# Synergistic exploitation of localized spectral-spatial and temporal information with DNNs for multisensor-multitemporal image-based crop classification

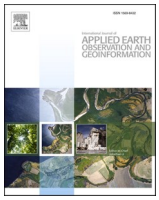
Gopal Singh Phartiyal, Dharmendra Singh, Hussein Yahia

International Journal of Applied Earth Observation and Geoinformation 125 (2023) 103595

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



# Synergistic exploitation of localized spectral-spatial and temporal information with DNNs for multisensor-multitemporal image-based crop classification

Gopal Singh Phartiyal [a], Dharmendra Singh [a], Hussein Yahia [b,*]

[a] Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India
[b] GeoStat, INRIA, Bordeaux Sud-Ouest 33405, France

ARTICLE INFO

*Keywords:*
Multisensor multitemporal
Spectral neighbourhood
bi-directional LSRMs
CNNs-RNNs

ABSTRACT

The challenge of performing efficient and reliable crop classification with multisensor multitemporal (MSMT) images in mixed land cover scenarios i.e. presence of small land parcels (area < 20,000-meter square) of crops and other land covers such as built-up or grasslands, is significant. Specially in countries (ex. India) where diverse crops are practiced in small land parcels. This challenge can be addressed if deep neural network (DNN) based models can exploit all three i.e. spatial, spectral, and temporal information of a crop, present in the MSMT images, efficiently and effectively. Therefore, this study presents a novel DNN based model that exploits all three information in a synergistic fashion to achieve improved crop classification. At first, the model increases the significance of local spectral information via a strategy that creates versions of spectral band set wherein neighbourhood of spectral bands is permuted. Then, the model utilizes three-dimensional convolutions, in a time-distributed fashion, to extract local spectral-spatial features. Finally, the model utilizes bidirectional long short-term memory or LSTM-RNNs to extract the temporal information embedded in the time-distributed feature-space created after the convolutions. The developed model is trained and evaluated on Sentinel-1 and Sentinel-2 MSMT data to achieve a 6-class classification including two major crops grown in the region. One of the proposed models namely *Perm-3D-CRNN-v1* showed a 97.77 % overall accuracy on test samples and reflected satisfactory on quantitative analysis. The localized spectral-spatial convolutions created prominent class-specific features whereas the bidirectional information flow in the recurrent layer improved the exploitation of crop-phenology type features making the model perform efficiently.

## 1. Introduction

In the recent years, India is strongly leaning heavily towards the use of open-source satellite images for crop monitoring and crop information extraction (Maurya et al., 2019; Murugan and Singh, 2018; National Remote Sensing Center, 2017; Phartiyal and Singh, 2018; Pravash, 2019; Ray and Neetu, 2017; Ray, n.d.; Sharma and Ghosh, 2023) however they still need to go far. A particular issue of 'mixed' land cover scenario, typical to Indian croplands remains unresolved during development of efficient crop monitoring applications with time-series/multitemporal (TS/MT) satellite images. The 'mixed' land cover scenario refers to a situation where small land parcels of multiple land cover/crop types are scattered over the study area. The diversity in crops grown by individual farmers having small land parcels creates such a situation. This is a common situation across all Indian croplands. Section 2.1 and section 2.3 demonstrates and discusses a similar situation. Numerous conventional and specific studies have been done on crop monitoring in these regions (Kumar et al., 2018; Maurya et al., 2018). However, simultaneous and synergistic exploitation of the spectral, spatial, and temporal information is key to address the challenge of crop monitoring with TS/MT satellite images in mixed land cover scenarios. In other terms, a more generic, efficient, scalable, robust approach is desirable.

In this context, recently, deep neural networks (DNN) with deep learning (DL) techniques are one of the most popular techniques (Xu et al., 2021; Zhong et al., 2019). In particular, CNNs and recurrent neural networks (RNNs) are the most popular variants utilized in satellite TS/MT image-based agriculture crop monitoring applications (Di

* Corresponding author.
*E-mail address:* hussein.yahia@inria.fr (H. Yahia).

https://doi.org/10.1016/j.jag.2023.103595
Received 30 April 2023; Received in revised form 31 October 2023; Accepted 27 November 2023
Available online 2 December 2023

Mauro et al., 2017; Ho Tong Minh et al., 2018; H. Li et al., 2020; Phartiyal et al., 2018; Rubwurm et al., 2017; Scarpa et al., 2018; Tang et al., 2020). One-, two-, and three-dimensional CNNs have been popular with TS/MT image-based crop classification. 1D-CNNs have been used to extract spectral and temporal features. Studies such as (Pelletier et al., 2019; Zhong et al., 2019) reflect to the utility of 1D-CNNs as feature extractors in both the spectral and temporal dimension. A study in (Kussul et al., 2017) established that 2D-CNNs achieved improved accuracy during crop classification over the 1D-CNNs which only extracted spectral features. Since then, numerous variants of 2D-CNNs have been developed for crop classification (Seydi et al., 2022). It is validated from various studies that 3D-CNNs are better spatiotemporal feature extractors than 2D or 1D CNNs for satellite time series images (Ji et al., 2020; Z. Li et al., 2020). However, studies have employed channel attention modules (Ji et al., 2020) or transformer hybrid modules (Z. Li et al., 2020) to separately focus on either, spectral or temporal information. More recently, (Teimouri et al., 2022) used multispectral-SAR (Sentinel-1 and Sentinel-2) multimodal time-series data. They considered R, G, B, and NIR from Sentinel-2 and VV, and VH from Sentiel-1. They prepared time-series for each feature separately. 6 time-steps for Sentinel-2 and 7 time-steps from Sentinel-1 are considered. A 7-class classification is performed. Separate 3D-CNN-based architectures are employed to extract the features from each channel/band. These studies establish that 3D-CNNs are better spatiotemporal feature extractors than 2D or 1D CNNs for satellite TS/MT images. Irrelevant to the type of CNN, they cannot exploit the temporal information due to lack of memory units. The temporal information is treated similar to as the spatial or spectral information which results in sub-optimal classification.

Recurrent neural networks or RNNs have the ability to exploit temporal dependence and therefore researchers have achieved good performance with RNNs in TS/MT image-based crop classification (Ho Tong Minh et al., 2018; Ienco et al., 2017; Ndikumana et al., 2018; Rubwurm et al., 2017; Sharma et al., 2018). A LTM-RNN model is proposed in (Ienco et al., 2017) for pixel as well as object-based land cover classification. This model is evaluated over two different multispectral datasets and the results indicated the potential of RNNs as classifiers in RS image data-based applications. More recently, authors in (Chen et al., 2022) reported a novel Im-BiLSTM model for crop classification with Sentinel-2 time-series data. Their model was able to address missing Sentinel-2 data from time-series. They employed an imputation strategy to figure out the missing values. In general, RNN architectures have been used for temporal and spectral-temporal feature-based crop classification. Attention and transformer models are employed to extract spectral features and then recurrent layers are used to extract temporal information. It is important to note from the above studies that RNNs lack the ability to exploit spatial information from TS/MT images. However, the potential of CNNs and RNNs can be put together to bridge the gap.

Therefore, since last few years, a large set of studies are utilizing CNNs and RNNs together in various fashion for TS/MT image-based crop classification (Benedetti et al., 2018; Chamorro Martinez et al., 2021; Gadiraju et al., 2020; Luo et al., 2020; Turkoglu et al., 2021; Yaramasu et al., 2020). In (Gadiraju et al., 2020), CNNs are used with RNNs in a parallel architecture to extract spectral-spatial–temporal features from a combination of MODIS time-series and NAIP high spatial resolution dataset. Recently, authors in (Turkoglu et al., 2021) developed a convRNN to encode and classify crop types at multiple levels. For the same pixel, the convRNNs provides three labels based on granularity. Similar studies are gaining attention but still there is scope for more. In summary, CNNs, RNNs, and CNNs-RNNs have been successful in analysing and processing single-sensor and multisensor TS/MT images for crop classification applications. Furthermore, the combined utilization of CNNs and RNNs, in-parallel or in-cascade fashion utilizes the best of both i.e. spatial-spectral feature extraction ability of CNNs and temporal feature extraction ability of RNNs. Therefore, the use of CRNN architectures in feature extraction and classification with TS/MT images for crop classification is highly motivating and a takeaway. Also, new models focussing on crop classification where cropland parcels are small or cropping pattern is highly diverse, are needful.

In this study, the use of CRNN architectures to exploit the spectral, spatial, and temporal information simultaneously is explored for multisensor-multitemporal (MSMT) image-based crop classification in Indian croplands. In fact, not many studies have attempted to develop DNN models for crop classification of Indian croplands i.e. mixed land cover scenarios (Gavade and Rajpurohit, 2020; Paul et al., 2022; Phartiyal and Singh, 2018; Sreedhar et al., 2022). Authors in (Sreedhar et al., 2022) developed a two-layer LSTM model for sugarcane mapping in India with multisensor-multitemporal satellite images. They used NDVI from Sentinel-2 and VH from Sentinel-1. Their model achieved a 99 % accuracy in identifying sugarcane in the region. However, their study focused on mapping sugarcane and not to achieve a generic crop classification. Authors in (Paul et al., 2022) used Sentinel-1 multi-temporal data for crop (5-crops) classification in the Indian agriculture landscape. They focused on pre-harvest classification however, they also reported performance with full temporal data. The employed a cascade of seven 2D-CNNs layers. They only utilized the CNN-based model for multi-temporal data which is sub-optima exploitation and is reflected in the overall accuracy (OA) of 89 % achieved by their nodal. (Phartiyal and Singh, 2018) provides a concise comparison of convolutional, recurrent and convolutional-recurrent layer architectures for MT image-based crop classification in Indian cropland scenarios. However, more such studies are still needed to set-up a stable DNN-based framework for crop classification in these regions. Another observation revealed from the review is that for crop classification in mixed land cover scenarios such as the one taken up in this study, exploiting the spectral, spatial, and temporal information simultaneously and directly is a necessity.

Models similar to CRNN models for TS/MT image processing reviewed previously can be explored and developed for crop classification in Indian cropland scenarios. Also, a focus of boosting the contribution of local spectral information during crop classification in mixed land cover scenarios could be beneficial.

This invites new studies which propose on development of novel CRNN models with focus on a judicious, simultaneous and more synergistic exploitation of the spectral, spatial, and temporal information on crops from MSMT imagery to achieve improved crop classification. Authors in (Phartiyal et al., 2020) proposed a novel strategy that strengthens the impact of spectral information while keeping the significance of the spatial and temporal information intact during land cover classification. They proposed a '*localized spectral convolutions on permuted spectral-neighbourhood*' strategy to boost contribution of localized spectral information during land cover classification. Their strategy proved successful during classification. However, their strategy does not involve consideration of temporal information. The addition of a complementary information (temporal in this case) makes the exploitation more challenging and requires more sophisticated and more integrated DNN models. Therefore, the current study aims to explore and extend the strategy developed in (Phartiyal et al., 2020) for MSMT image-based crop classification and aims to come up with focussed and efficient crop classification DNN models in the mixed class scenarios. Integrating their strategy with the power of CNNs and RNNs leads to a series of novel DNN models. These novel models involve; *-first* time-distributed permuted spectral band neighbourhood set generation for improving contribution of localized spectral information, *-second* time-distributed one-, two-, and three-dimensional localized spectral-spatial convolutions, and *-third* bi-directional RNNs for exploiting temporal information. The bidirectional RNNs capture the temporal profile whereas the time-distributed spectral-spatial convolutions capture the localized spectral and spatial profiles respectively of a particular crop. The bidirectional strategy in RNNs enables the influence of the past, present, and future crop information whilst deciding the present output. This is beneficial in time series data-based classification applications since the classification result can be observed from any time node. In short, in the study presented in this paper, an attempt is made; to

