

Modellen en toepassingen van AI in de hematologische diagnostiek

Alexander Coulie – 1^e jaar assistent klinische biologie
Supervisor Christine Van Laer

Overzicht

Inleiding

- **Toenemende interesse in AI**
- Welke AI modellen bestaan er?
- Datamining in machine learning

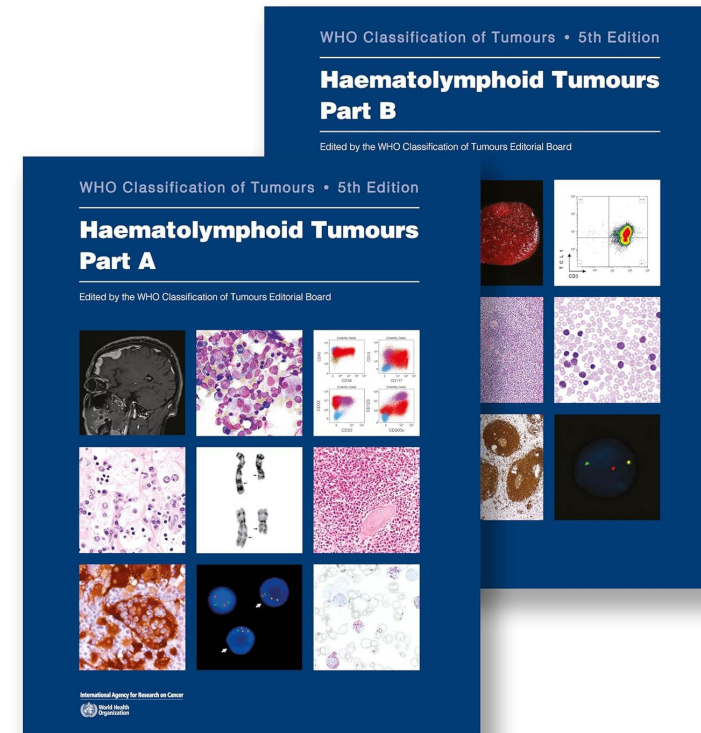
Onderzoeksvragen

1. Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?
2. Kan AI de diagnostiek van MDS verbeteren?

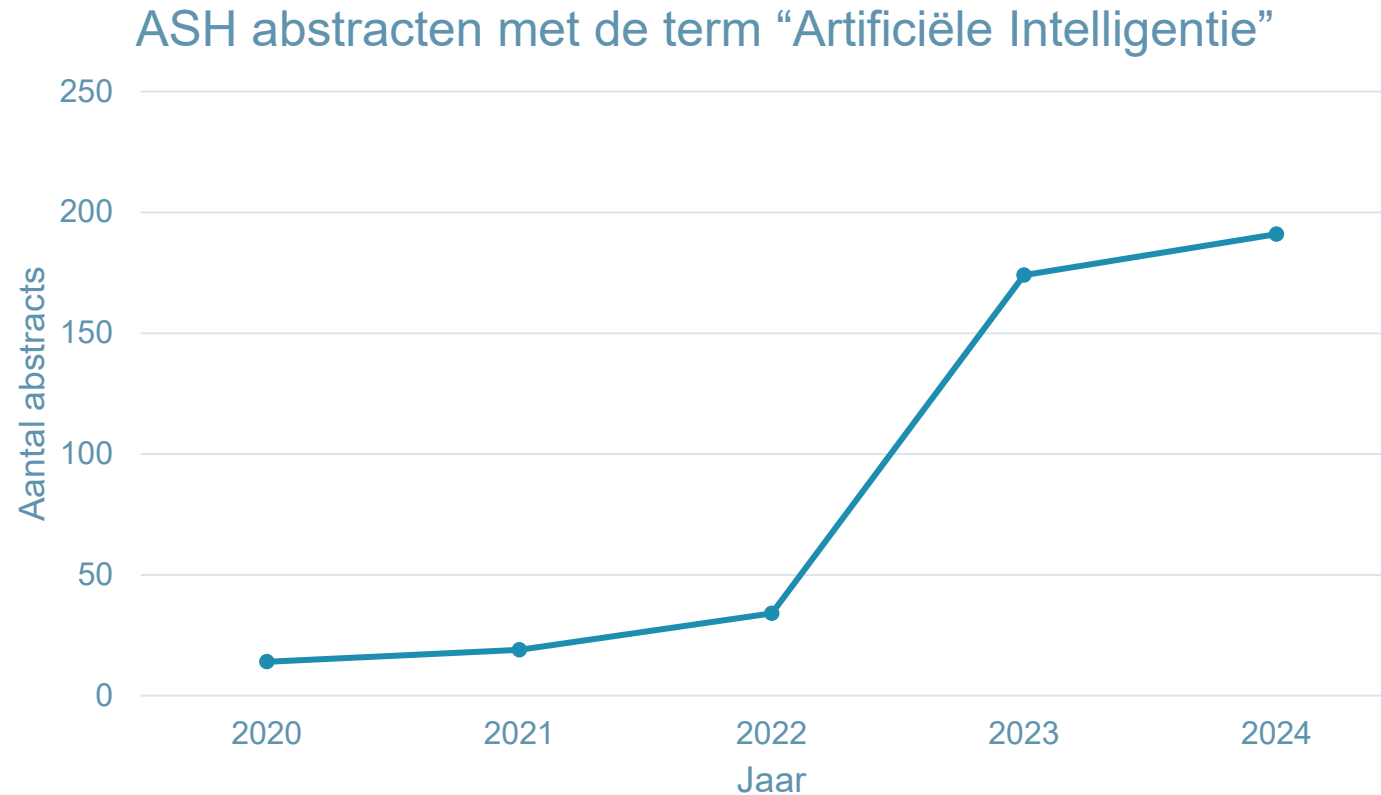
Conclusie & To Do

WHO 2022

- Diagnostisch
 - Genetisch/moleculair
 - Ondersteunende morfologie
- Rol van AI binnen de hematologie?

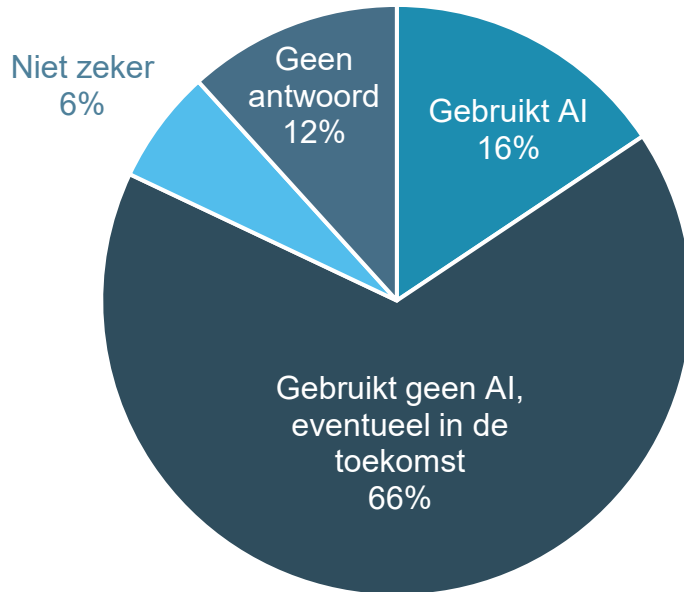


Toename van AI interesse in de hematologie



De waarde van AI in de laboratoriumgeneeskunde

Gebruik van AI



n = 302
Respondenten = 128



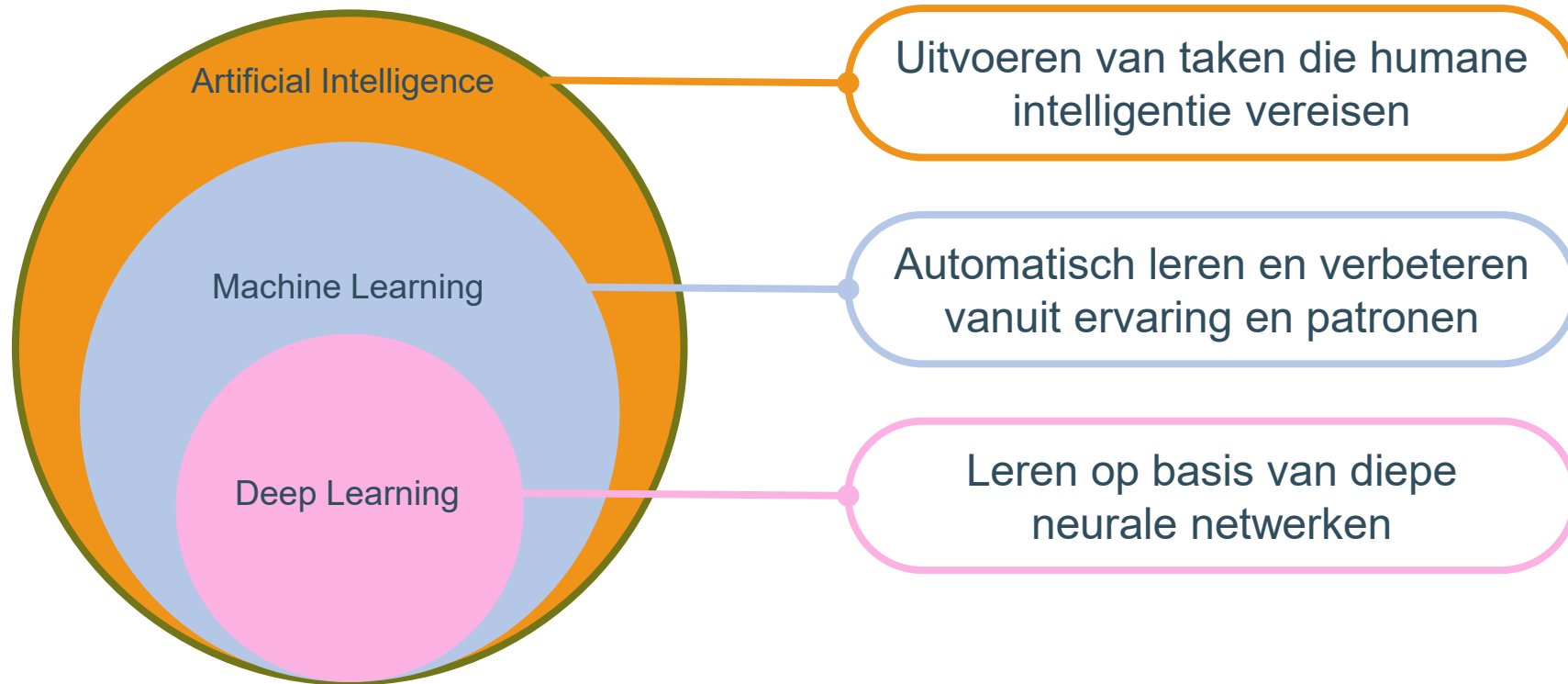
Organisatorische voordelen (efficiëntie)
Kwaliteit van zorg (nauwkeurigere diagnoses)
Kostenreductie
Betere workflow



Hoge kosten
Gebrek aan bewijs voor klinisch nut
Privacyzorgen
Beperkte kennis

Conclusie: Grote belangstelling, maar gebrek in kennis is aanwezig

AI vs. Machine learning vs. Deep learning





Arthur Samuel toont hoe machine learning kan dammen door patronen te herkennen (1962)

© IBM Watson Media 1962

Overzicht

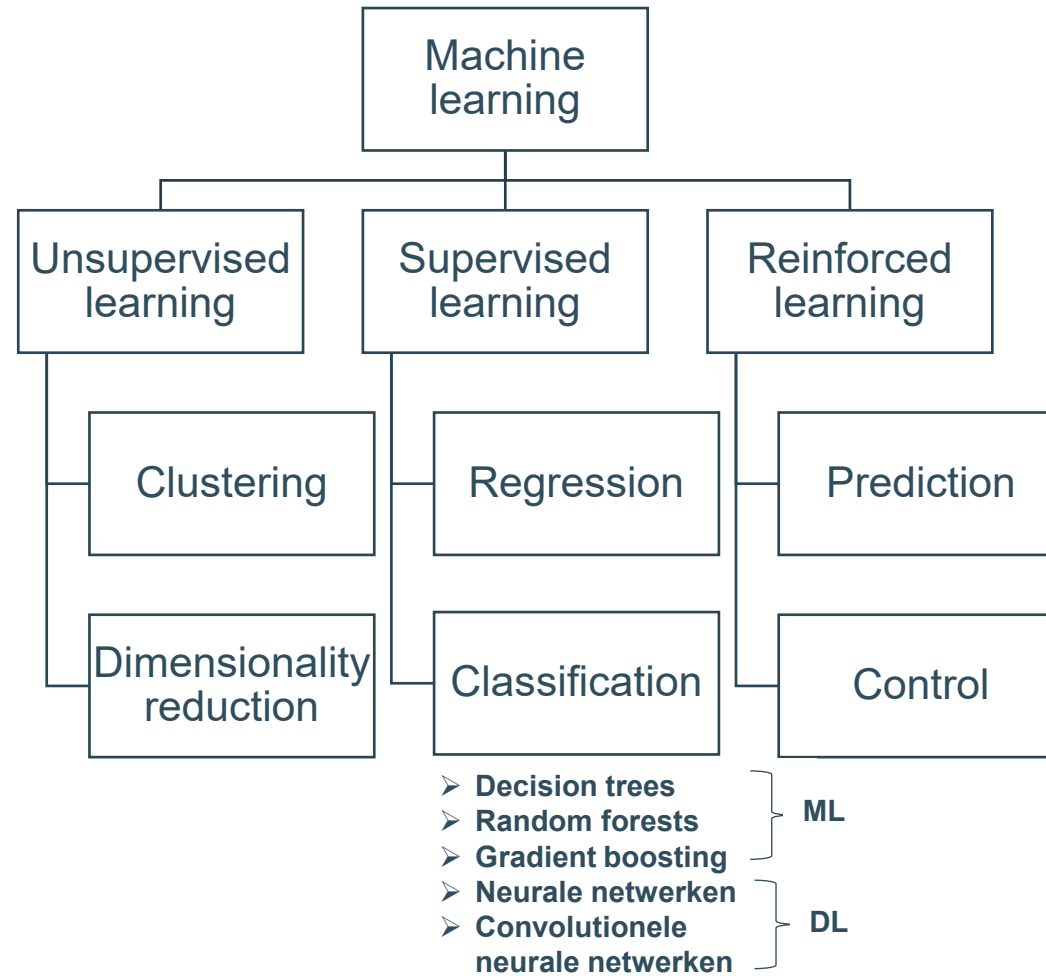
Inleiding

- Toenemende interesse in AI
- **Welke AI modellen bestaan er?**
- Datamining in machine learning

Onderzoeksvragen

1. Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?
2. Kan AI de diagnostiek van MDS verbeteren?

