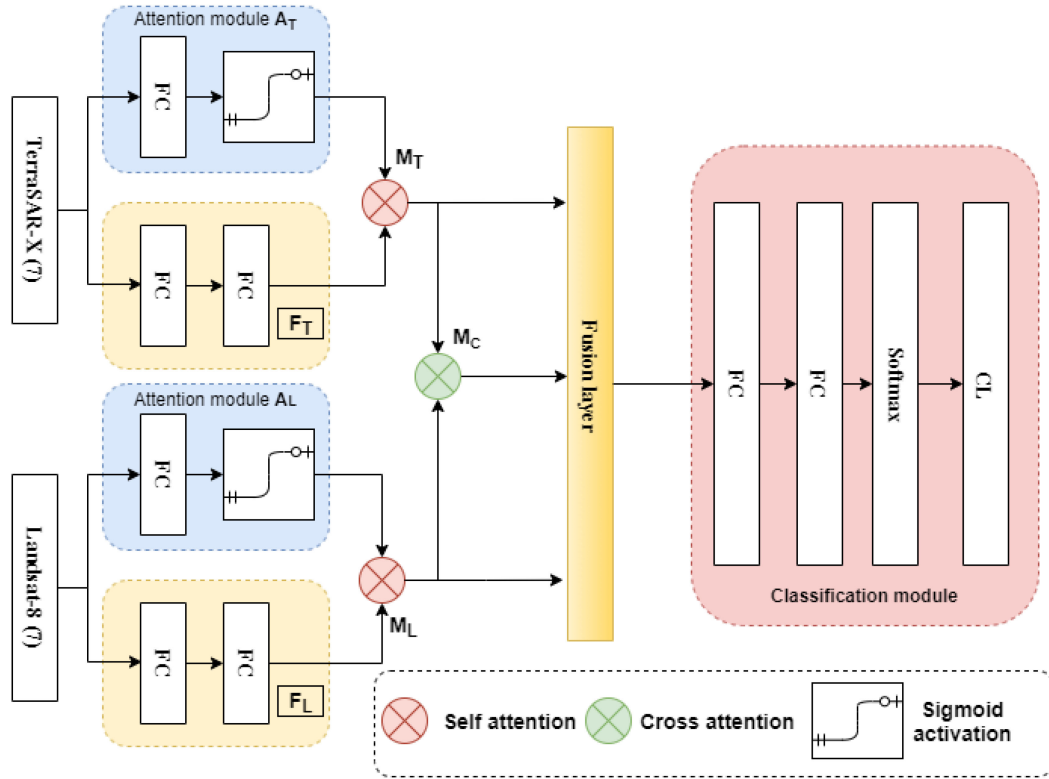


Fig. 3. Proposed fusion network architecture for crop-type classification. Here, “FC” denotes fully connected layer. Initially, we extract features from TanDEM-X and Landsat-8 data, represented as F_T and F_L , respectively. Simultaneously, the attention masks are generated through the attention modules, A_T and A_L , respectively. The attention masks are individually multiplied with the extracted features to get M_T and M_L . Thereafter, the common important feature, M_C is extracted using M_T and M_L . The concatenated features of M_T , M_L , and M_C are passed to the classification module for crop-type classification. Here, “softmax” represents the softmax activation layer and “CL” represents the classified output layer.

acquisition was around 5.8 m, which is considerably smaller than conventional TD-X data (> 30 m). This is because they were acquired as a part of the science phase (April–September 2015). Therefore, it allowed us to perceive variations in the coherence values for taller crops.

D. Model Overview and Network Architecture

The work intends to explore the spectral property from L-8 and the scattering property from TD-X data synergistically using the “attention framework.” The essential aspect of the attention framework is to highlight important features in a dataset to increase the class variance while enhancing the overall classification accuracy. We perform this classification enhancement by providing the TD-X and L-8 data points through two individual feature extractors and two individual self-attention modules. After that, we compute the characteristic features among the extracted features from TD-X and L-8 in a cross-attention manner. Following this process, all these features are transferred to a weighted fusion layer to enhance the information of the latent feature space. Subsequently, the resultant fused features are given to the classification module.

It is well known that optical sensors provide biochemical characteristics of a target whereas, SAR sensors provide geophysical features of a target. This proposed methodology combines these two pieces of information to provide valuable information

to classify diverse crop types. The overall architecture of the proposed fusion network is illustrated in Fig. 3. The proposed architecture consists of four modules. The first two modules are the attention modules of TD-X and L-8, A_T and A_L , respectively. The second module is a cross attention module, M_C among TD-X and L-8 features, and the last module is the classification process. These modules are discussed as follows.

TD-X Feature Extractor, F_T and L-8 Feature Extractor, F_L : Both F_T and F_L consists of three dense hidden layers for the extraction of spectral feature from L8 and TD-X. The output of each dense layer is nonlinearly transformed using a scaled exponential linear unit (Selu) activation function

$$\text{Selu} : \begin{cases} \text{scale} \times p & \text{if } p > 0 \\ \text{scale} \times \alpha \times (e^p - 1) & \text{if } p < 0 \end{cases} \quad (1)$$

where, scale and alpha are predefined constant: scale = 1.67 326 324 and alpha = 1.05 070 098. Therefore, these modules can be represented as follows: $F_T = f(w_{FT}, \mathbf{x}_T^i)$ and $F_L = f(w_{LT}, \mathbf{x}_L^i)$. w_{FT} and w_{LT} represent the weights of F_T and F_L , respectively. The outputs of F_T and F_L are 6D vectors.

Self Attention Modules, A_T and A_L : From the two input data, we also derive an attention module. A self-attention network helps in highlighting the salient features to appear in the forefront as required dynamically. Effectively, it reweighs the features according to some externally or internally assigned weights.

In this work, we use a soft-attention network wherein we give continuous weights using a sigmoid activation function.

A_T draws its attention mask from TD-X data and A_L draws its attention mask from L-8 data. These modules consist of two individual dense hidden layers with a sigmoid activation function. These modules are denoted as, $A_T = f(w_{AT}, \mathbf{x}_T^i)$ and $A_L = f(w_{AL}, \mathbf{x}_L^i)$. Here, w_{AT} and w_{AL} are the weights of the modules. The output of these modules are 6D vectors. These outputs are multiplied with F_T and F_L to provide highlighted features, expressed as, $M_T = F_T \otimes A_T$ and $M_L = F_L \otimes A_L$. Here, \otimes represent the broadcasted element-wise matrix multiplication operation, such that the resultant product retains the size of the matrix with higher dimension. Therefore, the outputs from M_T and M_L are again 6D vectors.

Cross Attention Layer, M_C : Following this procedure, we exploit the two self-attended data and design a cross-attention module. In the cross-attention network, we provide the TanDEM-X derived attended feature map to accentuate the spectral features of the Landsat-8 attended feature map. In effect, this aligns the SAR and the optical features while highlighting the salient features common to both the data streams. In this layer, the common highlighted features in M_T and M_L are enhanced. This module can be expressed as, $M_C = M_L \otimes M_T$. Similarly, the outputs from M_C is again a 6D vector.

Fusion Layer: In this layer, all the features from M_T , M_L , and M_C are stacked. The weighted fusion technique is then applied to generate a total of 12 output features. This layer is represented as, $Fus = f(w_{FUS}, [M_T, M_L, M_C])$. Here, w_{FUS} is the weight of the fusion layer.

Classification Module: The input to the classification module is the final fused features from the fusion layer. This module consists of four dense hidden layers. All the layers are treated by the Selu activation function except the last layer. We utilized the softmax activation in the last layer that is intended for classification purposes. This layer is represented as, $\mathcal{C} = f(w_c, Fus)$, where w_c are the weights of classification module. The output of the classification module is $1 \times Z$, where Z is the number of crop types.

Loss Function: We utilized the sparse categorical cross-entropy loss function to calculate the loss between the output and actual labels. This loss is then backpropagated to train the network architecture in an end-to-end fashion. The representation is the same for the sparse categorical cross-entropy and categorical cross-entropy loss functions. This loss function is defined as

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \mathbf{1}_{y_i \in C_c} \log(p_{\text{model}}[y_i \in C_c]) \quad (2)$$

where, N is number of observations, C indicates classes, $\mathbf{1}$ is the indicator function and $p_{\text{model}}[y_i \in C_c]$ is the predicted probability of observation i belonging to class c .

IV. RESULTS AND DISCUSSION

We utilize a temporal stack of TanDEM-X (TD-X) and Landsat-8 (L-8) data to classify crop fields over Seville, Spain. The temporal stack consists of five scenes of each sensor (TD-X

and L-8). In the case of L-8 data, we used the band reflectance values of coastal aerosol, blue, green, red, near-infrared, short wave infrared-1 (SWIR-1), and short wave infrared-2 (SWIR-2) while for TD-X data, we used the novel three decomposed power components: Odd bounce (P_s), even bounce (P_d), and diffused (P_v) and the coherence information in HH, VV polarization channels (Coh_{HH} , Coh_{VV}), as well as the Pauli channels (i.e., $\text{HH} + \text{VV}$ (Coh_{P1}) and $\text{HH} - \text{VV}$ (Coh_{P2})). The band reflectance values are generated using the ENVI software, whereas, the model-free decomposed power components are generated using the PolSAR tools plugin [29]. The cropping calendar is shown in Fig. 2. The temporal variations of spectral and scattering properties over the crop fields are shown in Figs. 4–7, respectively.

