Prediction of Crop Planting Map Using One-dimensional Convolutional Neural Network and Decision Tree Algorithm

Hui Li, Liping Di*, Chen Zhang, Li Lin, Liying Guo, Haoteng Zhao
Center for Spatial Information Science and Systems (CSISS), George Mason University
Fairfax, VA, USA
{hli47, * ldi, czhang11, llin2, lguo4, hzhao22}@gmu.edu

Abstract—**The crop type planting prediction map is an essential agro-geoinformation data source to explore and quantify agriculture cultivation distribution in the coming year, implying crop planting change tendency. This paper validates the feasibility of crop type prediction using a one-dimensional convolutional neural network (1D CNN) and decision tree algorithm. To construct the 1D CNN model, we encode and stack the historical Cropland Data Layer (CDL) into a 3D time series location matrix as the training dataset. According to the validation for the 2021 crop planting map in Cass County of Iowa, the prediction result owns high overall accuracy (0.927) and kappa coefficient (0.857). The major crop types, corn and soybean, have high prediction producer accuracy (0.9 – 0.95) and user accuracy (0.91-0.94). The minor crop alfalfa has lower accuracy (0.55-0.73). This approach provides an option to predict major crop type's planting maps for the next year.**

Keywords— crop map prediction, one-dimensional CNN, CDL, decision tree

I. Introduction

Agriculture planting continuously feeds the population and provides energy and raw material for humankind's society [1]–[3]. However, regional conflicts, global climate change, economic recession, and pandemics give some uncertainty to agricultural planting and grain markets. The timely crop maps are essential datasets for understanding the current agriculture planting situation – distribution and acreage that usefully monitor food supply security, agricultural strategy planning, and other domestic economic activities [4], [5]. Various Earth Observation satellite data, such as Landsat and ESA – Sentinel series imagery, are collected to illustrate the Earth's surface agricultural planting changes, acting as a crucial resource in remote sensing agricultural crop mapping technology that intends to classify diverse crop types and locate crop growing distribution [6], [7].

According to the phase of mapping, crop mapping can be classified into three categories: pre-season, in-season, and post-season maps [8]. The remote sensing post-season and in-season crop mapping technology have already been widely investigated by the agricultural crop type identification community [9], [10]: Cropland Data Layer (CDL), a well-known annual post-season crop mapping product, is produced by the United States Department of Agriculture (USDA) that monitors the US agriculture planting and free access to the public [11]; the remote sensing in-season crop mapping is blooming explored using in-season time series satellites data and machine learning techniques [12]–[16], as well as, other in-season vegetation maps - grassland [17] related to these techniques. Nevertheless, pre-season remote sensing crop mapping techniques are frequently pioneered in agricultural academic communities [18]–[20]. The crop map before the growing season predicts planting information that includes potential crop type, location distribution, and acreage estimation, serving for food security pre-evaluation, potential agriculture market competition analysis, gross yield estimation of specific grains, and agricultural policies prejudgment. However, massive challenges occurred when remote sensing technology forecasts what types of crop will grow on the ground fields based on the pre-season spectral data. Additionally, the survey of planting ambition is a tough task, collecting data from farmers that depends on a huge human resource effort.

The machine learning method gives us an option in agriculture activity prediction. Briefly, the can explore phenomena and discover the principle, imitating the human brain learning metric, which needs to accumulate experience from input data and distill the essential patterns of development or change to predict the next phase. Machine learning algorithms have been popularly utilized in crop yield forecasts [21]–[25]. Furthermore, to estimate the crop type and area on the specific ground in the coming growing season, Recurrent Neural Network and Artificial Neural Network were employed to predict the crop map based on the historical crop maps [8], [18].

In this study, we use one-dimensional Convolutional Neural Network (1D CNN) machine learning framework to predict the crop planting map for the coming year using CDL. Our objective is to predict the 2021 crop map using the historical CDL data (2008-2020) as training data to build 1D CNN. We use the bool approach to encode every crop type training data for extremely reserving the spatial information. Every crop type's planting probability map can be generated that will be integrated into the final prediction planting map by a decision tree method.