Conclusie & To Do



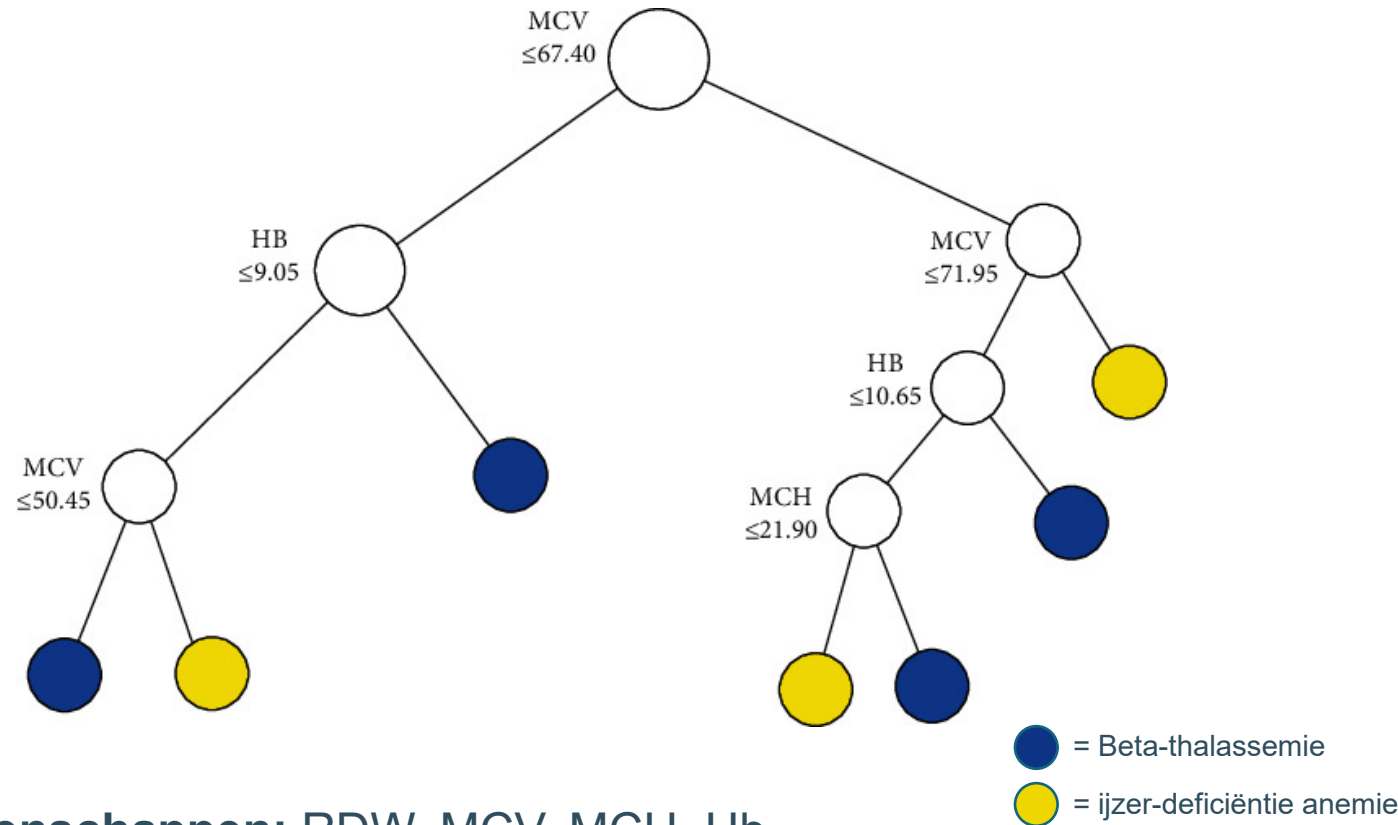
Supervised learning

Supervised learning



Op basis van gelabelde trainingsdata wordt een outcome voorspeld

Voorbeeld: Het voorspellen van beta thalassemie of ijzer deficiëntie anemie

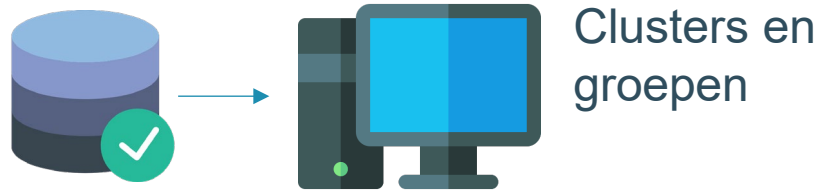


Eigenschappen: RDW, MCV, MCH, Hb

Label Beta thalassemie trait vs ijzer deficiëntie anemie

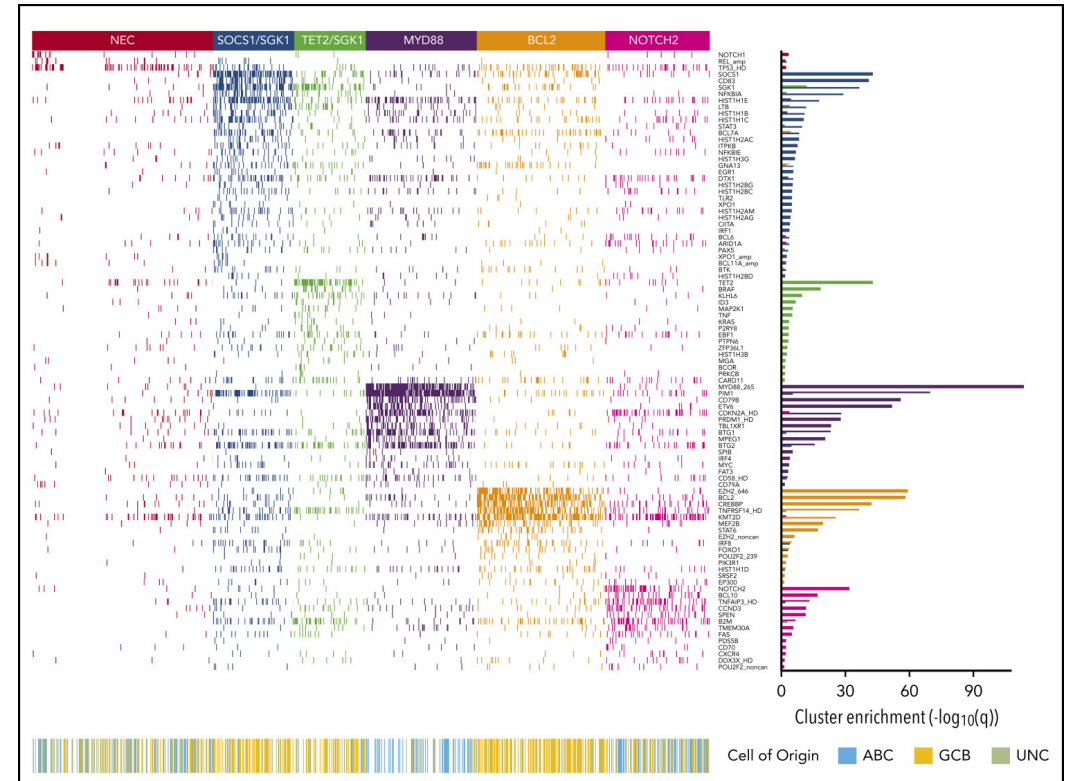
Unsupervised learning

Unsupervised learning



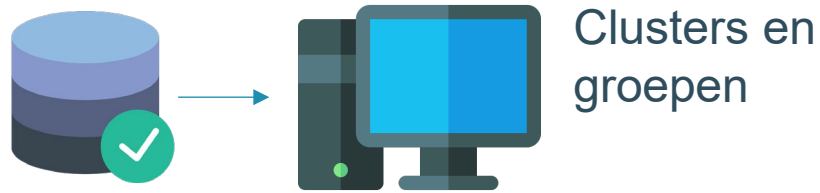
Het model leert op basis van groepen en patronen

Voorbeeld: Identificatie van gen expressie clusters bij B-cel lymfoom



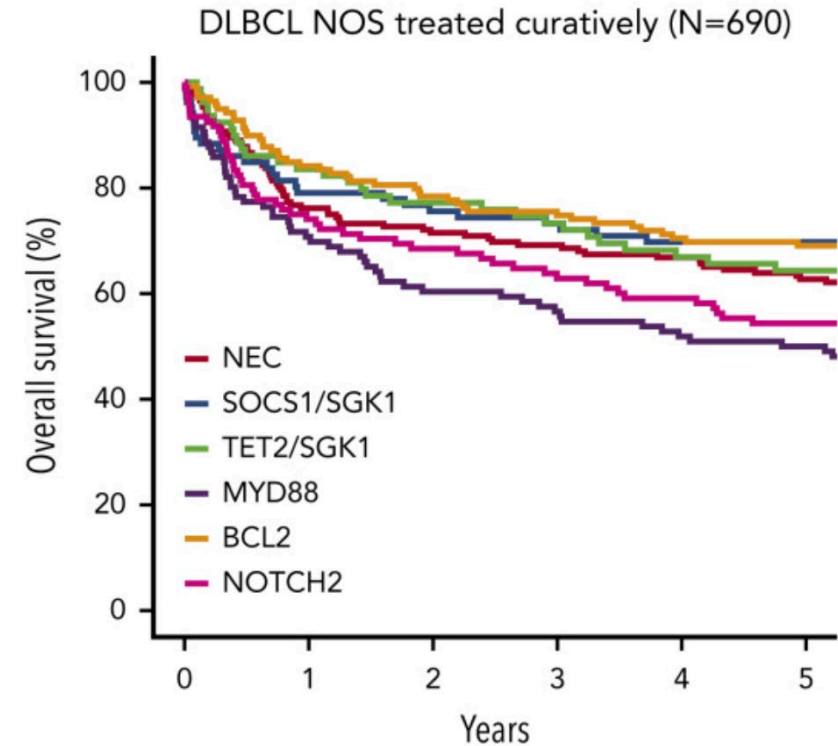
Unsupervised learning

Unsupervised learning



Het model leert op basis van groepen en patronen

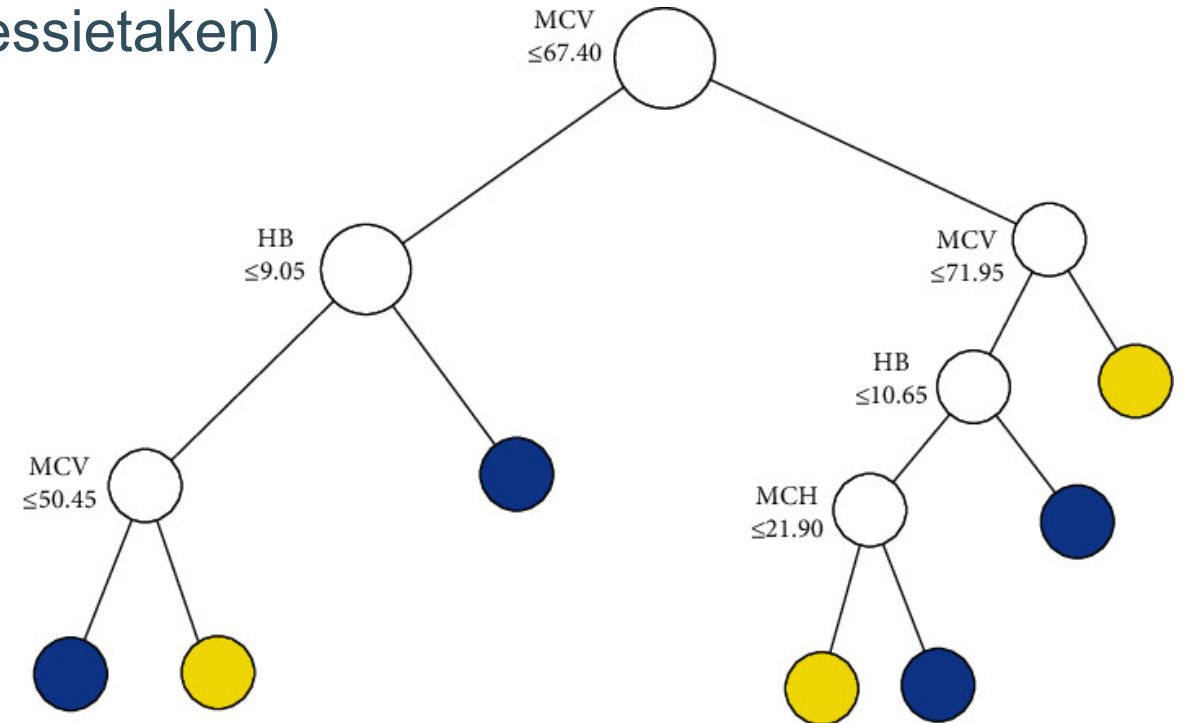
Voorbeeld: Identificatie van gen expressie clusters bij B-cel lymfoom



	0	1	2	3	4	5
NEC	172	131	123	118	114	107
SOCS1/SGK1	86	68	65	63	60	59
TET2/SGK1	79	66	60	57	52	50
MYD88	106	75	64	60	55	53
BCL2	139	117	109	104	98	96
NOTCH2	108	80	73	67	63	57

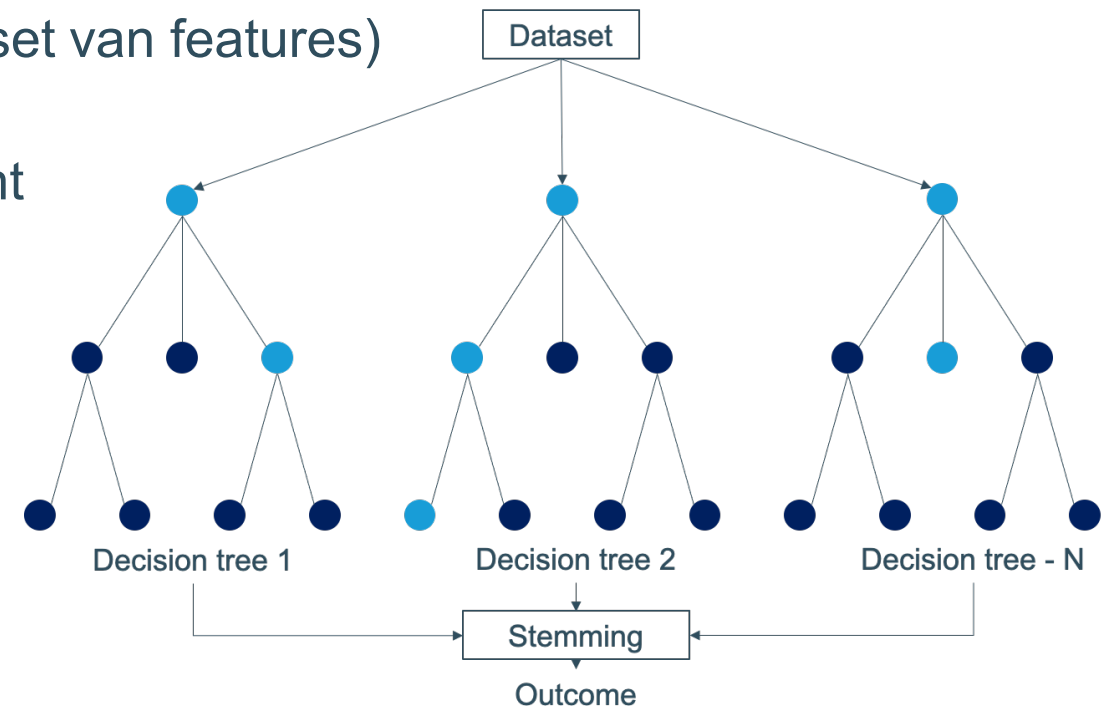
Beslissingsbomen (decision trees)

