

Bidirectional Long Short-Term Memory Networks for Automatic Crop Classification at Regional Scale using Tabular Remote Sensing Time Series

Harijs IJABS¹, Ēvalds URTĀNS²

¹University of Latvia, Faculty of Geography and Earth Sciences, Riga, Latvia

²Riga Technical University, Department of Artificial Intelligence and Systems Engineering, Riga, Latvia

harijs.ijabs@gmail.com, Evalds.Urtans@rtu.lv

Abstract. With the arrival of European Union's new Common Agricultural Policy (CAP 2020), a paradigm shift in subsidy control is underway. Member states are required to gradually transition from a system of on-the-spot checks, where the presence or absence of a crop is detected manually on the field, to a system of agricultural monitoring based on remote sensing data; primarily – Sentinel-1 and Sentinel-2. This paper presents a classification of regional crop types based on the Bidirectional Long-Short-Term Memory (BiLSTM) network. The approach is based on tabular time series of Sentinel-1 and Sentinel-2 sensor data over the entire territory of Latvia. Two types of LSTM architectures are evaluated in this paper – regular and bidirectional. An exhaustive grid search of network hyperparameters with 15 distinct crop types led to the conclusion that the bidirectional variant of LSTM yields the highest overall weighted test accuracy of 89.1%.

Keywords: remote sensing, crop type, neural networks, LSTM, classification, satellite data, sentinel data, Sentinel-1, Sentinel-2

List of abbreviations

B1, B2, B3 ...	Optical satellite spectral bands
BI	Brightness Index
BiLSTM	Bidirectional Long Short-Term Memory
FOI	Field of interest
LSTM	Long Short-Term Memory
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
RNN	Recurrent Neural Network
RF	Random Forest
RSS	Rural Support Service of Latvia
S1	Sentinel-1
S2	Sentinel-2
VH	Cross-polarisation of radar sensor
VV	Co-polarisation of radar sensor

1. Introduction

As projected by the UN (United Nations, 2019), the Earth's population will reach ten billion inhabitants by 2050. Along with the rising population also rises the need for effective management of the agricultural sector. Starting from 2020, as part of the new Common Agricultural Policy (ECA, 2021) EU member states are required to shift from control by on-the-spot checks, where inspectors physically visit fields of interest, to checks by monitoring, where majority of subsidy related decisions are based on analysis of remote sensing data, mainly Sentinel-1 and Sentinel-2 satellites of the Copernicus program. This approach allows minimizing the administrative burden for both the Paying Agency (PA) and farmers alike, as well as diminishes the environmental impact of field visits. By providing the PA with this information, crop type classification yields a novel approach to subsidy application control where one can remotely discern whether a farmer has indicated the appropriate crop type for his field, instead of, for example, a different type with a higher Euro per hectare subsidy rate.

In principle, classification of agricultural crops by remote sensing imagery is a process by which a multi-layered dataset of geospatial imagery is transformed into a thematic map or a geo-referenced dataset indicating the location and distribution of certain crop types. (Dhumal et al., 2013) For the classification to be sufficiently representative of the actual situation, the input data must adequately describe the phenological and spectral characteristics of the crop types in question. For each land cover the amount of reflected and emitted energy is different, this is defined as a spectral signature – it is a reflectance characteristic that describes land cover or crop type differences in various bandwidths of infrared and visible light. (Chen, et al., 2016)

Although LSTM and BiLSTM models have been used to classify crop types and characteristics from Sentinel data in previous studies and have already shown increased prediction accuracy over classical machine learning and statistical methods (Filho, et al., 2020; Portalés-Julià et al., 2021; Paris et al., 2020) the prospect of large-scale multi-sensor tabular time series classification is yet to be fully considered (Pluto-Kossakowska, 2021). The possible advantages of tabular data classification over pixel-based classification have been described during the past decade. These potential advantages include higher prediction accuracies (Trang et al., 2016), especially when working with multiple resolution imagery (Weih and Riggan, 2010), as well as improved data portability and reduced model training time, which contributes to a more efficient and exhaustive model optimization routine. Additionally, most of the current research exploring the performance of LSTM models in crop classification focus on a relatively small research area of a several thousand square kilometres (Rußwurm and Körner, 2018; Metzger et al., 2021) or explore comparatively fewer crop types (Zhao et al., 2021). This paper proposes an approach to utilize LSTM models on a regional scale with a wide variety of crop types, which introduces additional complexity to the problem as location differences (continental or maritime climate, soil properties, local weather patterns and agricultural practices) impact crop phenology (Gao and Zhang, 2021) and the number of considered crop types increases the difficulty of the classification task (Pluto-Kossakowska, 2021).