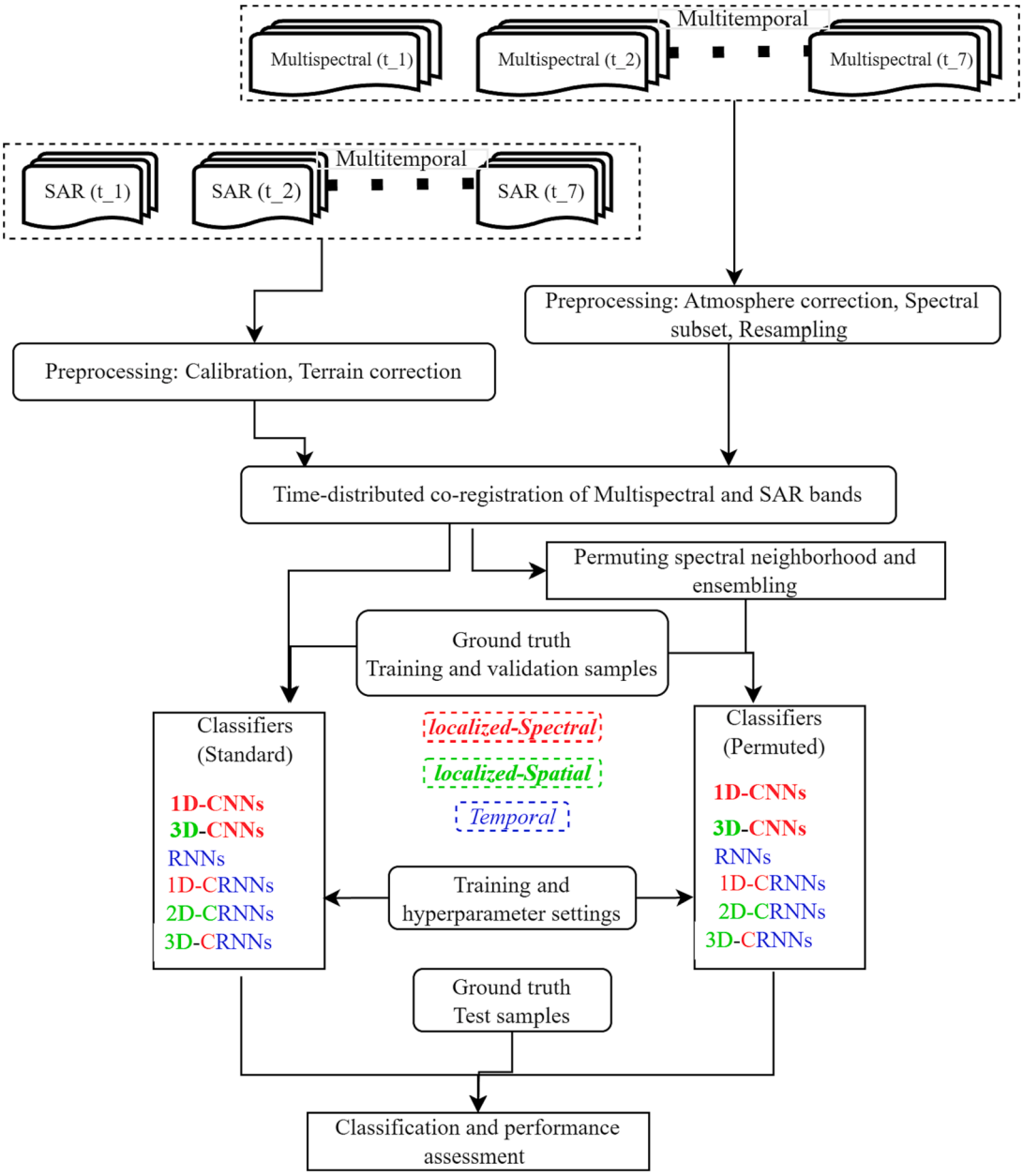



**Fig. 1.** Flowchart for the methodology opted during the study.

**Table 1**
Geographical extent (in coordinates) of study area.

| Study area | Coordinates (top left) | | Coordinates (bottom right) | |
|---|---|---|---|---|
| | Latitude | Longitude | Latitude | Longitude |
| | 29.84183 | 77.02115 | 29.78843 | 78.09080 |

develop novel CNN, RNN, and CRNN models that synergistically exploits spectral, spatial, and temporal information in agriculture crop classification with MSMT imagery.

This paper is organized as follows. Section 2 encompasses the materials used and methods developed in the study. Section 3 discusses on obtained results and its evaluation. Finally, section 4 concludes the study.

## 2. Materials and methods

The overall methodology is shown in Fig. 1.

### 2.1. Study area

The area selected for study is a cropland region in the Haridwar district of the Uttarakhand, India where Wheat, Sugarcane, Rice, Pulses, and Oilseeds are the major crops practiced seasonally. The region is considered based on accessibility to ground truth collection, crop season, and an exemplary mixed land cover scenario. The geographical extents of the region are provided in Table 1 and Fig. 2 shows the study area. The region is a land parcel which is managed by local farmers. Seasonal crops, majorly wheat in Rabi season (November-April) and Paddy in the Kharif season (June-October) are grown in this region. Sugarcane (annual) is the major commercial crop grown in this region. Various other crops are also grown in the region. However, the proportion of land cultivated under other crops such as Pulses and Oilseeds,

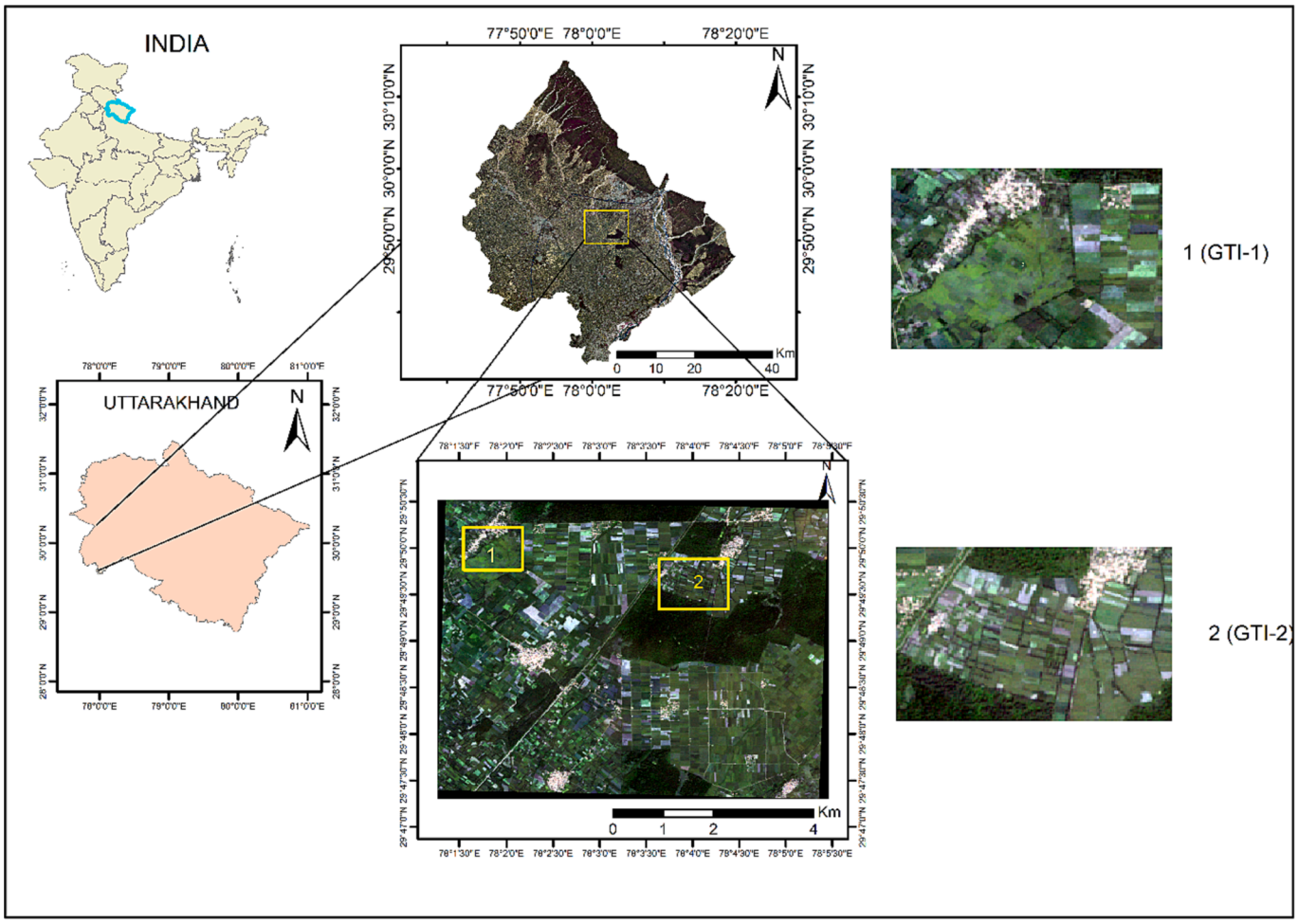



**Fig. 2.** Map highlighting the study area. Subset areas selected for visual inspection. 1: GTI-1, 2: GTI-2 (highlighted in yellow boxes and are shown in zoomed-in view alongside). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
Sentinel-1 and Sentinel-2 multitemporal data acquisition dates.

| Month | Acquisition dates | |
|---|---|---|
| | Sentinel-1 | Sentinel-2 |
| October | 3/10/2017 | 7/10/2017 |
| November | 20/11/2017 | 21/11/2017 |
| December | 2/12/2017 | 1/12/2017 |
| February | 12/2/2018 | 9/2/2018 |
| March | 8/3/2018 | 6/3/2018 |
| April-March | 1/4/2018 | 31/3/2018 |
| April | 13/4/2018 | 15/4/2018 |

**Table 3**
Band specifications of Sentinel-1 and Sentinel-2.

| Sensor | Band selection | Pixel Spacing (meters) |
|---|---|---|
| Sentinel-1 | VH | 10 |
| | VV | |
| Sentinel-2 | Band 2: Blue | 10 |
| | Band 3: Green | |
| | Band 4: Red | |
| | Band 8: NIR | |
| | Band 5: Vegetation Red Edge | 20 |
| | Band 6: Vegetation Red Edge | |
| | Band 7: Vegetation Red Edge | |
| | Band 8A: Vegetation Red Edge | |
| | Band 11: SWIR-1 | |
| | Band 11: SWIR-2 | |

vegetables, and fruits is quite small (<5%) compared to Sugarcane and Wheat. Adding to this, the region also has patches of forests, grasslands, and deciduous tree plantations (especially 'Popular'). Therefore, the region is suitable to be treated as a mixed class scenario for algorithm testing. The presence of multiple land covers will evaluate the generalization ability of the classification models. Areas marked (1, and 2) in the image are subsets of study area created for qualitative and visual inspection of the classification performance of classification models developed in the study.

### 2.2. Dataset

Sentinel-1 dual-pol SAR and Sentinel 2 multispectral sensor images are used together as a multisensor dataset in this study. Sentinel-1C band dual pol SAR GRD image products considered here are available at 10 m pixel spacing and at a revisit time of 6 days (12 if pass direction is considered). Similarly, Sentinel-2 multispectral (13-band) images are available at 10 m, 20 m, and 60 m spatial resolutions and at a revisit time period of approx. 5 days. Both sensor datasets are freely available for public use by the European Union Space Agency and providing complementary and consistent coverage throughout the year. Based on the growing season of wheat i.e. the Rabi season, Multiple cloud-free Sentinel-2 L1C satellite multitemporal images and Sentinel-1 ground range detected (GRD) are selected throughout the wheat growing Rabi season for the year 2017–18. Acquisition dates of each month for both sensors are listed in Table 2. The selection of acquisition dates is based on availability of cloud-free Sentinel-2 data and data present from each

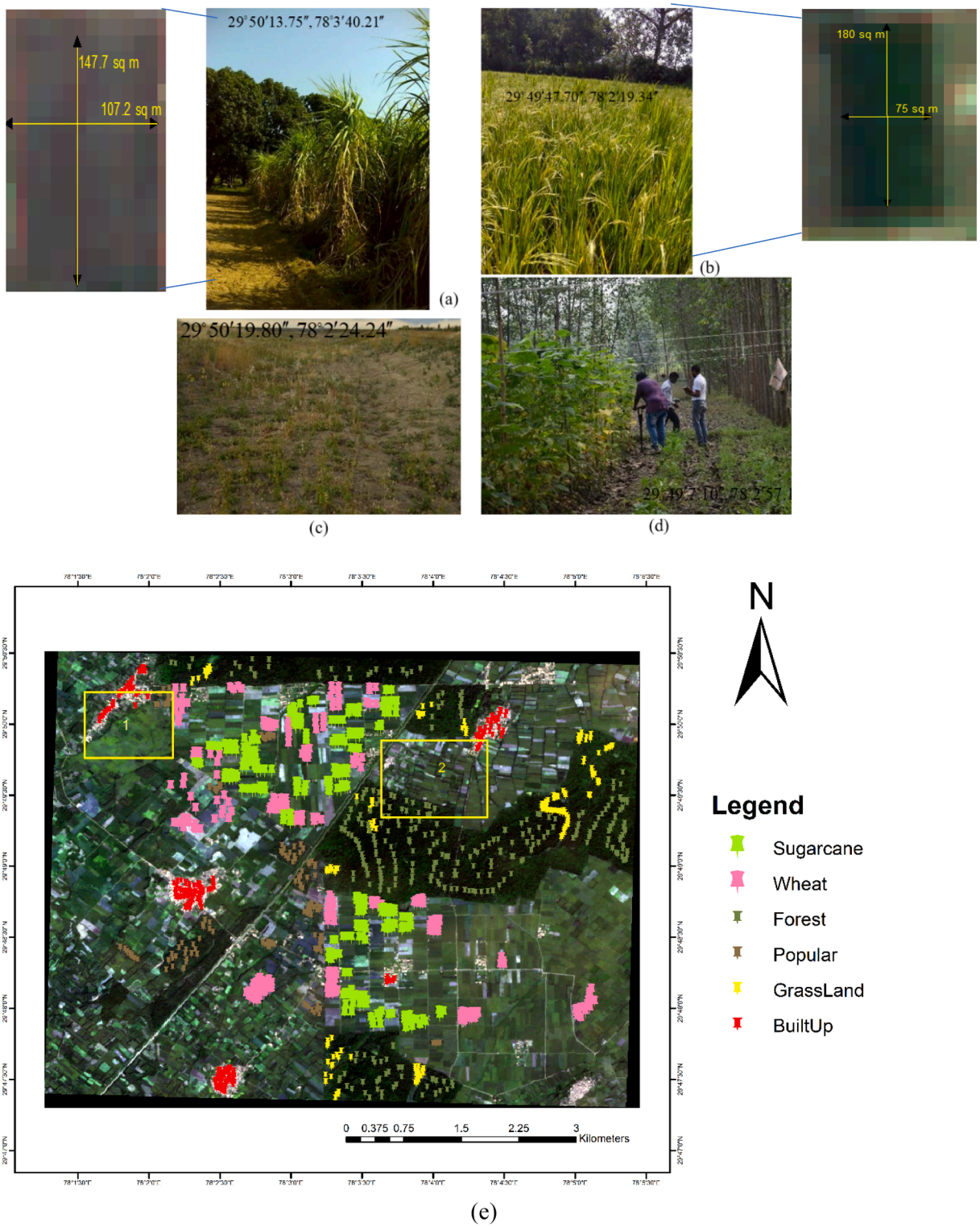



**Fig. 3.** Sample images collected during ground truth survey. (a) Sugarcane cropland, (b) Wheat cropland, (c) Grasslands, (d) Popular plantations. The dimensions of sample land parcel where Wheat and Sugarcane crops are grown are also presented (top-left and top-right), (e) Study area map with Ground truth samples overlaid.

month along the observation period. Also, Table 3 lists the specifications of bands considered from Sentinel-1 and Sentinel-2. Overall, 10 multi-spectral bands and 2 dual-pol SAR channels, and 7 time-steps from October 2017 to April 2018 are considered for the study.

