

Beata Hejmanowska<sup>1</sup>, Mariusz Twardowski<sup>2</sup>, Anna Żądło<sup>3</sup>

## An Application of the “Traffic Lights” Idea to Crop Control in Integrated Administration Control System

**Abstract:** The aim of the paper is to discuss the idea of marking agricultural parcels in the control of direct payments to agriculture. The method of using remote sensing to monitor crops and mark them according to the idea of “traffic lights” is introduced. Classification into a given “traffic lights” color gives clear information about the status of the parcel. The image classification was done on Sentinel-1 and Sentinel-2 datasets by calculating the NDVI and SIGMA time series in the season from autumn 2016 to autumn 2017. Two approaches are presented: semi-automated and automated classifications. Semi-automated classification based on NDVI\_index and SIGMA\_index. Automated classification was performed on NDVI by Spectral Angle Mapper method and on SIGMA by Artificial Neural Network (Multilayer Perceptron, MLP method). The following overall accuracy was obtained for NDVI\_SAM: 70.35%, while for SIGMA\_CNN it was: 62.01%. User accuracy (UA) values were adopted for traffic lights analysis, in machine learning: positive predictive value (PPV). The UA/PPV for rapeseed were in NDVI\_index method: 88.1% (6,986 plots), NDVI\_SAM: 85.0% (199 plots), SIGMA\_index: 61.3% (4,165 plots) and in SIGMA\_CNN: 88.9% (2,035 plots). In order to present the idea of “traffic lights”, a website was prepared using data from the NDVI\_index method, which is a trade-off between the number of plots and UA/PPV accuracy.

**Keywords:** crop monitoring, IACS, Sentinel-1, Sentinel-2, NDVI, sigma, CNN, SAM

Received: 9 August 2021; accepted: 25 September 2021

© 2021 Authors. This is an open access publication, which can be used, distributed and reproduced in any medium according to the Creative Commons CC-BY 4.0 License.

<sup>1</sup> AGH University of Science and Technology, Krakow, Poland,  
email: galia@agh.edu.pl, ORCID ID: <https://orcid.org/0000-0003-0230-8386>

<sup>2</sup> AGH University of Science and Technology, Krakow, Poland,  
email: misiekt@agh.edu.pl, ORCID ID: <https://orcid.org/0000-0001-7107-6190>

<sup>3</sup> AGH University of Science and Technology, Krakow, Poland,  
email: annazadlo888@gmail.com, ORCID ID: <https://orcid.org/0000-0001-6575-6763>

## 1. Introduction

The Integrated Administration and Control System (IACS) [1] is an information system for the management of payments to farmers in European Union countries under the principle of share management. Support for farmers comes from the European agricultural guarantee fund (EAGF). The main aims of IACS in the farm context are:

- to carry out transactions correctly,
- to recover unduly paid amounts,
- to support farmers in making correct applications.

It is also important to manage and control the support in a standardized way throughout the EU. National administrations provide pre-established information, check if farmers meet the conditions for income support, and update applications for the following year. To meet these goals, IACS consists of digital databases, such as [1]:

- Land Parcel Identification System (LPIS) – for the identification of plots in EU countries,
- Geospatial Aid Application (GSAA) – for farmers to graphically indicate the agricultural area for which they are applying,
- an integrated control system based on computational cross checks and physical on-farm controls.

### 1.1. IACS in Poland

The organization which implements IACS in Poland is the Agency for Restructuring and Modernization of Agriculture (ARMA) (in Polish: Agencja Restrukturyzacji i Modernizacji Rolnej). ARMA began the realization of the IACS goals in June 2001, and is responsible for [2]:

- a register of animals kept for farming purposes,
- a register of direct payments,
- documentation referring to the register of farms and to the subsidies granted and paid,
- documentation referring to controls and regulations conducted by the IACS.

ARMA is also responsible for keeping and updating the LPIS in Poland. This system is based on plans and cadastral documents, cartographic materials, geographic information system (GIS) and aerial or spatial imagery.

IACS in Poland consists of the non-IT part, created by ARMA, and an IT part created by the Asseco Poland company. Its main aim is to manage and control of the use of European Union funds allocated to farmers. The system prevents the occurrence of irregularities and abuses due to the use of advanced recoding and control mechanisms. The controls are carried out to assess their compliance with the standards: automatic controls, substantive correctness as well as consistency of the study and visual controls [3].

## 1.2. Analysis of the Problems. The Concept of a “Monitoring Approach” and the Idea of “Traffic Lights”

Three documents prepared by Joint Research Center (JRC) [4–6] describe the concept of checks by monitoring (CbM) substituting the current solution (on the spot checks) which is time consuming and required a lot of field visits. The aim of the ‘monitoring approach’ is to simplify, reduce the burden of controls and perform controls remotely which can be applied systematically. The idea needs to be developed for a specific area, customized to local requirements by adopting newly available technology such as cloud processing and machine learning (ML) algorithms. Examples of the application of ML methods in crop recognition can be found in the publications [7–12].

A working definition of monitoring was proposed by JRC and specified by the Commission Implementing Regulation (EU) 2018/746 amending Implementing Regulation (EU) No 809/2014 as [5]:

“Procedure of regular and systematic observation, tracking and assessment of all eligibility criteria, commitments and other obligations which can be monitored by Copernicus Sentinels satellite data or other data with at least equivalent value, over a period of time that allows to conclude on the eligibility of the aid or support requested” with, “where necessary, and in order to conclude on the eligibility of aid or support requested, appropriate follow-up activities”.

The data sources offered by ESA Sentinel-1 (S-1 radar) and Sentinel-2 (S-2 optical) satellites provide images across the territory of European Union with the nominal revisit every 5 days for S-2 and every 6 days for S-1. Sentinel images can be enhanced with additional imagery like high resolution or hyperspectral data and with data sources which is the evidence from farmers as geo-tagged images.

One of the important aspects of the project is assessing eligibility conditions met by declared agricultural parcels. The assessment is carried out in several stages [6]:

- The declared and actual parcels areas correspond.
- The monitored land should be compliant with the measure associated with the declared agriculture parcel requires.
- The evidence of an incompatibility that impacts payment.
- The term of conclusion on the payment to the holding can be made when a sufficient area has been confirmed.
- Noncompliance will be warned about early.
- Monitoring after payment is continued to screen any infringements and obligations to the scheme.

The farmer’s application form as well as external data such as the time series data from S-1 and S-2 are needed to assess if the declaration of what has been planted in the field and how it behaves is compatible with the truth on the ground. The assessment is based on a reductive approach. From the beginning of the application

year, information about the land is only gathered. Then, when the declaration clearly states which conditions need to be monitored and when monitored sources become available, parcels should be assigned using any logic code with respect to a particular scheme. The proposed coding is based on assigning colors to each plot depending on the conditions that are met and has been called “traffic lights”. “Flashing (blinking) lights” suggest that additional information or follow-up action is needed. The meaning of each color is as follows [6]:

- Black (no lights) – there is no actual declaration available, but the parcel is considered because it was declared in previous years. In many cases, older information can be relevant for a current declaration.
- White – the actual declaration is available, but the assessment is not yet complete.
- Flashing (blinking) yellow/ yellow – the parcel is assessed and probably the declared scheme/support measure is not in accordance with the requirements due to the absence of farmer action (warning alerts should be sent to the farmer).
- Flashing (blinking) blue – the judgment of an expert is required because the parcel has been assessed and probably the declared scheme/support measure is not in accordance with the requirements. To complement the monitoring, additional information is required.
- Yellow – the parcel has been assessed but the declared scheme/support measure cannot be confirmed or rejected because of insufficient evidence.
- Green – the parcel has been assessed and confirmed as compliant with the conditions of the declared scheme/support measure.
- Red – parcel has been assessed and confirmed as non-compliant with the conditions of the declared scheme/support measure.

