Enhanced crop classification through integrated optical and SAR data: a deep learning approach for multi-source image fusion

Niantang Liu, Qunshan Zhao, Richard Williams & Brian Barrett

Published online: 17 Jul 2023.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

Enhanced crop classification through integrated optical and SAR data: a deep learning approach for multi-source image fusion

Niantang Liu [a,b], Qunshan Zhao [c], Richard Williams [a] and Brian Barrett [a]

aSchool of Geographical and Earth Sciences, University of Glasgow, Glasgow, UK; bKey Laboratory of Agricultural Remote Sensing, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Ministry of Agriculture and Rural Affairs, Beijing, China; cUrban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow, UK

ABSTRACT

Agricultural crop mapping has advanced over the last decades due to improved approaches and the increased availability of image datasets at various spatial and temporal resolutions. Considering the spatial and temporal dynamics of different crops during a growing season, multi-temporal classification frameworks are well-suited for mapping crops at large scales. Addressing the challenges posed by imbalanced class distribution, our approach combines the strengths of different deep learning models in an ensemble learning framework, enabling more accurate and robust classification by capitalizing on their complementary capabilities. This research aims to enhance the crop classification of maize, soybean, and wheat in Bei'an County, Northeast China, by developing a novel deep learning architecture that combines a three-dimensional convolutional neural network (3D-CNN) with a variant of convolutional recurrent neural networks (ConvRNN). The proposed method integrates multi-temporal Sentinel-1 polarimetric features with Sentinel-2 surface reflectance data for multi-source fusion and achieves an overall accuracy of 91.7%, a Kappa coefficient of 85.7%, and F1 scores of 93.7%, 92.2%, and 90.9% for maize, soybean, and wheat, respectively. Our proposed model is also compared with alternative data augmentation techniques, maintaining the highest mean F1 score (87.7%). The best performer was weakly supervised with ten per cent of ground truth data collected in Bei'an in 2017 and used to produce an annual crop map for measuring the model's generalizability. The model learning reliability of the proposed method is interpreted through the visualization of model soft outputs and saliency maps.

1. Introduction

Crop mapping is essential for the assessment of the underlying factors for farming system changes and the management of crops. Northeast China has become one of

the main breadbaskets of the country, serving an increasingly important role in agricultural production and international trade of certain crops such as soybeans (Dong et al. 2016; Yang et al. 2019). Targeting the economic sustainability of agricultural development, however, the retrieval of quantitative information from the changes in the local croplands has been limited due to the annual crop rotation practice featured in this region (You et al. 2021). As such, accurate annual crop maps are still in high demand by local authorities in China to build near real-time crop monitoring mechanisms for early yield assessment of major crops at the county-level scale. Many studies have made considerable progress in the development of crop mapping systems by using satellite imagery with moderate spatial resolutions due to their coverage and regular repeat acquisitions (Boryan et al. 2011; Defourny et al. 2019; Inglada et al. 2015). Considering the spectral characteristics observed in commonly used optical satellite sensors such as Landsat, MODIS and Sentinel-2, many studies have investigated and quantified the dynamics (i.e. seasonal changes) of vegetation indices (VIs) and optical bands, using them as distinctive input features to accurately identify crop types throughout the growing seasons (Fan et al. 2014; Song et al. 2017; You and Dong 2020; Zheng et al. 2015; Zhong et al. 2016, 2016). In light of existing research in automated crop identification, our study seeks to develop a novel approach for enhancing crop mapping performance leveraging the potential of satellite remote sensing data, which can contribute towards addressing the pressing need for sustainable agricultural development in Northeast China.

Cloud cover and/or adverse weather conditions can limit the quality of optical acquisitions and impact upon crop monitoring capabilities, resulting in data loss within time-series of satellite acquisitions during the growing season (Griffiths, Nendel, and Hostert 2019; Kussul et al. 2018; Sonobe et al. 2014). Synthetic aperture radar (SAR) sensors are active remote sensors that can operate independently of weather conditions or solar illumination. SAR images provide unique radar-related information primarily responding to the biophysical properties of vegetation (e.g. Gao et al. 2018; Qu et al. 2020; Sun et al. 2019). Many studies have demonstrated the feasibility of using radar polarimetric features to detect crop types, generated by specific polarimetric decomposition algorithms that include Pauli, Cloude-Pottier, Freeman-Durden, H/A/α, Huynen, Yamaguchi Neumann, and Krogager (Gao et al. 2018; He et al. 2020; Liao et al. 2020; Xie et al. 2019). Both optical and radar data have respectively demonstrated their capability for crop mapping, and the fusion of image data from both data sources is increasingly explored to improve the crop mapping performance (e.g. Gao et al. 2018; Li et al. 2022; Liao et al. 2020; Moumni, Lahrouni, and Jung 2021; Sun et al. 2019; Van Tricht et al. 2018). The combination of optical and radar data provides complementary information that can reduce temporal gaps in data capture, which can contribute significantly to identifying crops in cloud-prone regions (Liao et al. 2020; Sun et al. 2019). Similarly, combining multi-sensor data yields richer information on certain crops to overcome the heterogeneity of some areas caused by mixed crops (Moumni, Lahrouni, and Jung 2021). Most previous studies have either stacked optical and radar data at the pixel level for crop classification (e.g. Gao et al. 2018; Liao et al. 2020; Moumni, Lahrouni, and Jung 2021; Van Tricht et al. 2018), or independently trained the image data from dual sources, e.g. Sentinel-1 and Sentinel-2, using separate models in parallel. Subsequently, the resultant outputs from each model are integrated into one learned feature sequence (Teimouri et al. 2022).

Previous studies employed machine learning models, such as Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) to identify crops based on multi-temporal observations (Bargiel 2017; Gao et al. 2018; Pelletier et al. 2016; Teluguntla et al. 2018; Zhong, Gong, and Biging 2014), however conventional machine learning models were not originally designed to process temporal data. Additionally, the enhanced representation of crop growth patterns requires phenological metrics defined with expertise in multi-temporal remote sensing data (You and Dong 2020), and those designed metrics are not always available until the end of the crop growth cycle (Xu et al. 2021). Although machine learning approaches improve classification performance with increasing dimensions of input variables and reduce the requirements for designating threshold-based classification rules, the temporal relationship in multi-temporal satellite data cannot be fully and automatically utilized. More recently, studies demonstrated that a series of deep learning networks could successfully explore the sequential relationships within time-series remote sensing data for crop classification (Crisóstomo de Castro Filho et al. 2020; Dou et al. 2021; Liao et al. 2020; Rußwurm and Körner 2020; Sun et al. 2020; Xu et al. 2020; Zhao et al. 2021; Zhong, Hu, and Zhou 2019). These deep neuron-based architectures include one-dimensional Convolutional Neural Networks (1D-CNNs), Long Short-Term Memory (LSTM), and variants or combinations of both architectures. Given that these architectures by the 1D-CNN or LSTM models are naturally fitted with extracting sequential dependencies within multi-temporal remote sensing data, these models generally outperform the nontemporal models such as RF in terms of classification performance for maize and soybean (Xu et al. 2020) and other crops (Liao et al. 2020; Rußwurm and Körner 2020; Zhong, Hu, and Zhou 2019). However, temporal models are not used for the extraction of spatial features from satellite imageries.

The spatial relationship, known as the spatial arrangement of the adjacent pixels represented by the data matrix in remote sensing images, is also a main consideration for crop classification with remote sensing data. Two-dimensional Convolutional Neural Networks (2D-CNNs) are used to extract multi-level spatial features from satellite data for crop classification (He et al. 2020; Kussul et al. 2017; Wei et al. 2019). A patch-based CNN architecture is designed for regional-level classification on medium-resolution satellite imagery by collecting a series of image patches as inputs instead of pixel-based samples used for machine learning models, 1D-CNNs or LSTM models (Sharma et al. 2017). 2D-CNNs only focus on the spatial dimension due to the multidimensional input (the image size and the channel-wise image bands), whereas the temporal dependencies are not considered. Therefore, three-dimensional Convolutional Neural Networks (3D-CNNs) are proposed for the extraction of spatio-temporal features from image data. Fewer studies have applied 3D-CNN-based architectures for crop classification (e.g. Adrian, Sagan, and Maimaitijiang 2021; Ji et al. 2018; Teimouri et al. 2022). Roy et al. (2019) showed that a hybrid 3D-2D CNN had an improved performance over using standalone 3D-CNN and 2D-CNN, respectively. Another approach to obtaining spatio-temporal features is Convolutional Recurrent Neural Networks (ConvRNNs), and the variants represented by different recurrent units have been used to identify a large number of crop classes in a hierarchical framework (Turkoglu et al. 2021). To the best of our knowledge, there is less research regarding the synergistic use of 3D-CNN, 2D-CNN, and convRNN architectures for crop classification.

Despite the findings in previous studies, annual crop mapping in Northeast China remains challenging due to the high intra-class variance and inter-class similarity of spectral qualities and phenology of crops in the region, which are influenced by varying climate conditions, geomorphic characteristics, and cropping systems (Wang, Azzari, and Lobell 2019). Additionally, regular and cloud-free time series acquisitions are often limited to agriculture monitoring at a large scale (Defourny et al. 2019). As a result, this study utilizes a small number of available optical acquisitions for large-scale crop mapping as supplementary sources for time series SAR data to develop models that enhance crop mapping accuracy. The study aims to develop a novel framework that combines 3D-CNN, 2D-CNN, and ConvRNN architectures for county-level crop mapping based on the fusion of multi-temporal optical and SAR images for Bei'an county in Northeast China in 2017 at a 10 m spatial resolution. This spatio-temporal model contributes to improved performance in identifying crops during the growing season and addressing imbalanced class distribution, which could lead to model bias towards majority classes. The resulting crop maps can be used for dynamic monitoring of interannual crop growth in the same area and provide annual crop inventory information for local authorities to evaluate land-use policies. In this study, the proposed model is assessed for crop mapping and juxtaposed with models presented in previous studies (Ji et al. 2018; Pelletier, Webb, and Petitjean 2019; Roy et al. 2019; Turkoglu et al. 2021). Additionally, the models are examined in relation to data augmentation techniques and evaluated across three randomly selected geographical locations. Subsequently, the optimally chosen model is employed to generate an annual crop map for Bei'an in 2017 through model inference.

2. Study area

Bei'an is a county located in the northeast part of Heilongjiang province in China (47°35′N ~ 48°33′N, 126°16′E ~ 127°53′E) (Figure 1). According to Bei'an Municipal People's
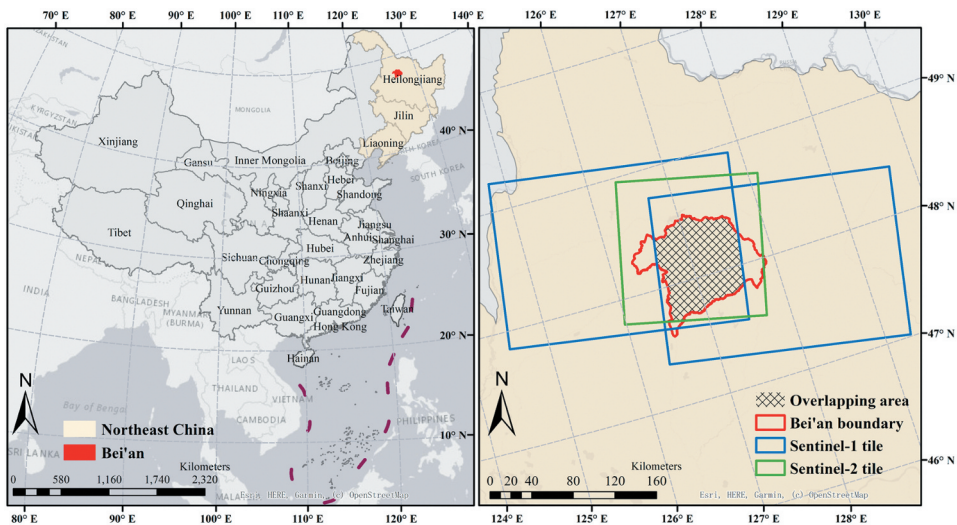

Figure 1. The study area in Bei'an. The multi-temporal Sentinel-1 and Sentinel-2 data are overlapped to capture the area that is covered by complete time-series acquisitions.

