APPLICATION OF DEEP LEARNING CROP CLASSIFICATION MODEL BASED ON MULTISPECTRAL AND SAR SATELLITE IMAGERY

Y. Qi, G. Bitelli, E. Mandanici, F. Trevisiol*

Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM) – University of Bologna, Viale del Risorgimento 2, 40136 Bologna (BO), Italy, yitian.qi@studio.unibo.it , (gabriele.bitelli, emanuele.mandanici, francesca.trevisiol2)@unibo.it

KEY WORDS: Crop classification, deep learning, Transformer, Land cover, Sentinel, Google Earth Engine.

ABSTRACT:

Classifying crops using satellite data is a challenge, especially since most crops have similar growth cycles. Due to their different characteristics and chlorophyll content, different crops exhibit subtle differences in their reflectance spectra. This study uses a data-driven approach to build a series of deep learning models to classify 36 different land covers in Steele County and Traill Country, North Dakota, US. A Google Earth Engine workflow was implemented to generate a composite layer containing Sentinel 1 and Sentinel 2 satellite data and surface crop data over the study area. 200,000 sample points were generated on this layer, 140,000 for training dataset, 30,000 for validation dataset and 30,000 for testing dataset. Each sample point contains the values of 12 months of SAR and spectral data. In this way, a two-dimensional feature matrix of the time dimension and spectral band dimension (bands refer to specific wavelengths of data in remote sensing imagery and other type of data like NDVI) is generated for each sample point. The training dataset of the model is composed of the feature matrix of these sample points, and the surface crops as labels correspond to the feature matrix. Since this is a dataset with two-dimensional features, this research uses four deep learning models: Dense Neural Network (DNN), Long short-term memory (LSTM), Convolutional neural network (CNN) and Transformer. Among them, the Transformer model based on the self-attention mechanism performed the best, with a comprehensive accuracy rate of 85%, and the classification accuracy rate of crops with more than 2,000 sample points in the training data set reached more than 90%.

1. INTRODUCTION

Application of satellite remote sensing are rising faster and faster in the last years due to the new availability of data and open-data satellite projects. Indeed, following the successful experience of the Landsat program, the European Earth Observation programme “Copernicus” launched the Sentinel missions to provide users with remotely sensed data for environmental monitoring purposes. In particular, the Sentinel-2 and Sentinel-1 missions have been collecting free of charge medium-resolution optical and SAR data with high revisit times since 2015. This huge amount of data represents a powerful source of information that needs to be exploited.

It is evident that local computers are limited in processing that amount of satellite images, due to personal computer memory limits. Therefore, the release of Google Earth Engine (GEE), the cloud-based platform designed to store and process large amounts of geospatial data powered by Google, was a major turning point for the analysis of satellite remote sensing images, enabling long time series analysis (Gorelick et al., 2017). From 2010, GEE makes available to users over 40 years of satellite imagery and other geospatial datasets such as digital terrain models, climate, weather, and demographic data. In addition to this huge data warehouse, GEE offers Google's computational capabilities and algorithms for data processing. The relative simplicity of this tool opens a new frontier for remotely sensed Big Data analysis (Casu et al., 2017), which would normally require significant computing and storage capacity, resulting in large hardware and software costs.

In the last years, the availability of data, together with the computational power provided by platforms like GEE, allowed the spread of the time series (TS) analysis of satellite data in the scientific community, especially in agricultural application. Indeed, time series analysis can be particularly effective for monitoring land cover characterized by seasonal pattern, such as crops and forests.

In this context, crop classification maps are essential tools in contemporary agriculture and land use monitoring, offering valuable insights into the distribution and extent of various crops and land use types. These maps play a critical role in agricultural planning, yield estimation, and natural resource management (Bégué et al., 2018). Indeed, solid knowledge on crop type distribution at a regional scale represents an important tool for decision makers, for example in the water management policies: constantly updated crop maps would allow authorities to determine optimal water allocation for sustainable irrigation. Furthermore, accurate crop type maps accounting the growth phases of vegetation in a near real time monitoring perspective are essential under the increasing thread of drought. Crop classification maps can also help in identifying areas of potential deforestation, urbanization, and other land use changes that may have significant environmental impacts.

