Science of Remote Sensing 7 (2023) 100085

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Utility of daily 3 m Planet Fusion Surface Reflectance data for tillage practice mapping with deep learning

Dong Luo [a], Hankui K. Zhang [a,*], Rasmus Houborg [b], Lina M.N. Ndekelu [a], Maitiniyazi Maimaitijiang [a], Khuong H. Tran [a], John McMaine [c]

[a] Geospatial Sciences Center of Excellence, Department of Geography and Geospatial Sciences, South Dakota State University, Brookings, SD, 57007, USA
[b] Planet Labs PBC, San Francisco, CA, 94107, USA
[c] Department of Agricultural and Biosystems Engineering, South Dakota State University, Brookings, SD, 57007, USA

ARTICLE INFO

Keywords:
Planet fusion
Tillage practice
Tillage date
Deep learning classification
Deep learning interpretation

ABSTRACT

Tillage practices alter soil surface structure that can be potentially captured by satellite images with both high spatial and temporal resolution. This study explored tillage practice mapping using the daily Planet Fusion surface reflectance (PF-SR) gap-free 3 m data generated by fusing PlanetScope with Landsat-8, Sentinel-2 and MODIS surface reflectance data. The study area is a 220 × 220 km^2 agricultural area in South Dakota, USA, and the study used 3285 PF-SR images from September 1, 2020 to August 31, 2021. The PF-SR images for the surveyed 433 fields were sliced into 10,747 training (70%) and evaluation (30%) non-overlapping time series patches. The training and evaluation patches were from different fields for evaluation data independence. The performance of four deep learning models (i.e., 2D convolutional neural networks (CNN), 3D CNN, CNN-Long short-term memory (LSTM), and attention CNN-LSTM) in tillage practice mapping, as well as their sensitivity to different spatial (i.e., 3 m, 24 m, and 96 m) and temporal resolutions (16-day, 8-day, 4-day, 2-day and 1-day) were examined. Classification accuracy continuously increased with increases in both temporal and spatial resolutions. The optimal models (3D CNN and attention CNN-LSTM) achieved ~77% accuracy using 2-day or daily 3 m resolution data as opposed to ~72% accuracy using 16-day 3 m resolution data or daily 24 m resolution data. This study also analyzed the feature importance of different acquisition dates for the two optimal models. The 3D CNN model feature importances were found to agree well with the tillage practice time. High feature importance was associated with observations during the fall and spring tillage period (i.e., fresh tillage signals) whereas the crop peak growing period (i.e., tillage signals weathered and confounded by dense canopy) was characterized by a relatively low feature importance. The work provides valuable insights into the utility of deep learning for tillage mapping and change event time identification based on high resolution imagery.

1. Introduction

Tillage, an agricultural activity to disturb soil and to reduce crop residue, has been used for thousands of years to facilitate seed bed preparation and reduce weeds and insects. However, tillage has also been found to accelerate water and nutrition loss in soil, and negatively affect carbon sequestration due to the increase surface runoff caused by tillage soil disturbance (Giglio et al., 2006; Mishra et al., 2010; Roger-Estrade et al., 2010; Zheng et al., 2014). More sustainable conservation tillage practices have been instituted to reduce soil disturbance. Studies have shown that conservation tillage practice can benefit sustainable agriculture by reducing carbon dioxide emission and soil erosion and increasing carbon sequestration (Alvarez and Alvarez, 2005; Busari et al., 2015; Melero et al., 2009; South et al., 2004). Consequently, the conventional tillage has shifted to conservation tillage to preserve more crop residue for long-term soil health, e.g., to reduce water and nutrition loss and to sequestrate more carbon. Tillage can be categorized into different types based on the percentage of crop residue covered on the topsoil, i.e., conventional tillage (less than 15% crop residue cover), reduced tillage (between 15 and 30% crop residue cover), and conservation tillage (at least 30% of crop residue cover) and no tillage (100% of crop residue cover) (USEPA, 2018).

Accurate mapping of different practices is needed for large scale assessment of erosion risk, agricultural carbon sequestration, and water

* Corresponding author.
E-mail address: hankui.zhang@sdstate.edu (H.K. Zhang).

https://doi.org/10.1016/j.srs.2023.100085
Received 15 March 2023; Received in revised form 19 April 2023; Accepted 23 April 2023
Available online 26 April 2023

quality altered by runoff of agricultural inputs. Studies also shows tillage can impact the crop yield (De Vita et al., 2007; Deines et al., 2019). Currently, data from medium spatial resolution satellite sensors such as Landsat have been used to identify low intensity and high intensity tillage or conservation tillage (Azzari et al., 2019; Beeson et al., 2020; Gao et al., 2022; South et al., 2004; Watts et al., 2009, 2011; Wang et al., 2022). However, one major drawback of medium spatial resolution (e.g., 30 m) remote sensing data is the inability to properly resolve texture information associated with different tillage practices. More recently, high or very high spatial resolution (i.e., less than 10 m) sensors (e.g., WorldView 2/3 and PlanetScope) have become a potential resource for tillage practice mapping (Quemada et al., 2018; M. Liu et al., 2022). For example, Hively et al. (2018) used WorldView 3 data with short-wave infrared (SWIR) band reflectance indices to analyze tillage systems, and their results showed promise in using WorldView data. Another study used an Unmanned Aerial Vehicle (UAV) system to detect conventional tillage and no-tillage practices in a cotton crop at the beginning of the growing season (Ashapure et al., 2019). These high spatial resolution datasets can effectively capture spatial texture information of different tillage practices. However, limited temporal resolution remains an issue because of frequent cloud contamination and difficulties associated with identifying optimal data acquisition dates for tillage mapping, which is further complicated by spatial variations in tillage time and rapid weathering of crop residue (Pacheco and McNairn, 2010; Zheng et al., 2014). Some studies have suggested that images acquired immediately before and after tillage is critical for tillage classification (Watts et al., 2011), which suggests a value add in having access to frequent time series data. Notably, Azzari et al. (2019) and Beeson et al. (2020) used Landsat time series data to map low-intensity and high-intensity tillage practices in the north central US region by compositing Landsat 5, 7, and 8 images to overcome cloud contamination issues. In addition, identification of tillage date is important as tillage time is important for soil moisture preservation (Arvidsson and Bölenius 2006; Tarkalson et al., 2016), weed control (Teasdale and Mirsky 2015; Smith 2006), and crop yield (Mamkagh et al., 2009; Wang et al., 2018). While high temporal resolution images can potentially provide information for tillage mapping, there is no studies utilizing such information for tillage date identification.

The PlanetScope constellation with 200+ CubeSats in low earth orbits delivers daily high spatial resolution images, which could prove valuable for tillage classification purposes. This data has the unique ability to track both high spatial and high temporal resolution surface changes, thus increasing the potential applications in the remote sensing community (Roy et al., 2021). However, due to sensor interoperability issues, cross-calibration challenges, and atmospheric contamination, deriving daily and temporally consistent analysis ready data from the PlanetScope constellation is associated with substantial challenges. Recently, Planet Labs PBC has implemented and improved the cubesat enabled spatio-temporal enhancement method (CESTEM) (Houborg and McCabe 2018) to cross-calibrate, harmonize, and fuse multi-sensor data streams leveraging publicly accessible satellites such as Sentinel-2, Landsat 8/9, Moderate Resolution Imaging Spectroradiometer (MODIS), and Visible Infrared Imaging Radiometer Suite (VIIRS) in concert with high resolution PlanetScope data (Planet Fusion Team, 2022). The resulting Planet Fusion product (PF-SR) delivers daily, gap free, radiometrically accurate, and temporally consistent 4-band surface reflectance (SR) data.

Machine learning has been extensively used to classify time series remote sensing data (Townshend et al., 1991; Wulder et al., 2018). Deep learning as a branch of machine learning is emerging as an effective approach in the remote sensing community for land use and land cover change mapping and analysis, vegetation classification, and agricultural applications (Chen et al., 2020; Kamilaris and Prenafeta-Boldú, 2018; Ma et al., 2019, 2022; Xu et al., 2018; Yan et al., 2022; Yuan et al., 2020). Compared to traditional machine learning, deep learning algorithms can take an image patch (or image patch time series) as input and learn its spatial and/or temporal information content because of its ability to learn hierarchical representations of features. Convolutional Neural Networks (CNNs), for example, have been successfully used to extract features from imagery by considering the spatial correlation of pixels in the image (Kattenborn et al., 2021; Thorp and Drajat, 2021; Turkoglu et al., 2021; Wang et al., 2018). Recent studies have adapted CNNs (such as one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D)) to process and interpret time series satellite images (Ben Hamida et al., 2018; Kattenborn et al., 2021; Xu et al., 2018). At the same time, recurrent neural networks such as Long Short-Term Memory (LSTM) have been adapted to process time series information at various spatial scales (Ienco et al., 2019; Rußwurm and Körner, 2020; Xu et al., 2021). For instance, LSTM has been used to estimate paddy rice production stages (Thorp and Drajat, 2021), and for land cover mapping (Ienco et al., 2019) using time series of Sentinel-2 images. Other studies have combined CNN and LSTM algorithms to classify time series remote sensing images (Masolele et al., 2021; J. J. Sun et al., 2019; Turkoglu et al., 2021). With the development of deep learning algorithms, the attention mechanism (Bahdanau et al., 2015; Luong et al., 2015) has emerged as an attempt to further improve the performance of the models (Li et al., 2020; Ofori-Ampofo et al., 2021). Various studies have applied the attention mechanism into deep learning models such as CNN and LSTM, demonstrating improved land cover classification accuracies (Masolele et al., 2021; Xu et al., 2021). Nevertheless, studies applying these deep learning algorithms to classify tillage practices do not currently exist to the best of our knowledge.

The objectives of this study are to (i) examine the feasibility of the Planet Fusion daily 3 m time series data for tillage practice classification; (ii) investigate the performance of different deep learning models for tillage practices classification and sensitivity to changes in spatial and temporal resolutions; and (iii) examine the deep learning interpretation capability for tillage time information extraction. This study used 3285 PF-SR daily images covering a 220 × 220 km^2 area nearby Sioux Falls, South Dakota, USA from September 1, 2020 to August 31, 2021. The 433 fields were surveyed to get ground truth tillage data and the PF-SR images for the surveyed fields were sliced into 10,747 non-overlapping time series patches each containing 365 × 32 × 32 3 m pixels. 70% and 30% patches were used for model training and evaluation, respectively, and the training and evaluation patches were from different fields for evaluation data independence. The performance of four deep learning models (i.e., 2D convolutional neural networks (CNN), 3D CNN, CNN-Long short-term memory (LSTM), and attention CNN-LSTM) in tillage practice mapping, as well as their sensitivity to different spatial (i.e., 3 m, 24 m, and 96 m) and temporal resolutions (16-day, 8-day, 4-day, 2-day and 1-day) were examined. This study also analyzed the feature importance of different acquisition dates for the best models.