Government (http://www.hljba.gov.cn/), the total area of Bei'an county is approximately 7149 km^2. Bei'an is subject to a cold and temperate continental monsoon climate. The average annual temperature is around 1.2°C with annual effective accumulated temperature ranging from 18.30°C to 23.50°C. Bei'an receives an average annual precipitation of 529 mm, with the majority of rainfall occurring during the summer months from June to August. The average total amount of annual surface water resources are approximately 1.156 billion cubic metres. Bei'an is geographically located in the transitional zone between Songnen Plain and the Khingan Mountains, which is regarded as one of the world's three Chernozem (black soil) belts. Given the favourable soil fertility, meteorological conditions, and regional temperature, this region serves as an ideal ecological habitat conducive to crop growth and agricultural yield. According to Heihe Social and Economic Statistics Yearbook (2018), the total crop sown area of Bei'an approximates 2190 km^2. Summer maize and soybean are the primary crop types, accounting for 29.5% and 61.8% of the total sowing area, respectively. In contrast, wheat, as one of the minority crop types in Bei'an, covers 2.9% of the total sown area. According to the local crop sowing scheme, the growing season of maize often spans from late April to late September, and soybeans are normally sown from early May to mid-September. These periods might vary annually due to crop rotation cycles in the study area over the years.

3. Datasets

3.1. Sentinel-1/2 datasets and pre-processing

In this study, both time-series Sentinel-1B Single Look Complex products (Interferometric Wide swath SLC) and Sentinel-2A/B (Level-1C) image datasets were acquired from the Sentinel Scientific Data Hub (https://scihub.copernicus.eu/dhus/#/home). Considering the local cropping practice in which the majority of crops were planted and harvested from early May to late September 2017, the image acquisitions were collected from NaN Invalid Date to NaN Invalid Date , corresponding to the vegetative growing cycle of the recorded staple crops in Bei'an. As such, twenty-three Sentinel-1 acquisitions and three Sentinel-2 acquisitions were collected. The selection of Sentinel-2 data was based on the criteria that the average percentage of cloud coverage for the acquisition candidates is less than 8%.

The pre-processing of time series Sentinel-1 images was completed using the Sentinel Application Platform (SNAP) developed by the European Space Agency (ESA). The standard pre-processing steps follow Qu et al. (2020) which typically include radiometric calibration, multi-temporal speckle filtering (Refined Lee) and geocoding. Backscatter values were converted to decibel (dB) scale, and the cross-ratio of the backscatter was calculated by subtracting VV from VH, in accordance with logarithm rules. Sentinel-1 operates as an inherently dual-polarized SAR platform, which can constrain the extent of polarimetric information that can be explored, compared to quad-polarimetric SAR systems providing fully polarimetric observations. However, quad-polarization satellite acquisitions often suffer from reduced swath coverage, revisit time, and accessibility. Hence, a compact polarimetric technique, m-chi decomposition (Raney et al. 2012), has demonstrated its utility for crop mapping using dual-pol data (Sonobe 2019). The m-chi decomposition parameters were also obtained using SNAP. Each type of Sentinel-1 data was resampled to 10 m spatial resolution.

The pre-processing of Sentinel-2 images consists of the transformation from top-of-atmosphere (TOA) Sentinel-2 Level-1C reflectance images to bottom-of-atmosphere (BOA) Level-2A using Sen2Cor. Additionally, Band 4 (Red, 10 m), Band 8A (Vegetation Red Edge, 20 m), and Band 11 (SWIR, 20 m) were selected for their sensitivity to differentiate soybean and maize in Northeast China (You et al. 2021). All selected bands were resampled to 10 m and collocated with SAR data in a time-series sequence. Finally, a global min/max normalization approach was applied to all input features using the scikit-learn package to hasten the convergence of deep learning algorithms.

3.2. Ground truth and partitioning

Ground surveys of the study area were conducted during July, August, and September 2017 by the Chinese Academy of Agricultural Sciences (CAAS). During the 2017 period, 21257 fields were surveyed to calculate the area of crop parcels, record crop categories and retrieve annual statistics during the agricultural household survey. For cropland parcels with various crop types, they were manually digitized and labelled based on 5-metre resolution RapidEye satellite imagery (NIR Infra-red, Red Edge, and Red composite), while Sentinel-2A images (SWIR, Narrow NIR, and Red composite) were used for drawing relatively large cropland parcel areas with uniform crop types. In total, the classes of interest for major crops were assigned unique labels, including maize and soybean. In the in-situ dataset provided by CAAS, a small number of polygons were also identified for wheat and unknown crops. The proportions of ground sample pixels for each class in 2017 and the distribution of crop parcel size are displayed in Figure 2. A cropland mask layer, produced for Bei'an in 2017, is used to exclude non-cropland areas in this study during the model inference stage for the generation of an annual crop map. The cropland distribution and extent barely changed during 2017–2019 due to cropland protection by policies in Northeast China (Liu et al. 2014; Ning et al. 2018).

Since the cropping and managing system for each crop parcel would be different, the pixels within the same crop polygon are strongly correlated and need to be isolated when assigned to training, validation, and testing data sets. i.e. pixels in each set should be mutually exclusive and not from the same crop parcels. Additionally, the class distributions in all sets should be identical (Rußwurm and Korner 2017). In most croplands, pixels in the same parcel are very homogenous and highly correlated. Allocating pixels in a

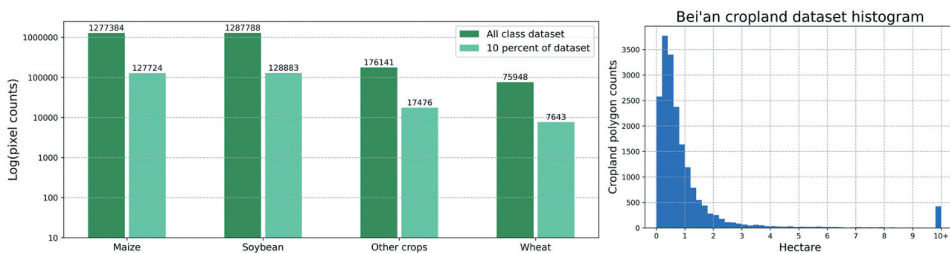**Figure 2.** The sample class distribution with the number of pixels (y-axis) at the logarithmic scale for the Bei'an dataset collected in 2017 (left), and 10% of the dataset is split into subsets for training, validation, and testing (left). The distribution of crop parcel size overall (right). The parcels large than 9 hectares are accumulated in the last bin in the histogram. The parcel size on average is 1.39 hectares.

parcel to different sets will violate the principle of independence. The model general-ization on truly unseen data would be affected because it is likely that models have seen at least parts of the image patches used for validation (Audebert, Le Saux, and Lefèvre 2019). Although the study area can be split into relatively large sub-regions, the crop types are not usually distributed evenly in the study area, which cannot ensure sub-regions with similar class distributions (Zhong, Hu, and Zhou 2019). In this study, each crop polygon was regarded as an entity. The parcels are further grouped using grids at 10-kilometre intervals so that crop parcels in the same grid are considered as a whole in dataset partitioning. Ten per cent of total ground samples are randomly selected to be used for model training, validation, and testing based on stratified sampling with the ratio of 60%, 20%, and 20%.

4. Methods

4.1. Methodology framework

The entire workflow in this study is depicted in Figure 3, outlining the four stages designed to evaluate deep learning approaches for crop mapping using the fusion of multi-source, time-series satellite data. The initial stage focuses on the pre-processing of multi-temporal satellite acquisitions, specifically targeting the overlapping area in Bei'an (Figure 1). The processed image data are then subdivided into small patches, aligning with the ground truth labels. These image patches then serve as input for the CNN models considered in this study. The experimental stage compares the performance of the proposed model with other state-of-the-art methods, given multiple model input scenarios. Particularly, an abla-tion experiment is conducted during model training and testing to determine the key input scenario for crop identification in Bei'an 2017. Following this, this study assesses the efficacy of implementing data augmentation techniques. The final stage involves generating a county-level crop map using the best performer and analysing the model learning out-comes. In the subsequent sections, the specifics of the experiment are introduced, present-ing aspects such as classification algorithms employed, the environment in which the
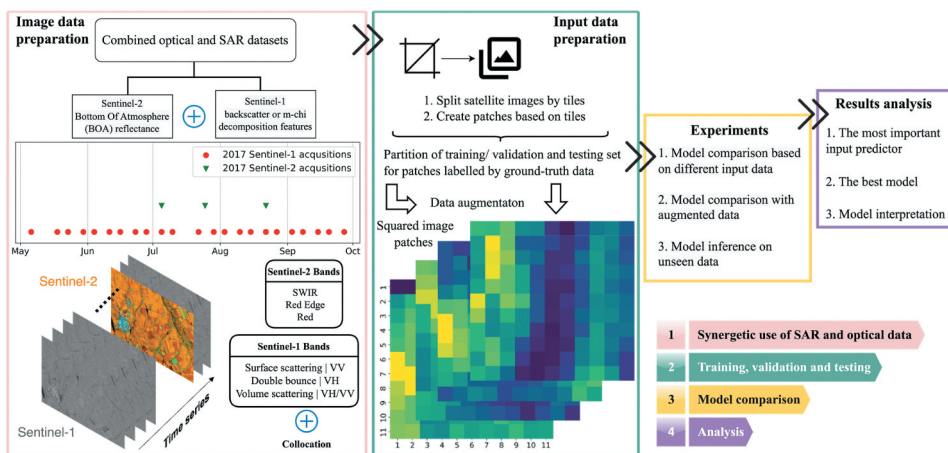

Figure 3. The overall workflow of the experiments.

models are deployed, compact polarimetric parameters, and augmentation techniques, to provide a comprehensive understanding of the methodology in this study.

4.2. Classification methods

4.2.1. 3D-CNN

Convolutional Neural Networks (CNNs), motivated by the animal's visual cortex is a deep learning technique to extract features in a way that considers spatial contexts between pixels instead of focusing on a single vector transformed by the typical multilayer neural networks (Sharma et al. 2017). CNNs, therefore, are also known as a dimensionality reduction method to handle multidimensional inputs in terms of its unique feature extraction pattern performed by convolutional kernels. Conventional two-dimensional CNNs (2D-CNN), however, are limited to the spatial features and may produce over-whelming parameters if the multidimensional inputs have large channels (spectral information) or time steps (temporal information) (Mäyrä et al. 2021). Conversely, the one-dimensional CNNs (1D-CNN) extract features from single-pixel temporal or spectral profiles of the input data without considering the spatial relationship between features. Although 2D-CNN can be combined with 1D-CNN to extract spatial-spectral or spatial-temporal information for improved results compared to dealing with information in only one dimension (Audebert, Le Saux, and Lefèvre 2019), the large number of model parameters will be needed by 2D-CNN. An alternative method to extract features simultaneously on both dimensions is three-dimensional CNNs (3D-CNN). The convolutional kernels in 3D-CNN are cubes and produce a feature map with volume rather than a two-dimensional image derived by 2D-CNN or a single vector by 1D-CNN. The three-dimensional convolving process can be written as follows:

$$Y_{i,j}^{x,y,d} = Relu\left(\sum_{n}\sum_{t=0}^{T}\sum_{p=0}^{S}\sum_{q=0}^{S} w_{(i-1),n}^{t,p,q} X_{(i-1),n}^{(x+p),(y+q),(d+t)} + b_{i,j}\right) \qquad (1)$$

Where $Y_{i,j}^{x,y,d}$ is the output values at 3D coordinate (x, y, d) on the jth feature cube in the ith layer, (x, y) is the spatial position and d indicates temporal index. $w_{(i-1),n}^{t,p,q}$ is the 3D kernel value at location (t, p, q) from the nth feature cube in the previous layer. Similarly, (p, q) is the spatial position, and t denotes the temporal indicator of the kernel. $X_{i-1,n}^{x,y,d}$ is the input at position (x, y, d) from the nth feature cube in the previous layer. $b_{i,j}$ is the bias vector on the jth feature cube in the ith layer. The size of the kernel is $T \times S \times S$ which is equivalent to length, height, and width. Empirically, the size of height equals width in CNNs.

