

Britt van Veggel, December 2020

Keurmerk herkenning en detectie

Questionmark Foundation ism Vantage AI



Agenda

Waar gaan we het allemaal over hebben?

1. Doel van de stage
2. Backstory neurale netwerken
3. Verloop process
4. Verheldering termen
5. Resultaten
6. Persoonlijke ervaring

Doel van de stage

Vanuit TU Delft

- Ervaring real life **toepassingen** van wiskunde;
- Bewust wording maatschappelijke **relevantie** van wiskunde;
- Ervaring binnen een **bedrijf**;
- Ervaring communiceren over wiskunde met **niet-wiskundigen**.
(Dat is vandaag)

Vanuit Questionmark

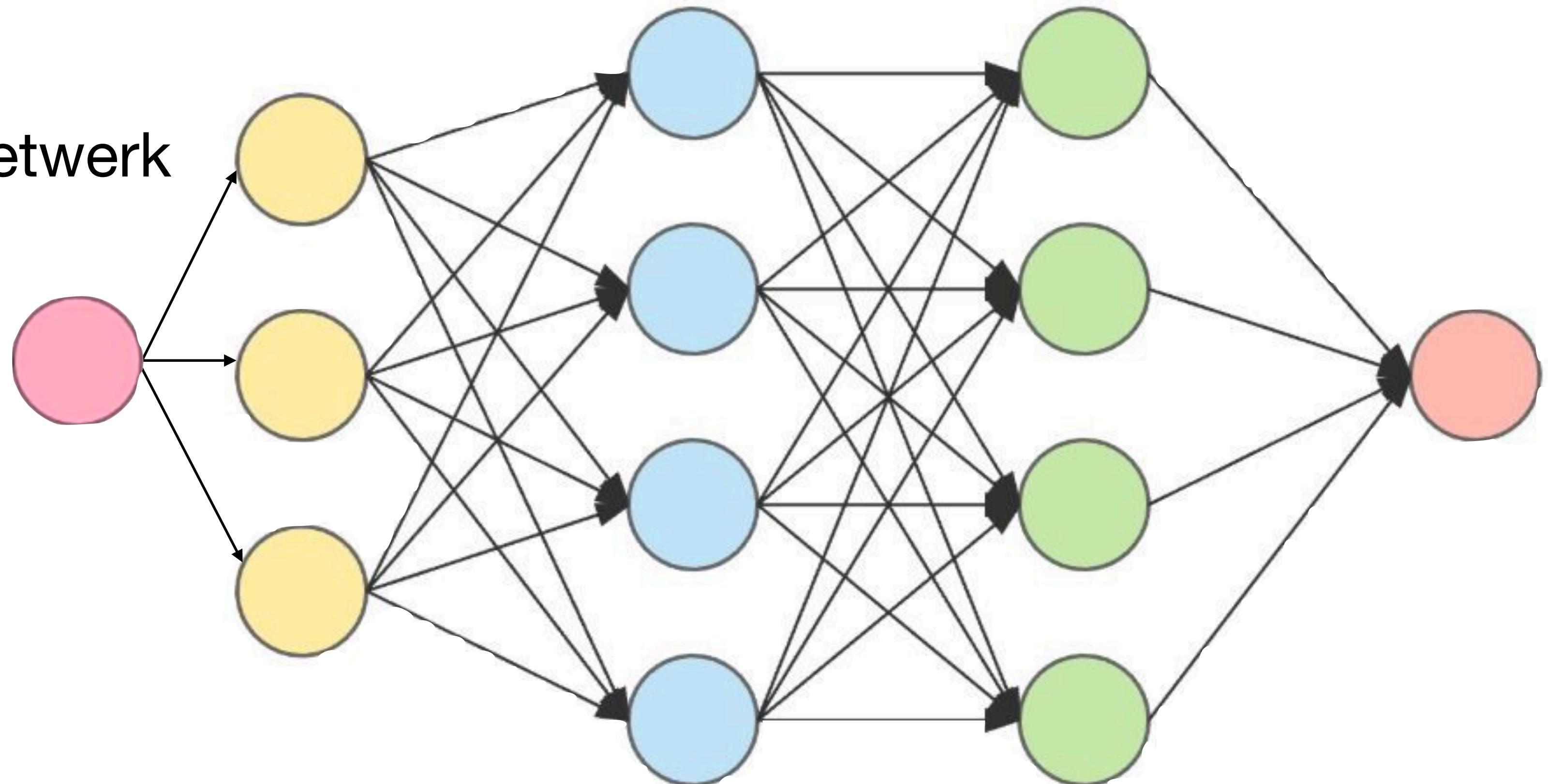
- Bouw software die **keurmerken** kan herkennen en detecteren op webshop foto's om zo de datakwaliteit te verbeteren



Neurale netwerken

Wat zijn dat nou precies?

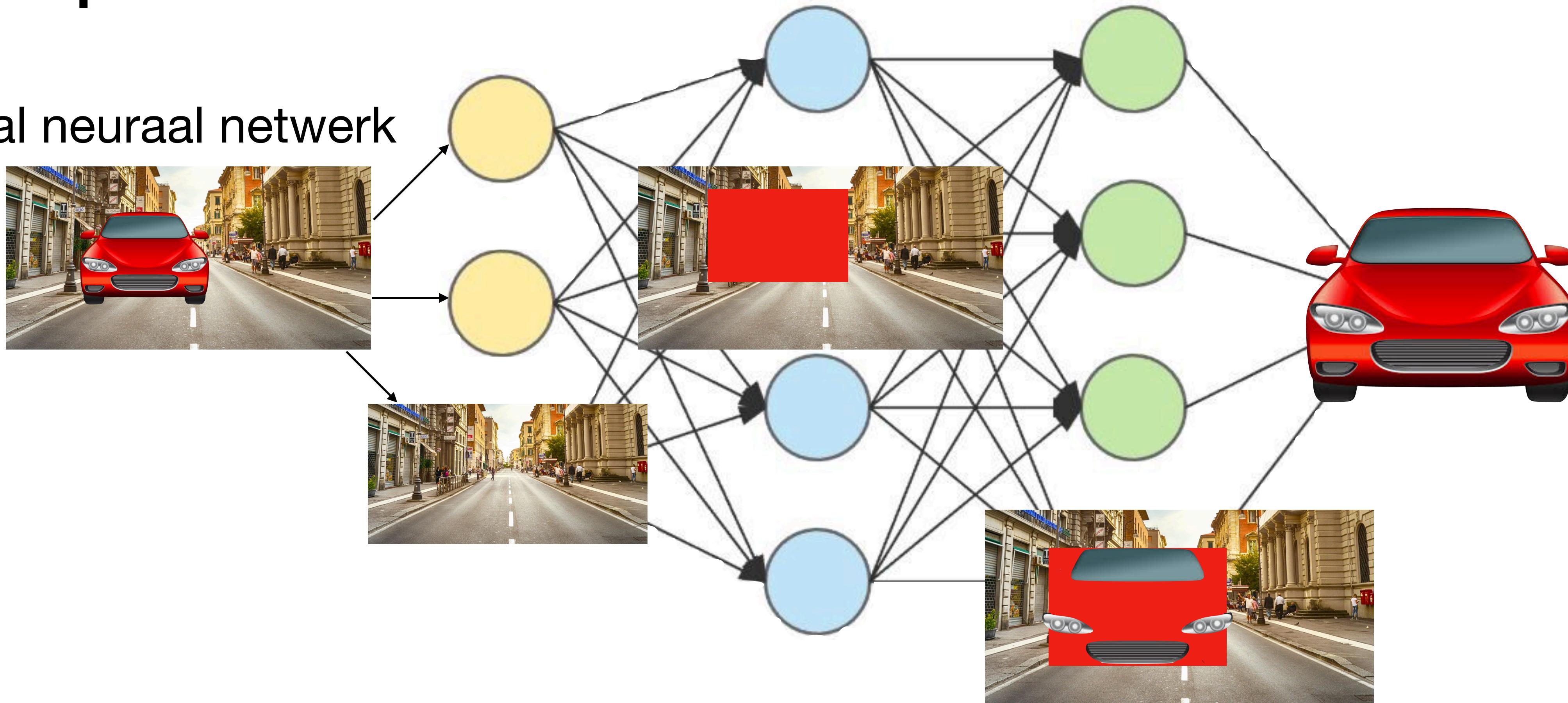
- Convolutional neuraal netwerk
- Trainen
- Data split
 - > Train
 - > Validation
 - > Test
- Evalueren



Neurale netwerken

Wat is dat nou precies?