The first two lights are meant to signal which parcels will be considered in the decision process. The next two flashing lights suggest that the process is ongoing in the case of the inconclusive yellow light cases. The last three represent a parcel’s state. It can be observed that green, red, yellow and flashing yellow lights correspond to work that is focused on an automation process without human control, while the flashing blue light needs human expert support in order to proceed.

The process of assessing “traffic lights” assumes that the parcel is monitored until it is possible to make a decision for the application year based on the markers. The decision is made when evidence for eligibility or ineligibility is observed. There is also uncertainty when observation or evidence are inconclusive or when the observation is delayed. The conclusive evidence depends on specific scheme and application. For example, for BPS (basic payment scheme) or EFA (ecological focus area) are considered otherwise, and different information are needed to evidence.

The general workflow can be as follow [6], and the more detailed one is shown in Figure 1 [5]:

1. The application made by the farmer determines the area of interest.
2. The parcel is assigned a green light if the required evidence is detected.
3. The parcel is assigned a red light if the conclusive counterevidence is detected.
4. The parcel is assigned a blinking yellow light if inconclusive counterevidence is detected. It may cause a request for additional information from the farmer. If it is sufficient, the light is reassigned as green but if it is not the light is reassigned as blinking blue and probably need more field data is required. The result of this activity should change the light to green or red.
5. The parcels that have not been assigned as red or blinking yellow are assigned as yellow and are treated as green lights.

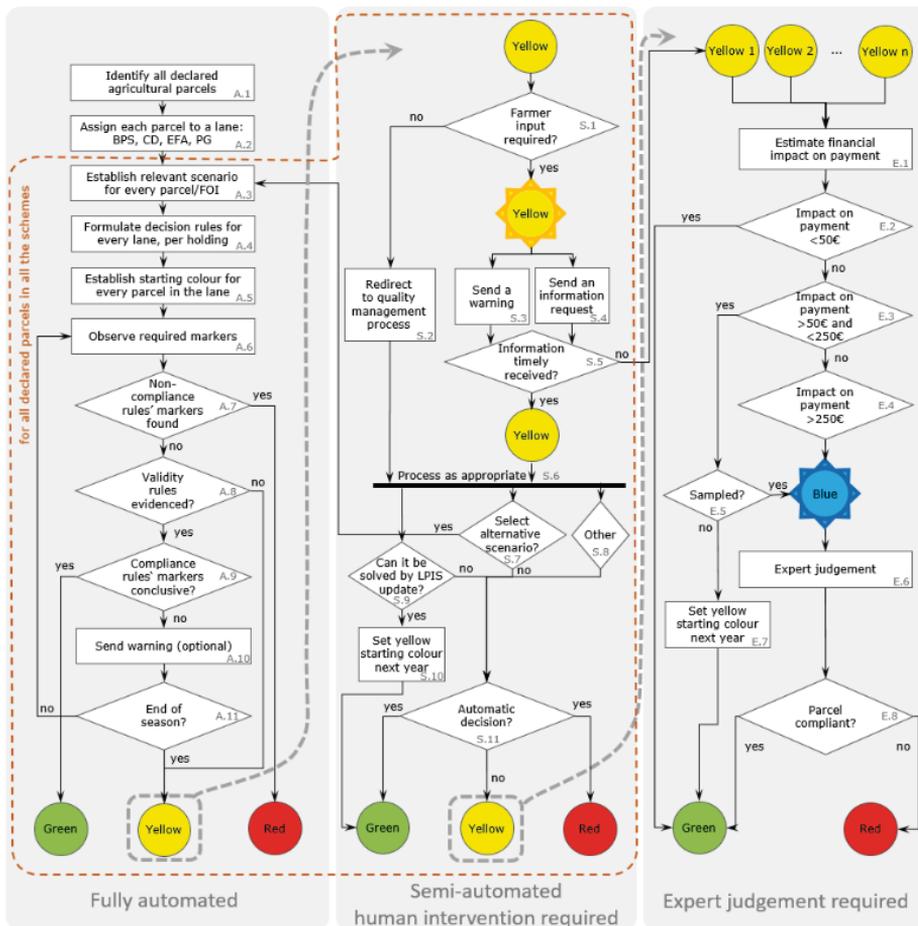


Fig. 1. The detailed workflow for how parcels could be assigned in “traffic lights” proposed by JRC

Source: [5]

Respectively, the evidence is justified based on three types of rules: *compliance rules* which indicate a compatibility between specific parcel and requirements and give a green light, *noncompliance rules* which indicate a contradiction between declaration and monitored data and give a red light, and the final *validity rules* which support the automatic process. General examples for *validity rules* could be observation changeability within a parcel which has ambiguities or observation of ploughing in a parcel that can cause changes to a scenario [5]. The rules are defined in general so paying agencies making a decision should choose appropriate criteria especially based on markers.

## 2. Material and Methods

### 2.1. Study Area

The recommendations of the area and shapes of parcels that can be monitored using Sentinel satellites was worked out by JRC [6]: the minimum size of parcel should be 0.5 ha and the monitoring will be done in areas where about 90% of the agricultural area is covered by agricultural parcels above the critical size.

Poland is spatially diverse in terms of its crop structure. Farms in northern Poland are typically remnants of former state farms and large parcels predominate. In southern Poland, the structure of crops is very fragmented and dominated by small plots often with an elongated shape.

The recommendation can be met in municipalities in northern Poland, while municipalities in the south do not even meet these requirements approximately.

The test area was selected in coordination with ARMA for 2 projects in 2018 and 2019 [13–15]. Initially, the commune of Brzeźno located in the north-western part of Poland was chosen as a representative test area. Due to the impossibility of obtaining unclouded time series within the whole Brzeźno commune, the search was extended (Fig. 2 – plots in green) and a different test area was selected for the S-2 and S-1 analyses (Fig. 2 – plots in gray and yellow respectively).

