APPLICATION OF TEMPORAL CONVOLUTIONAL NEURAL NETWORK FOR THE CLASSIFICATION OF CROPS ON SENTINEL-2 TIME SERIES

M. Račič[1]*, K. Oštir[1], D. Peressutti[2], A. Zupanc[2], L. Čehovin Zajc[3]

[1] Faculty of civil and geodetic engineering, University of Ljubljana, Slovenia - (matej.racic, kristof.ostir)@fgg.uni-lj.si
[2] Sinergise d.o.o., Ljubljana, Slovenia - (anze.zupanc, devis.peressutti)@sinergise.com
[3] Faculty of Computer and Information Science, University of Ljubljana, Slovenia - luka.cehovin@fri.uni-lj.si

KEY WORDS: deep learning, multi-temporal classification, sequence data, crop classification, Sentinel-2

ABSTRACT:

The recent development of Earth observation systems - like the Copernicus Sentinels - has provided access to satellite data with high spatial and temporal resolution. This is a key component for the accurate monitoring of state and changes in land use and land cover. In this research, the crops classification was performed by implementing two deep neural networks based on structured data. Despite the wide availability of optical satellite imagery, such as Landsat and Sentinel-2, the limitations of high quality tagged data make the training of machine learning methods very difficult. For this purpose, we have created and labeled a dataset of the crops in Slovenia for the year 2017. With the selected methods we are able to correctly classify 87% of all cultures. Similar studies have already been carried out in the past, but are limited to smaller regions or a smaller number of crop types.

1. INTRODUCTION

In the presented work we focus on the classification of crops, a task that is common with satellite data. This has been previously done with methods of varying complexity, such as traditional supervised classification methods, random forest (Breiman, 2001), support vector machines (Raj, SivaSathya, 2014) and recurrent neural networks (Rußwurm, Körner, 2018). But when dealing with temporal data, traditional approaches cannot take full advantage of such structured data because the order of the data has no effect on the model and thus time is not considered as a separate feature.

Deep learning offers a variety of approaches to resolve such tasks. In our work we investigate two architectures of deep neural networks for the classification of crops. Progress has already been made by several authors in the past, that have used the segmentation of satellite images using recurrent neural networks (Rußwurm, Körner, 2018) which are capable of processing temporal data. With such an approach it is not necessary to pre-process the data; the model e.g. learns to mask the clouds by training and optimizing the weights. However, such approaches are not without shortcomings. They have many parameters and each state depends on the previous one, which increases the learning time, and requires very large amounts of training data.

There have been also advances in architectures (Bai et al., 2018) that are able to deal with temporal information more efficiently. In this case, one of the main problems with deep learning remains - the need for very large amounts of well annotated data. Given the scale of these problems, we have limited ourselves to preparing the data, analysing and implementing selected architectures and comparing the results. The reference data used in this study was for Slovenian crops in the year 2017, shown in Figure 1. The dominant class in the area are meadows followed by maize. Region marked in red is dominated by vineyards and only further away we have meadows. Differently the region in black has small fields with various crop types clustered very close together.

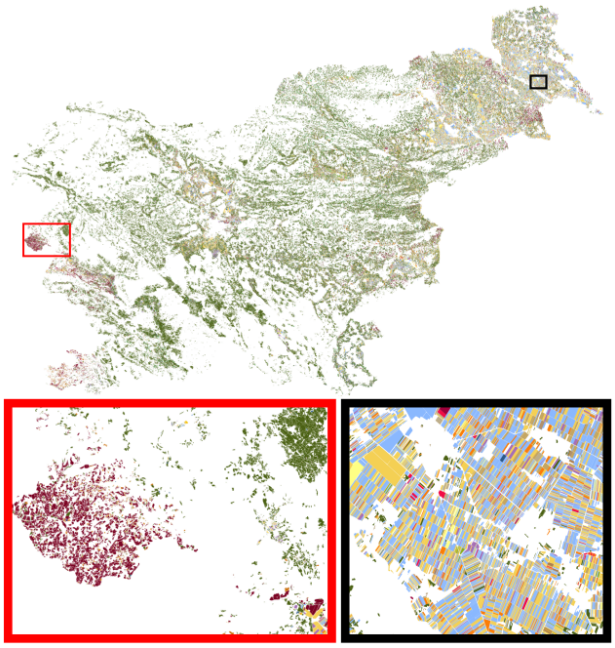

Figure 1. Crop coverage map in Slovenia for the year 2017. Enlarged areas show the diversification of crops in the country.