### 2.3. Ground truth

Based on the study regions and crop season selection, wheat is the major crop grown in the area (District Office of Economics and Statistics, 2022). Moreover, sugarcane is also grown throughout the year in this region (District Office of Economics and Statistics, 2022). From section 2.1, it is clear that although other crops are practiced but their

**Table 4**
Summary of ground truth survey timeline and samples collected during each survey.

| Survey dates | Sensors | Points collected | Classes observed |
|---|---|---|---|
| December 1, 2017 | Aerial Imagery: DJI Phantom 4 | 120 | Wheat, Sugarcane, Popular, Forest, Grassland |
| December 18, 2017 | Terrestrial imagery: Canon DSLR | 127 | |
| January 5, 2018 | GPS: Garmin Approach G5 | 164 | |
| January 19, 2018 | Data roaster: Data entry | 67 | |
| February 8, 2018 | | 88 | |
| March 9, 2018 | | 92 | |
| April 13, 2018 | | 112 | |

*Built-up samples are collected via visual inspection.
**Additional samples are collected via visual inspection in the vicinity of originally collected points during survey. True color and false color composites of Sentinel-2 together with Google Earth imagery are used for this purpose.

**Table 5**
Summary of class-wise sample split for classifier training and validation, and testing.

| Class | Ground truth samples | | |
|---|---|---|---|
| | | Training and validation | Testing* |
| Wheat | 300 | 247 | 53 |
| Sugarcane | 300 | 226 | 74 |
| Forest | 300 | 245 | 55 |
| Popular | 300 | 250 | 50 |
| Grassland | 300 | 232 | 68 |
| Built-up | 300 | 240 | 60 |
| Total | 1800 | 1440 | 360 |

*Based on random sampling.

proportions are very low (District Office of Economics and Statistics, 2022). Adding to this, the size of the land parcel in which a farmer grows these crops is also small as shown in Fig. 3. These two reasons hinder the consideration of these crops as output classes in this study area. Overall, six classes are considered namely: wheat, sugarcane, forests, deciduous plantation mostly 'Popular' plantations, grassland, and built-up. In the considered study area, no permanent "Bare Soil" land cover was present where 'permanent' refers to presence of bare soil throughout the observation time-period. Hence, Bare Soil is not considered as class in this study. This situation does not affect the 'mixed land cover scenario' context significantly since other classes are still mixing. The ground truth samples are either measured directly on the terrain or are collected via visual inspection from the dataset. At first, a crop-growing season long observation period is set. Then, various ground surveys are scheduled based on Sentinel satellite's visit calendar and is changed based on local weather. DJI Phantom 4 is used for capturing aerial images whereas Garmin GPS and Canon's DSLR cameras are used for capturing location information and terrestrial images. Summary of ground truth survey timelines, instrumentation and inventory used, and samples collected during each survey is provided in Table 4 and are overlaid over the study area map as shown if Fig. 3. Further, visual interpretations of aerial imagery along with Google Earth images is employed to increase the sample points in the vicinity of the originally collected sample points. Built-up class samples are also added manually via visual aid.

Sample images from the ground truth survey are also shown in Fig. 3. Two subset areas from the study area are selected for visual and qualitative evaluation. Fig. 2 and Fig. 3 shows the two marked areas. The subset areas are termed as ground truth image (GTI) i.e. GTI-1, and GTI-2. It is important to note here that the land parcels of individual crops are also quite small in size compared to the Europe and the US (Fritz et al., 2015; Lesiv et al., 2019). Fig. 3 depicts the dimension of an arbitrary wheat (Fig. 3, top-right) and sugarcane (Fig. 3, top-left) field selected from the study area. The area of this land parcel is approximately 15000-meter square or 0.01-kilometer square which is small in context to the 10-meter pixel spacing satellite images considered here for study keeping in mind that these are the best resolution available in open-source mode. This indicates that a single cropland is covered within few pixels (~25 pixels).This is compounded by the diverse cropping practices of neighbouring land parcels (District Office of Economics and Statistics, 2022). The challenge of crop classification with TS/MT satellite images in these mixed land cover scenarios is very unique. An 80/20 % ratio is set for splitting the samples into training and validation, and testing respectively. Details of class-wise sample split is provided in Table 5.

### 2.4. Data preprocessing

Both Sentinel-1 and Sentinel-2 datasets require preprocessing before classification. Sentinel-1 GRD datasets are terrain corrected and calibrated. Terrain correction is performed using the SRTM 1 arc-second digital elevation model (DEM) available for public use. Speckle filtering is not performed as speckle noise is already suppressed in GRD products. Sentinel-2 L1C datasets are atmospherically corrected using the 'sen2cor' processor (Main-Knorn et al., 2017). Further the 20 m resolution bands in the bottom of atmosphere (BOA) Sentinel-2 datasets are resampled to 10 m using the 'nearest neighbourhood' strategy. The Sentinel-1 and Sentinel-2 preprocessing is done using the ESA's SNAP image processing tool. Finally, 2 polarimetric channels i.e. VH, and VV from preprocessed Sentinel-1 and 10 multispectral bands from preprocessed Sentinel-2 datasets are selected and stacked together to form a single time-stamp multisensor (SAR-multispectral) dataset. In all, 7 analogous multisensor datasets are created corresponding to the 7 time-steps. The 7 multisensor datasets are stacked along the temporal dimension creating a multisensor-multitemporal (MSMT) dataset. This MSMT image dataset is utilized for crop classification.

### 2.5. Permuted spectral neighbourhood and localized spectral convolutions based models for MSMT image crop classification

The strategy of permuted spectral band neighbourhood with localized spectral convolutions in MSMT image classification is introduced and discussed here with presence of two complementary information i.e. spatial and temporal and the impact is critically analysed and assessed. Initially, authors in (Phartiyal et al., 2020) established the significance and advantage of, permuted spectral neighbourhood and its exploitation with localized spectral convolutions, over the conventional spectral convolutions. Their study involves exploitation of spectral and spatial information in satellite image-based land cover classification. It is important to note here that the dataset used in their study does not contain the 'temporal' information. Hence the impact of their strategy in MSMT image applications is still unknown and needs investigation. Section 2.5.1 provides a critical analysis and discussion on how to utilize this strategy in MSMT image-based crop classification. With the help of powerful and more so developing DNNs, it may be possible to exploit all three (spatial, spectral, and temporal) aspects more efficiently. Apart from few studies employing 3D-CRNNs (Ji et al., 2020; Tang et al., 2022), all other DNN models developed for high dimensional TS/MT image data applications focus more on temporal and localized spatial information over the localized spectral information. However, in context to satellite based earth observation images, the localized spectral information is, as important as the spatial and temporal information. The idea of utilizing the spectral information more efficiently while keeping the significance of the spatial and temporal information intact should benefit in satellite TS/MT image applications such as crop classification. Therefore, in this section, various novel CNN, RNN, and CRNN models

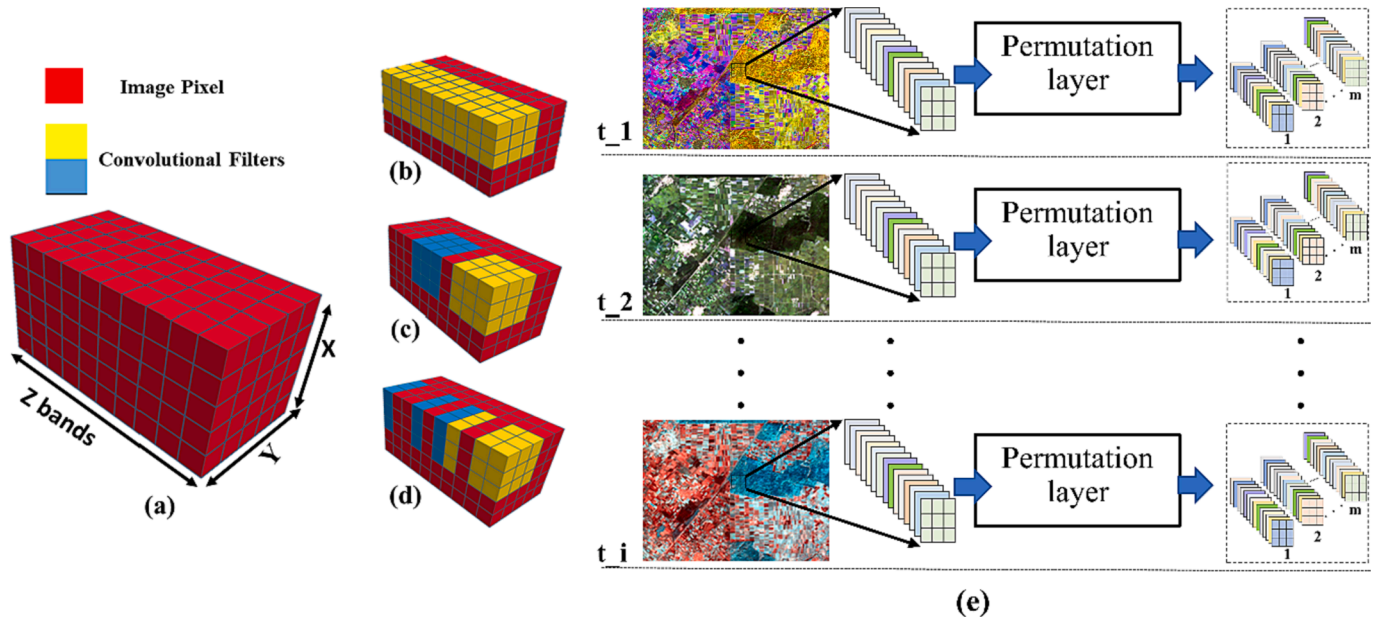



**Fig. 4.** (a) Data cuboid, (b) Conventional localized spatial convolution, (c) Localized spectral-spatial convolution with fixed spectral neighbourhood, (d) Localized spectral-spatial convolution with reordered and ensembled spectral neighbourhood (Phartiyal et al., 2020) and, (e) Process of employing the permuted spectral band set generation strategy on high-dimensional time-series images.

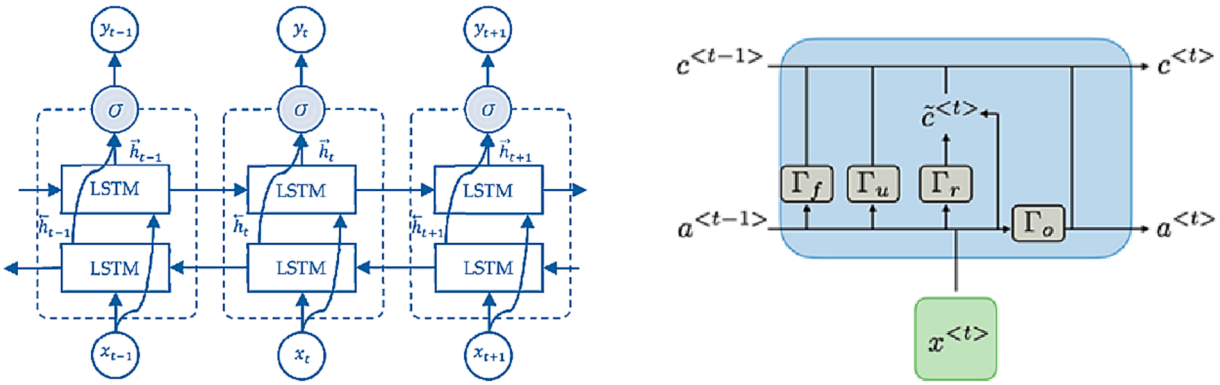



**Fig. 5.** (a) LSTM-based RNN architecture with bidirectional signal flow and (b) A LSTM unit.

**Table 6**
Architecture and hyperparameter settings of CNN models with permuted spectral convolution for MSMT images-based crop classification.

| Model | Layer | | | | |
|---|---|---|---|---|---|
| | Input | Permutation layer (P1) | Convolutional layers (C1, C2, C3) | Fully connected layer (FC1, FC2) | Output layer |
| Perm-1D-CNN-v1 | Input data size =7 × 12 (time, spectral)<br>Normalization = Batch Normalization | Permutated sets = 10<br>Kernel size = 12 (spectral)<br>Activation = Linear<br>Time-distributed: Yes | Filters = 20, 40, 60<br>Kernel size = 7, 21, 63 (spectral)<br>Activation = ReLU<br>Dropout fraction = 0.2<br>Time-distributed: Yes | Nodes = 50, 15<br>Activation = ReLU<br>Dropout fraction = 0.12 | Labels = 6<br>Activation = Softmax |
| Perm-3D-CNN-v1 | Input data size = 7 × 3 × 3 × 12(time, spatial, spatial, spectral)<br>Normalization = Batch Normalization | | Filters = 20, 30, 40<br>Kernel size = 3 × 3 × 7, 1 × 1 × 21, 1 × 1 × 63 (spatial, spatial, spectral)<br>Activation = ReLU<br>Dropout fraction = 0.21<br>Time-distributed: Yes | | |

are hypothesized for MSMT image-based crop classification that uses the concept of 'permuted spectral neighbourhood with localized spectral convolutions' are introduced.

#### 2.5.1. *Permuted spectral neighbourhood ensemble*

In order to increase the significance of spectral information in MSMT images-based crop classification, a *permutation layer* that generates permuted versions of the original spectral bandset and, then ensembles them in time-distributed fashion, is used.