2. Methodology

Initial data collection and pre-processing was handled by using Sen4CAP – an Earth Observation processing system. This system is a standalone processing chain which generates remote sensing products for agricultural monitoring applications, featuring pre-processing and data acquisition workflows for Sentinel-1, Sentinel-2 and Landsat-8 data. (Sen4CAP, 2022) All pre-processing steps related to Sentinel-1 backscatter and coherence metrics, as well as Sentinel-2 atmospheric correction were orchestrated using this system. Atmospheric correction of Sentinel-2 imagery was handled by the software CNES MAJA (Hagolle et al., 2016) version 3.2.2 TM, and Sentinel-1 pre-processing was handled by the SNAP toolbox (ESA, 2022).

2.1. Domain of the study

The domain of this research comprises the entire territory of Latvia - 64 589 km² (Figure 1). The decision to extend this research over the entire country was made because the underlying purpose of this study is to develop a model that would work as a tool for agricultural subsidy control at the national level. Thus, developing a model robust enough to handle local differences in climate and soil conditions was a crucial aspect when planning this research.

2.2. Earth Observation data

The entire spatiotemporal data stack is comprised of 470 Sentinel-1 and 1102 Sentinel-2 acquisitions. The study area is crossed by two Sentinel-1 orbits (number 131 and 160, both descending and ascending) and 18 Sentinel-2 MGRS tiles (Figure 1). The dataset spans a timeframe from 15.04.2021 until 21.09.2021. Sentinel-2 data with cloud coverage over 90% were automatically discarded by the MAJA pre-processing algorithm.

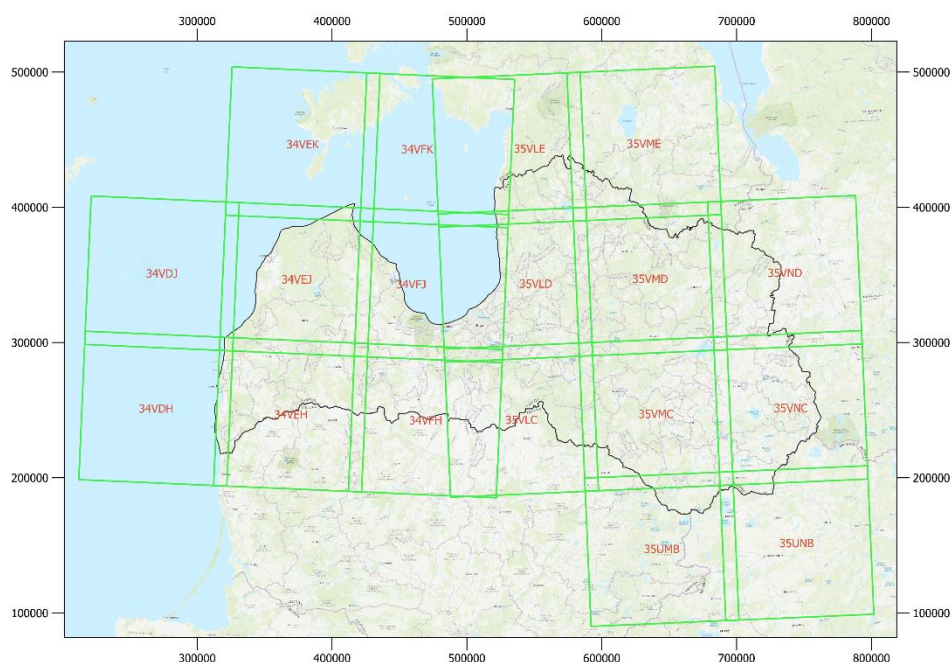


Figure 1. Study area and Sentinel-2 tile coverage

After the images were downloaded and pre-processed, a parcel extraction routine was performed by the Sen4CAP system wherein a zonal statistics approach was used to compute the mean metrics for each agricultural parcel at each of the acquisition dates. The aforementioned process yields a tabular dataset with rows representing agricultural parcels and columns representing the input features at each of the dates.