2. Study area, field survey data, PlanetScope data and land cover data

2.1. Study area and field survey data for training and evaluation

The study area is located near Sioux Falls in Minnehaha County of South Dakota (Fig. 1), covering an area of 2110 km^2 growing predominantly corn and soybean. The area is characterized by hot and humid summers and cold and dry winters. The monthly mean temperature ranges from −7.8 to 23.6 °C and the mean annual precipitation is ~707 mm. Farmers have used four different tillage practices in the region, i.e., conventional tillage, reduced tillage, mulch tillage, and no tillage (USDA-NRCS, 2019). The ground truth tillage data was collected in June 2021 for 460 fields along a route set up by the USDA Natural Resources Conservation Service (NRCS) for annual conservation farming survey (Fig. 1). The dates were selected by the NRCS field survey experts and corresponded to the crop seedling stage. Both crop type and crop residue can be identified at this stage and all the tillage activities have been

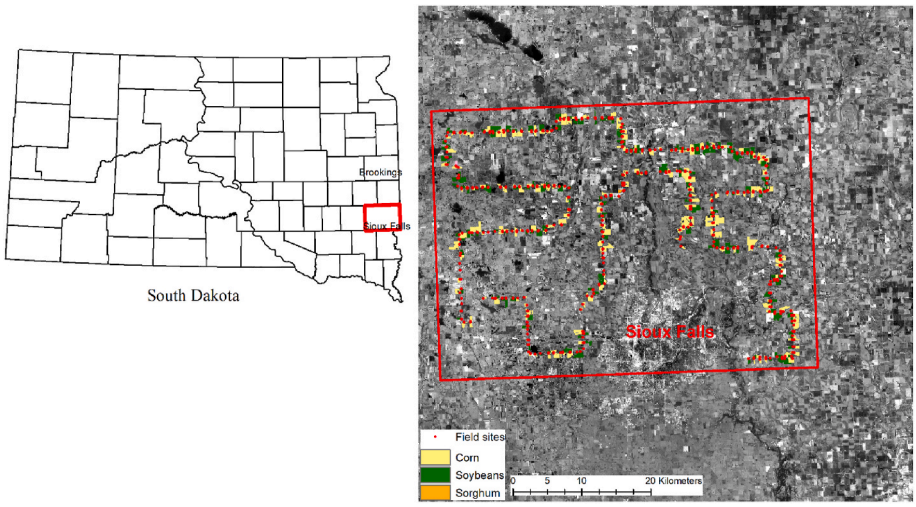


Fig. 1. The geographic location of the study area and surveying fields (red dots), and different crop type polygons used in the study. The background image is the Planet Fusion Surface Reflectance red band image acquired on September 15, 2020. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

completed for the growing season. Specifically, tillage mapping researchers and NRCS survey experts drove along the route and visited the 460 selected fields to record the tillage and crop types and to take photos. The tillage type was identified by visually determining how much crop residue was left in the field. After all the field data were collected, the researchers re-examined the tillage type record and corresponding photos for all the fields and removed those with potentially erroneous and ambiguous tillage types. Consequently, 433 out of the 460 fields were retained as training and evaluation samples. Each field was associated with a field boundary polygon using crop field boundaries generated by segmentation of 30 m Landsat annual time series and Crop Data Layer (CDL) data by Yan and Roy (2016) (https://lcluc.umd.edu/metadata/conterminous-united-states-conus-field-extraction). The field boundary polygons were then manually refined and edited using high spatial resolution images available in Google Earth Pro and Environmental Systems Research Institute (ESRI) World Imagery layer (Fig. 1). Among the 433 fields (i.e., polygons), conventional tillage, reduced tillage, mulch tillage, and no tillage was conducted in 71, 150, 107, and 105 fields, respectively. Fig. 2 shows example fieldwork photos and the corresponding Planet Fusion surface reflectance data for each type of tillage practice.

2.2. Planet fusion daily surface reflectance data

The Planet Fusion Surface Reflectance (PF-SR) product was used, which has 3 m spatial resolution and daily temporal resolution. Planet Fusion adopts an implementation of the CubeSat-Enabled Spatio-Temporal Enhancement Method (CESTEM; Houborg and McCabe, 2018) to leverage rigorously calibrated publicly accessible multispectral satellites (i.e., Sentinel-2, Landsat 8/9, MODIS, VIIRS) to work in concert with the higher spatial and temporal resolution data provided by Planet's medium resolution constellation of 200+ CubeSats. CESTEM is used to rigorously harmonize multi-sensor spectral data into a consistent radiometric surface reflectance standard for full fleet interoperability. The Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE version 3.7.7; Frantz, 2019) generates a harmonized Sentinel-2 and Landsat 8/9 BRDF adjusted Surface Reflectance (SR) product that is used as the cross-calibration reference target during the radiometric harmonization step. Planet Fusion processing also includes advanced temporally driven functionality related to geometric harmonization, cloud and shadow masking, gap-filling, and temporal filtering (Planet Fusion Team, 2022). The product contains the Planet Fusion Surface Reflectance data (PF-SR) and associated Quality Assurance data (PF-QA).

Planet Fusion delivers 4-band (blue: 0.45–0.51 μm, green: 0.53–0.59 μm, red: 0.64–0.67 μm, and near infrared: 0.85–0.88 μm) Surface Reflectance data in regularly gridded Universal Transverse Mercator (UTM) tiles each 24 km × 24 km (i.e., 8000 × 8000 3 m pixels). The study area intersected a total of 9 tiles and data were acquired from September 1st, 2020 to August 31st, 2021 to cover the full tillage operation, which is undertaken between the harvest of the previous growing cycle crop and the following sowing/cultivation operation (USEPA, 2018). A total of 3285 images were acquired through an author's affiliation to the Planet Labs PBC. The images are available to the public but there is a charge for accessing them.

2.3. Cropland data layer (CDL) data

The USDA cropland data layer (CDL) data was used to identify what crops have been planted in the year 2020. The year 2020 planted crops determine the crop residue spatial pattern, coverage amount in fall 2020 to spring 2021 and can potentially impact the tillage classification for our study period (September 1st, 2020 to August 31st, 2021). The year 2020 CDL crop types were thus used to evaluate the tillage classification accuracy for residue of each crop type. Note that CDL is not used as input for tillage classification as this study aims to examine how the PlanetScope images alone can classify tillage practices. The CDL product is an ongoing project sponsored by the USDA National Agricultural Statistics Service (NASS) (Boryan et al., 2011), and the product is based on training data derived from annual farmers' survey. The producer's and user's accuracies of CDL are higher than 96% for major crop types. Based on the CDL 2020 data, the study area field polygons were characterized by a mix of corn, sorghum, soybean, rye, oats, millet, alfalfa, and grassland with the majority being corn (231 fields) and soybean (193 fields) (Fig. 1).

3. Method

3.1. Overview

In this study, the four tillage practices (i.e., conventional tillage, reduced tillage, mulch tillage, and no tillage) were aggregated into two tillage categories (conventional and conservation) for classification. The two tillage categories classification conforms well with previous tillage classification studies in the literature (Ashapure et al., 2019; Azzari et al., 2019) and is consistent with the conventional and conservation

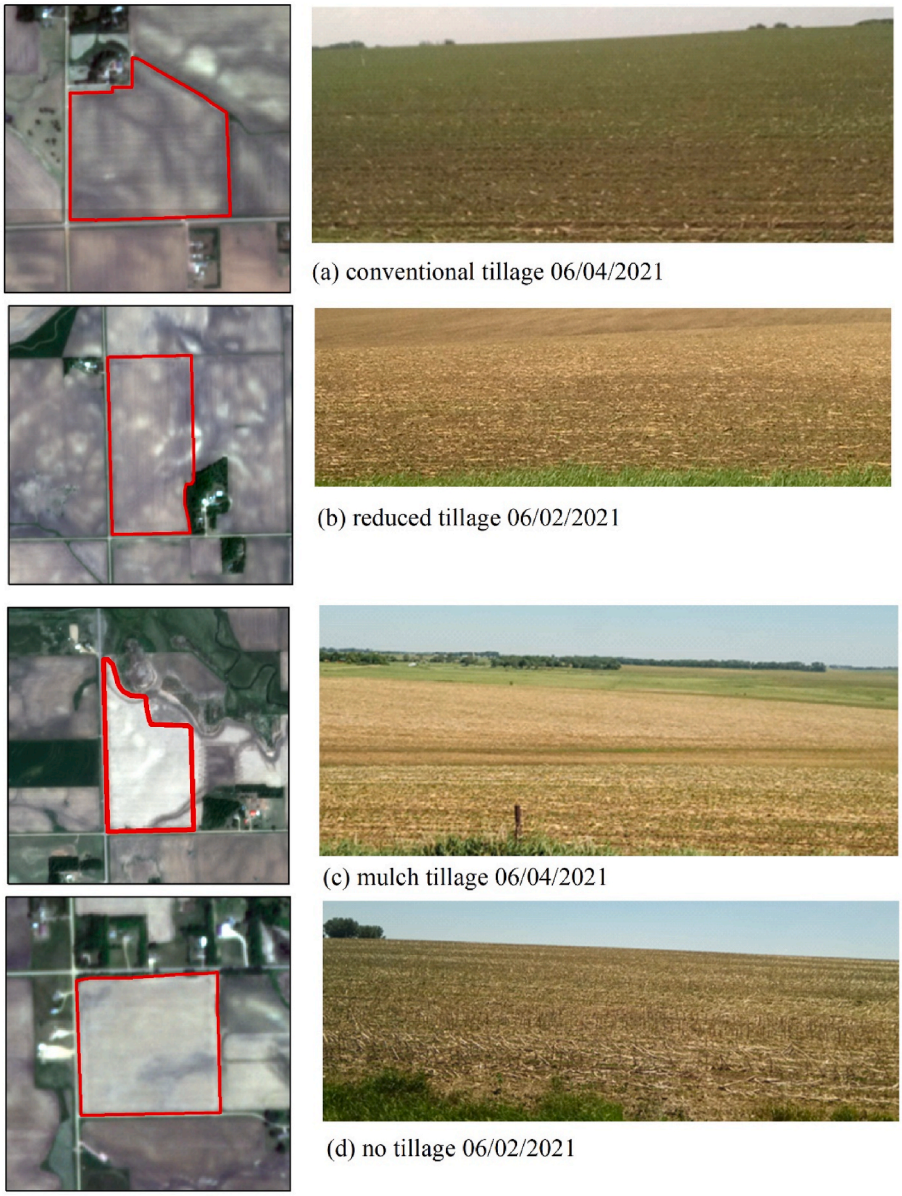


Fig. 2. Examples of Planet Fusion Surface Reflectance (RGB) imagery (left) and roadside photos (right) for conventional tillage practice (a), reduced tillage practice (b), mulch tillage practice (c) and no tillage practice (d).

tillage definition in sustainable agriculture. Specifically, the conventional tillage category is defined as <30% crop residue left on the surface after tillage such as mold board, disk, or chisel plow methods and thus includes conventional and reduced tillage practices. The conservation tillage category is defined as ≥30% crop residue left and includes mulch tillage and no tillage practices. All available bands (i.e., blue, green, red, and NIR) from the daily Planet Fusion Surface Reflectance product were used for classification.

Fig. 3 shows the flowchart of the proposed method that includes three major steps. (i) The first step is the extraction of training and evaluation time series PF-SR patches each with 32 × 32 × 365 × 4 reflectance values, i.e., 32 × 32 3 m pixels, 365 dates and 4 spectral bands. (ii) The second step is to use deep learning methods with capability to classify the four dimensional PF-SR time series patches (32 × 32 × 365 × 4) into a class label. We explored four deep learning models, i.e., 2D convolutional neural network (CNN), 3D CNN, CNN long short-term memory networks (LSTM), and attention CNN-LSTM with such capability. In addition, the sensitivity of the models to different spatial and temporal resolutions was analyzed by simulating a series of datasets with reduced spatial and temporal resolutions relative to the native PF-SR data. (iii) Finally, temporal feature importances were calculated from the trained deep learning models to estimate the timing of tillage practices. The principle behind this is that the important features (dates) should be those right after the tillage practice when the tillage signal is fresh and has not been weathering.