- Supervised learning (classificatie en regressietaken)
- Structuur
 - Root Node → startpunt
 - Branches → beslissingsregels
 - Leaf Node → outcome
- Voordelen
 - Visualisatie
 - Numerische en categorische data
 - Weinig data preprocessing
- Nadelen
 - Overfitting van de data
 - Gevoelig aan kleine veranderingen in de dataset

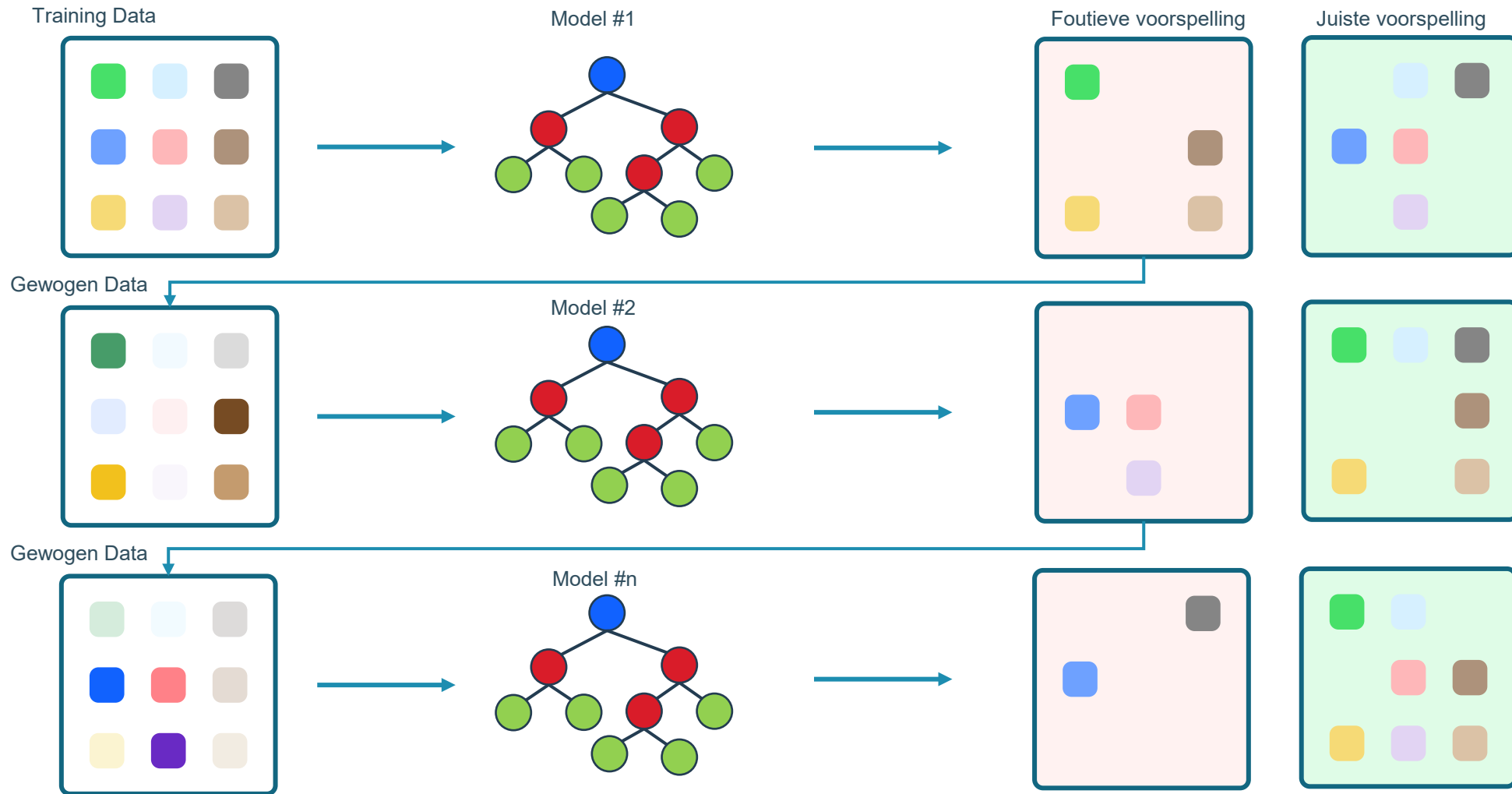


Random forests

- Supervised learning
- Ensemble-methode die meerdere decision trees traint
- Trek willekeurige subset van data (bootstrapping)
- Bouw per subset een decision tree (random subset van features)
- Combineer voorspellingen
 - Classificatie: meerderheid van de bomen wint
 - Regressie: gemiddelde van de bomen wint
- Voordelen
 - Hoge nauwkeurigheid
 - Robuster tegen overfitting
- Nadelen
 - Minder interpreteerbaar
 - Trager

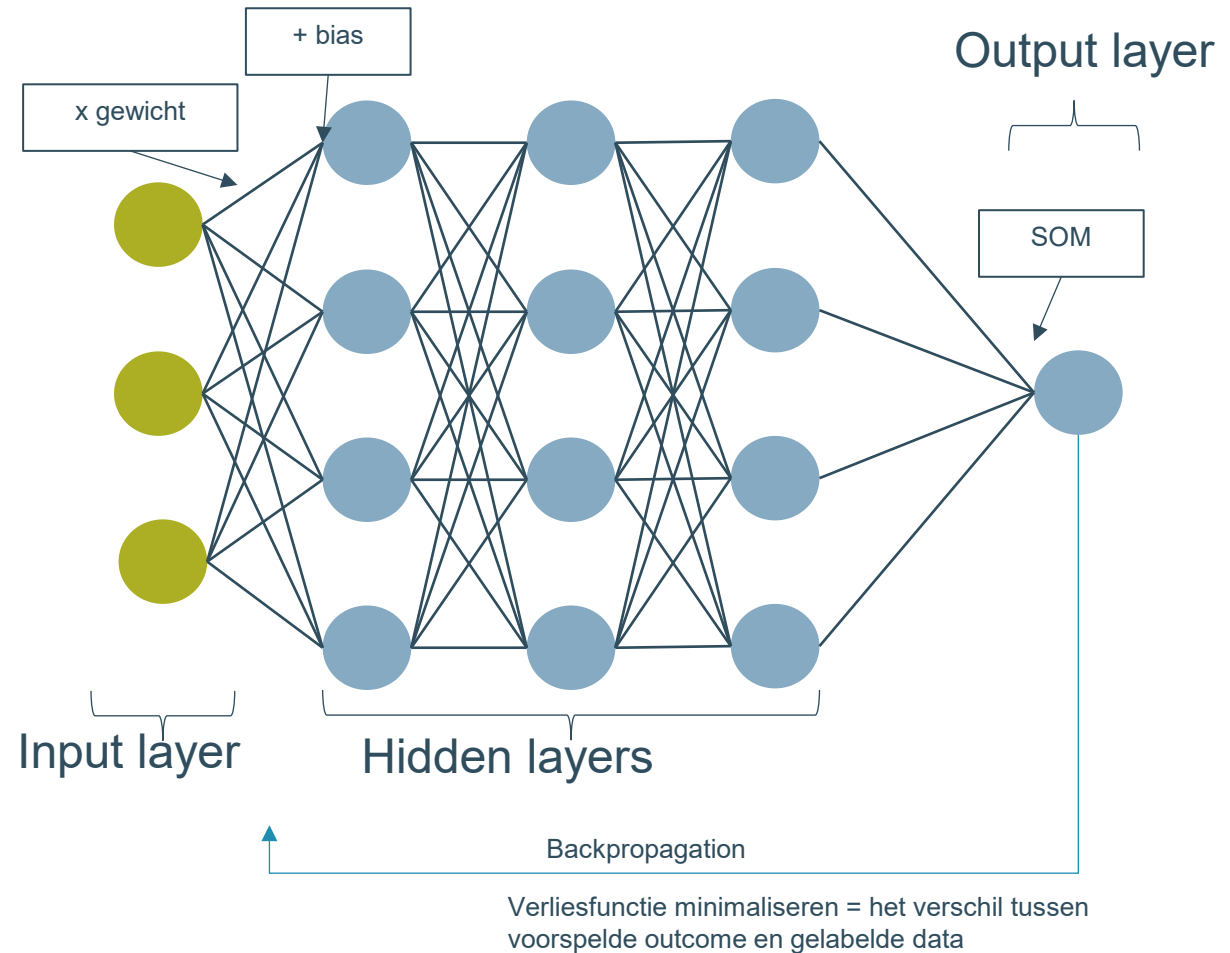


Gradient boosting



Neurale netwerken

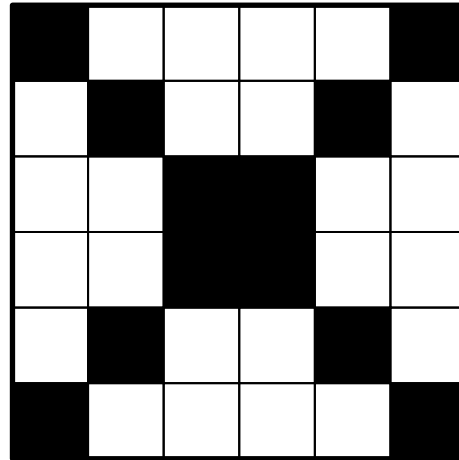
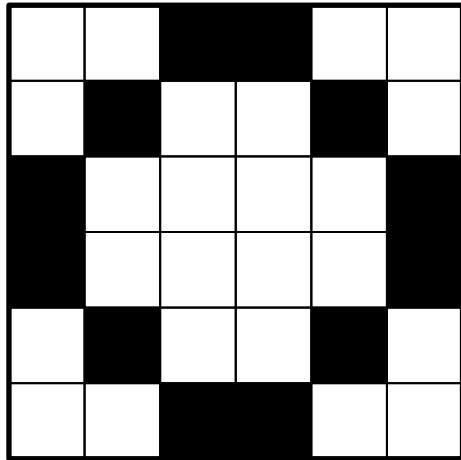
- Imitteert het menselijk brein
- Nodes & Layers
- Lineair regressiemodel
- Feed forward



Convolutionele neurale netwerken (CNN)

- Deep learning netwerk, speciaal voor beeld en patroonherkenning (cytomorfologie)
- Hoe werkt het?
 - Convolutie: filters schuiven over de input en detecteren randen, vormen en patronen
 - Pooling: data wordt 'samengevat' (bv max pooling) → minder data, maar belangrijkste info wordt behouden
 - Fully connected layers: Alle kenmerken worden gecombineerd en een voorspelling wordt gemaakt
- Voordelen
 - Automatische feature-extractie
 - Zeer accuraat voor image-processing
- Nadelen
 - Veel data en rekenkracht
 - Black box effect

Neuraal netwerk

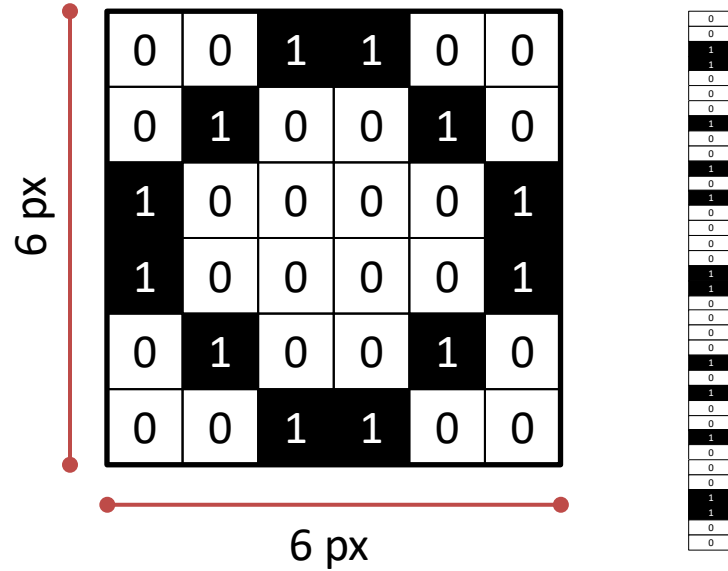


Neuraal netwerk

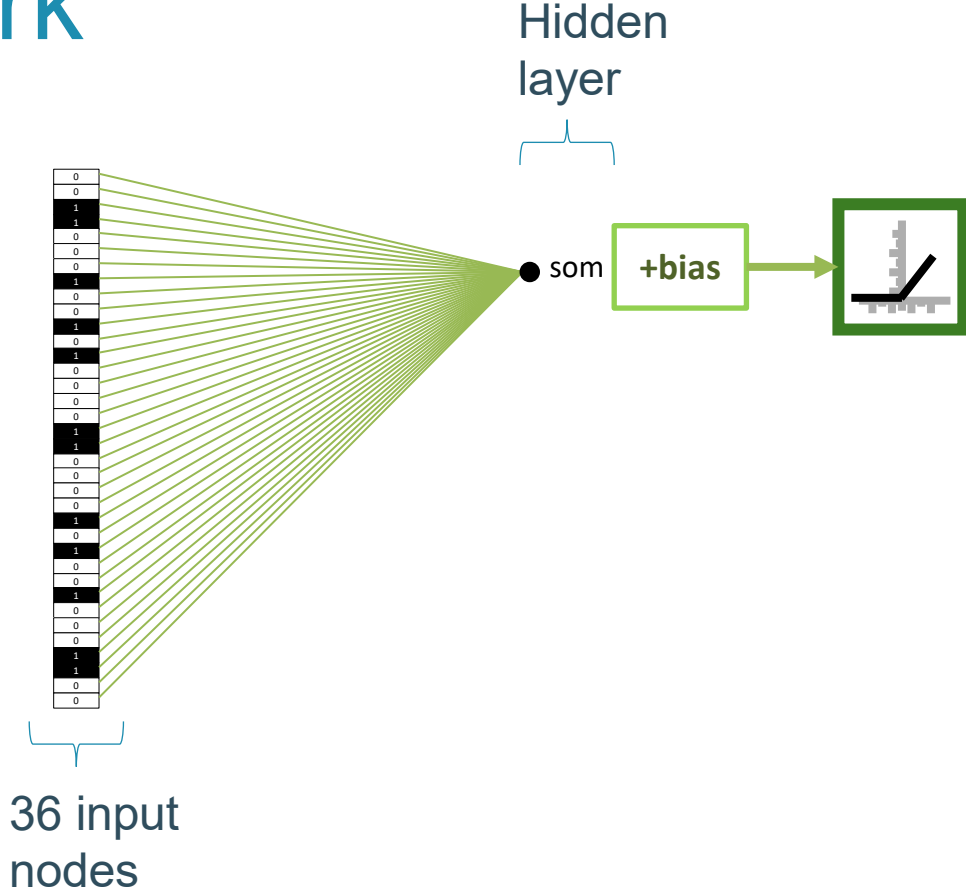
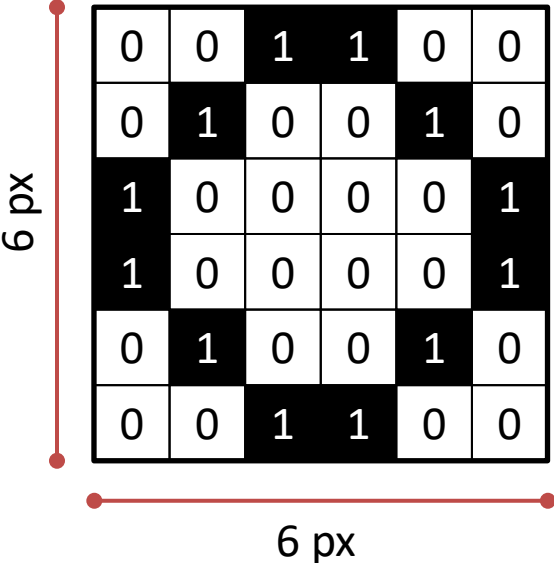
0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1