A. Temporal Analysis of Optical and SAR Descriptors

We can observe that during the acquisition time frame, all crop fields completed their preparation stage. Rice field was in the transition stage from land preparation to seeding stages. The carrot and wheat fields were subjected to early and mid-harvesting stages. Hence, depending on the different growth stages, changes in the reflectance values (for optical data), the scattering power components, and the interferometric coherence components are apparent from the plot. One can note from Fig. 4 that the trend of variations of coastal blue (CB), blue (B), green (G), and red (R) bands are pretty similar for most crops. However, comparatively the changes in R and G bands are more prominent than CB and B, which might be due to the sensitivity of these bands to leaf pigments.

For the majority of the crops, high reflectance in R is evident during the vegetation growth phase. However, rice shows a decreasing trend in the R reflectance. This lower trend might be due to the existence of the underneath water column, which might have absorbed this wavelength. Besides, the sensitivity in R and G for cotton is less than other crops, which might be due to the sparse canopy structure and the compound pinnate type leave structure.

Similarly, a clear distinction among growing, harvesting, and postharvest stages of wheat is evident from the plot. On 30-May, wheat was at the late growing stage. Hence, leaf pigments and the photosynthesis phenomenon absorbed most of the high-frequency wavelengths. Following this, the commencement of the harvesting stage increased the reflectance in R, G, B, and CB bands. Similar reflectance values in R, G, B, and CB continue throughout the postharvest stage of wheat.

In Fig. 5 we show the variations of near-infrared (NIR), short wave infrared-1 (SWIR-1), and short wave infrared-2 (SWIR-2). We know that NIR reflectance is highly dependent on the chlorophyll content and the mesophyll structure. Hence, for all the crops, the reflectance in NIR is high at peak vegetative phases. In particular, for corn, cotton, rice, and tomato, the dynamic ranges are also high. However, especially for cotton, the NIR reflectance follows a monotonically increasing trend due to its rapid vegetative growth within the observation window.

The NIR reflectance of corn and tomato is also high during 02-Jul as they have attained their peak vegetative stage. However,

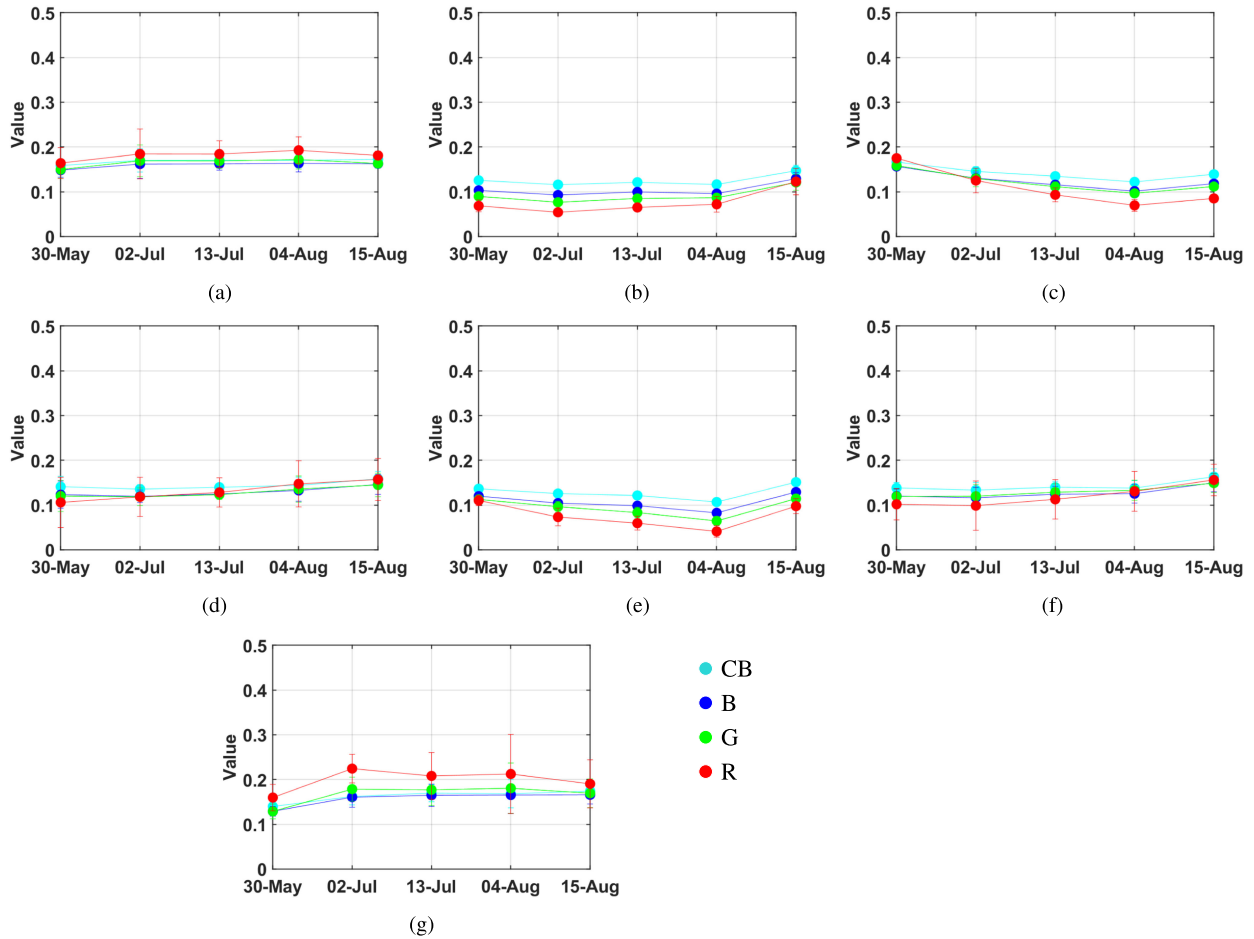


Fig. 4. Temporal variations of Coastal blue (CB), Blue (B), Green (G) and Red (R) bands for (a) Carrot, (b) Corn, (c) Cotton, (d) Quinoa, (e) Rice, (f) Tomato, and (g) Wheat crops.

toward the end of the observation period, the harvesting of these crops started. Hence, a decrease in the NIR reflectance is prominent from the plots. On the contrary, SWIR-1 and SWIR-2 are sensitive to soil moisture and crop water content. Therefore, we observe a marginal change in the dynamic range of the reflectance for all crops that depends on these two parameters.

In Fig. 6 the variation of scattering power components obtained from the dual co-polarimetric decomposition techniques is shown. It is interesting to note the variations of P_d , P_v and P_s for different crops. We can see that P_d is not sensitive for the carrot fields and its values are almost close to zero. However, P_s is marginally sensitive to the carrot fields. This might be due to the sparse canopy distribution of carrot crops. Due to the canopy structure, the backscatter values are highly affected by the underlying soil. Hence, P_s values are high throughout the season.

A similar trend in P_d is also evident for cotton, quinoa, tomato, and wheat crops. For the cotton crop, the observation window covers the progress of the vegetative stage. Hence, as the crop cover increased, the leaves of the top canopy layer generated a high amount of P_s power compared to P_v . However, depending on the randomness in the canopy structure, an increasing trend of P_v is also evident.

We observe a notable change in all the power components (P_d , P_v , and P_s) for rice. During the initial period, all the power components have similar values. However, depending on the field roughness, the P_v scattering power dominates. During 02-Jul, the advanced tillering stage of rice started. Hence, we observe an increase in the P_d power and an increase in the P_v scattering power due to increased scattering complexity from the random canopy structure of rice. We can further observe that the leaf structure also generated a significant P_s power in the time series.

We present the time series of single-pass interferometric coherence in Fig. 7. In the absence of temporal decorrelation, volume decorrelation [30] provides high sensitivity to vegetation height, i.e., with the increase in vegetation height, coherence decreases. This volume decorrelation information has been exploited in forests and crop studies to estimate vegetation height and recently for crop classification purposes [31]. Regarding the differences between polarimetric channels, they offer information about vegetation structure. However, they may also be affected by the signal-to-noise ratio (SNR), i.e., the lower the SNR, the lower the coherence. For example, in the time-series shown in Fig. 7, the lowest coherence is the one obtained from the 2nd Pauli channel (HH – VV) for all crop types except for rice.