For all these reasons, Earth Observation data available through Google Earth Engine has been recently often used for crop classification applications (Mutanga and Kumar, 2019). For example, in 2017, Andrii Shelestov et al. implemented the agricultural monitoring task in the Kyiv region of northern Ukraine on the Google Earth Engine. They used a random forest classifier to classify the crops in the research area (Shelestov et al., 2017a, 2017b). Indeed, with the completion of Google Earth Engine, especially the opening of the python API interface, the remote sensing dataset can be directly imported into some deep learning models written in python on the cloud platform of

[*] corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1515-2023 |

Google Earth Engine. In 2021, Ritika et al. used the python API of Google Earth Engine to draw a large-scale land cover land use (LCLU) map of Florida, USA, with an overall accuracy of 86% and Kappa value of 79% (Prasai et al., 2021). In this study, the deep learning (DP) models developed in recent years are applied to Senintel-1/2 data to draw large-scale crop maps through the Python API of Google Earth Engine: Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Transformer.

1.1 Deep learning models

Among the several available deep supervised learning, i.e. those DL algorithm dealing with labelled data, the most popular are convolutional neural networks (CNNs), deep neural networks (DNNs) and recurrent neural networks (RNNs) (Alzubaidi et al., 2021).

A DNN is a type of artificial neural network consisting of multiple (deeper number of) processing hidden layers of interconnected nodes (neurons) between input and output, that allows solving non-linear problems. DNNs are commonly used for various tasks, including image and speech recognition, natural language processing, and recommendation systems (Goodfellow et al., 2016).

CNNs are the most popular DL algorithm, designed for processing structured grid data, such as images or time series data. They utilize convolutional layers that apply filters to the input data, enabling the network to automatically learn spatial hierarchies of features. However, they are step-by-step algorithms that can only extract data features within the step size each time rather than the entire input feature. CNNs have been widely successful in image analysis tasks, including classification, object detection, and image segmentation (Lecun et al., 1998).

RNNs are widely employed as DP model, especially for speech processing tasks. They derive valuable information from the sequence of data, implementing sequential data in the networks. However, some limits of these networks are well known, such as their sensitivity to the exploding gradient and vanishing problems. In this context, LSTM and Transformer have been developed to overcome some of these problems.

In particular, LSTMs are a specific type of RNN, presented for the first time by Hochreiter and Schmidhuber to overcome the gradient disappearance problem in machine learning and capture long-term dependencies in sequential data (Hochreiter and Schmidhuber, 1997). LSTMs introduce a memory cell with multiple interacting gates that regulate the flow of information, allowing them to retain information over longer time steps (Hochreiter and Schmidhuber, 1997). Transformers are a type of neural network architecture that utilizes self-attention mechanisms to capture relationships between different elements of an input sequence. Unlike traditional sequential models, such as RNNs, Transformers can process the entire input sequence in parallel, making them highly parallelizable and efficient. Transformers have achieved remarkable success in various natural language processing tasks, such as machine translation, text generation, and question-answering systems (Vaswani et al., 2017).

Considering the state of the art of machine learning and DL applied to crop classification from satellite imagery, among the most widely used algorithms, the current mainstream crop classification model is the random forest. As demonstrated by several researches, the accuracy of the random forest classification model for crop classification is between 80-85% (Long et al., 2013; Ok et al., 2012; Tatsumi et al., 2015). Breaking through the bottleneck of 85% accuracy has now become the focus of crop classification model research in recent years. Lately, with the continuous development of DL algorithms, some of the presented models proved to be successful for such applications. Zhou et al. (2019) and Filho et al. (2020) respectively established the crop classification model of the Long-Short Term Memory (LSTM) algorithm based on SAR data. They found that when compared with the traditional method, the classification accuracy of the LSTM model has improved by 5% (Crisóstomo de Castro Filho et al., 2020; Sun et al., 2020). LSTM networks are well-suited for classification, processing, and forecasting based on time-series data, where there may be lags of unknown duration between important events in the time series. Based on this characteristic, when dealing with time series, especially when considering extended time intervals, the LSTM networks can often result in better performance compared with the standard RNN.

Since 2017, when Vaswani et al. proposed the Transformer, solely based on attention mechanisms, the advantage of this new model of reducing training time through parallel processing, seems promising for the crop classification task. Therefore, due to the excellent performance of the Transformer in processing sequential data, this study explores the application of the attention mechanism model in crop classification from satellite imagery. And a variant of the Transformer model for crop classification is established based on the attention mechanism. Since the variant is mainly a modification of the Transformer decoder part and the multi-head attention structure, the model is still called Transformer below. Moreover, since the crop classification problem considered in this study represents a typical multivariate time series classification problem, the LSTM was selected as a control model. The results obtained implementing the DNN and CNN are provided and compared.

All the models were tested on three different combinations of the input datasets: Sentinel-2, Sentinel-1 and their fusion.