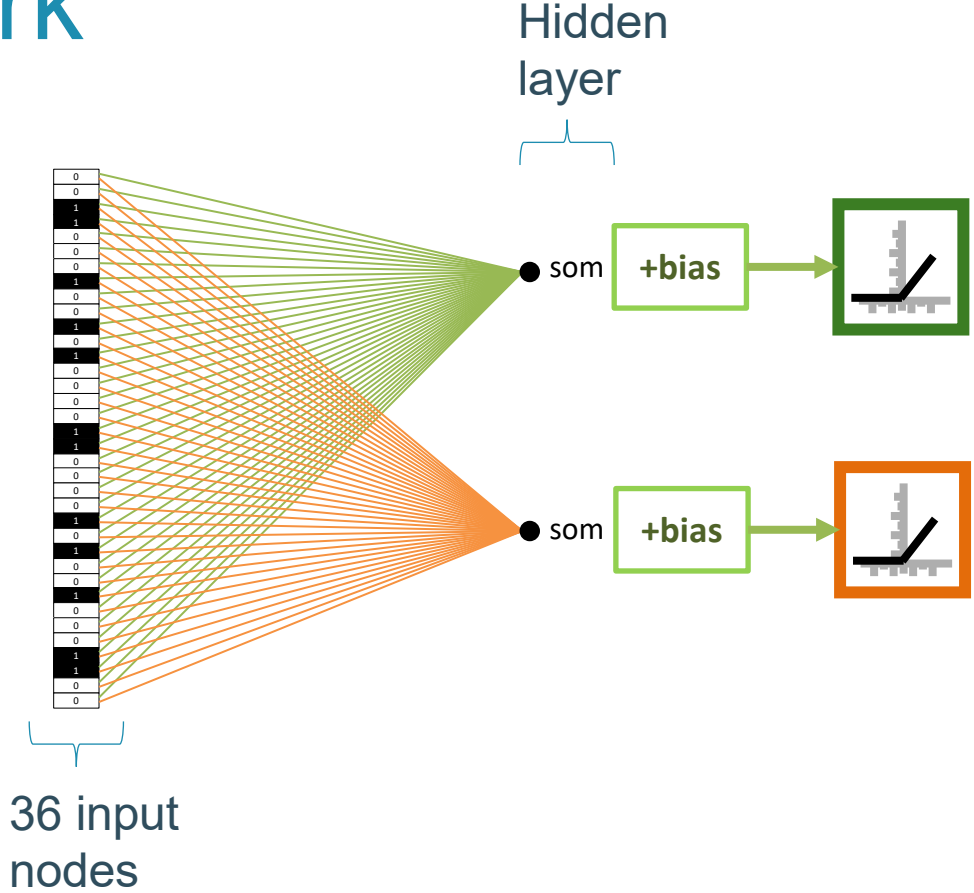
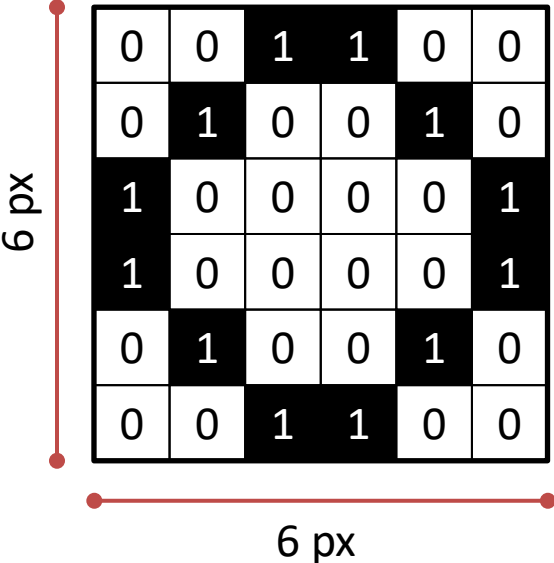
Neuraal netwerk



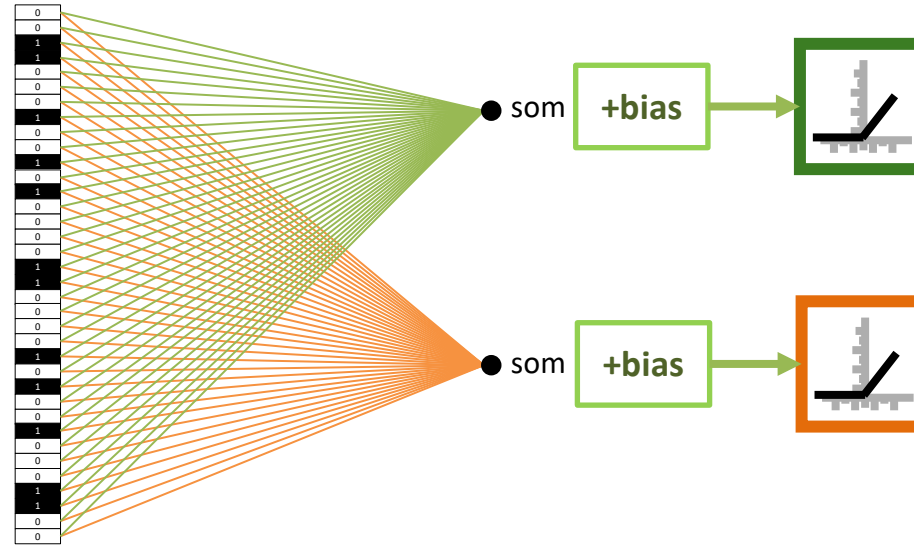
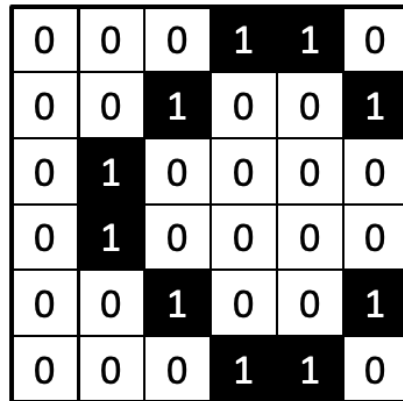
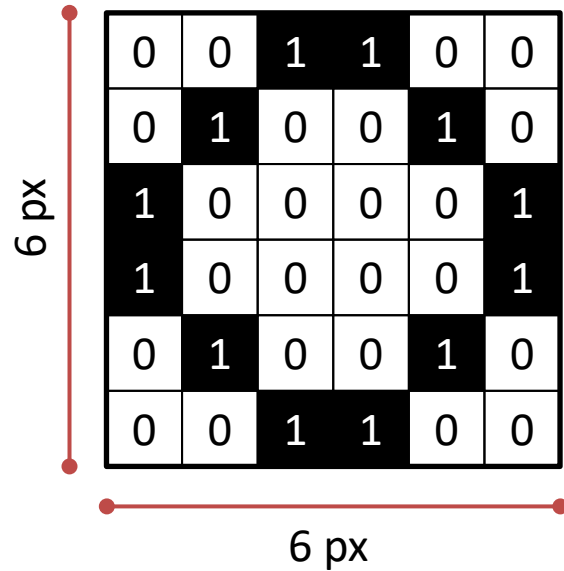
Neuraal netwerk



Neuraal netwerk

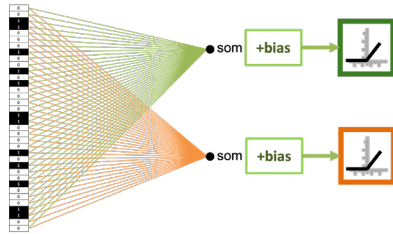


Neuraal netwerk



Convolutioneel Neuraal Netwerk (CNN)

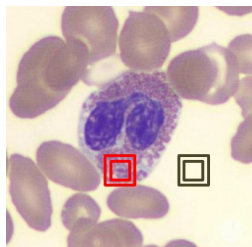
- Verminderen van aantal nodes



- Tolerant kleine verschuivingen van de afbeelding



- Maakt gebruik van correlaties tussen pixels



CNN

Input afbeelding

0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0

Filter (Kernel)

3 px

0	0	1
0	1	0
1	0	0

3 px

CNN

Input afbeelding

0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0

Na training

0	0	1
0	1	0
1	0	0

Backpropagation

1	0	1
0	0	1
1	1	0

Voor training

CNN

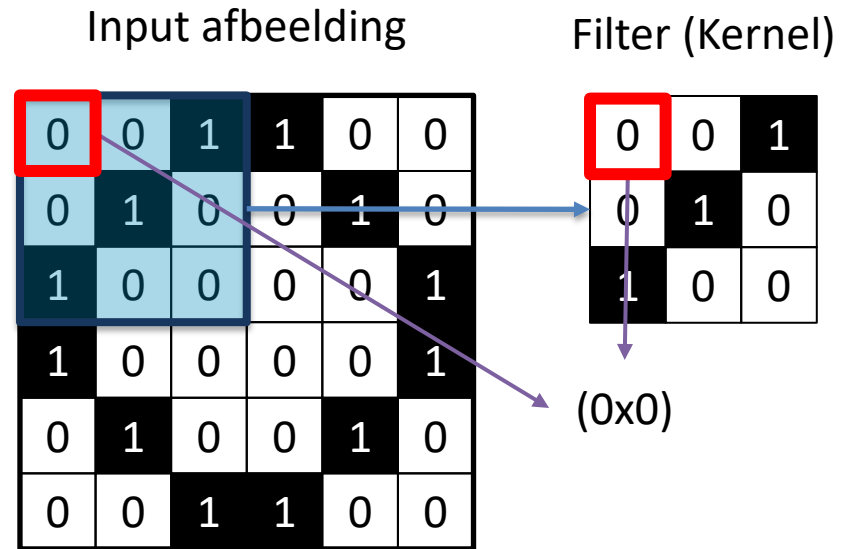
Input afbeelding

0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0

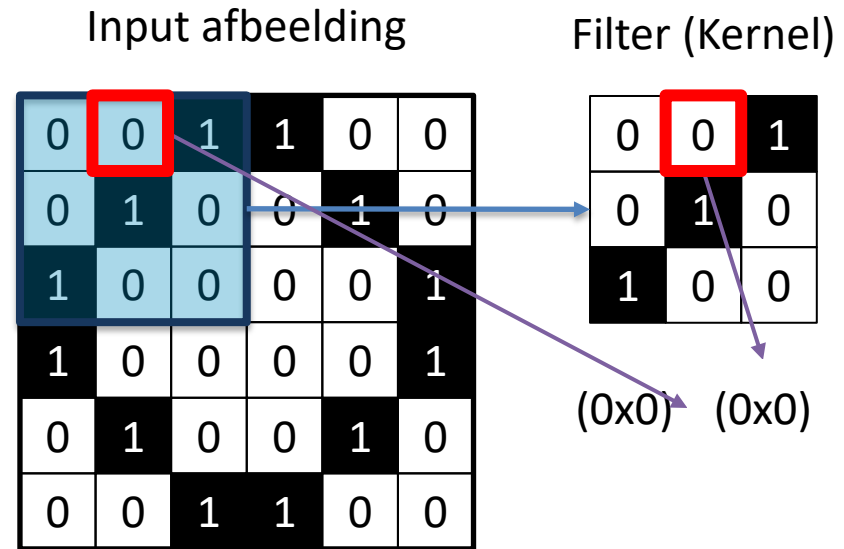
Filter (Kernel)