**Table 7**
Architecture and hyperparameter settings of RNN model with permuted input for MSMT images-based crop classification.

| Model | Layer | | | | |
|---|---|---|---|---|---|
| | Input | Permutation layer (P1) | Recurrent layer (R1) | Fully connected layer (FC1, FC2) | Output layer |
| Perm-RNN-v1 | Input data size =7 × 12 (time, spectral)<br>Normalization = Batch Normalization | Permutated sets = 10<br>Kernel size = 12 (spectral)<br>Activation = Linear<br>Time-distributed: Yes | Filters = 10<br>Kernel size = 7(time)<br>Activation = Tanh<br>Dropout fraction = 0.12<br>Bidirectional (merge): Yes (concatenate)<br>Input-output format: Many-to-many | Nodes = 50, 15<br>Activation = ReLU<br>Dropout fraction = 0.12 | Labels = 6<br>Activation = Softmax |

**Table 8**
Architecture and hyperparameter settings of CRNN with permuted spectral convolution-based models for MSMT images-based crop classification.

| Model | Layer | | | | | |
|---|---|---|---|---|---|---|
| | Input | Permutation layer (P1) | Convolutional layers (C1, C2, C3) | Recurrent layer (R1) | Fully connected layer FC1, FC2 | Output layer |
| *Perm-1D-CRNN-v1* | *Input data size = 7 × 12(time, spectral)*<br>*Normalization: Batch Normalization* | *Permutated sets = 10*<br>*Kernel size = 12 (spectral)*<br>*Activation = Linear*<br>*Time-distributed: Yes* | *Filters = 20, 40, 60*<br>*Kernel size = 7, 21, 63 (spectral)*<br>*Activation = ReLU*<br>*Dropout fraction = 0.21*<br>*Time-distributed: Yes* | *Filters = 10*<br>*Kernel size = 7(time)*<br>*Activation = Tanh*<br>*Dropout fraction = 0.11*<br>*Bidirectional (merge): Yes (concatenate)*<br>*Input-output format: Many-to-many* | *Nodes = 50, 15*<br>*Activation = ReLU*<br>*Dropout fraction = 0.12* | *Labels = 6*<br>*Activation = Softmax* |
| *Perm-2D-CRNN-v1* | *Input data size = 7 × 3 × 3 × 12(time, spatial, spatial, spectral)*<br>*Normalization: Batch Normalization* | | *Filters = 32*<br>*Kernel size = 3 × 3(spatial, spatial)*<br>*Activation = ReLU*<br>*Dropout fraction = 0.21*<br>*Time-distributed: Yes* | *Filters = 10*<br>*Kernel size = 7(time)*<br>*Activation = Tanh*<br>*Dropout fraction = 0.11*<br>*Bidirectional (merge): Yes (concatenate)*<br>*Input-output format: Many-to-many* | | |
| Perm-3D-CRNN-v1 | Input data size = 7 × 3 × 3 × 12(time, spatial, spatial, spectral)<br>Normalization: Batch Normalization | | Filters = 20, 30, 40<br>Kernel size = 3 × 3 × 7, 1 × 1 × 21 , 1 × 1 × 63 (spatial, spatial, spectral)<br>Activation = ReLU<br>Dropout fraction = 0.21<br>Time-distributed: Yes | Filters = 10<br>Kernel size = 7(time)<br>Activation = Tanh<br>Dropout fraction = 0.11<br>Bidirectional (merge): Yes (concatenate)<br>Input-output format: Many-to-many | | |

**Table 9**
Model training and hyperparameter settings.

| | |
|---|---|
| Training and validation parameters | Loss function = Categorical cross entropy<br>Learning rate = 0.012<br>Optimizer = Adam (Kingma and Ba, 2014)<br>10-fold cross validation strategy is employed during training.<br>Training-validation/Testing data split = 80 %/20 % |

This concept can be perceived more clearly from Fig. 4. Let us consider a multi-band image with X rows, Y columns, and Z bands as shown in figure Fig. 4**(a)**. During conventional two-dimensional convolutions, each filter convolves with entire spectral bandset at any given time. Figure Fig. 4**(b)** depicts this convolution process with a filter highlighted in yellow. This strategy induces bias based on scale of a particular feature or class proportion in the image (Phartiyal et al., 2020). To supress this issue, localized convolutions in the spectral dimension are proposed similar to as shown in Fig. 4**(c)**. The filters convolve locally with only a subset of the entire bandset. This resolves the original issue however in turn induces the concern of spectral neighbourhood. It means that all filters convolve with bands of a fixed length but the bands convolved are always contiguous. This limits the possibilities of plausible feature sub-spaces. Reordering of bands and further convolutions (see Fig. 4**(d)**) can address this issue to an extent. With time-series/multitemporal image data, this strategy can be realized in a time-distributed fashion as shown in figure Fig. 4**(e)**. This MSMT image data with time-distributed ensemble of permuted spectral band sets is used as input to all DNN models in this study. The new input is termed as 'ensemble' input.

### *2.5.2. Models exploiting permuted spectral neighbourhood strategy in MSMT image-based crop classification*

The output from the permutation layer is used as input to all DNN models proposed in this study.

a. CNN only models

***1D-CNNs:*** The first model uses only one-dimensional convolution layers. Since, the temporal depth remains unaltered in the permutation layer, one-dimensional convolution in the temporal dimension will yield similar results as standard 1D-CNN and is therefore insignificant to develop the model. However, with the increase in spectral depth, and permuted ordering of bands, localized spectral convolutions should yield unique features. Therefore, a series of localized spectral convolutions can be performed over the ensembled spectral bandset with the help of one-dimensional convolutional filters.

***2D-CNNs:*** Since the input MSMT image has one spectral, two spatial

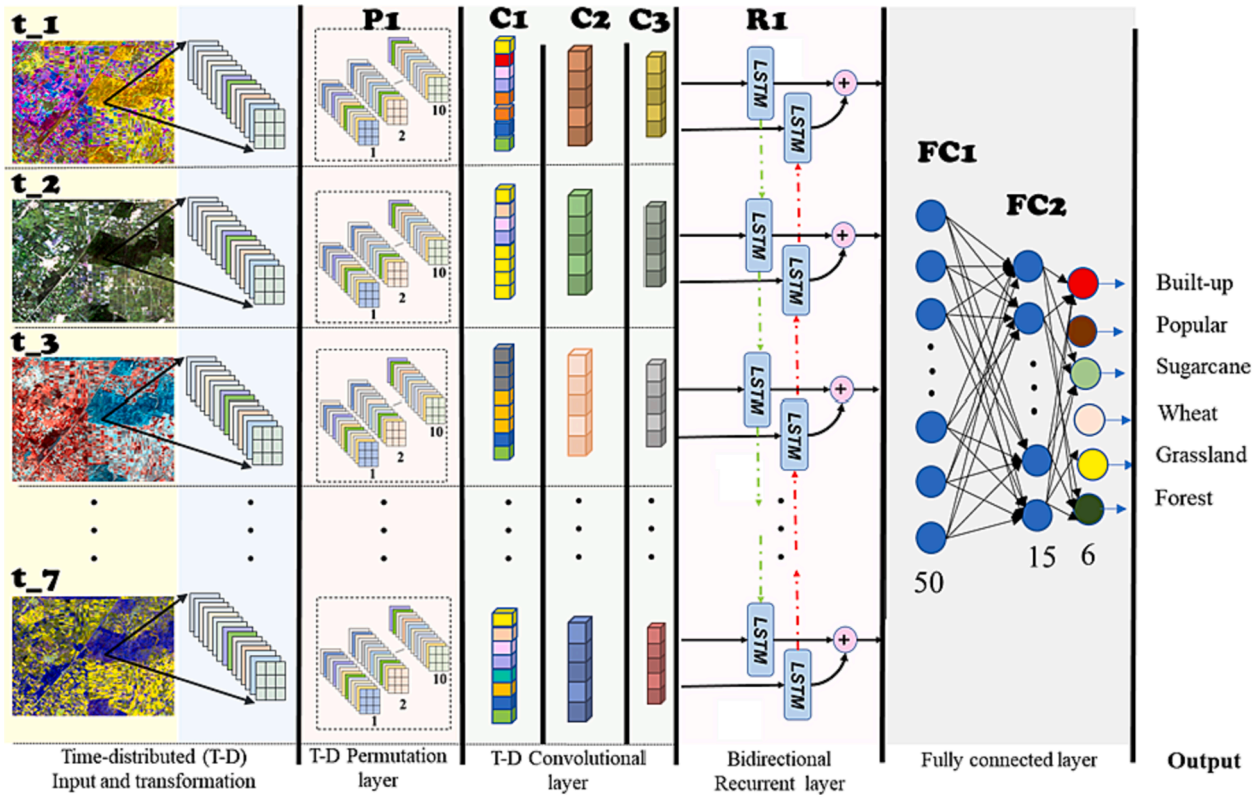



**Fig. 6.** Generic architecture an ensembled bandset with permuted spectral neighbourhood, time-distributed three-dimensional (spectral-spatial) convolution, and bidirectional recurrent layer-based DNN model. The horizontal dashed lines indicate time-distributed execution of pre-recurrent layer steps.

**Table 10**
Summary of the classification performance of the compared classifiers.

| Overall Accuracy OA (%) | | | | | |
|---|---|---|---|---|---|
| Standard | | | Permuted | | |
| Model | 10-fold cross validation | Testing | Model | 10-fold cross validation | Testing |
| 1D-CNN-v1 | 95.66 | 91.11 | Perm-1D-CNN-v1 | 99.70 | 94.44 |
| 3D-CNN-v1 | 97.45 | 93.88 | Perm-3D-CNN-v1 | 99.15 | 95.67 |
| RNN-v1 | 97.29 | 91.66 | Perm-RNN-v1 | 94.73 | 92.22 |
| 1D-CRNN-v1 | 89.70 | 91.66 | Perm-1D-CRNN-v1 | 96.85 | 94.99 |
| 2D-CRNN-v1 | 94.20 | 94.16 | Perm-2D-CRNN-v1 | 95.98 | 94.66 |
| 3D-CRNN-v1 | 96.00 | 95.55 | Perm-3D-CRNN-v1 | 99.10 | 97.77 |

and one temporal dimension, it is not possible for 2D-CNNs to process this 4-dimensional data simultaneously and directly. Therefore, the two-dimensional convolutions with localized spectral convolutions is not possible with 2D-CNNs.

***3D-CNNs:*** The more popular approach during utilization of 3D-CNNs in MSMT image applications is that the localized convolutions are employed along one temporal and two spatial dimensions whereas the spectral information is passed as *channels* (no localized convolutions in this dimension). In contrast, here, the 3D-CNN model can employ localized three-dimensional convolutions along one spectral and two spatial dimensions in a time distributed fashion. This strategy helps exploitation of local spectral information. The localization window size can be set different along the spatial and spectral dimensions in order to handle these two independently. In contrast to 1D-CNNs, this model gives smoother classification via suppression of noisy pixels. The temporal information is remains underutilized in this case.

b. RNN only models

RNN model used here is similar to the standard RNN model apart from the addition of a permutation layer that provides permuted spectral band set. Since, RNNs do not accommodate spatial-contextual information, per-pixel approach is considered. A simple LSTM unit can act as a recurrent cell. Several LSTM units constitute the recurrent layer and the architecture can be used in a many-to-many fashion. It means the input to the recurrent layer is a vector and output from the recurrent layer is a vector. Also, the flow of activation from each recurrent cell in the recurrent layer is considered in a bidirectional manner. It means that the activation from particular recurrent cell can move to next recurrent cell (next time stamp, usual) as well as to previous recurrent cell (previous time stamp). The activations from both (forward and backward) layers can be merged via strategies such as concatenation, averaging, etc. A graphical representation of the flow of activation in a recurrent layer with bidirectional LSTM recurrent cells is shown in Fig. 5(a) and a single LSTM unit architecture (Greff et al., 2017) is shown in Fig. 5(b). This is applicable and helpful in the MSMT images-based crop classification since the goal is improved classification performance and not forecasting (where future activation is not present). The bidirectional flow increases contribution of each recurrent sell (past or future) in classification. In this RNN model, the spectral information is passed to as channels and spatial context cannot be considered hence under-utilizing both information.