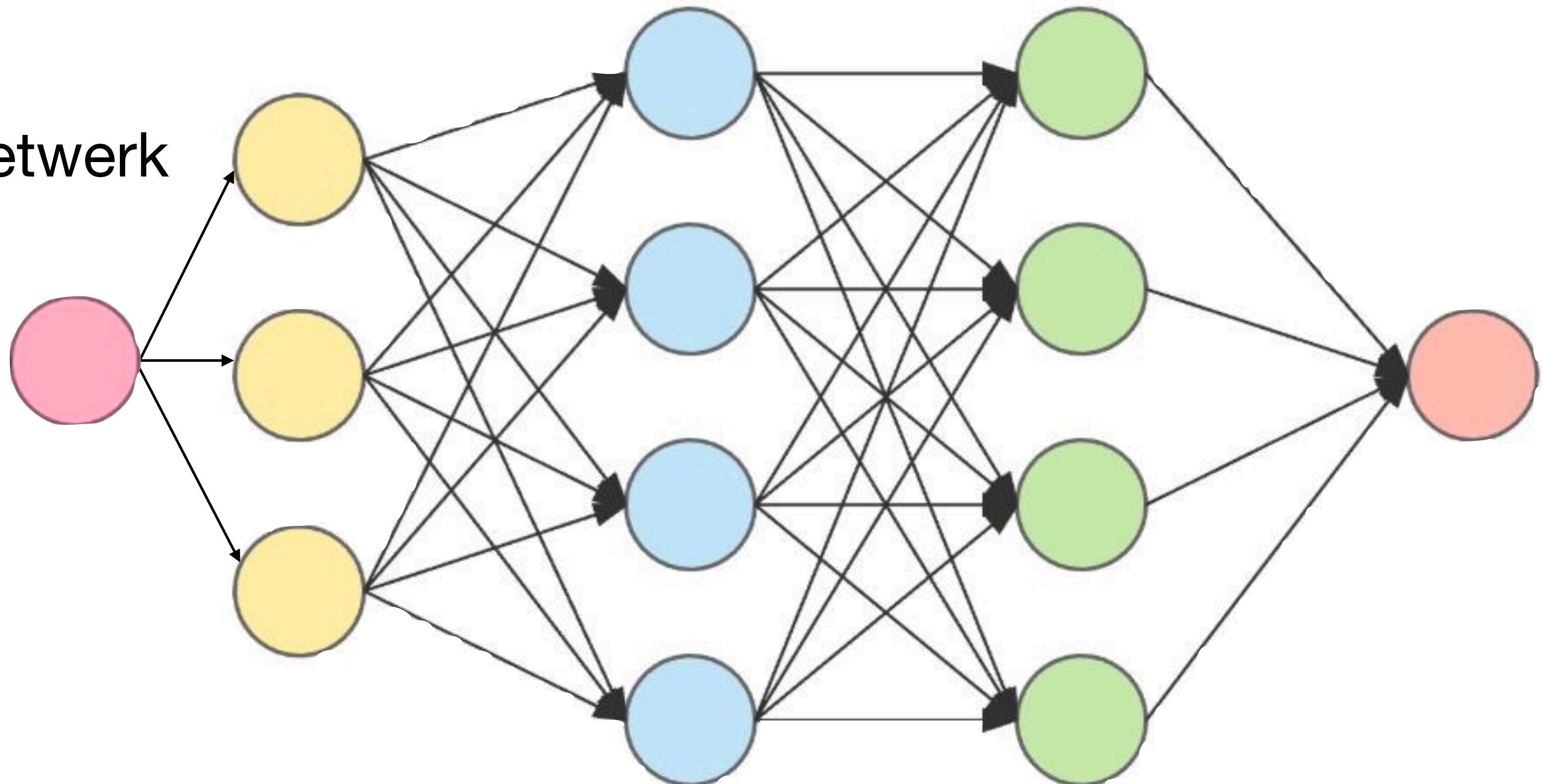
- Convolutional neuraal netwerk
- Trainen
- Data split
 - > Train
 - > Validation
 - > Test
- Evalueren



Neurale netwerken

Wat is dat nou precies?

- Convolutional neuraal netwerk
- Trainen
- Data split
 - > Train
 - > Validation
 - > Test
- Evalueren



Verloop process

Getting started

1. Tutorials, tutorials, tutorials
 - > Coursera
 - > Blogs & videos van Chiel
 - > Oefenopdrachten over bloemen, honden en krabbetjes
2. Wat voor dataset heeft Questionmark precies?
 - > Processing
3. ImageAI script van Xomnia
 - > Deze aan de praat krijgen
 - > Visualisatie maken van de data



Verloop process

Resultaten en conclusies eerste run

- Erg slechte resultaten
- Dataset niet passend bij beoogde doel
 - > Webshop:



> Dataset:



- Fouten in dataset (bio en ebio door elkaar gehusseld)



- Vreemd verdeeld (bijv: geen asc in train set, 2 asc in validation)

Aantal plaatjes: 1057
mAP: 0.24

asc : 0.0033
beterleven : 0.0000
bio : 0.8000
ebio : 0.0000
eko : 0.1837
fairtrade : 0.5032
msc : 0.0000
organic : 0.2956
planetproof : 0.5851
utz : 0.2292
vegan : 0.2857
weidemelk : 0.0000
mAP: 0.2405

Verloop process

Nieuwe data verkrijgen & prestatie beter meten

- Overzicht maken van data split
- Nieuwe data verkrijgen
 - > Op basis van webshop foto's uit QM Backend
 - > Tussen de 60 en 100 plaatjes per label
- Evaluation tools maken / zoeken
- En trainen maar!

Verloop process

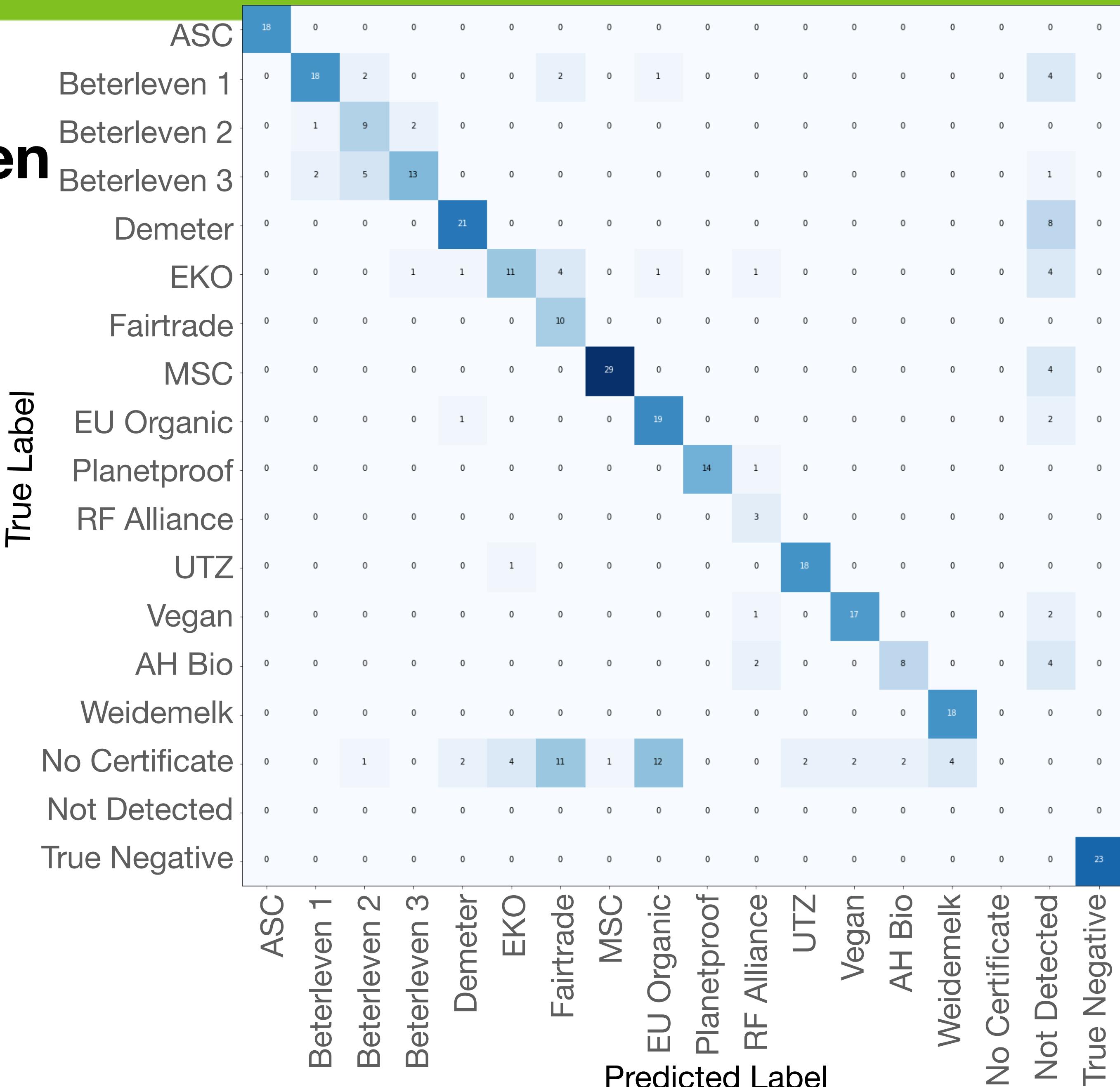
Nieuwe data & eerste resultaten

certificate	Alles	64% train	20% test	16% validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187

