



**AKADEMIA GÓRNICZO-HUTNICZA  
IM. STANISŁAWA STASZICA W KRAKOWIE**

**AGH UNIVERSITY OF SCIENCE  
AND TECHNOLOGY**

# Dokładność klasyfikacji w uczeniu maszynowym na przykładzie klasyfikacji pokrycia terenu na terenach górniczych

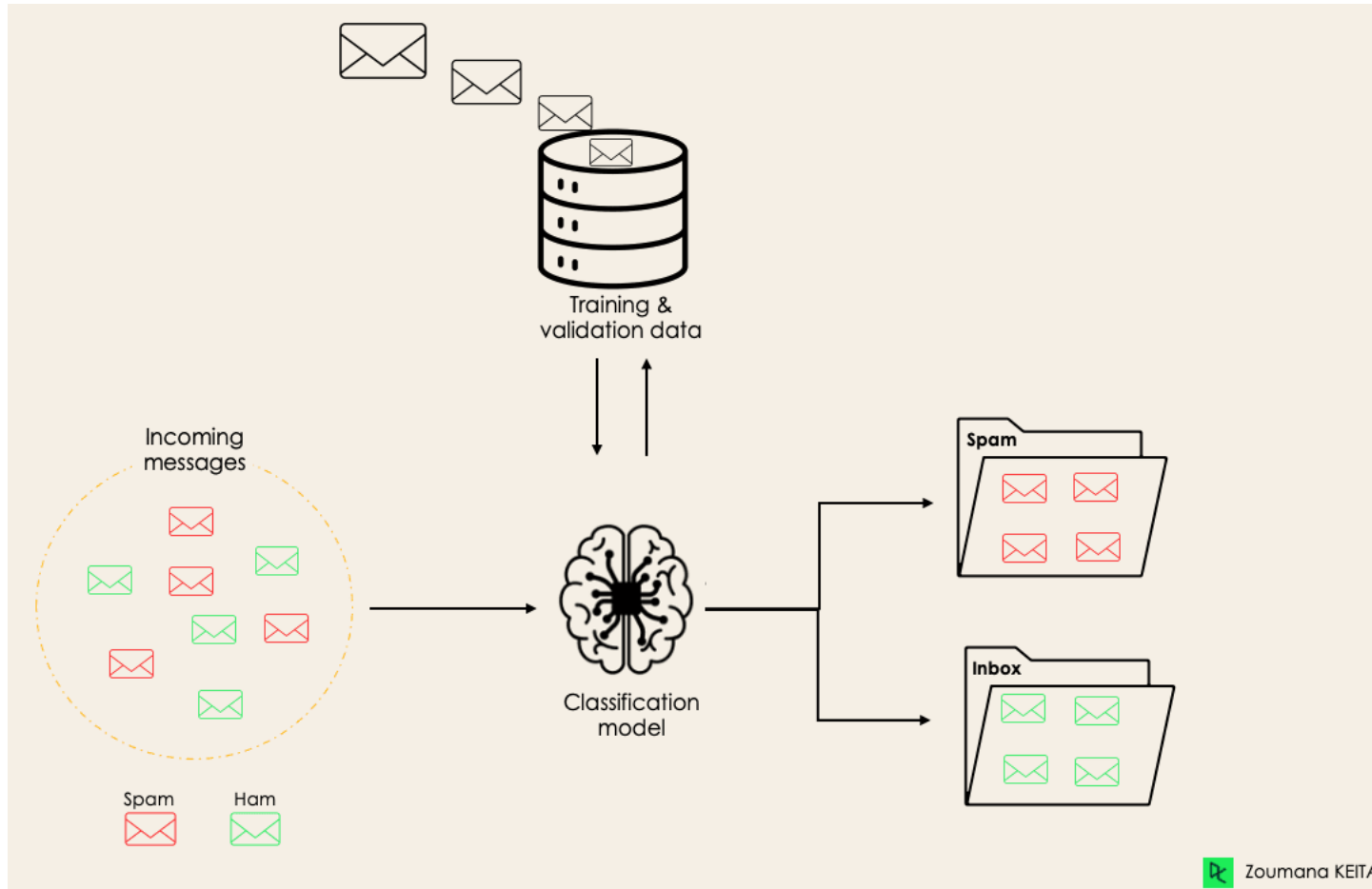
Beata Hejmanowska

Katedra Fotogrametrii Teledetekcji Środowiska i Inżynierii Przestrzennej  
Wydział Geodezji Górniczej i Inżynierii Środowiska