2. SATELLITE DATA

This research is focused on the use of Sentinel-2 data, which is openly accessible within the Copernicus program. Sentinel-2A and B together cover every area on Earth in at least 5 days in 13 bands. This high temporal resolution makes it possible to track seasonal trends, such as crop development, well. The most commonly used bands for vegetation mapping are the visual bands (2, 3, 4) and the near infrared band (8). These bands are

* Corresponding author

also the only ones available at 10 m, as others are acquired in 20 and 60 m and were re-sampled to 10 m resolution. With all the raw bands at the same resolution we reduce the complexity of further processing steps. We divided the area of Slovenia into squares of 1000 x 1000 pixels (i.e. 10 x 10 km), so it can also be processed by PC or laptop for simple analysis. In total approximately 300 patches were generated. Patches are visualised in Figure 2, yellow patches are used in training, the data colored in green was used for testing. The remaining patches were discarded as they had little to no crops. We separated the data spatially to ensure that the results were spatially generalised.

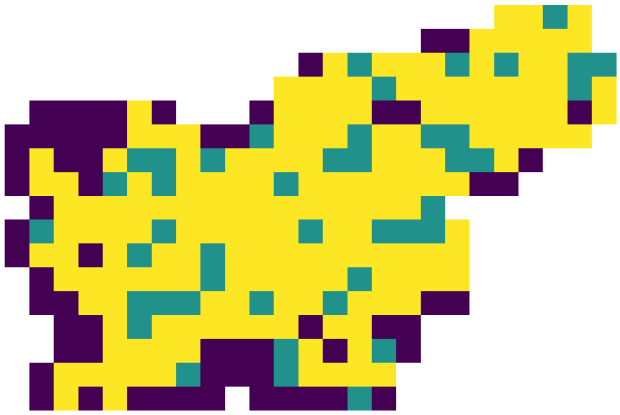

Figure 2. Data separation.

Data was downloaded from Sentinel HUb using the sentinelhub-py (Sinergise EO Research team et al., 2017) Python library and the study period was limited to the months from January to September of 2017, as this are the months when the changes in agricultural land are most visible. In subsequent months, in some areas winter crops for the next year are already being prepared.

All data was pre-processed using the eo-learn (Sinergise EO Research team et al., 2018) Python library to remove cloudy observations and construct indices which have been used also to classify crops also in related work (Pelletier et al., 2019). All values are normalised using min-max normalisation as suggested in (Pelletier et al., 2019). This normalisation subtracts the minimum value from each band and then divides it by its maximum. As this normalisation is highly sensitive to extreme values they further propose to use 2% and 98% percentile rather than the minimum and maximum value. This retain the temporal profile of the observed classes, as shown in Figure 3, and it retains all values within [-1,1]. After removing the clouds we are left with missing values in the time series. Which are most frequently weeks but can in some cases extend to a few months. Using linear interpolation we fill the gaps and provide a common time interval of the satellite data. This interpolation is very fast in comparison to alternatives, it is not computationally expensive and still retains enough information (Valero et al., 2016). But in case of larger gaps caused by clouds we now only have an average value between the measurements. This posses an issue when analysing seasonal trends of crops in cloudier regions. Which could be avoided by smoothing, but it comes with other challenges.

