



RSP Science Hub

International Research Journal on Advanced Science Hub  
2582-4376  
Vol. 05, Issue 05S May  
[www.rspsciencehub.com](http://www.rspsciencehub.com)



<http://dx.doi.org/10.47392/irjash.2023.S068>

International Conference on intelligent COMPUTing TEchnologies and Research (i-COMPUTER) 2023

## Crop Mapping using Multispectral Sentinel-2 Dataset

Ghantasala Mahathi <sup>1</sup>, Bala Charvitha Sumanjali <sup>1</sup>, Abhinaya P <sup>1</sup>, Venkatesan M <sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, National Institute of Technology, Karaikal, Puducherry, India

<sup>2</sup>Assistant Professor, Head of the Department, Department of Computer Science and Engineering, National Institute of Technology, Karaikal, Puducherry, India

Email: [mahathig2001@gmail.com](mailto:mahathig2001@gmail.com)

### Article History

Received: 26 February 2023

Accepted: 21 March 2023

### Keywords:

Remote Sensing;  
Crop Mapping;  
Deep Learning;  
Convolutional Neural Network (CNN);  
Transformer;  
Sentinel2

### Abstract

Accurate and timely information on crop distribution is crucial for decision-making in agriculture and ensuring global food security. Crop mapping using remote sensing data has become an essential tool for agricultural monitoring and management. The process of crop mapping involves the acquisition of multispectral data from satellites, pre-processing of the data and analysis to identify different crop types based on their spectral signatures. This information is then combined with ground truth data to create accurate crop mappings that show the location and extent of different crops within an area. In recent years, Convolutional Neural Network (CNN) models have been used for crop mapping using Sentinel-2 data. However, CNN models may not be effective in capturing the spatial dependencies between features extracted from multispectral data. To address this issue, we propose a transformer model. The proposed transformer model is compared with the CNN model to demonstrate its effectiveness and accuracy for crop mapping. This study demonstrates the potential of the Transformer model in capturing the spatial dependencies between features and efficiently processing long sequences of data, contributing to improved agricultural practices, resource management and food security.

### 1. Introduction

Crop mapping is a crucial task in precision agriculture, allowing farmers to make informed decisions about crop management and maximize their yield. (Gallo et al.) The use of satellite imagery has revolutionized this field, enabling large-scale and cost-effective crop monitoring. The Sentinel-2 satellite, launched by the European Space Agency (ESA), provides multispectral imagery with 13 spectral bands at a spatial resolution of up to 10 meters, making it a valuable resource for crop mapping. (Z. Li, Chen, and Zhang)

To extract useful information from this imagery, machine learning techniques have been widely used,

in this paper we have particularly concentrated on convolutional neural networks (CNNs) and transformer models. (Bradley et al.) CNN models are well-suited for analysing image data, as they can learn spatial features from the input data and classify different types of crops. On the other hand, Transformer models have shown remarkable performance in natural language processing (NLP) tasks but also achieved success in remote sensing image classification tasks. (Teke et al.)

In crop mapping, Transformer models can be applied by treating the satellite imagery as a sequence of pixels and extracting spatial relationships between these pixels using self-attention

mechanisms. This allows the model to capture global dependencies within the image and has shown promising results in image classification tasks. (Zhou, S. Li, and Shao)

Both CNN and Transformer models have their strengths and weaknesses. CNNs are better suited for feature extraction from image data, while Transformer models are more effective in capturing long-range dependencies and patterns in sequential data. The choice between the two models ultimately depends on the specific application and the nature of the data being analysed. (Feng *et al.*)

In summary, crop mapping using Sentinel-2 multispectral dataset can be achieved using both CNN and Transformer models. CNN models are well-suited for analysing image data and extracting spatial features, while Transformer models can effectively capture long-range dependencies within sequential data, such as image pixels. The choice between these two models ultimately depends on the specific requirements and nature of the data being analysed. (Nguyen, Robinson, and Galpern)

## 2. Materials and Methodology

In this study, a Convolutional neural network model (CNN model) and the Transformer model were constructed as a classification scheme for crop mapping which is done using multispectral sentinel-2 satellite data. Pre-processing is applied on the multispectral sentinel-2 image dataset which is obtained from the satellite data. The architecture for the CNN and Transformer models used here, is the classification scheme for multitemporal multisensor images from the satellite time series Sentinel-2 data. (Nowakowski *et al.*) The use of satellite imagery has made it possible to accurately map crops over large areas, which is essential for effective agricultural management and planning. This study demonstrates the potential of CNN and Transformer models for remote sensing applications and provides insights for developing more advanced models in the future, contributing to improved agricultural practices, resource management, and food security. (Saini and Ghosh)