### 2.2. Datasets

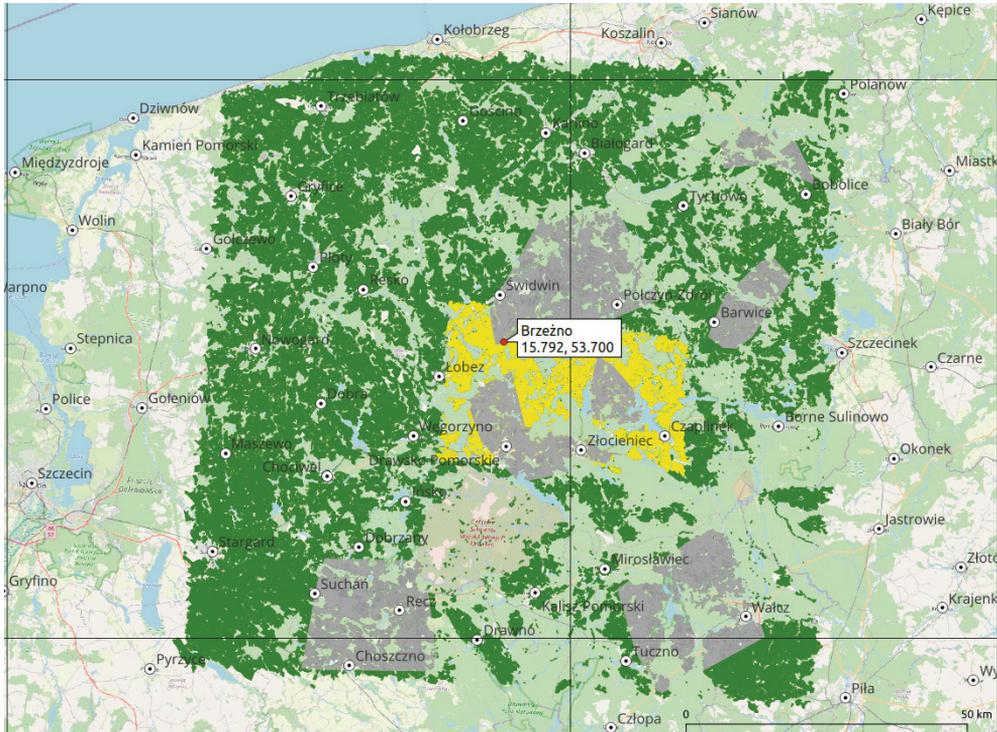
There were two types of datasets which are collected:

- 1) vector – a shapefile with the polygons defining the crops declared by the farmers,
- 2) raster – Sentinel-1 (S-1) and Sentinel-2 (S-2) time series.

Vector data containing information about the declaration was kindly provided by ARMA and was pre-filtered to remove parcels smaller than 0.3 ha.

Sentinel images were downloaded from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). S-1 mission use C-band Synthetic Aperture Radar which provide images regardless of weather and light conditions. Data are available on three processing level: Level-0 contain raw data, Level-1 produced as Single

Look Complex (SLC) and Ground Range Detected (GRD) and Level-2 contain components for Ocean Swell Spectra (OSW). Level-1 GRD was used in our research. Sentinel-2 provides multispectral data in the visible spectrum (VIS), red edge (RE), NIR and SWIR. The spatial resolution is 10 m for VIS and NIR, 20 m and 60 m for infrared bands and red-edge.



**Fig. 2.** Test area: all parcels in green obtained from ARMA, parcels selected in Sentinel-1 time series analysis in yellow (16,494), parcels selected in Sentinel-2 time series analysis in gray (27,803)

The data covered the season from autumn 2016 to the autumn of 2017.

9 images of S-1 were chosen for the Brzeżno area (for days: 22.03.2017, 03.04.2017, 21.05.2017, 05.06.2017, 08.06.2017, 23.07.2017, 27.08.2017, 24.09.2017, 26.09.2017) and 10 images of S-2 (for days: 03.09.2016, 12.11.2016, 12.12.2016, 12.03.2017, 22.03.2017, 10.04.2017, 21.05.2017, 29.08.2017, 11.09.2017, 28.09.2017).

Cloud cover is not an obstacle in radar registration. Therefore, one continuous test area in the center was chosen for analysis (16,494 plots, Fig. 2 – in yellow). For 4,165 parcels, the plant is given, for the remaining 12,329 parcels, the single area payment (SAPS) declared.

Acquisition of cloud-free S-2 time series in temperate climates is difficult, especially for large areas. It is not possible to acquire S-2 time series even for a single commune. Thus, a few fragments of unclouded areas on all 10 images were selected for testing. In these areas 27,803 plots were declared (Fig. 2 – in grey). For 6,986 parcels, the plant is given, for the remaining 20,817 parcels, the single area payment (SAPS) declared.

NDVI [16] was calculated from S-2 images, and SIGMA backscattering coefficient [17, 18] from S-1 images. For each declared plot selected for analysis, the average value of the NDVI/SIGMA of all pixels within a given plot was calculated from the NDVI/SIGMA image. The mean NDVI/SIGMA value of the successive dates of the time series was assigned to the plot as new attributes. Thus, each plot was assigned 10/9 new attributes.

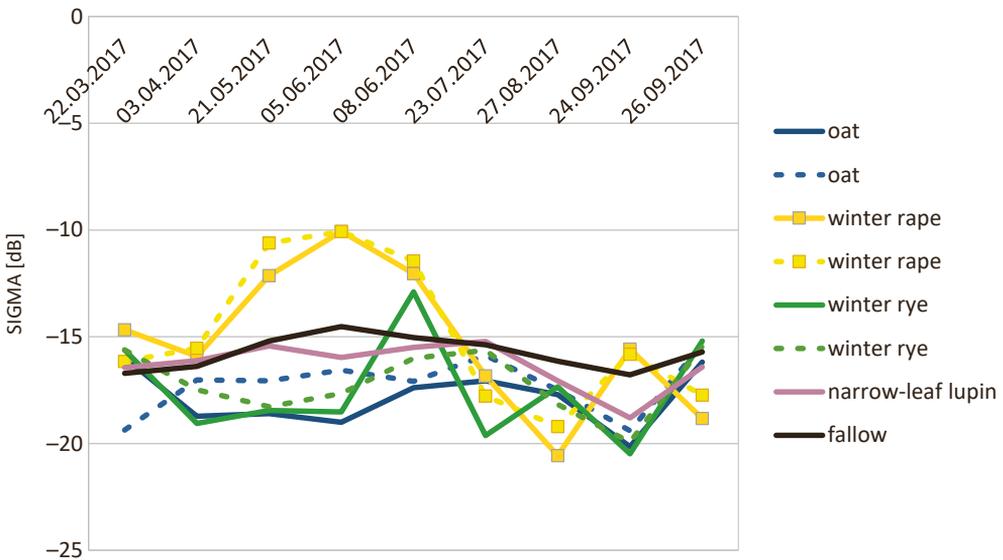


Fig. 3. Declared parcels with crops, selected parcels (in yellow) with "id" concerning NDVI layer

An enlarged part of the study area is shown in Figure 3. 8 sample plots were selected (Fig. 3 – in yellow, with id and plant label). Table 1 shows the 9 new attributes (SIGMA time series) and Table 2 (NDVI time series) for each selected parcel. Charts in Figures 4 and 5 present the time series of SIGMA/NDVI for each plant. The course of the curves makes it possible to define indicators based on which the classification can be done in order to separate the plots covered by particular crops.

