Learning Visual Patterns in Remote Sensing: An Overview of Agricultural Applications

Mateus Pinto da Silva, Mariana A. R. Schaefer
Hugo N. Oliveira, Julio C. S. Reis
Universidade Federal de Viçosa
Viçosa, Brazil
{mateus.p.silva,mariana.schaefer,hugo.n.oliveira, jreis}@ufv.br

Ian M. Nunes
Brazilian Institute of
Geography and Statistics
Rio de Janeiro, Brazil
ian.nunes@ibge.gov.br

Jefersson A. dos Santos
University of Sheffield
Sheffield,
United Kingdom
j.santos@sheffield.ac.uk

***Abstract*—Deep Learning based on Remote Sensing has become a powerful tool to increase agricultural productivity, mitigate the effects of climate change, and monitor deforestation. However, there is a lack of standardization and appropriate taxonomic classification of the literature available in the context of informatics. Against this background, this survey provides an overview of the relevant literature categorized into five main applications: Parcel Segmentation, Crop Mapping, Crop Yielding, Land Use and Land Cover, and Change Detection. We address notable trends, including transitioning from traditional to deep learning, convolutional models, recurrent and attention-based models, generative strategies, and self-supervised pre-training. The supplementary material also includes a comprehensive review of publicly available datasets for these applications. We hope that our work can be useful as a guide for future work in this context.**

***Index Terms*—Agricultural Remote Sensing, Deep Learning, Crop Mapping, Crop Yielding, Parcel Segmentation, Land Use, Land Cover, Change Detection**

I. Introduction

The Green Revolution (1960s and 1970s) led to a considerable increase in agricultural production through technologies such as pesticides, fungicides, herbicides, and genetically modified crops. This development was coupled with the widespread establishment of monocultures and improved machinery for practically all agricultural processes. Today, the population continues to grow, so another leap in agricultural productivity is needed [1]. This leap will be even harder to achieve, as it is essential to mitigate the negative impact of agribusiness on the environment. Moreover, such a leap must be implemented during climatic problems that are already occurring both passively (rising temperatures, drought) and actively (floods, hurricanes, etc.).

In this context, Remote Sensing (RS) is a candidate tool to help accomplish this new "green revolution" [2]. With RS, it's possible to check whether crops are being planted and harvested in the most suitable conditions and enables the use of precision agriculture [3]. In addition, RS makes it possible to actively monitor deforestation and check which use or cover has been identified in the deforested area [4]. It is also possible to anticipate food shortages predicting productivity for different regions and products [5]. A graphical depiction of some applications related to agricultural RS can be seen in Figure 1.

However, for large-scale crop monitoring, the price of expert data labeling can be prohibitive. The existence of different types of RS data makes the labeling process even more expensive. Therefore, machine learning (ML) is a relevant tool for speeding up large-scale RS tasks. The ability to extract deep correlations of variables and various modeling techniques allow the use of heterogeneous data [6], and it is possible to scale these methods for near real-time global monitoring. In this work, we focus on Deep Learning (DL) methods applied in agriculture. We also emphasize applications that enable more sustainable or productive agricultural management in addition to those that focus on preventing deforestation.

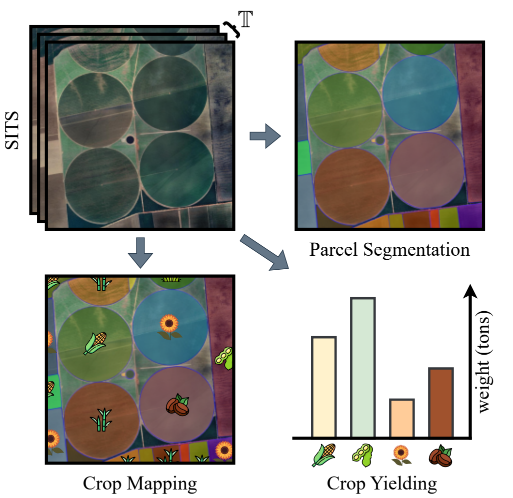
Further reviews have been produced on research areas that complement the areas covered in this review. The work conducted by Yuan et al. [7], for instance, serves as good introductory material for the field, with a focus on environmental applications. In contrast, Ma et al. [8] covered both urban and agricultural RS, but focuses on urban applications. The most cited application is Land Use and Land Cover (LULC), with Convolution Neural Networks (CNNs) being the most used architecture, five times more than the second-ranked model, AEs (autoencoders).