### 2.1. Data Set

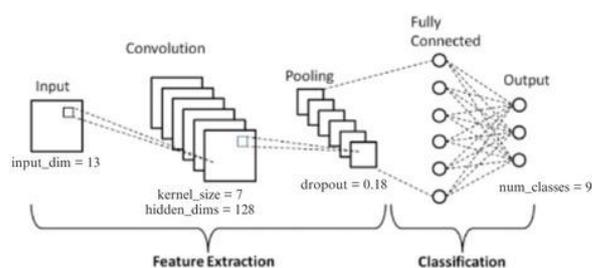
The dataset is composed of Sentinel-2 images. It is a time series satellite data extracted from January 1, 2017, to December 31, 2017. The categories considered are of crop type categories which contain 328

unique crop labels and are grouped into 23 groups. For the Breizh Crops dataset, we selected the 9 following crop categories: barley, wheat, rapeseed, corn, sunflower, orchards, nuts, permanent meadows, and temporary meadows. One of the nine crop type classes are labelled with each individual time series sample in the dataset.

### 2.2. Data Set Pre-processing

Crop types are inferred from the reflectance values of 13 spectral bands by Sentinel-2 data, collected every 2-5 days. Every time an image is collected, the mean over all pixels of one field parcel is calculated and stored as reflectance values that range between 0 and 10000. This is repeated on two processing levels named as L1C and L2A. L1C refers to all the Sentinel 2 data of region in Top-of-Atmosphere whereas L2A refers to the Bottom-of-Atmosphere, cloud filtered and atmospherically corrected data. Dataset images of the same sequence length of 45 are obtained by randomly sampling 45 observations from all available points. Temporal interpolations can also be performed on raw data to ensure the same sequence length of 45 in the dataset. Alternatively, the temporal context of observations of data points can be ignored to extract the statistical features and apply t-SNE embedding (t-distributed stochastic neighbour embedding) on statistical features for suitable transformation.

### 2.3. CNN Model Architecture



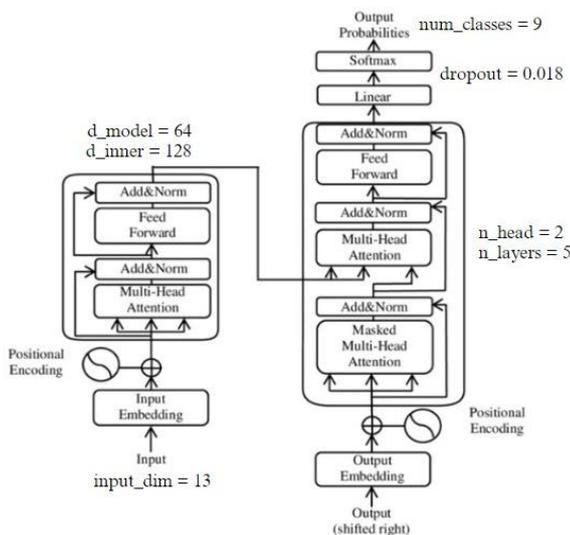
**FIGURE 1. Architecture of CNN Model (Phung and Rhee)**

The architecture of CNN model consists of feed forward neural networks using filters and pooling. The CNN model receives the input as images of dimension 13 and results in output with an approximate category along with its probability of prediction into nine classes. Each layer in the CNN model

detects various features in the image, many such layers are integrated to develop CNN model. A filter also known as kernel, starts as simple features, is applied to each input image and produces an output. This output gets efficient and more detailed after each layer of the CNN model. Through each layer in the architecture, the filters are applied so that features are identified that uniquely represent the input image. The output of each convolution layer becomes the input for the next layer in the architecture. Finally, the Fully-Connected layer in the last part of architecture represents the input image specifically. The outputs obtained from each layer move forward through multiple layers and crop type is identified. The optimal values of hyperparameters used to implement this architecture are mentioned in the attached Figure 1.

#### 2.4. Transformer Model Architecture

The Transformer model captures the spatial dependencies between the extracted features and which in result generates a crop map. In the architecture model, we use transformer model architecture, as it has a very efficient expressive ability in the formation of sequence information to model the multitemporal features which are taken as input.



**FIGURE 2.** Architecture of Transformer Model (Wikipedia)

#### 2.5. The Transformer module mechanism:

The transformer architecture is used for modelling sequence information. In transformer architecture,

a sequential encoding mechanism known as self-attention mechanism is utilized. This mechanism improves the expression ability, which is the relationship between the word sequences, that results in the effective performance of various multiple tasks. In addition to its efficient expression ability, self-attention is comparatively better in parallel ability because it inputs the entire sequence in one go, at a time for training, which in turn improves the training speed of a sequence model in an effective way. When using transformer architecture to build the model for classification, it has the ability to utilize all of the available sequence information to give label as output rather than giving sequence. Therefore, we make use of this encoder module for the classification task, in the transformer architecture without other mechanisms. Two sub-layers are present in the encoder module which are multi-head self-attention and the position-wise fully connected feed-forward network. Residual connections and layer normalizations are also present in each encoder. The input to this model is of 13 dimensions and applies a transformer encoder layer with an embedding dimension of 64 and multi-head attention layers with 2 heads and 5 layers as mentioned above in Figure 2. It applies layer normalization and ReLU activation function.