2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) | 979-8-3503-0351-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/Agro-Geoinformatics59224.2023.10233466

II. Data and Method

A. Study area

We select Cass County as the study area, which is located at the south and west of the Iowa, U.S., belongs to the Corn-Belt region, was used to plant corn, soybean, alfalfa, hay, oat, triticale, sorghum, winter wheat, rye, and barley, but only the top three crop types (i.e., corn, soybean, alfalfa) have considerable amount during past decades that is the main reason for choosing these three crop as prediction aims in this case. Fig 1 shows the location of the study area. The acreage changes of major agriculture land for these prediction aims during 2008-2020 is illustrated in Fig 2. Meanwhile, the area of Cass County is 360774 acreages with around 80% land use for these three crops. All above statistics are from the CropScape system [26].

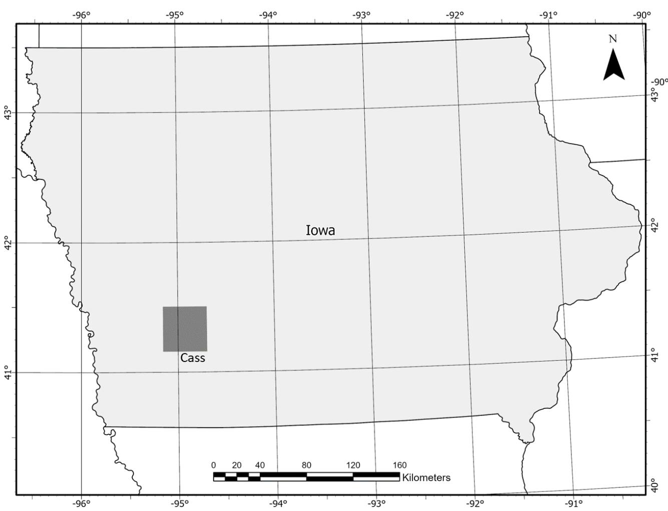

Fig 1 Study area location – Cass County, Iowa.

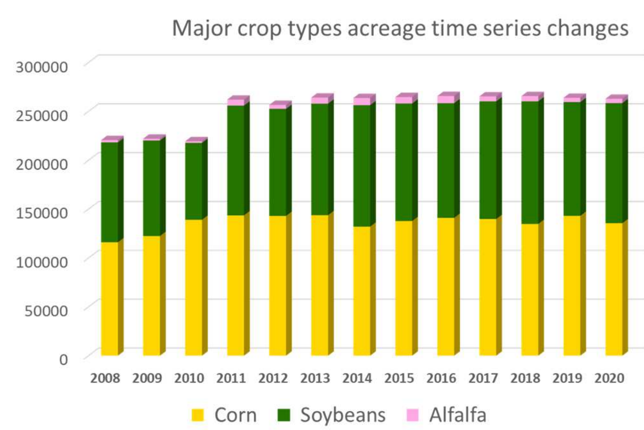

Fig 2 Major agriculture land use changes.

B. Official Cropland Data Layer

The USDA annually produces CDL for the Conterminous United States since 2008 to provide dozens of crops distribution data in a planting year. CDL is an official field-level raster cropland map with 30m spatial resolution, covering the entire CONUS with 48 states since 2008, including all major crop types (e.g., corn, soybean, cotton, rice, winter wheat) and non-crop types (e.g., developed land, open water, wetland, forest), is currently created by the decision tree classifier using Landsat 8 and 9, DEIMOS-1 and UK2, the ISRO ResourceSat-2 LISS-3, and Sentinel-2 satellite, is published to the public around February in the coming year after general crop growing season [11], [27]. The general producer accuracies of major crops in the large regions are greater than 80%, even in some regions reach to 90%, and the accuracy assessments error matrices are provided since 2008 [27], [28]. CDL has already been used to offer the crucial agro-geoinformation for acreage and yield estimates [28], disaster monitoring like flooding and drought [29], specific crop identification and extraction [30], and agricultural land use type change detection [31]. In this study, the CDL 2008-2020 acts as the inputting training data for the CNN machine learning classifier. All CDL data can be easily downloaded from CropScape system [26].