2. MATHERIALS AND METHODS

2.1 Multispectral data

Sentinel-2 mission of the European Copernicus programme provides wide-swath, medium resolution, optical imagery of the Earth surface since 2015 with high revisit time (5 days on average) (ESA, 2015). The Multi Spectral Instrument (MSI) on board the mission's twin satellites, Sentinel-2A/B, acquires images within 13 spectral bands, with geometric resolution ranging between 10 m to 60 m depending on the band wavelength. The 13 bands can be divided into three groups: the visible and near-infrared (VNIR) bands, the red-edge (RE) bands, the short-wave infrared (SWIR) bands. The VNIR bands have a spatial resolution of 10 meters and cover the wavelength range from 443 to 865 nm. The RE bands have a spatial resolution of 20 meters and cover the wavelength range from 705 to 745 nm. The SWIR bands have a spatial resolution of 20 or 60 meters and cover the wavelength range from 1190 to 2190 nm. For their spectral, geometric and temporal resolution, the Sentinel-2 data are widely used in various applications, such as land cover and land use mapping, vegetation monitoring, water quality assessment. Moreover, the free access policy of the dataset as well as its availability through the GEE data catalogue has spread its use in a variety of application. In the present study, Sentinel-2 Level-2A products, which are obtained from level-1C products by applying the Sen2Cor atmospheric correction algorithm, were used (Louis, 2016; Louis et al., 2016).

2.2 SAR data

The Sentinel-1 satellites of the Copernicus program are equipped with a C-band synthetic aperture radar (SAR) that can collect images day and night, regardless of weather conditions. The SAR has a ground range detected spatial resolution of 10 meters, as it is available through the Google Earth Engine. The satellites orbit the Earth in a near-polar orbit, with a 12-day repeat cycle and completing 175 orbits per cycle. The SAR instrument on Sentinel-1 operates in three different modes: Interferometric Wide Swath (IW), Extra Wide Swath (EW), and Stripmap (SM), each with different spatial resolutions and swaths. The SAR data used in this study were obtained from the IW mode, which provides a 250 km swath with a spatial resolution of 10 meters. SAR data are widely used in crop classification due to their ability to penetrate cloud cover, which represents a great benefit with respect to optical data, and to capture information on soil moisture and vegetation structure.

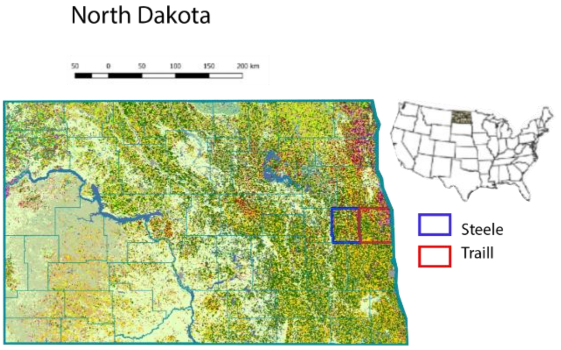
2.3 Reference data: cropland data layer

The Cropland Data Layer (CDL) is a dataset that provides information on crop cover across the continental United States. It is a collaborative effort between several organizations, including the United States Department of Agriculture, the National Agricultural Statistics Service, the Department of Research and Development, the Geospatial Information Division, and the Spatial Analysis Research Division.

The CDL dataset is created using satellite imagery and ground truth data from agricultural sources. It contains information on 133 different land cover types, including various types of crops, forests, and grasslands. The resolution of the CDL dataset is 30 meters, which means that each pixel in the dataset represents an area of 30 square meters on the ground. This high resolution allows for detailed analysis of crop cover patterns across the United States (USDA-NASS, 2023). Cropland Data Layer has an average crop accuracy of 92% (Lark et al., 2021)

2.4 Study Area

Since the CDL crop labelled ground truth dataset was available over the United States, the methodology was tested on two selected USA Counties: the Steele and Traill Counties in North Dakota (Figure 1). The two study areas cover a surface of 2227.6 km^2 and 1874.8 km^2, respectively. Over 91.4% of the surface is farmland. There are 36 different types of land cover, but the main crops are soybeans 36.5%, corn 23.5%, spring wheat 11.8%, dry beans 7.7%, grassland/pasture 4.7%, sugar beets 2.8%, and Barley 1.7%.