c. CNN-RNN or CRNN models

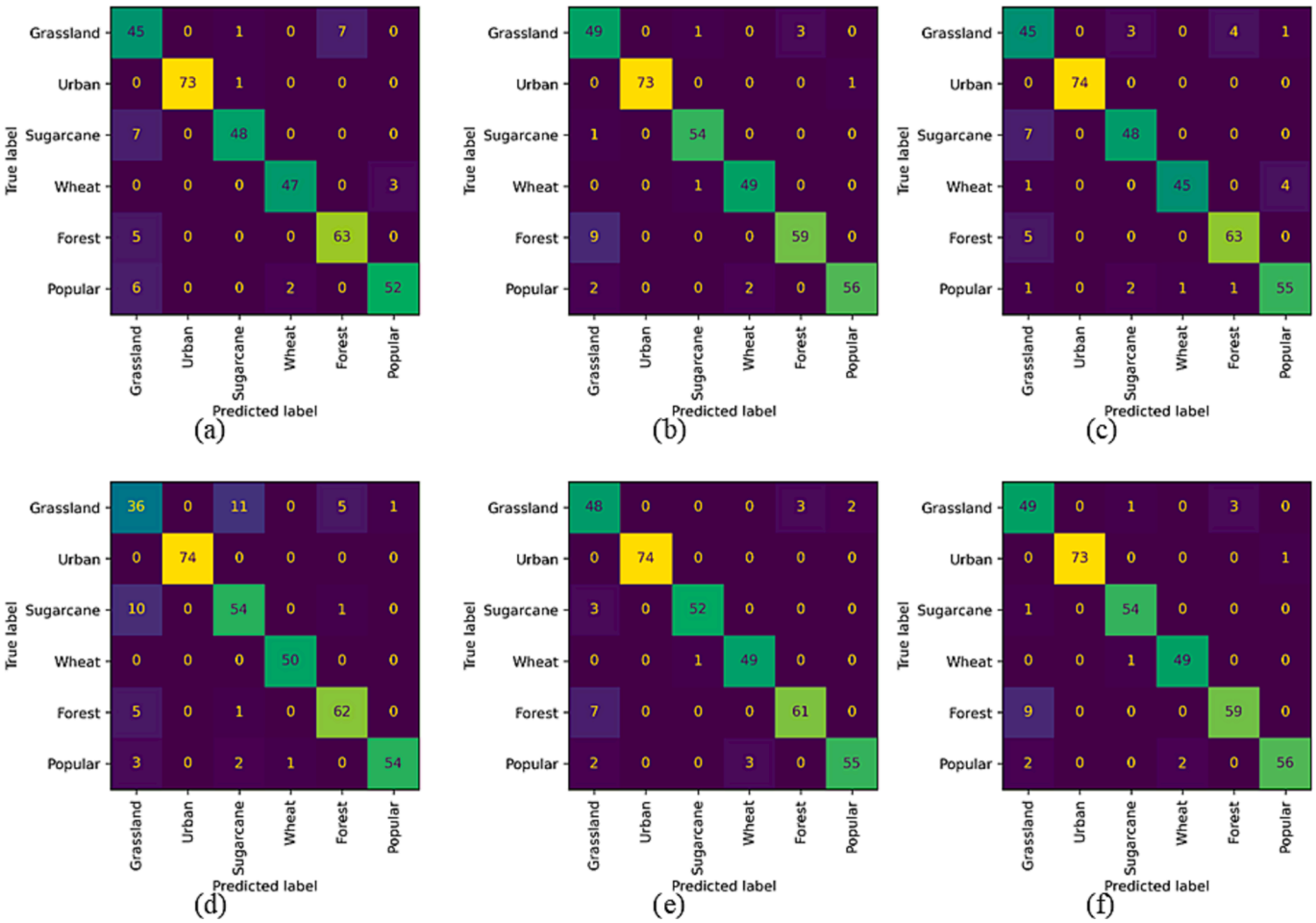



**Fig. 7.** Confusion matrices. (a) 1d-CNN-v1, (b) 3D-CNN-v1, (c) RNN-v1, (d) 1D-CRNN-v1, (e) 2D-CRNN-v1, (f) 3D-CRNN-v1.

Here, CNN and RNN architectures are used together to form CNN-RNN or CRNN models. It is clear from the previous discussions that the standalone CNN and RNN models are not able to efficiently exploit all three (spectral, spatial, and temporal) information simultaneously. At best, the standalone 3D-CNN model is able to exploit spectral-spatial information simultaneously but not all three. Alternatively, the standalone RNN model benefit from the temporal but is unable to extract localized spatial and spectral features. Therefore, an integrated utilization of CNN and RNN models is straightforward beneficial. In a generic CRNN model framework, CNNs are used to extract spectral and/or spatial information and RNNs are used to extract temporal information. The CNNs are used prior to RNNs in a time-distributed fashion to preserve the temporal information which is used by RNNs later. Three scenarios are possible.

***1D-CRNNs:*** In 1D-CRNN, one dimensional convolution layers are used in time-distributed fashion to extract local spectral features from MSMT images keeping the temporal information intact. Then, a bidirectional RNN layer can extract the temporal information. RNN layer characteristics are identical to that of standalone RNN model discussed earlier. Since, both the 1D-CNNs and RNNs cannot accommodate spatially-contextual information, this 1D-CRNN model is per-pixel based.

***2D-CRNNs:*** In 2D-CRNN model, two-dimensional convolutional layers are used which performs localized convolutions in the spatial domain with spectral information fed as channels. There are no localized spectral convolutions in this case. The convolutions are once again in time-distributed fashion. Then, a bidirectional RNN layer is used.

***3D-CRNNs:*** In 3D-CRNN, three-dimensional convolutional layers are opted to exploit the spectral and spatial information in time-distributed fashion and later RNN layer is opted to exploit the temporal information. With 3D-CNNs, localized convolutions are possible in the two spatial and one spectral dimension simultaneously whereas the bidirectional RNNs take care of the temporal information. Among others, this 3D-CRNN model strategy seems potentially equipped to handle MSMT imagery with enhanced spectral information.

In the above-mentioned CNN, RNN, and CRNN models, a fully connected layer followed by an output layer is used at the end. All these hypothesized models are developed and used for MSMT image-base crop classification in the next section along with few standard CNN and RNN models.

### *2.6. Model architectures, hyper-parameter settings, and training*

A total of 6 novel CNN, RNN, and CRNN models for MSMT image-based crop classification are developed here. The model architectures are listed in Table 6, Table 7, and Table 8 respectively. The developed models are namely; Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1. Generic hyper-parameter settings are listed in Table 9.

For the sake of understanding, a generic architecture explaining the Perm-3D-CRNN-v1 architecture (since it is extracting all three information) is shown in Fig. 6. The input data preparation, permuted spectral band set generation, and convolutions are employed in a time-distributed fashion (separated by dashes lines in Fig. 6). In order to determine the 'class' of a single pixel, a 3X3 spatial neighbourhood features are considered. Further, in order to obtain an ensemble of

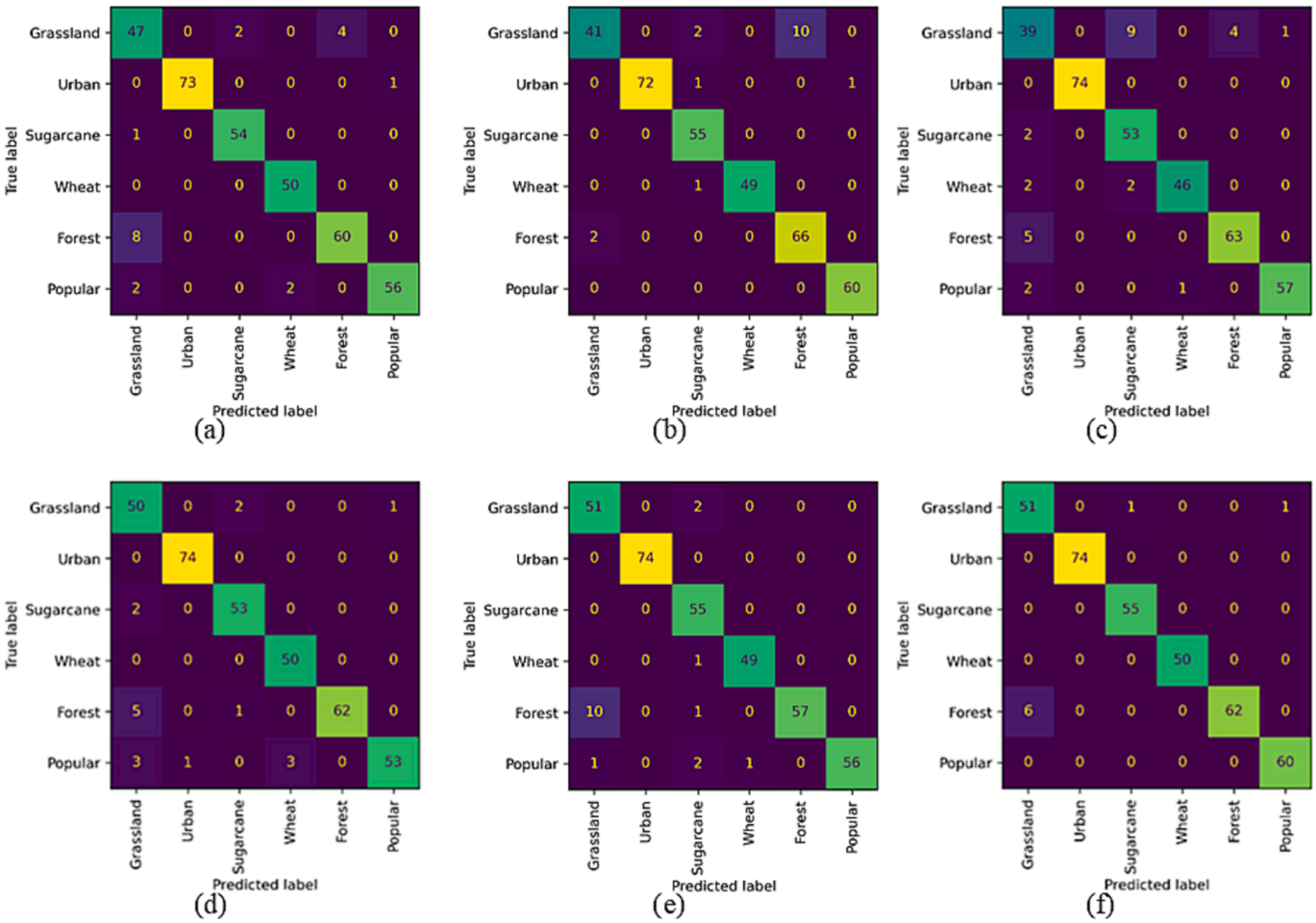



**Fig. 8.** Confusion matrices. (a) Perm-1D-CNN-v1, (b) Perm-3D-CNN-v1, (c) Perm-RNN-v1, (d) Perm-1D-CRNN-v1, (e) Perm-2D-CRNN-v1, and (f) Perm-3D-CRNN-v1.

permuted spectral band neighbourhood, 10 sets of spectral bands with permuted neighbourhood are ensembled. A three-layered convolution is then employed in time-distributed fashion. The first layer is a three-dimensional convolution with a 3X3X7 kernel where 3X3 represents localized spatial convolution and 7 indicates a localized spectral convolution. This results in a 1-dimensional feature vector. The next two convolutional layers help create more abstract features. The resulting features still contain temporal information which is then extracted with a bidirectional recurrent layer with 7 LSTM units (each for one time-stamp). The recurrent layer provides output in 'many-to-many' fashion resulting in 14 (7 from forward direction LSTM layer and 7 from backward direction LSTM layer) values per filter. Post the recurrent layer, the spectral-spatial–temporal information laced features are flattened and fed into a series of two fully-connected layers and then finally onto a 'SoftMax' layer where the pixel class is determined. Analogously, other models can be interpreted from Table 6, Table 7, and Table 8. All models are trained for 50 epochs with a 10-fold cross-validation strategy. A fully connected layer followed by an output layer is used in all models.

### 2.7. Evaluation

Confusion matrix, overall classification accuracy and F1-score are computed to evaluate the performances of the developed models. Confusion matrix is a matrix-type graphical representation of the distribution of samples over which a model is evaluated. the diagonal elements indicate correctly classified samples. whereas the off-diagonal elements indicate incorrectly classified samples. The overall accuracy (OA) is the ratio of the number of correctly classified samples to the number of the total samples. The F1-score is the harmonic mean of producer's accuracy and user's accuracy (Zhong et al., 2019). Further, qualitative evaluation is also carried out for a more generic out-of-sample performance.

## 3. Results and discussion

Classification results using the models developed in section 3 along with other standard classification models are discussed in this section. Both, qualitative and quantitative assessment are carried out. Overall, results from 12 different models are compared and discussed here. These are 1D-CNN-v1, 3D-CNN-v1, RNN-v1, 1D-CRNN-v1, 2D-CRNN-v1, and 3D-CRNN-v1 and their "permuted" counterparts i.e.; Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, Perm-3D-CRNN-v1.

### 3.1. Quantitative assessments

Table 10 shows quantitative assessment of the classification performance of these 12 models. In general, the permuted counterparts of the DNN models are showing better performance. Intuitively, this indicates to the significance and benefit of localized convolutions with permuted spectral neighbourhood over the global convolutions in the spectral dimension. Although, it is clear from Table 10 that the Perm-1D-CNN-v1 shows the best classification overall accuracy (OA) of 99.70 % on the training-validation set, but the Perm-3D-CRNN-v1 shows the best classification OA of 97.7 % on the test set.

Detailed assessment of the class-wise performance of these models is

**Table 11**
Classification performance quantitative assessment for standard models.

| Model | OA (test) | Class | Precision (a.k.a. Producer's accuracy) | Recall (a.k.a. User's accuracy) | F1-score |
|---|---|---|---|---|---|
| 1D-CNN-v1 | 91.11 | Grassland | 0.71 | 0.85 | 0.78 |
| | | Built-up | 1.00 | 0.99 | 0.99 |
| | | Sugarcane | 0.96 | 0.87 | 0.91 |
| | | Wheat | 0.96 | 0.94 | 0.95 |
| | | Forest | 0.90 | 0.93 | 0.91 |
| | | Popular | 0.95 | 0.87 | 0.90 |
| 3D-CNN-v1 | 93.88 | Grassland | 0.84 | 0.87 | 0.85 |
| | | Built-up | 1.00 | 0.99 | 0.99 |
| | | Sugarcane | 0.92 | 1.00 | 0.96 |
| | | Wheat | 0.96 | 0.96 | 0.96 |
| | | Forest | 0.93 | 0.91 | 0.92 |
| | | Popular | 0.98 | 0.90 | 0.94 |
| RNN-v1 | 91.66 | Grassland | 0.76 | 0.85 | 0.80 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.91 | 0.87 | 0.89 |
| | | Wheat | 0.98 | 0.90 | 0.94 |
| | | Forest | 0.93 | 0.93 | 0.93 |
| | | Popular | 0.92 | 0.92 | 0.92 |
| 1D-CRNN-v1 | 91.66 | Grassland | 0.82 | 0.68 | 0.74 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.79 | 0.98 | 0.88 |
| | | Wheat | 0.98 | 1.00 | 0.99 |
| | | Forest | 0.91 | 0.91 | 0.91 |
| | | Popular | 0.98 | 0.90 | 0.94 |
| 2D-CRNN-v1 | 94.16 | Grassland | 0.80 | 0.91 | 0.85 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.98 | 0.95 | 0.96 |
| | | Wheat | 0.94 | 0.98 | 0.96 |
| | | Forest | 0.95 | 0.90 | 0.92 |
| | | Popular | 0.96 | 0.92 | 0.94 |
| 3D-CRNN-v1 | 95.55 | Grassland | 0.88 | 0.87 | 0.88 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.93 | 0.98 | 0.96 |
| | | Wheat | 0.96 | 1.00 | 0.98 |
| | | Forest | 0.95 | 0.93 | 0.94 |
| | | Popular | 0.98 | 0.95 | 0.97 |

reflected in the confusion matrices provided in Fig. 7 and Fig. 8. Performance indicators such as user and producer accuracies are computed from corresponding confusion matrices and are tabulated in Table 11 and Table 12. The following major observations are made from Table 11 and Table 12.