0	0	1
0	1	0
1	0	0

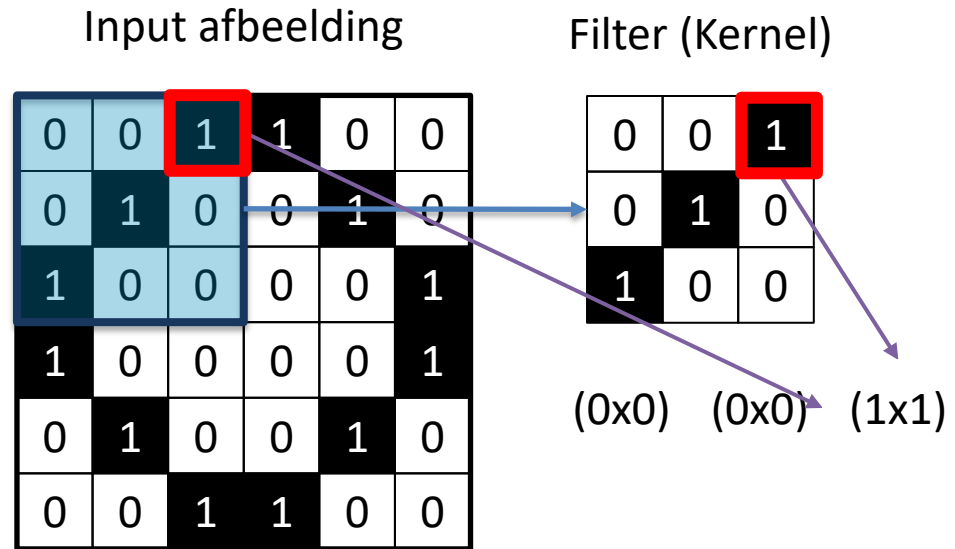
CNN



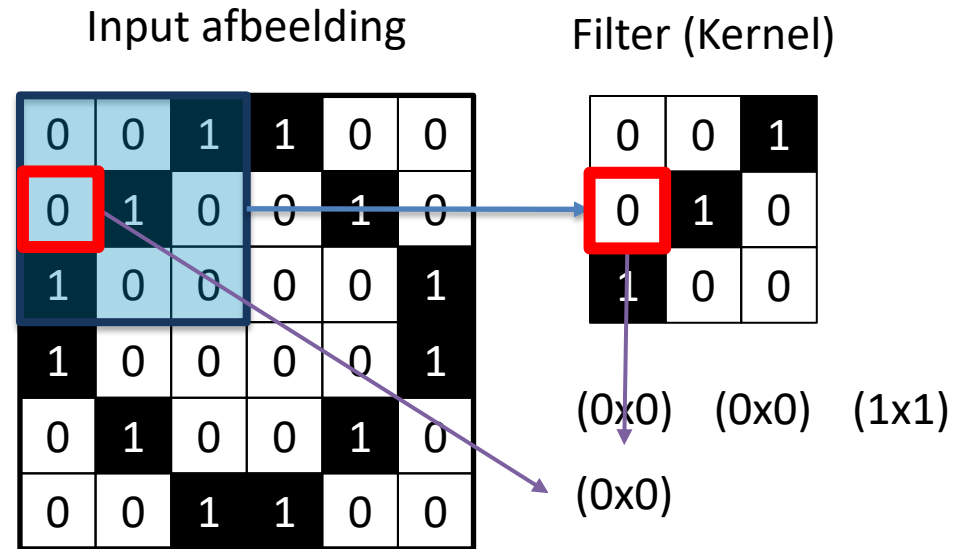
CNN



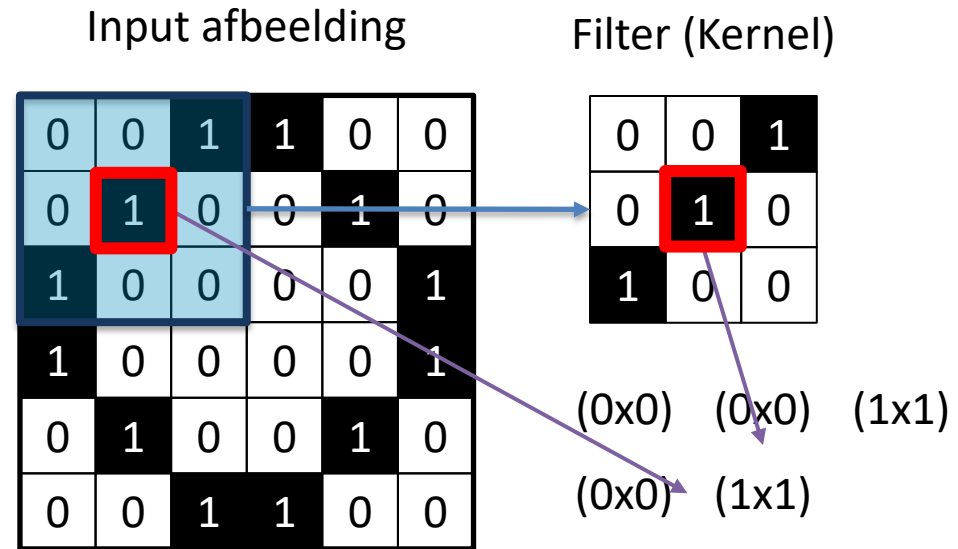
CNN



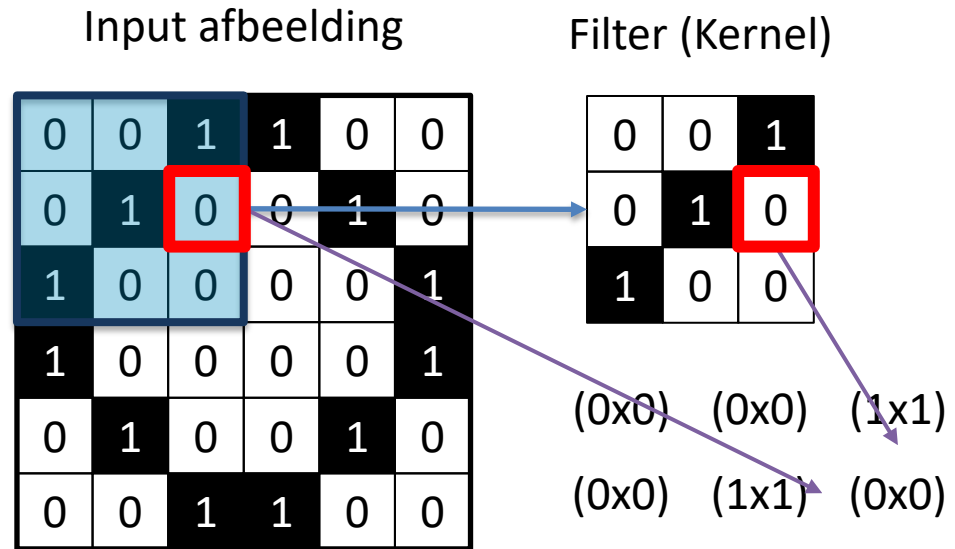
CNN



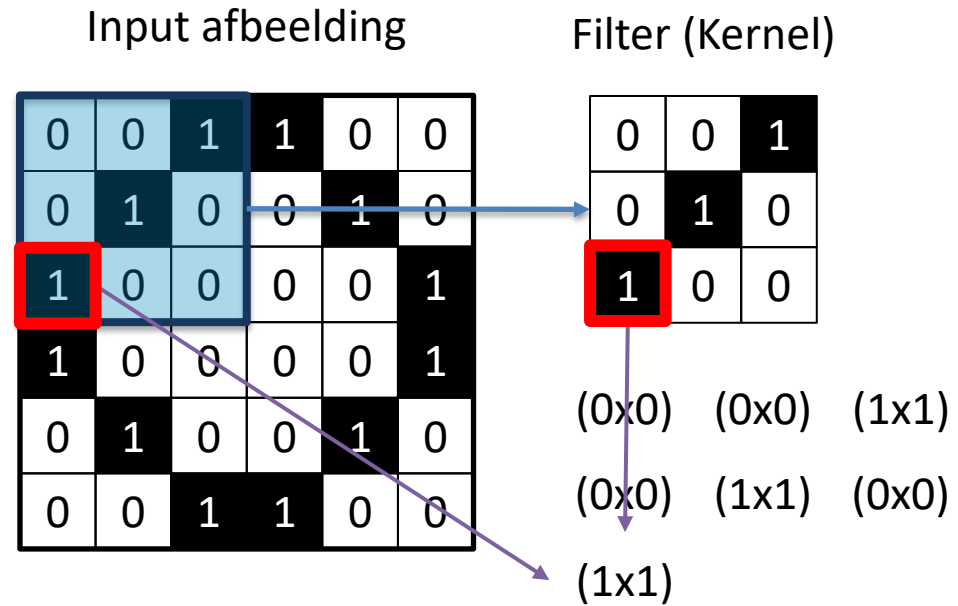
CNN



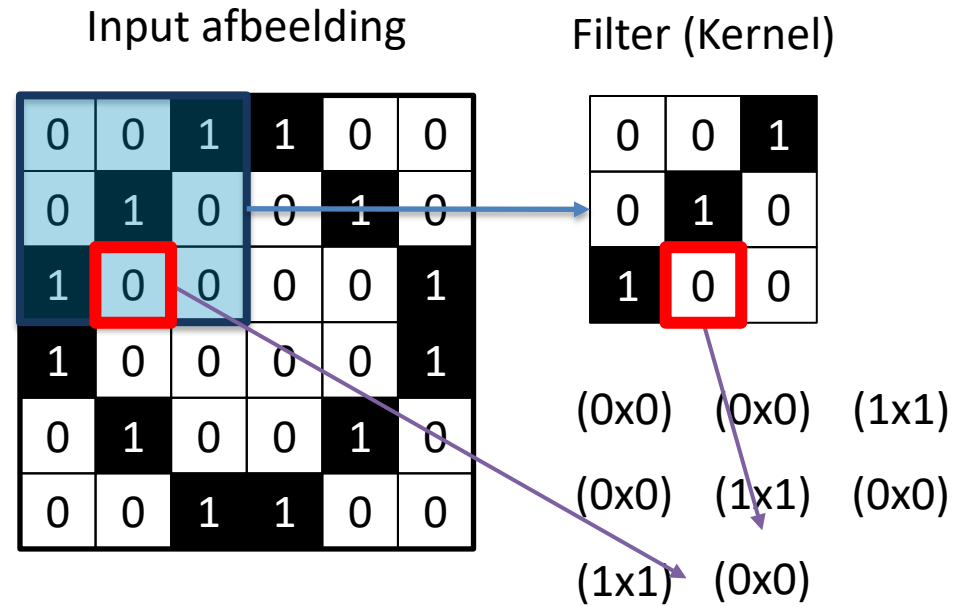
CNN



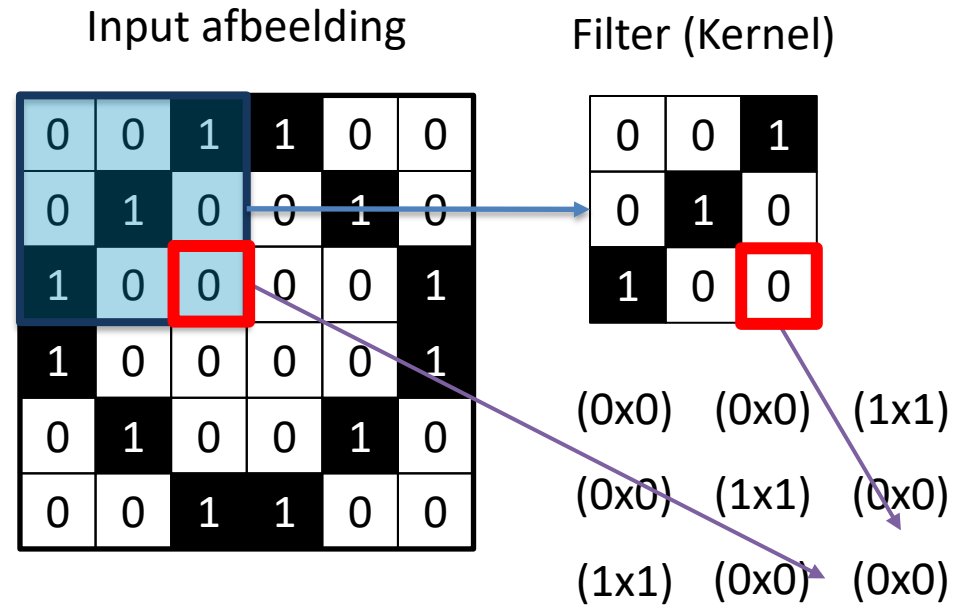
CNN



CNN



CNN



CNN

Input afbeelding

0	0	1	1	0	0
0	1	0	0	1	0
1	0	0	0	0	1
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0

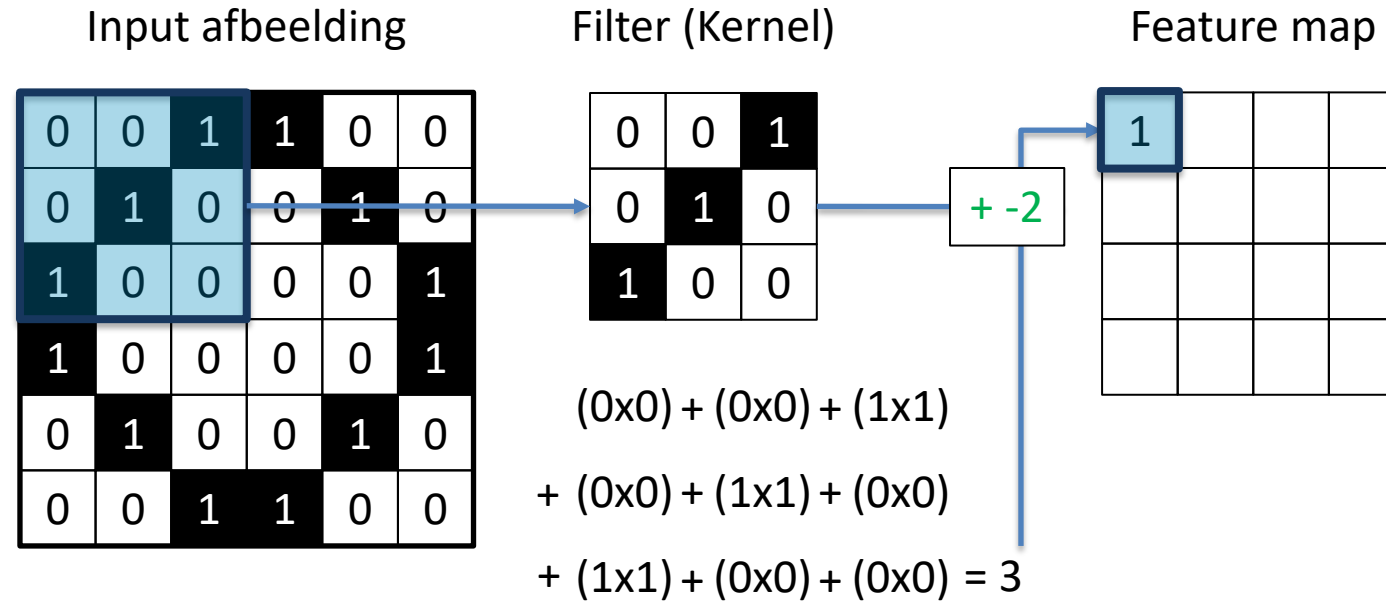
Filter (Kernel)

0	0	1
0	1	0
1	0	0

$$\begin{aligned} & (0 \times 0) + (0 \times 0) + (1 \times 1) \\ & + (0 \times 0) + (1 \times 1) + (0 \times 0) \\ & + (1 \times 1) + (0 \times 0) + (0 \times 0) = 3 \end{aligned}$$