The entire processing pipeline consists of four steps. First we erode the polygons, to remove the effect of the edge values. We used a buffer of size 7 m, with this we excluded pixels that

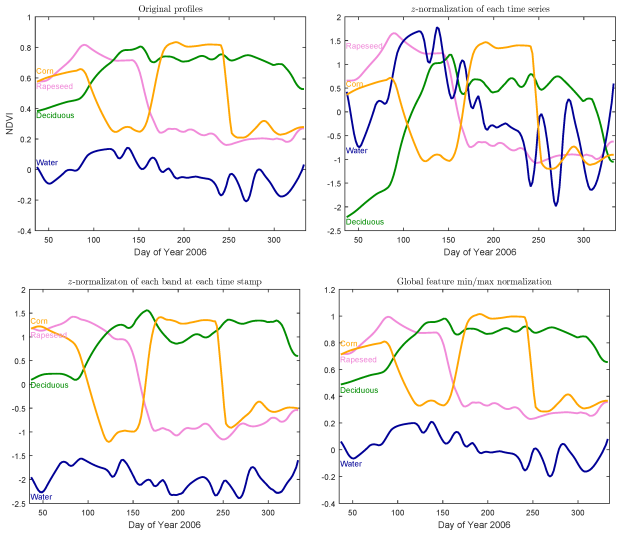

Figure 3. Normalisation taken from (Pelletier et al., 2019).

could potentially include other bordering classes or neighbouring fields. Then we transform the polygons into a matrix which corresponds to the size of the observed area. Lastly we randomly sample the pixels of each patch. Alternatively, weighted sampling could be used to attain equal distribution of all classes. We choose to better capture the data distribution and tackle the class imbalance at the training phase. The selected pixels were then interpolated to the 5 day interval, matching the Sentinel-2 revisit interval. Higher frequency of interpolation could provide more detailed trend without loosing some information. The downside of higher frequency would be the increased complexity both for data storage and computational power. With more time between each observation we are risking of missing sudden changes, such as sowing, that would be a indicate ripeness of crops and their collection.

3. REFERENCE DATA

The reference data was extracted from the database used for agricultural subsides, collected and managed by the Slovenian Agency for Agricultural Markets and Rural Development. Access to the data was granted within the project Perceptive Sentinel[1] funded by the EU.

Provided data consisted of 200 crops, the classification is very detailed, for most classes, several sub spices of crops are listed. Separating into such detailed groups is not always possible based on satellite imagery alone. Most groups are also were very few in number so joining them provided some larger classes that are better represented. For the purpose of this study crops were aggregated to 25 taxonomically similar groups. In Figure 1, we can see coverage of final crop classes in Slovenia. Some classes, such as hop and vineyards, are present only in certain regions, which effects the both training and results. When a class is not present in the training the model will predict that class at random and with low probability. In case of a class missing in the testing set we have to handle that separately. Whenever the class would be predicted, but not present, the prediction would be wrong. This can negatively effect the performance of the model.