C. Spatial Encoding of Time-series CDL

A sequence of value is set in CDL to label land use types; however, this is a fact that these nominal values cannot satisfy the machine learning method since these values lack spatial and temporal meaning to calculate. The original crop value of CDL thus should be processed spatial encoding before inputting the CNN workflow. The spatial encode needs to extremely remain geolocation property of the crop that is significantly to understand crop's spatial distribution and temporal change for machine learning model. In this paper, raster location map is employed to represent every crop type's distribution, separately. Every crop type has an independent location map that corresponds to every historical CDL. For a crop type location map in a certain year, setting 1 as pixel value if the pixel is occupied by this crop type but 0 if negative. Each historical CDL therefore can be encoded by three location maps – corn, soybean, and alfalfa. Every pixel in the location map corresponds to CDL pixel location either. Fig. 4 demonstrates an example, containing corn, soybean and alfalfa location maps that match CDL in a certain year. Following this method, every crop type time series location maps in 2008-2021 can be generated, respectively.

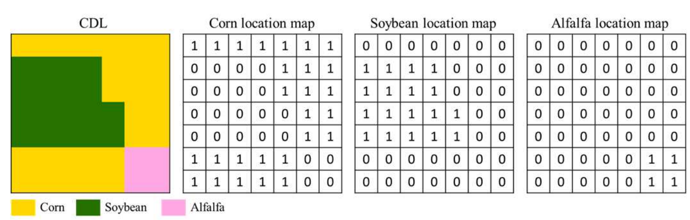

Fig 3 Corn, soybean, and alfalfa location maps correspond CDL.

D. Preparing training dataset and testing dataset

This part processes the crop type spatial encoded data into the training and testing dataset. To achieve them, every crop type's 3D time series location matrix should be created by flatting and merging. Fig 4 as an example shows training set, label set and validation dataset process for corn. Flatting and merging crop type location maps in 2008-2019 into a 2D matrix that is converted to a 3D matrix with Axis-X, Axis-Y, and Axis-Z as Fig 4 (a). Axis-X and Axis-Y construct a plane with 12×1 array that represent time series crop location status from 2008 to 2019, and each plane as an independent sequence ensures its prediction independence. Axis-Z means stacked planes. The 3D time series location matrix contains spatial and temporal information of crop type in the historical period. 2020 crop type location maps need to be flatted to a 2D matrix, labeling the 3D time series location matrix as Fig 4 (b). The 3D time series location matrix acts as training set, and 2020 label data can be the label set. 80% of them serve to train the model, and the rest 20% data set serve as validation data.

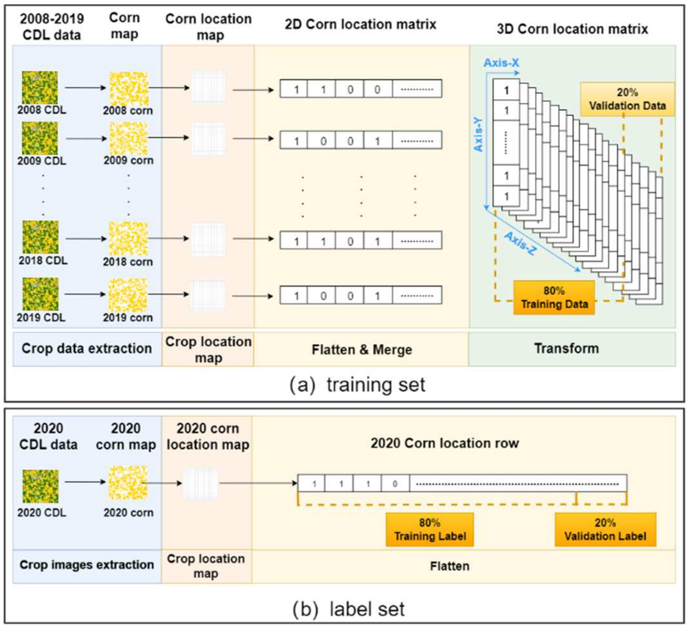

Fig 4 Training set, label set, and validation data process.