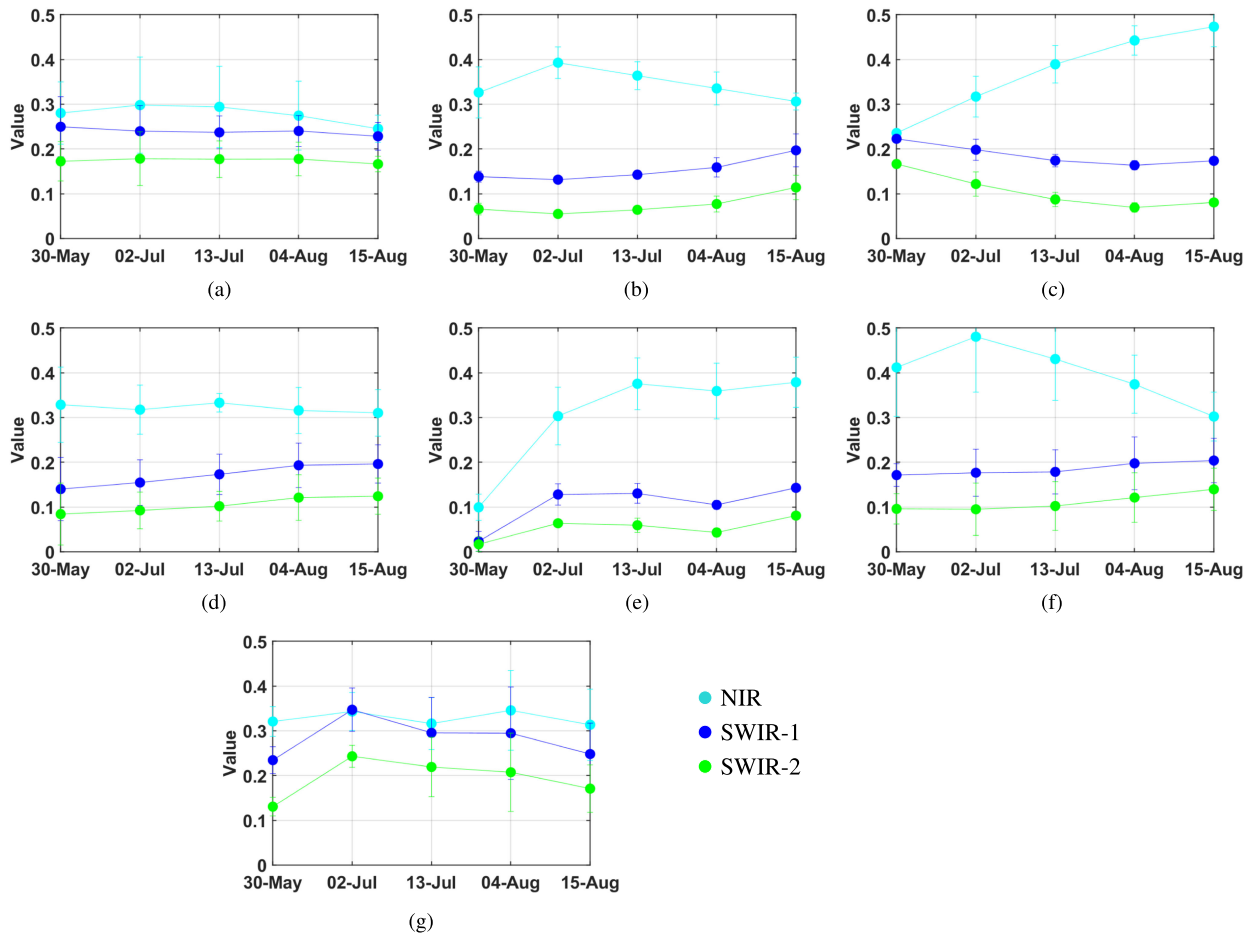


Fig. 5. Temporal variations of Near infrared (NIR), Short wave infrared-1 (SWIR-1), and Short wave infrared-2 (SWIR-2) bands for (a) Carrot, (b) Corn, (c) Cotton, (d) Quinoa, (e) Rice, (f) Tomato and (g) Wheat crops.

This observation is in agreement with the relative magnitude of the scattering components analyzed in Fig. 6.

We observe the largest excursion of single-pass interferometric coherence in Fig. 7 for rice, including an initial date with very low coherence due to the flooded condition of the fields. Cotton, carrot, and tomato fields exhibit high coherence all along the cultivation cycle. In contrast, corn and quinoa, which are tall crops at the observation period, are characterized by low coherence. Finally, winter wheat, which is harvested around the first acquisition date, also shows high coherence. We could attribute this observation due to the bare soil condition of the fields.

B. Classification Assessment

We base our classification results on several input features, with or without the attention module in the network architecture. In this regard, we first show the classification accuracy using the classification module individually for SAR and optical data. Following this, we use the fusion module and the classification module without any attention mechanisms. Later, we use the attention module with the fusion of optical and SAR scattering power component data. Finally, we show the classification accuracy using the fusion of optical, SAR scattering power components, interferometric coherence data, and the attention

TABLE I
OVERALL CLASSIFICATION ASSESSMENT OVER THE STUDY AREA

Feature set	OA (%)	κ	F1-score
Only SAR	83.41 \pm 2.06	0.82	0.84
Only Optical	87.36 \pm 1.76	0.88	0.87
Weighted fusion (without attention)	89.10 \pm 1.49	0.91	0.93
Weighted fusion (with attention)	92.10 \pm 1.01	0.94	0.94
Weighted fusion (attention + interferometric coherence map)	94.62 \pm 0.92	0.96	0.94

mechanism. In addition to this, we present the standard deviation of the classification accuracy with 20 repeated execution of the network architectures.

The classification results obtained from the abovementioned parameters are shown in Tables I–III. Table I shows the general assessment of the classification result using overall accuracy (OA), kappa coefficient (κ), and F1-score. Tables II and III presents the producer's and user's accuracy which provided detailed classification performance for each crop type.

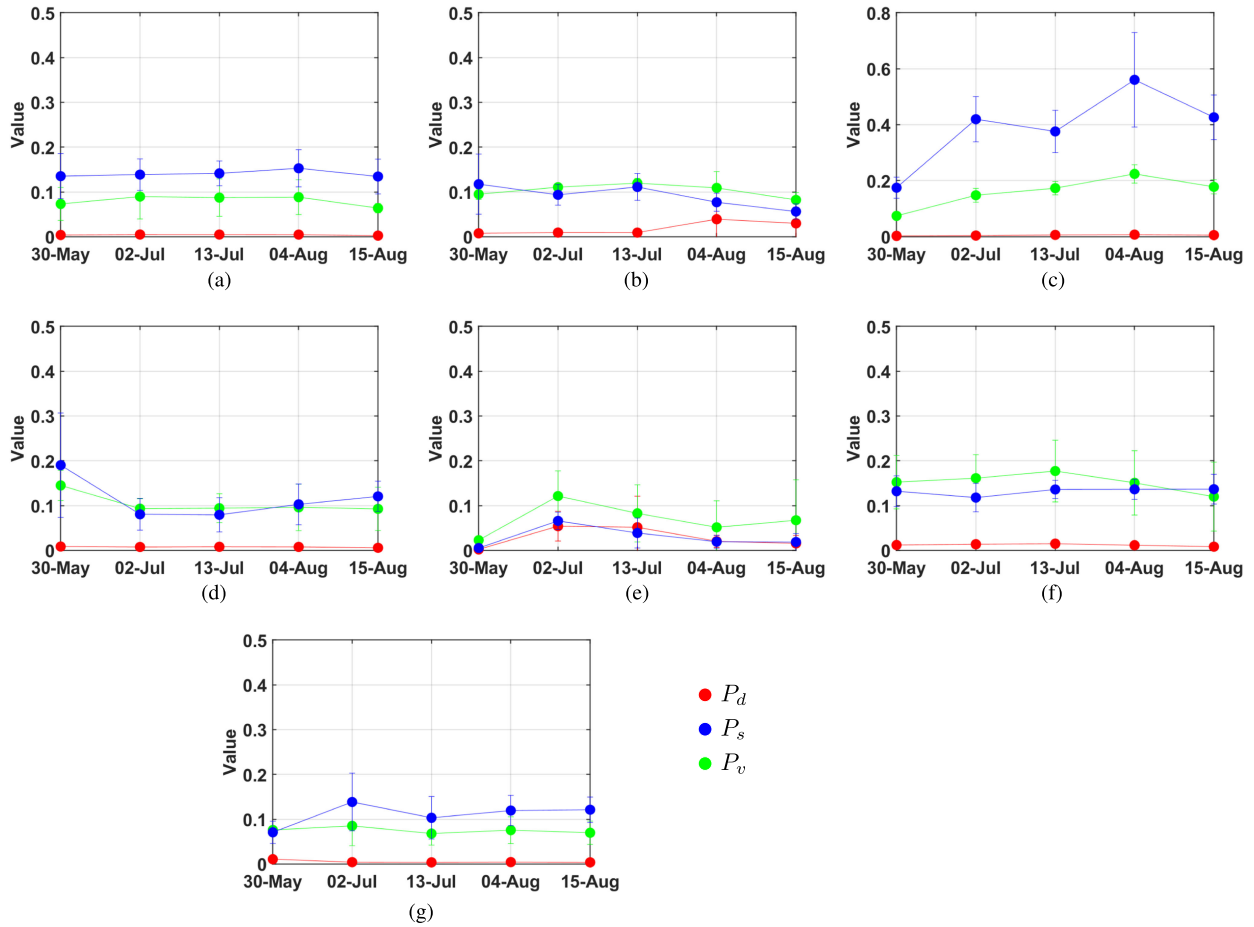


Fig. 6. Temporal variations of P_d , P_v and P_s bands for (a) Carrot, (b) Corn, (c) Cotton, (d) Quinoa, (e) Rice, (f) Tomato, and (g) Wheat crops.