[1] European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement 776115

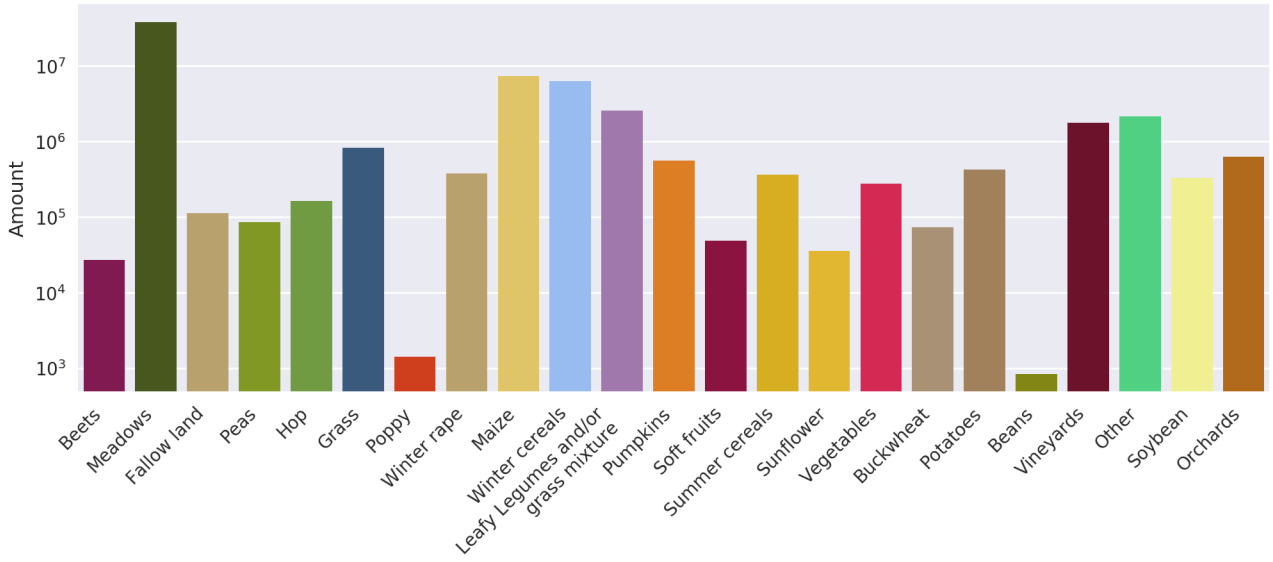

Figure 4. Crop distribution in Slovenia, the scale is logarithmic.

Figure 4 shows the distribution of crop classes in the data, including corresponding colors and names. Some classes were discarded as they presented less then 0.4% of crops in Slovenia. The remaining groups were:

- meadows,
- grassland,
- winter rape,
- maize,
- winter cereals,
- leafy legumes and/or grass mixture,
- pumpkins,
- summer cereals,
- vegetables,
- potatoes,
- vineyards,
- soybeans, and
- orchards.

Study	area in km^2	number of polygons
this work	20,273	803,201
Pelletier et at.	576	1,419
Rußwurm, Körner	4,284	137,000

Table 1. Overview of amount of data in (Pelletier et al., 2019) and (Rußwurm, Körner, 2018) approaches.

Related studies have been limited to smaller regions and/or polygon count as is presented in Table 1, where we compare area and number of polygons of each study. Further comparison to related work was not possible as in both studies RNN (Rußwurm, Körner, 2018) and TempCNN (Pelletier et al., 2019) reference data was provided by local agencies, which have made the data available only for the specific studies and not for sharing. Some differences are expected as Slovenia has smaller fields and consequently most pixels are on the edge, so we expect the data to contain more noise.

4. METHODS

In the first step, we used an algorithm similar to a random forest (Breiman, 2001), Since it has achieved good results in various classification tasks. The input of the training algorithm is a vector that includes spectral bands and indices for each observed point. In case of temporal information the vector size increases to $indices * temporalSteps$ and the temporal structure of the point is lost. We used a gradient boosting framework, that uses tree based learning algorithms. It differs from random forest algorithms in construction of trees. In every iteration we construct a new tree which minimises the error of the previous ones. Specifically, we choose LightGBM (Ke et al., 2017). It is faster, more efficient and simple to use than most similar implementations. The major advantages are in needing less RAM, can be speed up by using a GPU and offers many parameters that can be fine tuned to achieve desired performance.

We have compared it with two convolutional neural networks that are capable of processing temporal data. TempCNN was recently proposed and tested for classification of crops in South West France (Pelletier et al., 2019). As it had outperformed random forests, we expected it to outperform even gradient boosting methods as they to do not retain the temporal structure. The TempCNN architecture show in Figure 5 consist of three convolutional layers which are used to join the temporal information. Which is a fully connected layer that based on the condensed information provided by the previous layer predicts probability of the input belonging to the specified classes.