However, due to its release date, this review did not cover the current rapid growth of attention-based architectures in agricultural remote sensing applications.

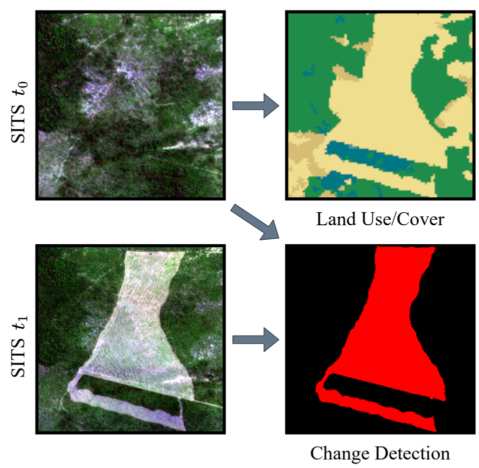
Li et al. (2022) conducted a study on multimodal Data Fusion (DF) in remote sensing, focusing on tasks like Classification, Cloud Removal, and Object Detection. The study utilized various types of data, including Medium Resolution (MR) images, High Resolution (HR) images, Light Detection and Ranging (LiDAR), and Synthetic Aperture Radar (SAR) data. The authors noted a rapid increase in publications on this topic but pointed out the absence of datasets for agricultural applications in their work.

Aleissaee et al. [9] mapped the use of Vision Transformers (ViTs) in RS. Primarily Very High Resolution (VHR) hyperspectral images are used, sometimes in combination with SAR data. The survey also noted that pre-training using RS data instead of natural images shows better results. However,

The authors would like to thank FAPESP (grants #2015/22308-2, #2017/50236-1, #2020/06744-5), Serrapilheira Institute (grant #R-2011-37776), CNPq, LNCC/MCTIC, IBGE/MPO, and CAPES for their financial support for this research.



(a)



(b)

Fig. 1: This figure shows illustrative examples of the applications covered in this work. In Figure 1a, we can observe parcel segmentation and crop mapping/yielding prediction illustrated over multitemporal images. In Figure 1b, two timestamps are used to compute land cover/usage and change detection. Most agricultural-related applications rely on multitemporal data across a range $\mathbb{T}$ of acquisition dates (a) or a pair t_0 and t_1 of dates (b).

this did not include unsupervised pre-training, also known as Self-Supervised Learning (SSL), and methods that operate on Satellite Image Time Series (SITS).

As noted in previous efforts [10], following the trend of traditional computer vision literature, RS researchers have started to adopt diffusion models, especially since 2023. Although the first applications were for data synthesis (i.e., Super Resolution and Cloud Removal), there are already applications for image interpretation. However, the authors did not specifically address agricultural applications.

Joshi et al. [11] reviewed Crop Mapping and Yield. In contrast to the previously mentioned surveys, mainly Low-Resolution (LR) and MR data are used here. Sentinel[1] data (optical and radar data) are usually used for crop mapping, while the most commonly used satellite for crop yield is MODIS[2], often in conjunction with weather and ground data. The most commonly applied methods are adaptations of CNNs and Recurrent Neural Networks (RNNs), with an emphasis on methods that explicitly deal with SITS. In addition, attention mechanisms are often built into these architectures.

Most recently, Kamilaris and Prenafeta-Boldú [12] published the latest survey on DL in agricultural RS. It mentions that there are few public agricultural remote sensing datasets, which continue to this day. However, due to the timing of the publication, no new techniques are presented.

We provide a comprehensive overview of publicly available datasets for agricultural RS for the mapped applications, in addition to the datasets made for SSL pretraining as separate Complementary Materials[3] due to space constraints. We also provide a taxonomy to organize the articles presented in this work in Figure 2.

The remaining sections of this paper discuss applications directly related to agriculture, such as Parcel Segmentation (Section II), Crop Mapping/Classification (Section III), and Crop Yielding (Section IV). This is followed by a review of related applications to agricultural settings, such as Land Cover/Use (Section V) and Change Detection (Section VI). Finally, Section VII presents concluding remarks.

II. Parcel Segmentation

Parcel segmentation, also known as cropland delineation, is the application that separates a given region into agricultural plots. This application operates over an image or time series of images, aiming to output a list of polygons circumventing agricultural tiles.