As per the Sen4cap data processing procedure, a seven day mean compositing procedure was applied to both Sentinel-1 and Sentinel-2 tabular time series in order to mitigate the impact of missing acquisitions and regularize the dataset. The final raw datasets thus span the entire season with a seven-day step for each of the input features.

Table 1. Features used in the dataset

Sensor	Feature
Sentinel-2	B3, B4, B5, B6, B7, B8, B11, B12, NDVI, NDWI, BI
Sentinel-1	<ul style="list-style-type: none"> • VV and VH polarised coherence • VV and VH polarised backscatter • Ratio between VV and VH backscatter

As indicated in Table 1, a total of 16 features were used for each of the composited acquisition dates. Alongside a selection of spectral bands which range from visible green light to short wave infrared wavelengths, the dataset includes mean parcel values of Normalized Difference Vegetation (NDVI), Normalized Difference Water (NDWI) and Brightness (BI) indexes.

An additional subset containing only Sentinel-1 time series was used to determine the performance of the BiLSTM architecture on a dataset with no optical sensor data. The purpose of this addition is to ascertain whether an operational solution based on this architecture could provide satisfactory results in an agricultural season with very little to no Sentinel-2 imagery due to frequent cloud cover, as entire summer months without a single clear Sentinel-2 image is not an uncommon occurrence in Latvia.

Preparation of dataset features

In order to fit the format requirements for our BiLSTM classifier, the initial dataset had to be pivoted to a tabular format where rows represent composited acquisition dates and columns represent the input features – mean values for sensor data. Therefore, each field of interest (FOI) is represented by a certain number of rows corresponding to the composited acquisition dates. This dataset consists of 21 dates; therefore, each FOI is represented by 21 rows of sensor data. For illustration purposes, Table 2 presents a partial sample of a single field. The dataset consists exclusively of scalar data, and was normalized with a feature range (MinMax) scaler.

Table 2. Partial example of a finalized input dataset

Date	B3	B4	B8	B11	NDVI	
4/15/2021	809.312	1018.96	1505.56	2423.06	0.193196	...
4/22/2021	1048.65	1363.7	2008.88	3045.51	0.190432	...
4/30/2021	1017.88	1363.81	2014.52	3116.77	0.193011	...
5/7/2021	1012.955	1352.43	2069.225	3142.86	0.210056	...
5/15/2021	1008.03	1341.05	2123.93	3168.95	0.227102	...
5/22/2021	1029.56	1283.98	2388.99	3192.47	0.303539	...
5/30/2021	1051.09	1226.91	2654.05	3216	0.372233	...
6/6/2021	938.3815	971.738	2963.79	2847.255	0.507477	...
...

2.3. In-situ data

A total of 379 842 agricultural parcels, provided by the Rural Support Service of Latvia (RSS), were used for feature extraction and sample labelling. The RSS dataset contains agricultural parcel coordinates and geometries for the year 2021 along with the crop type declared by the farmer. Parcel geometries were used as input arguments for the Sen4CAP system for sensor data time series extraction and the declared crop information was used to label each sequence with its respective crop type. Table 3 lists all the crop type groups featured in this study along with sample counts for each of the groups. Crop group selection was based on the predominant crop types present in the Latvian agricultural landscape.

Crop class	Support
Grasslands	211592
Winter wheat	49793
Oats	21587
Summer wheat	20709
Fallow	15854
Summer barley	12375
Winter rapeseed	11655
Orchards	8236
Legumes	6250
Winter Rye	5946
Potatoes	5944
Buckwheat	4194
Maize	2172
Summer rapeseed	2009
Winter barley	1526

Table 3. Crop type groups and sample counts

2.4. The BiLSTM model

The assumption for the classification task of this study is that each of the different types of crops has a distinct phenological cycle that is reflected in the sensor data time series. Therefore, this problem requires a robust model that can take into account the various temporal states (configurations of the 16 aforementioned features) of each sample, “remember” the input data at previous steps, and use all of this information to come to a conclusion as to which crop group this sample belongs to. A variant of the Recurrent Neural Network (RNN), called Bidirectional Long Short-Term Memory (BiLSTM), was chosen as the best fit.