TABLE II
PRODUCER'S ACCURACY FOR DIFFERENT CROP USING DIFFERENT INPUT PARAMETERS AND ATTENTION LAYER

Producer's accuracy (PA) in (%)							
Crop	Carrot	Corn	Cotton	Quinoa	Rice	Tomato	Wheat
Only SAR	13.70	90.60	96.47	61.82	82.60	78.13	49.71
Only Optical	14.02	91.46	95.76	66.44	91.41	88.64	68.12
Weighted fused (without attention)	29.66	96.39	96.29	55.89	98.39	87.34	79.70
Weighted fused (with attention)	35.72	95.00	97.39	83.96	96.39	91.93	82.24
Weighted fused (attention + interferometric coherence map)	68.23	97.60	96.23	85.49	98.72	95.49	91.87

Table I shows the OA for different classification processes. We can see that the OA using only SAR is 83.41% and the κ is 0.82. This classification accuracy is the lowest among all the other processes shown in the table. One possible reason might be the crop structural conditions. Most of the crop types were at the advanced vegetative to the preharvest stage. Hence, the vertical structure with random branch orientation is common for all crop types. Also, depending on the crop height, the penetration of the SAR signal through the crop canopy is evident. This signal

has high interaction with the underlying soil layer, increasing the P_s value for carrot, cotton, quinoa, and tomato. Besides, the random crop structure also generated a significant amount of P_v power for all the crop types. On the other hand, the classification accuracy using the only optical data is 87.36%. This might be due to the added separation capability of the crop biochemical characteristics. For example, the carrot field at the advanced phenology stage appears highly greenish, and the density increases. In contrast, reddish flowers appear for quinoa

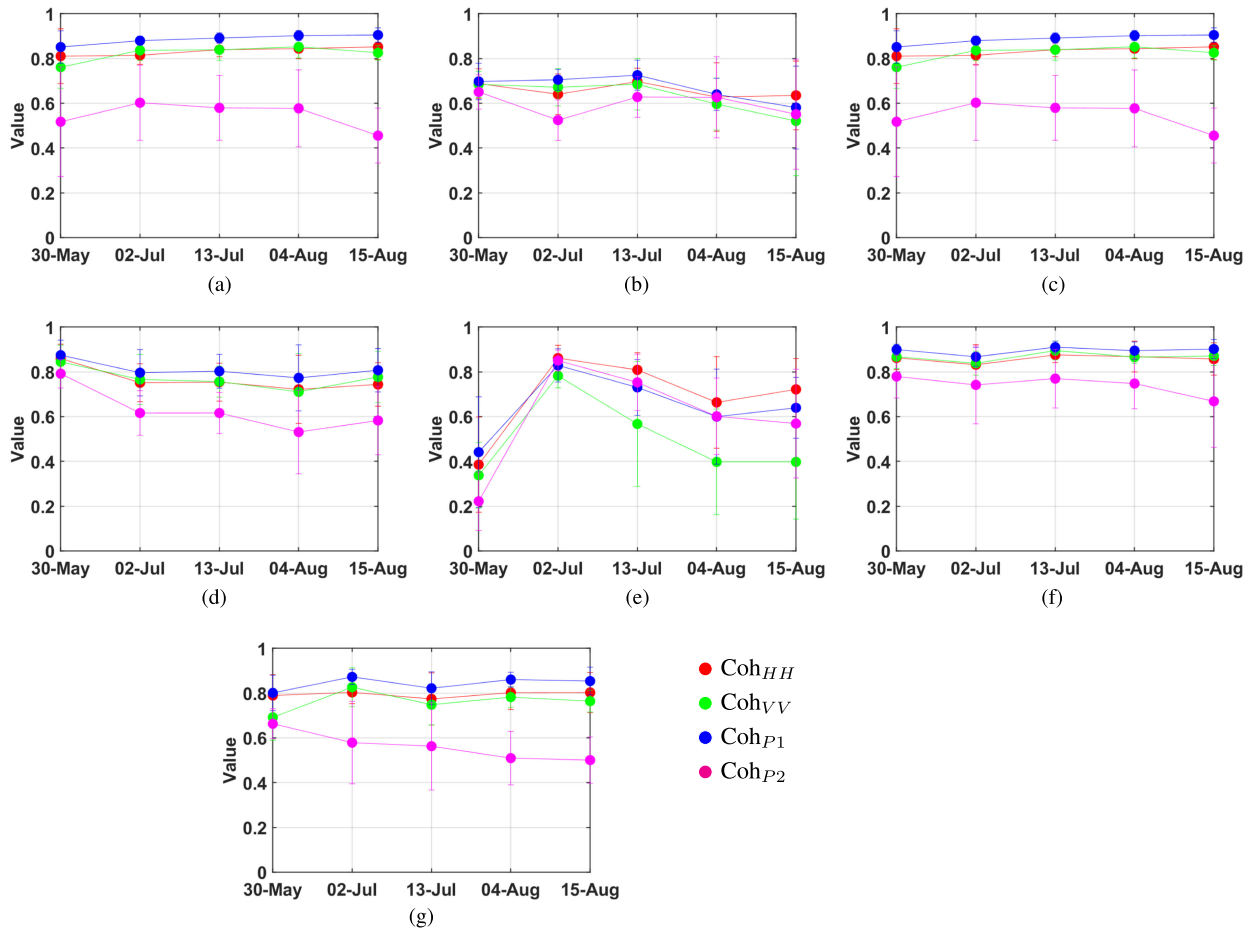


Fig. 7. Temporal variations of coherence in HH channel (Coh_{HH}), coherence in VV channel (Coh_{VV}), coherence in HH + VV Pauli channel (Coh_{P1}), and coherence in HH – VV Pauli channel (Coh_{P2}) bands for (a) Carrot, (b) Corn, (c) Cotton, (d) Quinoa, (e) Rice, (f) Tomato, and (g) Wheat crops.

TABLE III
USER'S ACCURACY FOR DIFFERENT CROP USING DIFFERENT INPUT PARAMETERS AND ATTENTION LAYER

Crop	User's accuracy (UA) in (%)						
	Carrot	Corn	Cotton	Quinoa	Rice	Tomato	Wheat
Only SAR	15.76	81.98	88.66	78.61	97.70	81.13	69.61
Only Optical	60.76	91.24	89.87	82.03	93.36	82.04	80.61
Weighted fused (without attention)	72.46	95.13	90.36	89.69	98.63	87.02	74.72
Weighted fused (with attention)	56.20	93.38	93.72	91.43	97.73	90.34	90.29
Weighted fused (attention + interferometric coherence map)	79.46	94.43	96.21	94.70	98.74	93.71	89.08

at the flowering stage. Hence, the marginal change in the PA and significant change in UA for optical data than SAR data is evident in Tables II and III. Similarly, at the advanced growth stage, the corn fields possess yellow fruit, and the fields have high reflectance in all the R, G, and B channels. Similar changes in the reflectance values occur with the appearance of fruits and flowers for tomato and wheat crops. Notably, the crop water content also changes significantly from one crop to another.

Therefore, depending on this factor, the reflectance values for NIR and SWIR regions also vary considerably. Due to these reasons, a 4% higher accuracy is evident with only the optical data.

Subsequently, the weighted fusion of optical and SAR data increased the classification accuracy by approximately 6% as compared to only SAR data. This is due to the consideration of both biophysical and biochemical changes that appear with

TABLE IV
COMPARISON WITH OTHER EXISTING FUSION TECHNIQUES AND CLASSIFIERS

Methods	OA (%)	κ	F1-score
Additive fusion + SVM	86.97 \pm 2.62	0.89	0.87
Additive fusion + RF	89.62 \pm 2.10	0.91	0.90
Additive fusion + proposed classification layer	89.79 \pm 1.10	0.94	0.93
Weighted fusion + SVM	89.60 \pm 2.02	0.90	0.89
Weighted fusion + RF	92.21 \pm 1.61	0.92	0.91
Weighted fusion + proposed classification layer	94.62 \pm 0.92	0.96	0.94

crop phenology advancement. The weighted fusion technique essentially embeds the crop structural and dielectric information with the scattering power components. In contrast, the crop chlorophyll content, vegetation water content, and mineral composition of crop canopy are captured by the reflectance values of optical data. Hence, the weighted fusion technique transforms the optical and SAR data to more meaningful target information content than only SAR or optical data. As a result, the κ score has also increased to 0.91 with the F1 score to 0.93, along with the increase of OA for the weighted fused product.

The comparison with other existing fusion technique and classifiers are shown in Table IV. We can observe that OA is higher when we used the weighted fusion technique. Also, the classification accuracy is higher for the proposed classification layer. Amongst the examined classifiers, Support vector machines (SVM) provide the least classification accuracies of 86.97% for additive fusion and 89.60% for weighted fusion. The classification accuracies using random forest with 500 trees are intermediate. The OAs are 89.62% for additive fusion and 92.21% for weighted fusion. We observe the highest classification accuracies for weighted fusion and the proposed classification layer.

This classification accuracy further increases with the inclusion of the attention module. A 9% increase in OA is evident as compared to only SAR data. The κ score is also increased by 0.12, while F1-score by 0.10. This enhancement in the accuracy might be because attention usually provides an insight into the complete feature vector and creates connections among each feature attribute within the entire feature vector. Alternatively, it can be described as the contribution of each feature attribute to the overall feature context. The attributes which are more descriptive for an input instance are granted more weight. Hence, it helps in highlighting the most relevant/important information. Since the attention network helps boost these essential feature constructs, we obtain improved overall discernibility amongst the classes. This procedure leads to an increase in the accuracy of the framework. Although there is an overall increase in the OA, UA, and PA for different crops, we observe an anomaly in the PA and UA for corn and rice as seen in Tables II and III. The decrease in the UA and PA could be due to comparable SAR and optical signature patterns after employing the attention mechanism.