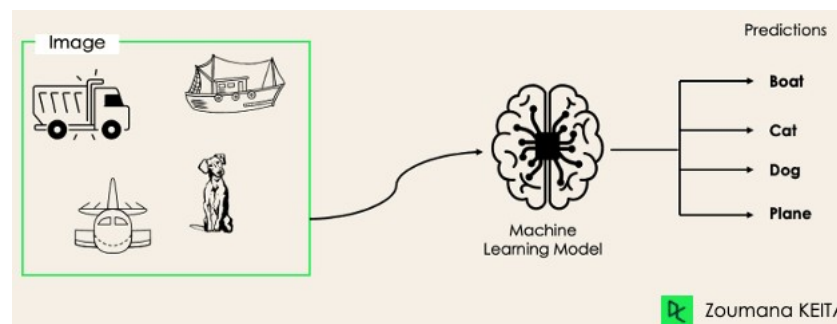
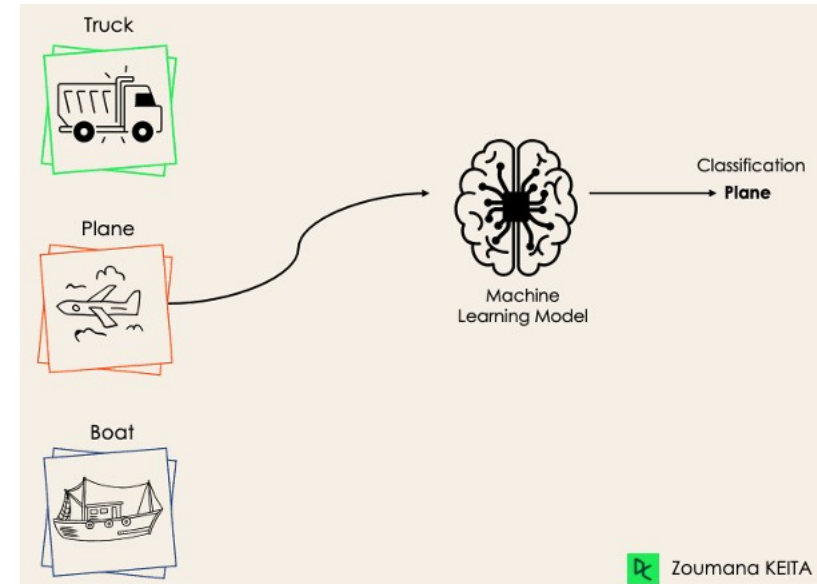
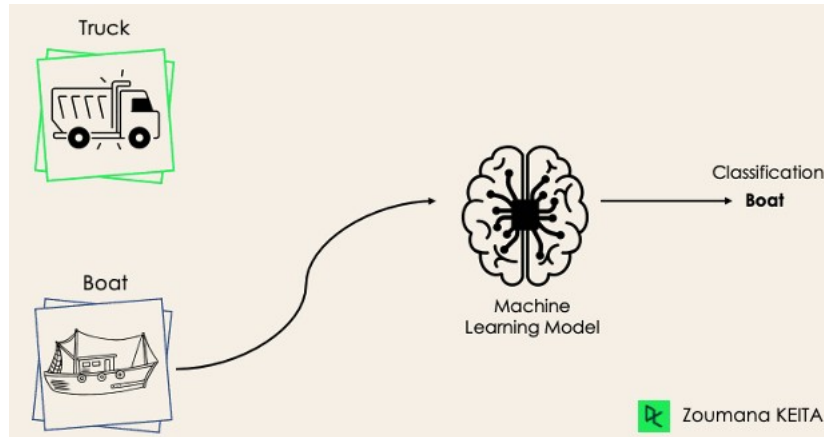


**Szkoła Eksploatacji  
Podziemnej**

# Problem klasyfikacji



# Klasyfikacja binarna i wielu klas



# Ocena dokładności klasyfikacji

## A bit of context

Imagine you are a healthcare startup, and want an AI assistant able to predict whether a given patient has a heart disease or not based on its health record. This is a binary classification problem where the model will predict

- 1, True or Yes if the patient has heart disease
- 0, False or No otherwise

## 1 Confusion matrix

A 2X2 matrix that nicely summarizes the number of correct predictions of the model. It also helps in computing different other performance metrics.