3.2. Training and evaluation data

The convolutional neural network (CNN) based deep learning models usually take as training input an image patch or image patch time series with multispectral bands and multiple dates ($n \times n \times t \times s$) (with n being the spatial patch window size, $s = 4$ spectral bands and $t =$ 365 dates in this study). The trained model will then produce output in the form of class labels (e.g., tillage type) for the image patch. In this study, the field polygons with the samples have various irregular shapes and areas ranging from 32,540 m^2 to 1,495,390 m^2. Fixed-size patches were extracted from each individual field polygon (Fig. 1) by slicing a 32 × 32 ($n = 32$) window across each polygon with step size 32 (i.e., no

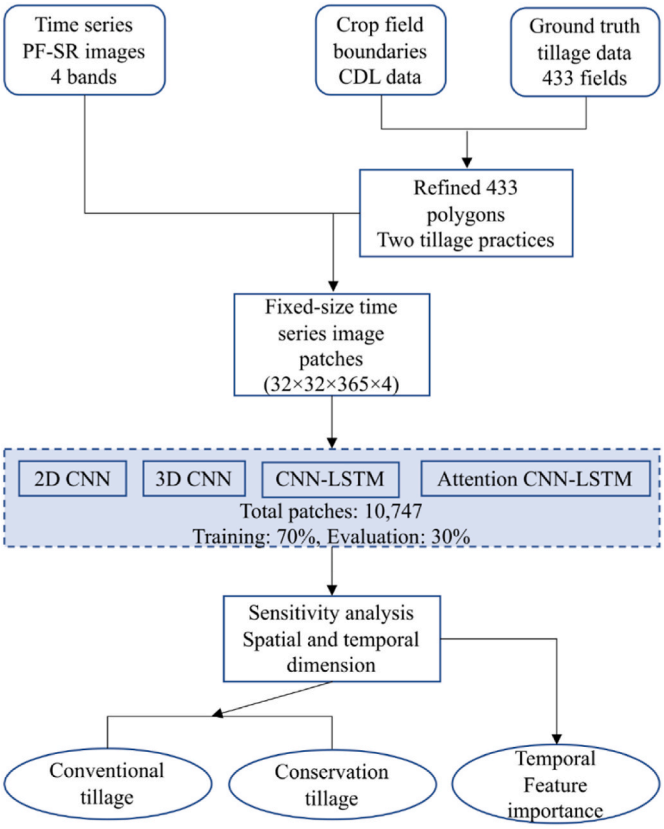


Fig. 3. Flowchart of the proposed method. The CNN and LSTM represent convolutional neural network and long short-term memory, respectively.

overlap between patches). Patches at the polygon edges were kept if at least 95% of the pixels were inside the field polygon. A total of 10,747 patches were generated in this way and subdivided into conventional (5722) and conservation (5025) tillage patches. There were 5821 and 4842 patches planted in 2020 as corn and soybean, respectively.

About 70% (7429 patches) and 30% (3008 patches) of the patches were used for training and evaluation of deep learning models, respectively. It is well established that evaluation patches coming from the same field with the training patches may boost the evaluation accuracy due to the training and evaluation data correlation. In this study, the training and evaluation patches were selected from different field polygons to avoid boosting the evaluation data accuracy. The 433 field polygons were first randomly split into training and evaluation polygons. All patches in the training and evaluation field polygons were used for training and evaluation, respectively (i.e., a field polygon was never used for both training and evaluation). 4% (310 patches) of training patches (sometimes termed as validation data) were randomly selected and used to tune the hyper-parameters of the deep learning models.

3.3. Deep learning models

Fig. 4 shows the four deep learning models used in this study that are two-dimensional (2D) CNN (Krizhevsky et al., 2012), three-dimensional (3D) CNN (Ji et al., 2013), 2D CNN-long short-term memory (LSTM) and attention 2D CNN-LSTM (Ma and Hovy, 2016; Interdonato et al., 2019; Masolele et al., 2021; Thorp and Drajat, 2021; Turkoglu et al., 2021; Wang et al., 2022). All the models take each time series patch with 32 × 32 × 365 × 4 reflectance values as input and classify it into a tillage type (either conventional or conservation). However, they are slightly different in how the input data are organized. The 2D CNN reshaped the 4D input time series image patch into a 3D image 32 × 32 × 1460. The 3D CNN directly took the 32 × 32 × 365 × 4 time series patch. The CNN-LSTM and attention CNN-LSTM need the input patch reshaped to be 365 × 32 × 32 × 4 to align with the structure requirement. The conventional CNN method classifying each patch to a single label is used rather than the fully convolutional networks (FCN) that can classify each pixel in the patch (Long et al., 2015; Ronneberger et al., 2015). This is because all pixels within a patch usually use the same tillage practice considering that the patch size (96 m × 96 m = 0.009 km^2) is much smaller than the typical US field size with mean value of 0.193 km^2 (Yan and Roy 2016).

3.3.1. 2D CNN

2D CNN can learn spatial correlation among pixels and normally takes three-dimensional image patches as input (e.g., 32 × 32 × *s* with *s* being the third dimension, normally the spectral bands in remote sensing image). The 4-dimensional input patch time series (32 × 32 × *t* × 4) was flattened into 3-dimensions by collapsing the spectral (*s*: 4 in this study) and temporal (*t*) dimensions into one dimension with length *t* × *s*, i.e., by stacking all the images acquired at *t* different dates into a single image with *t* × 4 bands. The 2D CNN usually contains 2 to hundreds of convolutional layers and followed by 2–3 fully connected layers. Each convolutional layer is usually associated with several (e.g., 64) convolutional filters/kernels (i.e., three-dimensional matrix with learnable weight and bias) to convolute with the 3D images. The convolutional layer generates the same number of 2D feature maps as the number of filters and these 2D feature maps can be considered as one single 3D feature map used as input for the next convolutional layer. The last CNN layer feature maps are flattened into a 1D feature vector for the fully connected layers. Each fully connected layer has a 2D weight matrix and 1D bias vector coefficients to linearly transform the input 1D-feature vector to generate another 1D feature vector. The last fully connected layer is used to predict class labels.

In this study, each of the four convolutional blocks in the 2D CNN model consisted of one convolutional layer, one batch normalization layer and one max pooling layer (Table 1). Batch normalization of feature map using mean and standard deviation has proven effective towards avoiding gradient issue and potentially improve model performance (Ioffe and Szegedy, 2015). The max pooling layer applies a sliding window across the feature maps and derives a single max value in each window to suppress irrelevant information. The filters in each 2D convolutional layer were set as: 64, 128, 256, 512; and filter size in each convolutional layer was 3 × 3. The first fully connected layer has 512 neutrons and the second fully connected layer has 2 so that the output has 2 values to represent class confidence for each class. To avoid overfitting, one dropout layer with 0.5 drop rate was added to the first fully connected layer (Fig. 4 and Table 1). This structure was built by testing a range of convolutional layer numbers (3,4,5,6) and the first convolution layer filter numbers (32, 64 and 128) and is consistent with the optimal used by Masolele et al. (2021) in the classification of tropical deforestation using Landsat 5 and 7 time series imagery over six years. The number of convolutional layer filters is doubled after each convolutional block following the conventional deep CNN models.

3.3.2. 3D CNN

The 3D CNN (Ji et al., 2013) is different from the 2D CNN as each convolutional layer contains 4D filters that are directly applied to the input 4D image (32 × 32 × *t* × 4) and derives a 4D feature map. Similarly, the 3D CNN has batch normalization, pooling, and fully connected layers. Due to the feature map dimensions, the 3D pooling layer replaces the 2D pooling and uses a 3D window to slide across the feature maps and derives a max value in each window. Then the 4D feature maps of the last convolutional layer were flattened to a 1D vector and fed to the fully connected layers. The structure of the 3D CNN model used in the study consisted of four 3D convolutional blocks. Each block was composed of one 3D convolutional, one batch normalization, and one 3D max pooling layer. The number of filters in each 3D convolutional layer was the same as the 2D CNN model (i.e., 64, 128, 256, 512) and the

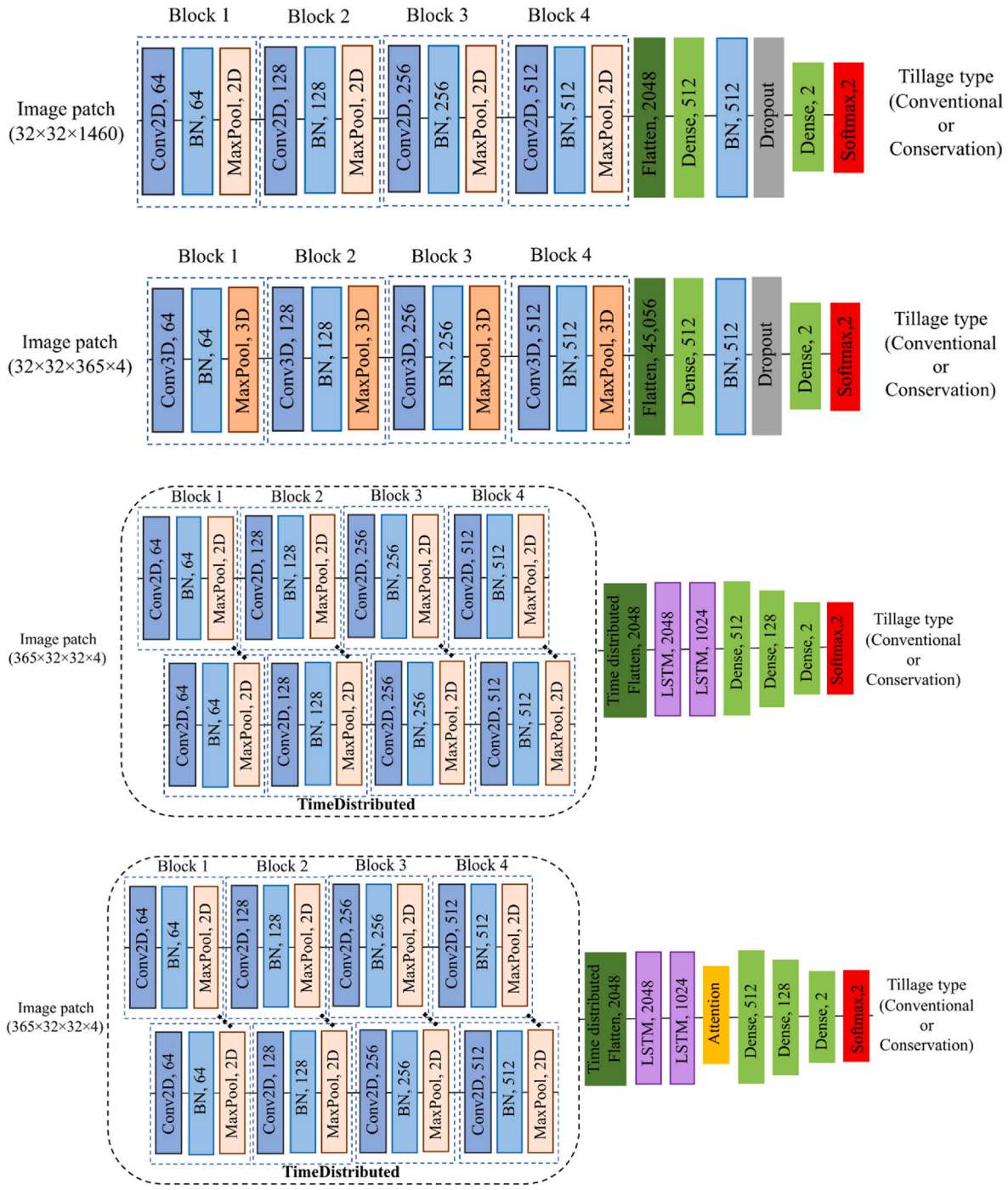


Fig. 4. The input, structure and output of the four deep learning models: the two-dimensional (2D) CNN (first row), three-dimensional (3D) CNN (second row), 2D CNN- long short-term memory (LSTM, third row) and attention 2D CNN-LSTM (the last row). Note despite all models take the same input data, the data organizations are different. BN represents batch normalization.

filter size was 3 × 3 × 3. Similar to the 2D CNN model, the flattened feature map also had two fully connected layers (i.e., 512 and 2 neurons, respectively) and one batch normalization layer and one dropout layer with 0.5 drop rate were added to the first fully connected layer to avoid overfitting (Fig. 4 and Table 1).