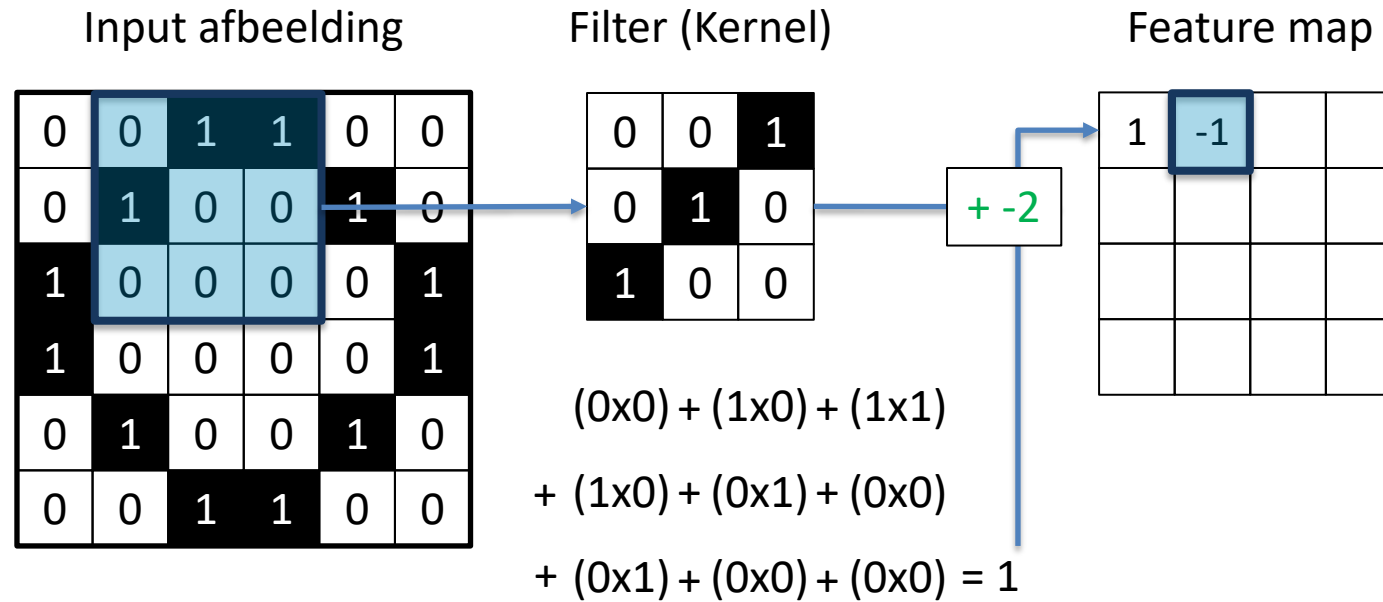
De filter is geconvuleerd met de input

CNN



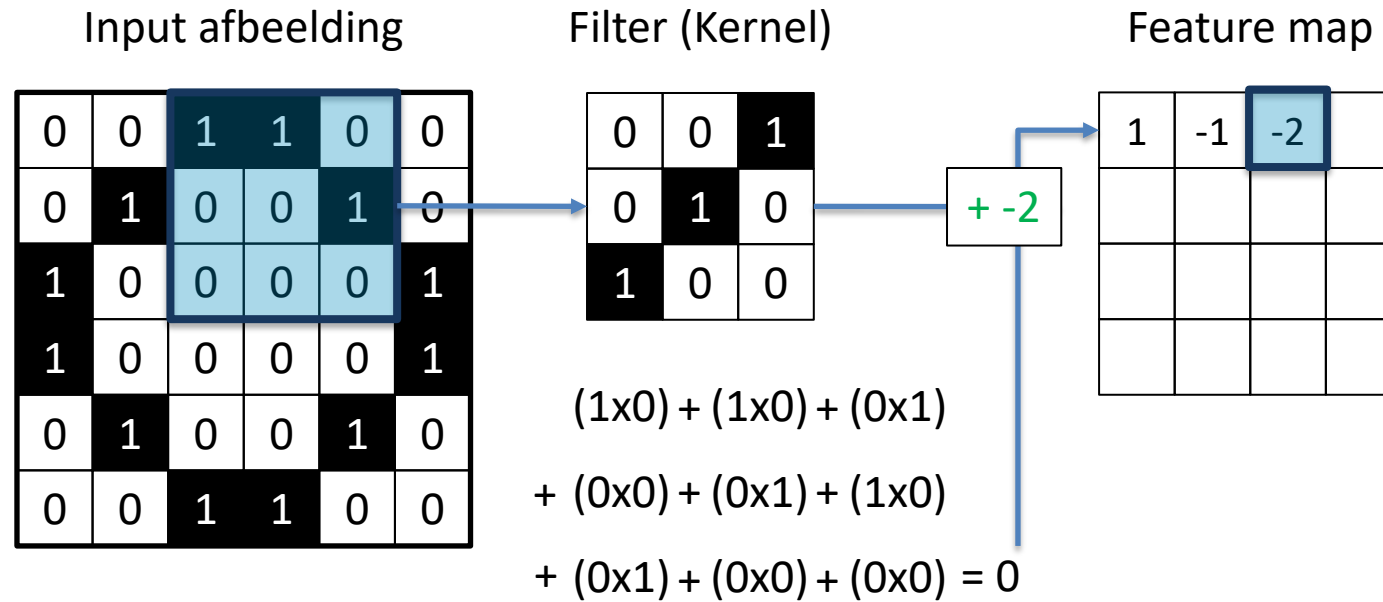
De filter is geconvuleerd met de input

CNN



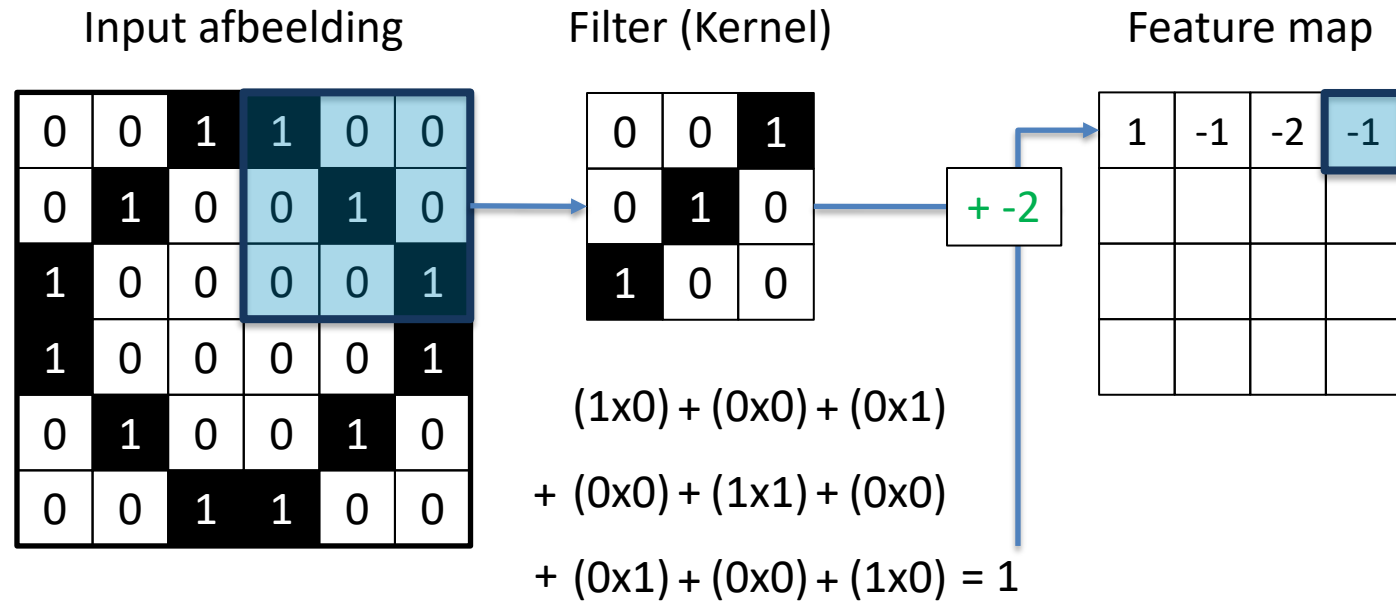
De filter is geconvuleerd met de input

CNN



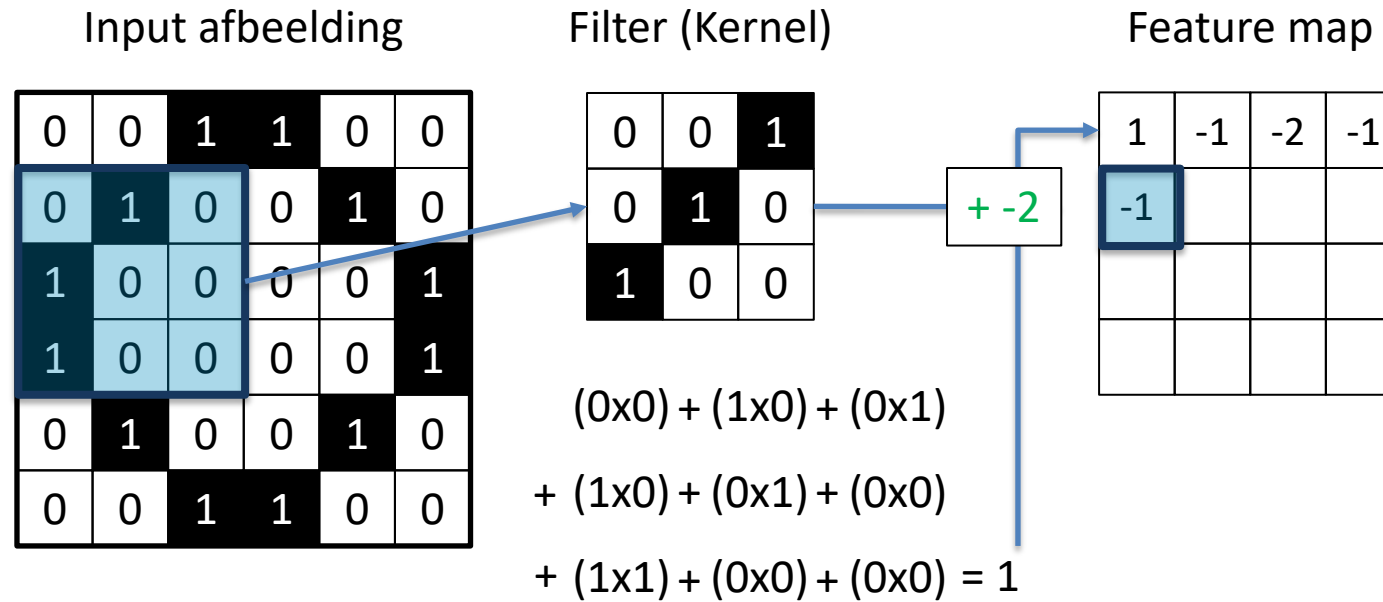
De filter is geconvuleerd met de input

CNN



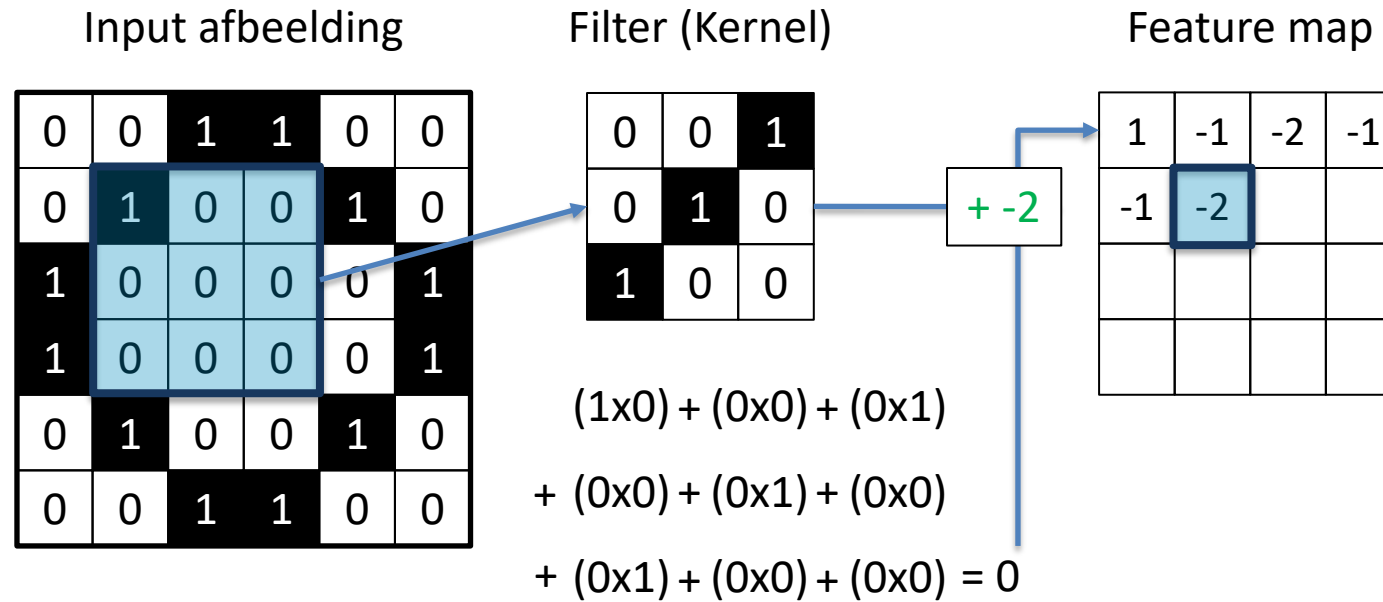
De filter is geconvuleerd met de input

CNN



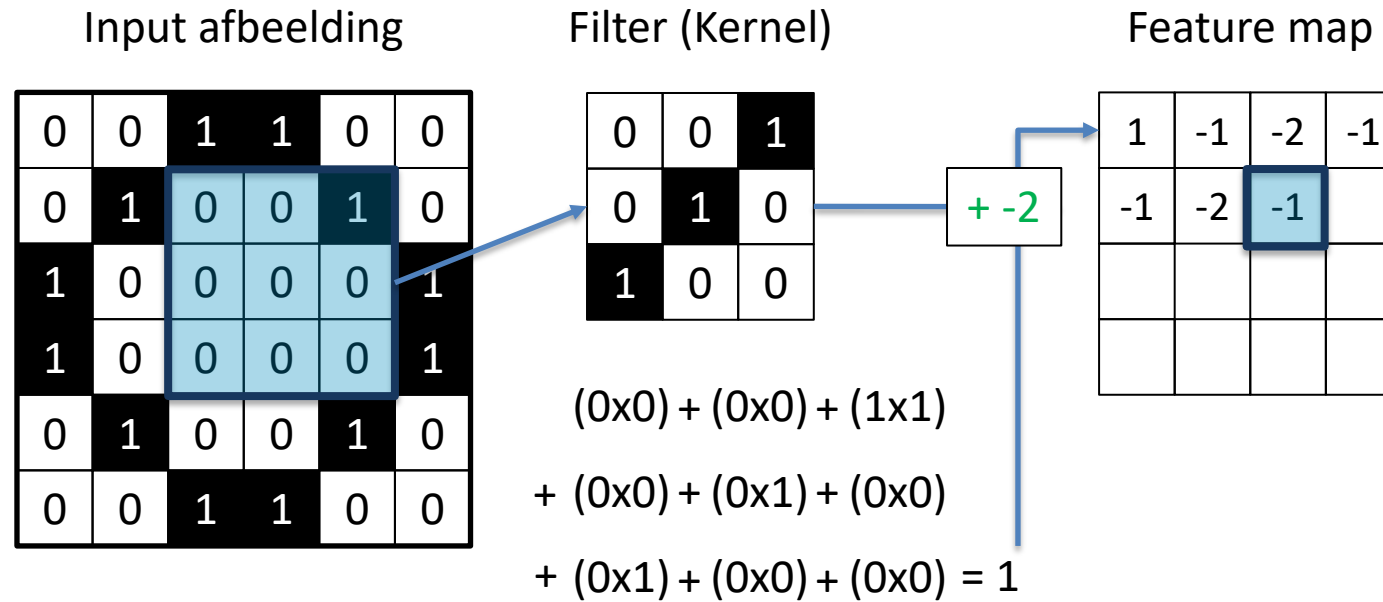