Finally, the addition of the interferometric coherence information in the weighted fusion of SAR and optical data achieves the highest classification accuracy with 94.62% and $\kappa = 0.96$. The increase in OA with respect to only SAR is approximately

11%. This is due to the inclusion of changes in all three dimensions. The polarimetric scattering power information and optical data provide relevance in the variation in the spatial context. In contrast, the coherence information captures the changes in height. A significant difference in the heights of different crop types is evident from the *in situ* data. Some of the crops are erectophile, and some are planophile in nature. Moreover, the changes in the height throughout the phenology period are significant. Hence, this additional information is gained through the fusion of interferometric data with polarimetric and optical data. Furthermore, the overall variation in the classification accuracy for each execution appeared stable, with a standard deviation of 0.92. Therefore, a difference of 11.21 is evident for the mean OA of the weighted fusion (attention + interferometric coherence map) and only SAR data. Assuming a significance level of 0.05, we can neglect the null hypothesis as the Z-value is 4.96 (>1.96). Similarly, Z-value between the weighted fusion and optical data is 3.66 (>1.96). We can also interpret the significance of the changes in terms of κ mean and standard deviation. We observe a Z-value of 5.29 between only SAR and weighted fusion and a Z-value of 4.87 between only optical and weighted fusion techniques. Hence, we can conclude that there is a significant change in classification accuracy after the fusion of interferometric, polarimetric, and optical data. The inclusion of the coherence elements at HH, VV, HH + VV, and HH - VV channels provides information about each crop's differential height, which might have improved the overall classification accuracy.

Concerning the values of PA and UA obtained for each crop type, the best classified crops are corn, cotton, rice, and tomato, with values above 90%. The lowest accuracy is obtained for the carrot crop. This observation is expected because the acquisition period of the TanDEM-X data is toward the end of its phenological cycle. During the acquisition period, the crop was in the late harvest stage. Hence variation in the temporal pattern might not be distinguishable in the classification process.

The first set of the process includes computing the scattering power components from the MF3CD technique [21] and the coherence information from HH, VV, HH + VV, and HH - VV channels. For corn, cotton, rice, and tomato, PA and UA are $\geq 80\%$ as most phenological changes are captured by the SAR backscatter coefficients within the observation window. In contrast, the PA and UA of wheat are 49.71% and 69.61%, which might be low because of the saturation in the scattering power components during the harvesting and postharvest preparation stages.

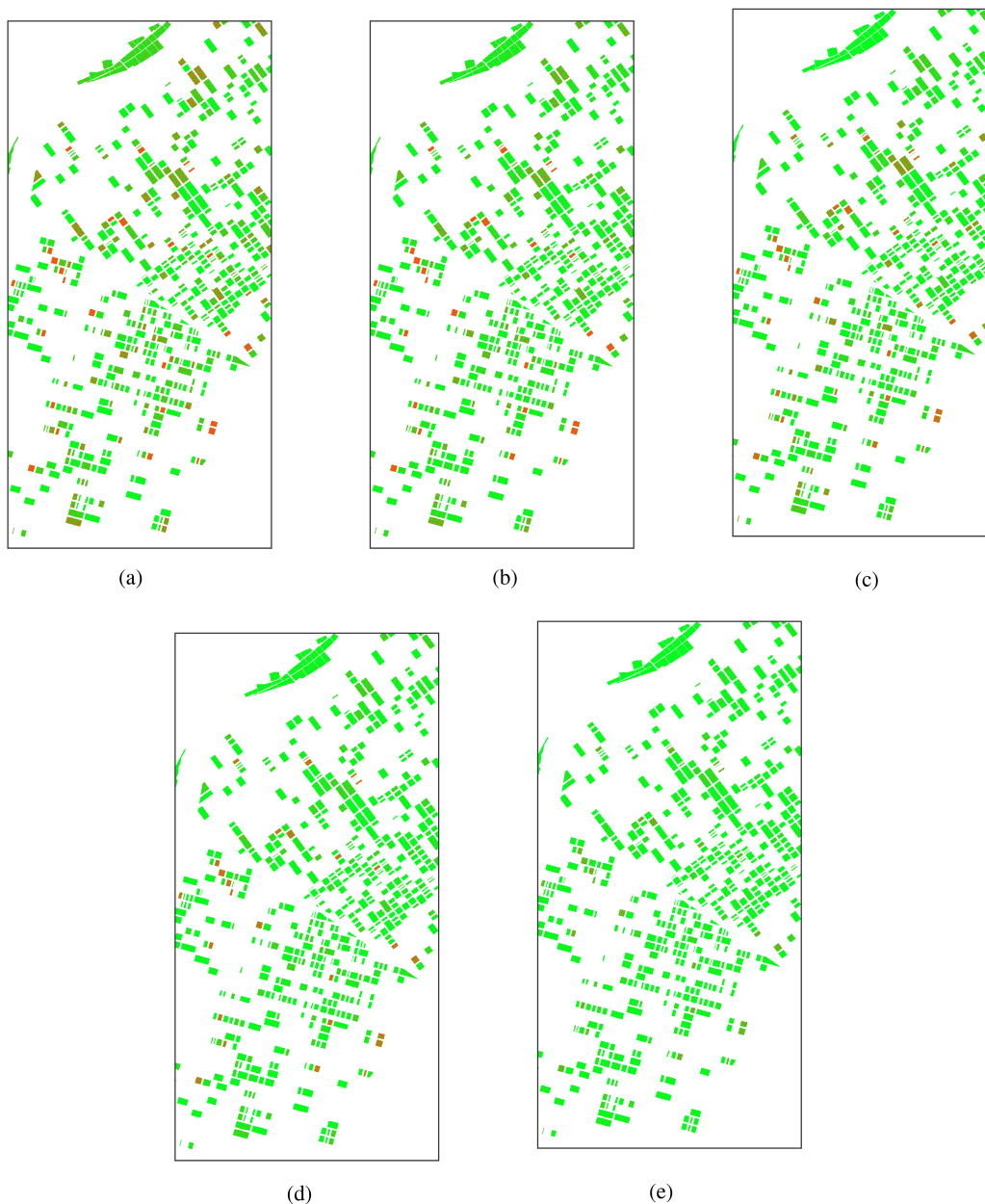


Fig. 8. Classification error map for different input parameters from SAR and optical data. The “red” color represents misclassified labels and “green” color represents the correctly classified labels.

The second processing step includes obtaining reflectance values from the optical bands of the Landsat-8 sensor. In this case, the UA and PA for all the crops have marginally increased, except for the carrot fields. As stated earlier, this field was in the period of late harvesting. Hence, it might have been restricted to low classification accuracy. The PA and UA of wheat have also increased. This might be due to the anomaly of the spectral signature from harvest to the postharvest stage that was likely missed by SAR backscatter coefficients.

The third and fourth sets include the classification using fused SAR and optical data without an attention mechanism. Please note that using the fusion framework, the PA and UA for all crops have increased significantly. This observation might be due to the embedding of biophysical and biochemical information in the fused products. As such, we observe an average increase of

approximately 6% to 10% for all crops. Interestingly, the use of the attention mechanism has also improved the classification accuracy of the fused product by 3% to 4%. Therefore, it is noteworthy that the attention mechanism can significantly focus on the input feature space within the network architecture. As a result, the accuracy of the fused product is considerably enhanced.

The last set of the process includes the fusion of the optical bands with SAR scattering components and the four coherence maps of HH, VV, HH + VV, and HH – VV channels with the attention mechanism. This particular set increased the UA and PA of corn, cotton, quinoa, rice, and tomato by $\approx 2\%$ to 3%. We observe an interesting increase for the carrot and wheat crops. The UA and PA of carrot is 79.46% and 68.23%, respectively. Besides, the UA and PA of wheat is 89.08% and

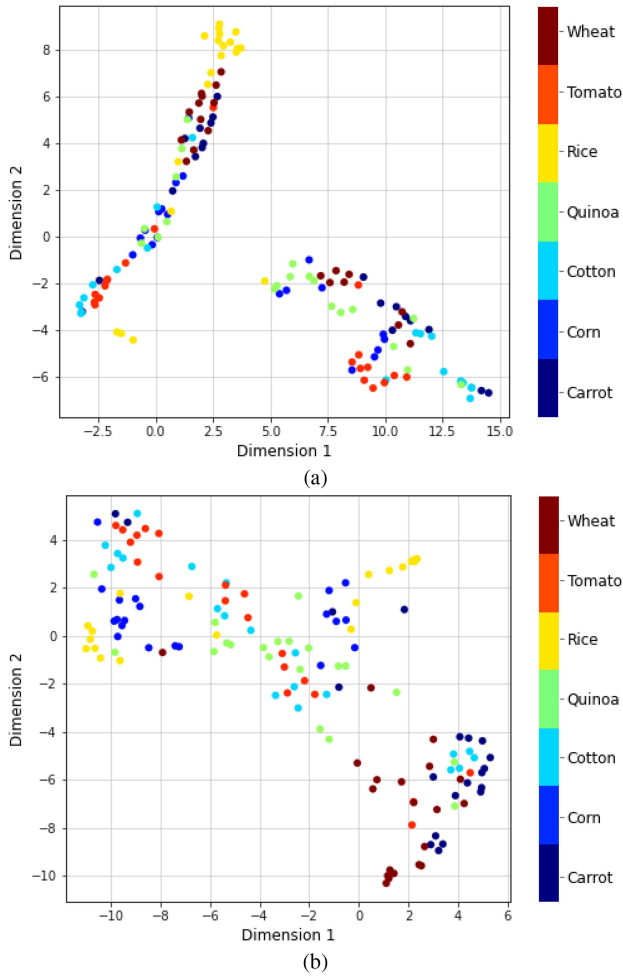


Fig. 9. Visualization of the data clusters obtained for different crop fields for (a) Tandem-X data and (b) Landsat-8 data using t-SNE plot.