While RNNs, being a temporal variant of feed-forward Artificial Neural Networks (ANN), are capable of retaining feature information in individual time steps, their ability to do so decreases as the length of the time series grows – this is known as the vanishing gradient problem. To overcome this problem, a variant of the RNN called Long Short-Term Memory Network (LSTM) is used. LSTM incorporates an additional dimension that describes the relationships between the output data sequences and the longer input data sequences. Moreover, the bi-directional variant of the LSTM model (BiLSTM) feeds the data not only forward (from start to end of time series), but also in the opposite direction (from end to start of time series), therefore resulting in generally higher classification and prediction accuracies (Siarni-Namini et al., 2019; Baldi et al., 1999). As part of this research, a comparison between LSTM and BiLSTM model performance was conducted in order to determine whether the later part of the time series yields additional information in crop phenology sequence classification, when using maximum temporal pooling.

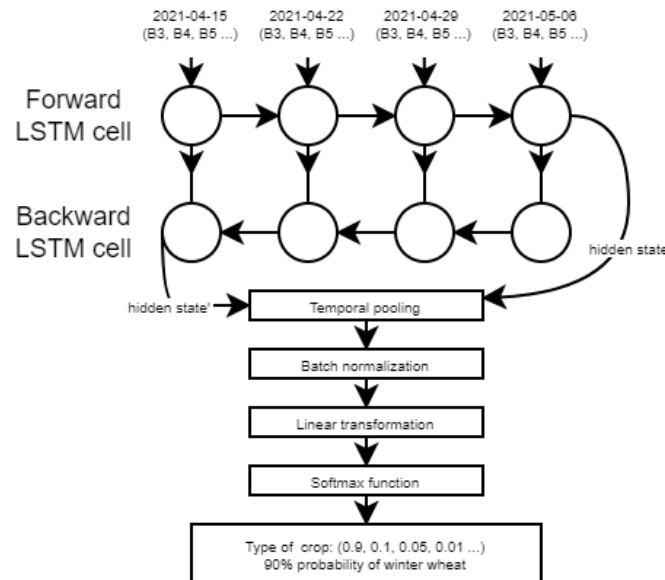


Figure 2. Schematic of the BiLSTM model applied in this study

Model training

The dataset was split into train and test subsets with a split of 80 and 20 percent respectively. The training dataset was used for model training and the validation dataset was used for evaluation of model performance after each training epoch. Model hyperparameter tuning was performed through a grid search method, where a complete training and evaluation sequence was completed for each of the hyperparameter combinations (Table 4), yielding 48 training runs.

Table 4. Model grid search hyperparameters

Hyperparameter	Values
Learning rate	1e-3, 1e-4, 1e-5
Batch size	32, 64, 128, 256
Hidden size	32, 64, 128, 256
Model type	LSTM, BiLSTM

The best performing hyperparameters for the BiLSTM model on the full dataset were a learning rate of 0.001, batch size of 32 and hidden size of 64. For LSTM model hidden size of 128 and same batch size and learning rate was chosen. As the classes in the training dataset were not balanced, a class ratio-based weighting loss function was implemented by adding class weights to the cross-entropy loss function of the model. Also weighted random sampling was tested, but class ratio-based loss function yielded better results for more represented classes which are more important in practical applications. Model accuracy was evaluated with metrics commonly used in classification tasks - precision, recall, F1 score and weighted accuracy. F1 score was chosen as the representative performance metric as it incorporates both precision and recall values for each class, thus describing how the classifier performs from the perspective of both user's and producer's accuracy.

3. Results

Table 5 reports the highest performing test F1 scores for each of the classes in the dataset, as well as overall weighted accuracies for each of the classification scenarios. The four scenarios were: BiLSTM model with both Sentinel-1 and Sentinel-2 data, BiLSTM with only Sentinel-1 data, LSTM with Sentinel-1 and Sentinel-2 data, and LSTM with just Sentinel-1 data.