In general, authors used primarily CNNs or variations for agricultural parcel segmentation. Xu et al. [13] used a U-Net with depthwise separable convolution layers called DSCUnet to raw segment and classify the agricultural areas then a Richer Convolutional Features Network was employed to fine delineate the boundaries of the parcels from agricultural classified regions obtained by the former step, using single VHR images from the Gaofen-2 satellite. The authors pointed to misclassification issues when the study area contains small buildings near the agricultural parcels. Xie et al. [14] propose a High-Resolution Network (HRNet), a CNN integrated with an attention module that retains feature information at different scales through parallel multiresolution branches. These branches are merged and fed into two modules: an object-contextual module, which enhances the representation of contextual information and outputs parcel edge results, and a connectivity attention module, designed to extract connection

[1] https://sentinels.copernicus.eu/web/sentinel/home
[2] https://modis.gsfc.nasa.gov/
[3] https://github.com/mateuspinto/rs-agri-survey, also in the IEEE Supplementary Materials

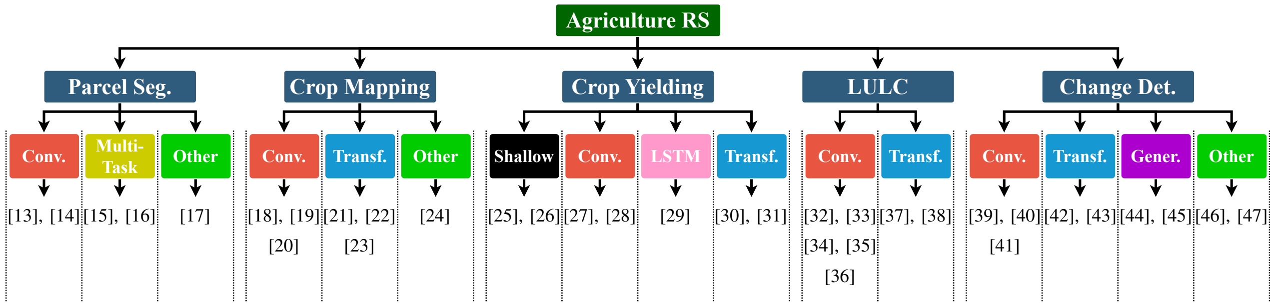


Fig. 2: Classification of the selected articles under the proposed categories of the taxonomy presented in this work. Each category can be further divided into more refined groups according to the methods' characteristics. Each method may fall under more than one group, as they are not mutually exclusive.

and directional information. The outputted parcels are post-processed to merge erroneously over-segmented plots.

A growing trend in parcel segmentation is the use of siamese/multi-stream architectures. Li et al. [15] used a two-branched convolutional architecture (TSANet): the first branch was responsible for drawing parcels, and the second one was to fine-delimit their boundaries. The two pieces of information were combined using a multitask loss formed by a Binary Cross-Entropy and a Dice Loss to deal with class imbalance and training instability. The method was tested for single-image segmentation using Gaofen-2 and Sentinel-2 satellites. Yan et al. [16] also used a dual-stream architecture with each branch using the same idea as the former paper but with SITS as input and weighting joining the same losses instead of a simple sum, which allows choosing between better pixel-level metrics or more consistent parcels. This approach is very similar to TSANet, and both were tested in the same study region (almost the whole Netherlands) with Sentinel-2, although on different Datasets, achieving similar results.

At last, Kerner et al. [17] showed that datasets do not generalize well to different regions and presented the possibility of simple Transfer Learning of a dataset from a different geographic region or satellite to a different target dataset improves the metrics in the target dataset. The authors showed that this approach works best for datasets with little labeled data, achieving an F1 Score improvement of up to 6%.

CNNs [14] seem to have reached their limit in this task, and researchers have begun testing new architectures, the most prominent of which focus on dual stream and/or multitasking [15], [16], removing noise from non-existent plots. The literature reports a great variety of available datasets for parcel segmentation, although there is little geographical variety [17].

III. Crop Mapping

Crop mapping is the application in which, given a parcel or a map, a classification per parcel or pixel is generated for certain crops, usually using a time series of RS images. Datasets from a given region may not generalize well to other areas due to climatic or agricultural calendar differences [48].