91.87%, respectively. This increase in UA and PA might be due to the difference in crop height before and after harvesting. The coherence information well captures the accumulation of stubble in the fields after harvesting that.

Therefore, it is shown that OA, PA, and UA increased significantly when we fused optical and SAR (scattering power components + interferometric coherence map) data along with an attention mechanism than using only SAR (scattering power components + interferometric coherence map) and optical data. We have also shown the classification error map in Fig. 8. We can observe that the fusion technique has reduced the error pixels significantly. To visualize the separation of the data points using optical and SAR (scattering power components + interferometric coherence map) data, we have presented the t-distributed stochastic neighborhood embedding (t-SNE) plots in Figs. 9 and 10. Fig. 9 shows the t-SNE plot of SAR and optical data points for different crops, respectively. Fig. 10 shows the t-SNE plot of optical and SAR (scattering power components + interferometric coherence map) fused data points for the same crop types.

t-SNE is a nonlinear dimensionality reduction technique to embed high dimensional space into lower dimensions for efficient visualization of data points [32]. The feature distribution

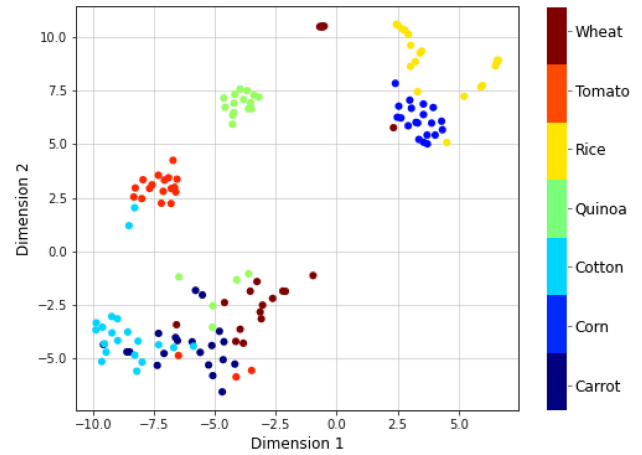


Fig. 10. Visualization of the data clusters obtained for different crop fields using t-SNE plot for optical and SAR (scattering power components + interferometric coherence map) fused product with attention module.

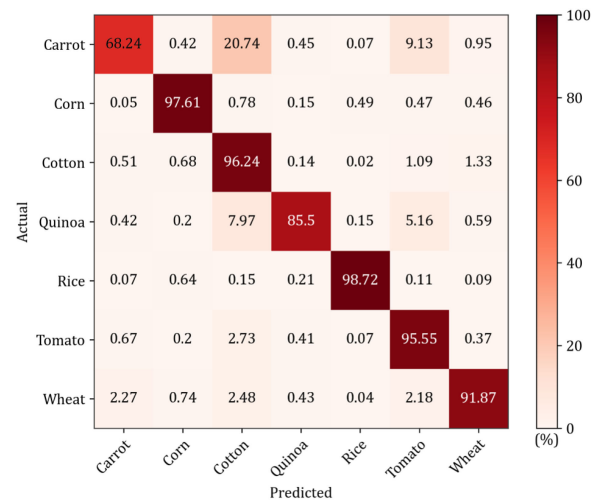


Fig. 11. Confusion matrix of optical and SAR (scattering power components + interferometric coherence map) fused product with attention module.

divergence is measured using the cost function in t-SNE. Within the high dimensional space, the objective of t-SNE is to pick similar points with high probability and dissimilar points with low probability. Therefore, t-SNE helps to visualize the mixing of high-dimensional representation of data points for different crop types.

We can see from Fig. 9 that the data points are highly cluttered for SAR features, while marginal separation among some crops is seen in the optical t-SNE plot. Hence, the OA is higher for optical features as compared to SAR features. On the other hand, in the case of fused products, i.e., optical and SAR (scattering power components + interferometric coherence map) with attention module, significant separation among the data points for different crops can be observed (Fig. 10). This is why classification accuracies are higher for the fused product than using only SAR and optical data. Besides, we observe a mixture of carrot and cotton, tomato, quinoa, and wheat. We also witness this mixing in the confusion matrix, shown in Fig. 11.

TABLE V
 DAY-WISE CLASSIFICATION REPORT USING OPTICAL AND SAR (SCATTERING POWER COMPONENTS + INTERFEROMETRIC COHERENCE MAP) FUSED PRODUCT WITH ATTENTION MODULE OVER THE STUDY AREA

Dates	Crop-fields					
	Carrot		Corn		Cotton	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
30-May	49.38	46.30	93.41	92.23	94.12	94.20
02-Jul	90.61	38.92	97.33	95.84	95.74	98.21
13-Jul	84.01	35.76	95.64	95.81	94.72	97.77
04-Aug	53.61	52.94	95.83	95.86	96.49	98.29
15-Aug	55.88	42.78	96.22	96.66	95.46	98.37

Here, PA: Producer's accuracy; UA: User's accuracy.

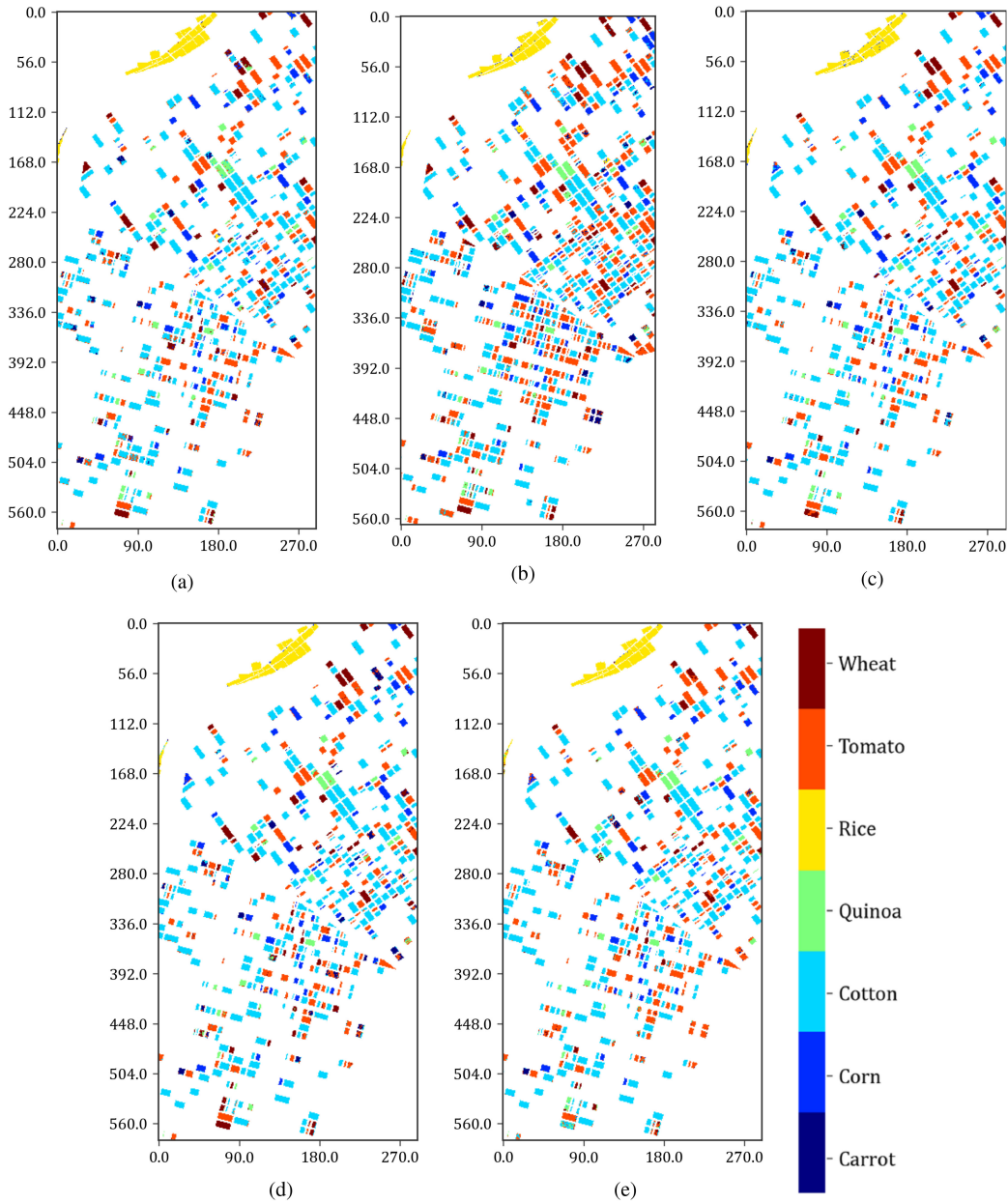


Fig. 12. Analysis ready classified crop-type map on different dates using optical and SAR (scattering power components + interferometric coherence map) fused product with attention module

TABLE VI
DAY-WISE CLASSIFICATION REPORT USING OPTICAL AND SAR (SCATTERING POWER COMPONENTS + INTERFEROMETRIC COHERENCE MAP)
FUSED PRODUCT WITH ATTENTION MODULE OVER THE STUDY AREA

Dates	Crop-fields							
	Quinoa		Rice		Tomato		Wheat	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
30-May	87.40	81.94	97.31	98.60	86.27	95.60	90.28	85.96
02-Jul	99.16	88.38	99.44	97.16	92.27	99.14	94.71	94.23
13-Jul	99.38	85.73	99.71	97.43	88.41	98.22	95.60	81.60
04-Aug	90.99	77.60	99.63	97.96	90.99	88.60	81.83	87.97
15-Aug	90.60	83.93	99.61	97.68	77.80	93.98	83.68	53.64

Here, PA: Producer's accuracy; UA: User's accuracy.