E. 1D CNN model

This study selects 1D CNN as the prediction model for the coming year crop planting map. 1D CNN owns a hidden convolutional layer to learn and predict time series variables that can be a big length sequence. Traditionally, the pooling layer follows the one-dimensional convolutional layer to reduce the amount of the input data, and the following flatten layer converts the multi-dimensional data to one-dimensional to flow into the fully connected layer. And in some cases, with huge input data, convolutional layer and pooling layer could be repeated and rearranged into the model structure.

In this case, the Axis-Y of 3D location matrix indicates time series crop type location change that can be recognized as independent input sequence with small size: thus, the pooling layer can be ignored. In the convolutional layer, the size 2 kernel and the relu activation are used. To reduce the overfitting of CNN in this case, the dropout with 0.2 is added after convolutional layer. There are two fully connected layers with Relu and SoftMax that will push model calculate the probability of predicting items. At the end of the model, the accuracy is validated by sparse categorical cross entropy method and optimized via Adam optimization algorithm. Fig 6 illustrates the 1D CNN structure for crop planting prediction. This study needs to separately operate 1D CNN for every crop type.

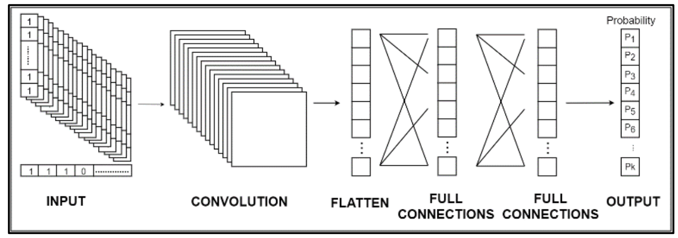

Fig 5 1D CNN structure

F. Producing predicted crop map

In this study, 1D CNN predicts every crop cultivation map for 2021, separately. The output of prediction is planting probability map whose every pixel has a probability value from 0 to 1, converting them to crop planting map is necessary for advanced application. Every pixel corresponds to three crop types probability values (corn, soybean, alfalfa), comparing them and renaming the crop type in this pixel if it is the maximum among three probabilities. Fig 6 shows the comparison and locking process.

The above process can generate crop type in most pixels; however, some pixel has two or three equal probabilities could not lock crop type that need to employ a simple neighborhood decision tree method to tackle further. For a pixel with equal probabilities in some crop types can be named confusing pixel, the spatial and temporal selector will be operated. Firstly, the pixel spatial connecting eight neighborhoods will be made category statistics and comparisons. The crop type with max count will be locked in the pixel as Fig 7 (a). Secondly, the pixel will be recognized as other or noise if there are no available neighborhoods as Fig 7 (b). Thirdly, if these crop types in the neighborhoods have the same count, the temporal selector will be switched on - the higher frequency crop type in 2008-2020 will be locked in this pixel as Fig 7 (c). Finally, the 2021 crop map is predicted by these workflow steps. The confusion matrix between prediction and CDL will be calculated to validate the accuracy.

Fig 6 Crop type probability comparing and locking process.

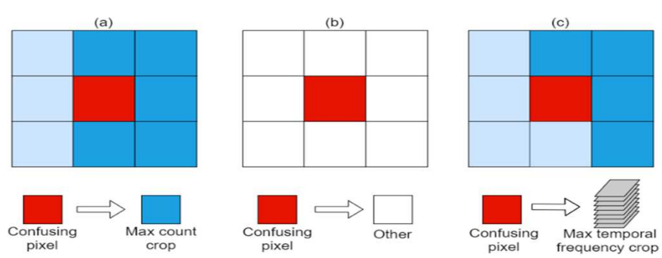

Fig 7 Confusing pixel process (the red pixel is confusing pixel, dark blue and slight blue color means different crop types, the white color pixel is other land use types).

III. RESULTS AND DISCUSSIONS

This section describes and validates the 2021 crop map prediction in Cass County. The 1D CNN model is constructed by 2008-2020 historical training dataset, which predicting 2021 crop type map based on 2009-2020 3D time series location matrix. Fig 8 illustrates the prediction result and corresponding 2021 CDL for corn, soybean, and alfalfa in Cass County, as well as the difference map between two maps.