**Table 1.** Examples of SIGMA values as attributes of parcels

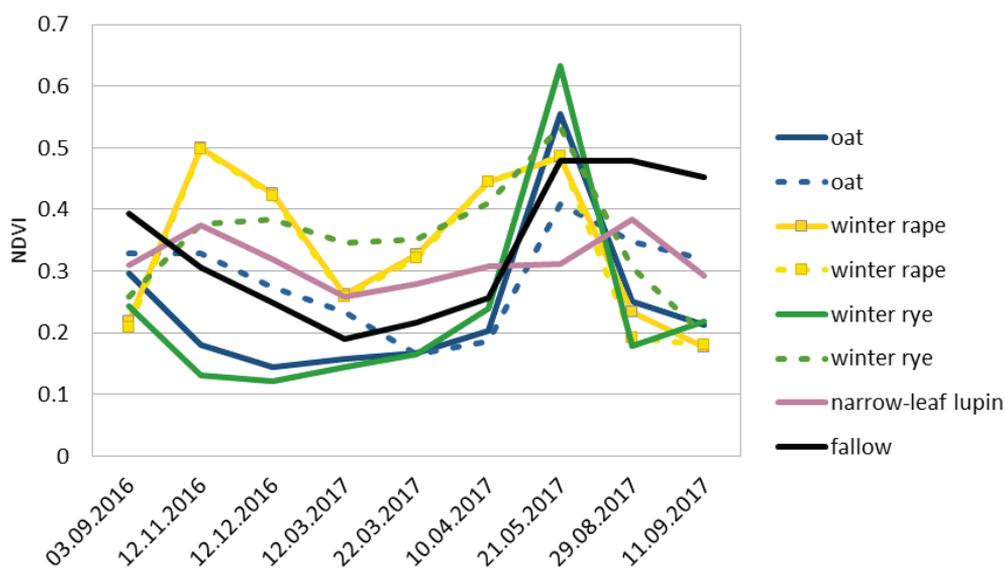
Crop/id/ data	Oat/3842	Oat/767	Winter rape/7892	Winter rape/3844	Winter rye/11163	Winter rye/16308	Narrow-leaf lupin/5633	Fallow/3449
22.03.2017	-15.986	-19.365	-14.681	-16.158	-15.616	-15.617	-16.447	-16.71
03.04.2017	-18.72	-17.022	-15.915	-15.54	-19.064	-17.483	-16.117	-16.394
21.05.2017	-18.607	-17.07	-12.142	-10.611	-18.456	-18.284	-15.425	-15.197
05.06.2017	-19.001	-16.567	-10.069	-10.076	-18.521	-17.652	-15.968	-14.526
08.06.2017	-17.383	-17.085	-12.041	-11.446	-12.901	-16.01	-15.493	-15.037
23.07.2017	-17.066	-15.874	-16.836	-17.781	-19.626	-15.635	-15.221	-15.372
27.08.2017	-17.714	-17.491	-20.565	-19.207	-17.346	-18.173	-17.063	-16.151
24.09.2017	-20.146	-19.402	-15.578	-15.8241	-20.483	-19.937	-18.801	-16.785
26.09.2017	-16.192	-15.347	-18.837	-17.738	-15.1996	-15.452	-16.412	-15.72
Area [m <sup>2</sup> ]	36,358.34	36,136.73	71,813.27	57,008.71	19,670.64	20,950.04	32,039.14	81,920.31



**Fig. 4.** SIGMA time series charts of selected parcels (rape\_index\_S1=SIGMA 05.06.2017 ≥ -13.7)

**Table 2.** Examples of NDVI values as attributes of parcels

Crop/id/ data	Oat/82	Oat/23	Winter rape/154	Winter rape/84	Winter rye/205	Winter rye/293	Narrow- leaf lupin/123	Fallow/65
03.09.2016	0.297	0.329	0.219	0.208	0.244	0.257	0.309	0.394
12.11.2016	0.180	0.328	0.500	0.498	0.131	0.377	0.374	0.305
12.12.2016	0.144	0.274	0.425	0.422	0.121	0.383	0.320	0.249
12.03.2017	0.157	0.233	0.263	0.258	0.144	0.345	0.258	0.190
22.03.2017	0.166	0.164	0.327	0.321	0.165	0.351	0.279	0.217
10.04.2017	0.204	0.186	0.445	0.445	0.239	0.410	0.308	0.257
21.05.2017	0.555	0.408	0.485	0.487	0.632	0.534	0.311	0.479
29.08.2017	0.250	0.347	0.234	0.191	0.178	0.307	0.384	0.478
11.09.2017	0.213	0.320	0.176	0.181	0.218	0.198	0.293	0.453
28.09.2017	0.277	0.317	0.207	0.202	0.321	0.370	0.330	0.447
Area [m <sup>2</sup> ]	36,358.34	36,136.73	71,813.27	57,008.71	19,670.64	20,950.04	32,039.14	81,920.31

**Fig. 5.** NDVI time series charts of selected parcels  
(rape\_index\_S2=NDVI12112016-NDVI3092016)

### 2.3. Classification Methods

Crop recognition using S-1 and S-2 data was performed using the object-oriented classification method by two methods: semi-automatic, based on time series charts, and automatically, using NDVI and SIGMA values of parcel attributes.

### 2.4. Semi-automatic Classification

For semi-automatic classification of NDVI/SIGMA datasets, the variability curves over the time were calculated for each crop (as in Figs. 3, 4). By analyzing the charts, it is possible to see the characteristic moments of the phenological development of the crop and agrotechnical procedures performed. Peaks and rapid changes of the NDVI/SIGMA curve are the basis to define classification criteria which can help to separate one crop from others, especially when the values are much different in specific time.

Let us analyze the curves for winter rape in (Figs. 4, 5). In Figure 4, the maximum SIGMA value can be observed between 21.05.2017 and 06.06.2017. This allows the use of simple thresholding to separate rape from other crops (in our case  $\text{rape\_index\_S\_1} = \text{SIGMA}_{05.06.2017} \geq -13.7$ ). In the NDVI plot (Fig. 5) one can notice the unique course of the rapeseed curve and the large increase in NDVI values from 0.2 to 0.5 between 03.09.2016 and 12.11.2016. Therefore, a formula can be given for:  $\text{rape\_index\_S2} = \text{NDVI}_{12112016} - \text{NDVI}_{3092016}$ , with a threshold ( $\text{rape\_index} > 0.3$ ) to separate the plots covered by rape.

In the second approach, automatic classification of the SIGMA and NDVI time series was performed. Classifications were not made on individual plants but rather groups of plants. Crop structure for the test area: S-1, SIGMA and S-2, NDVI are presented in Figure 6. Five classes have been selected for automatic classification, taking into account the number of declarations for a given crop (crops with low numbers have been removed): grass, winter rye, potato, rapeseed, other cereals, maize.

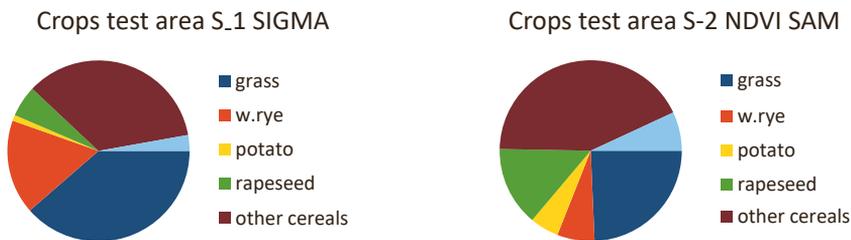


Fig. 6. Crop structure in the test area

SIGMA and NDVI were developed separately, so the test areas are different. Finally, a set of 2,904 plots was selected for SIGMA time series classification, which was divided into two independent subsets, taking: 30% as training and 70% as test. In the NDVI classification, on the other hand, a collection of 662 plots was adopted: 70% training and 30% testing.

### 2.5. Automatic Classification

Automatic classification of SIGMA time series was performed by means of the neural network (Multilayer Perceptron) method with 2 hidden layers in NeuroLab Python library [19]. Different scaling methods were selected during data preparation: *maxscale, normalize, scale, and stdscale*.