TABLE VII
DAY-WISE CLASSIFICATION REPORT USING OPTICAL AND SAR (SCATTERING
POWER COMPONENTS + INTERFEROMETRIC COHERENCE MAP) FUSED
PRODUCT WITH ATTENTION MODULE OVER THE STUDY AREA

Dates	Accuracy metrics		
	OA (%)	κ	F1-score
30-May	91.00	0.87	0.91
02-Jul	95.00	0.93	0.94
13-Jul	94.00	0.91	0.93
04-Aug	93.10	0.89	0.92
15-Aug	89.02	0.86	0.89

We have also shown the day-wise UA, PA, and OA using the weighted fused optical and SAR (scattering power components + interferometric coherence map) with attention module in Tables V–VII. Besides, an analysis-ready classification map for each day is provided in Fig. 12. The analysis-ready map shows good agreement with the *in situ* data. The day-wise accuracies show an increasing trend in OA from 30-May to 04-Aug due to crop morphological changes at different phenological stages. However, on 15-Aug, OA is reduced as compared to previous dates. This lower value of OA might be due to the similar response of most crops during their harvest or postharvest stages. Therefore, the weighted fusion technique with optical and SAR (scattering power components + interferometric coherence map) data may constitute a perfect complement to all the efforts performed so far with time-series data.

V. CONCLUSION

This article proposes a novel fusion technique using an attention-based network architecture for optical and SAR data. The optical dataset includes the reflectance of seven bands of the Landsat-8 sensor. The seven optical bands are coastal blue (CB), blue (B), green (G), red (R), near-infrared (NIR), short wave infrared-1 (SWIR-1), and short wave infrared-2 (SWIR-2). The SAR data include the three scattering power components from the novel model-free three-component dual copolarimetric decomposition (MF3CD) technique and the single-pass coherence map of HH, VV, HH + VV (P1), and HH – VV (P2) channels.

The results show that the scattering power components, coherence information, and the reflectance of optical bands are sensitive to the changes in the crop phenological stages in

time series. Depending on the crop geometry and structural properties, variations in the response of SAR observable are prominent. For example, rice shows a wider variation in the P_d scattering power due to the flooded ground. Similarly, NIR, R, and P_s power changes are evident toward the end of the season due to the harvest stage. Moreover, it is also exciting to observe crop height differences and variations in the coherence information. However, a few crops were nearing their harvest stage within the acquisition window, due to which a marginal variation of these parameters is evident from the plots.

For the classification accuracy, the fusion of SAR and optical data outperforms the accuracy obtained using SAR or optical data individually. Furthermore, an additional $\approx 3\%$ overall accuracy is achieved while we included interferometric coherence information in the fused product. Moreover, the use of the attention mechanism can focus on the essential features within the network architecture. As a result, the network is stable in terms of standard deviation and improves the classification accuracy. Therefore, the attention mechanism in the fusion of SAR and optical data shows a promising improvement in the classification result.

Although the results are very promising, a denser and wider time series of SAR and optical data might reduce confusion among certain crop types, particularly for low height and narrow cultivation time crops. Moreover, dense temporal data can inform cultivation practices and short revisit times to monitor and map diverse crop types throughout the season. Future studies might also include radar images acquired at C- and L-band, such as RADARSAT-2 and ALOS-2 satellites. C-band could discriminate the initial growth stages from advanced growth stages, while L-band might effectively discriminate the crop types at advanced phenological stages. These multifrequency analyses might provide better understanding of crop phenology and crop-type map to the end-user community.

ACKNOWLEDGMENT

The authors would like to thank E. Navarro (FERAGUA), M. P. González-Dugo (IFAPA), and L. Mateos (IAS-CSIC) for providing the reference data maps. S. Dey would like to thank the RSS Cloud Toolbox team, European Space Agency (ESA), for providing the computation facilities. The authors would like to thank the Overleaf¹ team for providing the latex editing platform.

¹[Online]. Available: <https://overleaf.com/>

REFERENCES

- [1] C. Pohl and J. L. Van Genderen, "Review article multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.
- [2] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.
- [3] B. Brisco, R. Brown, and M. Manore, "Early season crop discrimination with combined SAR and TM data," *Can. J. Remote Sens.*, vol. 15, pp. 44–54, 1989.
- [4] I. Sandholt, "The combination of polarimetric SAR with satellite SAR and optical data for classification of agricultural land," *Geografisk Tidsskrift-Danish J. Geography*, vol. 101, no. 1, pp. 21–32, 2001.
- [5] J. Qi, C. Wang, Y. Inoue, R. Zhang, and W. Gao, "Synergy of optical and radar remote sensing in agricultural applications," *Proc. Ecosystems' Dyn., Agricultural Remote Sens. Model., Site-Specific Agriculture*, vol. 5153, pp. 153–158, 2003.
- [6] X. Blaes, L. Vanhalle, and P. Defourny, "Efficiency of crop identification based on optical and SAR image time series," *Remote Sens. Environ.*, vol. 96, no. 3/4, pp. 352–365, 2005.
- [7] H. McNairn, C. Champagne, J. Shang, D. Holmstrom, and G. Reichert, "Integration of optical and synthetic aperture radar (SAR) imagery for delivering operational annual crop inventories," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 5, pp. 434–449, 2009.
- [8] N. Torbick *et al.*, "Integrating SAR and optical imagery for regional mapping of paddy rice attributes in the Poyang Lake Watershed, China," *Can. J. Remote Sens.*, vol. 37, no. 1, pp. 17–26, 2011.
- [9] G. Hong, A. Zhang, F. Zhou, L. Townley-Smith, B. Brisco, and I. Olthof, "Crop-type identification potential of Radarsat-2 and MODIS images for the Canadian prairies," *Can. J. Remote Sens.*, vol. 37, no. 1, pp. 45–54, 2011.
- [10] A. Sheoran and B. Haack, "Classification of California agriculture using quad polarization radar data and Landsat thematic mapper data," *GISci. Remote Sens.*, vol. 50, no. 1, pp. 50–63, 2013.
- [11] S. Skakun, N. Kussul, A. Y. Shelestov, M. Lavreniuk, and O. Kussul, "Efficiency assessment of multitemporal C-band Radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3712–3719, Aug. 2016.
- [12] N. Kussul, G. Lemoine, F. J. Gallego, S. V. Skakun, M. Lavreniuk, and A. Y. Shelestov, "Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1 A data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2500–2508, Jun. 2016.
- [13] J. Inglada, A. Vincent, M. Arias, and C. Marais-Sicre, "Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series," *Remote Sens.*, vol. 8, no. 5, 2016, Art. no. 362.
- [14] S. Park and J. Im, "Classification of croplands through fusion of optical and SAR time series data," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLI-B7, pp. 703–704, 2016.
- [15] D. Bargiel and S. Herrmann, "Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight TerraSAR-X data," *Remote Sens.*, vol. 3, no. 5, pp. 859–877, 2011.
- [16] R. Sonobe, H. Tani, X. Wang, N. Kobayashi, and H. Shimamura, "Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data," *Remote Sens. Lett.*, vol. 5, no. 2, pp. 157–164, 2014.
- [17] C. Hütt and G. Waldhoff, "Multi-data approach for crop classification using multitemporal, dual-polarimetric TerraSAR-X data, and official geodata," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 62–74, 2018.
- [18] S. Mirzaee, M. Motagh, H. Arefi, and M. Nooryazdan, "Classification of agricultural fields using time series of dual polarimetry TerraSAR-X images," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. XL-2/W3, pp. 191–196, 2014.
- [19] R. Sonobe, H. Tani, X. Wang, N. Kobayashi, and H. Shimamura, "Discrimination of crop types with TerraSAR-X-derived information," *Phys. Chem. Earth, Parts A/B/C*, vol. 83, pp. 2–13, 2015.
- [20] M. Busquier, J. M. Lopez-Sanchez, and D. Bargiel, "Added value of coherent copolar polarimetry at X-band for crop-type mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 819–823, May 2020.
- [21] S. Dey *et al.*, "Rice phenology mapping using novel target characterization parameters from polarimetric SAR data," *Int. J. Remote Sens.*, vol. 42, no. 14, pp. 5519–5543, 2021.
- [22] G. Wang, L. Guo, and H. Duan, "Wavelet neural network using multiple wavelet functions in target threat assessment," *Sci. World J.*, vol. 2013, pp. 632437, 2013.
- [23] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10705–10714.
- [24] S. Dey, A. Bhattacharya, D. Ratha, D. Mandal, and A. C. Frery, "Target characterization and scattering power decomposition for full and compact polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3981–3998, May 2021.
- [25] J. F. Gallardo, *The Soils of Spain*. Berlin, Germany: Springer, 2015.
- [26] L. Wang, J. J. Qu, X. Hao, and Q. Zhu, "Sensitivity studies of the moisture effects on MODIS SWIR reflectance and vegetation water indices," *Int. J. Remote Sens.*, vol. 29, no. 24, pp. 7065–7075, 2008.
- [27] S. Dey, A. Bhattacharya, A. C. Frery, C. López-Martínez, and Y. S. Rao, "A model-free Four component scattering power decomposition for polarimetric SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3887–3902, 2021.
- [28] R. Barakat, "Degree of polarization and the principal idempotents of the coherency matrix," *Opt. Commun.*, vol. 23, no. 2, pp. 147–150, 1977.
- [29] N. Bhogapurapu, S. Dey, D. Mandal, A. Bhattacharya, and Y. S. Rao, "Polsar tools: A QGIS plugin for generating SAR descriptors," *J. Open Source Softw.*, vol. 6, no. 60, 2021, Art. no. 2970.
- [30] M. Martone, P. Rizzoli, and G. Krieger, "Volume decorrelation effects in TanDEM-X interferometric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1812–1816, Dec. 2016.
- [31] M. Busquier, J. M. Lopez-Sanchez, A. Mestre-Quereda, E. Navarro, M. P. González-Dugo, and L. Mateos, "Exploring TanDEM-X interferometric products for crop-type mapping," *Remote Sens.*, vol. 12, no. 11, 2020, Art. no. 1774.
- [32] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.