Table 5. F1 score comparison between models and datasets

	BiLSTM S1+S2	LSTM S1+S2	BiLSTM S1	LSTM S1	RF S1+S2
Winter rapeseed	0.976	0.978	0.969	0.957	0.67
Winter Wheat	0.956	0.954	0.904	0.838	0.84
Grasslands	0.954	0.952	0.926	0.900	0.86
Rye	0.865	0.868	0.771	0.634	0.65
Legumes	0.855	0.840	0.804	0.694	0.47
Buckwheat	0.851	0.814	0.684	0.459	0.49
Maize	0.849	0.790	0.765	0.524	0.32
Winter Barley	0.830	0.807	0.728	0.519	0.26
Summer rapeseed	0.799	0.739	0.786	0.707	0.31
Potatoes	0.762	0.721	0.594	0.456	0.31
Summer Wheat	0.757	0.720	0.670	0.454	0.5
Oats	0.751	0.701	0.630	0.516	0.28
Summer Barley	0.699	0.639	0.580	0.351	0.22
Fallow	0.626	0.609	0.519	0.362	0.19
Orchard	0.487	0.412	0.191	0.017	0.20
Weighted OA	89.43%	88.34%	84.50%	78.67%	66%

The BiLSTM classifier with both Sentinel-1 and Sentinel-2 data was the highest performing of the four scenarios, yielding a weighted accuracy of 89.43%, improving on the LSTM model by +1%. The LSTM classifier outperforms BiLSTM in two classes – winter rapeseed and rye – but only by a narrow margin. As expected, the LSTM classifier is less accurate. Highest prediction accuracies were attained for winter rapeseed, winter wheat and grasslands with F1 values above 0.95. Datasets with only S1 features performed comparatively poorly, as was expected. However, the performance in the BiLSTM S1 scenario still provides an overall weighted accuracy of 84.5% which provides some basis for the assumption that this method may be used operationally (albeit in a more limited capacity) even in particularly cloudy summer seasons where a clear Sentinel-2 image of the entire country may be expected once per month. As a benchmark a traditional Random Forest classification approach was tested as well, however, its performance was well below that of the other classifiers, most likely due to the fact that random forest classifiers do not retain knowledge of previous states in a time series. When conducting 10 repeated training runs with a randomized train-test split, no significant change in test accuracy was observed for the BiLSTM model with Sentinel 1 and Sentinel 2 data with confidence above 95% and z-score above 1.99. Also, BiLSTM results are statistically significant when comparing with LSTM and RF results with p-value below 1e-7.

4. Discussion

By analysing nearly four hundred thousand samples of multi-temporal sensor data, as well as exploring the predictive capability of this data, some assumptions can be made by looking at the confusion matrix presented in table 6.

The largest misclassification proportional to the sample size can be observed between various types of summer crops, as well as between classes of high inherent spectral variability – fallow lands, grasslands and orchards. The difficulty to correctly delineate types of summer crops from one another possibly stems from the fact that crops such as summer barley, summer wheat and oats all share similar phenologic cycles and spectral signatures, thus illustrating the issue that at current scale this data does not offer sufficient information to separate some crop types. An investigation into what could be done to improve classification performance of spectrally and phenologically similar crop types would serve as a potential next step to further increase the operational effectiveness of this model.

As for the confusion between fallows, orchards and grasslands, an assumption could be made that these classes have high internal variability both spectrally and temporally, and thus they overlap with one another. An orchard might consist of various tree types, with a different growing cycle for each plot or even segments within a field, introducing an increased cardinality within a single sample. The same can be said for grassland fields, where each parcel can contain different combinations of plant species. Moreover, fallow lands can resemble a grassland for most of the season, as they develop similarly to temporary grasslands for most of the season, and are ploughed whenever the farmer decides, thus sharing an inconsistent temporal profile with grasslands, which are mown on an irregular or even a once-per-season basis. A potential solution for this issue would be further separation of the orchard class into distinct species, as well as separating temporary from permanent grasslands.

It is worth noting that the BiLSTM architecture performs only marginally better (+1.1% of test accuracy) on the full dataset, which seems to imply that the input features might not yield a large amount additional information when analysed backwards (from end to start of the time series). However, a much larger improvement in accuracy from the BiLSTM model is achieved when classifying only Sentinel-1 data (+6%), which might imply that radar backscatter and coherence data contain more information on the backward pass.

The fact that winter crops are classified more accurately is also noteworthy. This might be explained by the fact that these crops have a significantly higher vegetation index at the start of the season, as they have been sown before the winter and therefore exhibit a very distinct temporal profile and do not mix with grasslands, which could be another reason why summer crops perform worse. Some inter-class confusion between winter crops can be observed, especially between winter rye and winter wheat, but not so much winter rapeseed, most likely due to its distinct yellow hue during the flowering season.