4.2.2. ConvSTAR

Convolutional recurrent neural network (ConvRNN) is a variant of sequence modelling that is built with convolutional operations in state transitions instead of matrix multiplications for handling spatio-temporal data. A typical convolutional Long Short-Term Memory (ConvLSTM) is designed to capture the spatio-temporal correlations for precipitation forecast (Shi et al. 2015), which outperforms the general LSTM structure in which spatial information is not considered. Given that stacking multiple ConvRNN layers contributes to feature extraction, a novel recurrent cell, namely STAckable Recurrent cell (STAR), is developed to reduce the exploding gradient effects and the number of trainable

parameters (Turkoglu et al. 2021). The convolutional version of STAR (ConvSTAR) is the modification of ConvLSTM in which the input and output gate are removed, which can be written as:

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + B_f) \tag{2}$$

$$Z_t = tanh(W_{xz} * X_t + B_z) \tag{3}$$

$$H_t = \tanh(f_t \bullet Z_t + H_{t-1} \bullet (1 - f_t)) \tag{4}$$

Where σ is the sigmoid activation function, $*$ denotes the convolution operator and $\bullet$ indicates the Hadamard product (elementwise). The input X_t is firstly non-linearly projected through the activation function in Z_t. In addition, the previous state H_{t-1} and new inputs are linearly combined in the gating module which is the determinant of the state-to-state flow to create a new hidden state. W and B are matrices for weight and bias, respectively. The hidden state H_t is the output of a single ConvSTAR layer, which can be used for classification, or be used as the new inputs for the next layer or other decoders. Figure 4 illustrates a ConvSTAR cell that integrates Equation (2), Equation (3) and Equation (4). The code of the ConvSTAR layer was adapted in Tensorflow format from the Pytorch repository (https://github.com/0zgur0/STAckable-Recurrent-network.git).

4.2.3. Synergic use of 3D-CNN and ConvSTAR

Recent studies have applied the combination of 3D-CNN and convolutional recurrent networks for univariant and multivariate time series forecasting. The 3D-CNN layers and attention ConvLSTM layers are utilized sequentially for multispectral soybean prediction (Nejad et al. 2022). The 3D-CNN layer can also be fed with features produced by ConvLSTM, serving as the model output layer for predicting urban expansion in image segmentation (Boulila et al. 2021). The enhanced performance on human action
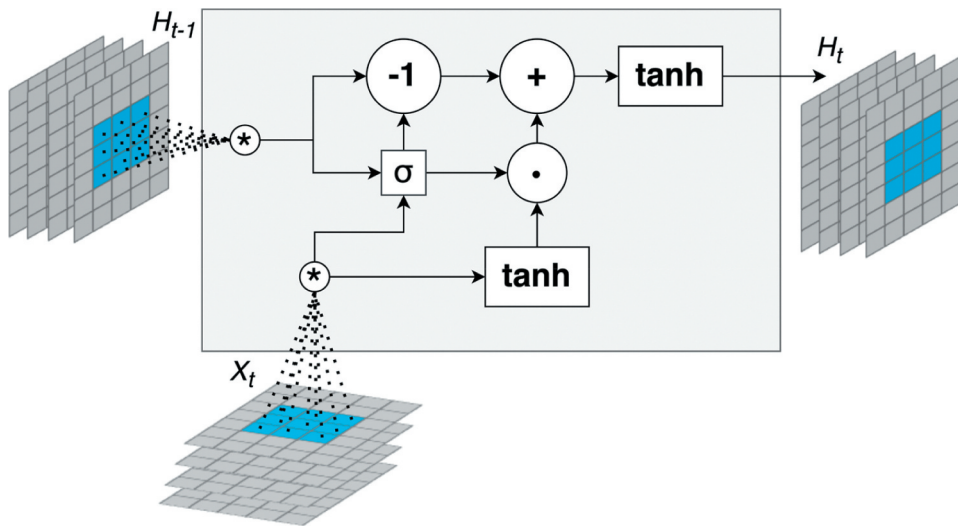
Figure 4. The structure of a ConvSTAR cell.

recognition was also proved by the combined use of these two deep learning approaches (Wang et al. 2021). In this study, a similar hybrid feature learning framework, 3D-ConvSTAR, is proposed to improve crop classification performance (Figure 5). The proposed network consists of three stages. The first step is made of a three-layer 3D-CNN with the optimal kernel size $3 \times 3 \times 3$, considering that three convolutional layers demonstrated effectiveness over two-layer and four-layer networks (Ji et al. 2018). Each layer has 32, 32, and 64 filters, respectively. The feature maps after each 3D convolutional operation are not shrunk by applying zero padding. The convolutional cubes are moved during one step. As previously introduced in Section 4.2.1, 3D-CNN is used to extract spatio-temporal features simultaneously. Next, the output tensors from the 3D-CNN module are reshaped

Figure 5. The architecture of 3D-ConvSTAR.

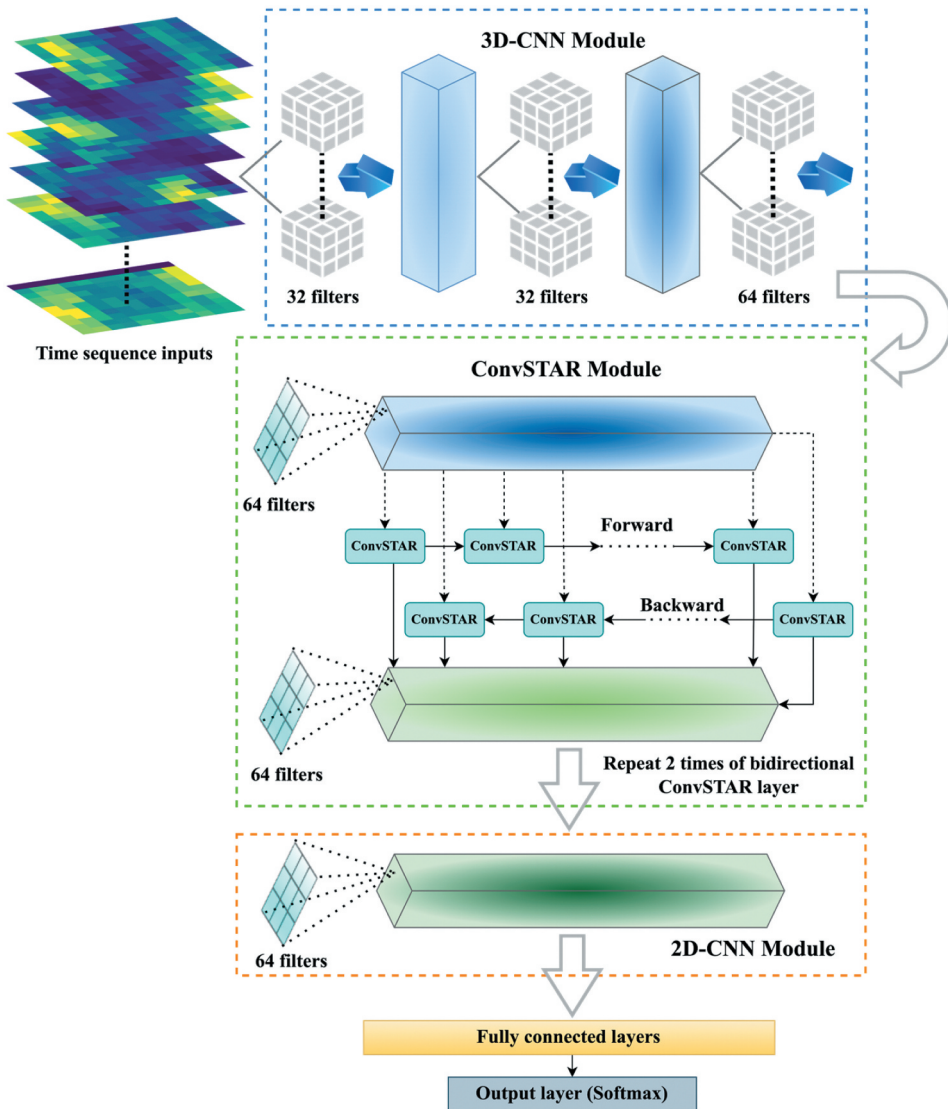and fed into a three-layer bidirectional ConvSTAR unit. Bidirectional recurrent cells preserve temporal information from both the future and past, alleviating temporal bias towards data in later time steps (Rußwurm and Körner 2018). The kernel size for ConvSTAR is 3×3 and the number of kernels for each layer is set to 64. Followed by ConvSTAR layers is a shallow 1-layer 2D-CNN with 64 3×3 kernels to take in the final hidden state from the previous layer for further extracting discriminative feature maps on the spatial dimension. It also can perform dimensionality reduction to some extent so that the number of model parameters can be optimized since it reduces the size of the feature maps and preserves the main information captured by the previous layers (Mäyrä et al. 2021; Roy et al. 2019). The last part of 3D-ConvSTAR is constructed with three fully connected (FC) layers. The last two dense layers have 256 and 128 units respectively and both are followed by a dropout layer with a factor of 0.4 to prevent networks from overfitting. The activation function Rectified Linear Unit (ReLU) (Nair and Hinton 2010) is applied after CNN, convolutional recurrent, and FC layers to augment the nonlinearity of outputs and control model systematic errors (bias). Pooling layers are not applied after all layers in the proposed model since it could cause loss of information at multiple dimensions (Li, Huang, and Ji 2019).

4.3. M-chi decomposition

A similar methodology was developed for single-transmitted dual-receive polarization data that transmits at circular polarization and receives at horizontal and vertical polarization (Raney et al. 2012). This decomposition methodology, originally for compact polarimetric radar data, is based on the 2×2 covariance matrix, which is not applicable to general quad-pol data. This method is often characterized by the form composed of four-element Stokes parameters, which provides potential for hybrid polarimetric and dual-pol data. One application is the m-chi decomposition, in which the observed field is characterized by the degree of polarization (m) as:

$$m = \frac{\sqrt{S_2^2 + S_3^2 + S_4^2}}{S_1}, \ 0 \le m \le 1 \tag{5}$$

Where $S_{1,2,3,4}$ represents four Stokes parameters in the total power over an image field, and m refers to the degree of polarization. $Chi(\chi)$, a Poincaré variable, denotes the field's ellipticity and circularity, which can be expressed through:

$$\sin 2\chi = -\frac{S_4}{mS_1} - 45° \le \chi \le +45° \tag{6}$$

Based on these two variables calculated from the Stokes parameters, three target scattering parameters P_s (Single-bounce scattering), P_d (Double-bounce scattering) and P_v (Volume scattering) for m-chi decomposition can be expressed as follows:

$$P_s = \sqrt{\frac{mS_1(1 - \sin 2\chi)}{2}} \tag{7}$$

$$P_d = \sqrt{\frac{mS_1(1 + \sin 2\chi)}{2}} \tag{8}$$

$$P_v = \sqrt{S_1(1 - m)} \tag{9}$$

Previous studies have shown that hybrid polarimetric data are close to and occasionally equivalent to the analysis of quad-pol data e.g. using the conventional Freeman-Durden decomposition (Ainsworth, Kelly, and Lee 2009; Nord et al. 2008). Even though all required four-element Stokes parameters can be derived from dual-polarized data such as Sentinel-1 SLC products, the challenge imposed on this usage is the separability between single and double bounce targets due to the ellipticity angle from dual-pol data on the verge of zero which could decrease the difference between polarimetric scattering types. Despite this challenge, one scattering type between the mix of all scattering mechanisms often dominates agricultural targets at a certain growth stage, and the scattering type of dominance could change with the development of the canopy (McNairn et al. 2009). Therefore, all scattering types of m-chi decomposition are considered in this study as input predictors to classify crops across full growth stages.