CNNs are one of the most used strategies for Crop Mapping, as in Gallo *et al.* [18] used a 3D CNN Features Pyramid Network (FPN) to classify SITS crops pixel-wise using multi-spectral images from Sentinel-2, using a CELoss to construct the Crop Parcels and an MSELoss to determine in which time interval it is more probable to be Crop Season. Integrating attention mechanisms in a CNN is the strategy adopted by Farmonov et al. [19] to extract long-term features and classify crops at the pixel level, achieving 97.89% of Overall Accuracy (OA) with hyperspectral images. The method proposed by Yaramasu *et al.* [20] used a bidirectional ConvLSTM as a temporal encoder with a pre-trained ImageNet VGG11 as a spatial encoder to predict crop mapping using crop rotation from previous years with Landsat SITS as training.

Transformers have recently gained popularity due to their ability to learn representations of data from any position within a sequence, handling effectively varied sizes of time series. Yuan et al. [21] leverage generative pretraining for Bidirectional Encoder Representations for Transformers (BERT) models via masking a pixel SITS with high-value pixels to simulate clouds and training the network for reconstructing it, followed by fine-tuning to incorporate semantic information about crop types, and it used the Positional Encoder (PE) of the Transformer with the Day Of the Year (DOY) to integrate information from the agricultural calendar. Xu et al. [22] surpassed the performance of this masking-based strategy by pretraining using a similarity-based contrastive SSL Loss and using an NDVI Seq2Vec instead of a max-pooling to reduce the temporal dimension to obtain a representation vector, achieving an improvement of 1% in OA. Abbas et al. [23] proposed a temporal AE capable of learning a crop mapping representation from a given region, then retrained in another area with a different agricultural calendar, done by transformers PE output computed on the source region is used as a proxy to quantify the temporal shift concerning the PE output obtained on the target region.

Russwurm *et al.* [24] presented a classification head that can be attached to any SITS crop mapping model, enabling In-Season Crop Mapping. This classification head outputs the probability of the classification being correct given the number of timestamps compared to the full season, trained with an Earliness-Rewarded Loss, turning the model into a multimodal approach. The methodology was tested with an LSTM model, being able to predict the Crops using from 16% to 40% of the SITS.

Crop Mapping is a task that involves SITS from one or more different data sensors. Thus, CNNs [18] and LSTMs [20] have been gradually being replaced by Transformers [22], [23], given their ability to learn representations of data from long distances in a sequence. Furthermore, due to the abundance of unlabeled RS data, SSL [21], [22] began to be adopted and brought significant results, including the need for less labeled data.

IV. Crop Yielding

Crop yield prediction is a regression application that typically uses the geometry of a crop region and growth cycle dates to forecast production. Several methods combine satellite images with other sources of information, such as weather data and soil information. Regarding data availability, there are a few datasets, some with a limited data volume, and most are private.

Shallow Learning methods are still widely used. Sabo *et al.* [25] compared DL and shallow methods for crop yielding, concluding that the former is the best-performing method, suggesting that, for larger data sets, deep learning methods should perform better. Lang *et al.* [26] conclude that the LSTM model performed better than shallow methods.

CNN BiGRU network with attention layers proposed by Lu *et al.* [27] to predict soybeans production using RS, climate data, and photosynthetic-related parameters. Leveraging training data from the USA and using it in Argentina, Huber *et al.* [28] presented Deep Transfer Learning techniques to overcome catastrophic forgetting and negative transfer problems.

Jeong *et al.* [29] used an LSTM forwarded by 1D CNN architecture on MODIS SITS and meteorological data time series, fused with geographic data (such as coordinates and country identifiers) processed by a Dense Network for early South and North Korea rice yield prediction achieving a RMSE of $0.61\,\mathrm{Mg\,ha^{-1}}$.

Improving temporal dependencies, Liu *et al.* [30] presented a Transformer architecture followed by a convolutional layer using satellite and environmental data, also presenting a graphical analysis of the explainability of attention mechanisms, predicting rice yield two months before harvest. Lin *et al.* [31] used a Multi-modal ViT with Sentinel 2 and weather data, claiming that the method is climate-aware and presented better results than the Convolutional LSTM hybrid models, also comparing SimCLR pre-training and a multi-modal pre-training. The first type of pre-training presented an improvement of less than 0.3 in RMSE (since SimCLR can not process climate data), and the second presented a significant improvement greater than 1.2 in the same metric.