Automatic NDVI time series classification was performed by the algorithm adopted from image classification, namely the Spectral Angle Mapper (SAM). In this method, each plot is represented by a vector. The attribute values (NDVIs) create the coordinates of the vector. Thus, each plot has 10 coordinates (10 NDVI values, as in Table 2). Reference vectors were determined for each class from the training set, based on the average values of the coordinates of all vectors of a given crop. Then the angle between each of the reference vectors and the vector of each plot was calculated. The plot was assigned to the class for which the angle was the smallest.

### 2.6. Accuracy Assessment

The accuracy assessment was performed for each method by the calculation of the full confusion matrix and binary confusion matrix, also called the error matrix. In the full confusion matrix, classification results are compared to the true on the ground information. The binary confusion matrix is used in machine learning approach and classifies the results into four classes:

- 1) TP true positive – means that the parcel was classified as a declared crop,
- 2) TN true negative – means that the parcel was not classified as a declared crop and the declaration is different to the classification result,
- 3) FP false positive – means that the parcel was classified as a given crop, but the declaration is different,
- 4) FN false negative – means that parcel was not classified as a given crop but declared as that crop.

The most important parameter was calculated and shown in Table 3.

**Table 3.** Accuracy assessment parameters

Name	Description	Formula
Producer accuracy (PA) / sensitivity / true positive rate (TPR)	Number of correctly classified parcels in a given group of parcels declared to the number of all analyzed parcels in this group	$\frac{TP}{TP + FN}$
Specificity / true negative rate (TNR)	Number of parcels correctly classified as not the given group to the number of all negative cases	$\frac{TN}{TN + FP}$

Table 3. cont.

User accuracy (UA) / precision / positive predictive value (PPV)	Number of correctly classified parcels in a given group for all parcels classified as a given group	$\frac{TP}{TP + FP}$
Accuracy (ACC)	Number of correctly classified parcels as and as not a given group to the number of all parcels analyzed	$\frac{TP + TN}{TP + TN + FP + FN}$
F1 score	The measure of a test accuracy	$\frac{2 \cdot TP}{2 \cdot TP + FP + FN}$
Overall accuracy (OA) / percent of correct precision	Number of correctly classified parcels to the number of all parcels analyzed	$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + TN_i + FP_i + FN_i)}$

Source: [20, 21]

## 2.7. Classification Result vs. "Traffic Lights"

As a result of automatic classification, we obtained information about the class to which the plot has been classified and the values of the accuracy metrics as in Table 3. Overall accuracy and F1 score relate to classification accuracy in general. The other metrics refer to individual classes (crops/crop groups). According to the diagram in Figure 1, as a result of automatic classification, plots should be assigned one of three lights: green, red or yellow. A green light means that the classification result coincides with the declared crop, if it is different than a red light is "flashing". In cases of doubt, a yellow light is used and the procedure continues as shown in Figure 1: semi-automatic procedure and expert judgment. Note that the results of the automatic classification confirming or negating the correctness of the declaration are not processed further. However, it should be taken into account that no classification is perfect and there is no 100% certainty that a decision is correct, whether it is green or red.

In our study, we focused only on the first step of scheme Figure 1, and not in a complete way. Automatic and semi-automatic classification was performed (semi-automatic does not mean the second stage shown in this diagram). The outcome of the classification either confirmed the declaration or it did not. If the declaration is confirmed, a green light can be switched on, otherwise a red light is illuminated. In both cases, there is a risk of error. A farmer could be given the benefit of the doubt and assume that running a green light is less risky than running a red light. Therefore, we decided that a red light is not automatically switched on for FP and FN. Further action in the FP and FN cases may depend on the values of the accuracy metrics and classification reliability. For example, if we have high user accuracy, i.e., a small commission error, we can light a red light for FP with high

confidence. In the case of lower producer accuracy and a higher omission error, we would need to verify the FNs to avoid incorrectly putting the red light on for them. Therefore, even at the first stage, we decided to assign the color blue in case where an expert decision is need (for example a color composition analysis).

### 3. Results and Discussion

This subsection presents the results of NDVI and SIGMA time series classification for crop control using the semi-automatic method, SAM method and neural networks. Finally, classification results are presented in the form of a “traffic lights” map for rapeseed.

#### 3.1. Classification Results – Semi-automatic Method

The result of semi-automatic classification based on NDVI can be analyzed in Figure 7 (p. 147) and accuracy analysis of SIGMA and NDVI in Table 4, where there are:

- TP, TN, FP, FN,
- ML metrics: accuracy (ACC), specificity (TNR),
- metrics in traditional approach:
  - UA(PPV) together with commission error (CE)/false discovery rate (FDR) – percentage of the number of parcels declared as another crop then rape falsely classified as rape,
  - PA(TPR) together with omission error (OE)/false negative rate (FNR) – percentage of the number of parcels declared as rape falsely classified in another class.

**Table 4.** Reliability of the classification using the semi-automated method for rape

Classification method	TP	TN	FP	FN	ACC	TNR	UA(PPV)	CE(FDR)	PA(TPR)	OE(FNR)
SIGMA_index	152	4,143	96	22	0.973	0.977	<b>0.613</b>	<b>0.387</b>	<b>0.874</b>	<b>0.126</b>
NDVI_index	252	6,319	34	381	0.941	0.995	<b>0.881</b>	<b>0.119</b>	<b>0.398</b>	<b>0.602</b>

In the case of binary classification of originally multiple classes, it makes no sense to give the values of accuracy (ACC) and specificity (TNR), often used in machine learning, because they always reach values above 90–95% (in our case for rapeseed: 97.3% and 94.1% for SIGMA and NDVI respectively). Also, it makes no sense to give OA, and especially not to confuse it with accuracy.

For crop inspection, metrics of particular interest are over- and under-estimation error (commission and omission error) and the corresponding user/producer accuracy (in machine learning: PPV and TPR). Table 4 shows these metrics for

rapeseed. Commission error presents the percentage of the number of parcels declared as a crop other than rape or falsely classified as rape.

The SIGMA\_index method had a significant commission error: 38.7% and a small omission error: 12.6%.

In the NDVI\_index method, the reverse was observed: the commission error was small: 11.9%, while the omission error was very large: 60.2%.

This means that the reliability of the classification of oilseed rape based on SIGMA was much higher than using NDVI.

### 3.2. Classification Results – Automatic Methods

The neural network classification made on SIGMA dataset gives different results depending on the method of data standardization. The *maxscale*, *normalize*, *scale*, and *stdscale* were used and the *scale* gives the best results. Tables 5 and 6 present binary classification matrices for classification of SIGMA values by neural network method for both training and test dataset. Note the high mean values of validation ACC and TNR, which are above 99%, on the test data are also high above 80% except for cereals (ACC = 71.35%, TNR = 77.39%). In general, all metrics for the validation process average above 95%.

For the test dataset, the producer accuracy (PA/TPR) as well as the user accuracy (UA/PPV) and F1 are low and on average are respectively: 53.15%, 51.92% and 47.76%. Eventually it can be given the overall accuracy for validation is 98.50% and for test 62.01%. When analyzing the metrics for individual crops, only grass and rapeseed have acceptable values (F1 82.27% and 83.84% respectively). For potatoes, the classification result is completely unacceptable (TPR, PPV and F1 are 0). Despite this fact, the ML metrics of accuracy (ACC) and specificity (TNR) are above 95% (97.94% and 99.10% respectively).