4.4. Model implementation

Considering the model inputs, we adopted the typical remote sensing scene classification approach and extracted square image patches centred around a labelled crop category. Square patches for each crop class are generated with the size of 9×9, 11×11, 15×15 and 21×21. Larger image patches will cover multiple polygons from different crop species but could lead to greater model parameters. Therefore, the optimal size of image patches is selected to 11×11 after examinations for classification purposes. Therefore, the complete shape of input image patches for the proposed model is $11 \times 11 \times 26 \times 3$ (width $\times$ height $\times$ sequence $\times$ channel). Patches without any ground truth information and missing values of pixels were eliminated.

The implementations of models are customized to train the network with reference to the Adam optimizer (Kingma and Ba 2014), and cross-entropy loss function (Botev et al. 2013). The learning rate for the optimizer is set to 0.001, and weight decay is regularized at 0.0001. A batch size of 128 is used during the training stage, and the model is saved with the best validation accuracy for later model inference. Another regularization strategy, namely Early stopping, is used to prevent the models from overfitting since it will terminate the model training progress once the model validation performance is relatively stabilized during the iterative training process.

The proposed 3D-ConvSTAR is compared with deep learning architectures applied in other studies: a 1D-CNN-based architecture, namely temporal CNN (TCNN) (Pelletier, Webb, and Petitjean 2019), a typical 3D-CNN (Ji et al. 2018), a hybrid 3D-2D CNN (Roy et al. 2019) and a 3-layer ConvSTAR (Turkoglu et al. 2021). This study reproduced the implementations of the models in the aforementioned studies. The modelling environment is implemented in Python 3.7.15 with Tensorflow backend (2.5.0) and Keras library (2.1.1) for model construction and generalization under two graphic devices of NVIDIA Quadro P4000 (8 GB RAM per GPU), and two processors of Intel (R) Xeon (R) Silver 4114 CPU (2.20 GHz/2.19 GHz).

4.5. Data augmentation

Imbalanced class distribution in real-world label datasets often leads to bias in supervised classification approaches, where machine learning models, under typical model training schemes, are prone to weigh importance in favour of majority classes (Dong, Gong, and Zhu 2018; Ren et al. 2018). Underrepresented crop classes in certain places may have the same or even higher value (either financial or ecological) as the crops that occur more frequently (Turkoglu et al. 2021). The annual statistics record that the total sown area for wheat is 6,309 hectares compared to 64,564 and 135,401 hectares for maize and soybean in Bei'an 2017 (Heihe Social and Economic Statistics Yearbook 2018). Consequently, class-balanced model evaluation is crucial for agricultural mapping applications, as fine-structured agricultural systems in local areas could lead to crop classes with the imbalanced distribution.

One approach to address this issue is typically called inverse frequency weighting (Cui et al. 2019). The training samples of minority classes are assigned higher weighting factors than majority classes for calculating training loss, which neutralizes the model bias towards majority classes. The underlying side effect of weighting samples could weaken model performance on majority classes to which lower weights are given. Another approach to balance the dataset is either via oversampling minority classes (Ling and Sheng 2008) or undersampling majority classes (He and Garcia 2009). Both techniques depend on a trade-off for the number of samples across all classes. Missing data for dominant classes due to undersampling could severely affect model performance, considering that deep learning and machine learning methods are data-driven and data-hungry. Therefore, undersampling is not used in this study.

This study applied an oversampling technique followed by a rotation of image patches. Each resampled sample patch is randomly rotated within 180 degrees and flipped horizontally. Oversampling minor classes, however, is not always a panacea, as it might not significantly improve mapping minority classes when the sample size is too small. A recent data augmentation method called mix-up, designed to linearly combine labelled images for model training, has proven successful in mapping tree species (Mäyrä et al. 2021). In this study, the input image patches of different crop classes and corresponding labels (categorical encoding) are mixed up, respectively, to generate blended or synthetic datasets, i.e. an image patch may contain 20% soybean and 80% wheat. The class distribution of the output via mix-up is also balanced. This study proposes a joint learning structure to combine deep learning models to counter imbalanced class distribution and compares it with the aforementioned data augmentation techniques, including oversampling, inverse weighting, and mix-up, for mapping minority classes.

4.6. Model interpretation

The interpretability of deep learning models on crop mapping tasks is still limited considering that the extracted higher-level features are outputted by a hidden learning process in the operating mechanism held by deep learning approaches, which is often called a 'black box' (Heo, Joo, and Moon 2019). The explanation of deep learning methods benefits users in understanding the intricate patterns of crop growth and evaluating model reliability on crop mapping (Xu et al. 2021; Zhang and Hu 2019).

Previous studies investigated deep learning methods by visualizing intermediate layers of the networks for monitoring the model learning process and temporal learning patterns of certain crops (Xu et al. 2020; Zhong, Hu, and Zhou 2019). Another approach to interpreting deep learning models is based on gradient-based explanations (Bastings and Filippova 2020; Mäyrä et al. 2021; Rußwurm and Körner 2020; Xu et al. 2021). Typically, the deep learning model input and corresponding labels are fitted with neural-network-based functions that are differentiable and perform nonlinear transformations. According to the gradient descent algorithm, the model weights during feature extraction can be iteratively updated and optimized to minimize the difference between predicted outputs and corresponding true input values. This study computed the gradients of the predicted scores for each crop type with respect to input image patches for the proposed model via vanilla backpropagation. A gradient for each crop is composed of an array of partial derivatives and it signifies the correlation between the changes in the input features and the corresponding prediction score. The highest magnitude of the gradient indicates the most influential pixels for the process of identifying certain crops. The prediction score is the model soft output derived by the softmax function at the last layer of the proposed model (Figure 5) and it suggests the confidence degree of the proposed model to the classification results of each crop category in this study. The gradients for each class can be visualized via saliency maps. Considering that the spatial dimension of the sample patches used in this study is only 11×11, the assessment of the important input features at spatial dimension may not decisively and accurately demonstrate the locations that contribute to spatial importance for crop mapping. Therefore, the results are more of a performance check for the proposed model.

4.7. Evaluation metrics

For the accuracy assessment of each network, overall accuracy (OA), Cohen's kappa coefficient (Kappa), and F1 score were selected as the performance indicators in this study. The OA is calculated to evaluate the overall model performance. Overall accuracy is calculated by aggregating the number of correctly classified values $n_i^{correct}$ based on the number of classes C and dividing by the total number of samples N in Eq. (10):

$$OA = \frac{\sum_{i=1}^{C} n_i^{correct}}{N} \qquad (10)$$

The Kappa coefficient, also known as Cohen's Kappa, is a widely used metric in deep learning and remote sensing studies to assess the performance of classification models, particularly in the context of crop classification (Congalton 1991; Foody 2004). It is computed from the empirical probability of observed agreement, also known as OA and expected agreement p_e in Eq. (11) and (12). p_e is calculated by n_i^p, the total number of predicted labels, and n_i^t, the total number of ground truth labels. Kappa values range from -1 to 1, with values closer to 1 interpreted as a high level of agreement between the predicted and ground truth labels and values closer to 0 indicating that the agreement is no better than chance. In remote sensing and crop classification studies, a Kappa coefficient is often used alongside other performance metrics, such as the overall accuracy,

producer's accuracy, and user's accuracy, to provide a comprehensive evaluation of the classification model (Congalton 1991):

$$Kappa = \frac{OA - p_e}{1 - p_e} \tag{11}$$

$$p_e = \frac{\sum_{i=1}^{C} n_i^p n_i^t}{N^2} \tag{12}$$

The F1 score is used to measure classification performance grouped into categories since the sample data were imbalanced in this study. It relates to the harmonic mean of the producer's accuracy (*Recall*) and user's accuracy (*Precision*) respectively (Stehman 2001) and is determined as follows:

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i}, i \in C \tag{13}$$

5. Results

5.1. Classification results

It is found that the deep learning models using m-chi decomposition features yielded better classification results than using backscatter and its cross-ratio. See Table 1. Especially for TCNN, this 1D-CNN-based architecture benefited from polarimetric features significantly, increasing the OA and Kappa by >20% and >30%, respectively. With regard to the models considering both spatial and temporal dimensions, using m-chi decomposition features slightly improved the classification accuracy over backscatter, but these models outperform TCNN. Compared with other models, the proposed method, 3D-ConvSTAR, achieved the highest OA on both backscatter (82.0%) and m-chi decomposition (89.4%). On the other perspective, incorporating few optical acquisitions into the sequential radar dataset contributes to improved performance for all models compared to using standalone multi-temporal SAR data, and the deep learning models with the combination of multispectral bands and polarimetric features performed the best among all scenarios. Under this circumstance, the proposed method, 3D-ConvSTAR, outperformed other approaches in terms of the highest OA (91.7%) and Kappa (85.7%). The second-best performer is the standalone 3D-CNN under the same scenario.

Table 1. The comparison of model performance based on multiple composite features. The best scores for each metric are highlighted in bold, and the second best are underlined.

Features	Models	TCNN	3D-CNN	3D-2D CNN	ConvSTAR	3D-ConvSTAR
Backscatter	OA (%)	56.7	82.3	79.2	81.9	82.0
	Kappa (%)	20.9	68.6	62.9	68.0	68.2
m-chi	OA (%)	77.2	85.0	84.1	85.8	89.4
	Kappa (%)	59.4	73.6	71.8	75.2	81.6
Optical+backscatter	OA (%)	72.6	87.0	87.4	88.9	87.9
	Kappa (%)	51.0	77.5	78.1	80.5	79.0
Optical+m-chi	OA (%)	86.4	90.5	89.0	90.0	**91.7**
	Kappa (%)	75.8	83.5	81.1	82.6	**85.7**

Table 2. The comparison of model performance in each class. The best score for each column is highlighted with bold and the second best is underlined.

Models	Features	Maize F1%)	Soybean F1%)	Wheat F1%)	Other crops F1%)	Mean F1%)
TCNN	Backscatter	62.2	56.2	2.5	0.6	30.4
(Pelletier, Webb, and Petitjean 2019)	m-chi	80.9	79.0	37.6	38.2	58.9
	Optical+backscatter	78.1	71.2	63.2	1.1	53.4
	Optical+m-chi	91.6	88.1	43.1	36.2	64.8
3D-CNN	Backscatter	86.1	84.5	50.4	41.2	65.6
(Ji et al. 2018)	m-chi	89.0	86.5	60.8	47.7	71
	Optical+backscatter	90.6	88.4	75.8	50.4	76.3
	Optical+m-chi	93.5	91.3	69.0	70.0	81.0
3D-2D CNN	Backscatter	82.5	81.9	6.5	45.5	54.1
(Roy et al. 2019)	m-chi	87.8	86.3	12.1	54.1	60.1
	Optical+backscatter	91.1	89.5	68.1	48.7	74.4
	Optical+m-chi	93.4	91.1	77.6	62.2	81.1
ConvSTAR	Backscatter	85.5	84.4	30.8	45.9	61.7
(Turkoglu et al. 2021)	m-chi	90.0	88.3	22.2	55.5	64.0
	Optical+backscatter	93.5	90.0	76.2	38.1	74.5
	Optical+m-chi	**93.8**	**92.3**	79.7	52.5	79.6
3D-ConvSTAR	Backscatter	85.6	83.6	69.4	38.0	69.2
	m-chi	92.1	90.4	81.8	61.0	81.3
	Optical+backscatter	91.7	89.6	83.9	45.6	77.7
	Optical+m-chi	93.7	92.2	**90.9**	**74.0**	**87.7**

Table 2 presents the F1-score yielded by deep learning models for each crop type, based on different combinations of input features. The polarimetric parameters overall are better than backscatter according to the F1 scores derived by each model for crop types. Especially m-chi decomposition features lead to significant improvement for TCNN on differentiating crops compared to using backscatter. For the models that consider the spatio-temporal dimension, m-chi decomposition features are still reliable predictors over backscatter for identifying maize and soybean, and 3D-ConvSTAR its advantage on the minority class by producing 81.8% of F1 for wheat that is the highest accuracy among the model results with connection to applying m-chi decomposition. For the use of a combination of multispectral bands and SAR features, ConvSTAR produced the highest F1 scores for maize (93.8%) and soybean (92.3%), which are only 0.1% higher than the same measurements derived by 3D-ConvSTAR. However, 3D-ConvSTAR also yielded the leading performance on less frequent classes, including wheat (90.9%) and other crops (74.0%), resulting in the highest mean F1 (87.7%) among all types of input features used. Therefore, the best-performing model for crop classification with an imbalanced dataset is the proposed 3D-ConvSTAR with the combined features of m-chi decomposition and multispectral bands.