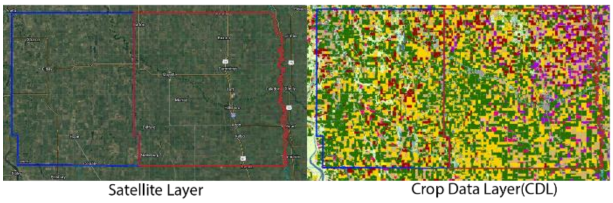


Figure 1. The study area Steele and Traill with crop cover

2.5 Methodology

The methodology adopted in this study can be summarized in the following sequential steps:

- Retrieve all Sentinel-1 and Sentinel-2 data for the 2021 study area on Google Earth Engine (GEE).
- Preprocess the data, by performing cloud/snow/shadow masking followed by fine co-registration.
- Sample the training, validating, and testing pixels. SAR and spectral data corresponding to the pixels are extracted in the data layer in GEE.
- Generate time series features for each pixel, and the time series take each month as the node to take the average value of the SAR and multispectral data of the month.
- Transform the multivariate time series into two-dimensional arrays.
- Build 4 deep learning classification models based on dense neural network (DNN), long short-term memory neural network (LSTM), convolutional neural network (CNN), and Transformer.
- Predict the test dataset and extract the confusion matrix.
- Based on the results, evaluate the performance of four classification models for different crops.

The data preparation procedures for extracting SAR data layers (VV, VH bands of Sentinel-1), extracting optical data layers (12 spectral bands of Sentinel-2), merging data layers, and generating sample points on the combined data layer are completed through the Google Earth Engine platform. The first extraction of sampled pixels allowed to create the feature matrix. These steps where conducted in the Python environment, and the feature matrices are stored in the form of DataFrame, a data structure containing labelled axes (rows and columns). The row label represents the time node. The column label describes the bands of SAR and optical sensors. Then we will separate the DataFrame data matrix into a NumPy two-dimensional matrix and a corresponding label representing the land cover type.

2.5.1 Feature Matrix

As shown in the schema of Figure 2, more than 140,000 random points were generated to extract monthly SAR data and monthly multispectral Time series at sampled pixels location. All the extracted data, from both the S1 and S2 image collections, were normalized, ensuring that all values are scaled between zero and one. This normalization resulted in the formation of a feature matrix as shown in the diagram of Figure 2. The yellow-boxed VV and VH data represent the SAR data collected by Sentinel-1 satellite, while the red-boxed bands represent the spectral range captured by Sentinel-2 satellite, ranging from 443.9 nm to 2202.4 nm. We merged the monthly SAR and multispectral data of extracted from the sample pixels to create a multivariate time series dataset. This dataset was further transformed into a feature matrix, where the vertical axis represents time, and the horizontal axis represents the backscattering values of SAR and the spectral reflectance among the 13 S2 bands.

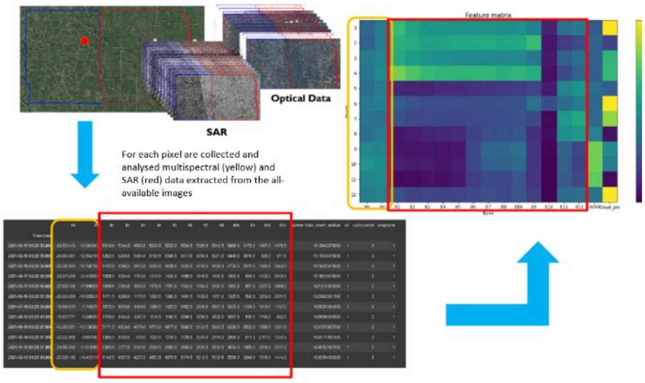

Optical Data
SAR
For each pixel are collected and analysed multispectral (yellow) and SAR (red) data extracted from the all-available images
Feature matrix

Figure 2. Generation of the feature matrix

This feature matrix has 12 rows and 17 columns. The 12 rows correspond to the 12 months in a year, representing the SAR data and spectral data recorded for the sample point during each month. Among the 17 columns, two columns represent the VV and VH bands from Sentinel-1 satellite, while the other 13 columns represent the 13 spectral bands from Sentinel-2 satellite. Additionally, two columns contain the NDVI values (Normalized Difference Vegetation Index) calculated from the spectral data, and the percentage of monthly average cloud cover provided by Google Earth Engine. In summary, the feature matrix consists of monthly SAR data, spectral data, NDVI values, and cloud cover information, providing a comprehensive representation of the characteristics of the sample point over a year.

2.5.2 Transformer

As shown in the diagram of Figure 3, the Transformer is a model that follows an encoder-decoder architecture. The encoder maps the input sequence represented by symbols (x1, ..., xn) to a continuous representation sequence z = (z1, ..., zn). Given z, the decoder generates symbol outputs (y1, ..., ym) one element at a time. At each step, the model is autoregressive, using previously generated symbols as additional input when generating the next one. The Transformer adheres to this overall architecture, employing stacked self-attention and point-wise, fully connected layers for both the encoder and decoder, as depicted in the left half of the diagram in Figure 3.