Verloop process

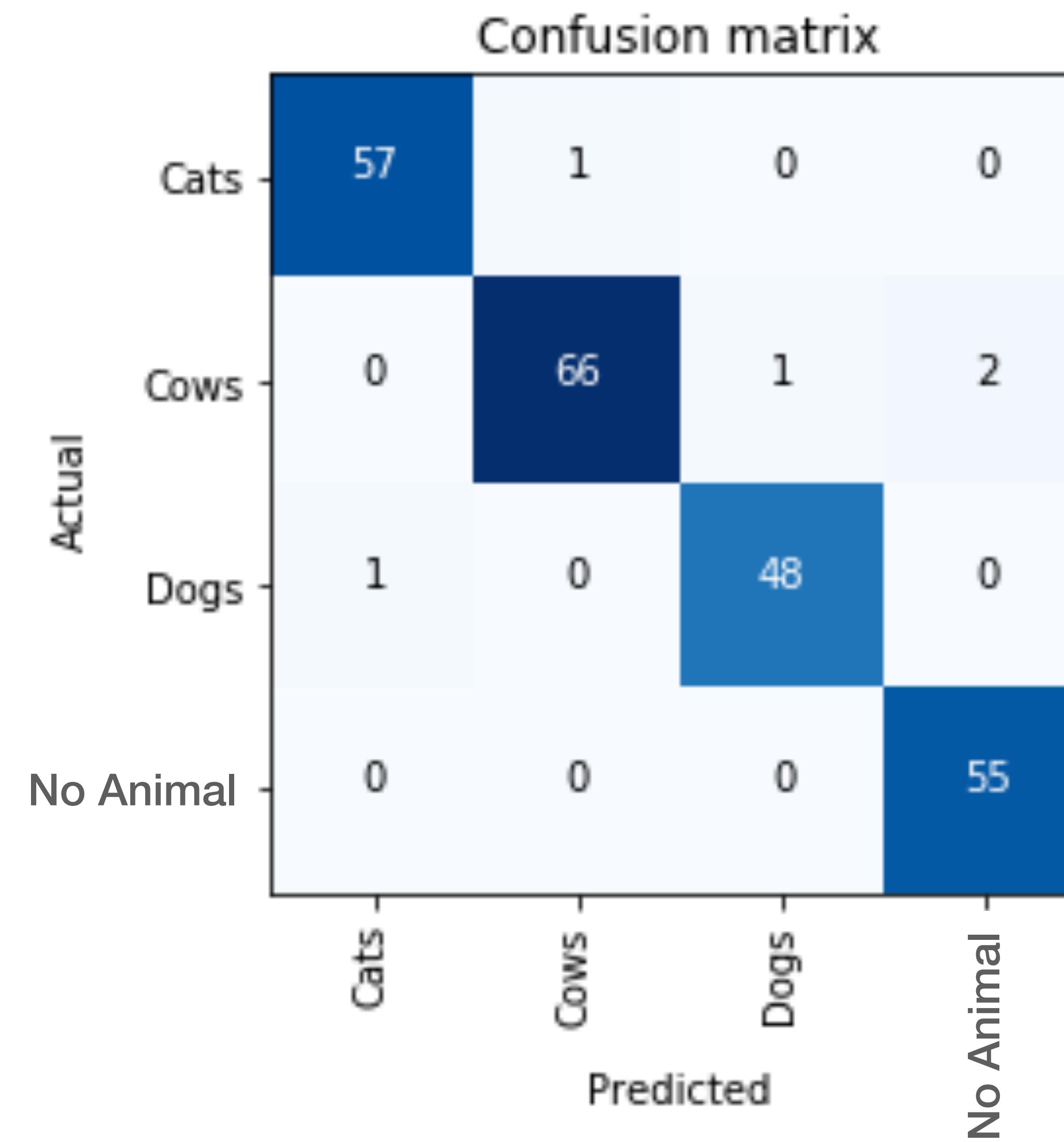
Nieuwe data & eerste resultaten

		64%	20%	16%
certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187



Verheldering van wat termen

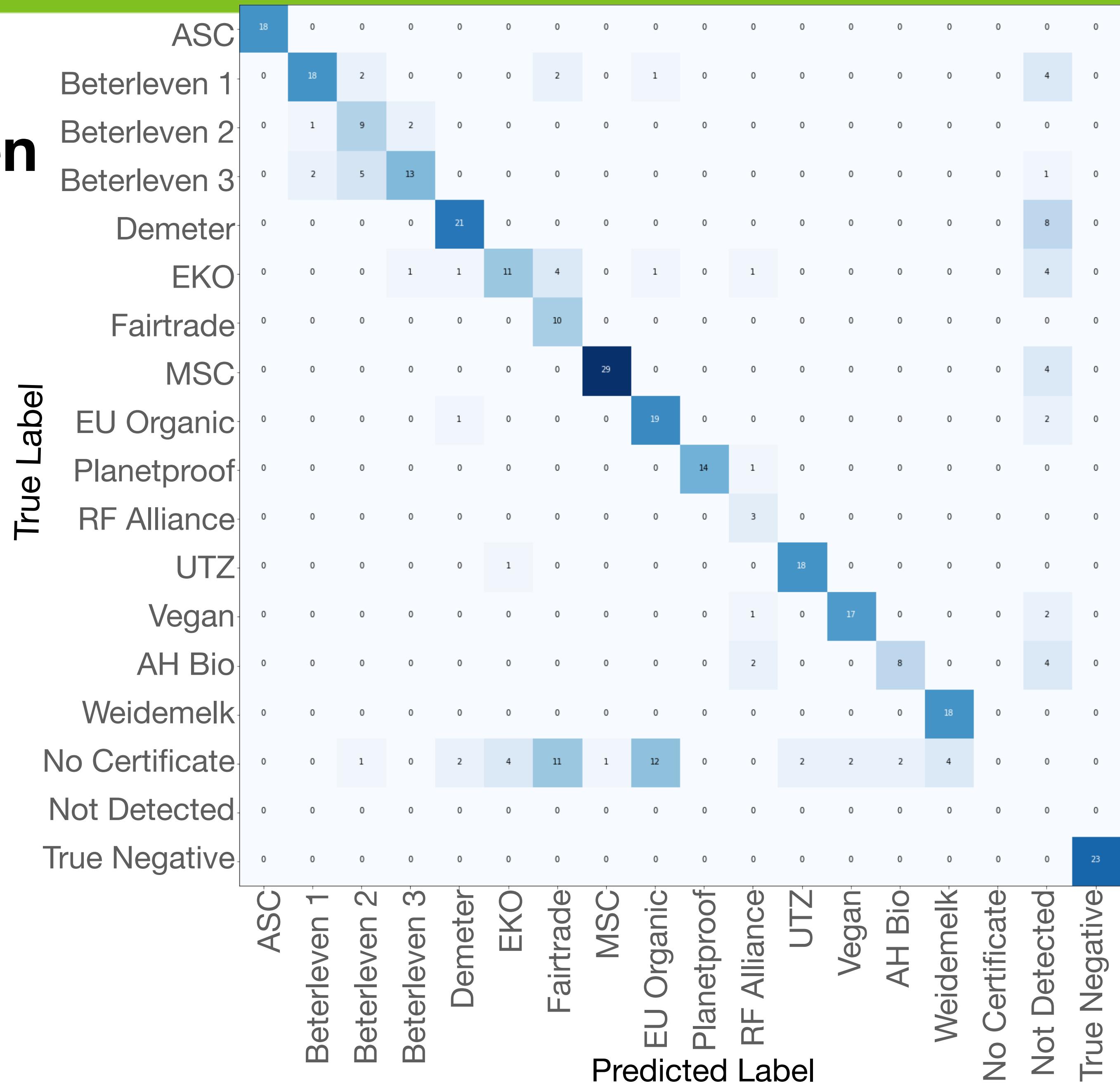
De Confusion Matrix



Verloop process

Nieuwe data & eerste resultaten

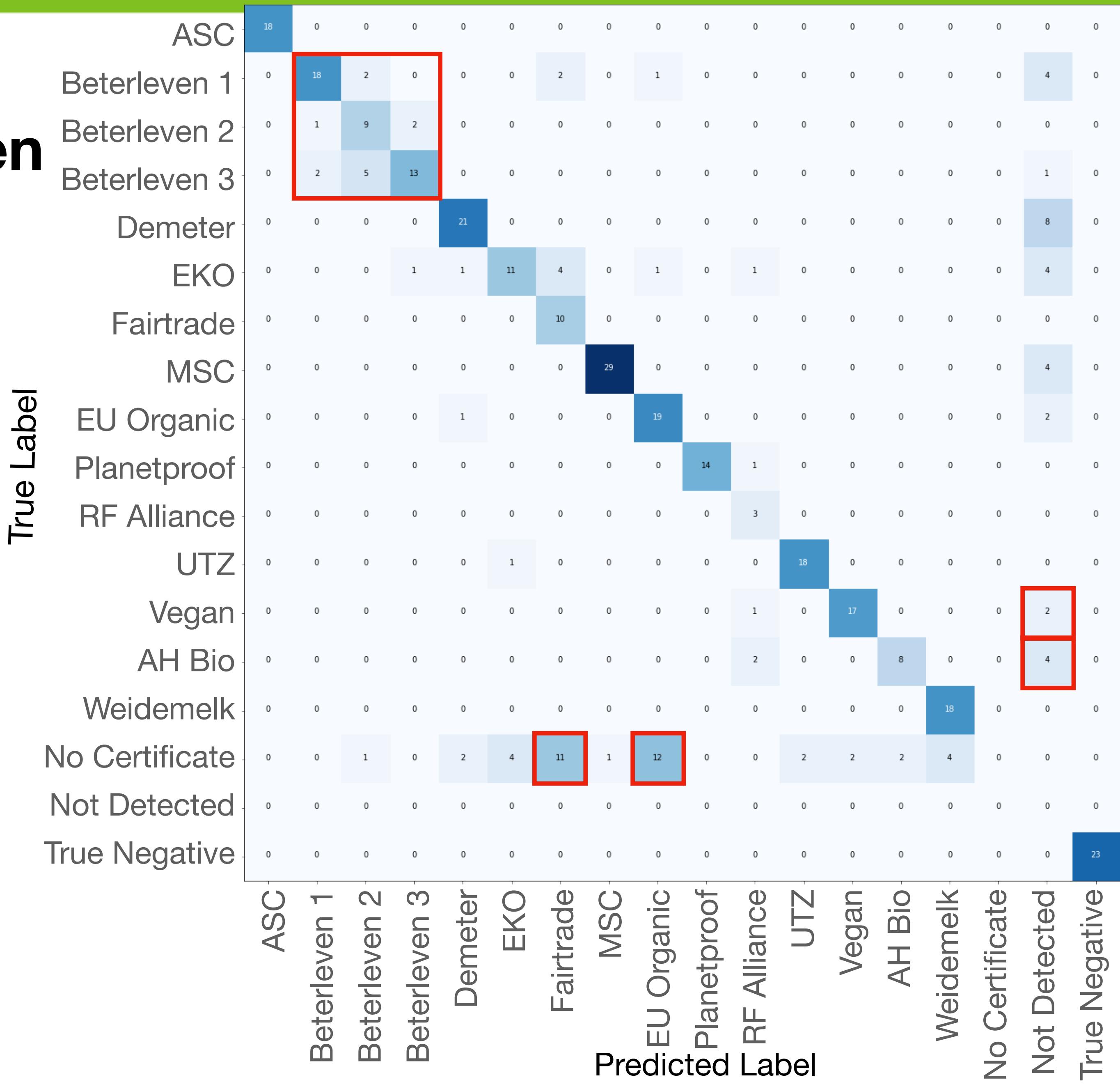
		64%	20%	16%
certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187



Verloop process

Nieuwe data & eerste resultaten

	Alles	train	test	validation
	64% 20% 16%			
certificate				
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187