The classification performance using 'Optical+m-chi' as input features varies across data augmentation techniques, with some models performing better or worse than the baseline where no data augmentation is applied in the training data (Table 3). The oversampling method generally performed better than other augmentation methods. TCNN has been improved with oversampling significantly from 64.8% to 83.5% on mean F1. In contrast, balanced loss and mix-up reduced the overall performance for 3D-2D CNN and 3D-ConvSTAR. Some data augmentation methods for the certain model, such as 3D-ConvSTAR, marginally increased identification performance for majority crops including maize and soybean by 0.05% and 0.01% respectively while lowered the performance for minority classes. Furthermore, each method is compared based on model predictions for different

Table 3. The comparison of model performance with applying data augmentation techniques. The best measurements for each column are highlighted in bold, followed by underlines indicating the second-best performance.

Models	Inputs	Methods	Maize F1%)	Soybean F1%)	Wheat F1%)	Other crops F1%)	Mean F1%)
TCNN	Optical+m-	Oversampling	92.7	90.7	83.9	66.8	83.5
(Pelletier, Webb, and Petitjean	chi	Balanced loss	91.5	88.7	64.1	49.1	73.4
2019)		Mix-up	91.6	89.9	63.3	50.0	73.7
			91.6	88.1	43.1	36.2	64.8
3D-CNN	Optical+m-	Oversampling	92.6	90.2	88.4	59.7	82.7
(Ji et al. 2018)	chi	Balanced loss	94.0	92.2	86.1	71.1	85.9
		Mix-up	93.9	91.9	66.3	65.5	79.4
			93.5	91.3	69.0	70.0	81.0
3D-2D CNN	Optical+m-	Oversampling	92.2	90.3	87.1	60.6	82.6
(Roy et al. 2019)	chi	Balanced loss	92.7	88.2	74.6	53.4	77.2
		Mix-up	91.9	90.4	62.2	62.0	76.6
			93.4	91.1	77.6	62.2	81.1
ConvSTAR	Optical+m-	Oversampling	92.8	90.7	**91.8**	71.2	86.6
(Turkoglu et al. 2021)	chi	Balanced loss	93.0	90.4	85.3	55.4	81.0
		Mix-up	**94.2**	91.4	85.8	71.5	85.7
			93.8	**92.3**	79.7	52.5	79.6
3D-ConvSTAR	Optical+m-	Oversampling	93.8	91.2	85.6	67.6	84.6
	chi	Balanced loss	93.6	90.7	89.5	63.6	84.4
		Mix-up	**94.2**	**92.3**	88.5	71.0	86.5
			93.7	92.2	90.9	**74.0**	**87.7**

geographical locations within Bei'an, taking into account the highest mean F1 score achieved by certain models that incorporate data augmentation techniques (Figure 6).

A close examination of the classification accuracies reveals that the proposed 3D-ConvSTAR method outperforms the other models in most cases. At Site A, the 3D-ConvSTAR model yields an accuracy of 89.4%, which is higher than the accuracies derived by the TCNN with oversampling (84.3%), 3D-CNN with balanced loss (85.4%), the 3D-2D CNN with oversampling (89.0%) and ConvSTAR with oversampling (87.1%) models. At Site B, the 3D-ConvSTAR method and ConvSTAR with oversampling both achieve a classification accuracy of 95.9%. While the 3D-2D CNN with oversampling model exhibits a slightly higher accuracy (96.5%), the 3D-ConvSTAR model outperforms the other two methods, namely TCNN with oversampling (90.3%) and 3D-CNN with balanced loss (96.0%). Finally, at Site C, the proposed 3D-ConvSTAR model demonstrates the second-highest classification accuracy (97.3%), only surpassed by the 3D-2D CNN with oversampling (97.5%). Nevertheless, the difference is marginal and 3D-ConvSTAR method outperforms the rest of the scenarios. In summary, the proposed 3D-ConvSTAR demonstrates competitive performance in comparison to the other deep learning models fed with augmented data for crop mapping across all three sites. It also performs the highest mean F1 score of 87.7% on the testing dataset compared with the models using augmented data (Table 3). Therefore, the proposed method was used to predict unseen data for the creation of a thematic map.

As mentioned in Figure 2, ten per cent of total ground truth samples were used for dataset split and sixty percent of which was used for model training. Then, the pre-trained 3D-ConvSTAR produced the annual crop map during model inference for Bei'an in 2017, shown in Figure 7. The predicted results were also compared with the total ground truth dataset in the confusion matrix illustrated in Figure 8. The cell in the top left corner shows that 96.58% of maize instances were correctly classified as
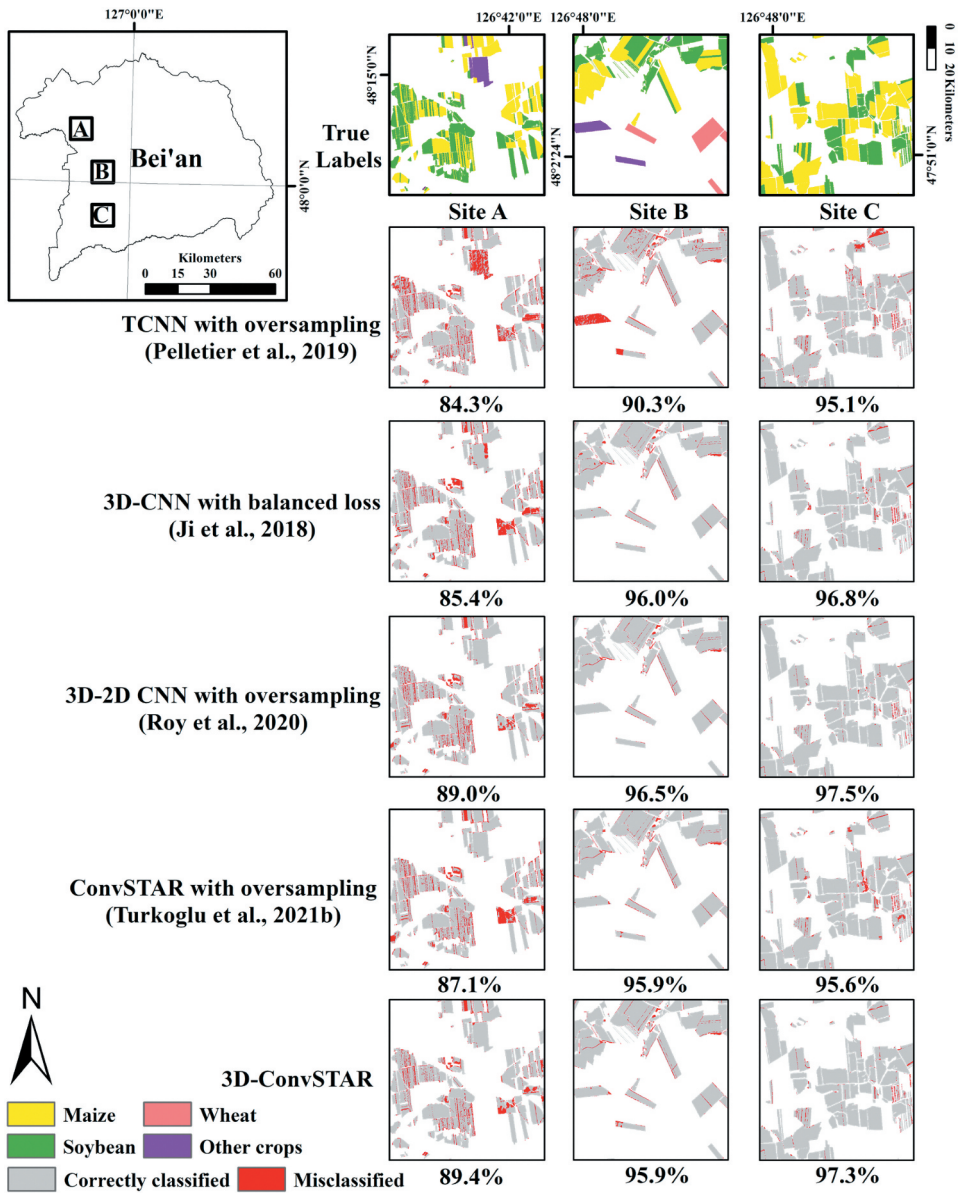
Figure 6. Comparison of classification performances between the models with applying data augmentation techniques and the proposed method across various sites within Bei'an. Percentages indicate the proportion of correctly classified samples with respect to ground truth labels.

maize, while 2.92% of them were misclassified as soybean and 0.00% as wheat. Similarly, the cell in the bottom right corner shows that 86.87% of instances for other crops were correctly classified as themselves, while 11.18% of them were misclassified as soybean. In general, the model performs relatively well in distinguishing maize, soybean, and wheat while it struggles more with identifying other crops.
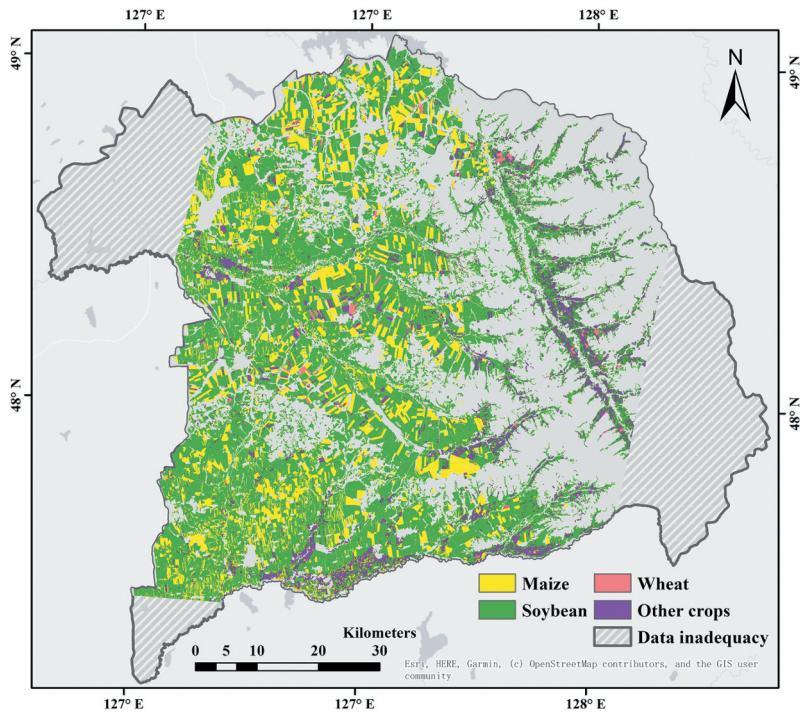
Figure 7. The annual crop map for Bei'an 2017. It was produced by 3D-ConvSTAR, weakly supervised with ten per cent of all ground truth samples. The areas not designated as cropland were excluded using a cropland mask introduced in Section 3.1.

5.2. Model interpretation

The prediction scores, as outlined in Section 4.5, are soft outputs produced by the final layer of the 3D-ConvSTAR model (illustrated in Figure 5). Figure 9 visually demonstrates the confidence level of the proposed model in its crop classification performance on the testing dataset. The results indicate that 3D-ConvSTAR exhibits a higher level of confidence in its mapping of maize, soybean, and wheat crops, as compared to other crops. This is evident from the concentration of prediction scores for most samples of the three classes, which averagely hover around 90%. The proposed model is less confident in accurately identifying other crops with reference to the relatively lower mean prediction score (71%), which is consistent with the misclassification presented by the F1 score in Table 2, despite the fact that the 'other crops' category has a larger number of training samples than wheat (see Figure 2).