In the context of this research, the input consists of a multivariate time series with 12-time steps and 17 variables. The Transformer model constructed in this study comprises 6 encoders and 6 decoders. Each encoder consists of a stack of 6 identical layers. Each layer contains two sub-layers: a multi-head self-attention mechanism and a simple position-wise fully connected feed-forward network. We apply residual connections around each of these sub-layers, followed by layer normalization. In other words, the output of each sub-layer is computed as LayerNorm(x + Sublayer(x)), where Sublayer(x) represents the function implemented by the sub-layer itself. To facilitate these residual connections, all sub-layers in the model, as well as the embedding layer, generate outputs of dimensionality model = 17. Here the 17 dimensions correspond to the 17 variables in the feature matrix. The output of this model is a vector of length 34, where each element represents the probability of a specific crop.

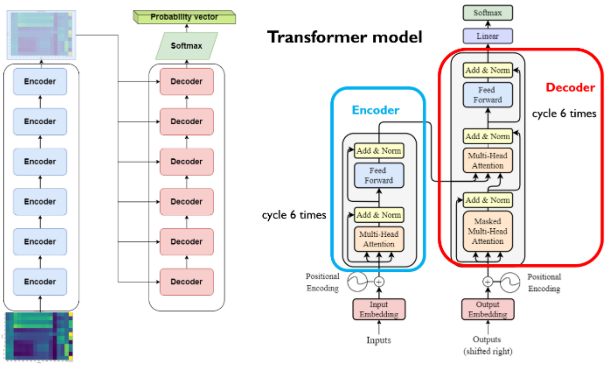


Figure 3. The Transformer Model

The structure of Self-Attention, as illustrated in the schema of Figure 4, involves the use of matrices Q (query), K (key), and V (value) for computations. In practice, Self-Attention receives input in the form of a feature matrix (12x17) representing a specific sample point or the output of the previous Encoder block. Q, K, and V are derived through linear transformations of the input during the Self-Attention process. Let X represent the input matrix for Self-Attention, and Q, K, and V can be computed using linear transformation matrices WQ, WK, and WV. The calculation, as depicted in the diagram, involves representing each row of X, Q, K, and V as the data for each time step. Once matrices Q, K, and V are obtained, the output of Self-Attention can be computed using the formula presented in the diagram.

In the formula, the inner product of each row vector in matrices Q and K is calculated. To prevent excessively large inner products, they are divided by the square root of d. Multiplying Q by the transpose of K results in a matrix of size 17x17, where 17 represents the number of bands for each time step. This matrix captures the attention strengths between different bands. After obtaining QK^T, the softmax function is applied to compute attention coefficients for each word with respect to other words. Upon obtaining the softmax matrix, it can be multiplied by V to yield the final output matrix Z.

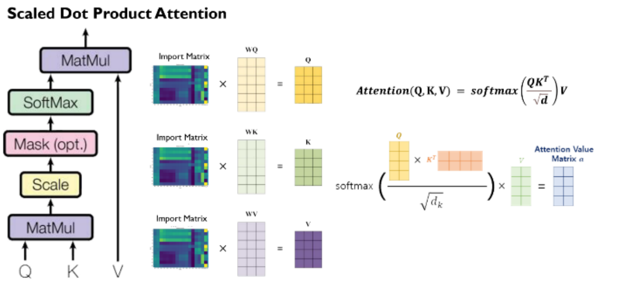


Figure 4. Scaled Dot Product Attention

The Multi-Head Attention structure, as illustrated in the diagram below (Figure 5), is formed by combining multiple Self-Attention mechanisms. From the diagram, it can be observed that Multi-Head Attention consists of multiple Self-Attention layers. The input, X, is passed through h different Self-Attention mechanisms, resulting in h output matrices, Z. In the diagram, the case of h=8 is shown, where 8 output matrices, Z, are obtained. After obtaining the 8 output matrices, Multi-Head Attention concatenates them together and passes them through a Linear layer, yielding the final output matrix, Z. It is evident from the diagram that the dimensions of the output matrix, Z, from Multi-Head Attention are the same as the input matrix, X.

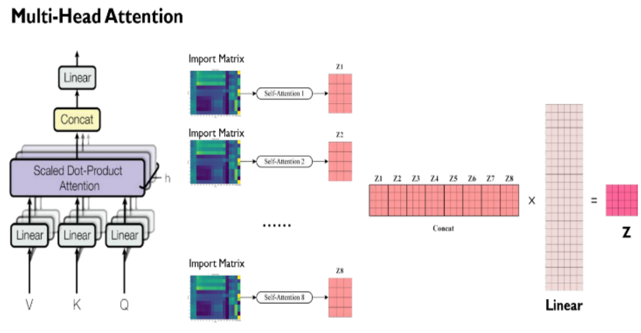


Figure 5. Multi-Head Attention

3. RESULTS AND DISCUSSION

The control group of this study involved the use of solely SAR data from Sentinel-1 and spectral data from Sentinel-2, which were then utilized with four different deep learning models, namely DNN, LSTM, CNN, and Transformer.