Predicti Reality	Yes	No
Yes	True Positives (TP)	False Negatives (FN)
No	False Positives (FP)	True Negatives (TN)

Type I Error

Type II Error

Type I & II Errors can be used interchangeably when referring to False Positives and False negatives respectively

# Problem klasyfikacji – klasyfikacja binarna

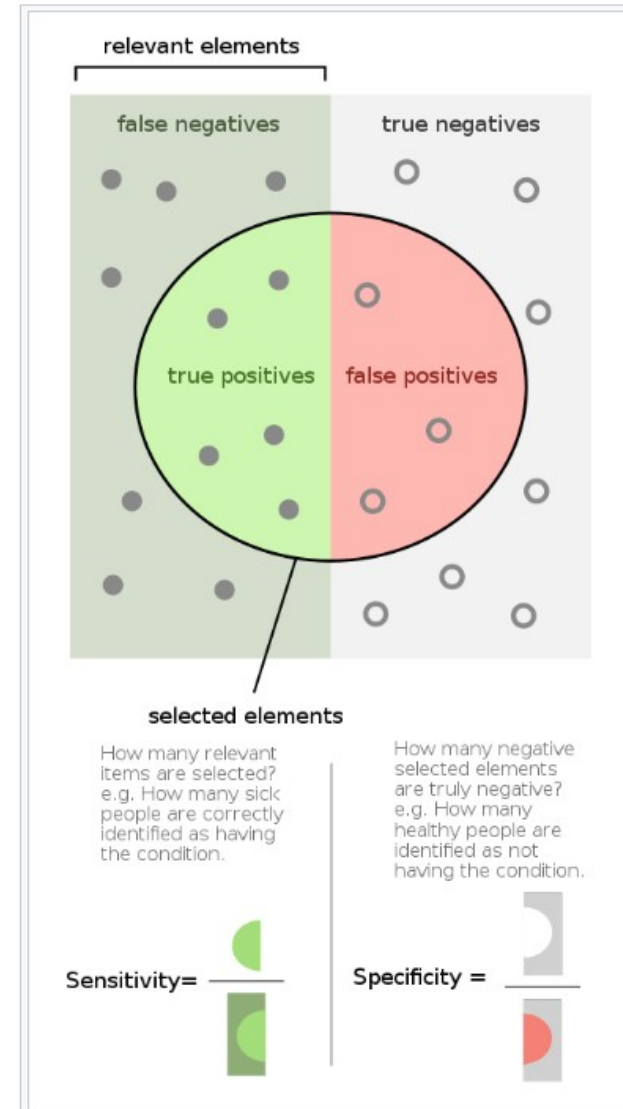
**Czułość testu** to stosunek wyników prawdziwie dodatnich do sumy prawdziwie dodatnich i fałszywie ujemnych. Czułość 100% w przypadku testu medycznego oznaczałaby, że wszystkie osoby chore zostaną rozpoznane. Pojęcie interpretuje się jako zdolność testu do prawidłowego rozpoznania choroby tam, gdzie ona występuje.

**Swoistość testu** to stosunek wyników prawdziwie ujemnych do sumy prawdziwie ujemnych i fałszywie dodatnich. Swoistość 100% oznaczałaby, że wszyscy ludzie zdrowi w wykonanym teście diagnostycznym zostaną oznaczeni jako zdrowi. Test o wysokiej swoistości cechuje niski błąd pierwszego rodzaju.

## Sensitivity and specificity

Tabela krzyżowa testu diagnostycznego:

		Stan (na przykład choroba)		
		prawdziwy	fałszywy	
Wynik testu	dodatni	prawdziwie dodatni	fałszywie dodatni	→ wartość predykcyjna dodatnia
	ujemny	fałszywie ujemny	prawdziwie ujemny	→ wartość predykcyjna ujemna
		↓ czułość	↓ swoistość	




Sensitivity and specificity - The left half of the image with the solid dots represents individuals who have the condition, while the right half of the image with the hollow dots represents individuals who do not have the condition. The circle represents all individuals who tested positive.

# ML metryki dokładności

## 2 Accuracy

We get accuracy by answering this question: "out of the predictions made by the model, what percentage is correct?"

$$\text{Accuracy} = \frac{TP + TN}{\text{Total number observation}}$$

 Zoumana KEITA

## 3 Precision

We get precision by answering this question: "out of all the YES predictions, how many of them were correct?"

$$\text{Precision} = \frac{TP}{TP + FP}$$

## 4 Recall / Sensitivity

It aims to answer this question: "how good was the model at predicting real Yes events?", which can be considered as the flip of the precision.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

## 5 Recall / Specificity

It aims to answer this question: "how good was the model at predicting real No events?".

$$\text{Specificity} = \frac{TN}{TN + FP}$$

## 6 F1 Score

Sometimes used when dealing with imbalanced data set, meaning that there are more of one class/label than there are of the other. It corresponds to the harmonic mean of the precision and recall.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



# Problem klasyfikacji

## 7 AUC – ROC Curve

AUC– ROC generates probability values instead of binary 0/1 values. It should be used when your data set is roughly balanced.

Using ROC for imbalanced data sets lead to incorrect interpretation.

ROC curves provide good overview of trade-off between the TP rate and FP rate for binary classifier using different probability thresholds.