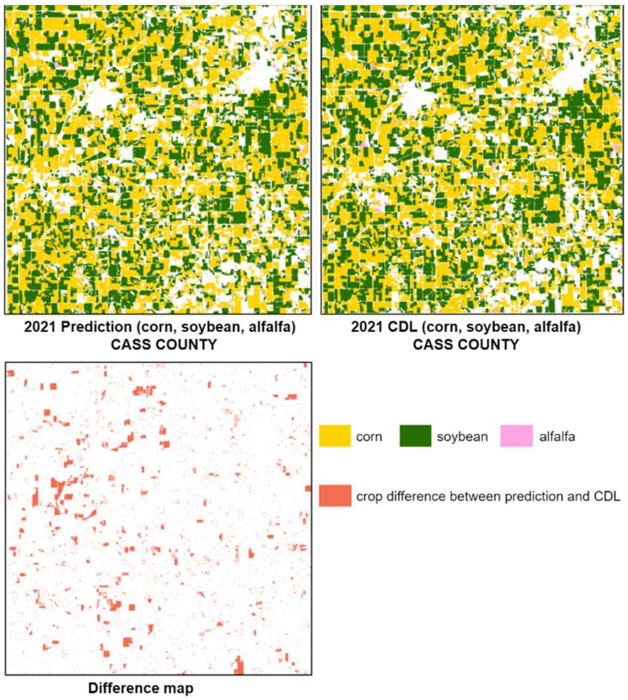

Fig 8 Prediction map, CDL 2021 in Cass County, and difference map between two maps (corn, soybean, alfalfa).

To assess the accuracy of the prediction, we use CDL 2021 as the reference data and randomly select 48107 points to calculate the confusion matrix of prediction map – Table I. The corn and soybean own higher producer and user accuracy that are greater than 0.9. Alfalfa shows lower accuracy in these two aspects. In summary, the prediction map has 0.927 overall accuracy and 0.857 kappa coefficient. The corn and soybean occupy massive agriculture land, but alfalfa grow on small size of fields from 2008 to 2020 in CASS county that can be seen in Fig.2. This bias time series crop cultivation pattern directly impacts quantity of each crop type's training data in our 1D CNN prediction model that is the probable reason for unbalance prediction accuracy in corn, soybean, and alfalfa. These phenomena confirm that the approach can be used to predict major crop types planting map for the coming year.

TABLE I. CONFUSION MATRIX BETWEEN PREDICTION AND CDL

	Corn	Soybean	Alfalfa	Total	U_Accuracy	Kappa
Corn	23846	2067	149	26062	0.915	
Soybean	1037	20432	119	21588	0.946	
Alfalfa	80	43	334	457	0.731	
Total	24963	22542	602	48107		
P_Accuracy	0.955	0.906	0.555		0.927	
Kappa						0.857

This paper uses the location maps to encode the crop types distribution of historical CDL that are constructed into 3D time series location matrix, extremely containing spatial and temporal information of the crop types history cultivation. At the same time, however, it increases the complexity of the input data for the machine learning model. Moreover, the training dataset inherits all properties from historical CDL that still includes some misclassification and adds uncertainty to the learning process [10]. Therefore, the further refinement of CDL is one of the improvement aspects of prediction approach, previous relevant studies provide some valuable material [10], [32]–[34]. 1D CNN for time series prediction in this paper owns simple structure without pooling layer and fast fitting speed. Every crop type prediction needs to separately operate 1D CNN model, probably pushing efficiency decrement to some certain extent with increment of crop type. In addition, other environmental factors should be considered in the prediction process. For example, soil moisture could represent soil water storage for crop growth [35] and be transferred into quantitative format to join in the neural network calculation.

IV. CONCLUSION

This paper innovatively encodes historical CDL crop types via location map and stacks them into the 3D time series location matrix to calculate the spatial and temporal pattern of crop type cultivation using 1D CNN. The corn, soybean, and alfalfa 2021 distribution map in Cass County of Iowa was predicted. According to the confusion matrix with 2021 CDL, prediction result owns high overall accuracy and kappa coefficient. The single crop type prediction accuracy probably depends on the quantity of historical planting area. In this case, corn and soybean have high prediction accuracy (> 0.9). This approach provides an option with reasonable accuracy to predict major crop types (corn, soybean) planting map in the next year.