The average gradient magnitudes are represented by saliency maps, as shown in Figure 10. These maps represent the most important spatial locations for each crop type in the image samples of the testing dataset. The shape of the pixel chunks with the highest importance for maize may complement the part of the lowest importance for soybean and vice versa, indicating that maize and soybean in the samples are mostly intercropped. The pixel importance for wheat shows that the field shapes in the samples are mostly separated. However, the pixel importance for other crops is scattered without forming a clear shape, which corresponds to the relatively lower mapping accuracy for those crops.
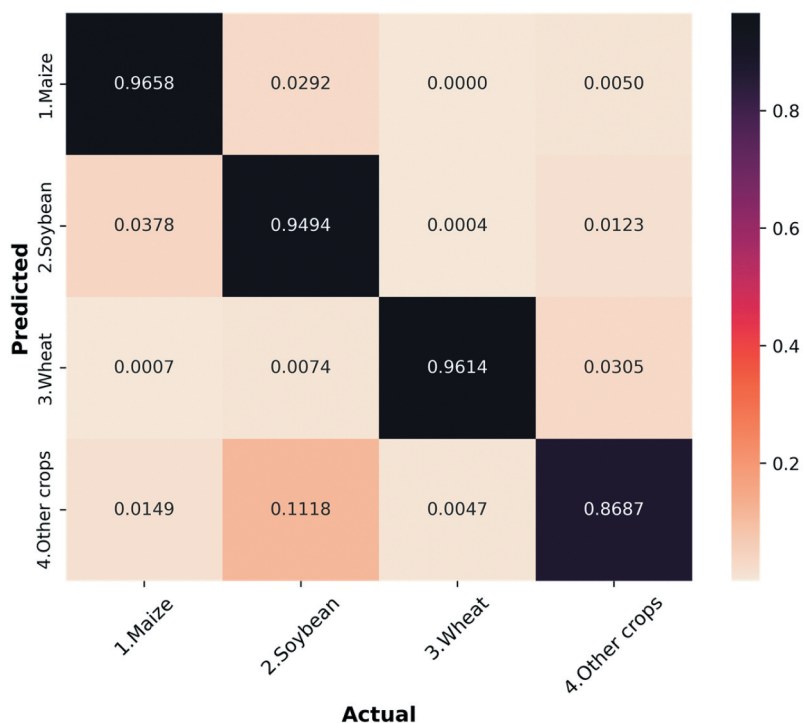
Figure 8. The confusion matrix for the comparison between predicted labels derived by 3D-ConvSTAR and all ground truth labels.
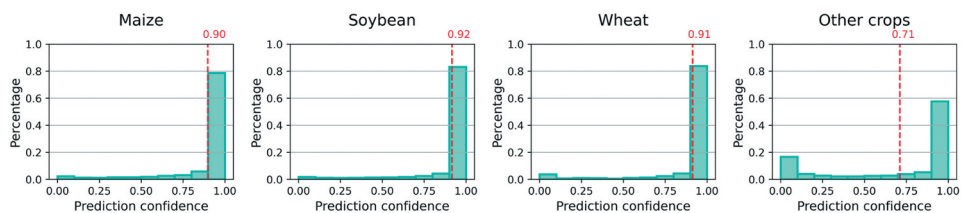

Figure 9. The prediction score distribution for each crop, derived by the last dense layer of 3D-ConvSTAR. The red dashed lines indicate average prediction scores.
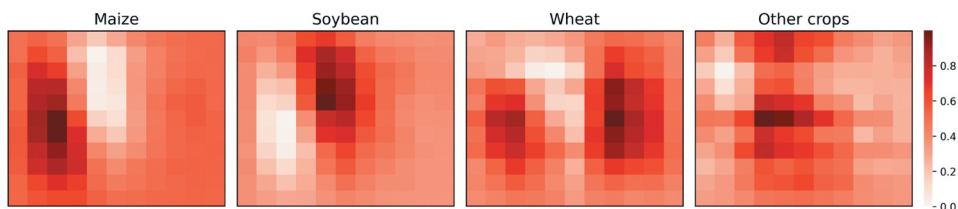

Figure 10. The saliency maps represented by the average magnitude of gradients for each crop. 1500 image patches were randomly extracted from the testing dataset and fed into 3D-ConvSTAR to generate saliency maps for illustration.

6. Discussion

In this study, various deep learning architectures for patch-based crop mapping are evaluated using SAR and fused SAR-optical data. The proposed method functions as an ensemble deep learning by synergistically connecting 3D-CNN and ConvSTAR, reaching the highest performance overall among all models in terms of the OA (91.7%) and Kappa (85.7%) for the identification and classification of different crops (see Table 1). This study also validated the effectiveness of using SAR polarimetric decomposition parameters of m-chi for identifying certain crops over using SAR backscatter, which confirmed the findings by previous studies (e.g. De, Kumar, and Rao 2014; Sonobe 2019). The integration of a few optical acquisitions and time-series SAR image data does improve the classification performance overall compared to standalone SAR data, but maize and soybean are not increased significantly (<10% on average) regarding the F1-score derived by all models (Table 2). These results are likely to be explained by the fact that maize and soybean are dominant crops in the county that provides an enormous number of labelled ground truth (Figure 2), so both crops can be well-trained by those data-driven models. In contrast, the fusion of optical and m-chi decomposition features enhanced minority classes surprisingly by all models, especially for wheat. The proposed method with multi-source fusion yields the highest F1-score (90.9%) for wheat compared to using backscatter (69.4%) and polarimetric features (81.8%). This indicates that multispectral information contributes mostly to the enhancement of mapping for minority crops.

We discovered that crop mapping performance can be enhanced not only by fusing SAR-optical datasets, which provide spatio-temporal, polarimetric, and spectral characteristics related to different crop structures (Gao et al. 2018; Van Tricht et al. 2018), but also by utilizing multispectral information as a reliable complementary source owing to the synergistic nature of SAR and optical data. This study demonstrates that Sentinel-1 and Sentinel-2 imagery exhibit a mutually complementary effect, increasing the sensitivity of both sensors to specific crop class characteristics throughout the growing season. Sentinel-2 data can be associated with the quantitative analysis of chlorophyll and moisture content in crop leaves, with spectral bands such as the Vegetation Red Edge being particularly useful for differentiating certain crops, confirming previous findings by Guerschman et al. (2003) and You and Dong (2020). Sentinel-1 data is sensitive to morphological variations, as it provides biophysical, structural, and agronomic characteristics, and is strongly correlated with the structural development of crops during the growing season (Adrian, Sagan, and Maimaitijiang 2021; Sonobe 2019).

In addition to the SAR backscatter, polarimetric parameters can reflect in-depth scattering properties of crops due to scattering mechanisms with robust physical interpretability, making them useful for crop mapping (Gao et al. 2018; He et al. 2020; Liao et al. 2020; Xie et al. 2019). Polarimetric features directly relate to the underlying physical properties of the crops. These parameters can be used to quantify the contribution of different scattering mechanisms and provide insights into the crops' biophysical, structural, and agronomic properties, such as crop type, plant density, leaf area index, and growth stage. In this study, we capitalized SAR scattering diversity, such as surface scattering, volume scattering, and double-bounce scattering (See Section 4.2) that are responsible for the interactions between the electromagnetic waves emitted by SAR sensors and the various components of a crop's structure. Sentinel-1 data, however, is

limited to insufficient decomposition methods, since it is a dual-polarized SAR sensor, which restricts the available decomposition methods. While quad-polarimetric data can be analysed using fully polarimetric decomposition algorithms, this study found that m-chi decomposition, originating from compact-pol planforms, effectively maps crops using dual-polarization data and synergizes well with optical bands. It is important to note that fully polarimetric SAR systems often have reduced swath coverage and relatively inconsistent temporal frequency, posing challenges for crop mapping across extensive areas (Sonobe 2019).

With respect to the ablation study regarding applying different input features, the proposed deep learning network showcases the advantage of the proposed model that outperformed the deep learning methods with standalone architectures (TCNN, 3D-CNN, and ConvSTAR), and combined architectures such as 3D-2D CNN proposed in other studies for crop mapping under the same input predictors, as detailed in Table 2. The model performance varies significantly across different crop classes, with wheat and other crops generally exhibiting lower F1 scores compared to maize and soybean. The 3D-ConvSTAR improved all performance overall, in particular per-class performance in separating the maize, soybean, and wheat, providing a beneficial method for local industries due to the commercial interests in these crops. We also investigated the comparative analysis of various deep learning models with data augmentation techniques for crop mapping to deal with imbalanced class distribution (Table 3). Examining the F1 scores for individual crop classes, it is evident that the 3D-ConvSTAR model using the mix-up consistently outperforms other combinations, achieving 94.2% for maize, and 92.3% for soybean, but reducing performance for wheat and other crops. All models produced similar results for maize and soybean after data augmentation techniques are applied, and oversampling generally outperforms other augmentation methods. Wheat and other crops generally exhibit lower F1 scores compared to maize and soybean. This finding, while preliminary, may imply that this data augmentation technique could be more useful for enriching training samples when data collection is a major challenge in certain research fields. For example, it is particularly well-suited to augment imagery data collected from the human nervous system (Smucny et al. 2022), and increase the airborne training sample size for mapping species (Mäyrä et al. 2021). However, it may not be useful to overcome imbalanced class distribution.

The proposed network, in terms of average F1 score, outperforms other approaches that rely on data augmentation techniques, primarily because it effectively integrates the temporal nature of remote sensing data into a more sophisticated input space while accounting for the spatial relationships between features along the time series. This leads to a better separation of crop types with homogeneous representations (Figure 7 and 8). However, the proposed method is prone to generating a higher number of training parameters compared to alternative methods, resulting in increased model training time. This issue is particularly due to the connections between the learned features produced by the 3D-CNN, ConvSTAR, and the subsequent shallow CNNs implemented by the 2D-CNN, as these settings can lead to an increased number of training parameters. All classifiers exhibit suboptimal performance for the 'other crops' category, which can be attributed to the mixture of various crop types. Each of these unknown crop types is only represented by a relatively small sample size in the training data, thereby limiting the model's ability for identification. Although the 'other crops' category has more training

samples than wheat (Figure 2), it may still be underrepresented compared to samples for maize, soybean, and wheat in the dataset. This could lead the model to focus on the distinctive class labels and consequently perform poorly in the 'other crops' category with mixed labels for unknown crop types. The 'other crops' category may encompass a wide range of crop types, each with distinct spectral, polarimetric, and temporal signatures. For example, there is only one field parcel for rice in the ground truth data collection, so it was labelled as other crops in this study. This increased diversity may make it more challenging for the deep learning model to accurately identify and classify these crops. In contrast, maize, soybean, and wheat may have more consistent and easily distinguishable characteristics, allowing for a higher F1 score. Additionally, the features extracted from the combination of SAR and optical data might be more informative and discriminative for maize, soybean, and wheat than for the 'other crops' category. The complexity of the features for the 'other crops' category might be higher, making it more difficult for the model to learn and correctly classify these samples, which leads to the lower F1 score observed.

The results overall demonstrate the advantages of using the combination of polarimetric and multispectral data from Sentinel-1 and Sentinel-2. Data fusion provides additional information for classification algorithms to exploit crops' structural details, while also offering supplementary polarimetric and spectral properties. The discrepancy in F1 scores between the wheat and 'other crops' categories in crop mapping highlight the need for further research into optimizing model architectures and methodologies to enhance crop mapping across all classes. Moreover, future work should focus on integrating fully polarimetric data with optical data to further improve crop mapping accuracy by applying popular deep learning architectures, such as Fully Convolutional Neural Networks (FCN), which perform pixel-wise segmentation on images. For instance, the 3D U-Net architecture can be employed to extract the spatio-temporal features of crops (Adrian, Sagan, and Maimaitijiang 2021; Ji et al. 2018). Investigating the contribution of SAR texture information combined with optical data to semantic segmentation for crop mapping also enables further exploration. Recent studies have employed a self-attention-based convolutional recurrent network to learn temporal dependencies of multivariable time series (Fu et al. 2022) and combined 3D-CNN with an attention-based recurrent network for crop yield prediction (Nejad et al. 2023). Both studies assessed the feasibility of attention mechanisms in extracting attentive spatio-temporal features. Consequently, future research could involve the integration of 3D-CNN with attention-based convolutional recurrent networks, such as ConvSTAR, for crop mapping and comparison with architectures for semantic segmentation. More importantly, the model's robustness should be further evaluated for predicting crop types in different years. Model behaviours may be influenced by interannual variability within the same region, and recurrent structures have shown promise in capturing crop phenological characteristics and enhancing model generalization (Xu et al. 2021). Assessing the model's spatial transferability is also a critical aspect of future research, given the potential application of the model in diverse geographical contexts. This could facilitate the design of efficient strategies for improved applicability, potentially contributing to the optimization of agricultural practices and crop mapping on a global scale. One such strategy refers to training the model with representative crop

datasets that can accurately reflect the complexity and heterogeneity of the agricultural landscape. Alternatively, the model's parameters could be adjusted to accommodate the unique conditions of specific locations.