Verloop process

Nieuwe data & eerste resultaten

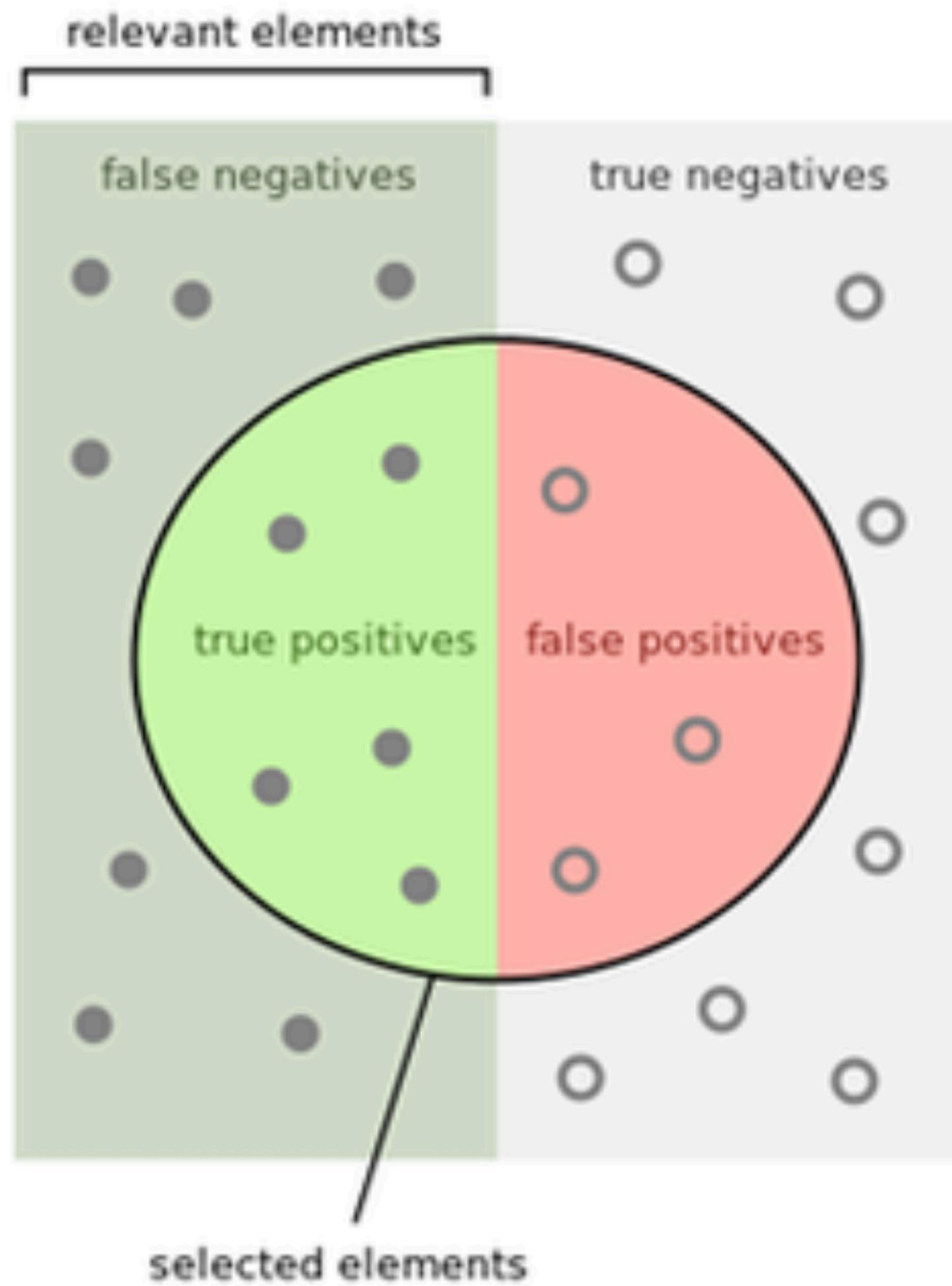
Ingebouwde Evaluatie ImageAI

certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187

asc: 1,0000
beterleven1: 0,6430
beterleven2: 0,7221
beterleven3: 0,7500
bio: 0,0000
demeter: 0,8045
eko: 0,5500
fairtrade: 0,8766
msc: 0,5556
organic: 0,8827
planetproof: 0,8516
rfalliance: 1,0000
utz: 0,8464
vegan: 0,0000
weidemelk: 0,8109
mAP: 0,6862

Opheldering termen

Accuracy, Precision en Recall



How many of the predictions of a certificate were correct?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How often did the model recognise an actual label?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

How many of the predictions were correct?

$$\text{Accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives}}$$

Verloop process

Nieuwe data & eerste resultaten

certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187

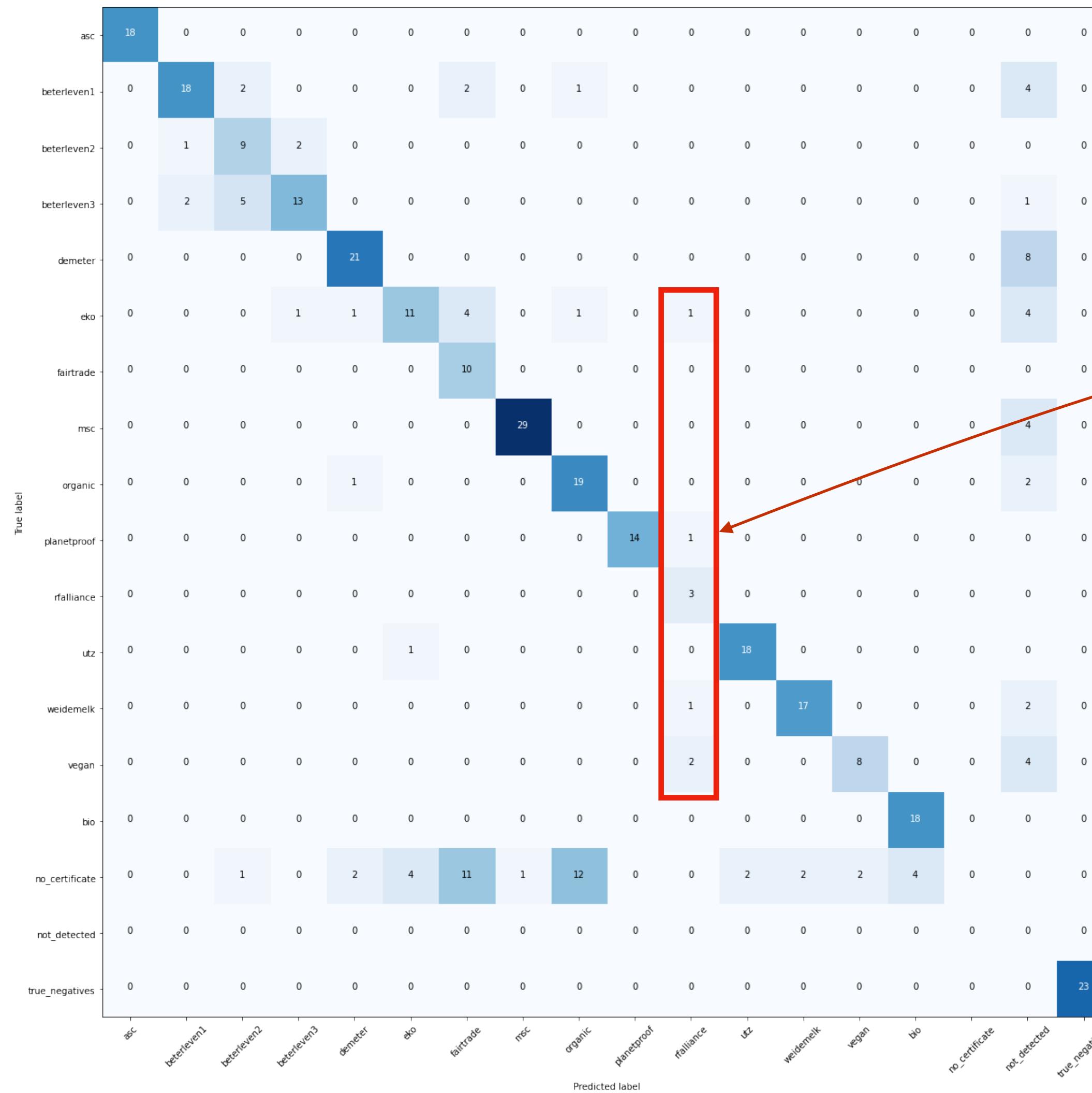
Ingebouwde Evaluatie ImageAI

asc: 1,0000
beterleven1: 0,6430
beterleven2: 0,7221
beterleven3: 0,7500
bio: 0,0000
demeter: 0,8045
eko: 0,5500
fairtrade: 0,8766
msc: 0,5556
organic: 0,8827
planetproof: 0,8516
rfalliance: 1,0000
utz: 0,8464
vegan: 0,0000
weidemelk: 0,8109
mAP: 0,6862

Verloop process

Nieuwe data & eerste resultaten

certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187

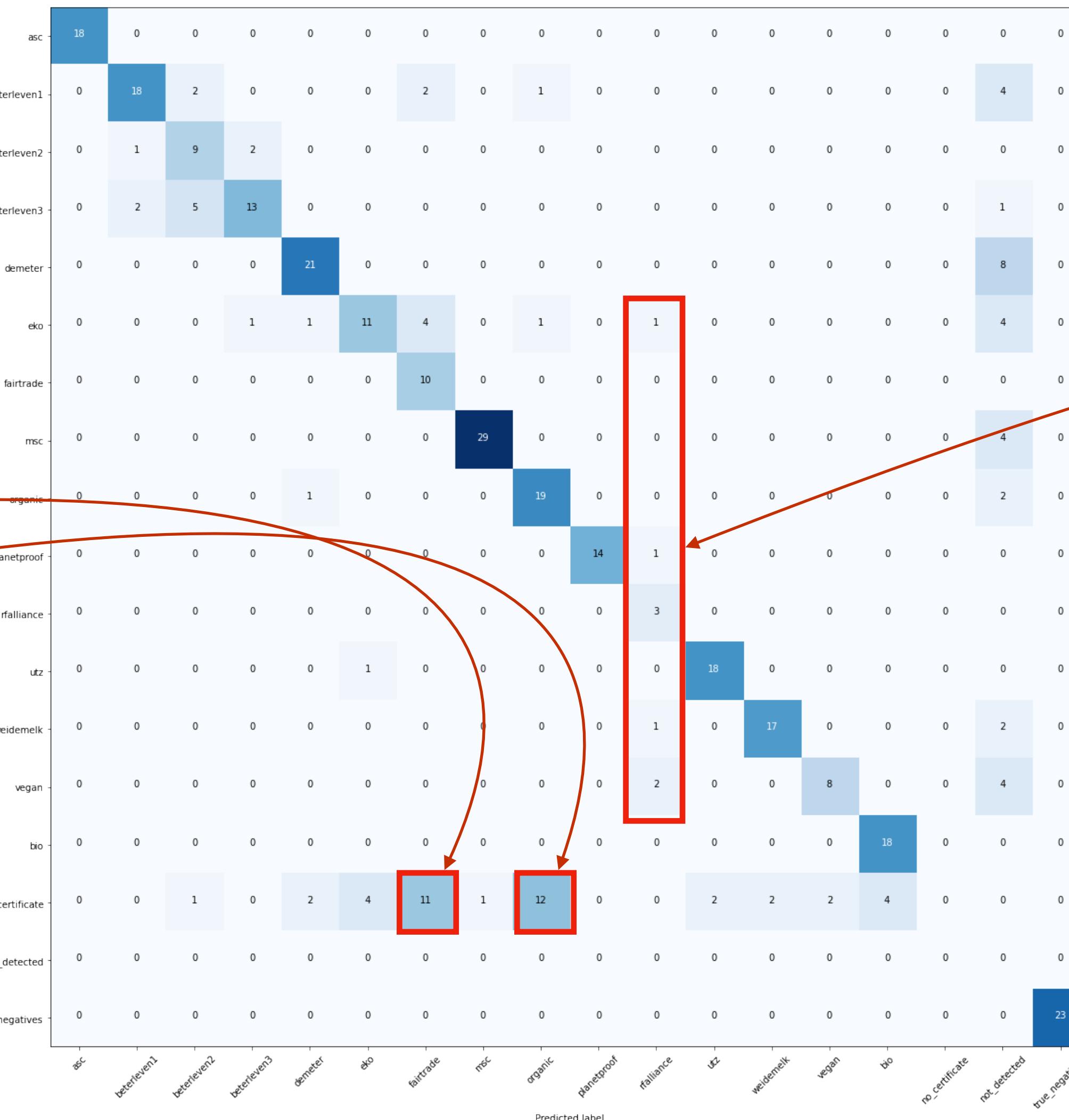


asc: 1,0000
 beterleven1: 0,6430
 beterleven2: 0,7221
 beterleven3: 0,7500
 bio: 0,0000
 demeter: 0,8045
 eko: 0,5500
 fairtrade: 0,8766
 msc: 0,5556
 organic: 0,8827
 planetproof: 0,8516
 rfalliance: 1,0000
 utz: 0,8464
 vegan: 0,0000
 weidemelk: 0,8109
 mAP: 0,6862