Data	DNN	LSTM	CNN	Transformer
S1&S2	0.8443	0.8315	0.8426	0.8504
S2	0.8426	0.8400	0.8419	0.8421
S1	0.7253	0.7119	0.7258	0.5306

Table 1. Accuracy table of 4 deep learning models

Table 1 depicts the results of the experiment where three different datasets were used to train and test the four deep learning models. The first row of the table represents the fusion data of Sentinel-1 and Sentinel-2, the second row shows the data of only Sentinel-2, and the third row demonstrates the data of only Sentinel-1. The four columns on the right side of the table correspond to the four deep learning models (DNN, LSTM, CNN, Transformer). The analysis of the results shows that when the fusion data of Sentinel-1 and Sentinel-2 were used, the Transformer model achieved the best overall accuracy rate of 85.04% on the test set. The DNN model achieved the second-best result with an accuracy rate of 84.43%, followed by the CNN model with 84.26%, and finally the LSTM model with 83.15%.

Moreover, when only the spectral data of Sentinel-2 was utilized, the overall accuracy rate was not significantly different from the fusion data of Sentinel-1 and Sentinel-2. However, there was a slight increase of 0.75% in the accuracy of the LSTM model, whereas the accuracy of the other three models decreased by less than 1%. Lastly, when only the SAR data of Sentinel 1 was used, the accuracy rates of DNN, LSTM, and CNN models decreased by approximately 12%, whereas the Transformer model had the most significant decline, dropping to an accuracy rate of 53.06%. These results demonstrate that the fusion of SAR and spectral data can enhance the overall accuracy of the deep learning models, and the Transformer model performs better than the other models in this context.

To ease the interpretation, the results are presented focusing on seven land cover types, which are the most present in the study area, for the confusion matrix (Figure 6): Soybeans, Cron, Spring wheat, Dry beans, Grassland, Sugar beets, and Barley. Although our training set comprised 140,000 sample points, it is important to note that these were randomly selected from the study area. Out of the 33 different land covers, only the number of sample points for the aforementioned seven crops exceeded 2,000 individuals.

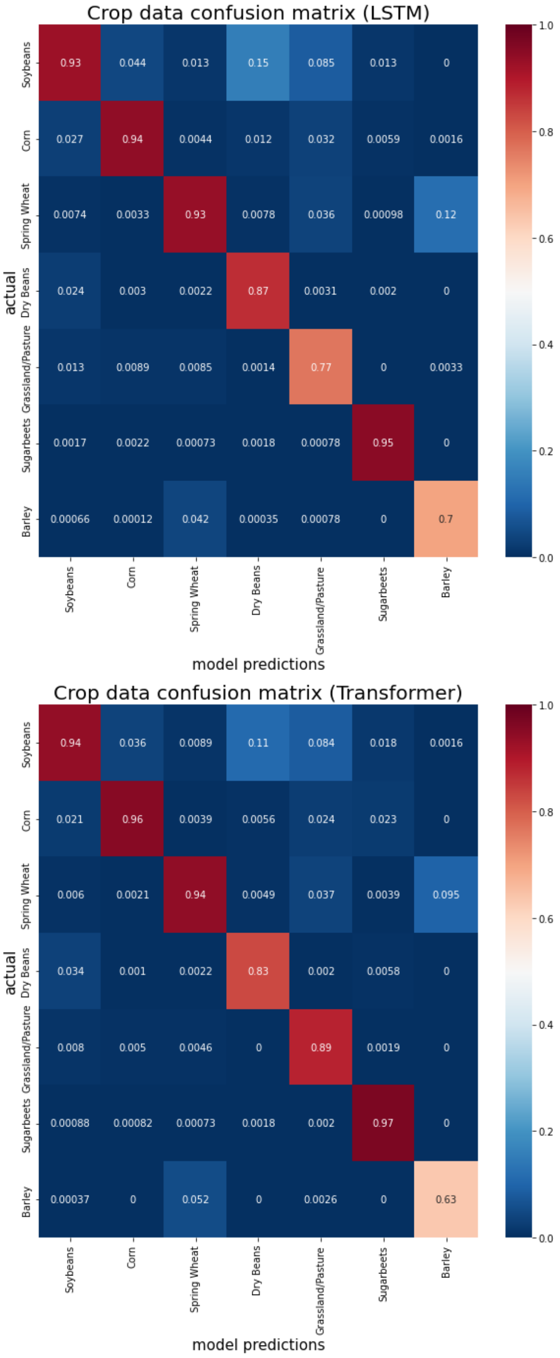


Figure 6. Confusion Matrix of LSTM (top) and Transformer Model (bottom) with seven crops types