7. Conclusions

In this research, we proposed a workflow for multi-temporal crop mapping based on the fusion of Sentinel-1 polarimetric parameters and Sentinel-2 multispectral reflectance, combined with various deep learning architectures. The proposed 3D-ConvSTAR, which connects 3D-CNN layers and convolutional recurrent layers, delivers enhanced classification performance for crop mapping in comparison to the architectures designed in previous studies. Additionally, the designed architecture is robust when training the dataset with imbalanced class distribution and outperforms other data augmentation techniques. This study demonstrates that crop mapping can be conducted with high accuracy using the proposed 3D-ConvSTAR in terms of overall accuracy and F1 score for each crop class. Although the implemented architecture is likely not the optimal solution, given the training parameters overload, it still manages to produce accurate and valuable results for separating the crops with significant commercial value in Bei'an. While the proposed network exhibits superior performance in terms of crop type separation and accounting for the temporal and spatial relationships in remote sensing data, it is essential to address the challenges posed by the increased number of training parameters and the inherent limitations in classifying under-represented crop types. Future research should focus on optimizing the network architecture and exploring alternative approaches to improve classification accuracy across all crop types while minimizing the computational cost associated with training the model. The model's generalization for crop mapping needs further assessment based on interannual and spatial transferability.

Acknowledgements

Disclosure statement

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.

ORCID

Niantang Liu 🆔 http://orcid.org/0009-0002-7083-0950
Qunshan Zhao 🆔 http://orcid.org/0000-0002-5549-9457
Richard Williams 🆔 http://orcid.org/0000-0001-6067-1947
Brian Barrett 🆔 http://orcid.org/0000-0002-4380-4020

References

Adrian, J., V. Sagan, and M. Maimaitijiang. 2021. "Sentinel SAR-Optical Fusion for Crop Type Mapping Using Deep Learning and Google Earth Engine." *ISPRS Journal of Photogrammetry and Remote Sensing* 175:215–235. https://doi.org/10.1016/j.isprsjprs.2021.02.018.

Ainsworth, T. L., J. P. Kelly, and J. S. Lee. 2009. "Classification Comparisons Between Dual- Pol, Compact Polarimetric and Quad-Pol SAR Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (5): 464–471. https://doi.org/10.1016/j.isprsjprs.2008.12.008.

Audebert, N., B. Le Saux, and S. Lefèvre. 2019. "Deep Learning for Classification of Hyperspectral Data: A Comparative Review." *IEEE Geoscience and Remote Sensing Magazine* 7 (2): 159–173. https://doi.org/10.1109/MGRS.2019.2912563.

Bargiel, D. 2017. "A New Method for Crop Classification Combining Time Series of Radar Images and Crop Phenology Information." *Remote Sensing of Environment* 198:369–383. https://doi.org/10.1016/j.rse.2017.06.022.

Bastings, J., and K. Filippova. 2020. "The Elephant in the Interpretability Room: Why Use Attention as Explanation When We Have Saliency Methods?" *ArXiv*. https://doi.org/10.48550/arXiv.2010.05607.

Boryan, C., Z. Yang, R. Mueller, and M. Craig. 2011. "Monitoring US Agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program." *Geocarto International* 26 (5): 341–358. https://doi.org/10.1080/10106049.2011.562309.

Botev, Z. I., D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer. 2013. "The Cross-Entropy Method for Optimization." In *Handbook of Statistics - Machine Learning: Theory and Applications*, 35–59. Vol. 31. Elsevier. https://doi.org/10.1016/B978-0-444-53859-8.00003-5.

Boulila, W., H. Ghandorh, M. A. Khan, F. Ahmed, and J. Ahmad. 2021. "A Novel CNN-LSTM-Based Approach to Predict Urban Expansion." *Ecological Informatics* 64:101325. https://doi.org/10.1016/j.ecoinf.2021.101325.

Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. https://doi.org/10.1016/0034-42579190048-B.

Crisóstomo de Castro Filho, H., O. Abílio de Carvalho Júnior, O. L. Ferreira de Carvalho, P. Pozzobon de Bem, R. dos Santos de Moura, A. Olino de Albuquerque, C. Rosa Silva, P. H. Guimaraes Ferreira, R. Fontes Guimarães, and R. A. Trancoso Gomes. 2020. "Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series." *Remote Sensing* 12 (16): 2655. https://doi.org/10.3390/rs12162655 .

Cui, Y., M. Jia, Lin, T. Y., Song, Y., and Belongie, S. 2019. "Class-balanced loss based on effective number of samples." *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach, CA. 9268–9277.

Defourny, P., S. Bontemps, N. Bellemans, C. Cara, G. Dedieu, E. Guzzonato, and B. Koetz. 2019. "Near Real-Time Agriculture Monitoring at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems Around the World." *Remote Sensing of Environment* 221:551–568. https://doi.org/10.1016/j.rse.2018.11.007.

De, S., V. Kumar, and Y. S. Rao 2014. "Crop Classification Using RISAT-1 Hybrid Polarimetric SAR Data." In EUSAR 2014; 10th European Conference on Synthetic Aperture Radar, Berlin, Germany. 1–4. VDE.

Dong, Q., S. Gong, and X. Zhu. 2018. "Imbalanced Deep Learning by Minority Class Incremental Rectification." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (6): 1367–1381. https://doi.org/10.1109/TPAMI.2018.2832629.

Dong, J., X. Xiao, G. Zhang, M. A. Menarguez, C. Y. Choi, Y. Qin, and B. Moore. 2016. "Northward Expansion of Paddy Rice in Northeastern Asia During 2000–2014." *Geophysical Research Letters* 43 (8): 3754–3761. https://doi.org/10.1002/2016GL068191.

Dou, P., H. Shen, Z. Li, and X. Guan. 2021. "Time Series Remote Sensing Image Classification Framework Using Combination of Deep Learning and Multiple Classifiers System." *International Journal of Applied Earth Observation and Geoinformation* 103:102477. https://doi.org/10.1016/j.jag.2021.102477.

Fan, C., B. Zheng, S. W. Myint, and R. Aggarwal. 2014. "Characterizing Changes in Cropping Patterns Using Sequential Landsat Imagery: An Adaptive Threshold Approach and Application to Phoenix,

Arizona." *International Journal of Remote Sensing* 35 (20): 7263–7278. https://doi.org/10.1080/01431161.2014.967891.

Foody, G. M. 2004. "Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy." *Photogrammetric Engineering & Remote Sensing* 70 (5): 627–633. https://doi.org/10.14358/PERS.70.5.627.

Fu, E., Y. Zhang, F. Yang, and S. Wang. 2022. "Temporal Self-Attention-Based Conv-LSTM Network for Multivariate Time Series Prediction." *Neurocomputing* 501:162–173. https://doi.org/10.1016/j.neucom.2022.06.014 .

Gao, H., C. Wang, G. Wang, J. Zhu, Y. Tang, P. Shen, and Z. Zhu. 2018. "A Crop Classification Method Integrating GF-3 PolSar and Sentinel-2A Optical Data in the Dongting Lake Basin." *Sensors* 18 (9): 3139. https://doi.org/10.3390/s18093139.

Griffiths, P., C. Nendel, and P. Hostert. 2019. "Intra-Annual Reflectance Composites from Sentinel-2 and Landsat for National-Scale Crop and Land Cover Mapping." *Remote Sensing of Environment* 220:135–151. https://doi.org/10.1016/j.rse.2018.10.031.

Guerschman, J. P., J. M. Paruelo, C. D. Bella, M. C. Giallorenzi, and F. Pacin. 2003. "Land Cover Classification in the Argentine Pampas Using Multi-Temporal Landsat TM Data." *International Journal of Remote Sensing* 24 (17): 3381–3402. https://doi.org/10.1080/0143116021000021288.

He, H., and E. A. Garcia. 2009. "Learning from Imbalanced Data." *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

He, C., B. He, M. Tu, Y. Wang, T. Qu, D. Wang, and M. Liao. 2020. "Fully Convolutional Networks and a Manifold Graph Embedding-Based Algorithm for Polsar Image Classification." *Remote Sensing* 12 (9): 1467. https://doi.org/10.3390/rs12091467.

Heihe Social and Economic Statistics Yearbook. 2018. *Heihe Social and Economic Statistics Yearbook*. Beijing: China Statistical Publishing House.

Heo, J., S. Joo, and T. Moon. 2019. "Fooling Neural Network Interpretations via Adversarial Model Manipulation." *ArXiv*. https://doi.org/10.48550/arXiv.1902.02041.

Inglada, J., M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, and B. Koetz. 2015. "Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery." *Remote Sensing* 7 (9): 12356–12379. https://doi.org/10.3390/rs70912356 .

Ji, S., C. Zhang, A. Xu, Y. Shi, and Y. Duan. 2018. "3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images." *Remote Sensing* 10 (1): 75. https://doi.org/10.3390/rs10010075 .

Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." *ArXiv*. https://doi.org/10.48550/arXiv.1412.6980 .

Kussul, N., M. Lavreniuk, S. Skakun, and A. Shelestov. 2017. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14 (5): 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Kussul, N., L. Mykola, A. Shelestov, and S. Skakun. 2018. "Crop Inventory at Regional Scale in Ukraine: Developing in Season and End of Season Crop Maps with Multi-Temporal Optical and SAR Satellite Imagery." *European Journal of Remote Sensing* 51 (1): 627–636. https://doi.org/10.1080/22797254.2018.1454265 .

Liao, C., J. Wang, Q. Xie, A. A. Baz, X. Huang, J. Shang, and Y. He. 2020. "Synergistic Use of Multi-Temporal RADARSAT-2 and Venμs Data for Crop Classification Based on 1D Convolutional Neural Network." *Remote Sensing* 12 (5): 832. https://doi.org/10.3390/rs12050832.

Li, C., W. Chen, Y. Wang, Y. Wang, C. Ma, Y. Li, and W. Zhai, W. Zhai. 2022. "Mapping Winter Wheat with Optical and SAR Images Based on Google Earth Engine in Henan Province, China." *Remote Sensing* 14 (2): 284. https://doi.org/10.3390/rs14020284.

Li, H., J. Huang, and S. Ji 2019. "Bearing Fault Diagnosis with a Feature Fusion Method Based on an Ensemble Convolutional Neural Network and Deep Neural Network." Sensors, 19(9), 2034.

Ling, C. X., and V. S. Sheng. 2008. "Cost-Sensitive Learning and the Class Imbalance Problem." *Encyclopedia of Machine Learning* 2011. 231–235.

Liu, J., W. Kuang, Z. Zhang, X. Xu, Y. Qin, J. Ning, and W. Chi. 2014. "Spatiotemporal Characteristics, Patterns, and Causes of Land-Use Changes in China Since the Late 1980s." *Journal of Geographical Sciences* 24 (2): 195–210. https://doi.org/10.1007/s11442-014-1082-6.

Mäyrä, J., S. Keski-Saari, S. Kivinen, T. Tanhuanpää, P. Hurskainen, P. Kullberg, and P. Vihervaara. 2021. "Tree Species Classification from Airborne Hyperspectral and LiDar Data Using 3D Convolutional Neural Networks." *Remote Sensing of Environment* 256:112322. https://doi.org/10.1016/j.rse.2021.112322.