The literature indicates that complementary data can be used in addition to RS, such as weather/climate data, soil data, and crop species information. According to Sabo *et al.* [25], shallow models are better suited and capable of producing good results with less data than DL models. DL tackles temporal dependencies better, and architectures like LSTM and Transformers are gaining traction in the community [29], [30]. Transfer learning and generation of synthetic data are trends [28].

V. Land Use and Land Cover

Land use and land cover (LULC) is a segmentation task that provides information to help understand the landscape. Producing LULC maps is one of the most common tasks using RS data. We can identify several possible uses in agriculture, such as checking changes in water bodies, mapping natural vegetation, soil deterioration, deforestation, or artificial structures.

SL methods are still used for LULC, especially for LR/MR multispectral satellite imagery. Dou *et al.* [32] argued that the limited quantity of multispectral bands in Landsat Satellites leads CNNs to problems in stable features from the spectral domain, using CNNs to generate Deep Features to use as the input to Shallow Learning Classifiers producing probabilities to use as the input to another CNN and finally producing LULC predictions with OA of 88.95%.

Sun *et al.* [33] introduced an end-to-end FCN for hyperspectral image (HSI) segmentation that classifies all pixels in an HSI cube simultaneously. Zhan *et al.* [34] enhanced accuracy through multiscale feature reconstruction and interclass attention weighting, tackling issues like ambiguous boundaries and intraclass variance. Zhu *et al.* [35] added a Global Joint Attention to enhance spectral and spatial feature discrimination. Similarly, Farmonov *et al.* [36] applied a Wavelet-Attention on a CNN with a spectral attention mechanism to improve crop type mapping accuracy.

Sun *et al.* [37] combined CNN and Transformer models, using 3D and 2D convolutions for low-level feature extraction, a Gaussian-weighted tokenizer for feature transformation, and a transformer encoder for learning deep feature representations. Yao *et al.* [38] used Transformers in a multimodal approach to include heterogeneous RS data, using parallel branches of position-shared ViTs extended with separable convolution modules, each branch made to process a type of data. Achieving 88.95% of OA using hyperspectral and LiDAR data and balancing the weights of each branch, and 87.71% using hyperspectral data only.

LULC studies large and heterogeneous area datasets, and the target mapping labels usually include natural or anthropized classes. Even considering only farms, one can identify several classes, such as water bodies, buildings, roads, farmlands, and forests. The application has been using SL for LR/MR or multispectral satellites due to the low number of features possible to be generated [32]. As for HR/VHR or hyperspectral satellites, the research area has been following closely the area of computer vision, replacing traditional CNNs [33] with versions with attention modules [49] or even using Transformers [37], [38].

VI. Change Detection

In RS, Change Detection is the classification of dissimilarities in a given region, given two (or more) timestamps as input [50]. Such a task is vital for active monitoring in near real-time. Amongst CNN models, Ye *et al.* [39] used symmetric and siamese networks to compare different timestamps. U-Net-like architectures [40] tend to perform well, but transformer-based models perform better and take advantage of large volumes of data. In addition to optical data, Deep Cascade Network for Change Detection [41] uses SAR data and enables the method to avoid clouds.

Transformers were used in Change Detection due to their ability to function as a global and generic feature extractor. Chen *et al.* [42] used a symmetric architecture that mixes a ResNet without as the backbone and a Transformer as an encoder/decoder to highlight the most relevant semantic features of each image underlying modified objects in each dataset. Yan *et al.* [43] achieved better results using a symmetric Swin Transformer as the backbone with a Progressive Attention Module method for pixel-wise classification of image changes.

More recently, diffusion models like Bandara *et al.* [44] used a U-Net pre-trained in SSL way to remove synthetic noise added to RGB Sentinel-2 images and finetuned a simple classifier that uses features generated for a before and after image to predict the Change Detection map. The authors argued that such generative pre-training makes the methodology more invariant both to problems in data generation and storage (noise/blur in image capture, etc) and to variations inherent to the environment (such as seasonal change). Jia *et al.* [45] a diffusion U-Net is trained end-to-end fashion to generate the Change Detection map, achieving a slight improvement of 0.06% OA.

Other architectures and paradigms used, like unsupervised methods by Du *et al.* [46] that used a symmetric network model for pseudo-label generation, or Tang *et al.* [47] that used Graph Neural Networks (GNN) and Metric Learning.