- A value below 0.5 indicates a poor classifier
- A value of 0.5 means random classifier
- Value over 0.7 corresponds to a good classifier
- 0.8 indicates a strong classifier
- We have 1 when the classifier perfectly predicts everything.

### Strategies to choose the right metric

#### Choose accuracy

- The cost of FP and FN are roughly equal.
- The benefit of TP and TN are roughly equal.

#### Choose Precision

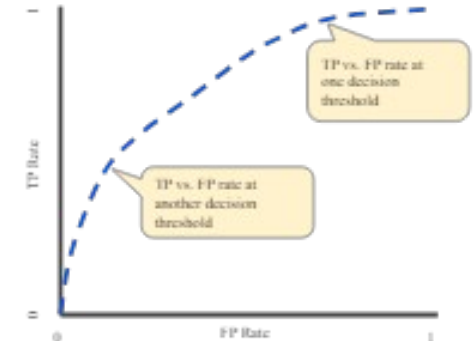
- The cost of FP is much higher than a FN.
- The benefit of a TP is much higher than a TN.

#### Choose recall

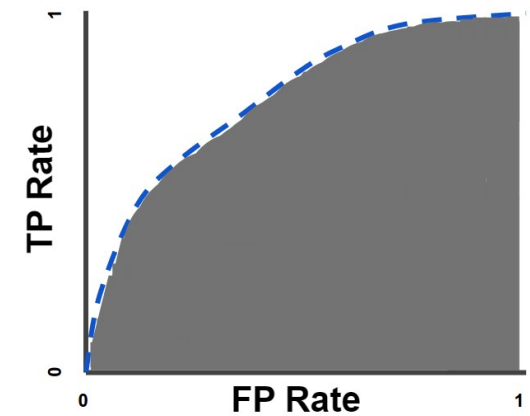
- The cost of FN is much higher than a FP.
- The cost of a TN is much higher than a TP.

### ROC AUC & Precision – Recall curves

- Use ROC when the dealing with balanced data sets.
- Use precision-recall for imbalanced data sets.



ROC receiver operating characteristic curve



AUC area under the curve

Współczynnik dodatni (TPR) jest synonimem rozpoznawalności, więc został określony w ten sposób:

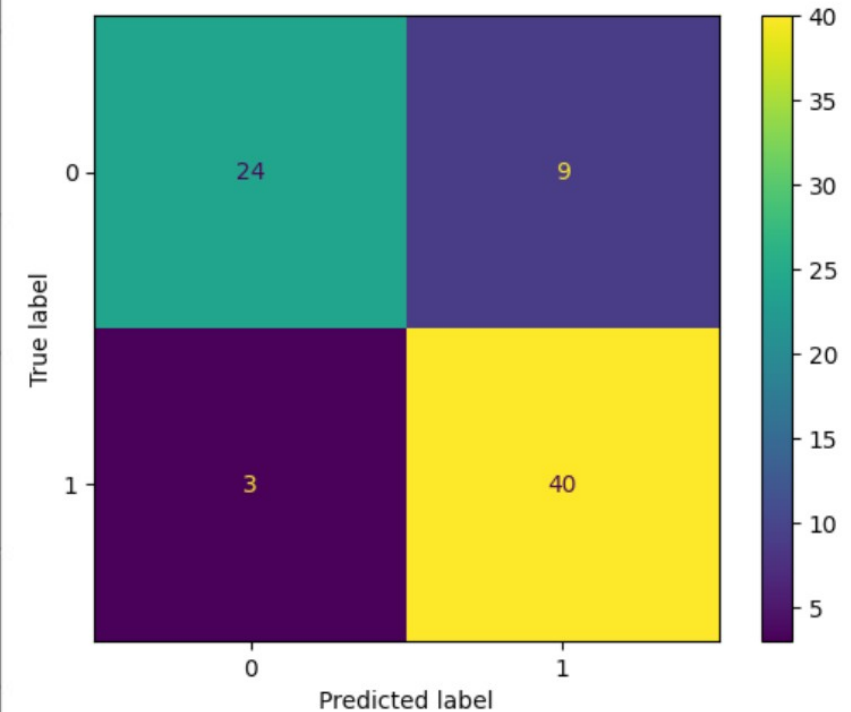
$$TPR = \frac{TP}{TP + FN}$$

Współczynnik dodatni (FPR) jest zdefiniowany w ten sposób:

$$FPR = \frac{FP}{FP + TN}$$

# Metryki dokładności

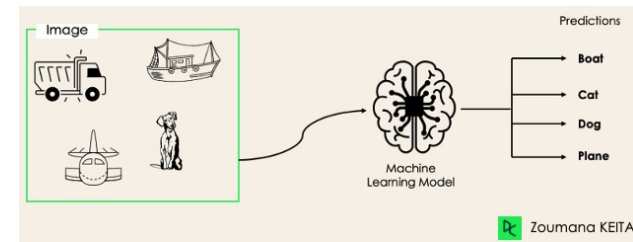
Name	Formuła
Producer accuracy (PA) Sensitivity True positive rate (TPR)	$\frac{TP}{TP + FN}$
Specificity True negative rate (TNR)	$\frac{TN}{TN + FP}$
User accuracy (UA) Precision Positive predictive value (PPV)	$\frac{TP}{TP + FP}$
Accuracy (ACC)	$\frac{TP + TN}{TP + TN + FP + FN}$
F1 score	$\frac{2TP}{2TP + FP + FN}$
Overall accuracy (OA) Percent of correct precision	$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + TN_i + FP_i + FN_i)}$



ACC=OA=0.84



# Macierz błędów



Predicted/True	C1	C2	C3	C4	C5	C6	PA
C1	87	3	0	0	7	0	0.8969
C2	3	90	2	1	0	0	0.9375
C3	0	6	45	0	0	0	0.8823
C4	0	1	0	29	3	0	0.8787
C5	0	2	0	0	23	1	0.8846
C6	0	0	0	1	2	14	0.8235
UA	0.9667	0.8824	0.9574	0.9355	0.6571	0.9333	

Binary confusion matrix for class C1.  $OA=ACC=(87+220)/(87+3+10+220)=0.9594$

TP	FP	87	3
FN	TN	10	220

Binary confusion matrix for all classes, with metrics.

Class	TP	TN	FP	FN	<i>tpr</i>	<i>tn</i>	<i>acc</i>	<i>ppv</i>	<i>f1</i>
C1	87	220	3	10	0.8969	0.9865	0.9594	0.9667	0.9305
C2	90	212	12	6	0.9375	0.9464	0.9438	0.8824	0.9091
C3	45	267	2	6	0.8824	0.9926	0.975	0.9574	0.9184
C4	29	285	2	4	0.8788	0.993	0.9813	0.9355	0.9063
C5	23	282	12	3	0.8846	0.9592	0.9531	0.6571	0.7541
C6	14	302	1	3	0.8235	0.9967	0.9875	0.9333	0.875
average					0.8839	0.9791	0.9667	0.8887	0.8822

**OA=0.9000**

**ACC średnie=0.9667**

Name	Formula
Producer accuracy (PA) Sensitivity True positive rate ( <i>TPR</i> )	$\frac{TP}{TP+FN}$
Specificity True negative rate ( <i>TNR</i> )	$\frac{TN}{TN+FP}$
User accuracy (UA) Precision Positive predictive value ( <i>PPV</i> )	$\frac{TP}{TP+FP}$
Accuracy ( <i>ACC</i> )	$\frac{TP+TN}{TP+TN+FP+FN}$
F1 score	$\frac{2TP}{2TP+FP+FN}$
Overall accuracy (OA) Percent of correct precision	$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i+TN_i+FP_i+FN_i)}$

# Klasyfikacja pokrycia terenu na terenach górniczych



04.06.2023



09.07.2023



18.08.2023

- 1) Las iglasty
- 2) Las mieszany
- 3) Zabudowa
- 4) **Uprawy: zboża przed żniwami lub wiosną przed wykonaniem zabiegów agrotechnicznych**
- 5) **Gleba odkryta**
- 6) Trwałe użytki zielone
- 7) Drogi
- 8) Woda
- 9) **Uprawy w fazie wegetacji**
- 10) **Obszary przemysłowe, kamieniołomy, odkrywki**

## Automatyczna klasyfikacja obrazu

- Zbadanie możliwości wytrenowania **sieci neuronowej** do klasyfikacji wybranych typów pokrycia terenu **niezależnie od daty rejestracji i w dowolnym miejscu w kraju bez dodatkowych pól treningowych.**
- Wybrany został model sieci Unet, który zaimplementowano we framework-u pytorch, z wykorzystaniem języka programowania Python.