3.3.3. CNN-LSTM

The CNN-LSTM model is a combination of the 2D CNN and recurrent neural network (RNN) that can directly take the $t \times 32 \times 32 \times 4$ image patch time series as input. The 2D CNN is used to transform each image ($32 \times 32 \times 4$) into a feature vector (e.g., with length l') and the time series feature vectors ($l' \times t$) will then be fed into RNN that was developed to handle sequential data (Bengio Y, et al., 1994). The hidden layer features at time t are derived as a function of both the previous layer feature values and the current layer feature values at previous time step. The hidden layer features at time t bear all the information from all the past steps and the hidden layer features at the last time step were used for classification. Due to the structure, the calculated gradient values in RNN training can be extremely small (gradient vanishing) or large (gradient exploding), which will stop meaningful model coefficient updates/training.

To avoid gradient vanishing and exploding issues, Long Short-Term

Table 1
The structures of the deep learning models used in the study. The two numbers following the model name are the total number of model parameters for the 365-day and 16-day classification, respectively. The other numbers indicate the convolutional filters and filter size for the convolutional layers, the number of hidden units for the fully connected and LSTM layers, the drop rate for the dropout layer, the number of hidden units and heads for the multi-head attention layer. A max pooling and a batch normalization layer follow each convolutional layer.

Layers	2D CNN	3D CNN	CNN-LSTM	Attention CNN-LSTM
	(3.4 million)	(27.7 million)	(48.3 million)	(55.1 million)
	(2.6 million)	(5.7 million)	(48.3 million)	(55.1 million)
1st convolutional layer	2D (64, 3)	3D (64, 3)	2D (64, 3)	2D (64, 3)
2nd convolutional layer	2D (128, 3)	3D (128, 3)	2D (128, 3)	2D (128, 3)
3rd convolutional layer	2D (256, 3)	3D (256, 3)	2D (256, 3)	2D (256, 3)
4th convolutional layer	2D (512, 3)	3D (512, 3)	2D (512, 3)	2D (512, 3)
Flatten layer	Flatten	Flatten	Time Distributed Flatten	Time Distributed Flatten
1st LSTM layer	NA	NA	2048	2048
2nd LSTM layer	NA	NA	1024	1024
Multi-head attention	NA	NA	NA	Multi-head attention (1024, 8)
1st fully connected layer	512 with batch normalization		512	
Dropout	0.5		NA	
2nd fully connected layer	NA		128	
Last fully connected layer	2 with SoftMax activation			

Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Chung et al., 2014) algorithms have been developed and used for land use and land cover classification (Paoletti et al., 2018; Z. Z. Sun et al., 2019; Yuan et al., 2020). We used LSTM based deep learning algorithms (CNN-LSTM and Attention CNN-LSTM) in this study as we experimentally found little difference between the LSTM and GRU. The CNN-LSTM model in this study had four 2D convolutional layers and 2D max pooling layers (Fig. 4 and Table 1), i.e., similar to the plain 2D CNN model (Section 3.3.1), and followed by two LSTMs with 2048 and 1024 units, respectively. This structure is determined after testing different number of LSTM layers (1, 2, and 3) and different hidden units (512, 1024, and 2048). Then three fully connected layers with 512, 128 and 2 feature vectors were added to the output of the last LSTM layer to predict class labels as described for the plain 2D CNN model (Section 3.3.1).

3.3.4. Attention CNN-LSTM

The attention CNN-LSTM model was also used to examine the recent development of the attention models in handling time series data (Bahdanau et al., 2015; Luong et al., 2015). Attention mechanism has been used in the remote sensing community (Ienco et al., 2019) and have achieved a higher accuracy compared with non-attention RNN models (Masolele et al., 2021; Xu et al., 2021). Attention mechanism takes the time series hidden layer feature vectors as input to derive a context vector (or a weighting vector) that captures the relevant importance of the hidden state at each time step. The structure of the attention CNN-LSTM model was similar to the CNN-LSTM model but a multi-head attention mechanism was added to the second LSTM layer. In this study, a multi-head self-attention was adapted to combine the feature vectors returned by the second LSTM at each time step, which is a key component of the popular Transformer model (Vaswani et al., 2017). The hidden unit and head numbers of the attention layer were set as 1024 (same as the last LSTM unit) and 8 (after testing 4 and 8), respectively. The output of the multi-head self-attention layer also followed three fully connected layers as done for the CNN-LSTM model to predict class labels (Fig. 4 and Table 1).

3.3.5. Training parameters and evaluation metrics

All the deep learning models were implemented using Keras and TensorFlow. The rectified Linear Unit (ReLU) activation function (Glorot et al. 2011) was used in all the convolutional and fully connected layers (Table 1). The sparse categorical cross entropy was used as the loss function. The RMSprop optimizer was chosen over the Adam optimizer based on experimental results. A dynamic learning rate exponential decay schedule (Gotmare et al., 2019) was used in this study as studies have proved that learning rate schedule is better than a fixed learning rate (Li and Arora, 2019; Rußwurm and Körner, 2020). The initial learning rate was set as 0.0001 after testing three different rates (i.e., 0.00001, 0.0001, and 0.001) and all other optimizer parameters were set as default following the convention. The batch size was set as 32 (16, 24, and 32 tested) and the epoch was set as 50 (50, 100 and 150 tested). The experiments were undertaken on a server with 4 NVIDIA Tesla V100 PCIe GPUs each with 32 GB memory.

A confusion matrix was derived to compare ground observed and model predicted tillage practice data. The producer's accuracy, user's accuracy, and F1-score for each tillage practice were computed to quantitatively analyze classification results. In addition, the optimal model was used to classify all crop field polygons within the study area (i.e., 9 Planet image tiles each with 8000 × 8000 3 m pixels). Specifically, 62,500 time series image patches each with $32 \times 32 \times t \times 4$ reflectance values were exacted from 9 tile images. Then the optimal deep learning model was applied to each time series patch to classify it as conventional or conservation tillage practice. The class label for each CDL field polygon was derived as the majority of the classified labels of all the 32 × 32 patches intersected with the polygon.

3.4. Classification sensitivity to spatial and temporal resolution analysis

The PF-SR time series data were degraded to lower spatial and temporal resolution images to analyze the impact of the different resolutions on the tillage classification results. For the spatial resolution impact analysis the input 3 m pixel image patch time series ($32 \times 32 \times t \times 4$) was degraded to 24 m by 24 m ($4 \times 4 \times t \times 4$) and 96 m by 96 m ($1 \times 1 \times t \times 4$) using average pixel aggregation. A more advanced pixel degradation using a sensor specific point spread function (PSF) (Huang et al., 2002; Che et al., 2021) was not considered here as there is no sensor PSF specifications. The purpose of this study was to examine the spatial resolution effect rather than some lower spatial resolution data acquired by any specific sensor. The 3D CNN model was used to evaluate the spatial resolution sensitivity, and we have experientially confirmed that other models followed a similar pattern. For the 24 m image patch time series the same 3D CNN model (Table 1) was used by removing the max pooling layer from the first and second 3D convolutional blocks to adapt to the smaller input patch size. For the 96 m image patch time series, the spatial filter size was down-adjusted from $3 \times 3 \times 3$ to $1 \times 1 \times 1$ to equal a 1D CNN. This can be referred to as a 1D CNN model associated with a 1D convolutional layer and 1D max pooling layer.

For the temporal resolution analyses, the input time series data (i.e., a total of 365 days) were reduced into 6 temporal resolutions corresponding to data acquired on 16 (16-day temporal resolution), 32 (8-day), 64 (4-day), 128 (2-day), 256 (1-day), and 365 (1-day) days,

respectively. These days are simply selected from all the 365 days to ensure the selected dates are evenly distributed. Images acquired on the 50 earliest (from September 1, 2020 to October 20, 2020) and 59 latest (from July 4, 2021 to August 31, 2021) days were removed to ensure an even distribution of the selected days. As a result, a total of 6 datasets were created for each spatial resolution (Table 2). The sensitivity to temporal resolution was examined using all the four deep learning models introduced in Section 3.3. This is because the LSTM and 3D CNN use different techniques to handle time series and may respond differently to the length of the time series. For example, the LSTM model has been associated with performance degradation when translating sentences with >35 words by examining a series of word lengths from 4 to 80 (Sutskever et al., 2014). This is a limitation of the LSTM in processing long time series that is not an issue for the 3D CNN model.

3.5. Temporal feature importance analysis for tillage time identification

Tillage time plays an important role in soil thermal and moisture status and crop growth conditions (Mamkagh et al., 2009; Wang et al., 2018). There is no literature on how to detect tillage time using remote sensing time series data. One objective of this study was to identify tillage time via a temporal feature importance analysis given the dense time series of the Planet Fusion data and the fact that remote sensing images close to the tillage dates are more important for tillage classification (Watts et al., 2011). The feature importance analysis represents a core deep learning interpretation technique (Montavon et al., 2018; Voosen, 2017; Ge et al., 2022). In this study, the importance of the different dates were ranked to highlight critical dates for tillage classification via gradient backpropagation (Rußwurm and Körner, 2020; Xu et al., 2021). The gradients of deep learning models can be backpropagated to calculate partial derivatives of each input variable. The trained deep learning models were used to calculate feature importance based on class-specific Jacobian derivatives. The derivatives for each patch time series were a $32 \times 32 \times 256 \times 4$ array for each tillage class. The absolute derivative values were averaged over the spatial (32×32) and spectral (4) dimensions to use only the derivative values as a function of time as metrics of the temporal feature importance.

4. Results

4.1. Planet fusion reflectance time series for different tillage practices

Fig. 5 shows spatially averaged (solid line) surface reflectance time series representing conventional tillage (blue lines) and conservation tillage (red lines) practice samples for each of the four spectral bands (i.e., blue, green, red, and NIR). The associated standard deviations are indicated by the dashed lines. The smoothness of the seasonal band-specific reflectance dynamics demonstrates the high quality and temporal consistency of the PF-SR product with no rapid reflectance fluctuations due to cloud or snow contamination. As expected, the visible bands have much lower reflectance magnitudes compared to the NIR band in the growing season months (September 2020 and June, July and August 2021), which reflects the typical spectral signature of a healthy crop. The reflectance increased with wavelength from the blue to the NIR domain during the pre-growing season (March to April 2021), which corresponds with the typical spectral signature of soil (Jacquemoud et al., 1992).

Conservation tillage generally shows higher reflectance than conventional tillage during non-growing season months (i.e., from November 2020 to May 2021). This is expected as fields with conservation tillage have more crop residue relative to fields with conservation tillage (Fig. 2) combined with crop residue generally being more reflective than bare soil surfaces (Gao et al., 2022; Pacheco and McNairn, 2010). The reflectance time series during the growing season months (September 2020 and June, July, and August 2021) were very similar between conventional tillage and conservation tillage practices, as the reflectance were dominated by the dense crop canopy rather than the soil surface. Noteworthy is the large and partly overlapping reflectance standard deviations between the two tillage practices outside of the growing season (Fig. 5), which suggests that using simple reflectance magnitudes (i.e., spectral information) may not always be enough to accurately classify/separate conventional and conservation tillage practices.

4.2. Model accuracy and sensitivity to spatial and temporal resolutions

Fig. 6 shows the overall accuracy of the tillage classification as a function of the number of the temporal resolution for the four deep learning models. The 3D CNN, CNN-LSTM, and attention CNN-LSTM models performed better than the 2D CNN model, which is expected as the 2D CNN is not designed to process data with a third dimension. The flattening of the time series images into a single band for 2D CNN is evidently not an effective method for time series classification, which is also consistent with findings from Chen et al. (2020). The CNN-LSTM model and attention CNN-LSTM model performed better than the 3D CNN model for the lower temporal resolution cases (i.e., 16–128) with the attention CNN-LSTM slightly outperforming CNN-LSTM (Fig. 6). This is expected as the attention mechanism can assign day-specific importance scores to the different images and prioritize specific images.