Change Detection is widely explored in the literature and has several large and well-known datasets enabling the comparison between models. Diffusion Models [44], [45] have more recently appeared as an attempt to create more general models than transformer-based models [42], [43] that are gradually replacing CNN ones [39], [40].

VII. Conclusion

In this survey, we have presented a new taxonomy to categorize DL Techniques for agricultural RS. We have provided an overview of the applications of Crop Mapping, Crop Yielding, LULC, and Parcel Segmentation and discussed the methods used in each. We show that there have been many changes in the research field, such as the introduction of DL [13] instead of SL models [25], confirming the statements of [12]. This is partly due to the greater availability of data from HR/VHR or Hieperspectral satellites, as the increase in information collected by SL models [26] is not well accounted for. Although the literature generally seems to be moving towards Transformer models [9] to deal with time series, CNNs [19] fitted with attention modules have been widely used for this and some diffusion models [44], [45] are appearing on the horizon.

SSL also appears to be a new trend. Of the five mapped applications, three already represent adaptations of this methodology (Crop Mapping [21], [22], Crop Yielding [31] and Change Detection [44], [45]), with the use of generative and contrastive pre-training, both based on techniques already used in the field of computer vision, as well as more specific techniques developed to solve RS problems. It is also worth noting that, according to our research, no LULC work uses SSL pretraining, although this seems to be possible due to the type of architecture used in the latest methods in the field [38].

However, some problems have also been identified in the literature. The datasets of most of the applications illustrated here (Crop Mapping, Crop Yielding, and Parcel Segmentation) do not generalize well between regions or are sparse [17], [28], which makes direct comparisons between the methods difficult. In addition, the satellites are very different, with different sensor types and resolutions, and a method that works well for one region may not work well for another [32]. We expect the community's interest in RS Deep Learning to increase, considering that it can help increase agricultural production while having less or no impact on the environment.

References

[1] D. Llewellyn, "Does Global Agriculture Need Another Green Revolution?" *Engineering*, vol. 4, no. 4, pp. 449–451, 2018.

[2] Z. Ahmed, A. Shew, L. Nalley, M. Popp, V. S. Green, and K. Brye, "An Examination of Thematic Research, Development, and Trends in Remote Sensing Applied to Conservation Agriculture," *International Soil and Water Conservation Research*, vol. 12, no. 1, pp. 77–95, 2024.

[3] S. Liaghat, S. K. Balasundram *et al.*, "A Review: The Role of Remote Sensing in Precision Agriculture," *American Journal of Agricultural and Biological Sciences*, vol. 5, no. 1, pp. 50–55, 2010.

[4] D. S. Candra, "Deforestation Detection Using Multitemporal Satellite Images," in *IOP Conference Series: Earth and Environmental Science*, vol. 500, no. 1. IOP Publishing, 2020, p. 012037.

[5] J. A. Quinn, W. Okori, and A. Gidudu, "Increased-Specificity Famine Prediction Using Satellite Observation Data," in *ACM Symposium on Computing for Development*, 2010, pp. 1–6.

[6] J. Li, D. Hong, L. Gao, J. Yao, K. Zheng, B. Zhang, and J. Chanussot, "Deep Learning in Multimodal Remote Sensing Data Fusion: A Comprehensive Review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102926, 2022.

[7] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang *et al.*, "Deep Learning in Environmental Remote Sensing: Achievements and Challenges," *Remote Sensing of Environment*, vol. 241, p. 111716, 2020.

[8] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019.

[9] A. A. Aleissaee, A. Kumar, R. M. Anwer, S. Khan, H. Cholakkal, G.-S. Xia, and F. S. Khan, "Transformers in Remote Sensing: A Survey," *Remote Sensing*, vol. 15, no. 7, p. 1860, 2023.

[10] Y. Liu, J. Yue, S. Xia, P. Ghamisi, W. Xie, and L. Fang, "Diffusion Models Meet Remote Sensing: Principles, Methods, and Perspectives," *arXiv preprint arXiv:2404.08926*, 2024.

[11] A. Joshi, B. Pradhan, S. Gite, and S. Chakraborty, "Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review," *Remote Sensing*, vol. 15, no. 8, p. 2014, 2023.

[12] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[13] L. Xu, D. Ming, T. Du, Y. Chen, D. Dong, and C. Zhou, "Delineation of Cultivated Land Parcels Based on Deep Convolutional Networks and Geographical Thematic Scene Division of Remotely Sensed Images," *Computers and Electronics in Agriculture*, vol. 192, p. 106611, 2022.