- 1) Las iglasty
- 2) Las mieszany
- 3) Zabudowa
- 4) **Uprawy: zboża przed żniwami lub wiosną przed wykonaniem zabiegów agrotechnicznych**
- 5) **Gleba odkryta**
- 6) Trwałe użytki zielone
- 7) Drogi
- 8) Woda
- 9) **Uprawy w fazie wegetacji**
- 10) **Obszary przemysłowe, kamieniołomy, odkrywki**

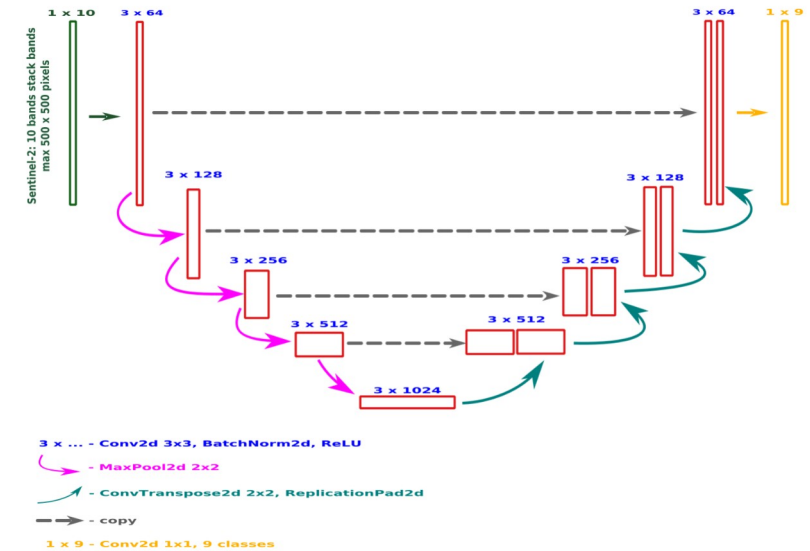
# Trenowanie sieci neuronowej

## Dane treningowe i walidacyjne:

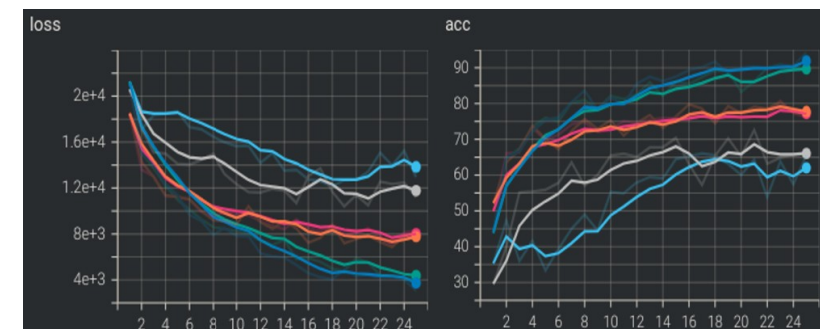
- Strzegom\_20210619, 20220619, 20220719, 20221012, 20230209, 20230301
- Kolbuszowa\_20210327, 20210411, 20210509, 20210728, 20210906
- Krakow\_20220603, 20220603, 20220603

## Dane testowe:

- strzegom\_20230709
- kolbuszowa\_20210725



Schemat sieci UNet

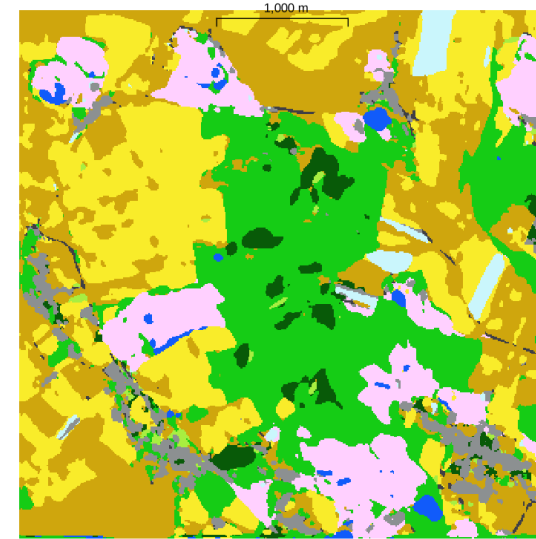
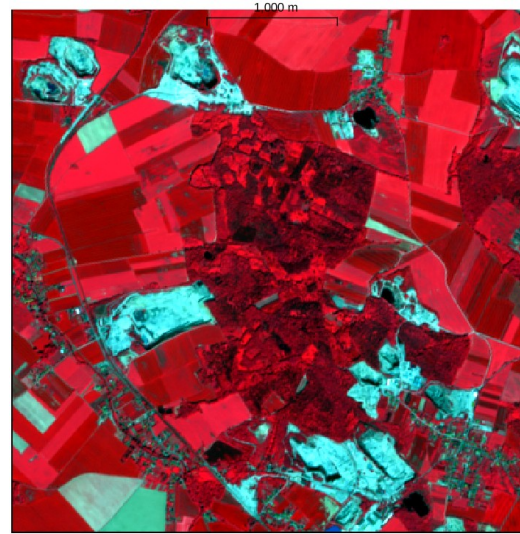