De filter is geconvuleerd met de input

CNN



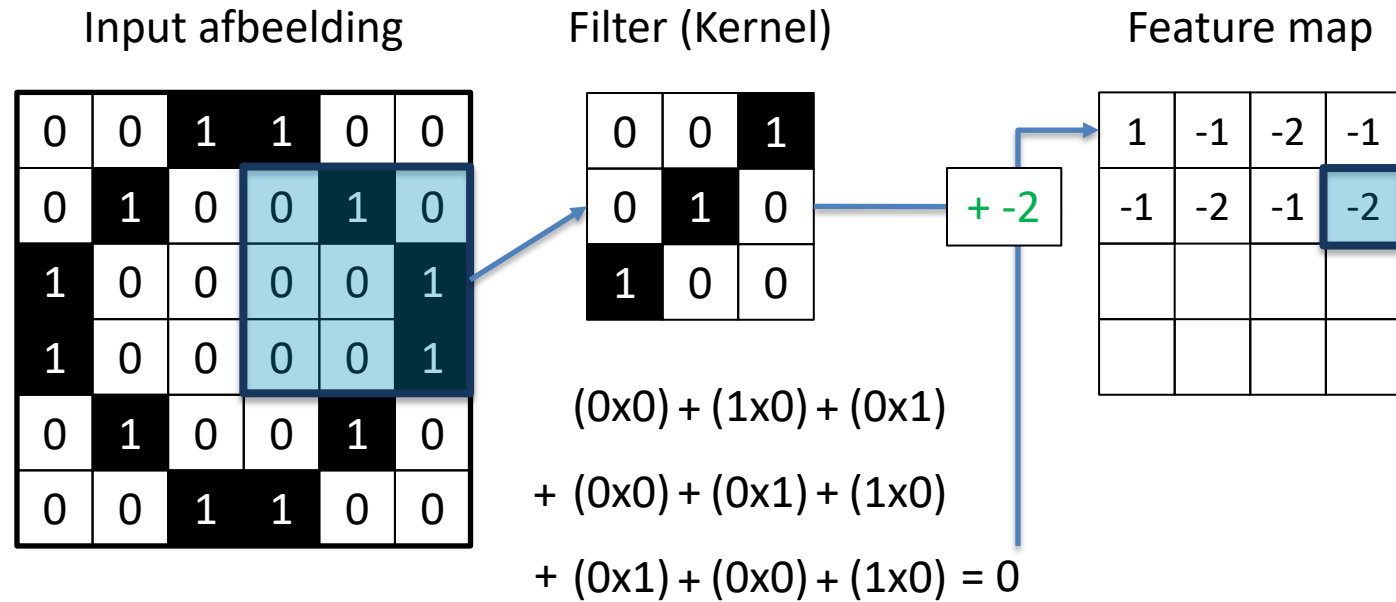
De filter is geconvuleerd met de input

CNN



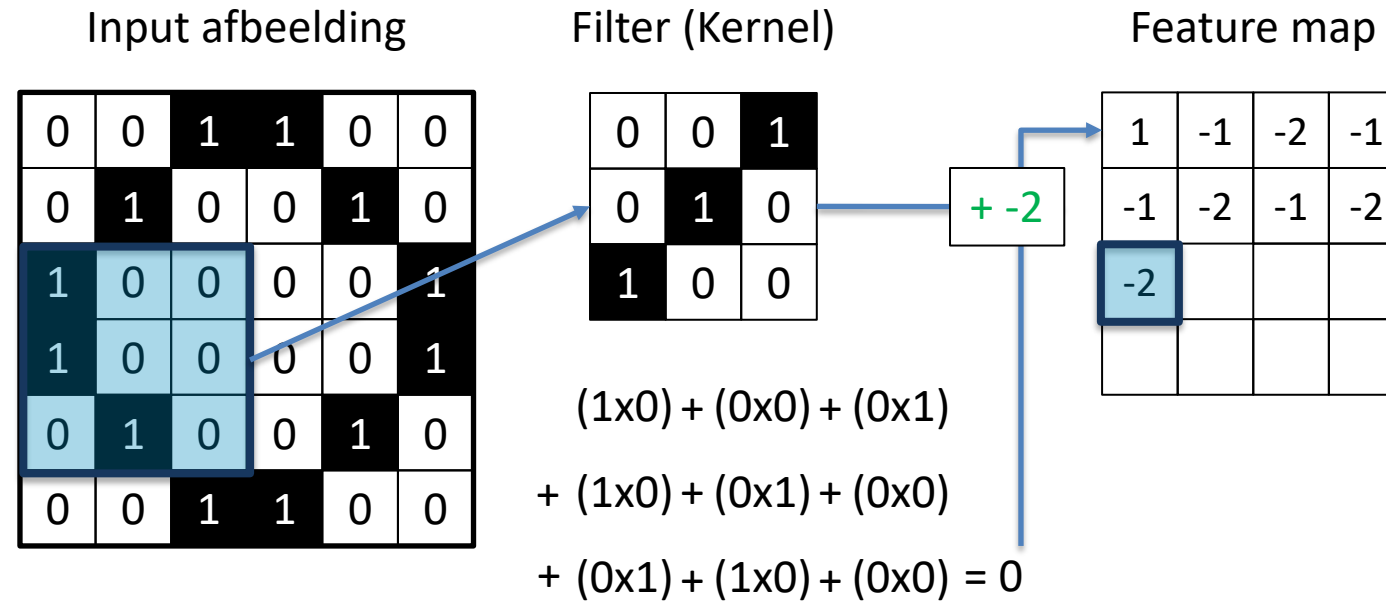
De filter is geconvuleerd met de input

CNN



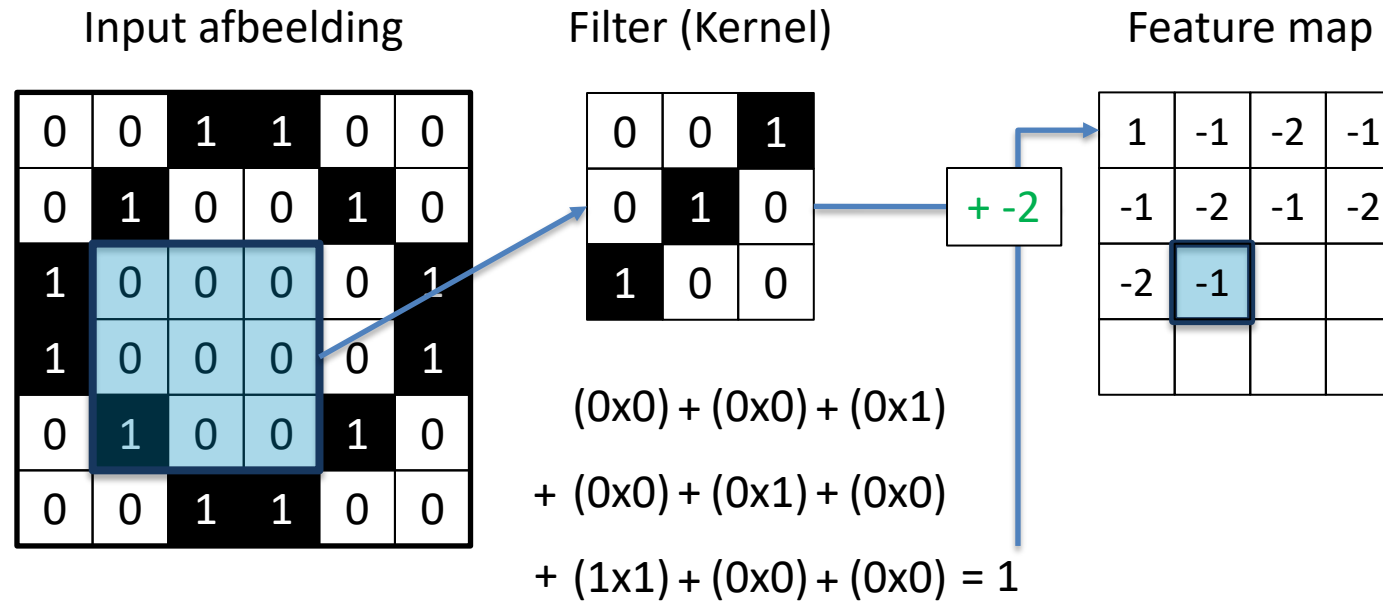
De filter is geconvuleerd met de input

CNN



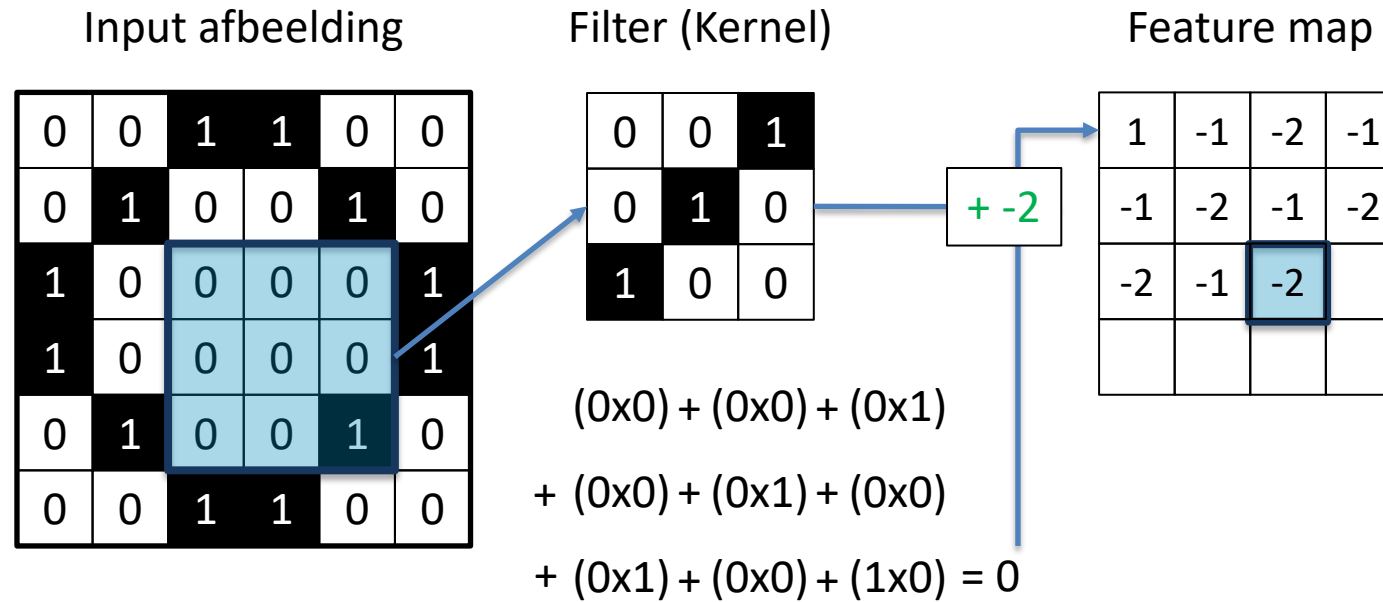
De filter is geconvuleerd met de input

CNN



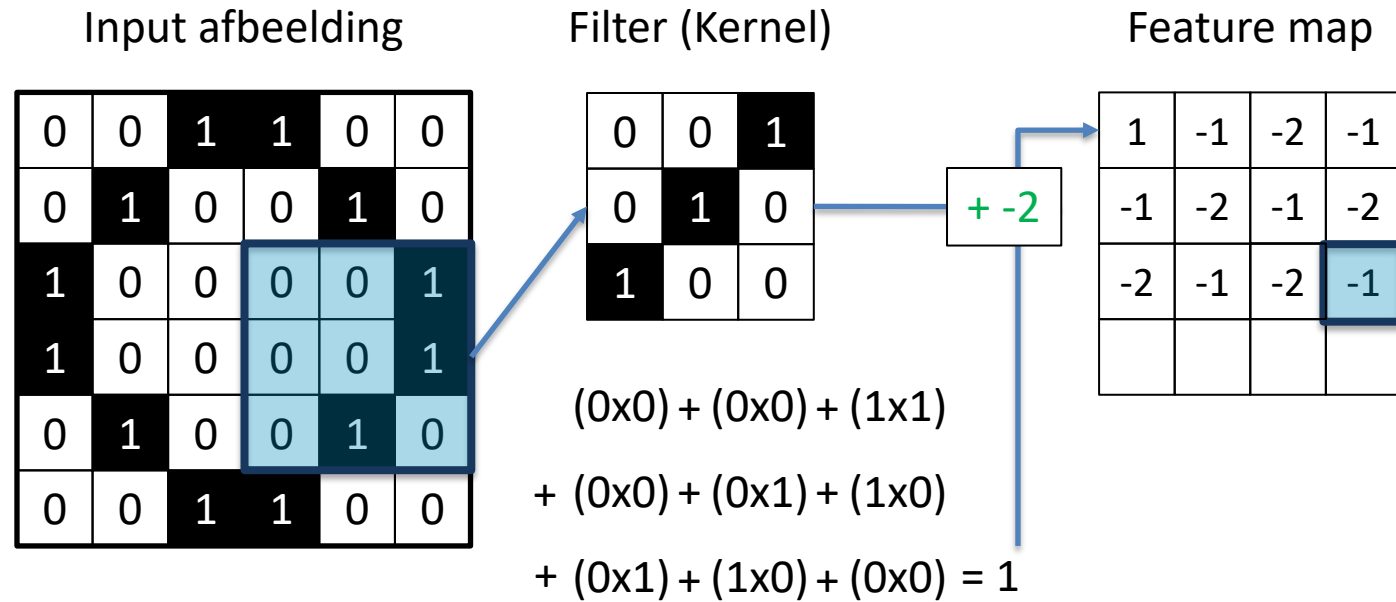