Verloop process

Nieuwe data & eerste resultaten

certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187

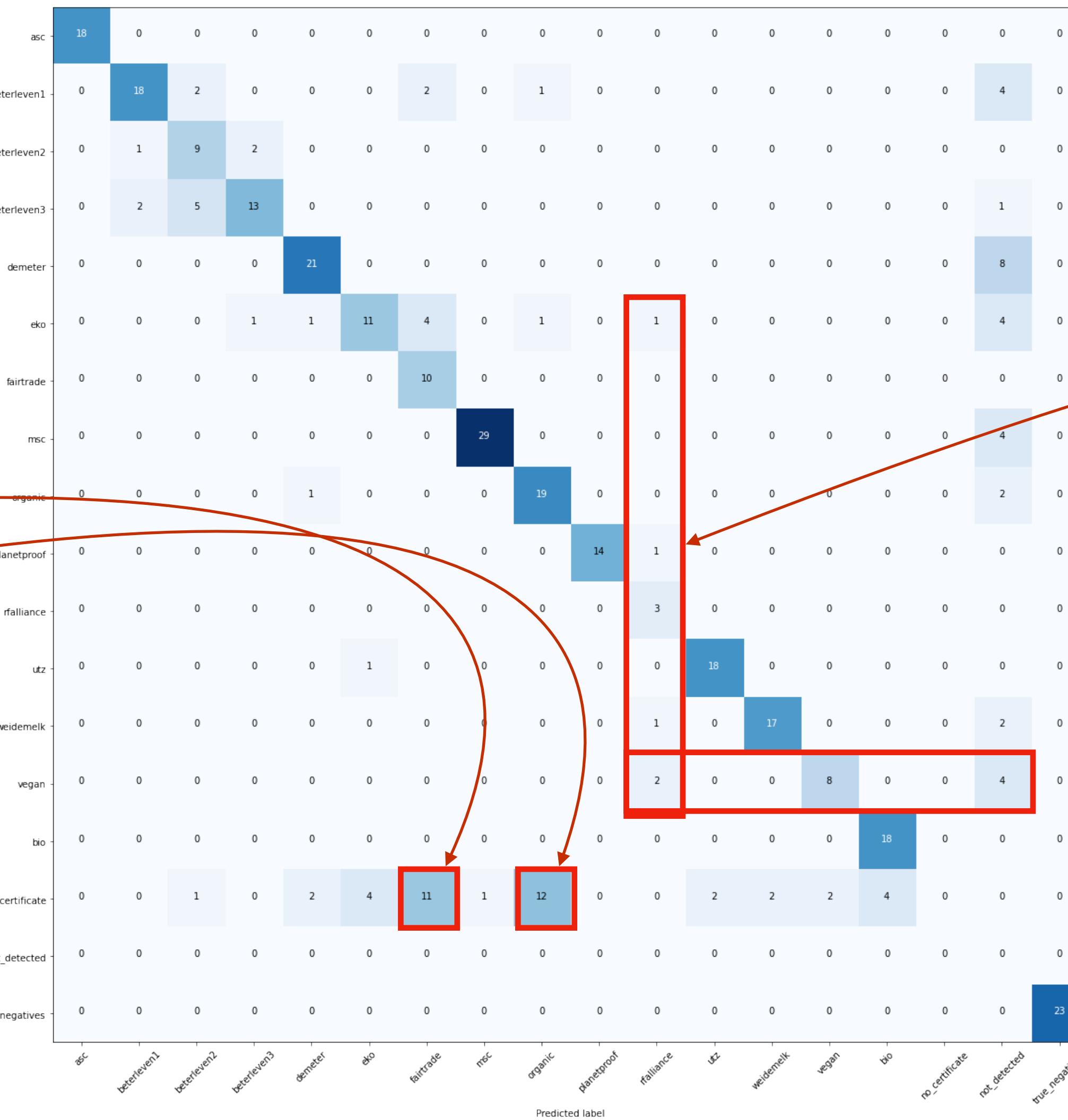


asc: 1,0000
 beterleven1: 0,6430
 beterleven2: 0,7221
 beterleven3: 0,7500
 bio: 0,0000
 demeter: 0,8045
 eko: 0,5500
 fairtrade: 0,8766
 msc: 0,5556
 organic: 0,8827
 planetproof: 0,8516
 rfalliance: 1,0000
 utz: 0,8464
 vegan: 0,0000
 weidemelk: 0,8109
 mAP: 0,6862

Verloop process

Nieuwe data & eerste resultaten

certificate	Alles	train	test	validation
asc	68	41	16	11
beterleven1	90	55	22	13
beterleven2	66	42	17	7
beterleven3	72	52	13	7
bio	52	35	11	6
demeter	97	69	17	11
eko	89	65	9	15
fairtrade	164	110	29	25
msc	89	59	19	11
organic	242	163	40	39
planetproof	65	39	16	10
rfalliance	38	28	5	5
utz	78	51	16	11
vegan	43	27	9	7
weidemelk	93	60	22	11
total	1346	896	261	187



asc: 1,0000
 beterleven1: 0,6430
 beterleven2: 0,7221
 beterleven3: 0,7500
 bio: 0,0000
 demeter: 0,8045
 eko: 0,5500
 fairtrade: 0,8766
 msc: 0,5556
 organic: 0,8827
 planetproof: 0,8516
 rfalliance: 1,0000
 utz: 0,8464
 vegan: 0,0000
 weidemelk: 0,8109
 mAP: 0,6862

Verloop process

Verbeteringen model mbt data, training en evaluatie

- Data
 - > AH Bio en Vegan eruit
 - > Gebalanceerde hoeveelheden Fairtrade en EU Organic
 - > Preprocessing
- Training
 - > Data augmentatie
 - > Loss function
 - > Volledige training op AWS ipv CoLab
- Evaluatie
 - > Kwantitatiever recall en accuracy in kaart brengen

Verheldering van wat termen

Balanceren Fairtrade en EU Organic

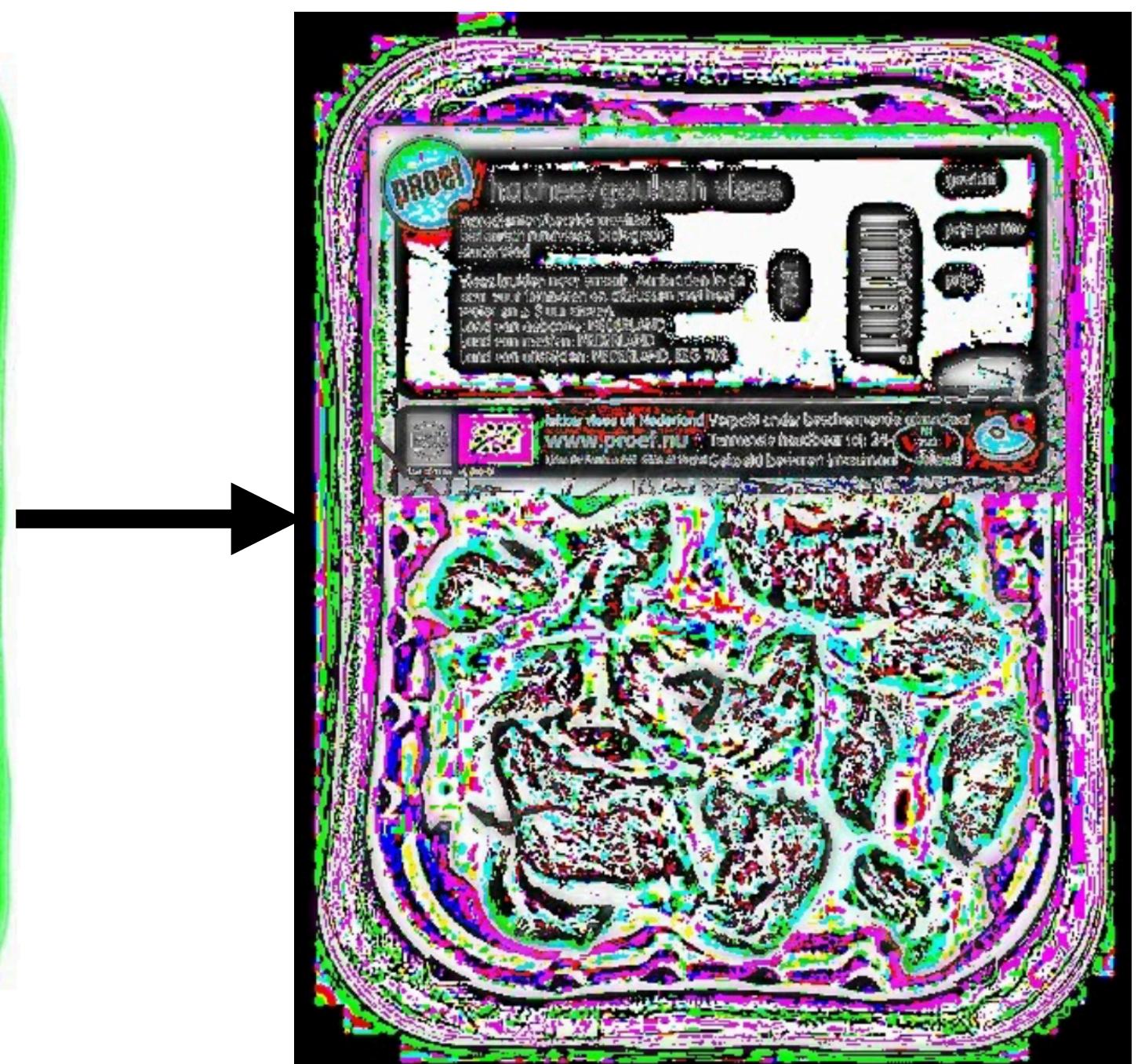
Plaatjes uitkrassen om niet andere labels te verwijderen