Przykładowe wykresy wartości funkcji straty i dokładności

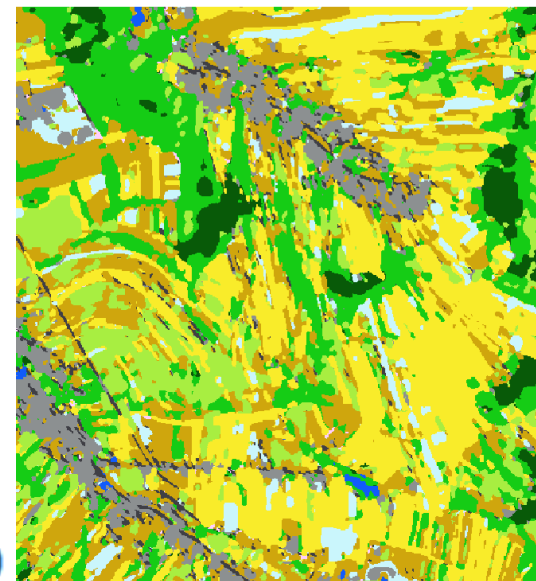
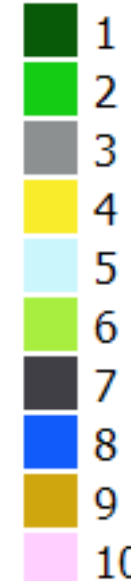
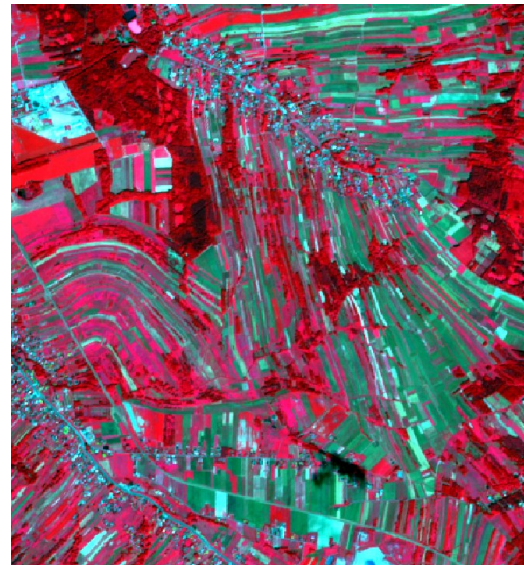
**Korelacja....**  
**Podejście pikselowe, obiektowe**