The main focus of this work was to explore the potential of combining large scale sensor time series with state-of-the-art deep learning methods in order to develop a methodology for a crop type classification accurate enough to use in operational rural service administration. With an overall weighted accuracy of 89.4%, the underlying goal of this study appears to be at least partially achieved, but not without its set of challenges and considerations. While the BiLSTM model yields the highest accuracy, the

improvement is marginal, and might be introduced by random patterns inherent to this specific dataset. The issue of inter-class confusion also one that should not be understated, as, for example, different types of summer crops are applicable to different payment rates per hectare. Bearing these considerations and limitations in mind, this method still proves to be a solid foundation for performing regional-scale multi-class crop type classification tasks in order to support effective and objective agricultural subsidy administration.

References

- Baldi, P., Brunak, S., Frasconi, P., Pollastri, G. (1999). Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics*, 15(11), 46.
- Chen, Y., Wang, Q., Wang, Y., Duan, S.-B., Miao Zhong, X., Li, Z.-L. (2016). A Spectral Signature Shape-Based Algorithm for. *International Journal of Geo-Information*, 2016(5), 154.
- Dhumal, R. K., Rajendra, Y., Kale, K., Mehrotra, S. (2013). Classification of Crops from remotely sensed Images: A review. *International Journal of Engineering Research and Applications*, 3(3), 758-761.
- ECA (2021). *The new common agricultural policy: 2023-27*. https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/new-cap-2023-27_en
- ESA (2022). *SNAP*. <https://step.esa.int/main/toolboxes/snap/>
- Filho, H. C., Júnior, O. A., Carvalho, O. L., Bem, P. P., de Moura, R. d., de Albuquerque, A. O., Arnaldo Trancoso, R. G. (2020). Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series. *Remote Sensing*, 12(16), 2655.
- Gao, F., Zhang, X. (2021). Mapping Crop Phenology in Near Real-Time Using Satellite Remote Sensing: Challenges and Opportunities. *Journal of Remote Sensing*, 2021, 14.
- Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R. (2016). MAJA Algorithm Theoretical Basis Document. SPIE Remote Sensing.
- Metzger, N., Turkoglu, M. O., D'Aronco, S., Wegner, J. D., Schindler, K. (2021). Crop Classification Under Varying Cloud Cover With Neural Ordinary Differential Equations. *IEEE Transactions on Geoscience and Remote Sensing, Early Access*, 1-12.
- Paris, C., Weikmann, G., Bruzzone, L. (2020). Monitoring of agricultural areas by using Sentinel 2 image time series and deep learning techniques. Proc. SPIE 11533, Image and Signal Processing for Remote Sensing XXVI, 115330K (21 September 2020).
- Pluto-Kossakowska, J. (2021). Review on Multitemporal Classification Methods of Satellite Images for Crop and Arable Land Recognition. *Agriculture*, 11(10), 999.
- Portalés-Julià, E., Campos-Taberner, M., Javier García-Haro, F., Gilabert, M. A. (2021). Assessing the Sentinel-2 Capabilities to Identify Abandoned Crops Using Deep Learning. *Agronomy*, 11(4), 654.
- Rußwurm, M., Körner, M. (2018). Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS Int. J. Geo-Inf.*, 7(4), 129.
- Siami-Namini, S., Tavakoli, N., Siami Namin, A. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data).
- Trang, N. T., Toan, L., Ai, T. T., Vu Giang, N., Viet Hoa, P. (2016). Object-Based vs. Pixel-Based Classification of Mangrove Forest Mapping in Vien An Dong Commune, Ngoc Hien District, Ca Mau Province Using VNREDSat-1 Images. *Advances in Remote Sensing*, 5(4), 284-295.
- United Nations (2019). *World Population Prospects*. https://reliefweb.int/sites/reliefweb.int/files/resources/WPP2019_Highlights.pdf

- Weih, Jr., R. C., Riggan, Jr., N. D. (2010). Object-based Classification vs. Pixelbased classification: Comparative Importance of Multi-Resolution Imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-4/C7.
- Zhao, H., Duan, S., Liu, J., Sun, L., Reymondin, L. (2021). Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information. *Remote Sensing*, 13(14), 2790.

Received June 3, 2022, revised September 16, 2022, accepted September 19, 2022