REFERENCES

[1] F. P. Carvalho, "Agriculture, pesticides, food security and food safety," Environmental Science & Policy, vol. 9, no. 7, pp. 685–692, Nov. 2006, doi: 10.1016/j.envsci.2006.08.002.

[2] D. Fróna, J. Szenderák, and M. Harangi-Rákos, "The Challenge of Feeding the World," Sustainability, vol. 11, no. 20, Art. no. 20, Jan. 2019, doi: 10.3390/su11205816.

[3] C. J. Rhodes, "Feeding and Healing the World: Through Regenerative Agriculture and Permaculture," Science Progress, vol. 95, no. 4, pp. 345–446, Dec. 2012, doi: 10.3184/003685012X13504990668392.

[4] M. E. Brown, "Remote sensing technology and land use analysis in food security assessment," Journal of Land Use Science, vol. 11, no. 6, pp. 623–641, Nov. 2016, doi: 10.1080/1747423X.2016.1195455.

[5] C. Qu and X. Hao, "Agriculture Drought and Food Security Monitoring Over the Horn of Africa (HOA) from Space," in 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), Aug. 2018, pp. 1–4. doi: 10.1109/Agro-Geoinformatics.2018.8476128.

[6] S. Bontemps et al., "'Sentinel-2 for agriculture': Supporting global agriculture monitoring," in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2015, pp. 4185–4188. doi: 10.1109/IGARSS.2015.7326748.

[7] L. P. Dutrieux, C. C. Jakovac, S. H. Latifah, and L. Kooistra, "Reconstructing land use history from Landsat time-series: Case study of a swidden agriculture system in Brazil," International Journal of Applied Earth Observation and Geoinformation, vol. 47, pp. 112–124, May 2016, doi: 10.1016/j.jag.2015.11.018.

[8] C. Zhang, L. Di, L. Lin, and L. Guo, "Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps," Computers and Electronics in Agriculture, vol. 166, p. 104989, Nov. 2019, doi: 10.1016/j.compag.2019.104989.

[9] P. Hao, L. Di, C. Zhang, and L. Guo, "Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples," Science of The Total Environment, vol. 733, p. 138869, Sep. 2020, doi: 10.1016/j.scitotenv.2020.138869.

[10] L. Lin et al., "Validation and refinement of cropland data layer using a spatial-temporal decision tree algorithm," Sci Data, vol. 9, no. 1, p. 63, Dec. 2022, doi: 10.1038/s41597-022-01169-w.

[11] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program," Geocarto International, vol. 26, no. 5, pp. 341–358, Aug. 2011, doi: 10.1080/10106049.2011.562309.

[12] D. M. Johnson and R. Mueller, "Pre- and within-season crop type classification trained with archival land cover information," Remote Sensing of Environment, vol. 264, p. 112576, Oct. 2021, doi: 10.1016/j.rse.2021.112576.

[13] H. Li, L. Di, C. Zhang, L. Lin, and L. Guo, "Improvement Of In-season Crop Mapping For Illinois Cropland Using Multiple Machine Learning Classifiers," in 2022 10th International Conference on Agro-geoinformatics (Agro-Geoinformatics), Jul. 2022, pp. 1–6. doi: 10.1109/Agro-Geoinformatics55649.2022.9859153.

[14] C. Zhang et al., "Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data," Agricultural Systems, vol. 201, p. 103462, Aug. 2022, doi: 10.1016/j.agsy.2022.103462.

[15] Z. Yu et al., "Selection of Landsat 8 OLI Band Combinations for Land Use and Land Cover Classification," in 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Jul. 2019, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2019.8820595.

[16] L. R. Defante, O. F. Vilpoux, and L. Sauer, "Rapid expansion of sugarcane crop for biofuels and influence on food production in the first producing region of Brazil," Food Policy, vol. 79, pp. 121–131, Aug. 2018, doi: 10.1016/j.foodpol.2018.06.005.