1. Classes which were unchanged during the data observation period such as Built-up, are correctly classified with very good user and producer accuracies (~99 % for both) by all the models.
2. The two major crops i.e. Wheat and Sugarcane are best classified with Perm-3D-CRNN-v1 achieving 100 % UA and 100 % PA for Wheat and 100 % UA and 98 % PA for Sugarcane. Both Wheat and Sugarcane are better classified with CRNN models (standard and permuted variants) in contrast to CNN-only and RNN-only counterparts. Especially for Sugarcane, the variation between UA and PA is high for CNN-only and RNN-only models reflecting inconsistency. This can be observed from F1-score of 0.89 with both RNN-v1 and Perm-RNN-v1. Alternatively, the Perm-3D-CNN-v1 provides a 100 % UA and a 93 % PA with a significant variation of 7 % among the two.
3. The most challenging class is the Grassland class. Almost all the models have performed the least on segregating this class with Perm-RNN-v1 performing the worst (F1-score of 0.76) and Perm-3D-CRNN-v1 performing the best (F1-score of 0.93). Either the spectral response or the life-span of Grassland matches with Sugarcane, Wheat, and Forest classes to an extent. The misclassifications of Grassland samples with Wheat, Sugarcane, and Forest as shown in Fig. 7 and Fig. 8 are clear evidence of the same.
4. The impact of localized spatial convolutions on classification performance can be observed from Table 11 and Table 12. The 2D and 3D variants of CNN (in both, standard and permuted cases) are improving class-wise accuracies. For example, the average F1-scores (averaged over all classes) for Perm-1D-CRNN-v1 is 0.945 whereas the average F1-scores for Perm-3D-CRNN-v1 is 0.98 with increase of approx. 4 %.

It is important to note here that the reported accuracies are on training-validation/testing samples which is although out-of-sample accuracy however still is "sample" accuracy and not "population" accuracy. This is one of the main reasons for including a detailed "qualitative assessment" section in the article. Evaluation of models' performance in these study areas must include qualitative assessments. The next section provides such an assessment.

### 3.2. *Qualitative assessments*

The overall classification results from all the considered classifiers is presented in Fig. 9. Overall, wheat and sugarcane are present in major proportions. Forest also cover a significant proportion whereas popular,

**Table 12**
Classification performance quantitative assessment for proposed novel models.

| Model | OA | Class | Precision (a.k.a. Producer's) | Recall (a.k.a. User's) | F1-score |
|---|---|---|---|---|---|
| Perm-1D-CNN-v1 | 94.44 | Grassland | 0.81 | 0.89 | 0.85 |
| | | Built-up | 1.00 | 0.99 | 0.99 |
| | | Sugarcane | 0.96 | 0.98 | 0.97 |
| | | Wheat | 0.96 | 1.00 | 0.98 |
| | | Forest | 0.94 | 0.88 | 0.91 |
| | | Popular | 0.98 | 0.93 | 0.96 |
| Perm-3D-CNN-v1 | 95.13 | Grassland | 0.95 | 0.77 | 0.85 |
| | | Built-up | 1.00 | 0.97 | 0.99 |
| | | Sugarcane | 0.93 | 1.00 | 0.96 |
| | | Wheat | 1.00 | 0.98 | 0.99 |
| | | Forest | 0.87 | 0.97 | 0.92 |
| | | Popular | 0.98 | 1.00 | 0.99 |
| Perm-RNN-v1 | 92.22 | Grassland | 0.78 | 0.74 | 0.76 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.83 | 0.96 | 0.89 |
| | | Wheat | 0.98 | 0.92 | 0.95 |
| | | Forest | 0.94 | 0.93 | 0.93 |
| | | Popular | 0.98 | 0.95 | 0.97 |
| Perm-1D-CRNN-v1 | 94.99 | Grassland | 0.83 | 0.94 | 0.88 |
| | | Built-up | 0.99 | 1.00 | 0.99 |
| | | Sugarcane | 0.95 | 0.96 | 0.95 |
| | | Wheat | 0.94 | 1.00 | 0.97 |
| | | Forest | 1.00 | 0.91 | 0.95 |
| | | Popular | 0.98 | 0.88 | 0.93 |
| Perm-2D-CRNN-v1 | 94.66 | Grassland | 0.95 | 0.77 | 0.85 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.94 | 0.96 | 0.95 |
| | | Wheat | 1.00 | 0.96 | 0.98 |
| | | Forest | 0.93 | 1.00 | 0.97 |
| | | Popular | 0.91 | 0.97 | 0.94 |
| Perm-3D-CRNN-v1 | 97.77 | Grassland | 0.89 | 0.96 | 0.93 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.98 | 1.00 | 0.99 |
| | | Wheat | 1.00 | 1.00 | 1.00 |
| | | Forest | 1.00 | 0.91 | 0.95 |
| | | Popular | 0.98 | 1.00 | 0.99 |

water, built-up, and grassland are present in smaller proportions. Overall qualitative assessment suggest that different models are showing significantly varied proportions of land cover classes. This may be due to the "length of observation" and incurred natural changes in that period. However, the study is focussed in crop information more-so than in any other class. Therefore, any model that shows the presence of crop class should be preferred. This is also in-line with the quantitative assessments that most models are showing classification performance in the same order. A deeper investigation is needed to understand the performance of the models.

Information collected in the form of subsets/polygons during the crop growing season are used as visual ground truth to evaluate the performance of these models in mixed class scenarios. The visual interpretation is also an evaluation of the generalization performance of the models. The subset areas selected for visual inspection are marked in Fig. 2 (highlighted with yellow rectangular boxes). These subset areas are termed as ground truth images (GTI, refer section 2.3).

GTI-1 is a good example of mixed class scenario. The land parcels (crop and non-crop) are quite small in area (less than 5000 m sq.). In Fig. 10, the cropland highlighted in yellow box in the ground truth image (GTI-1) is a sugarcane field and is classified as sugarcane by the Perm-3D-CRNN-v1. In contrast, the Perm-1D-CNN-v1 model focused more the spectral information due to the permuted spectral band ensemble than on the temporal information and hence, classified the sugarcane field on the basis of a particular time stamp and not on the basis of the entire temporal profile. This analysis also indicates to the overfitting nature of convolutional only models in case of MSMT images-based crop classification. The efficient exploitation of the spectral, spatial, and temporal information present in the MSMT images by the Perm-3D-CRNN-v1 is also evidenced from the correct classification of popular plantations in the subset image (highlighted in red box). On the other hand, Perm-1D-CNN-v1misclassifies it as built-up.

The GTI in Fig. 11 (GTI-2) shows crop fields with lots of tree lines in between (highlighted in red lines) and forests. Once again, an efficient exploitation of the spectral, spatial, and temporal information in the MSMT images is done by the Perm-3D-CRNN-v1 model as it retains the spectral and structural identities of tree line present in between crop field while suppressing the temporary fallow lands spectral attributes at the same time.

In summary, the detailed visual interpretations highlight the potentially overfitting nature of the otherwise quantitatively the best model i.e. Perm-1D-CNN-v1 in case of MSMT image-based crop classification. Also, these interpretations highlight the good generalization ability (out-of-sample performance) of the Perm-3D-CRNN-v1 model which is an important attribute of any classifier. It is clear from these interpretations that the Perm-3D-CRNN-v1 has proved effective and reliable in crop classification in mixed land cover scenarios. Hence, it is safe to assume that the Perm-3D-CRNN-v1 model is the best classifier among all the developed and compared models.

The quantitative results look satisfactorily good for all models, but qualitative analysis reflects insights about mixed pixels and deviates slightly from the former. This is may be due to the following reasons.

- The reported accuracy is on cross-validation/testing samples which is although out-of-sample accuracy however still is "sample" accuracy and not "population" accuracy. This is one of the main reasons for including a detailed qualitative assessment. Evaluation of models' performance in these study areas must include such assessments.
- Presence of small land parcels/portion of a particular class within other class of larger proportions. For example, popular plantations are practiced in the region but in small land parcels (say 60 m by 100 m parcel). These plantations are surrounded by land parcels where "Wheat" is practiced (land parcel is still small but number of such parcels is very high). The models classify this small popular plantation parcel correctly as "Popular" class. This small popular class segment may be interpreted as wrong/incorrect classification because of a large parcel surrounding it is classified as "Wheat". This situation is common in these study regions as farmers have small land parcels and are independent to grow anything. Similar situations may be imagined with other classes. Also, averaging the post-classification result might deviate the crop-acreage estimates and in-turn crop yield estimates.

### *3.3. Comparison with recent models in MSMT image-based crop classification.*

Not many studies are present that have performed crop classification keeping the following three conditions in mind

- On study areas having mixed land cover
- with multisensor and multitemporal RS data
- Use of state-of-the-art DNNs.

Many of the similar studies are either directly employing traditional DNNs onto the time-series/multitemporal RS data or are a minor variant of traditional ones (De MacEdo et al., 2020; Qu et al., 2020; Yaramasu et al., 2020; Zhao et al., 2019; Zhou et al., 2019). We have already compared the proposed models with few of the traditional models in our study (refer section 3.1 and 3.2) and hence are not considered here.

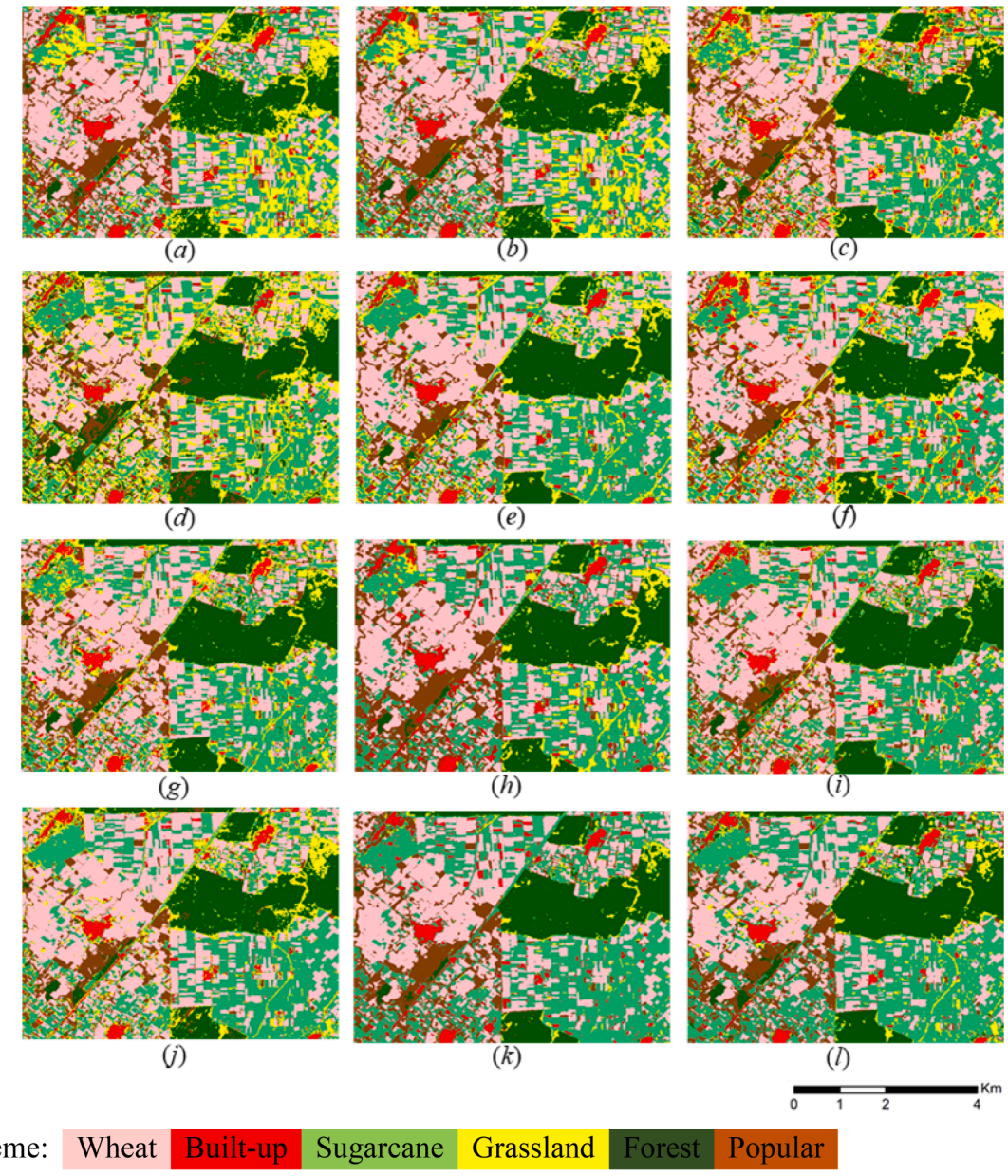



**Fig. 9.** Classified image; (a) 1D-CNN-v1, (b) 3D-CNN-v1, (c) RNN-v1, (d) 1D-CRNN-v1, (e) 2D-CRNN-v1, (f) 3D-CRNN-v1, (g) Perm-1D-CNN-v1, (h) Perm-3D-CNN-v1, (i) Perm-RNN-v1, (j) Perm-1D-CRNN-v1, (k) Perm-2D-CRNN-v1, and (l) Perm-3D-CRNN-v1.