[14] Y. Xie, S. Zheng, H. Wang, Y. Qiu, X. Lin, and Q. Shi, "Edge Detection with Direction Guided Postprocessing for Farmland Parcel Extraction," *IEEE JSTARS*, vol. 16, pp. 3760–3770, 2023.

[15] M. Li, J. Long, A. Stein, and X. Wang, "Using a Semantic Edge-Aware Multi-Task Neural Network to Delineate Agricultural Parcels from Remote Sensing Images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 200, pp. 24–40, 2023.

[16] S. Yan, X. Yao, J. Sun, W. Huang, L. Yang, C. Zhang, B. Gao, J. Yang, W. Yun, and D. Zhu, "TSANet: A Deep Learning Framework for the Delineation of Agricultural Fields Utilizing Satellite Image Time Series," *Computers and Electronics in Agriculture*, vol. 220, p. 108902, 2024.

[17] H. Kerner, S. Sundar, and M. Satish, "Multi-Region Transfer Learning for Segmentation of Crop Field Boundaries in Satellite Images with Limited Labels," *arXiv preprint arXiv:2404.00179*, 2024.

[18] I. Gallo, R. La Grassa, N. Landro, and M. Boschetti, "Sentinel 2 time series analysis with 3d feature pyramid network and time domain class activation intervals for crop mapping," *ISPRS International Journal of Geo-Information*, vol. 10, no. 7, p. 483, 2021.

[19] N. Farmonov, K. Amankulova, J. Szatmári, A. Sharifi, D. Abbasi-Moghadam, S. M. M. Nejad, and L. Mucsi, "Crop type classification by desis hyperspectral imagery and machine learning algorithms," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 16, pp. 1576–1588, 2023.

[20] 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, 2020.

[21] Y. Yuan, L. Lin, Q. Liu, R. Hang, and Z.-G. Zhou, "SITS-Former: A Pre-Trained Spatio-Spectral-Temporal Representation Model for Sentinel-2 Time Series Classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 106, p. 102651, 2022.

[22] Y. Xu, Y. Ma, and Z. Zhang, "Self-Supervised Pre-Training for Large-Scale Crop Mapping Using Sentinel-2 Time Series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 207, pp. 312–325, 2024.

[23] A. Abbas, M. Linardi, E. Vareille, V. Christophides, and C. Paris, "Towards Explainable AI4EO: An Explainable Deep Learning Approach for Crop Type Mapping Using Satellite Images Time Series," in *IGARSS*. IEEE, 2023, pp. 1088–1091.

[24] M. Rußwurm, N. Courty, R. Emonet, S. Lefèvre, D. Tuia, and R. Tavenard, "End-to-End Learned Early Classification of Time Series for In-Season Crop Type Mapping," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 445–456, 2023.

[25] F. Sabo, M. Meroni, F. Waldner, and F. Rembold, "Is Deeper Always Better? Evaluating Deep Learning Models for Yield Forecasting with Small Data," *Environmental Monitoring and Assessment*, vol. 195, no. 10, p. 1153, 2023.

[26] P. Lang, L. Zhang, C. Huang, J. Chen, X. Kang, Z. Zhang, and Q. Tong, "Integrating Environmental and Satellite Data to Estimate County-level Cotton Yield in Xinjiang Province," *Frontiers in Plant Science*, vol. 13, p. 1048479, 2023.

[27] J. Lu, H. Fu, X. Tang, Z. Liu, J. Huang, W. Zou, H. Chen, Y. Sun, X. Ning, and J. Li, "GOA-Optimized Deep Learning for Soybean Yield Estimation Using Multi-Source Remote Sensing Data," *Scientific Reports*, vol. 14, no. 1, p. 7097, 2024.

[28] F. Huber, A. Inderka, and V. Steinhage, "Leveraging Remote Sensing Data for Yield Prediction with Deep Transfer Learning," *Sensors*, vol. 24, no. 3, p. 770, 2024.

[29] S. Jeong, J. Ko, and J.-M. Yeom, "Predicting Rice Yield at Pixel Scale Through Synthetic Use of Crop and Deep Learning Models with Satellite Data in South and North Korea," *Science of The Total Environment*, vol. 802, p. 149726, 2022.