The accuracies continuously increased with increasing number of images for classification up to 256, which reflect the critical importance of high temporal resolution for discriminating the two different tillage practices. However, beyond 256 images a drop in accuracy was observed for all the models (Fig. 6). A plausible explanation for the decreased accuracy may be that the 256-day period from October 21, 2020 to July 4, 2021 is simply more suitable for tillage classification relative to using the full year as the growing season period will provide limited additional information. Significant crop growth (green up started early June) will negatively affect tillage identification. The observed accuracy drop (from 256 to 365 days) was more pronounced for the LSTM based models relative to the CNN based models (Fig. 6). This may be because the LSTM can't handle long time series (>100 time steps) very well as it was originally developed in machine translation to process sentences with <100 words (Cho et al., 2014; Sutskever et al., 2014).

Table 3 shows the confusion matrix (i.e., overall model accuracies) from the attention CNN-LSTM model using 256 PF-SR images. The user's and producer's accuracies were all greater than 68%. Accuracies were higher for conventional relative to conservation tillage practice for the producer's accuracy, user's accuracy, and F1-socre (Table 3). This may be because of a more pronounced soil disturbance during tillage operation (Fig. 2). Table 3 also breaks down the confusion matrix for corn and soybean crop residue. For corn, the producer's accuracies for both conventional and conservation tillage were similar (i.e., ranging from 70% to 74%). However, user's accuracies were significantly higher

Table 2
The input data dimensions of the reduced spatial and temporal resolution ($n \times n \times t \times s$).

Number of acquisition days (Temporal resolution)	16 days (16-day temporal resolution)	32 days (8- day temporal resolution)	64 days (4- day temporal resolution)	128 days (2- day temporal resolution)	256 days (daily temporal resolution)	365 days (daily temporal resolution))
3 m	$32 \times 32 \times 16 \times 4$	$32 \times 32 \times 32 \times 4$	$32 \times 32 \times 64 \times 4$	$32 \times 32 \times 128 \times 4$	$32 \times 32 \times 256 \times 4$	$32 \times 32 \times 365 \times 4$
24 m	$4 \times 4 \times 16 \times 4$	$4 \times 4 \times 32 \times 4$	$4 \times 4 \times 64 \times 4$	$4 \times 4 \times 128 \times 4$	$4 \times 4 \times 256 \times 4$	$4 \times 4 \times 365 \times 4$
96 m	$1 \times 1 \times 16 \times 4$	$1 \times 1 \times 32 \times 4$	$1 \times 1 \times 64 \times 4$	$1 \times 1 \times 128 \times 4$	$1 \times 1 \times 256 \times 4$	$1 \times 1 \times 365 \times 4$

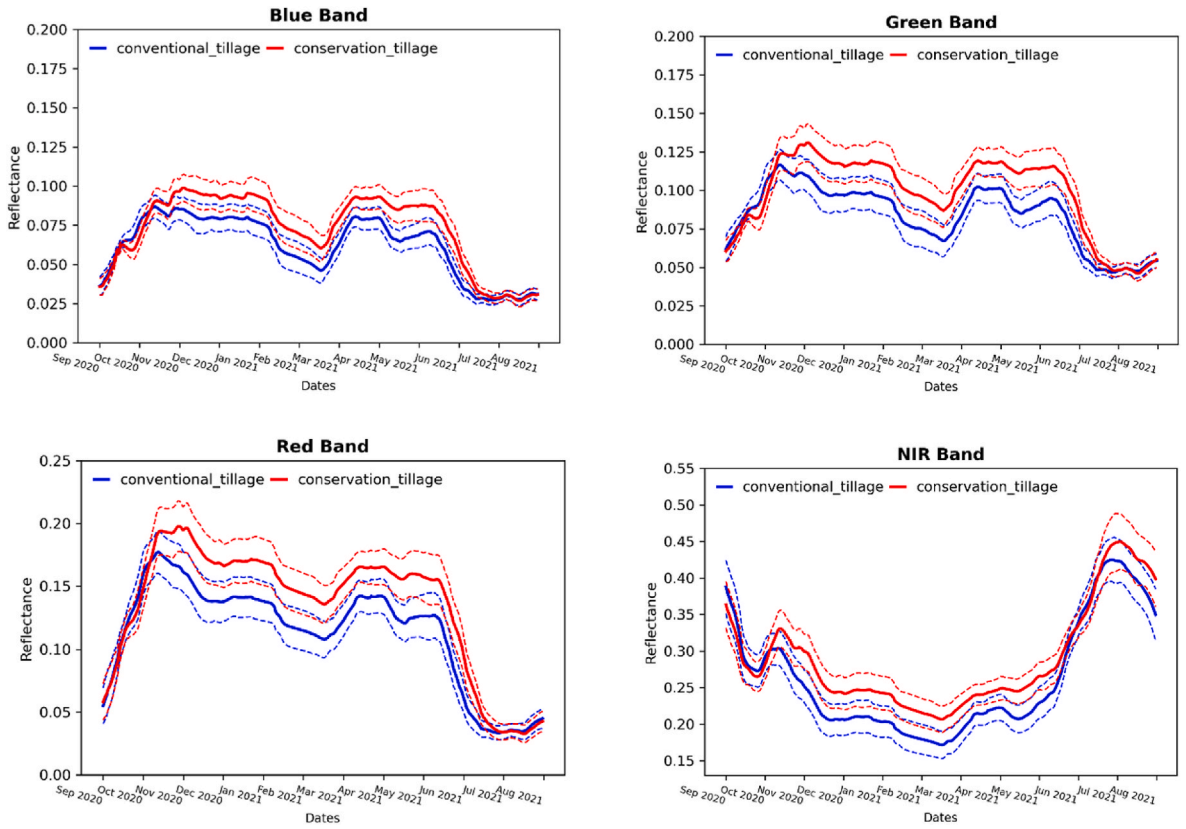


Fig. 5. Time series of band-specific surface reflectance mean (solid lines) and standard deviation (dashed lines) derived from all of the conventional (blue lines) and conservation (red lines) tillage practice samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

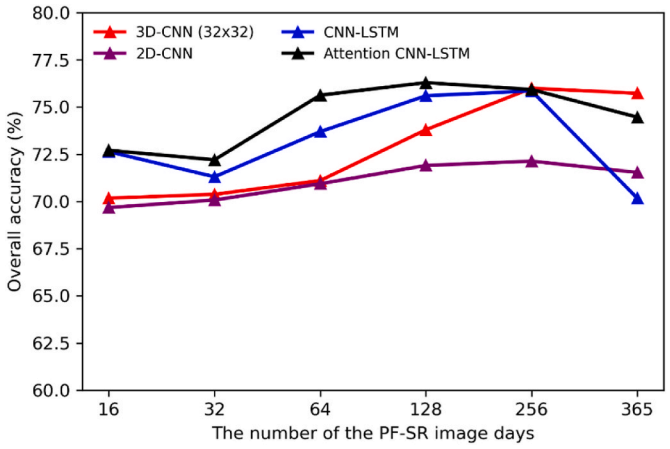


Fig. 6. Overall accuracy (%) of the evaluation dataset for different deep learning models as a function of the number of images (temporal resolution) used for classification.

(>20%) for conservation tillage. A similar tendency is observed for the F1-score value (Table 3). A reasonable explanation is that corn has more residue than soybean after harvest, which can lead to enhanced reflectance disturbances. For soybean, the producer's accuracy, user's accuracy, and F1-score were highest for the conventional tillage class. Compared to corn residue, soybean residue can be more difficult to identify and this could be the reason for the lower performances for conservation tillage. This is consistent with the previous findings by Pacheco and McNairn (2010).

Fig. 7 shows the model (3D CNN) overall accuracies of the tillage practice classification as a function of different spatial resolutions. The 3D CNN models using the original 3 m spatial resolution and degraded 24 m spatial resolution performed much better than the degraded 96 m spatial resolution dataset. It's a clear indication that a high spatial resolution is important for tillage classification with deep learning models. The 3D CNN model with the original 3 m spatial resolution imagery (32 × 32) consistently produced the highest overall accuracy, which is expected given the enhanced spatial information content relative to the coarser resolution experiments.

Table 4 shows the training and evaluation computation time (second) for each deep learning model. It is expected that with increasing number of images used for classification, the deep learning model training and evaluation time increased. Notably, although the 3D CNN model and attention CNN-LSTM model had similar overall accuracy, the 3D CNN model computation time for training (855.30–8102.50s) and evaluation (13.78–69.04s) is much less than the attention CNN-LSTM model (1512.26–12191.85s for training and 16.51–109.33s for evaluation). In general, CNN based deep learning models (2D CNN and 3D CNN) need much less computation time than LSTM based deep learning models (CNN-LSTM and attention CNN-LSTM). This is reasonable as noted in Section 3.3, in the CNN based models the features for all the image dates can be calculated in parallel. However, for the LSTM based models, the features at one date is the input of the features for another date and can only be calculated in sequence rather than in parallel.

Fig. 8 shows the tillage practice map in the study area (i.e., all the 9 PF-SR image tiles in Fig. 1). The result was generated by the 3D CNN model trained with 256 days of PF-SR images as the model performed the best. There were 36.81% and 63.19% cropland patches classified as conventional and conservation tillage, respectively. The results are in line with Beeson et al. (2020) and reflected that the USDA Environmental Quality Incentives Program (EQIP) has encouraged farmers to

Table 3
The classification confusion matrices (including producer's and user's accuracies and F1-score) for the attention CNN-LSTM model using 256 PF-SR images.

		Classification			
		Conventional	Conservation	Producer (%)	F1-score
All crops	Conventional	1400	311	81.82	0.7946
	Conservation	413	884	68.16	0.7095
	User (%)	77.22	73.97		
Corn	Conventional	446	163	73.23	0.6622
	Conservation	292	711	70.89	0.7576
	User (%)	60.43	81.35		
Soybean	Conventional	935	141	86.90	0.8842
	Conservation	104	172	62.32	0.5840
	User (%)	89.99	54.95		

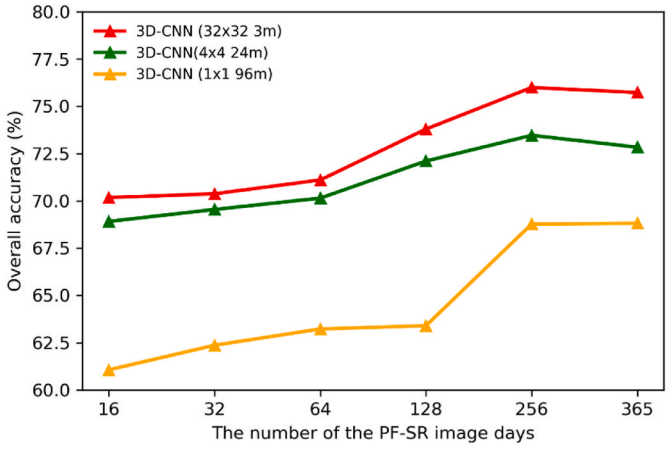


Fig. 7. Overall model accuracy (%) for the 3D CNN model for different spatial resolutions (3 m, 24 m and 96 m) as a function of the number of the PF-SR images used for classification. Note that the 96 m 3D CNN model can be regarded as 1D CNN model as a single pixel (1 × 1) was used as input.

adopt conservation practices in recent years. The conservation tillage practices (red color) is dominated the north and southwest of the Sioux Fall city. Conventional tillage practices occurred more often in the east part of the study area.

4.3. Temporal feature importance analysis for tillage time identification

The temporal feature importance of the best model in the LSTM family (i.e., attention CNN-LSTM model) and the best model in the CNN family (i.e., 3D CNN model) are shown in Figs. 9 and 10, respectively. The feature importance was very low at the beginning and end of the study period, which coincided with the growing season with the crop canopy covering the soil surface making it impossible to identify tillage practices. Peak feature importance is observed during November 2020 and May 2021 when the soil surface was fully exposed and tillage is usually undertaken. The two different tillage practices (i.e., conventional and conservation) showed similar temporal feature importance patterns. However, the temporal dynamics were smoother for the conservation tillage practice, which may be the result of reduced soil disturbance from this type of tillage practice.