Compared to TCN (Bai et al., 2018), that was proposed as an alternative to RNN when working with temporal. This approach has not yet been tested on satellite imagery. Main advantage over the RNN is computational power and memory needed. States are not depended of the previous ones as is the case with RNN, which makes backpropagation faster and learning more memory efficient. The method in some cases outperforms RNN, especially when longer history is needed. The architecture is entirely made of convolutional layers, which are well optimised to be run on GPUs. An example of such architecture is shown in Figure 6 with the blue lines showing the captured information of each filter and layer. Architecture used in this task has two more convolution layers, that can be interpret similarly. Additional layers are required so the entire input vector is covered. One of key differences from the previous approach is the dilatation on each layer. In each layer the filter uses bigger dilatation, that grows exponentially with the depth of the network and effectively expands the receptive field of the net-

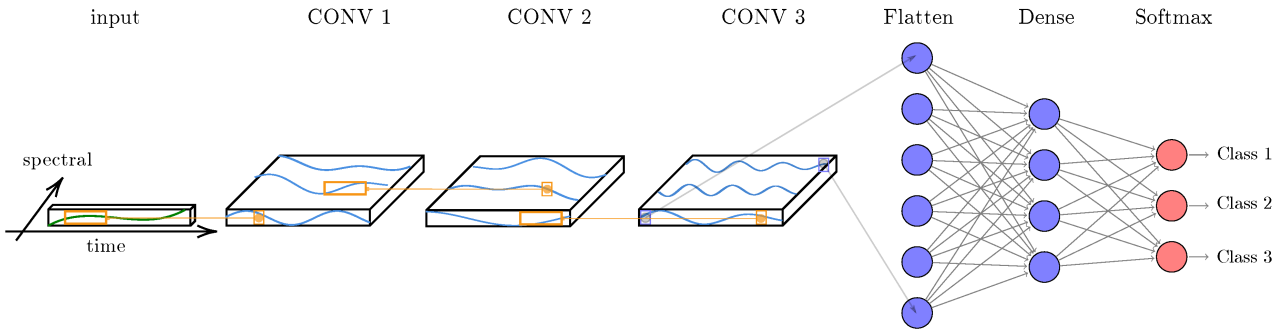

Figure 5. TempCNN architecture according to (Pelletier et al., 2019).

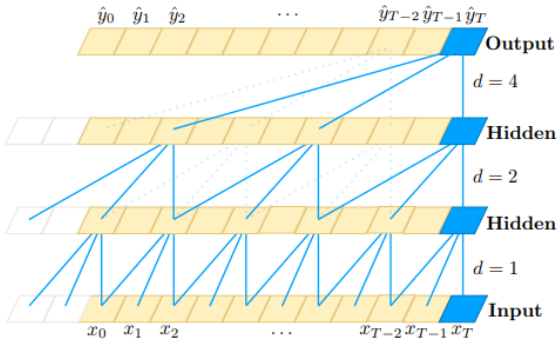

Figure 6. Example of TCN architecture with dilatation $d = 1, 2, 4$ and filter size 3 taken from (Bai et al., 2018).

work. In Figure 6 we can also see that by using dilatation we only overlap on neighbouring values. With this changes networks are more efficient and we can have large effective history, without requiring a lot of memory or computational power during training. As the same filters are applied thought the entire layer and can be run in parallel.

5. EVALUATION

To evaluate performance of each approach we first divide the data into smaller parts which represent the dataset. In case of multi-class classification we have to make sure all classes are present in both training and testing dataset. With this we have a supervised learning problem, as we have classes corresponding to all input sequences. Throughout the training process the method adapts the network weights to map the input values to the desired classes on the output. Many different metrics are available to asses the performance of methods.

Results can be displayed in a confusion matrix. In case of binary classification the table has two rows and two columns. Which can be expanded to include more classes with additional columns and rows, one for each class. In all cases columns contain classes predicted by the models and rows present the reference class which each example belongs to. Most commonly accuracy is used which represents the percentage of correctly classified samples (True Positive) against all samples. Recall measures how many samples (TP) of the class were correctly classified as belonging to the class divided by all samples of the class in the data (TP+False Negative). Metric that combines both is F1 which offers a single value to present the two. As we have multiple classes we measure all the metrics per each class. Usually during training we monitor overall accuracy. Which is accuracy weighted by the number of samples. It is most informative when all classes are equally represented. This is not always true in real life examples. In our case, the models quickly learned to classify meadows and achieved over 70% accuracy but performed poorly on other classes. We weighted all classes equally during the training of the model and monitored the macro accuracy.