#### 2.6. Output Layer for Supervised Classification:

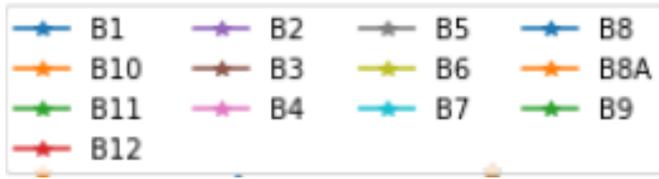
Input sequence is passed through the multilayer transformer encoder module (also through the feature extraction layer) to get the final activation of the encoder. Category features are then extracted by utilizing the multilayer encoder module result which is an output sequence. We derive the deep correlation patterns of multi temporal sequence, by obtaining multilayer encoder modules which are derived from the transformer. We add feed-forward layers followed by SoftMax layers to predict the label of categories of crops.

### 3. Results and Discussion

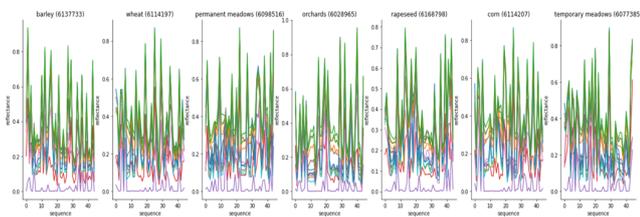
After performing the model construction for CNN (Convolutional Neural Networks) and Transformer models, training is done with the respective Breizh dataset. To understand the working of models with the satellite time series data, a random sample is taken from the dataset and is sent through the models to get the predicted output value. The results of the accurate classification by model for each crop

label, are displayed in the form of bar graphs. The predicted output is depicted along the vertical axis and compared with actual output class.

The graph below represents the reflectance values over the sequence length for each crop label, where a random sample of each crop is considered.



**FIGURE 3.** Color codes

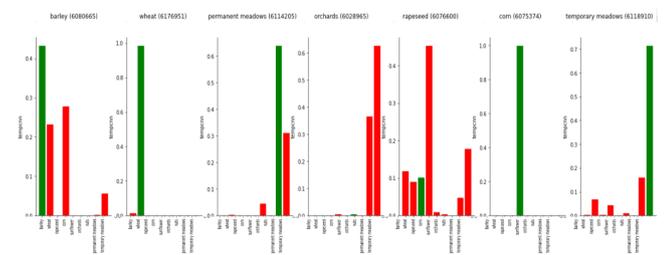


**FIGURE 4.** Graph of reflectance over sequence length for each crop

The reflectance by 13 bands is depicted in the graph. A sequence length of 45 is considered, which indicates that 45 reflectance values measured at different wavelengths and bands are taken as input data for each crop field. Each reflectance value corresponds to a specific band wavelength and the time series represents how the reflectance at each wavelength changes over time.

**3.1. CNN model performance:**

A random id sample for each crop is taken to predict the crop class probabilities from input reflectance data and depicted in bar charts below.



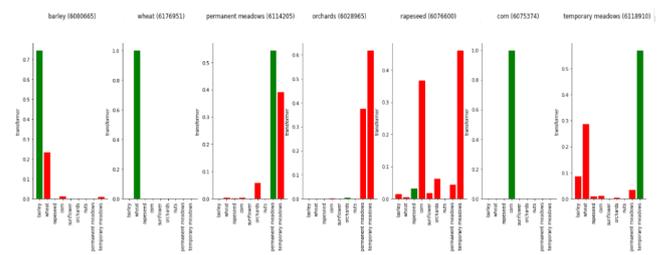
**FIGURE 5.** Predicted probability of crop label vs crop labels

The accuracy of crop identification is fairly high in case of extensively cultivated crops such as wheat

and corn. Less common crops such as orchards and rapeseed were classified with less accuracy. Domination in agricultural areas by corn, wheat, meadow, barley might lead to clear distinction of phenological changes in crop over time. Broad categories such as orchards seemed not clearly distinguishable and hence, low accuracy in crop classification probability.

**3.2. Transformer model performance:**

Random crop samples for each crop class label is processed through a transformer model and predicted probabilities of crop are depicted in bar graphs below.