**Table 5.** Binary confusion matrix for SIGMA classification – training dataset (OA = 98.50%)

Crop	TP	TN	FP	FN	ACC	TPR	TNR	PPV	F1
grass	341	517	7	4	0.9873	0.9884	0.9866	0.9799	0.9841
w_rye	134	732	0	3	0.9965	0.9781	1.0000	1.0000	0.9889
potato	10	856	1	2	0.9965	0.8333	0.9988	0.9091	0.8696
rape	41	827	1	0	0.9988	1.0000	0.9988	0.9762	0.9880
other cereals	304	558	4	3	0.9919	0.9902	0.9929	0.9870	0.9886
maize	26	842	0	1	0.9988	0.9630	1.0000	1.0000	0.9811
<b>Mean</b>	<b>143</b>	<b>7223</b>	<b>2</b>	<b>2</b>	<b>0.9950</b>	<b>0.9588</b>	<b>0.9970</b>	<b>0.9754</b>	<b>0.9667</b>

**Table 6.** Binary confusion matrix for SIGMA classification – test dataset (OA = 62.01%)

Crop	TP	TN	FP	FN	ACC	TPR	TNR	PPV	F1
grass	608	1165	95	167	0.8713	0.7845	0.8746	0.8649	<b>0.8227</b>
w_rye	110	1623	59	243	0.8516	0.3116	0.8698	0.6509	0.4215
potato	0	1993	24	18	<b>0.9794</b>	<b>0.0000</b>	<b>0.9910</b>	<b>0.0000</b>	<b>0.0000</b>
rape	96	1902	12	25	0.9818	0.7934	0.9870	0.8889	<b>0.8384</b>
other cereals	408	1044	278	305	<b>0.7135</b>	0.5722	<b>0.7739</b>	0.5948	0.5833
maize	40	1675	305	15	0.8428	0.7273	0.9911	0.1159	0.2000
<b>Mean</b>	<b>210</b>	<b>1567</b>	<b>129</b>	<b>129</b>	<b>0.8730</b>	<b>0.5315</b>	<b>0.9240</b>	<b>0.5192</b>	<b>0.4776</b>

Tables 7 and 8 present binary confusion matrices for the classification of NDVI values by means of the SAM method for both the training and test dataset. In this case, one can see a smaller discrepancy between the metrics calculated from the training and test data. The ACC and TNR values are high for both datasets (above 90%), but not as high as for neural networks, where they were above 99%.

**Table 7.** Binary confusion matrix for NDVI classification – training dataset (OA = 72.79%)

Crop	TP	TN	FP	FN	ACC	TPR	TNR	PPV	F1
grass	103	303	37	20	0.8769	0.8374	0.8912	0.7357	0.7833
w_rye	9	421	14	19	0.9287	0.3214	0.9678	0.3913	0.3529
potato	5	425	13	20	0.9287	0.2000	0.9703	0.2778	0.2326
rape	55	389	15	4	0.9590	0.9322	0.9629	0.7857	0.8527
other cereals	135	224	43	61	0.7754	0.6888	0.8390	0.7584	0.7219
maize	30	427	4	2	0.9870	0.9375	0.9907	0.8824	0.9091
<b>Mean</b>	<b>56</b>	<b>364</b>	<b>21</b>	<b>21</b>	<b>0.9093</b>	<b>0.6529</b>	<b>0.9370</b>	<b>0.6385</b>	<b>0.6421</b>

**Table 8.** Binary confusion matrix for NDVI classification – test dataset (OA = 70.35%)

Crop	TP	TN	FP	FN	ACC	TPR	TNR	PPV	F1
grass	29	142	19	9	0.8593	0.7632	0.8820	0.6042	<b>0.6744</b>
w_rye	3	174	9	13	0.8894	0.1875	0.9508	0.2500	0.2143
potato	3	187	3	6	0.9548	0.3333	0.9842	0.5000	0.4000
rape	34	158	6	1	0.9648	0.9714	0.9634	0.8500	<b>0.9067</b>
other cereals	58	93	19	29	0.7588	0.6667	0.8304	0.7532	<b>0.7073</b>
maize	13	182	3	1	0.9799	0.9286	0.9838	0.8125	<b>0.8667</b>
<b>Mean</b>	<b>23</b>	<b>156</b>	<b>10</b>	<b>10</b>	<b>0.9012</b>	<b>0.6418</b>	<b>0.9407</b>	<b>0.6283</b>	<b>0.6282</b>

For the test dataset, the producer accuracy (PA/TPR) as well as the user accuracy (UA/PPV) and F1 are low and on average are respectively: 65.29%, 63.85% and 64.21%. Ultimately, the overall accuracy for the training dataset can be given as 72.79% while the test dataset is 70.35%. When analyzing the metrics for individual crops, only winter rye and potato have unacceptable values (F1 21.43 and 40.00% respectively).

The automatic classification score of both SIGMA and NDVI should be considered as poor based on OA (62.01% and 70.35%, respectively). However, it is important to note the varying accuracy in the classification of individual crops. In both classifications the lowest accuracy was obtained for winter rye and potatoes. In both cases, the explanation could be the small number of declarations: potatoes. In contrast, winter rye was misclassified for SIGMA despite not being a marginal class. Overall NDVI classification scored better than SIGMA except for winter rye and potatoes, and for maize and rape F1 was satisfactory (86.67% and 90.67% respectively). Even the accuracies for grass and other cereals were relatively high (67.44% and 70.73% respectively).

### 3.3. Results of Using “Traffic Lights”: the Example of Rapeseed

We used the idea of “traffic lights” to illustrate our classification results. The best classified crop was selected: rapeseed. Table 9 compares the metrics obtained by the four methods for rapeseed. The four cases discussed differ in area, number of plots analyzed, and method of classification.

In all cases, the values commonly used in ML, ACC (accuracy) and TNR (specificity), are very high at above 94% and should be considered unreliable because

they do not correspond to the values of metrics traditionally used in remote sensing (compare TPR/PA and PPV/UA columns). Ideally, both of these metrics should have high values. But regardless, there are still cases of FP and FN that do not necessarily reflect reality.

If we do not want to mark all FPs and FNs in red, or check all these cases visually, the following approach could be proposed. The most reliable classification result can be assumed due to, for example, the user accuracy PPV/UA. This implies a minimum commission error, i.e. if a plot is classified as a given crop then there is high probability that it is indeed that crop. This means that FP cases are really mis-declarations, i.e. they can be highlighted in red. FN cases, on the other hand, require visual interpretation and are highlighted in blue.

For the purpose of marking controlled plots according to the idea of “traffic lights”, the semi-automatic NDVI\_index method was chosen (results in the first row of in Table 9). This represents a compromise between the method for which the highest PPV/UA value was obtained and the maximum number of classified plots (NDVI\_index; PPV/UA = 88.1%, no of plots = 6,986, SIGMA\_CNN, PPV/UA = 88.9%, no of plots plots = 2,035).