Fig. 9 shows the date-specific feature importance of the attention CNN-LSTM model using 256 Planet Fusion images. Unlike the 3D CNN model, the attention CNN-LSTM exhibited very high feature importance towards the end of the study period. This may be because only the last time step hidden layer features is used for classification and the last time step features tend to 'forget' information of the beginning time steps. The feature importance was higher during the fall period (i.e., October and November 2020) compared to the winter fallow period (i.e., December 2020 to February 2021). This may indicate the 2020 fall tillage practices. Furthermore, the CNN-LSTM feature importance associated with conservation tillage practice was lower and less fluctuating compared to the conventional tillage practice (Fig. 9). This observation corresponds to the lower degree of soil disturbance associated with conservation tillage.

The 3D CNN model-based feature importances were more physically realistic compared to the attention CNN-LSTM model-based results (Fig. 10). The 3D CNN model creates a 3D filter to learn information from all temporal resolutions (i.e., 256 days) and uses 3D max pooling recursively to remove irrelevant information and retain the essential features from all temporal dimensions. In contrast, the LSTM model is developed from natural language processing (e.g., language translation) to handle sequential-to-sequential prediction (Devlin et al., 2019). Adapting the LSTM model to tillage classification of sequential-to-one prediction could result in unrealistically high feature importances at the end of a time series as only the hidden units of the last time step were used for classification. Furthermore, the feature importance results from the 3D CNN model were in good agreement with the tillage times reported by the USDA (USEPA, 2018).

Fig. 11 shows the 3D CNN model-based date feature importances for a specific field with corn residue and under conventional tillage. The feature importances show trends similar to those of the patch-averaged values (Fig. 10) with peaks occurring in November 2020 and May 2021. To further investigate these two peaks, Fig. 12 shows temporal sequences of true color (RGB) PF images of the specific field polygon around the first (November 1 to November 30, 2020) and second (May 1 to May 31, 2021) peak. Only images with >80% observed pixels (i.e., not gap-filled) were selected for visual display. The spring tillage likely occurred between the first (May 6, 2021) and second image (May 8,

Table 4
Training and evaluation computation time (second) for each deep learning model.

Temporal resolution (days)		365	256	128	64	32	16
3D CNN	Train	8102.50	5792.80	3151.93	1857.59	1193.45	855.30
	Evaluation	69.04	52.74	33.52	22.03	15.67	13.78
2D CNN	Train	2803.31	2195.26	1272.00	877.60	692.23	597.55
	Evaluation	51.78	41.01	24.19	19.14	14.75	12.94
CNN-LSTM	Train	11894.34	8669.75	4842.28	2898.42	1915.14	1451.38
	Evaluation	107.39	74.24	46.57	28.45	19.45	16.20
Attention CNN-LSTM	Train	12191.85	8934.47	4950.18	2996.85	2003.46	1512.26
	Evaluation	109.33	75.94	45.81	29.55	19.60	16.51

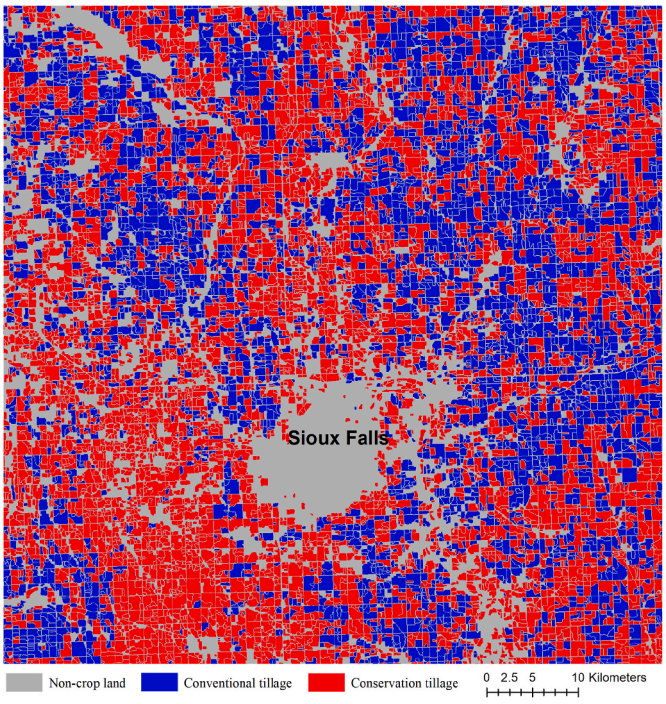


Fig. 8. The tillage practice map after 2020 harvest classified using the 3D CNN model using 256 days of PF-SR images.

2021) dates of the six spring images shown in Fig. 12b as evidenced by the soil textures changes and reflectance drops from the first to the second images. Similarly, the fall tillage likely occurred between the first (November 2, 2020) and fourth image (November 6, 2020) dates of the six fall images shown in Fig. 12a as evidenced by soil textures changes and reflectance drops. However, it is not clear whether tillage has been undertaken in the second and third images (November 3, 2020 and November 4, 2020) as they did exhibit subtle texture changes compared to the first image. It should be noted that in fall season right after harvest and before tillage there could be other crop residue management activities such as bailing and grazing (personal communication with local farmers). The images in Fig. 12 are associated with the highest feature importances, which supports the utility of a feature importance analysis

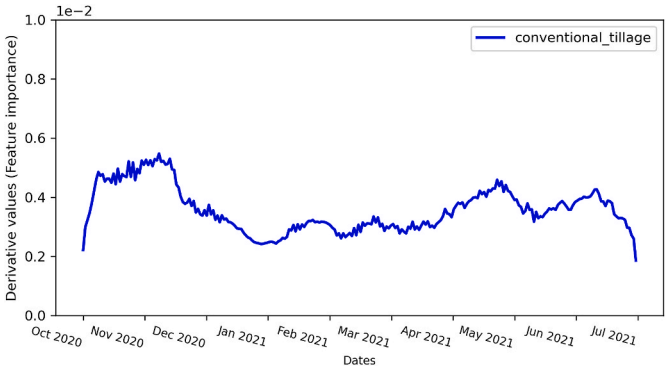


Fig. 11. 3D CNN model feature importances for a specific field with corn residue and conventional tillage practice.

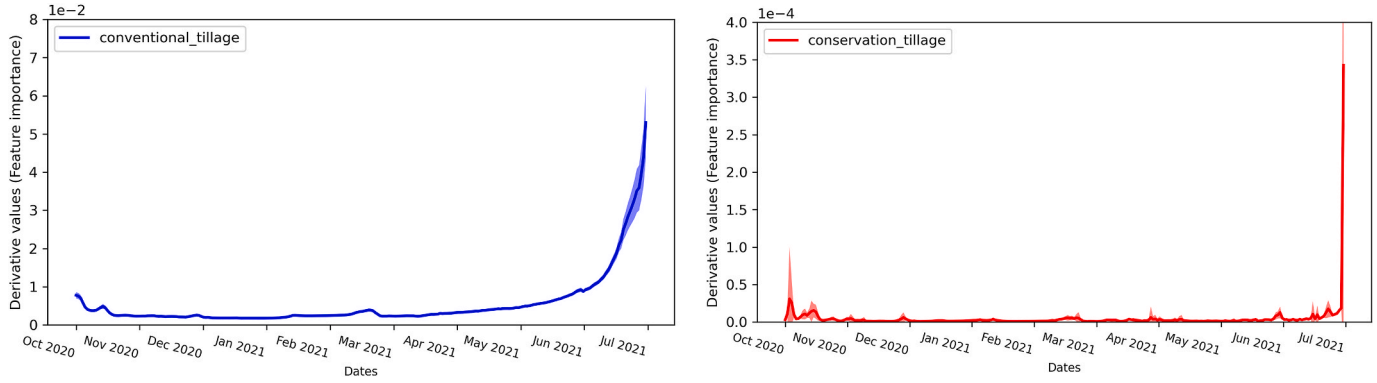


Fig. 9. The attention CNN-LSTM date feature importance averaged over all patches for conventional (left) and conservation (right) tillage practice. The model was trained using 256 PF-SR images and buffer area represents the associated confidence interval.

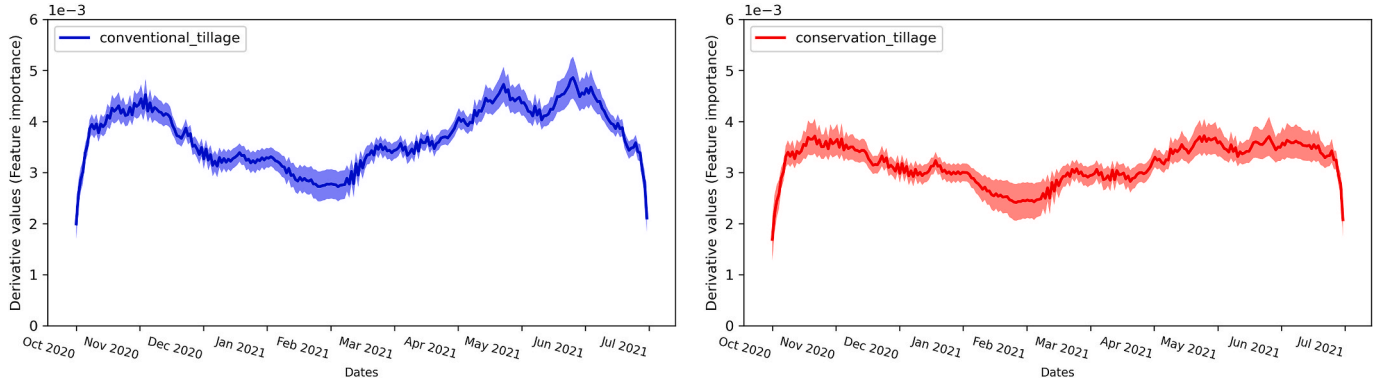
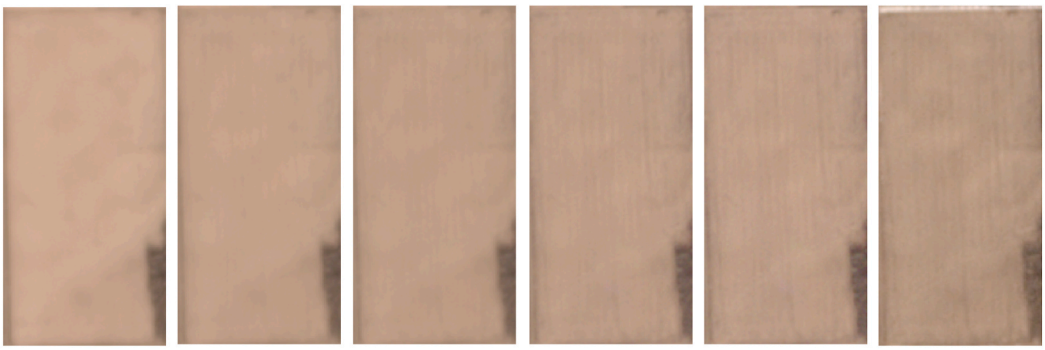


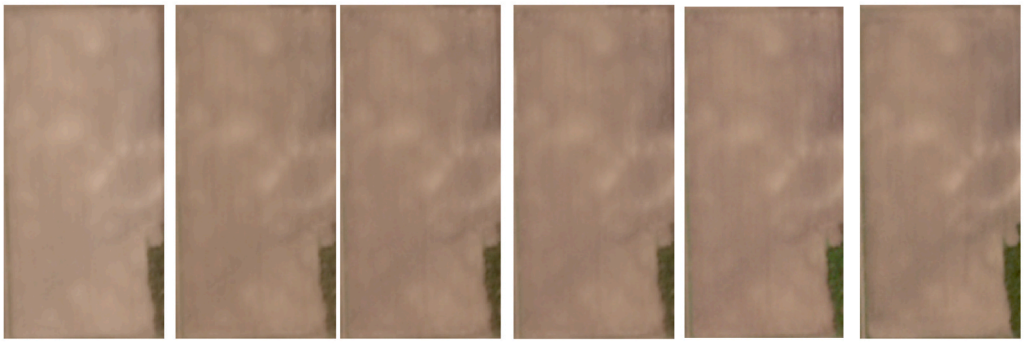
Fig. 10. The 3D CNN date-specific feature importance averaged over all patches for conventional (left) and conservation (right) tillage practice. The model was trained using 256 PF-SR images and buffer area represents the associated confidence interval.