**FIGURE 6.** Predicted probability of crop label vs crop labels by Transformer model

Phenological characteristics can be easily traced to single specific crop types such as wheat, corn, etc and hence they are predicted with the highest accuracy, around 90%. Barley and meadows are extensively grown in the area and easier to distinguish. Less frequent crop classes such as orchards and rapeseed, are grouped into a broader range of vegetation types which makes them difficult to distinguish, resulting in low prediction accuracy.

The above table compares the percent of correct predictions by CNN and Transformer models for each crop label. The Transformer model shows better results in predicting the probabilities of crop class labels in comparison with the CNN model. This might be due to the fact that the Transformer model may be able to capture the spatial and temporal dependencies between different time points and regions. Also, the Transformer model is known to process sequential data effectively which is a key characteristic of time-series remote sensing data for crop mapping. Since different crops have unique spectral signatures that require special attention, the attention mechanism in transformers allows them to focus on specific parts of the input sequence for

**TABLE 1. Prediction Accuracy by CNN and Transformer for each crop**

| Crops             | Prediction Accuracy-CNN | Prediction Accuracy-Transformer |
|-------------------|-------------------------|---------------------------------|
| Barley            | 0.45                    | 0.75                            |
| Wheat             | 0.95                    | 0.98                            |
| Permanent Meadows | 0.65                    | 0.55                            |
| Orchards          | 0.05                    | 0.05                            |
| Rapeseed          | 0.10                    | 0.08                            |
| Corn              | 0.98                    | 1.00                            |
| Temporary Meadows | 0.70                    | 0.68                            |

effective crop mapping. The spectral reflectance values of crops vary significantly across time and field therefore transformers work better with long range dependencies between different parts of input sequence.

#### 4. Conclusion

In conclusion, both Convolutional Neural Network (CNN) and Transformer models have shown promising results in crop mapping. While CNN models have been traditionally used for image processing tasks, Transformer models have recently gained popularity in various natural language processing and computer vision tasks.

In terms of accuracy, both models have shown comparable results in crop mapping. However, Transformer models offer a more flexible approach in capturing long-term dependencies and encoding spatial information through self-attention mechanisms. On the other hand, CNN models excel in learning hierarchical representations of the input data through convolutional layers, which is useful in capturing local features.

Overall, the choice between CNN and Transformer models for crop mapping ultimately depends on the specific requirements and nature of the dataset. Researchers and practitioners can leverage the strengths of each model to achieve better results in their respective applications.

#### 5. Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

#### References

- Bradley, Bethany A, et al. "A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data". *Remote Sensing of Environment* 106.2 (2007): 137–145.
- Feng, Siwen, et al. "Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.9 (2019): 3295–3306.
- Li, Zhengtao, Guokun Chen, and Tianxu Zhang. "A CNN-Transformer Hybrid Approach for Crop Classification Using Multitemporal Multisensor Images". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020): 847–858.
- Nguyen, Lan H, Samuel Robinson, and Paul Galpern. "Medium-resolution multispectral satellite imagery in precision agriculture: mapping precision canola (*Brassica napus* L.) yield using Sentinel-2 time series". *Precision Agriculture* 23.3 (2022): 1051–1071.
- Nowakowski, Artur, et al. "AI Opportunities and Challenges for Crop Type Mapping Using Sentinel-2 and Drone Data". *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (2021): 258–261.
- Phung, V H and E J Rhee. "A Deep Learning Approach for Classification of Cloud Image Patches on Small Datasets". *J. Inf. Commun. Converg. Eng* 16 (2018): 173–178.
- Saini, R and Sanjay Ghosh. "CROP CLASSIFICATION ON SINGLE DATE SENTINEL-2 IMAGERY USING RANDOM FOREST AND SUPPORT VECTOR MACHINE". *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2018).
- Teke, Mustafa, et al. "A short survey of hyperspectral remote sensing applications in agriculture". *2013 6th International Conference on Recent Advances in Space Technologies (RAST)* (2013): 171–176.
- Wikipedia. "Transformer (machine learning model)". *Wikipedia* (2023).

Zhou, Zhuang, Shengyang Li, and Yuyang Shao. "Crops Classification from Sentinel-2A Multispectral Remote Sensing Images Based on Convolutional Neural Networks". *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (2018): 5300–5303.



© Ghantasala Mahathi et al. 2023 Open Access.

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and

reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Embargo period:** The article has no embargo period.

**To cite this Article:** , Ghantasala Mahathi, Bala Charvitha Sumanjali , Abhinaya P , and Venkatesan M. "Crop Mapping using Multispectral Sentinel-2 Dataset." *International Research Journal on Advanced Science Hub* 05.05S May (2023): 507–512. <http://dx.doi.org/10.47392/irjash.2023.S068>