[17] R. Pazúr, N. Huber, D. Weber, C. Ginzler, and B. Price, "A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data," Earth Syst. Sci. Data, vol. 14, no. 1, pp. 295–305, Jan. 2022, doi: 10.5194/essd-14-295-2022.

[18] R. Yaramasu, V. Bandaru, and K. Pnvr, "Pre-season crop type mapping using deep neural networks," Computers and Electronics in Agriculture, vol. 176, p. 105664, Sep. 2020, doi: 10.1016/j.compag.2020.105664.

[19] A. Dupuis, C. Dadouchi, and B. Agard, "Predicting crop rotations using process mining techniques and Markov principals," Computers and Electronics in Agriculture, vol. 194, p. 106686, Mar. 2022, doi: 10.1016/j.compag.2022.106686.

[20] A. Yao and L. Di, "Machine Learning-based Pre-season Crop Type Mapping: A Comparative Study," in 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Jul. 2021, pp. 1–4. doi: 10.1109/Agro-Geoinformatics50104.2021.9530356.

[21] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 151, pp. 61–69, Aug. 2018, doi: 10.1016/j.compag.2018.05.012.

[22] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," IEEE Access, vol. 8, pp. 86886–86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

[23] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," Sci Rep, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41598-020-80820-1.

[24] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 177, p. 105709, Oct. 2020, doi: 10.1016/j.compag.2020.105709.

[25] R. Shrestha et al., "Regression based corn yield assessment using MODIS based daily NDVI in Iowa state," in 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Jul. 2016, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2016.7577657.

[26] W. Han, Z. Yang, L. Di, and R. Mueller, "CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support," Computers and Electronics in Agriculture, vol. 84, pp. 111–123, Jun. 2012, doi: 10.1016/j.compag.2012.03.005.

[27] USDA NASS, "USDA - National Agricultural Statistics Service - Research and Science - CropScape and Cropland Data Layers," 2023. https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section3_17.0 (accessed Apr. 18, 2023).

[28] L. Yan and D. P. Roy, "Automated crop field extraction from multi-temporal Web Enabled Landsat Data," Remote Sensing of Environment, vol. 144, pp. 42–64, Mar. 2014, doi: 10.1016/j.rse.2014.01.006.

[29] C. G. Boryan, Z. Yang, A. Sandborn, P. Willis, and B. Haack, "Operational Agricultural Flood Monitoring with Sentinel-1 Synthetic Aperture Radar," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Jul. 2018, pp. 5831–5834. doi: 10.1109/IGARSS.2018.8519458.

[30] C. Boryan, Z. Yang, P. Willis, and A. Sandborn, "Early Season Winter Wheat Identification Using Sentinel -1 Synthetic Aperture Radar (Sar) and Optical Data," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Jul. 2019, pp. 6239–6242. doi: 10.1109/IGARSS.2019.8898511.

[31] K. Copenhaver, Y. Hamada, S. Mueller, and J. B. Dunn, "Examining the Characteristics of the Cropland Data Layer in the Context of Estimating Land Cover Change," ISPRS International Journal of Geo-Information, vol. 10, no. 5, Art. no. 5, May 2021, doi: 10.3390/ijgi10050281.

[32] L. Lin et al., "Improvement and Validation of NASA/MODIS NRT Global Flood Mapping," Remote Sensing, vol. 11, no. 2, Art. no. 2, Jan. 2019, doi: 10.3390/rs11020205.

[33] H. Zhao, L. Di, Z. Sun, E. Yu, C. Zhang, and L. Lin, "Validation and Calibration of HRLDAS Soil Moisture Products in Nebraska," in 2022 10th International Conference on Agro-geoinformatics (Agro-Geoinformatics), Jul. 2022, pp. 1–4. doi: 10.1109/Agro-Geoinformatics55649.2022.9858974.

[34] Md. S. Rahman et al., "Comparison of selected noise reduction techniques for MODIS daily NDVI: An empirical analysis on corn and soybean," in 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Jul. 2016, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2016.7577661.

[35] H. Zhao et al., "Impacts of Soil Moisture on Crop Health: A Remote Sensing Perspective," in 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Jul. 2021, pp. 1–4. doi: 10.1109/Agro-Geoinformatics50104.2021.9530318.