Verheldering van wat termen

Preprocessing

“Universelere” plaatjes om zo de keurmerken meer te laten opvallen



Verheldering van wat termen

Data Augmentatie

Plaatjes een beetje vervormen en veranderen van kleur om zo meer plaatjes te hebben



Verheldering van wat termen

Loss Function

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

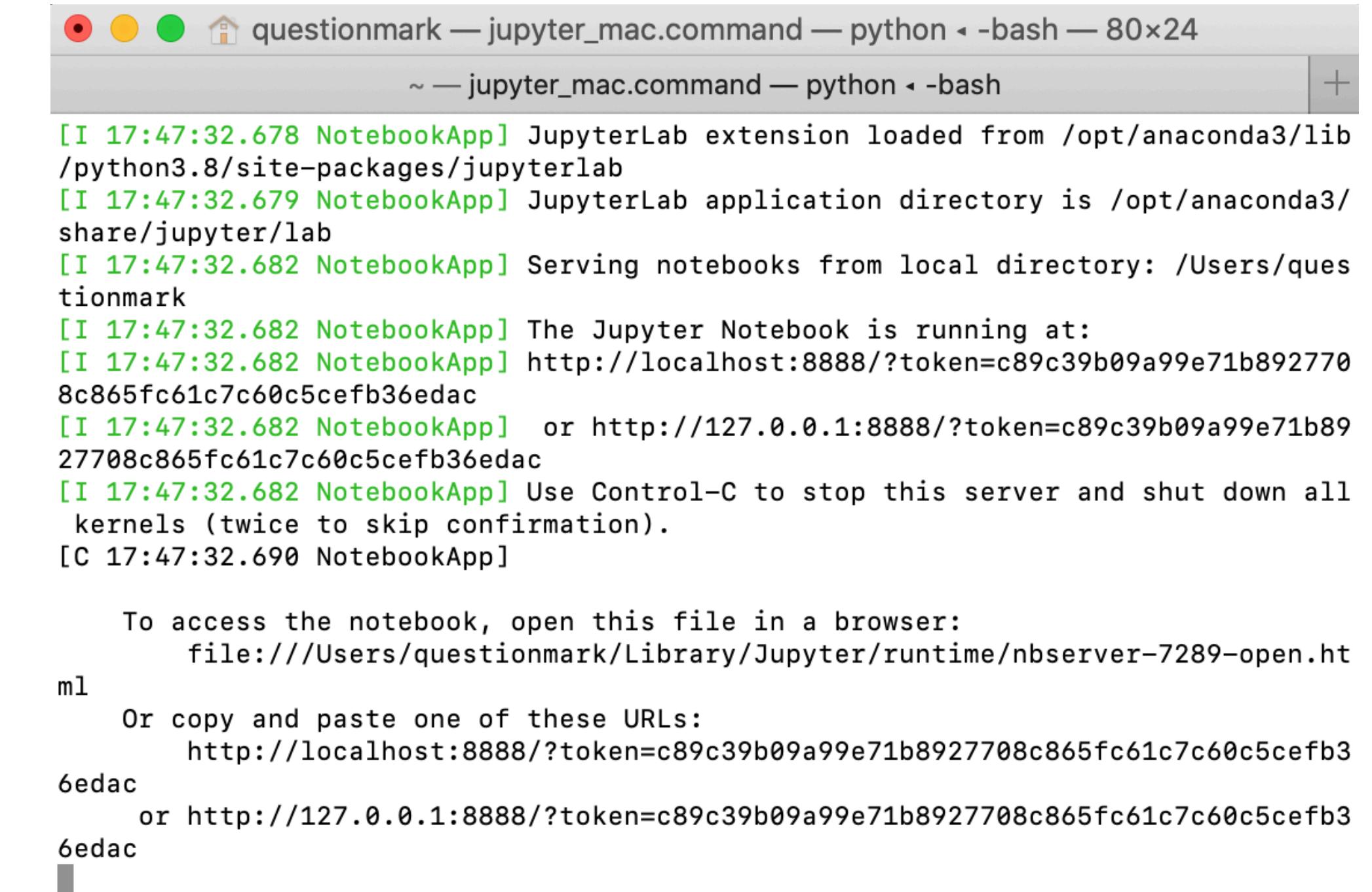
$$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

$$+ \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2$$

Verheldering van wat termen

Trainen op AWS computer

- Zware berekeningen kosten veel rekenkracht en tijd
- Mijn computer heeft niet zo'n sterke rekenkracht
- Google CoLab beperkt de rekenkracht en rekentijd op een onvoorspelbaar moment
- AWS verhuurt rekenkracht en -tijd +



```
[I 17:47:32.678 NotebookApp] JupyterLab extension loaded from /opt/anaconda3/lib/python3.8/site-packages/jupyterlab
[I 17:47:32.679 NotebookApp] JupyterLab application directory is /opt/anaconda3/share/jupyter/lab
[I 17:47:32.682 NotebookApp] Serving notebooks from local directory: /Users/questionmark
[I 17:47:32.682 NotebookApp] The Jupyter Notebook is running at:
[I 17:47:32.682 NotebookApp] http://localhost:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
[I 17:47:32.682 NotebookApp] or http://127.0.0.1:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
[I 17:47:32.682 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 17:47:32.690 NotebookApp]

To access the notebook, open this file in a browser:
file:///Users/questionmark/Library/Jupyter/runtime/nbserver-7289-open.html

Or copy and paste one of these URLs:
http://localhost:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
or http://127.0.0.1:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
```

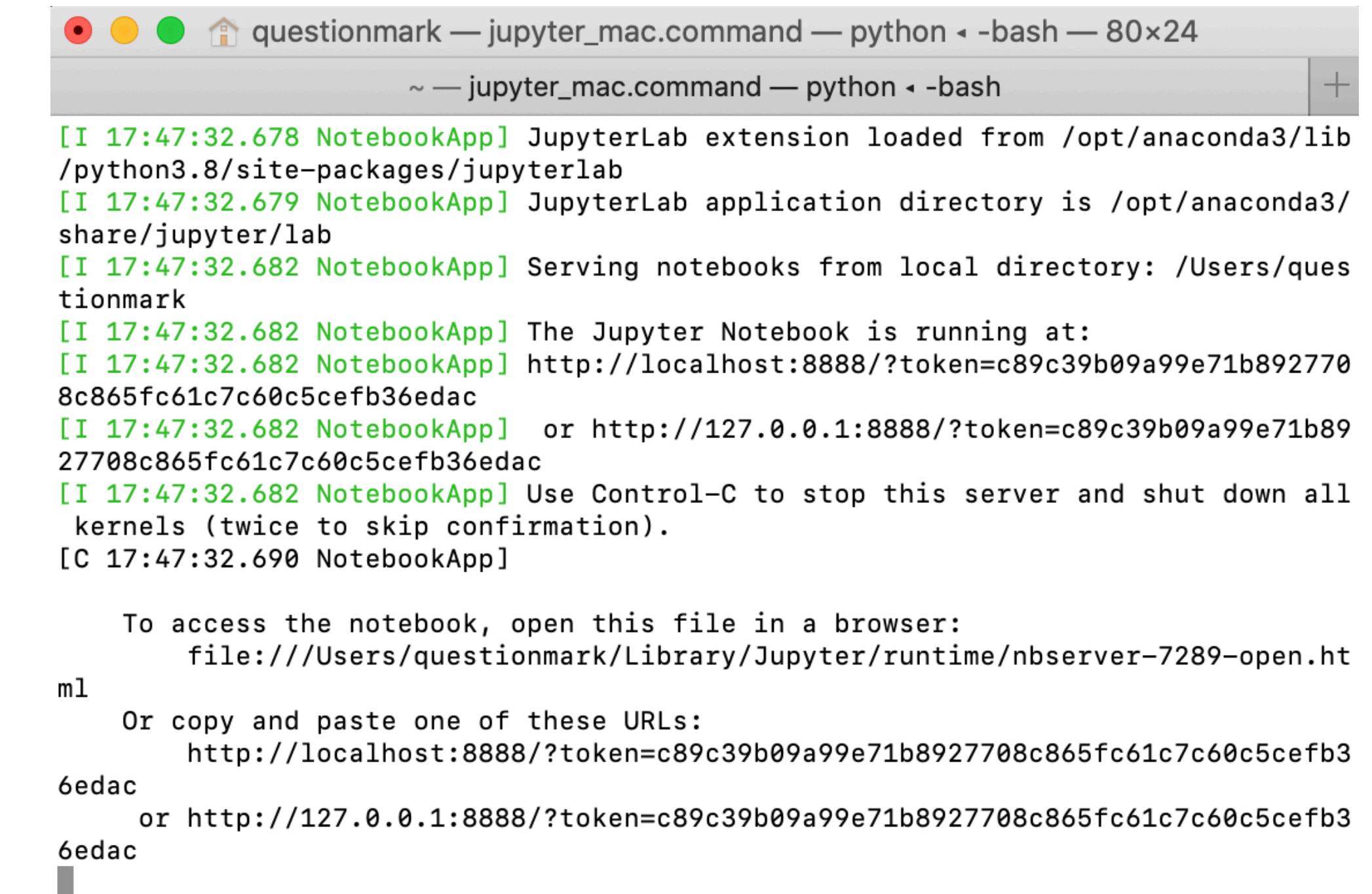
Een goede training doe je
op een AWS computer

Verheldering van wat termen

Trainen op AWS computer

- Zware berekeningen kosten veel rekenkracht en tijd
- Mijn computer heeft niet zo'n sterke rekenkracht
- Google CoLab beperkt de rekenkracht en rekentijd op een onvoorspelbaar moment
- AWS verhuurt rekenkracht en -tijd +

Een goede training doe je
op een AWS computer



```
[I 17:47:32.678 NotebookApp] JupyterLab extension loaded from /opt/anaconda3/lib/python3.8/site-packages/jupyterlab
[I 17:47:32.679 NotebookApp] JupyterLab application directory is /opt/anaconda3/share/jupyter/lab
[I 17:47:32.682 NotebookApp] Serving notebooks from local directory: /Users/questionmark
[I 17:47:32.682 NotebookApp] The Jupyter Notebook is running at:
[I 17:47:32.682 NotebookApp] http://localhost:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
[I 17:47:32.682 NotebookApp] or http://127.0.0.1:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
[I 17:47:32.682 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 17:47:32.690 NotebookApp]

To access the notebook, open this file in a browser:
file:///Users/questionmark/Library/Jupyter/runtime/nbserver-7289-open.html

Or copy and paste one of these URLs:
http://localhost:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
or http://127.0.0.1:8888/?token=c89c39b09a99e71b8927708c865fc61c7c60c5cefb36edac
```

Vind ik ook moeilijk

Resultaten

Wat is er getest?