6. RESULTS

The class distribution in Slovenia is shown in Figure 4. The landscape is dominated by meadows, which account for 60% of the data. In some regions there are very specific groups of crops such as hop and vineyards. Based on the class distribution, we could achieve an overall accuracy of 60% with the prediction of the class meadows for all pixels. So in Table 2, we focus on per class accuracy. In general, the results are comparable for most classes. The average F1 score is between 51%-53%. All methods have high success in classifying meadows, maize and winter cereals. Difficulties occur in classifying grassland, vegetables, summer cereals, potatoes and orchards. This is probably due to the overlapping of the temporal pattern for the classes. Meadows are similar to grassland and leafy legumes and/or grass mixture. Vegetables contains a lot of different vegetables types, which seems to results in lower performance. Even with some classes having low F1 score we still achieve high weighted average of 87% as the data distribution is in favor of meadows. With the difficulty mainly in classes with fewer samples the overall performance is promising.

As can be seen in Table 2, Neural networks outperform LightGBM, but only by one to two percent in the F1 score. LightGBM surpasses both other methods in the classification of summer cereals. The two neural networks achieve similar results, differences are visible in hops classification, while TempCNN achieves a lower accuracy but higher recall, which is more important because we want our predictions to be correct more often. The reason for the lower F1 result could be that the TCN has four times fewer parameters.

Weighted average at the bottom of the Table 2 represents the accuracy for all crops based on the number of samples. Both neural networks correctly classify 87% of all pixels. Since the test and training data were spatially separated, we assume that the score represents the model's ability to generalise. The models could be fine-tuned with appropriate data for each region or year. We expect that the model could achieve similar scores in countries with similar geography and for the same crop types. With high quality reference data, networks can achieve good performance on well represented crops. Problems occur when we have similar classes or mixture reference is provided as was the case in vegetables.

	LightGBM			TempCNN			TCN		
	Accuracy	Recall	F1	Accuracy	Recall	F1	Accuracy	Recall	F1
Meadows	95	71	81	98	87	93	97	89	93
Hop	30	87	44	82	92	87	87	58	70
Grassland	5	28	8	2	54	5	0	0	0
Winter rape	82	87	84	75	98	85	84	93	88
Maize	95	87	91	95	90	92	93	90	92
Winter cereals	92	85	89	93	91	92	93	88	90
Leafy legumes and/or grass mixture	23	41	30	27	63	38	22	57	32
Pumpkins	64	73	68	54	89	68	73	65	69
Summer cereals	18	54	27	15	52	23	7	56	12
Vegetables	3	54	27	5	7	6	8	6	7
Potatoes	8	55	14	39	40	39	37	17	24
Vineyards	47	67	55	21	94	34	51	70	59
Soybeans	86	81	83	55	98	70	68	97	80
Orchards	5	47	9	7	34	12	9	24	13
Average	46	66	51	48	71	53	52	57	52
Weighted average		72			87			87	

Table 2. Classification score per crop type.

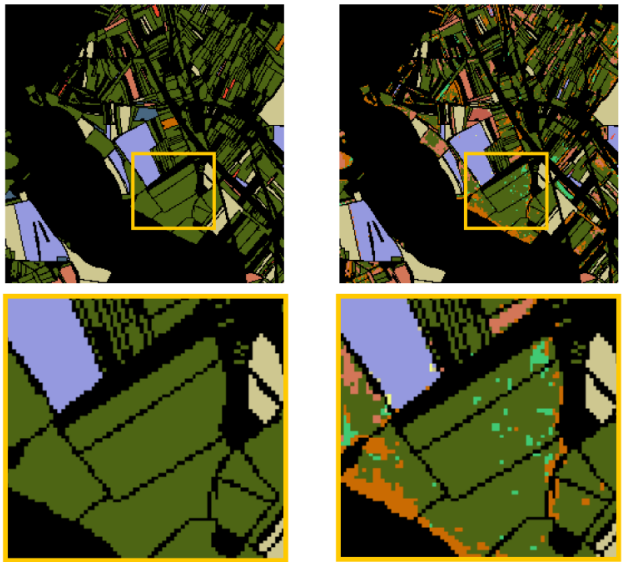

Figure 7. Reference on the left and result of TempCNN on the right.