# Obszar testowy

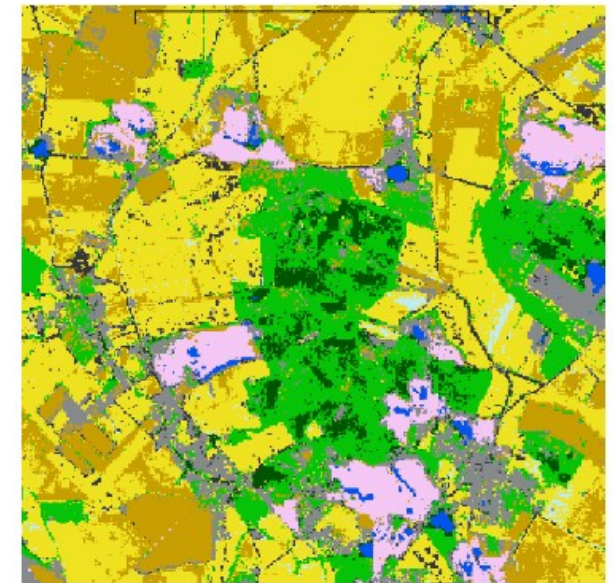
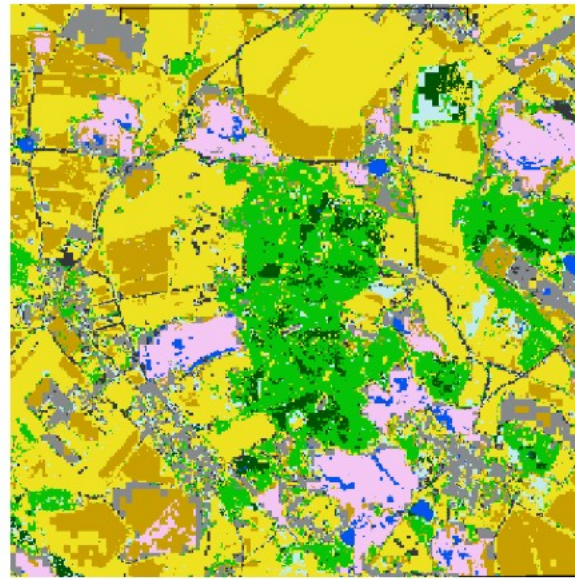
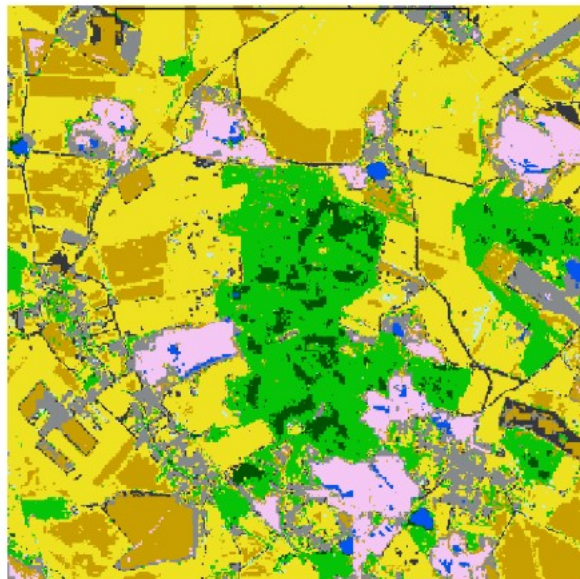
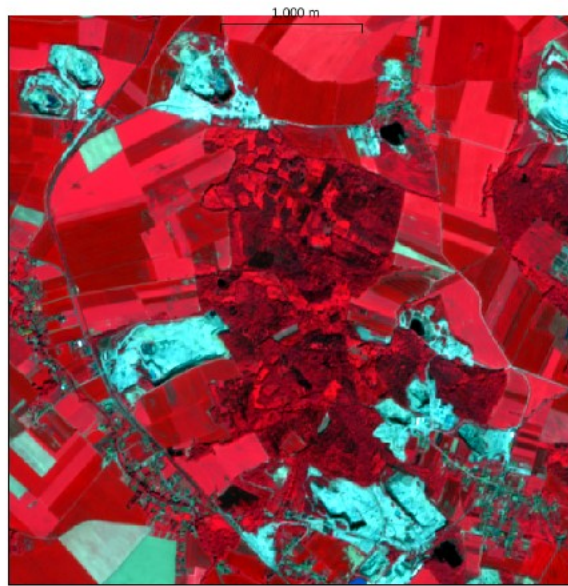


FCC - strzegom\_20230709



FCC - kolbuszowa\_20210725





# Metryki dokładności

## Metryki dokładności UNet

	TPR/PA	TNR	PPV/UA	ACC	F1
1	83.9	99.31	82.87	<b>98.92</b>	86.97
2	84.52	98.57	89.26	<b>97.56</b>	85.30

### Porównanie metryk dokładności

1 – Strzegom, pixel, 2 – Strzegom object

	pixel	object
strzegom_20230709	<b>94.6</b>	<b>89.0</b>

OA overall accuracy

Dokładność klasyfikacji Unet (PA, UA, F1):

- klasa 10 (obszar kopalni) wynosi średnio 99,01%.
- niezarośnięte obszary (gołe gleby) wynoszą 92,3%.
- klasy 4 i 9 (uprawy roczne) wynoszą średnio 94,8%.

## Metryki dokładności CART 2023-01-01;2023-07-30

Klasa	TPR/PA	TNR	PPV/UA	ACC	F1
1	80.00	97.14	57.14	97.78	66.66
2	68.57	92.86	72.73	96.33	70.59
3	98.57	90.71	73.40	88.10	84.14
4	51.43	82.86	72.00	93.33	60.00
5	0.00	92.14	0.00	93.82	0.00
6	0.00	100.00	0.00	100.00	0.00
7	50.00	95.00	71.43	98.46	58.82
8	80.00	98.57	100.00	100.00	88.89
9	40.00	88.21	44.44	94.00	42.10
10	75.00	98.21	100.00	100.00	85.71
średnia	<b>54.36</b>	<b>93.57</b>	<b>59.11</b>	<b>96.18</b>	<b>55.69</b>

**METRYKA "ACCURACY" W ANALIZIE KLASYFIKACJI WIĘKSZEJ LICZBY KLAS NIŻ DWIE JEST NIEWIARYGODNA - ZAWSZE PRZYJMUJE BARDZO WYSOKIE WARTOŚCI**

## Wnioski

Wydaje się, że można wytrenować sieć UNet do klasyfikacji dowolnego obrazu wielospektralnego na poziomie 90% **bez pól treningowych w dowolnym miejscu w kraju**

Nie należy używać do oceny dokładności metryki **“accuracy”** zamiast **“overall accuracy”** **ryzyko przeszacowania dokładności 10% lub więcej**