[30] Y. Liu, S. Wang, J. Chen, B. Chen, X. Wang, D. Hao, and L. Sun, "Rice Yield Prediction and Model Interpretation Based on Satellite and Climatic Indicators Using a Transformer Method," *Remote Sensing*, vol. 14, no. 19, p. 5045, 2022.

[31] F. Lin, S. Crawford, K. Guillot, Y. Zhang, Y. Chen, X. Yuan, L. Chen, S. Williams, R. Minvielle, X. Xiao *et al.*, "MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer," in *ICCV*, 2023, pp. 5774–5784.

[32] P. Dou, H. Shen, C. Huang, Z. Li, Y. Mao, and X. Li, "Large-scale land use/land cover extraction from landsat imagery using feature relationships matrix based deep-shallow learning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 129, p. 103866, 2024.

[33] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 2810–2825, 2021.

[34] Z. Zhan, Z. Xiong, X. Huang, C. Yang, Y. Liu, and X. Wang, "Multi-scale feature reconstruction and inter-class attention weighting for land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[35] Q. Zhu, W. Deng, Z. Zheng, Y. Zhong, Q. Guan, W. Lin, L. Zhang, and D. Li, "A spectral-spatial-dependent global learning framework for insufficient and imbalanced hyperspectral image classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 11 709–11 723, 2021.

[36] N. Farmonov, K. Amankulova, J. Szatmári, A. Sharifi, D. Abbasi-Moghadam, S. M. M. Nejad, and L. Mucsi, "Crop type classification by desis hyperspectral imagery and machine learning algorithms," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 16, pp. 1576–1588, 2023.

[37] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[38] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (exvit) for land use and land cover classification: A multimodal deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[39] Y. Ye, M. Wang, L. Zhou, G. Lei, J. Fan, and Y. Qin, "Adjacent-level Feature Cross-Fusion with 3D CNN for Remote Sensing Image Change Detection," *IEEE TGRS*, 2023.

[40] L. Miao, X. Li, X. Zhou, L. Yao, Y. Deng, T. Hang, Y. Zhou, and H. Yang, "SNUNet3+: A Full-Scale Connected Siamese Network and a Dataset for Cultivated Land Change Detection in High-Resolution Remote-Sensing Images," *IEEE TGRS*, 2023.

[41] Y. Gao, F. Gao, J. Dong, and S. Wang, "Change Detection from Synthetic Aperture Radar Images Based on Channel Weighting-based Deep Cascade Network," *IEEE JSTARS*, vol. 12, no. 11, pp. 4517–4529, 2019.

[42] H. Chen, Z. Qi, and Z. Shi, "Remote Sensing Image Change Detection with Transformers," *IEEE TGRS*, vol. 60, pp. 1–14, 2021.

[43] T. Yan, Z. Wan, P. Zhang, G. Cheng, and H. Lu, "TransY-Net: Learning Fully Transformer Networks for Change Detection of Remote Sensing Images," *IEEE TGRS*, 2023.

[44] W. G. C. Bandara, N. G. Nair, and V. M. Patel, "DDPM-CD: Denoising Diffusion Probabilistic Models as Feature Extractors for Change Detection," 2024.

[45] J. Jia, G. Lee, Z. Wang, L. Zhi, and Y. He, "Siamese Meets Diffusion Network: SMDNet for Enhanced Change Detection in High-Resolution RS Imagery," 2024.

[46] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised Deep Slow Feature Analysis for Change Detection in Multi-Temporal Remote Sensing Images," *IEEE TGRS*, vol. 57, no. 12, pp. 9976–9992, 2019.

[47] X. Tang, H. Zhang, L. Mou, F. Liu, X. Zhang, X. X. Zhu, and L. Jiao, "An Unsupervised Remote Sensing Change Detection Method Based on Multiscale Graph Convolutional Network and Metric Learning," *IEEE TGRS*, vol. 60, pp. 1–15, 2021.

[48] K. K. Gadiraju and R. R. Vatsavai, "Remote sensing based crop type classification via deep transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4699–4712, 2023.

[49] X. Zheng, H. Sun, X. Lu, and W. Xie, "Rotation-invariant attention network for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 31, pp. 4251–4265, 2022.

[50] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep Learning-Based Change Detection in Remote Sensing Images: A Review," *Remote Sensing*, vol. 14, no. 4, 2022.