McNairn, H., J. Shang, X. Jiao, and C. Champagne. 2009. "The Contribution of ALOS PALSAR Multipolarization and Polarimetric Data to Crop Classification." *IEEE Transactions on Geoscience and Remote Sensing* 47 (12): 3981–3992. https://doi.org/10.1109/TGRS.2009.2026052.

Moumni, A., A. Lahrouni, and H.-S. Jung. 2021. "Machine Learning-Based Classification for Crop-Type Mapping Using the Fusion of High-Resolution Satellite Imagery in a Semiarid Area." *Scientifica* 2021:1–20. https://doi.org/10.1155/2021/8810279.

Nair, V., and G. E. Hinton. 2010. "Rectified Linear Units Improve Restricted Boltzmann Machines." *In Proceedings of the 27th international conference on machine learning (ICML-10)*, Haifa, Israel. 807–814.

Nejad, S. M. M., D. Abbasi-Moghadam, A. Sharifi, N. Farmonov, K. Amankulova, and M. László. 2022. "Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16:254–266. https://doi.org/10.1109/JSTARS.2022.3223423.

Ning, J., J. Liu, W. Kuang, X. Xu, S. Zhang, C. Yan, and J. Ning 2018. "Spatiotemporal Patterns and Characteristics of Land-Use Change in China During 2010–2015." *Journal of Geographical Sciences*, *28*, 547–562.

Nord, M. E., T. L. Ainsworth, J. S. Lee, and N. J. Stacy 2008. "Comparison of Compact Polarimetric Synthetic Aperture Radar Modes." *IEEE Transactions on Geoscience and Remote Sensing*, 47(1), 174–188

Pelletier, C., S. Valero, J. Inglada, N. Champion, and G. Dedieu. 2016. "Assessing the Robustness of Random Forests to Map Land Cover with High Resolution Satellite Image Time Series Over Large Areas." *Remote Sensing of Environment* 187:156–168. https://doi.org/10.1016/j.rse.2016.10.010.

Pelletier, C., G. I. Webb, and F. Petitjean. 2019. "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series." *Remote Sensing* 11 (5): 523. https://doi.org/10.3390/rs11050523 .

Qu, Y., W. Zhao, Z. Yuan, and J. Chen. 2020. "Crop Mapping from Sentinel-1 Polarimetric Time-Series with a Deep Neural Network." *Remote Sensing* 12 (15): 2493. https://doi.org/10.3390/rs12152493 .

Raney, R. K., J. T. Cahill, G. W. Patterson, and D. B. J. Bussey. 2012. "The M-Chi Decomposition of Hybrid Dual-Polarimetric Radar Data with Application to Lunar Craters." *Journal of Geophysical Research: Planets* 117 (E12). https://doi.org/10.1029/2011JE003986.

Ren, M., W. Zeng, B. Yang, and R. Urtasun 2018, July. "Learning to Reweight Examples for Robust Deep Learning." In International conference on machine learning, Stockholm, Sweden. 4334–4343. PMLR.

Roy, S. K., G. Krishna, S. R. Dubey, and B. B. Chaudhuri. 2019. "HybridSn: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification." *IEEE Geoscience and Remote Sensing Letters* 17 (2): 277–281. https://doi.org/10.1109/LGRS.2019.2918719.

Rußwurm, M., and M. Korner 2017. "Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-Spectral Satellite Images." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, Honolulu, Hawaii, USA. 11–19.

Rußwurm, M., and M. Körner. 2018. "Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders." *ISPRS International Journal of Geo-Information* 7 (4): 129. https://doi.org/10.3390/ijgi7040129.

Rußwurm, M., and M. Körner. 2020. "Self-Attention for Raw Optical Satellite Time Series Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 169:421–435. https://doi.org/10.1016/j.isprsjprs.2020.06.006.

Sharma, A., X. Liu, X. Yang, and D. Shi. 2017. "A Patch-Based Convolutional Neural Network for Remote Sensing Image Classification." *Neural Networks* 95:19–28. https://doi.org/10.1016/j.neunet.2017.07.017.

Shi, X., Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo. 2015. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting." *Advances in Neural Information Processing Systems* 1: 802–810.

Smucny, J., G. Shi, T. A. Lesh, C. S. Carter, and I. Davidson. 2022. "Data Augmentation with Mixup: Enhancing Performance of a Functional Neuroimaging-Based Prognostic Deep Learning Classifier in Recent Onset Psychosis." *NeuroImage: Clinical* 36:103214. https://doi.org/10.1016/j.nicl.2022.103214.

Song, X. P., P. V. Potapov, A. Krylov, L. King, C. M. Di Bella, A. Hudson, and M. C. Hansen. 2017. "National-Scale Soybean Mapping and Area Estimation in the United States Using Medium Resolution Satellite Imagery and Field Survey." *Remote Sensing of Environment* 190:383–395. https://doi.org/10.1016/j.rse.2017.01.008.

Sonobe, R. 2019. "Parcel-Based Crop Classification Using Multi-Temporal TerraSAR-X Dual Polarimetric Data." *Remote Sensing* 11 (10): 1148. https://doi.org/10.3390/rs11101148.

Sonobe, R., H. Tani, X. Wang, N. Kobayashi, and H. Shimamura. 2014. "Random Forest Classification of Crop Type Using Multi-Temporal TerraSAR-X Dual-Polarimetric Data." *Remote Sensing Letters* 5 (2): 157–164. https://doi.org/10.1080/2150704X.2014.889863.

Stehman, S. V. 2001. "Statistical Rigor and Practical Utility in Thematic Map Accuracy Assessment." *Photogrammetric Engineering and Remote Sensing* 67 (6): 727–734.

Sun, C., Y. Bian, T. Zhou, and J. Pan. 2019. "Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region." *Sensors* 19 (10): 2401. https://doi.org/10.3390/s19102401.

Sun, Z., L. Di, H. Fang, and A. Burgess. 2020. "Deep Learning Classification for Crop Types in North Dakota." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:2200–2213. https://doi.org/10.1109/JSTARS.2020.2990104.

Teimouri, M., M. Mokhtarzade, N. Baghdadi, and C. Heipke. 2022. "Fusion of Time-Series Optical and SAR Images Using 3D Convolutional Neural Networks for Crop Classification." *Geocarto International* 37 (27): 1–18. https://doi.org/10.1080/10106049.2022.2095446.

Teluguntla, P., P. S. Thenkabail, A. Oliphant, J. Xiong, M. K. Gumma, R. G. Congalton, and A. Huete. 2018. "A 30-M Landsat-Derived Cropland Extent Product of Australia and China Using Random Forest Machine Learning Algorithm on Google Earth Engine Cloud Computing Platform." *ISPRS Journal of Photogrammetry and Remote Sensing* 144:325–340. https://doi.org/10.1016/j.isprsjprs.2018.07.017.

Turkoglu, M. O., S. D'Aronco, G. Perich, F. Liebisch, C. Streit, K. Schindler, and J. D. Wegner. 2021. "Crop Mapping from Image Time Series: Deep Learning with Multi-Scale Label Hierarchies." *Remote Sensing of Environment* 264:112603. https://doi.org/10.1016/j.rse.2021.112603.

Turkoglu, M. O., S. D'Aronco, J. D. Wegner, and K. Schindler. 2021. "Gating Revisited: Deep Multi-Layer RNNs That Can Be Trained." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (8): 4081–4092. https://doi.org/10.1109/TPAMI.2021.3064878.

Van Tricht, K., A. Gobin, S. Gilliams, and I. Piccard. 2018. "Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium." *Remote Sensing* 10 (10): 1642. https://doi.org/10.3390/rs10101642.

Wang, S., G. Azzari, and D. B. Lobell. 2019. "Crop Type Mapping without Field-Level Labels: Random Forest Transfer and Unsupervised Clustering Techniques." *Remote Sensing of Environment* 222:303–317. https://doi.org/10.1016/j.rse.2018.12.026.

Wang, T., J. Li, M. Zhang, A. Zhu, H. Snoussi, and C. Choi. 2021. "An Enhanced 3DCNN-ConvLstm for Spatiotemporal Multimedia Data Analysis." *Concurrency & Computation: Practice & Experience* 33 (2): e5302. https://doi.org/10.1002/cpe.5302.

Wei, S., H. Zhang, C. Wang, Y. Wang, and L. Xu. 2019. "Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model." *Remote Sensing* 11 (1): 68. https://doi.org/10.3390/rs11010068 .

Xie, Q., J. Wang, C. Liao, J. Shang, J. M. Lopez-Sanchez, H. Fu, and X. Liu. 2019. "On the Use of Neumann Decomposition for Crop Classification Using Multi-Temporal RADARSAT-2 Polarimetric SAR Data." *Remote Sensing* 11 (7): 776. https://doi.org/10.3390/rs11070776.

Xu, J., J. Yang, X. Xiong, H. Li, J. Huang, K. C. Ting, and T. Lin. 2021. "Towards Interpreting Multi-Temporal Deep Learning Models in Crop Mapping." *Remote Sensing of Environment* 264:112599. https://doi.org/10.1016/j.rse.2021.112599.

Xu, J., Y. Zhu, R. Zhong, Z. Lin, J. Xu, H. Jiang, and T. Lin. 2020. "DeepCropmapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean Mapping." *Remote Sensing of Environment* 247:111946. https://doi.org/10.1016/j.rse.2020.111946.

Yang, L., L. Wang, J. Huang, L. R. Mansaray, and R. Mijiti. 2019. "Monitoring Policy-Driven Crop Area Adjustments in Northeast China Using Landsat-8 Imagery." *International Journal of Applied Earth Observation and Geoinformation* 82:101892. https://doi.org/10.1016/j.jag.2019.06.002.

You, N., and J. Dong. 2020. "Examining Earliest Identifiable Timing of Crops Using All Available Sentinel 1/2 Imagery and Google Earth Engine." *ISPRS Journal of Photogrammetry and Remote Sensing* 161:109–123. https://doi.org/10.1016/j.isprsjprs.2020.01.001.

You, N., J. Dong, J. Huang, G. Du, G. Zhang, Y. He, and X. Xiao, Y. Di, X. Xiao. 2021. "The 10-M Crop Type Maps in Northeast China During 2017–2019." *Scientific Data* 8 (1): 41. https://doi.org/10.1038/s41597-021-00827-9.

Zhang, P., and S. Hu. 2019. "Fine Crop Classification by Remote Sensing in Complex Planting Areas Based on Field Parcel." *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering* 35 (21): 125–134.

Zhao, H., S. Duan, J. Liu, L. Sun, and L. Reymondin. 2021. "Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information." *Remote Sensing* 13 (14): 2790. https://doi.org/10.3390/rs13142790.

Zheng, B., S. W. Myint, P. S. Thenkabail, and R. M. Aggarwal. 2015. "A Support Vector Machine to Identify Irrigated Crop Types Using Time-Series Landsat NDVI Data." *International Journal of Applied Earth Observation and Geoinformation* 34:103–112. https://doi.org/10.1016/j.jag.2014.07.002.

Zhong, L., P. Gong, and G. S. Biging. 2014. "Efficient Corn and Soybean Mapping with Temporal Extendability: A Multi-Year Experiment Using Landsat Imagery." *Remote Sensing of Environment* 140:1–13. https://doi.org/10.1016/j.rse.2013.08.023.

Zhong, L., L. Hu, L. Yu, P. Gong, and G. S. Biging. 2016. "Automated Mapping of Soybean and Corn Using Phenology." *ISPRS Journal of Photogrammetry and Remote Sensing* 119:151–164. https://doi.org/10.1016/j.isprsjprs.2016.05.014.

Zhong, L., L. Hu, and H. Zhou. 2019. "Deep Learning Based Multi-Temporal Crop Classification." *Remote Sensing of Environment* 221:430–443. https://doi.org/10.1016/j.rse.2018.11.032.

Zhong, L., L. Yu, X. Li, L. Hu, and P. Gong. 2016. "Rapid Corn and Soybean Mapping in US Corn Belt and Neighboring Areas." *Scientific Reports* 6 (1): 1–14. https://doi.org/10.1038/srep36240.