It is evident that the diagonal of the confusion matrix exhibits accuracy rates above 90% for all LSTM and Transformer models employed in crop classification. Conversely, the accuracy rates for the classification of Dry Beans, Grassland, and Barley are relatively lower. Specifically, the accuracy rates for Dry Beans and Grassland across all four models hover around 80%; however, the classification accuracy of Barley is notably low, with the DNN model exhibiting accuracy as low as 55%.

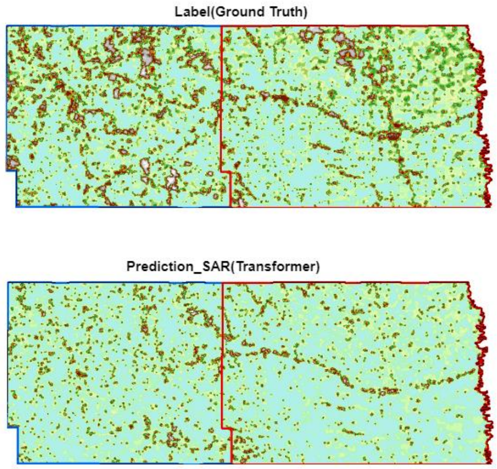


Figure 7. Crop Classification Maps based on SAR data by Transformer Models

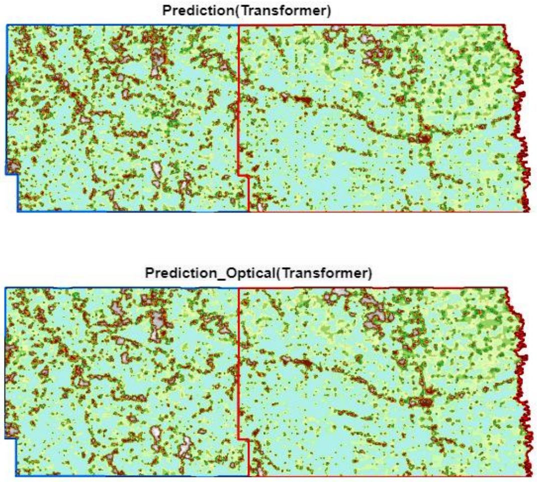


Figure 8. Crop Classification Maps based on optical data by Transformer Models

Based on the analysis of the four crop classification maps, we can see that the use of fusion data from Sentinel-1 and Sentinel-2 yields the most accurate results when input into the Transformer model. This suggests that the combination of SAR and optical data can improve the accuracy of crop classification compared to using only one type of data.

Moreover, it is interesting to note that using only Sentinel-2 optical data (Figure 8) in the lower right map still yields relatively accurate results, albeit slightly less accurate than the fusion data map. This highlights the importance of optical data in crop classification, as it captures important surface features such as vegetation and land use.

On the other hand, using only Sentinel-1 SAR data in the lower left map results in the least accurate classification map (Figure 7). This suggests that SAR data alone may not be sufficient for accurate crop classification, as it may not capture certain surface features that are important for classification.

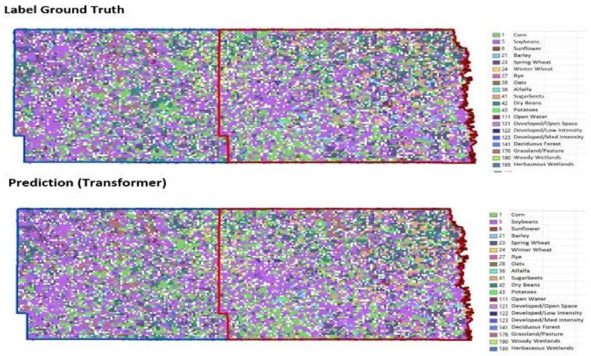


Figure 9. Crop classification map resulting from Transformer

Based on the comparison between the crop classification map generated by the Transformer and the actual crop map, it can be observed that they exhibit a high degree of similarity (Figure 9). The results indicate that the Transformer model achieves an overall accuracy of 85%, while major crops such as corn, sugar beets, and spring wheat achieve accuracy rates surpassing 90%, consistent with previous findings.