A careful screening based in an 'AND' operation between the keywords; *crop classification, CNNs, RNNs, SAR, optical, and time-series/multitemporal*, the following recent studies are found best suited for comparison.

a. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification: Authors in (Teimouri et al., 2022) used multispectral-SAR (Sentinel2-Sentinel1) multimodal time-series data. They considered R, G, B, and NIR from Sentinel-2 and VV, and VH from Sentiel-1. They prepared time-series for each feature separately. 6 time-steps for Sentinel-2 and 7 time-steps from Sentinel-1 are considered. A 7-class classification is performed. Separate 3D-CNN-based architectures are employed to extract the features from each band/channel. Out of different spatial kernel sizes, they reported 7X7 as best and therefore is considered here. Also, the 3SI-3D-CNN variant (local convolution with a window size of 3 in temporal dimension) gave better result and therefore is considered here for comparison.
b. A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification (Chen et al., 2022): Authors reported a novel Im-BiLSTM model for crop classification with Sentinel-2 time-series data. Their model was able to address missing Sentinel-2 data from time-series. They employed an imputation strategy to figure out the missing values. The missing values are treated as learning parameters in a deep learning environment along with model weights. Since cloudy data is not considered in this study, preparing could mask and imputing missing data case is not covered. However, the standard BiLSTM applied on 7 time-stamp 12-feature time-series data by (Chen et al., 2022) is selected here for comparison with multisensor data.
c. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences (Chamorro Martinez et al., 2021): They employed various deep learning models for crop

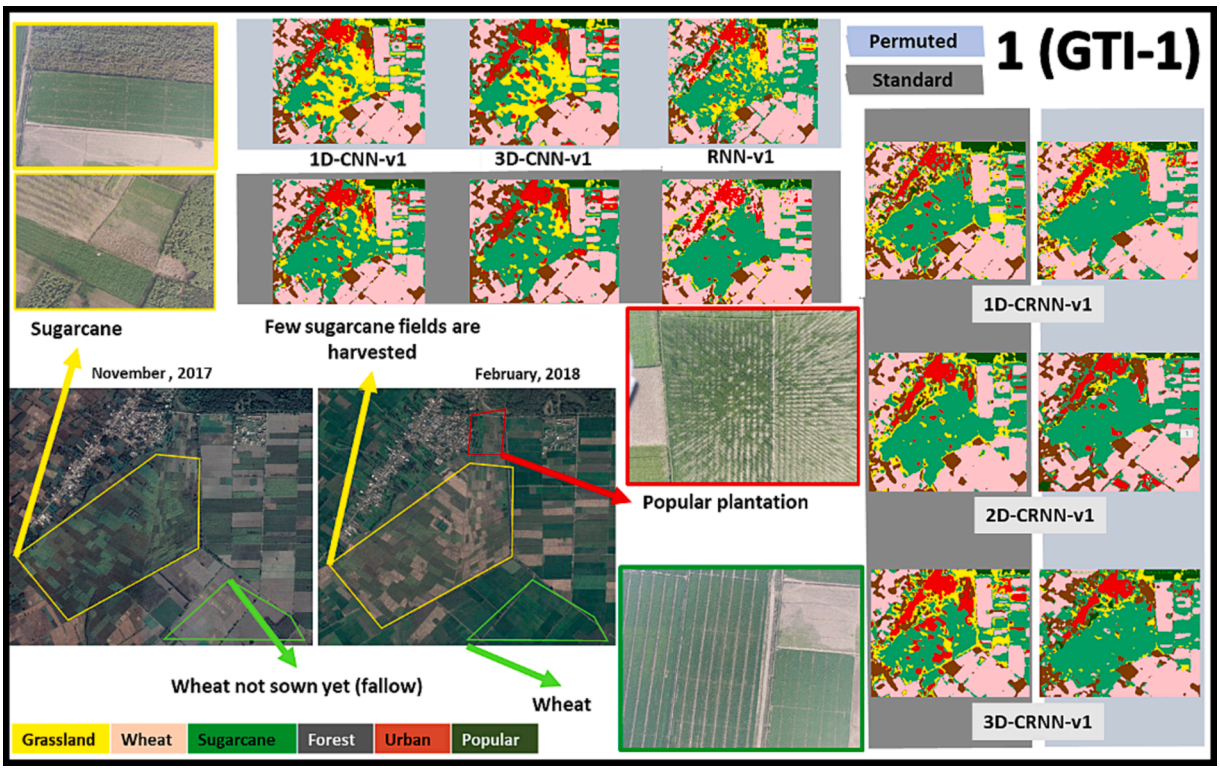



**Fig. 10.** Visual interpretation and evaluation of classification performance of the compared classification models for GTI-1. Permuted models: In blue background, Standard models: In green background. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

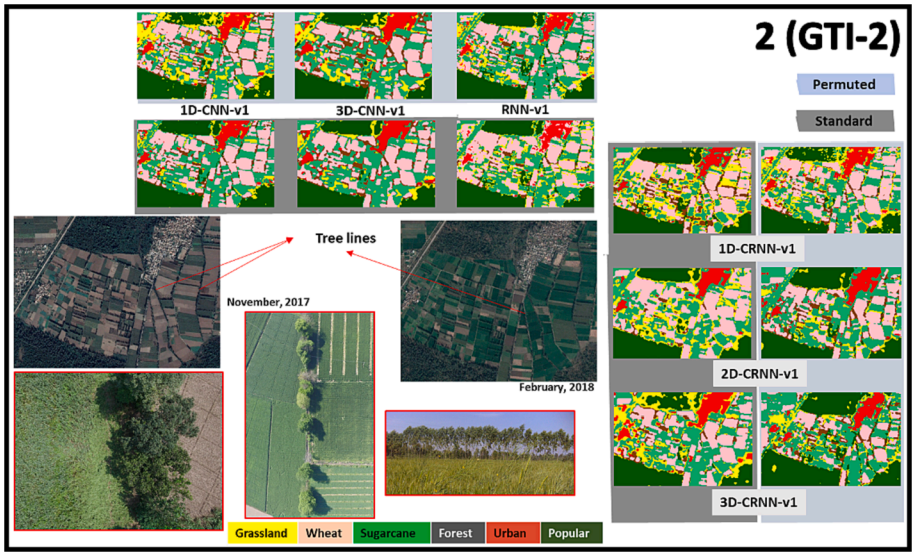



**Fig. 11.** Visual interpretation and evaluation of classification performance of the compared classification models for GTI-2. Permuted models: In blue background. Standard models: In green background. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

recognition with Sentinel-1 multi-temporal data. Their developed DNN models involved CNNs, RNNs, and semantic segmentation architectures (ex. U-net) for identifying 14 different crops. Most of these models are only applicable to input data with larger spatial size such as 432 taken by most of these models. These models cannot be realized here since the land parcels are small in the current study. The 'UConvLSTM' and the 'BConvLSTM' models are the only ones that can be realized here. Among these two, the BConvLSTM is employing a bidirectional LSTM layer first followed by a one-dimensional convolutional layer and is reported to have better performance (87 %) in different datasets (Chamorro Martinez et al., 2021). Therefore, the BConvLSTM is considered for comparison here with multisensor data

d. Sugarcane crop classification using time series analysis of optical and SAR Sentinel images: A deep learning approach (Sreedhar et al., 2022). They developed a two-layer LSTM model for sugarcane mapping in India. They used NDVI from Sentinel-2 and VH from Sentinel-1. We trained the same two-layer LSTM model with 12 features instead for comparison here.

**Table 13**
Quantitative assessment of the performance of the short-listed DNN models for time-series/multitemporal data-based crop classification.

| Model | OA | Class | Precision (a.k.a. Producer's) | Recall (a.k.a. User's) | F1-score |
|---|---|---|---|---|---|
| (Teimouri et al., 2022) | 93.33 | Grassland | 0.84 | 0.72 | 0.78 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.96 | 1.00 | 0.98 |
| | | Wheat | 0.96 | 1.00 | 0.98 |
| | | Forest | 0.83 | 0.96 | 0.89 |
| | | Popular | 1.00 | 0.90 | 0.95 |
| (Chen et al., 2022) | 91.67 | Grassland | 0.76 | 0.85 | 0.80 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.91 | 0.87 | 0.89 |
| | | Wheat | 0.98 | 0.90 | 0.94 |
| | | Forest | 0.93 | 0.93 | 0.93 |
| | | Popular | 0.92 | 0.92 | 0.92 |
| (Chamorro Martinez et al., 2021) | 90.0 | Grassland | 0.70 | 0.83 | 0.76 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.87 | 0.73 | 0.79 |
| | | Wheat | 0.94 | 0.98 | 0.96 |
| | | Forest | 0.96 | 0.94 | 0.95 |
| | | Popular | 0.91 | 0.88 | 0.90 |
| (Sreedhar et al., 2022) | 93.66 | Grassland | 0.86 | 0.72 | 0.78 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.98 | 0.98 | 0.98 |
| | | Wheat | 0.94 | 1.00 | 0.97 |
| | | Forest | 0.94 | 0.93 | 0.93 |
| | | Popular | 0.85 | 0.95 | 0.90 |
| (Paul et al., 2022) | 94.10 | Grassland | 0.78 | 0.87 | 0.82 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.92 | 0.87 | 0.90 |
| | | Wheat | 0.94 | 1.00 | 0.97 |
| | | Forest | 1.00 | 0.90 | 0.95 |
| | | Popular | 0.92 | 0.93 | 0.93 |
| Perm-3D-CRNN-v1 | 97.77 | Grassland | 0.89 | 0.96 | 0.93 |
| | | Built-up | 1.00 | 1.00 | 1.00 |
| | | Sugarcane | 0.98 | 1.00 | 0.99 |
| | | Wheat | 1.00 | 1.00 | 1.00 |
| | | Forest | 1.00 | 0.91 | 0.95 |
| | | Popular | 0.98 | 1.00 | 0.99 |

e. Generating pre-harvest crop maps by applying convolutional neural network on multi-temporal Sentinel-1 data (Paul et al., 2022). They used Sentinel-1 multi-temporal data for a 5-crop classification in the Indian agriculture landscape. They focused on pre-harvest classification however, they also reported performance with full temporal data. They employed a cascade of seven two-dimensional convolutional layers on a 7x7 patch with the temporal and polarimetric features stacked together. For example, if 7 time-stamps and 3 polarimetric features are used, the input is a 7x7x21 data cuboid. We employed the model variant utilizing the full-length multi-temporal data for comparison here. Also, multi-sensor data is used instead.

It is important to note here that though few of the models considered here for comparison were originally tested for single-sensor data, but here they are trained on multisensory data. Table 13 provides the summary of the quantitative assessment of the performance of the shortlisted models along with the best performing model proposed here. The model in (Paul et al., 2022) is providing the best OA of 94.10 % among the shortlisted candidates. However, the proposed Perm-3D-CRNN-v1 model has a 97.77 % OA and is comparatively better than the model in (Paul et al., 2022). This is because the proposed model is synergistically focusing on all three information i.e. spectral, spatial, and temporal whereas the model in (Paul et al., 2022) loses temporal information exploitation ability without RNNs though temporal information is extracted by 2D-CNNs. The proposed model also strengthens the impact of spectral information while keeping the significance of the spatial and temporal information intact. However, studies on diverse and multiple study areas would set a clear comparison.

**Table 14**
Comparison of classification performance of models utilizing RNNs with unidirectional and bidirectional strategies.

| Model | RNNs (OA %) | | | |
|---|---|---|---|---|
| | Unidirectional | | Bidirectional (B = backward, F = forward) | |
| | Last node (M2O) * 10 filters per node, 1 node | All nodes (M2M) ** 10 filters per node, 7 nodes | Last node (M2O) 10 filters per node,2 units (1B, and 1F) B and F concatenated | All nodes (M2M) 10 filters per node,14 units (7B, and 7F) B and F concatenated |
| Perm-RNN-v1 | 87.34 | 89.66 | 92.01 | 92.22 |
| Perm-1D-CRNN-v1 | 89.40 | 88.05 | 91.25 | 94.99 |
| Perm-2D-CRNN-v1 | 87.86 | 88.17 | 94.07 | 94.66 |
| Perm-3D-CRNN-v1 | 94.86 | 94.98 | 95.16 | 97.77 |

M2O: Many-to-One.
M2M: many-to-Many.

### 3.4. Unidirectional vs bidirectional RNNs

Few of the models in the study also employed RNNs in bidirectional fashion. It is important to establish the significance and advantage of utilizing bidirectional strategy over the unidirectional strategy in the case of MSMT image-based crop classification. A tabulated comparison of the performance of the models (with or without bidirectional strategy) is reported here. Four novel models utilizing RNNs i.e. Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1 are evaluated on both, unidirectional and bidirectional, strategies while keeping all other hyperparameters identical. Table 14 provides the overall classification accuracies over the test set. Alternatively, the models are evaluated on whether the output from the recurrent layer is taken only at the last node i.e. in many-to-one (M2O) architecture or, from each node i.e. in many-to-many (M2M) architecture. Analysing Table 14, the following observations are made.

- The bidirectional strategy is more efficient than the unidirectional strategy in MSMT image-based crop classification since it exploits the full potential of the temporal information (past, and future at each node) contained in the MSMT images. This strategy may not hold true for forecasting applications due to absence of future temporal information however in this case the strategy holds beneficial.
- The M2M architecture shows slightly better effect over the M2O architecture in models with bidirectional strategy than when the strategy is unidirectional. Intuitively, it seems that with bidirectional and M2M, more features are possible that upon concatenation results more stable output. A deeper dive into features derived at this stage may reveal more insights. Readers can find this as scope to develop new ways (for example, a weighted approach) to combine the features from a bidirectional and M2M recurrent layer in contrast to the 'concatenation' strategy used here.