India.

His research interests include land cover classification and agricultural crop mapping and monitoring using synthetic aperture radar data.



She is working on cross-modal retrieval of remote sensing (RS) and multimedia data with label scarcity. Her research interests include content-based retrieval of RS data using graph convolution networks.



His research interests include soil moisture estimation using synthetic aperture radar data.

Subhadip Dey (Graduate Student Member, IEEE) received the B.Tech degree in agricultural engineering from Bidhan Chandra Krishi Viswavidyalaya, Harinhat, India, and the M. Tech degree in aquacultural engineering, Department of Agricultural and Food Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India. He is currently working on crop mapping and monitoring using SAR data towards the Ph.D. degree with the Microwave Remote Sensing Lab, Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai,

Ushasi Chaudhuri (Graduate Student Member, IEEE) received the bachelor's degree in electronics and telecommunication engineering from Mumbai University, Mumbai, India, and the master's degree in visual computing working on cross-sensor data retrieval using AI techniques from IIT Kharagpur, Kharagpur, India, in 2017. She is currently working on cross-sensor data retrieval using advanced AI techniques toward the Ph.D. degree with the Centre of Studies in Resources Engineering (CSRE), IIT Bombay, Mumbai, India.

Narayanarao Bhogapurapu (Graduate Student Member, IEEE) received the B.E degree in civil engineering from Andhra University, Visakhapatnam, India, and the M. Tech degree in remote sensing and GIS, from the Department of Civil Engineering, National Institute of Technology Warangal, Warangal, India. He is currently working on soil moisture estimation using SAR data toward the Ph.D. degree with the Microwave Remote Sensing Lab, Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India.



Juan M. Lopez-Sanchez (Senior Member, IEEE) received the ingeniero M.S. and doctor ingeniero Ph.D. degrees in telecommunication engineering from the Technical University of Valencia, Valencia, Spain, in 1996 and 2000, respectively.

From 1998 to 1999, he worked as a Predoctoral Grantholder with the Space Applications Institute, Joint Research Centre of the European Commission, Ispra, Italy. Since 2000, he leads the Signals, Systems and Telecommunication Group of the University of Alicante, Alicante, Spain, where he has been a Full Professor, since November 2011. He has coauthored more than 90 papers in refereed journals and more than 130 papers and presentations at international conferences and symposia. His research interests include microwave remote sensing for inversion of biophysical parameters, polarimetric and interferometric techniques, SAR imaging algorithms, and applications of radar remote sensing in agriculture and geophysics.

Dr. Lopez-Sanchez was the recipient of the Indra Award for the best Ph.D. thesis about radar in Spain, in 2001. From 2006 to 2012, he was the Chair of the Spanish Chapter of the IEEE Geoscience and Remote Sensing Society.



Biplab Banerjee received the M.E. degree in computer science and engineering from Jadavpur University, Kolkata, India, in 2010, and the Ph.D. degree in image analysis from the Indian Institute of Technology (IIT) Bombay, Mumbai, India, in 2015.

He was a Postdoctoral Researcher with the University of Caen Basse-Normandy, France, and the Instituto Italiano di Tecnologia, Genova, Italy. He then was an Assistant Professor with the Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India, between 2016 and 2018. He is currently an Assistant Professor with the Indian Institute of Technology Bombay.

His research interests include computer vision and machine learning.

Dr. Banerjee was the recipient of the Encellence in Ph.D. Thesis Award for his thesis from the IIT Bombay.



Avik Bhattacharya (Senior Member, IEEE) received the integrated M.Sc. degree in mathematics from the Indian Institute of Technology, Kharagpur, India, in 2000, and the Ph.D. degree in remote sensing image processing and analysis from Télécom ParisTech, Paris, France, and the Ariana Research Group, Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, Nice, France, in 2007.

He is currently an Associate Professor with the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay (IITB), Mumbai, India. Before joining IITB, he was a Canadian Government Research Fellow with the Canadian Centre for Remote Sensing (CCRS) in Ottawa, ON, Canada. He is currently leading the Microwave Remote Sensing Lab (www.mrslab.in) at CSRE, IITB. His current research interests include SAR polarimetry, statistical analysis of polarimetric SAR images, applications of Radar Remote Sensing in Agriculture, Cryosphere, Urban and Planetary studies.

Dr. Bhattacharya was the recipient of the Natural Sciences and Engineering Research Council of Canada Visiting Scientist fellowship at the Canadian national laboratories, from 2008 to 2011. He is the Editor-in-Chief for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He has been the Guest Editor for the special issue on *Applied Earth Observations and Remote Sensing in India* in IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, 2017. He was one of the Guest Editors for the special stream on Advanced Statistical Techniques in SAR Image Processing and Analysis in IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 2018. He is the Founding Chairperson of the IEEE Geoscience and Remote Sensing Society Chapter of the Bombay Section.



Dipankar Mandal (Member, IEEE) received the B.Tech. degree in agricultural engineering from Bidhan Chandra Krishi Viswavidyalaya, Haringhata, India, in 2015, and the M.Tech and Ph.D. dual degrees in geoinformatics and natural resources engineering from the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India, in 2020.

He was a Visiting Researcher with the Agriculture and Agri-Food Canada (AAFC), Ottawa, ON, Canada, and Carleton University, Ottawa, ON, Canada, from October 2018 to February 2019. As a Visiting Researcher, he contributed to synthetic aperture radar (SAR) intercomparison experiment for crop biophysical parameter estimation within the Joint Experiment for Crop Assessment and Monitoring (JECAM) network of GEO Global Agricultural Monitoring. His research interests include applications of SAR polarimetry for crop classification, vegetation biophysical parameter estimation, deriving radar vegetation indices, and yield forecasting.

Dr. Mandal was a recipient of the Shastri Research Student Fellowship Award by the Shastri Indo-Canadian Institute, India, from 2018 to 2019. He was the recipient of the IEEE Geoscience and Remote Sensing Society India Best Ph.D. thesis award, in 2020, and the 'Excellence in Ph.D. Research' from Indian Institute of Technology Bombay, in 2021, for his doctoral work entitled "Retrieval of Biophysical Parameters for Agricultural Crops using Polarimetric SAR Data" and outstanding contribution toward improvement of diverse techniques for agricultural applications utilizing synthetic aperture radar remote sensing data.



Yalamanchili Subrahmanyeswara Rao (Member, IEEE) received the M.Sc. degree in physics from Andhra University, Visakhapatnam, India, in 1982, and the Ph.D. degree in passive microwave remote sensing of soil moisture from the Indian Institute of Technology (IIT), Bombay, Mumbai, in 1992.

He joined the Centre of Studies Resources Engineering, IIT Bombay, in 1985, as a Senior Research Assistant, and then became a Research Scientist, in 1999. During 2005–2009, he was a Senior Research Scientist, and then an Associate Professor, from 2009 to 2014. He is currently continuing as a Professor. He worked in both passive and active microwave remote sensing for several applications viz. soil moisture, flood mapping, and land use/land cover. He has participated in several spaceborne campaigns for collecting synchronous ground-truth data, and has experience in handling various datasets for several applications. His research interests include application of polarimetry for geo-physical parameter retrieval and SAR interferometry for DEM and displacement map generation.