In Figure 7 a visualisation of a reference area and the corresponding prediction of TempCNN is shown. TempCNN achieves higher recall which means it is more frequently correct. The method correctly predicts the majority of the classes. Issues are most common on the edges of the fields, but not exclusively. Meadows are in some cases predicted as pumpkins or orchards. This could be a problem from the definition of the class as orchards commonly have some space in between filled by meadows. Pumpkins grow more in width and can be overshadowed by tall grass which in turn causes confusion between the classes. As we know where the polygons are, we could achieve better classification visualisation by taking the most frequently predicted class. More intriguing would be to expand the method to include additional spatial information which is available in satellite imagery. It would most certainly remove the confusion within the fields, as it is uncommon for a single observation to belong to a different class then its neighbours. Which is often the case in the observed area.

7. CONCLUSIONS

Machine learning and remote sensing data are becoming more and more widely accessible and are thus gaining importance in many applications. Several machine learning algorithms have been used in the remote sensing community since decades, but only recently the availability of dense high resolution satellite image time series enabled the application of more advanced methods. In this paper we used Sentinel-2 data for classification of crops in Slovenia for the growth year 2017. We have compared three approaches, the baseline LightGBM and two deep learning approaches to handling temporal data.

Both TempCNN and TCN achieved comparable results for classification. TempCNN has been proven to work well by us and (Pelletier et al., 2019), while the evaluated TCN architecture offers an alternative when we have less data, computing power or time available. Both methods achieve 52%-53% F1 score for selected crop types and would perform equally good when presented with well annotated data.

For future work both methods could be extended to the use of spatial information (context). These models would potentially be more robust and would remove noise in individual polygons, i.e fields. As both models achieve similar performance, TCN would be more suited due to its lower computational speed. It has fever parameters which increase drastically with inclusion of another dimension to the data. Clouds still pose a major challenge in classification of land use and land cover, and radar images could provide additional information for periods and areas with high cloud cover. Deep learning offers various ways for multi-sensor merging, each having their advantages and drawbacks.

ACKNOWLEDGEMENTS

The authors acknowledge the project No. J2-9251 M3Sat – Methodology of Multitemporal Multisensor Satellite Image Analysis and research core funding No. P2-0406 Earth observation and geoinformatics were financially supported by the Slovenian Research Agency (ARRS). Access to reference data was granted within the EU project Perceptive Sentinel, H2020 grant No. 776115. Matej Račič is supported by ARRS young researcher funding No. 0792-N-53604. Luka Čehovin Zajc is supported by ARRS grant Z2-1866.

REFERENCES

Bai, S., Kolter, J. Z., Koltun, V., 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *CoRR*, abs/1803.01271. http://arxiv.org/abs/1803.01271.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 3146–3154.

Pelletier, C., Webb, G. I., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing*, 11(5). https://www.mdpi.com/2072-4292/11/5/523.

Raj, K. J., SivaSathya, S., 2014. Svm and random forest classification of satellite image with ndvi as an additional attribute to the dataset. *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, Springer, 95–107.

Rußwurm, M., Körner, M., 2018. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, 7(4). http://www.mdpi.com/2220-9964/7/4/129.

Sinergise EO Research team et al., 2017. sentinelhub-py. `https://github.com/sentinel-hub/sentinelhub-py`.

Sinergise EO Research team et al., 2018. eo-learn. `https://github.com/sentinel-hub/eo-learn`.

Valero, S., Pelletier, C., Bertolino, M., 2016. Patch-based reconstruction of high resolution satellite image time series with missing values using spatial, spectral and temporal similarities. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2308–2311.