Resultaten Vergeleken

Alteration	Date	Epochs	mAP	mAP	mAR	accuracy	Negatives	Negatives
			ImageAI				AP	AR
Default		65	0,761	0,931	0,832	0,770	0,244	0,392
		53	0,759	0,873	0,837	0,716	0,284	0,442
Exclusief Bio & Vegan	51	0,794	0,922	0,822	0,793	0,345	0,527	
Fairtrade & EU Organic	93	0,740	0,880	0,855	0,780	0,390	0,533	
Gebalanceerd	65	0,743	0,903	0,851	0,785	0,400	0,542	
	46	0,733	0,905	0,858	0,803	0,390	0,542	
Preprocessed	48	0,690	0,922	0,769	0,717	0,208	0,301	
	46	0,685	0,924	0,783	0,736	0,216	0,293	
Loss Function Weights	51	0,770	0,923	0,786	0,739	0,344	0,452	
	48	0,768	0,922	0,781	0,734	0,340	0,434	
Loss Function Weights	67	0,606	0,938	0,748	0,730	0,212	0,310	
	50	0,603	0,943	0,749	0,743	0,214	0,319	
Augment zoom	45	0,591	0,463	0,687	0,398	0,157	0,227	
	33	0,739	0,937	0,752	0,730	0,191	0,300	
Augment colors	55	0,759	0,930	0,817	0,775	0,289	0,440	
	41	0,601	0,562	0,748	0,477	0,204	0,278	
Meer keurmerkloze data (RFalliance verwijderd)	41	0,765	0,947	0,831	0,819	0,563	0,684	
	34	0,755	0,945	0,831	0,822	0,540	0,684	

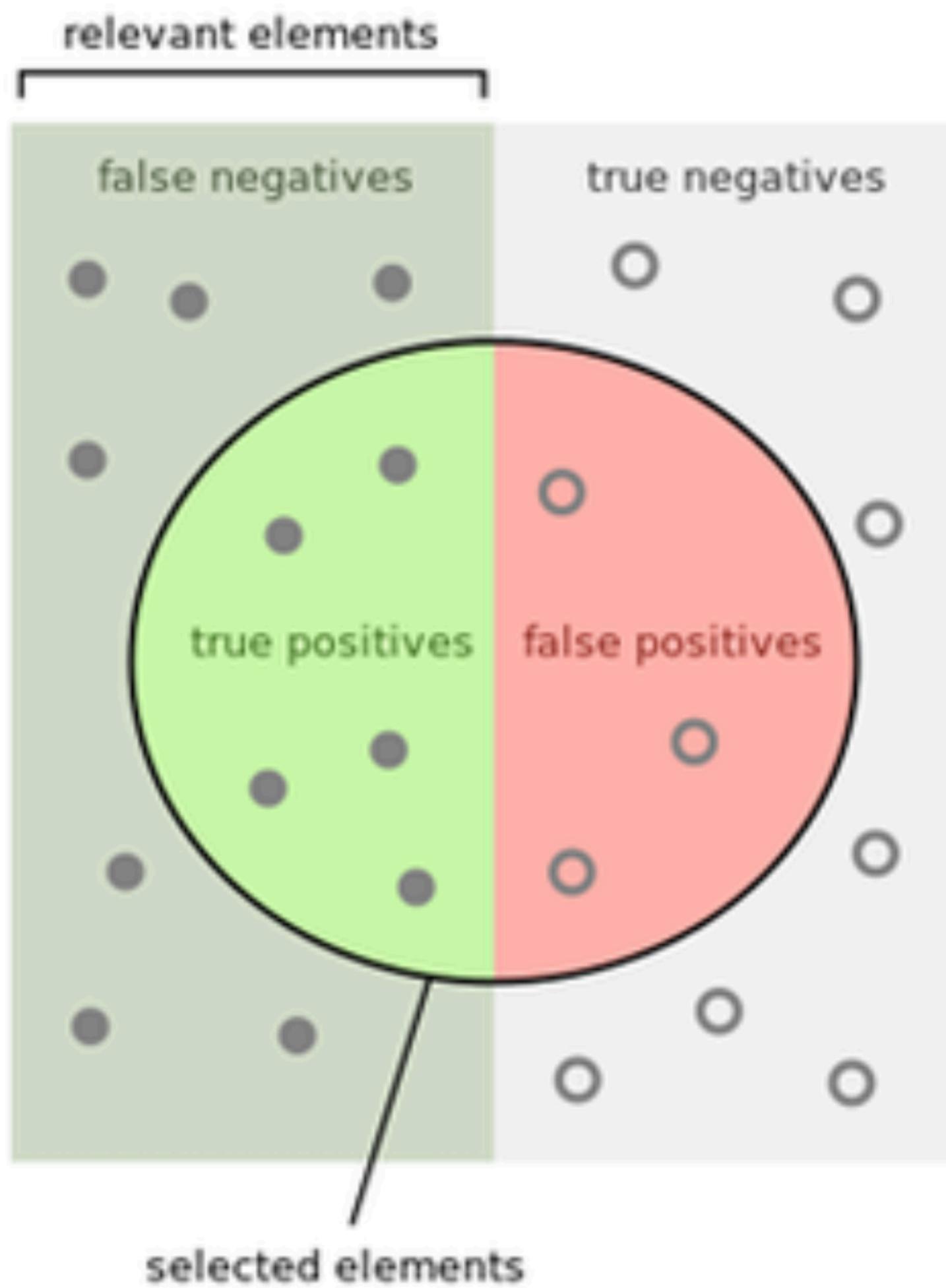
Resultaten

De Belangrijke details

<i>Alteration</i>	Date	mAP	mAR	accuracy	Negatives AP	Negatives AR
Default		0,931	0,832	0,770	0,244	0,392
Exclusief Bio & Vegan		0,922	0,822	0,793	0,345	0,527
Fairtrade & EU Organic Gebalanceerd		0,905	0,858	0,803	0,390	0,542
Preprocessed		0,924	0,783	0,736	0,216	0,293
Loss Function Weights		0,923	0,786	0,739	0,344	0,452
Loss Function Weights		0,943	0,749	0,743	0,214	0,319
Augment zoom		0,463	0,687	0,398	0,157	0,227
Augment colors		0,930	0,817	0,775	0,289	0,440
Meer keurmerkloze data (RFalliance verwijderd)		0,947	0,831	0,819	0,563	0,684

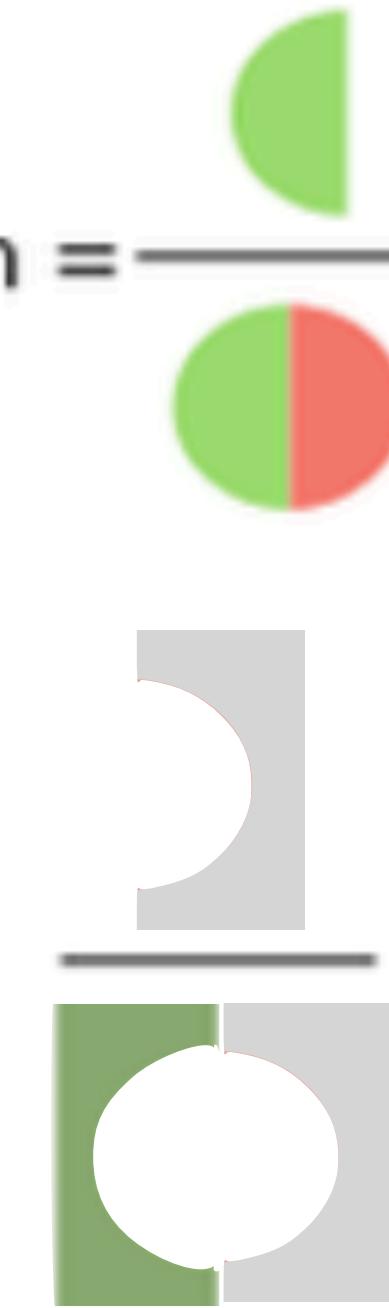
Opheldering termen

Accuracy, Precision en Recall



How many of the predictions of a certificate were correct?

Precision =



nAP =

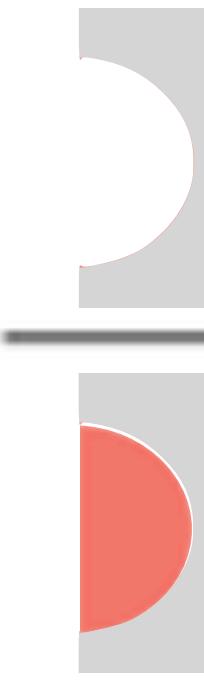


How often did the model recognise an actual label?

Recall =



nAR =



How many of the predictions were correct?

Accuracy =



Resultaten

De Belangrijke details

<i>Alteration</i>	Date	mAP	mAR	accuracy	Negatives AP	Negatives AR
Default		0,931	0,832	0,770	0,244	0,392
Exclusief Bio & Vegan		0,922	0,822	0,793	0,345	0,527
Fairtrade & EU Organic Gebalanceerd		0,905	0,858	0,803	0,390	0,542
Preprocessed		0,924	0,783	0,736	0,216	0,293
Loss Function Weights		0,923	0,786	0,739	0,344	0,452
Loss Function Weights		0,943	0,749	0,743	0,214	0,319
Augment zoom		0,463	0,687	0,398	0,157	0,227
Augment colors		0,930	0,817	0,775	0,289	0,440
Meer keurmerkloze data (RFalliance verwijderd)		0,947	0,831	0,819	0,563	0,684

Resultaten

Wat verbeterd?