**Table 9.** Comparison of classification accuracy for rapeseed

Classification's method	No of parcels	TP	TN	FP	FN	ACC	TPR	TNR	PPV	F1
NDVI_index	6,986	252	6,319	34	381	<b>0.941</b>	0.398	<b>0.995</b>	<b>0.881</b>	0.548
NDVI_SAM OA = 70.35%	199	34	158	6	1	<b>0.965</b>	0.971	<b>0.963</b>	<b>0.850</b>	0.907
SIGMA_index	4,165	152	4,143	96	22	<b>0.973</b>	0.874	<b>0.977</b>	<b>0.613</b>	0.720
SIGMA_CNN OA = 62.01%	2,035	96	1,902	12	25	<b>0.982</b>	0.793	<b>0.987</b>	<b>0.889</b>	0.838

The idea of “traffic lights” can be illustrated for rapeseed as the following:

- green – plots classified as rapeseed and are really rapeseed (TP),
- red – plots classified as non-rapeseed and declared as rapeseed (FP),
- blue – plots classified as oilseed rape and declared as another crop (FN).

An example application of the “traffic lights” idea can be found in the Internet [22] and also in Figure 6. On the color compositions (Figs. 7, 8), it is possible to correctly analyze verified plots (Fig. 6 – in green) and problematic plots (Fig. 6 – in red and blue). Compositions were created from S-2 images dated 21.05.2017. In the natural colors (Fig. 7 – channels 432), the rapeseed is light green, and in the false color composite (Fig. 8 – FCC channels: 843) it is pink.

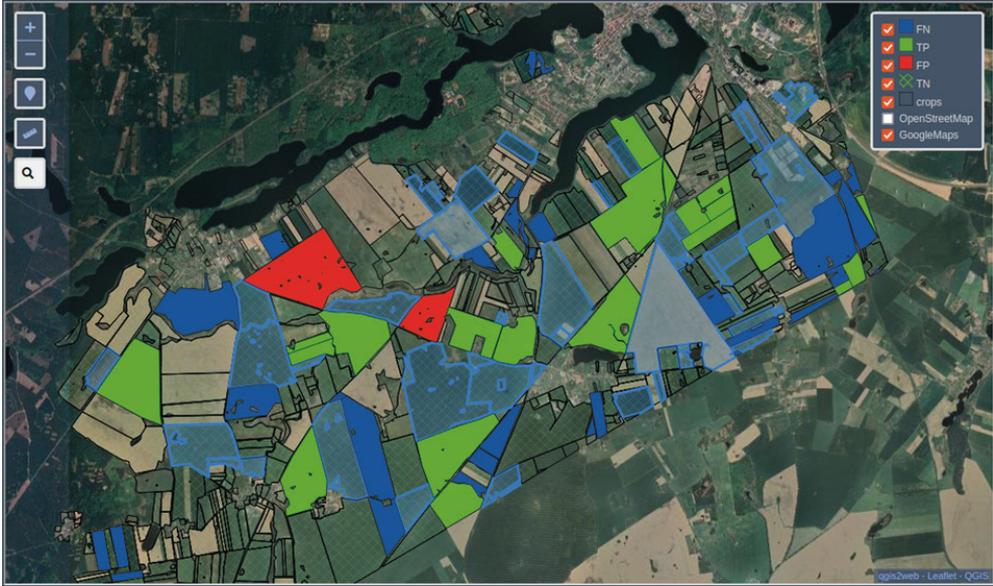


Fig. 7. Results of using "traffic lights": the example of rapeseed



Fig. 8. Composition in natural colors 21.05.2017 – rapeseed in green

Plots in blue (FN) can be easily accepted as rapeseed based on their color compositions. This represents the omission error of the semi-automatic classification based on NDVI index. In contrast, the plots in red (FP) actually on the S-2 compositions of the day are different from other plots covered with rapeseed.

In the conclusion of the presented example, it should be stated that such a detailed analysis is not necessary during a SAPS inspection. In this case, merely the confirmation of agricultural activity is sufficient for marking the parcel in green, regardless of whether there is maize or wheat on it.



**Fig. 9.** False Color Compost (FCC) 21.05.2017 – light red colored rapeseed

#### 4. Conclusion

In this paper we wanted to discuss the idea of “traffic lights” in IACS on a selected test area. Classification was done on NDVI and SIGMA time series calculated from the Sentinel-1 and Sentinel-2 satellite images registered in the season of autumn 2016 to autumn 2017. Classification was performed by automatic and semi-automatic methods.

Automatic classification of 6 crop groups resulted in the following accuracy (OA) on independent test fields:

- SIGMA CNN 62.01%,
- NDVI SAM 70.35%.

The accuracy of the validation on the training fields was:

- SIGMA CNN 98.50%,
- NDVI SAM 72.79%.

For comparison, it is possible to refer to the crop classification accuracy reported in the literature. It depends on a number of factors: the type of crop being classified, the climate zone, the design of accuracy analysis and metrics used. Indeed, metrics computed during validation, i.e., computed based on samples drawn from training set instead of learning-independent test set, are often reported as accuracy.

In this case, the accuracies are very high. A second factor that artificially inflates the accuracy is the provision of the average ACC value as OA.

However, we can cite the results of studies for which the accuracy analysis was performed in a manner similar to our study, with AO's obtained in Belgium of 82% [10], Australia 84.2% [11], South Africa 82% [12] or Poland 69% [7], 81% [15]. In this context, the classification accuracy presented in this paper is moderate, but consistent with similar studies on the verification of declarations in Poland [7]. In both cases, the time series of indices were classified. On the other hand, our later research on all Sentinel-2 channels allowed us to obtain a higher accuracy of 81% [15], similar to the above-cited results of foreign researchers.

The aim of this paper was to try to implement the idea of “traffic lights” regardless of classification accuracy. Finally, the best classified crop, rapeseed, was chosen and a semi-automated classification method with following accuracy metrics:

- SIGMA – PPV/UA 61.3%, TPR/PA 87.4%, F1 72.0%,
- NDVI– PPV/UA 88.1%, TPR/PA 39.8%, F1 54.8%.

The results of the analysis based on NDVI and the metric of PPV/UA 88.1% is available on the website [22] (an example in Figure 7). It should be mentioned that the example presented on this page is too detailed for SAPS control and more appropriate for voluntary coupled support (VSC).

Another issue discussed in the paper was accuracy metrics. In conclusion, it should be stated that the metrics of accuracy (ACC) and specificity / true negative rate (TNR) used in binary classification and popular in machine learning, should be considered unreliable for multi-class classification. In all cases, they reach very high values around 90% or higher and give an artificial impression of high classification accuracy. This is due to the high proportion of “true negative” (TN), which includes all other classes not correctly classified into a class.

Classification accuracy analysis has been of interest for many years [21, 23]. Although many researchers have proposed different accuracy indices, the traditional accuracy metrics of OA, PA and UA, are still considered as the most reliable in remote sensing [23].

However, nowadays, metrics automatically calculated in machine learning classification (sensitivity/specificity and accuracy) are increasingly reported in remote sensing. These metrics are designed to evaluate binary tests, e.g., the evaluation of medical tests with only positive-patient sick (TP), positive-patient healthy (FP), negative-patient sick (FN) and negative-patient healthy (TN). These metrics are inadequate for multi-class classification, and in particular, average accuracy (ACC) is not equated to overall accuracy (OA) [15]. The impact of ignored classes on the classification result can be seen in [24].