(a)



Derivate value	0.0046	0.0045	0.0048	0.0050	0.0045	0.0051
date	11/02/2020	11/03/2020	11/04/2020	11/06/2020	11/07/2020	11/20/2020
Tillage	Before	Not sure	Not sure	After	After	After

(b)



Derivate value	0.0042	0.0039	0.0037	0.0037	0.0038	0.0034
date	05/06/2021	05/08/2021	05/10/2021	05/11/2021	05/14/2021	05/22/2021
Tillage	Before	After	After	After	After	After

Fig. 12. PF daily images (RGB) of a corn field with conventional tillage practice showing tillage-driven soil disturbances over time. The derivate values (feature importances), acquisition dates and timing (before/after) relative to the tillage event are listed below the images. (a) PF image sequence from November 1 to November 30, 2020, (b) PF image sequence from May 1 to May 31, 2021.

for identifying potential tillage date.

5. Discussion

The model classification sensitivity to spatial and temporal resolutions was analyzed by simulating data with a series of different spatial and temporal resolutions from the original 3 m patch time series data. As expected, the overall classification accuracies with the original 3 m spatial resolution (32 × 32 pixel patch) data were better compared to degraded 24 m and 96 m spatial resolution data. Notably, the overall accuracy with 3 m spatial resolution data was around 8% higher than the overall accuracy with 96 m spatial resolution data using the 3D CNN model regardless of the temporal resolution (Figs. 6 and 7). This suggests, not surprisingly, that high (e.g., 3 m) spatial resolution information is important for tillage identification with deep learning models. The deep learning models were also sensitive to temporal resolution (Figs. 6 and 7), but the sensitivities were not consistent between the various models. The overall classification accuracies were seen to first increase with increasing temporal resolution before leveling off at 2-day (for LSTM models) or 1-day (CNN models) resolution. This is related to fundamental model differences between CNN and LSTM, which uses local convolution and global time series information accumulation, respectively to handle time series data (Cho et al., 2014; Ienco et al., 2019). Based on these results, the recommendation would be to use a 3D CNN with 1-day temporal resolution or attention CNN-SLTM model with 2-day temporal resolution for remote sensing-based tillage

identification. The significant value of high frequency observations (i.e., daily or every second day) is likely related to typically very short time periods between harvest and tillage practices (e.g., ~within 10 days in South Dakota, personal communication with local farmers).

The study used 3285 PF-SR daily images covering a 220 × 220 km^2 area in the US Great Plains. Generalization studies to different areas are limited by the training data availability. The farming practices are considered privacy and protected by laws in the US and the authors conducted field survey with the company of the USDA experts not only on tillage practice identification but also on data sharing policy. This combined with limited resources prohibited large area field survey. However, since we used the training and evaluation patches coming from different fields to achieve the evaluation data independence, we expect similar accuracy can be achieved at other areas as long as the area specific training data are available. The overall classification accuracies achieved in this study are similar to existing studies using medium spatial resolution satellite data (such as Landsat or Sentinel-2 data) to classify tillage practice (Azzari et al., 2019; Beeson et al., 2020; Watts et al., 2011). Even though the spatial and temporal resolution from PlanetScope data is significantly higher than what is currently achievable from public medium resolution satellite sensors (10–30 m and 5–16 day revisit cycle), the accuracy from PlanetScope data may be impacted by the lack of shortwave infrared bands as tillage practice is related to soil moisture that influences bands in the shortwave infrared domain. At the same time, the spectral signatures of crop residue typically show a decreasing trend between 1600 nm and 2300 nm (Hively et al., 2018; Quemada et al., 2018). In addition, there is another unique feature of this study that could negatively affect the classification accuracy. This study considers residues for all the crop types for tillage classification which is different from previous studies focusing on only one or two crop type residue.

The current study did not evaluate recently developed fully attention-based models (e.g., Transformer) as they were developed for three-dimensional single image classification and have not been directly applied to four dimensional patch time series classification. Future work is encouraged to adapt the Transformer-like models, e.g., Visual Transformer (Li et al., 2020; Y Liu et al., 2022; Ofori-Ampofo et al., 2021; Sainte Fare Garnot et al., 2020; Yan et al., 2022), or self-supervised models (Ayush et al., 2021; Manas et al., 2021; Yuan and Lin, 2021) for tillage classification.

6. Conclusion

In this study, daily 3 m Planet Fusion Surface Reflectance (PF-SR) time series were used for the first time to classify conventional and conservation tillage practice. The benefit of using time series surface reflectance images for identifying tillage practices have been previously demonstrated using a random forest classifier (Azzari et al., 2019; Watts et al., 2011). Deep learning models (e.g., CNNs) were explored in this study given their successful application for land cover classification (Ienco et al., 2019; Interdonato et al., 2019; Masolele et al., 2021; Yan et al., 2022; Zheng et al., 2022). In this study a total of 10,747 time series patches (each with 32 × 32 × 365 × 4 reflectance values) were derived from 433 polygons with field collected tillage data. These patches were classified into conventional and conservation tillage and 30% of these patches from the fields that are never used for training were used for model evaluation. Four deep learning models (i.e., 3D CNN, 2D CNN, CNN-LSTM, and attention CNN-LSTM) were tested in this study with the overall accuracies ranging from 68% to 77% as a function of differences in spatial and temporal resolution. The highest accuracy (77%) was achieved using 3D CNN and attention CNN-LSTM models applied to the 3 m spatial resolution data with daily or two-day temporal resolutions. The attention CNN-LSTM model consistently outperformed the CNN-LSTM model, which highlights the importance of the attention mechanism to re-weight the input time series features. The 3D CNN performed better than the 2D CNN model indicating that the use of a multiple-layer convolution network (in 3D CNN) is more effective than using a fully connected layer (in 2D CNN) to handle time series data. The spatial and temporal information from the PF-SR time series images were shown to play an important role on tillage classification.

This study for the first time tried to identify the tillage date using deep learning based interpretation capability through date-specific feature importance. Knowing the timing of tillage practices is important since tillage practices from agricultural activities can affect ecosystem services such as soil organic matter, soil erosion, and water quality. The logic behind the tillage date identification using classification model is that the images acquired right after the tillage should contribute more to tillage classification when the tillage signal is strong and has not been weathering. Consequently, date-specific feature importance provide a promising mechanism towards estimating the timing of tillage practices. This study lacks in-field collected tillage timing information for evaluation purposes and thus the results were only evaluated using visually image time series over selected fields. Collecting the tillage time data should be prioritized for future work on the topic.

Code availability

The python manipulation codes are available at https://github.com/lwind18/DL-tillage-time-series. The training and evaluation samples used in this study cannot be made publicly available to conform to the USDA NCRS policy.

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

The authors do not have permission to share data.

Acknowledgement

We thank Daniel Wehmeyer, Mary Lou Lacey, Casey Wenzlaff, and Andrew Ganske from the US Department of Agriculture (USDA) Natural Resources Conservation Service (NCRS) South Dakota office and Anthony Bly with South Dakota State Univeristy (SDSU) Extension for taking the authors to field survey. The study was supported by the SDSU Agricultural Experiment Station Hatch Project (Grant number 328100). Funding for initial data collection was supported in part by USDA NRCS South Dakota, Project number NR206740XXXXG002, Project title Roadmap to Water Resilience - Valuing Water as a Resource for Improved Ag Land Profitability and Reduction of Downstream Flood Risk.

References

Alvarez, R., Alvarez, R., 2005. A review of nitrogen fertilizer and conservation tillage effects on soil organic carbon storage. Soil Use Manag. 21, 38–52. https://doi.org/10.1079/sum2005291.

Ashapure, A., Jung, J., Yeom, J., Chang, A., Maeda, M., Maeda, A., Landivar, J., 2019. A novel framework to detect conventional tillage and no-tillage cropping system effect on cotton growth and development using multi-temporal UAS data. ISPRS J. Photogrammetry Remote Sens. 152, 49–64. https://doi.org/10.1016/j.isprsjprs.2019.04.003.

Arvidsson, J., Bölenius, E., 2006. Effects of soil water content during primary tillage–laser measurements of soil surface changes. Soil Tillage Res. 90 (1–2), 222–229.

Ayush, K., Uzkent, B., Meng, C., Tanmay, K., Burke, M., Lobell, D., Ermon, S., n.d. Geography-Aware Self-Supervised Learning 10181–10190.

Azzari, G., Grassini, P., Edreira, J.I.R., Conley, S., Mourtzinis, S., Lobell, D.B., 2019. Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. Remote Sens. Environ. 221, 417–429. https://doi.org/10.1016/j.rse.2018.11.010.