<i>Alteration</i>	Date	mAP	mAR	accuracy	Negatives AP	Negatives AR
Default		0,931	0,832	0,770	0,244	0,392
Exclusief Bio & Vegan		0,922	0,822	0,793	0,345	0,527
Fairtrade & EU Organic Gebalanceerd		0,905	0,858	0,803	0,390	0,542
Preprocessed		0,924	0,783	0,736	0,216	0,293
Loss Function Weights		0,923	0,786	0,739	0,344	0,452
Loss Function Weights		0,943	0,749	0,743	0,214	0,319
Augment zoom		0,463	0,687	0,398	0,157	0,227
Augment colors		0,930	0,817	0,775	0,289	0,440
Meer keurmerkloze data (RFalliance verwijderd)		0,947	0,831	0,819	0,563	0,684

Resultaten

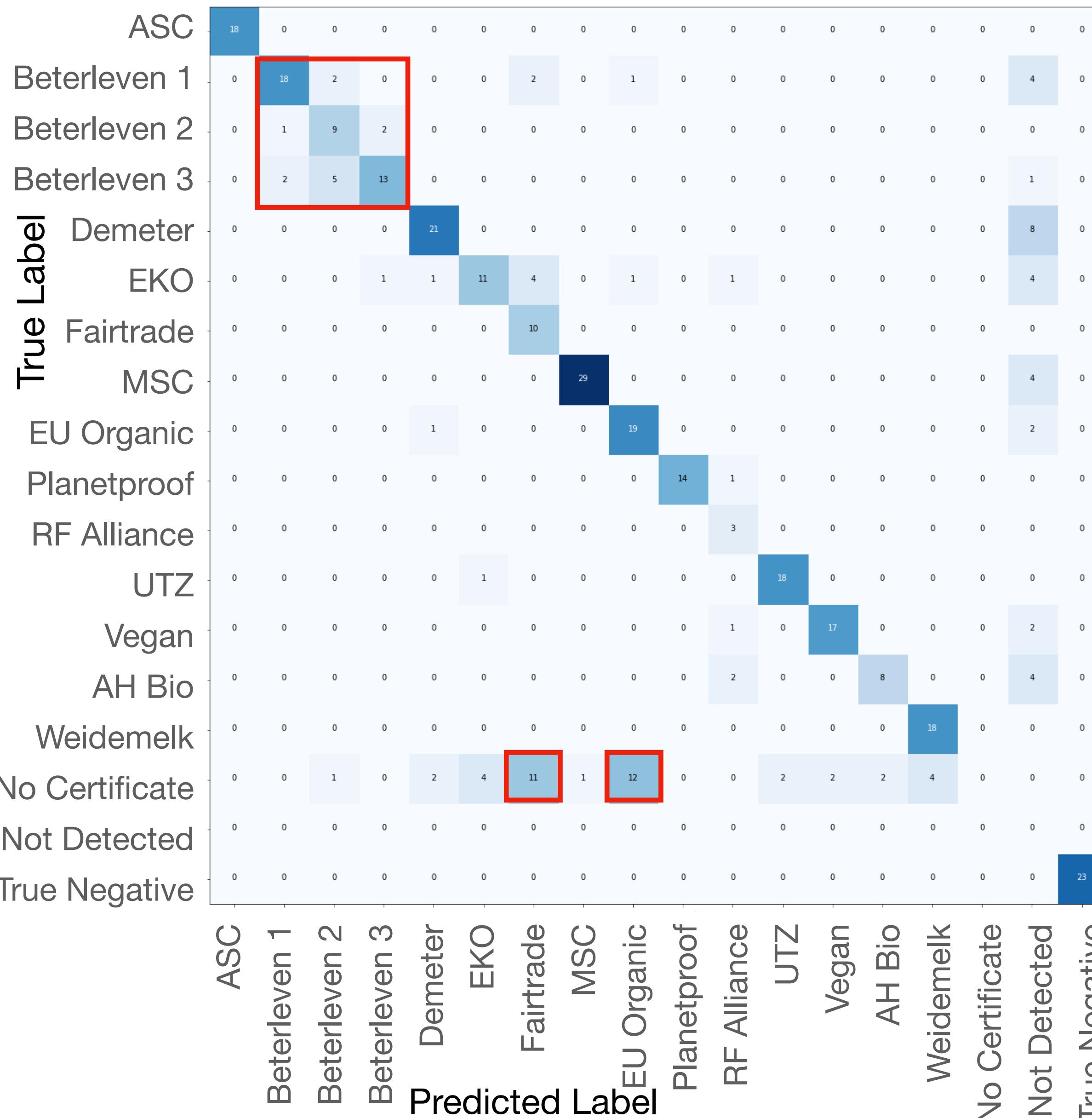
Final Results

- Bio, Vegan en RF Alliance labels
- Buitenproportioneel veel Fairtrade & EU Organic

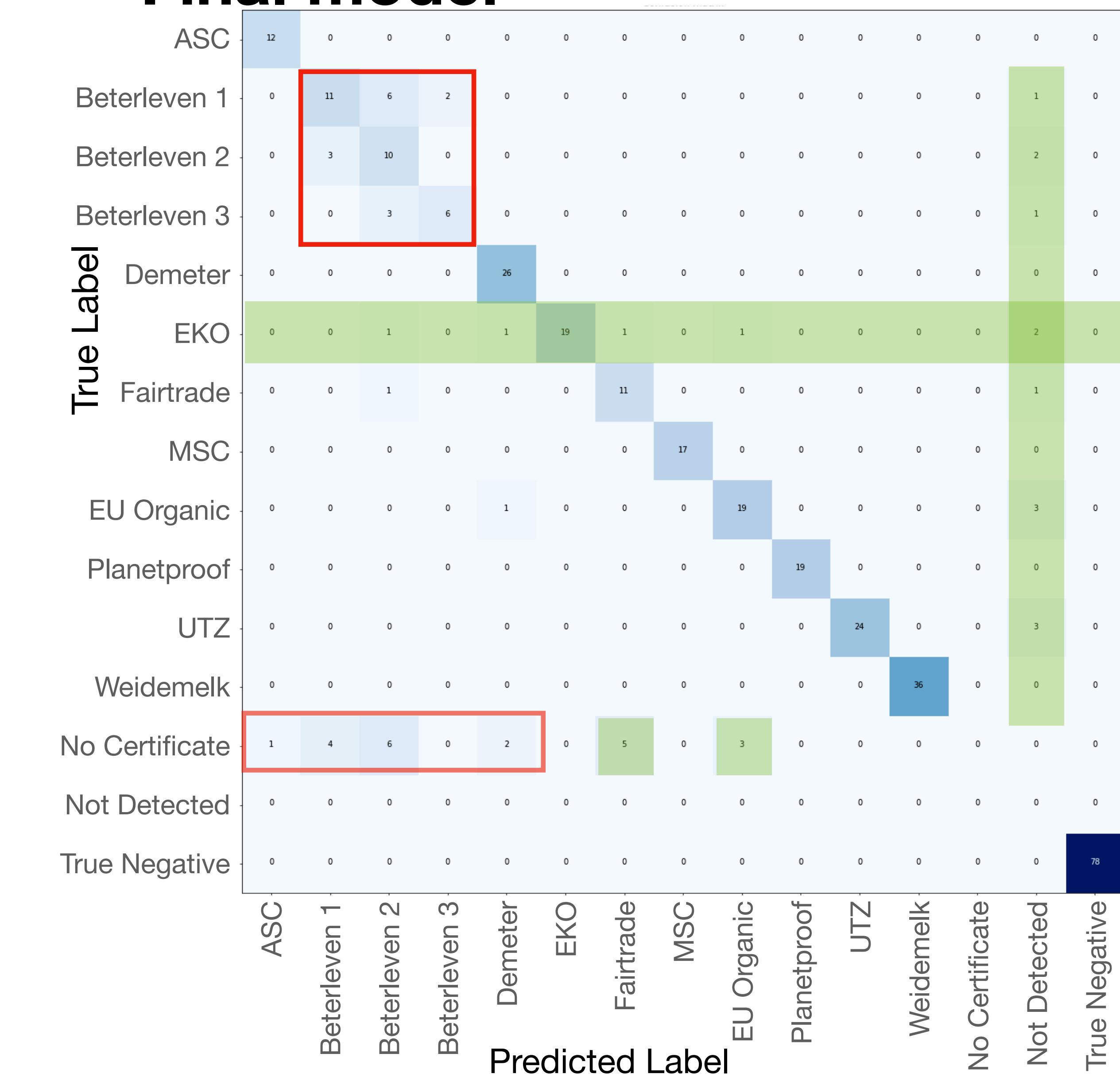
Epochs	mAP ImageAI	mAP	mAR	accuracy	Negatives	Negatives
		AP	AR			
Default	65	0,761	0,931	0,832	0,770	0,244
Final	47	0,807	0,897	0,845	0,842	0,703

Resultaten

Default model



Final model



Persononlijke ervaring

Wat heb ik geleerd?

- HEEL VEEL!
- (Convolutioneel) **neuraal netwerk**
 - >Theorie
 - > Praktijk
 - > **Factoren** die invloed hebben op **prestatie**
 - > **Evaluieren** prestatie
- Hoe gaat een **development process** voor software
- Gestructureerde programmeren (in Python)
- Hoe werk ik met open source
- Trainen in een cloud
- Hoe opereert een stichting en in het bijzonder Questionmark

Persononlijke ervaring

Zie ik mezelf dit later doen?

- + Bij een idealistisch instituut werken: JA
- + Bij een onderzoeksinstuut werken: JA
- + Efficiente algorithmes bedenken en implementeren: JA
- Open source neurale netwerken implementeren: NEE
- In een terminal prutsen: LIEVER NIET

Questions?



Bedankt allemaal!!

- Willem, bedankt voor de begeleiding, het vertrouwen en alle lekkere theetjes die ik bij jou en Talita thuis heb mogen proeven :) Ik heb het echt heel erg gezellig en fijn gevonden om samen met jou hieraan te mogen werken!
- Bedankt, Questionmark, voor de kans en het kijkje in de voedselindustrie, wetenschappelijk onderzoek doen, campagne voeren en bovenal de gezelligheid. Ik ga de tijd bij jullie niet snel vergeten (en waarschijnlijk ook niet snel beëindigen ;))
- Bedankt Vantage AI en in het bijzonder Chiel voor de technische ondersteuning. Ik heb ongelofelijk veel geleerd!
- Mijn lieve wiskunde vriendjes die mij af en toe nog geholpen hebben bij een acute error!



Guacamole