4. CONCLUSION

Based on the results of this study, it can be concluded that the use of deep learning models, particularly the Transformer model, in crop classification can be an effective approach, that should be further developed. The use of fusion data features matrices from multiple sources such as Sentinel-1 and Sentinel-2 can improve the accuracy of crop classification maps. However, the performance of the deep learning models is affected by various factors such as the size and distribution of the research area, the type of crops, and the quality of the input data. Therefore, it is necessary to carefully evaluate and compare the performance of different models on specific datasets before making a final choice. Further research can be conducted to improve the performance of the deep learning models in crop classification. This may involve the use of additional data sources, such as climate and soil data, to provide a more comprehensive understanding of crop growth patterns. Additionally, incorporating other advanced techniques such as transfer learning and ensemble learning can also be explored to enhance the accuracy and robustness of the models.

The four deep learning models tested in the study, DNN, LSTM, CNN, and Transformer, have all shown promising results in crop classification using remote sensing data. Each model has its own strengths and weaknesses, and the choice of model may depend on the specific needs and characteristics of the dataset.

The results of this study showed that the Transformer model outperformed the other three models, achieving an overall accuracy of 86% in the study area of the United States, with higher values if considering specific crops.
The research also found that the performance of the models varied for different crop types. Specifically, the Transformer model performed well for corn, soybeans, and wheat, while the DNN and LSTM models were better for alfalfa and cotton. The CNN model showed relatively poor performance for most crop types.

Overall, these findings confirm that the use of deep learning models for crop classification can improve the accuracy and efficiency of crop mapping, especially when using a fusion of Earth Observation data. However, the choice of model should be based on the specific characteristics of the study area and crop types.

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, Journal of Big Data. Springer International Publishing. https://doi.org/10.1186/s40537-021-00444-8

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., Ferraz, R.P.D., Lebourgeois, V., Lelong, C., Simões, M., Verón, S.R., 2018. Remote sensing and cropping practices: A review. Remote Sens. 10, 1–32. https://doi.org/10.3390/rs10010099

Casu, F., Manunta, M., Agram, P.S., Crippen, R.E., 2017. Big Remotely Sensed Data: tools, applications and experiences. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.09.013

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

ESA, 2015. SENTINEL-2 User Handbook 64. https://doi.org/GMES-S1OP-EOPG-TN-13-0001

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Lark, T.J., Schelly, I.H., Gibbs, H.K., 2021. Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the United States Using the USDA Cropland Data Layer. Remote Sens. 13, 968. https://doi.org/10.3390/rs13050968

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791

Long, J.A., Lawrence, R.L., Greenwood, M.C., Marshall, L., Miller, P.R., 2013. Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest. GIScience Remote Sens. 50, 418–436. https://doi.org/10.1080/15481603.2013.817150

Louis, J., 2016. Sentinel 2 MSI - Level 2A Product Definition, European Space Agency, (Special Publication) ESA SP.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., 2016. Sentinel-2 SEN2COR: L2A processor for users, in: Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016. Prague, pp. 9–13.

Mutanga, O., Kumar, L., 2019. Google Earth Engine Applications. Remote Sens. 11, 591. https://doi.org/10.3390/rs11050591

Ok, A.O., Akar, O., Gungor, O., 2012. Evaluation of random forest method for agricultural crop classification. Eur. J. Remote Sens. 45, 421–432. https://doi.org/10.5721/EuJRS20124535

Prasai, R., Schwertner, T.W., Mainali, K., Mathewson, H., Kafley, H., Thapa, S., Adhikari, D., Medley, P., Drake, J., 2021. Application of Google earth engine python API and NAIP imagery for land use and land cover classification: A case study in Florida, USA. Ecol. Inform. 66, 101474. https://doi.org/10.1016/j.ecoinf.2021.101474

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017a. Large scale crop classification using Google earth engine platform, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 3696–3699. https://doi.org/10.1109/IGARSS.2017.8127801

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017b. Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. Front. Earth Sci. 5. https://doi.org/10.3389/feart.2017.00017

Sun, Z., Di, L., Fang, H., Burgess, A., 2020. Deep Learning Classification for Crop Types in North Dakota. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 2200–2213. https://doi.org/10.1109/JSTARS.2020.2990104

Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Taipe, C.L.R., 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. Comput. Electron. Agric. 115, 171–179. https://doi.org/10.1016/j.compag.2015.05.001

USDA-NASS, 2023. USDA National Agricultural Statistics Service Cropland Data Layer [WWW Document]. URL https://data.nal.usda.gov/dataset/cropscape-cropland-data-layer

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. https://doi.org/10.48550/arXiv.1706.03762

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: State-of-the-Art Natural Language Processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6