Bahdanau, D., Cho, K.H., Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–15.
Beeson, P.C., Daughtry, C.S.T., Wallander, S.A., 2020. Estimates of conservation tillage practices using landsat archive. Rem. Sens. 12, 1–18. https://doi.org/10.3390/RS12162665.
Ben Hamida, A., Benoit, A., Lambert, P., Ben Amar, C., 2018. 3-D deep learning approach for remote sensing image classification. IEEE Trans. Geosci. Rem. Sens. 56, 4420–4434. https://doi.org/10.1109/TGRS.2018.2818945.
Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26, 341–358. https://doi.org/10.1080/10106049.2011.562309.
Busari, M.A., Kukal, S.S., Kaur, A., Bhatt, R., Dulazi, A.A., 2015. Conservation tillage impacts on soil, crop and the environment. Int. Soil Water Conserv. Res. 3, 119–129. https://doi.org/10.1016/j.iswcr.2015.05.002.
Che, X., Zhang, H.K., Liu, J., 2021. Making Landsat 5, 7 and 8 reflectance consistent using MODIS nadir-BRDF adjusted reflectance as reference. Remote Sens. Environ. 262, 112517 https://doi.org/10.1016/j.rse.2021.112517.
Chen, T.H.K., Qiu, C., Schmitt, M., Zhu, X.X., Sabel, C.E., Prishchepov, A.V., 2020. Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: a semantic segmentation solution. Remote Sens. Environ. 251, 112096 https://doi.org/10.1016/j.rse.2020.112096.
Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: encoder–decoder approaches. Proc. SSST 2014 - 8th Work. Syntax. Semant. Struct. Stat. Transl. 103–111. https://doi.org/10.3115/v1/w14-4012.
Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling 1–9.
De Vita, P., Di Paolo, E., Fecondo, G., Di Fonzo, N., Pisante, M., 2007. No-tillage and conventional tillage effects on durum wheat yield, grain quality and soil moisture content in southern Italy. Soil Tillage Res. 92 (1–2), 69–78.
Deines, J.M., Wang, S., Lobell, D.B., 2019. Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt. Environ. Res. Lett. 14 (12), 124038.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf. 1, 4171–4186.
Frantz, D., 2019. FORCE-Landsat + Sentinel-2 analysis ready data and beyond. Rem. Sens. 11 https://doi.org/10.3390/rs11091124.
Gao, L., Zhang, C., Yun, W., Ji, W., Ma, J., Wang, H., Li, C., Zhu, D., 2022. Mapping crop residue cover using adjust normalized difference residue index based on sentinel-2 MSI data. Soil Tillage Res. 220, 105374 https://doi.org/10.1016/j.still.2022.105374.
Ge, Y., Zhang, X., Atkinson, P.M., Stein, A., Li, L., 2022. Geoscience-aware deep learning: a new paradigm for remote sensing. Science of Remote Sensing 5, 100047.
Giglio, L., van der Werf, G.R., Randerson, J.T., Collatz, G.J., Kasibhatla, P., 2006. Global estimation of burned area using MODIS active fire observations. Atmos. Chem. Phys. 6, 957–974. https://doi.org/10.5194/acp-6-957-2006.
Gotmare, A., Shirish Keskar, N., Xiong, C., Socher, R., 2019. A closer look at deep learning heuristics: learning rate restarts, warmup and distillation. 7th Int. Conf. Learn. Represent. ICLR 2019.
Hively, W.D., Lamb, B.T., Daughtry, C.S.T., Shermeyer, J., McCarty, G.W., Quemada, M., 2018. Mapping crop residue and tillage intensity using WorldView-3 satellite shortwave infrared residue indices. Rem. Sens. 10 https://doi.org/10.3390/rs10101657.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
Houborg, R., McCabe, M.F., 2018. A cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing Planet, landsat and MODIS data. Remote Sens. Environ. 209, 211–226. https://doi.org/10.1016/j.rse.2018.02.067.
Huang, C., Townshend, J.R., Liang, S., Kalluri, S.N., DeFries, R.S., 2002. Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. Rem. Sens. Environ. 80 (2), 203–212.
Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogrammetry Remote Sens. 158, 11–22. https://doi.org/10.1016/j.isprsjprs.2019.09.016.
Interdonato, R., Ienco, D., Gaetano, R., Ose, K., 2019. DuPLO: a DUal view Point deep Learning architecture for time series classificatiOn. ISPRS J. Photogrammetry Remote Sens. 149, 91–104. https://doi.org/10.1016/j.isprsjprs.2019.01.011.
Jacquemoud, S., Baret, F., Hanocq, J.F., 1992. Modeling spectral and bidirectional soil reflectance. Rem. Sens. Environ. 41 (2–3), 123–132.
Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3D Convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 221–231. https://doi.org/10.1109/TPAMI.2012.59.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.
Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogrammetry Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.
Li, Z., Arora, S., 2019. An Exponential Learning Rate Schedule for Deep Learning, pp. 1–29.
Li, Z., Chen, G., Zhang, T., 2020. A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 13, 847–858. https://doi.org/10.1109/JSTARS.2020.2971763.
Liu, M., Chai, Z., Deng, H., Liu, R., 2022. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 15, 4297–4306. https://doi.org/10.1109/JSTARS.2022.3177235.
Liu, Y., Rao, P., Zhou, W., Singh, B., Srivastava, A.K., Poonia, S.P., et al., 2022. Using Sentinel-1, Sentinel-2, and Planet satellite data to map field-level tillage practices in smallholder systems. PLoS One 17 (11), e0277425. https://doi.org/10.1371/journal.pone.0277425.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.
Luong, M.T., Pham, H., Manning, C.D., 2015. Effective approaches to attention-based neural machine translation. Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process. 1412–1421. https://doi.org/10.18653/v1/d15-1166.
Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogrammetry Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.
Ma, X., Hovy, E., 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. 54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap. 2, 1064–1074. https://doi.org/10.18653/v1/p16-1101.
Ma, X., Wang, Q., Tong, X., Atkinson, P.M., 2022. A deep learning model for incorporating temporal information in haze removal. Rem. Sens. Environ. 274, 113012.
Mamkagh, A.M., Salmerón, A., García, Y., Real, A.I., 2009a. Effect of tillage time and plastic mulch on growth and yield of okra (Abelmoschusesculentus) grown under rain-fed conditions. Int. J. Agric. Biol. 11 (4), 453–457.
Mamkagh, A.M., Salmerón, A., García, Y., Real, A.I., 2009b. Effect of tillage time and plastic mulch on growth and yield of okra (Abelmoschusesculentus) grown under rain-fed conditions. Int. J. Agric. Biol. 11 (4), 453–457.
Masolele, R.N., De Sy, V., Herold, M., Marcos Gonzalez, D., Verbesselt, J., Gieseke, F., Mullissa, A.G., Martius, C., 2021. Spatial and temporal deep learning methods for deriving land-use following deforestation: a pan-tropical case study using Landsat time series. Remote Sens. Environ. 264, 112600 https://doi.org/10.1016/j.rse.2021.112600.
Melero, S., López-Garrido, R., Murillo, J.M., Moreno, F., 2009. Conservation tillage: short- and long-term effects on soil carbon fractions and enzymatic activities under Mediterranean conditions. Soil Tillage Res. 104, 292–298. https://doi.org/10.1016/j.still.2009.04.001.
Mishra, U., Ussiri, D.A.N., Lal, R., 2010. Tillage effects on soil organic carbon storage and dynamics in Corn Belt of Ohio USA. Soil Tillage Res. 107, 88–96. https://doi.org/10.1016/j.still.2010.02.005.
Montavon, G., Samek, W., Müller, K.R., 2018. Methods for interpreting and understanding deep neural networks. Digit. Signal Process. A Rev. J. 73, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011.
Ofori-Ampofo, S., Pelletier, C., Lang, S., 2021. Crop type mapping from optical and radar time series using attention-based deep learning. Rem. Sens. 13, 1–17. https://doi.org/10.3390/rs13224668.
Pacheco, A., McNairn, H., 2010. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. Remote Sens. Environ. 114, 2219–2228. https://doi.org/10.1016/j.rse.2010.04.024.
Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J. Photogrammetry Remote Sens. 145, 120–147. https://doi.org/10.1016/j.isprsjprs.2017.11.021.
Planet Fusion Team, 2022. Planet fusion monitoring technical specification. Version 1.0.0, San Francisco, CA. https://support.planet.com/hc/en-us/articles/4406292582673-Planet-Fusion-Monitoring-Technical-Specification.html.
Quemada, M., Hively, W.D., Daughtry, C.S.T., Lamb, B.T., Shermeyer, J., 2018. Improved crop residue cover estimates obtained by coupling spectral indices for residue and moisture. Remote Sens. Environ. 206, 33–44. https://doi.org/10.1016/j.rse.2017.12.012.
Roger-Estrade, J., Anger, C., Bertrand, M., Richard, G., 2010. Tillage and soil ecology: partners for sustainable agriculture. Soil Tillage Res. 111, 33–40. https://doi.org/10.1016/j.still.2010.08.010.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, III 18. Springer International Publishing, Munich, Germany, pp. 234–241. October 5-9, 2015, Proceedings, Part.
Roy, D.P., Huang, H., Houborg, R., Martins, V.S., 2021. A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. Remote Sens. Environ. 264, 112586 https://doi.org/10.1016/j.rse.2021.112586.
Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. ISPRS J. Photogrammetry Remote Sens. 169, 421–435. https://doi.org/10.1016/j.isprsjprs.2020.06.006.
Sainte Fare Garnot, V., Landrieu, L., Giordano, S., Chehata, N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. 12322–12331. https://doi.org/10.1109/CVPR42600.2020.01234.
Smith, R.G., 2006. Timing of tillage is an important filter on the assembly of weed communities. Weed Sci. 54 (4), 705–712.

South, S., Qi, J., Lusch, D.P., 2004. Optimal classification methods for mapping agricultural tillage practices. Remote Sens. Environ. 91, 90–97. https://doi.org/10.1016/j.rse.2004.03.001.

Sun, J., Di, L., Sun, Z., Shen, Y., Lai, Z., 2019. County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19, 1–21. https://doi.org/10.3390/s19204363.

Sun, Z., Di, L., Fang, H., 2019. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. Int. J. Rem. Sens. 40, 593–614. https://doi.org/10.1080/01431161.2018.1516313.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 4, 3104–3112.

Tarkalson, D.D., Bjorneberg, D.L., Moore, A., 2016. Fall and spring tillage effects on sugarbeet production. J. Sugar Beet Res. 52 (3 & 4), 30–38.

Teasdale, J.R., Mirsky, S.B., 2015. Tillage and planting date effects on weed dormancy, emergence, and early growth in organic corn. Weed Sci. 63 (2), 477–490.

Thorp, K.R., Drajat, D., 2021. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia. Remote Sens. Environ. 265, 112679 https://doi.org/10.1016/j.rse.2021.112679.

Townshend, J., Justice, C., Li, W., Gurney, C., McManus, J., 1991. Global land cover classification by remote sensing: present capabilities and future possibilities. Rem. Sens. Environ. 35 (2–3), 243–255.

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.

USDA-NRCS, 2019. 2019 South Dakota Cropping Systems Inventory. USDA-NRCS Tech. Resour. L. Use.

USEPA, 2018. Tillage intensity and conservation cropping in the United States United States department of agriculture. United states dep. Agric. For. 3–4.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017-Decem 5999–6009.

Voosen, P., 2017. The AI detectives. Science 357, 22–27. https://doi.org/10.1126/science.357.6346.22.

Wang, H., Zhang, X., Du, S., Bai, L., Liu, B., 2022. Mapping Annual Urban Evolution Process (2001–2018) at 250 m: A normalized multi-objective deep learning regression. Remote Sens. Environ. 278, 113088 https://doi.org/10.1016/j.rse.2022.113088.

Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y., Zhu, K., 2018. Change detection based on Faster R-CNN for high-resolution remote sensing images. Remote Sens. Lett. 9, 923–932. https://doi.org/10.1080/2150704X.2018.1492172.

Wang, X., Yue, Y., Noor, M.A., Hou, H., Zhou, B., Ma, W., Zhao, M., 2018. Tillage time affects soil hydro-thermal properties, seedling growth and yield of maize (Zea mays L.). Appl. Ecol. Environ. Res. 16 (5), 6007–6023.

Wang, X., Yue, Y., Noor, M.A., Hou, H., Zhou, B., Ma, W., Zhao, M., 2018. Tillage time affects soil hydro-thermal properties, seedling growth and yield of maize (Zea mays L.). Appl. Ecol. Environ. Res. 16 (5), 6007–6023.

Watts, J.D., Lawrence, R.L., Miller, P.R., Montagne, C., 2009. Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. Remote Sens. Environ. 113, 1843–1852. https://doi.org/10.1016/j.rse.2009.04.015.

Watts, J.D., Powell, S.L., Lawrence, R.L., Hilker, T., 2011. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. Remote Sens. Environ. 115, 66–75. https://doi.org/10.1016/j.rse.2010.08.005.

Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. Int. J. Rem. Sens. 39, 4254–4284. https://doi.org/10.1080/01431161.2018.1452075.

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.

Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. ISPRS J. Photogrammetry Remote Sens. 144, 423–434. https://doi.org/10.1016/j.isprsjprs.2018.08.005.

Yan, J., Liu, J., Wang, L., Liang, D., Cao, Q., Zhang, W., Peng, J., 2022. Land-cover classification with time-series remote sensing images by complete extraction of multiscale timing dependence. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 15, 1953–1967. https://doi.org/10.1109/JSTARS.2022.3150430.

Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86. https://doi.org/10.1016/j.rse.2015.10.034.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241, 111716 https://doi.org/10.1016/j.rse.2020.111716.

Yuan, Y., Lin, L., 2021. Self-supervised pretraining of transformers for satellite image time series classification. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14, 474–487. https://doi.org/10.1109/JSTARS.2020.3036602.

Zheng, B., Campbell, J.B., Serbin, G., Galbraith, J.M., 2014. Remote sensing of crop residue and tillage practices: present capabilities and future prospects. Soil Tillage Res. 138, 26–34. https://doi.org/10.1016/j.still.2013.12.009.

Zheng, Z., Zhong, Y., Tian, S., Ma, A., Zhang, L., 2022. ChangeMask: deep multi-task encoder-transformer-decoder architecture for semantic change detection. ISPRS J. Photogrammetry Remote Sens. 183, 228–239. https://doi.org/10.1016/j.isprsjprs.2021.10.015.