In conclusion, it can be said that, regardless of the automatic classification method and the accuracy achieved, one should consciously choose the appropriate accuracy metric to minimize the risk of error. After all, not all of the cases of FP and FN

are in fact a mismatch between the declaration and the actual crop. It also seems necessary to check some lights depending on the procedure adopted.

### Acknowledgements

The authors acknowledge the support of the Agency for Restructuring and Modernisation of Agriculture (ARMA) in Poland for funding 2 projects [13, 14] and providing data concerning farmer declarations.

This research was conducted as part of the AGH IDUB project: Integration of remote sensing data for control in the system of direct agricultural subsidies (IACS).

### Author Contribution

Author 1: conceptualization, methodology, Sentinel-1, Sentinel-2 acquisition and pre-processing, validation, writing – final version preparation.

Author 2: accuracy calculations, website preparation.

Author 3: literature review concerning IACS and “traffic lights”, website preparation, writing – original draft preparation.

## References

- [1] European Commission: *Integrated Administration and Control System (IACS)*. [https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/financing-cap/financial-assurance/managing-payments\\_en](https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/financing-cap/financial-assurance/managing-payments_en) [access: 3.08.2021].
- [2] Gov.pl: *IT SYSTEMS Integrated Administration and Control System (IACS)*. <https://www.gov.pl/web/arimr-en/it-systems> [access: 3.08.2021].
- [3] Asseco: *ARiMR – Administration and Control System (IACS)*. <https://pl.asseco.com/en/case-study/arimr-administration-and-control-system-iacs-32/> [access: 3.08.2021].
- [4] Devos W., Lemoine G., Milenov P., Fasbender D.: *Technical guidance on the decision to go for substitution of OTSC by monitoring*. Publications Office of the European Union, 2018. <http://dx.doi.org/10.2760/693101>.
- [5] Devos W., Lemoine G., Milenov P., Fasbender D., Loudjani P., Wirnhardt C., Sima A., Griffiths P.: *Second discussion document on the introduction of monitoring to substitute OTSC: rules for processing application in 2018–2019*. Publications Office of the European Union, 2018. <http://dx.doi.org/10.2760/344612>.
- [6] Devos W., Fasbender D., Lemoine G., Loudjani P., Milenov P., Wirnhardt C.: *Discussion document on the introduction of monitoring to substitute OTSC – Supporting non-paper DS/CDP/2017/03 revising R2017/809*. Publications Office of the European Union, 2017. <http://dx.doi.org/10.2760/258531>.
- [7] Musiał J., Bojanowski J.: *Assessing potential of the Sentinel-2 imagery for monitoring of agricultural fields in Poland*. 25th MARS Conference, Praga, 2019. [http://www.igik.edu.pl/upload/Poster\\_EOStat\\_final\(1\).pdf](http://www.igik.edu.pl/upload/Poster_EOStat_final(1).pdf) [access: 3.08.2021].

- 
- [8] Csillik O., Belgiu M., Asner G.P., Kelly M.: *Object-Based Time-Constrained Dynamic Time Warping Classification of Crops Using Sentinel-2*. Remote Sensing, vol. 11, no. 10, 2019, 1257. <https://doi.org/10.3390/rs11101257>.
- [9] Hütt C., Waldhoff G., Bareth G.: *Fusion of Sentinel-1 with official topographic and cadastral geodata for crop-type enriched LULC mapping using FOSS and open data*. ISPRS International Journal of Geo-Information, vol. 9, no. 2, 2020, 120. <https://doi.org/10.3390/ijgi9020120>.
- [10] Van Tricht K., Gobin A., Gilliams S., Piccard I.: *Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping. A case study for Belgium*. Remote Sensing, vol. 10, no. 10, 2018, 1642. <https://doi.org/10.3390/rs10101642>.
- [11] Brinkhoff J., Vardanega J., Robson A.: *Land cover classification of nine perennial crops using Sentinel-1 and -2 data*. Remote Sensing, vol. 12, no. 1, 2020, 96. <https://doi.org/10.3390/rs12010096>.
- [12] Maponya M., van Niekerk A., Mashimbye Z.: *Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning*. Computers and Electronics in Agriculture, vol. 169, 2020, 105164. <https://doi.org/10.1016/j.compag.2019.105164>.
- [13] Hejmanowska B., Mikrut S., Głowienka E., Michałowska K., Kramarczyk P., Pirowski T.: *Expertise on the use of Sentinel 1 and 2 images to monitor the agricultural activity of ARIMR beneficiaries*. 2018. [http://home.agh.edu.pl/~galia/img/Raport\\_ARIMR\\_AGH\\_2018\\_EN\\_haslo.pdf](http://home.agh.edu.pl/~galia/img/Raport_ARIMR_AGH_2018_EN_haslo.pdf).
- [14] Hejmanowska B., Mikrut S., Głowienka E., Kramarczyk P., Pirowski T.: *The use of hyperspectral data to monitor the agricultural activity of the ARMA beneficiaries and support its business processes*. 2019. [http://home.agh.edu.pl/~galia/img/Raport\\_ARIMR\\_AGH\\_2019\\_EN\\_haslo.pdf](http://home.agh.edu.pl/~galia/img/Raport_ARIMR_AGH_2019_EN_haslo.pdf).
- [15] Hejmanowska B., Kramarczyk P., Głowienka E., Mikrut S.: *Reliable Crops Classification Using Limited Number of Sentinel-2 and Sentinel-1 Images*. Remote Sensing, vol. 13, no. 16, 2021, 3176. <https://doi.org/10.3390/rs13163176>.
- [16] Rouse J., Haas R., Schell J., Deering D., Harlan J.: *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*. NASA/GSFC Type III Final Report, Greenbelt, MD, 1974.
- [17] Laur H.: *ERS-1 SAR Calibration: Derivation of Backscattering Coefficient  $\sigma^0$  in ERS-1 SAR PRI Products*. ESA/ESRIN, Issue 1, Rev. 0, October 1992.
- [18] Filipponi F.: *Sentinel-1 GRD Preprocessing Workflow*. Proceedings, vol. 18, no. 1, 2019, 11. <https://doi.org/10.3390/ECRS-3-06201>.
- [19] Kussul N., Lavreniuk M., Skakun S., Shelestov A.: *Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data*. IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, 2017, pp. 778–782. <https://doi.org/10.1109/LGRS.2017.2681128>.
- [20] Fawcett T.: *An introduction to ROC analysis*. Pattern Recognition Letters, vol. 27, iss. 8, 2006, pp. 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>.

- 
- [21] Congalton R.G.: *A review of assessing the accuracy of classifications of remotely sensed data*. Remote Sensing of Environment, vol. 37, iss. 1, 1991, pp. 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B).
- [22] Zespół Geoinformacji, Fotogrametrii i Teledetekcji Środowiska: *Application of the “traffic lights” idea in crops control in Integrated Administration Control System*. <https://twiki.fotogrametria.agh.edu.pl/c5www/index.php/katedra/gll2021> [access: 3.08.2021].
- [23] Morales-Barquero L., Lyons M.B., Phinn S.R., Roelfsema C.M.: *Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources*. Remote Sensing, vol. 11, no. 19, 2019, 2305. <https://doi.org/10.3390/rs11192305>.
- [24] Foody G.M.: *Impacts of ignorance on the accuracy of image classification and thematic mapping*. Remote Sensing of Environment, vol. 259, 2021, 112367. <https://doi.org/10.1016/j.rse.2021.112367>.