The extensive investigation of the classification of the proposed models covered over various sections reflect the superiority of the Perm-

3D-CRNN-v1 model as a crop classification model with MSMT images. Authors would like to highlight the potential of this model in other applications and encourage readers to utilize these models.

## 4. Conclusion

This paper presents novel, CNN and RNN models for crop classification with MSMT imagery in mixed land cover scenarios. The small-sized cropland parcels and diverse cropping pattern makes crop classification difficult in these scenarios. A synergistic exploitation of the spectral, spatial, and temporal information is therefore deemed necessary. Novel models are hypothesized focusing on a synergistically exploitation the spectral, spatial, and temporal information. Preferences given to temporal and spatial characteristics of MSMT image data by the standard DNN models during crop classification is observed and stated. The underutilization of spectral characteristics of MSMT images is highlighted. Therefore, at first, the novel models employ a unique strategy of permuting spectral band neighbourhood and generate permuted spectral handsets to increase the significance of spectral information during classification. Then, localized spectral convolutions are employed along with localized spatial convolutions in order to exploit spectral-spectral sub-spaces. This strategy provides localized spectral-spatial features that appears to be more significant features. Both, permuting the spectral band neighbourhood, and localized spectral-spatial convolutions are realized in a time-distributed fashion to retain the temporal information which is later exploited by recurrent layers. The bidirectional activation flow is used in the RNN layer which enables the impact of both, past and future 'crop status' on classification. This strategy again appears fruitful in crop classification with time-series images. Overall, 6 novel CNN and RNN models are proposed and developed during the study namely Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1.

Comprehensive analysis and evaluation of the performance of these novel models along with 6 standard CNN and RNN models in MSMT images-based crop classification is conducted with Sentine-1 and Sentinel-2 multisensor data collected over seven time-stamps. A 12-band, 7 time-stamp MSMT data is prepared for study. All the models are trained on areas having crop fields in a challenging i.e. mixed landscape. Wheat and sugarcane are the major crops practiced in the area however it also contains forests, built-up, grasslands, and deciduous plantations (Popular). The study area is a good representation of common crop classification scenarios particularly in India. The performance of these models in this study area is a good indicator of their generalization ability. 6 land cover classes with wheat and sugarcane as crops-of-concern are selected for classification. 7The classification performance is analysed both quantitatively and qualitatively. Quantitative analysis of the classification results indicates to superior performance of the Perm-3D-CRNN-v1, which provides 97.7 % classification accuracy over the test set. This performance is supported by the qualitative assessments via visual inspection. The local spectral-spatial information of land covers is efficiently captured via the time-distributed three-dimensional convolutions of the Perm-3D-CRNN-v1 and later the recurrent layer with bidirectional LSTMs captures the temporal information. This approach with the increased contribution of the spectral information via permuted and ensembled spectral bands helps in synergistic exploitation of the spectral, spatial, and temporal information for crop classification in mixed land cover scenarios. Other models are either focussing on any one or at-most two attributes simultaneously resulting in sub-optimal performances. Therefore, the Perm-3D-CRNN-v1 is the best model among all the classifiers. The analysis also indicated to the satisfactory performance of standard CNN, RNN and CRNN models in MSMT images-based crop classification whereas their counterparts with 'permuted spectral band sets' showed comparatively improved results. This study sets-up a baseline for further studies where the significance of spectral information can be further analysed and how its contribution can be increased in other ways. In future, high-dimensional spectral features can be transformed/reshaped to have more impact during classification. The proposed methodology and model architectures can work for any number of crop classification scenarios as far as sufficient training is provided. Even the methodology can be extended to other classification problems that includes high-dimensional remote sensing imagery.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References


Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R.G., Dupuy, S., 2018. 'M3 fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 4939–4949. https://doi.org/10.1109/JSTARS.2018.2876357.

National Remote Sensing Center, 2017. Agriculture [WWW Document]. Indian Sp. Res. Organ. URL https://nrsc.gov.in/Agriculture (accessed 10.24.18).

Chamorro Martinez, J.A., Cué La Rosa, L.E., Feitosa, R.Q., Sanches, I.D.A., Happ, P.N., 2021. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. ISPRS J. Photogramm. Remote Sens. 171, 188–201. https://doi.org/10.1016/j.isprsjprs.2020.11.007.

Chen, B., Zheng, H., Wang, L., Hellwich, O., Chen, C., Yang, L., Liu, T., Luo, G., Bao, A., Chen, X., 2022. A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification. Int. J. Appl. Earth Obs. Geoinf. 108 https://doi.org/10.1016/j.jag.2022.102762.

De MacEdo, M.M.G., Mattos, A.B., Oliveira, D.A.B., 2020. Generalization of Convolutional LSTM Models for Crop Area Estimation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 1134–1142. https://doi.org/10.1109/JSTARS.2020.2973602.

Di Mauro, N., Vergari, A., Basile, T.M.A., Ventola, F.G., Esposito, F., 2017. End-to-end learning of deep spatio-temporal representations for satellite image time series classification. CEUR Workshop Proc. 1972.

District Office of Economics and Statistics, 2022. Uttarakhand Statistical Report-District Haridwar [WWW Document]. Uttarakhand State Goverment. URL https://cdn.s3waas.gov.in/s33416a75f4cea9109507cacd8e2f2aefc/uploads/2022/08/2022080617.pdf (accessed 5.25.23).

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Chang. Biol. 21, 1980–1992. https://doi.org/10.1111/gcb.12838.

Gadiraju, K.K., Ramachandra, B., Chen, Z., Vatsavai, R.R., 2020. Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery, in. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 3234–3242. https://doi.org/10.1145/3394486.3403375.

Gavade, A.B., Rajpurohit, V.S., 2020. A hybrid optimization-based deep belief neural network for the classification of vegetation area in multi-spectral satellite image. Int. J. Knowledge-Based Intell. Eng. Syst. 24, 363–379.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J., 2017. LSTM: Search Space Odyssey. IEEE Trans. Neural Networks Learn. Syst. 28, 2222–2232.

Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land Cover Classification via Multioral Spatial Data by Recurrent Neural Networks. arXiv 14, 1685–1689.

Ji, S., Zhang, Z., Zhang, C., Wei, S., Lu, M., Duan, Y., 2020. Learning discriminative spatiotemporal features for precise crop classification from multi-temporal satellite images. Int. J. Remote Sens. 41, 3162–3174. https://doi.org/10.1080/01431161.2019.1699973.

Kumar, P., Prasad, R., Gupta, D.K., Mishra, V.N., Vishwakarma, A.K., Yadav, V.P., Bala, R., Choudhary, A., Avtar, R., 2018. Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data. Geocarto Int. 33, 942–956. https://doi.org/10.1080/10106049.2017.1316781.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I. ul H., Singha, K., Choudhury, S.B.,

Chetri, T., Malek, Ż., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturn, T., Karner, M., McCallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., Fritz, S., 2019. Estimating the global distribution of field size using crowdsourcing. Glob. Chang. Biol. 25, 174–186. https://doi.org/10.1111/gcb.14492.

Li, Z., Chen, G., Zhang, T., 2020b. A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 847–858. https://doi.org/10.1109/JSTARS.2020.2971763.

Li, H., Ghamisi, P., Rasti, B., Wu, Z., Shapiro, A., Schultz, M., Zipf, A., 2020a. A multi-sensor fusion framework based on coupled residual convolutional neural networks. Remote Sens. 12, 1–21. https://doi.org/10.3390/RS12122067.

Luo, C., Meng, S., Hu, X., Wang, X., Zhong, Y., 2020. Cropnet: Deep Spatial-Temporal-Spectral Feature Learning Network for Crop Classification from Time-Series Multi-Spectral Images. Int. Geosci. Remote Sens. Symp. 4187–4190 https://doi.org/10.1109/IGARSS39084.2020.9324097.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. In: Bruzzone, L. (Ed.), Image and Signal Processing for Remote Sensing XXIII. SPIE, pp. 37–48. https://doi.org/10.1117/12.2278218.

Maurya, A.K., Singh, D., Singh, K.P., 2018. Development of fusion approach for estimation of vegetation fraction cover with drone and sentinel-2 data. International Geoscience and Remote Sensing Symposium. 7448–7451. https://doi.org/10.1109/IGARSS.2018.8517613.

Maurya, A.K., Murugan, D., Singh, D., Singh, K.P., 2019. A Step for Digital Agriculture by Estimating Near Real Time Soil Moisture with Scatsat-1 Data. In: International Geoscience and Remote Sensing Symposium. IEEE, Yokohama, Japan, pp. 5698–5701. https://doi.org/10.1109/IGARSS.2019.8898433.

Ho Tong Minh, D., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P., 2018. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR sentinel-1. IEEE Geosci. Remote Sens. Lett. 15, 465–468. https://doi.org/10.1109/LGRS.2018.2794581.

Murugan, D., Singh, D., 2018. Development of an approach for monitoring sugarcane harvested and non-harvested conditions using time series Sentinel-1 data. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, Valencia, pp. 5308–5311.

Ndikumana, E., Minh, D.H.T., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue. France. Remote Sens. 10, 1217. https://doi.org/10.3390/rs10081217.

Paul, S., Kumari, M., Murthy, C.S., Kumar, D.N., 2022. Generating pre-harvest crop maps by applying convolutional neural network on multi-temporal Sentinel-1 data. Int. J. Remote Sens. 43, 6078–6101. https://doi.org/10.1080/01431161.2022.2030072.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 11 https://doi.org/10.3390/rs11050523.

Phartiyal, G.S., Bordu, N., Singh, D., Yahia, H., Daoudi, K., 2020. Permuted Spectral and Permuted Spectral-Spatial CNN Models for PolSAR-Multispectral Data based Land Cover Classification. Int. J. Remote Sens. 42, 1096–1120.

Phartiyal, G.S., Brodu, N., Singh, D., Yahia, H., 2018. A mixed spectral and spatial Convolutional Neural Network for Land Cover Classification using SAR and Optical data, in: EGU. Vienna, p. 12647.

Phartiyal, G.S., Singh, D., 2018. Comparative Study on Deep Neural Network Models for Crop Classification Using Time Series PolSAR and Optical Data. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 425, 675–681. https://doi.org/10.5194/isprs-archives-XLII-5-675-2018.

Pravash, P., 2019. Application of Space technoloy in Agriculture. Smart Agri Post 52.

Qu, Y., Zhao, W., Yuan, Z., Chen, J., 2020. Crop mapping from Sentinel-1 polarimetric time-series with a deep neural network. Remote Sens. 12 https://doi.org/10.3390/RS12152493.

Ray, S., Neetu, 2017. Crop Area Etimation with remote Sensning, in: Handbook on Global Strategy to Improve Agricultural and Rural Statistics. GSARS, Rome, pp. 151–261.

Ray, S.S., n.d. Remote Sensing Applications: Indian Experience.

Rubwurm, M., Korner, M., Rußwurm, M., Körner, M., 2017. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images, in: Computer Vision and Pattern Recognition Workshops. Honululu, HI, pp. 1496–1504. https://doi.org/10.1109/CVPRW.2017.193.

Scarpa, G., Member, S., Vitale, S., Member, S., Cozzolino, D., 2018. Target - Adaptive CNN - Based Pansharpening. IEEE Trans. Geosci. Remote Sens. 56, 5443–5457.

Seydi, S.T., Amani, M., Ghorbanian, A., 2022. A Dual Attention Convolutional Neural Network for Crop Classification Using Time-Series Sentinel-2 Imagery. Remote Sens. 14 https://doi.org/10.3390/rs14030498.

Sharma, V., Ghosh, S.K., 2023. Evaluating the Potential of 8 Band Planetscope Dataset for Crop Classification Using Random Forest and Gradient Tree Boosting by Google Earth Engine. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLVIII-M-1–2023, 325–330. https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-325-2023.

Sharma, A., Liu, X., Yang, X., 2018. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. Neural Networks 105, 346–355. https://doi.org/10.1016/j.neunet.2018.05.019.

Sreedhar, R., Varshney, A., Dhanya, M., 2022. Sugarcane crop classification using time series analysis of optical and SAR sentinel images: a deep learning approach. Remote Sens. Lett. 13, 812–821. https://doi.org/10.1080/2150704X.2022.2088254.

Tang, P., Du, P., Xia, J., Zhang, P., Zhang, W., 2022. Channel Attention-Based Temporal Convolutional Network for Satellite Image Time Series Classification. IEEE Geosci. Remote Sens. Lett. 19 https://doi.org/10.1109/LGRS.2021.3095505.

Tang, W., Long, G., Liu, L., Zhou, T., Jiang, J., Blumenstein, M., 2020. Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline.

Teimouri, M., Mokhtarzade, M., Baghdadi, N., Heipke, C., 2022. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification. Geocarto Int. 37, 15143–15160. https://doi.org/10.1080/10106049.2022.2095446.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J. D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. Remote Sens. Environ. 264, 112603 https://doi.org/10.1016/j.rse.2021.112603.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y., Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. Remote Sens. Environ. 264 https://doi.org/10.1016/j.rse.2021.112599.

Yaramasu, R., Bandaru, V., Pnvr, K., 2020. Pre-season crop type mapping using deep neural networks. Comput. Electron. Agric. 176 https://doi.org/10.1016/j.compag.2020.105664.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., Feng, M., 2019. Evaluation of three deep learning models for early crop classification using Sentinel-1A imagery time series-a case study in Zhanjiang. China. Remote Sens. 11 https://doi.org/10.3390/rs11222673.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. https://doi.org/10.1016/j.rse.2018.11.032.

Zhou, Y., Luo, J., Feng, L., Yang, Y., Chen, Y., Wu, W., 2019. Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data. Giscience Remote Sens. 56, 1170–1191. https://doi.org/10.1080/15481603.2019.1628412.