De filter is geconvuleerd met de input

CNN



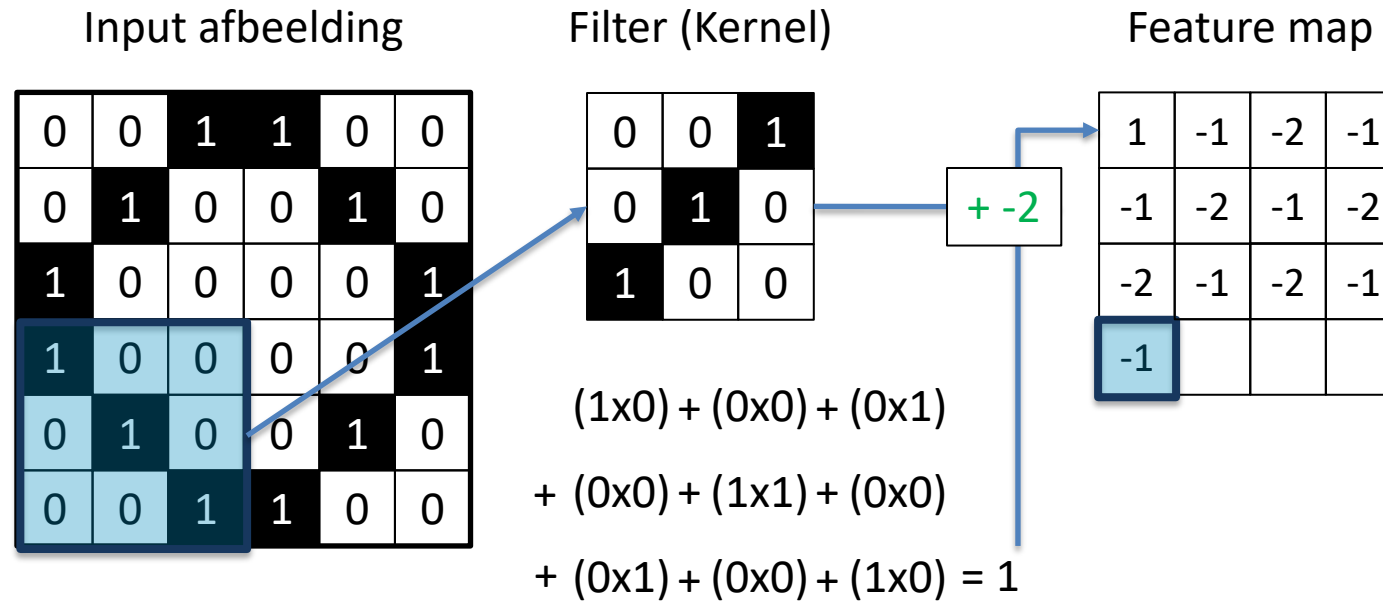
De filter is geconvuleerd met de input

CNN



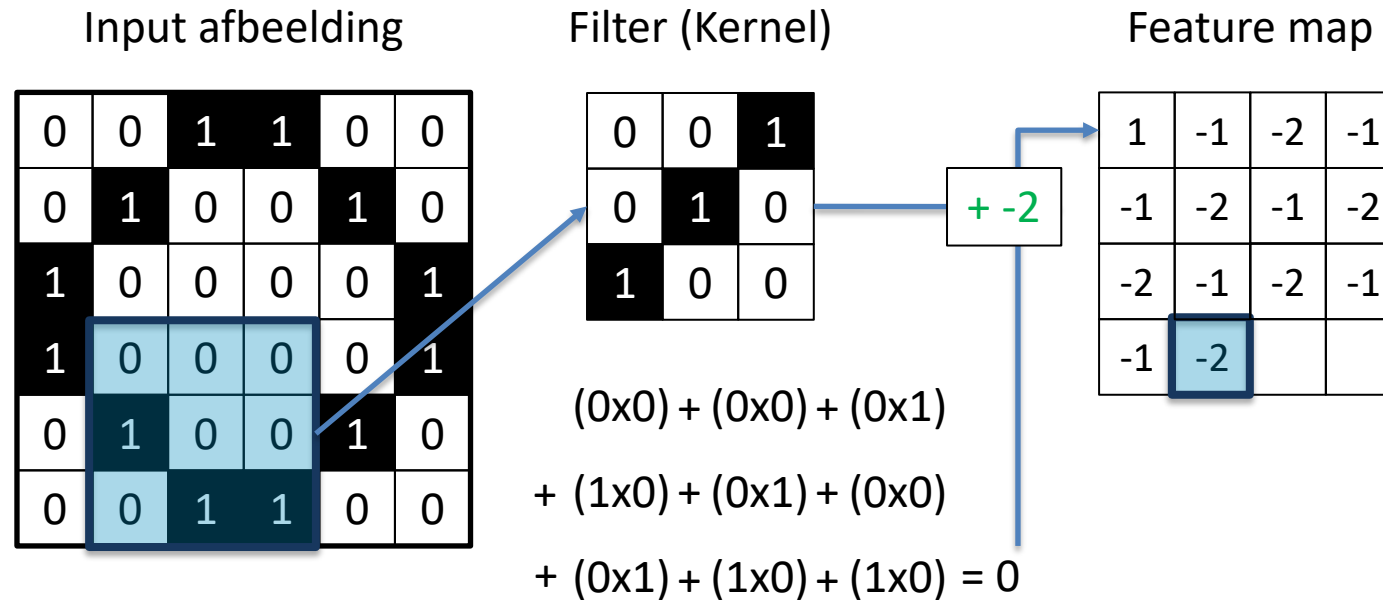
De filter is geconvuleerd met de input

CNN



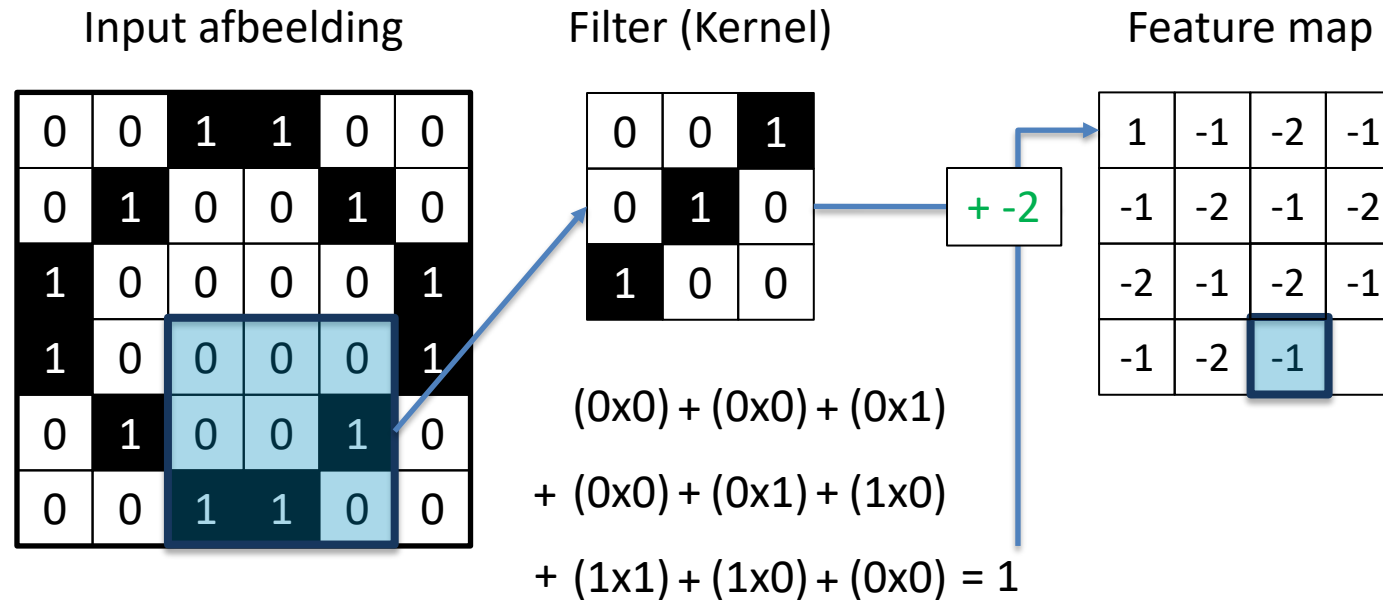
De filter is geconvuleerd met de input

CNN



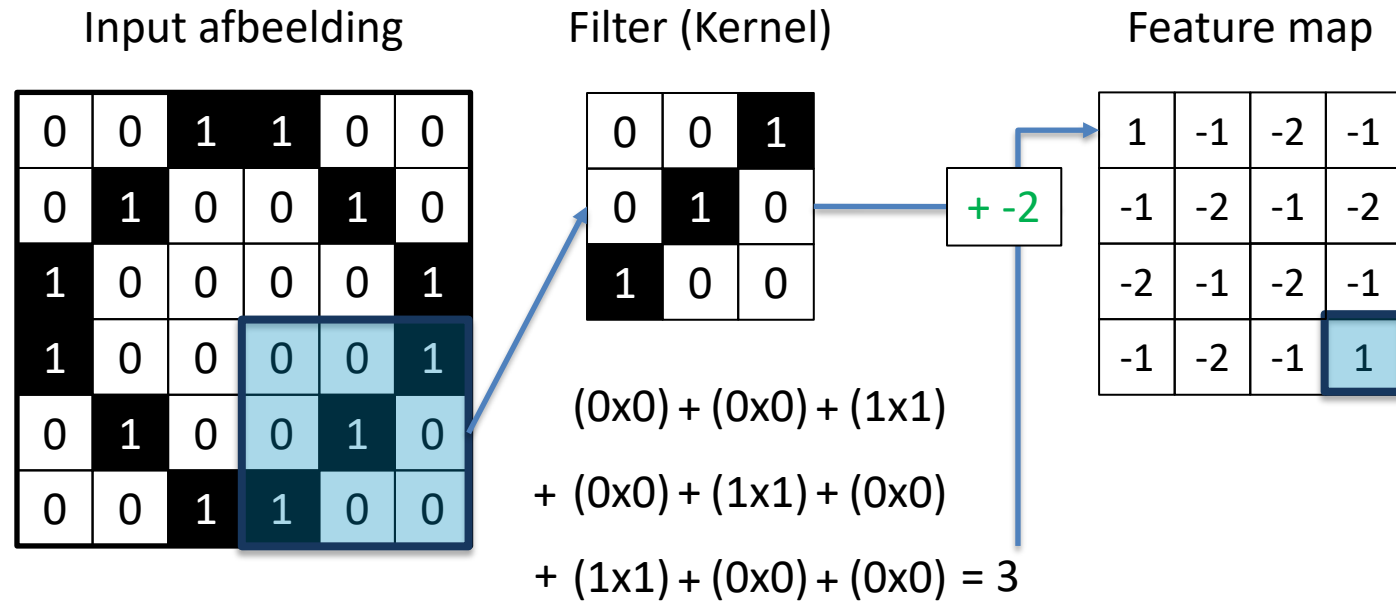
De filter is geconvuleerd met de input

CNN



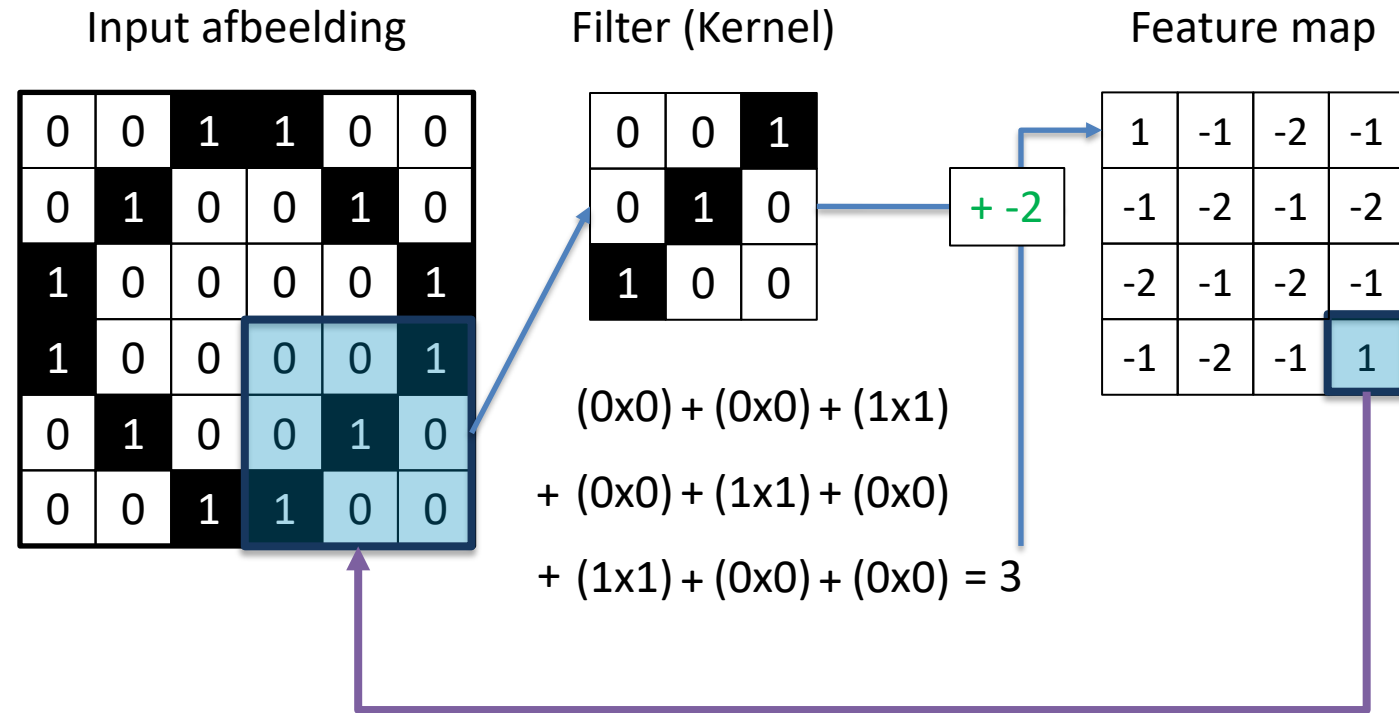
De filter is geconvuleerd met de input

CNN



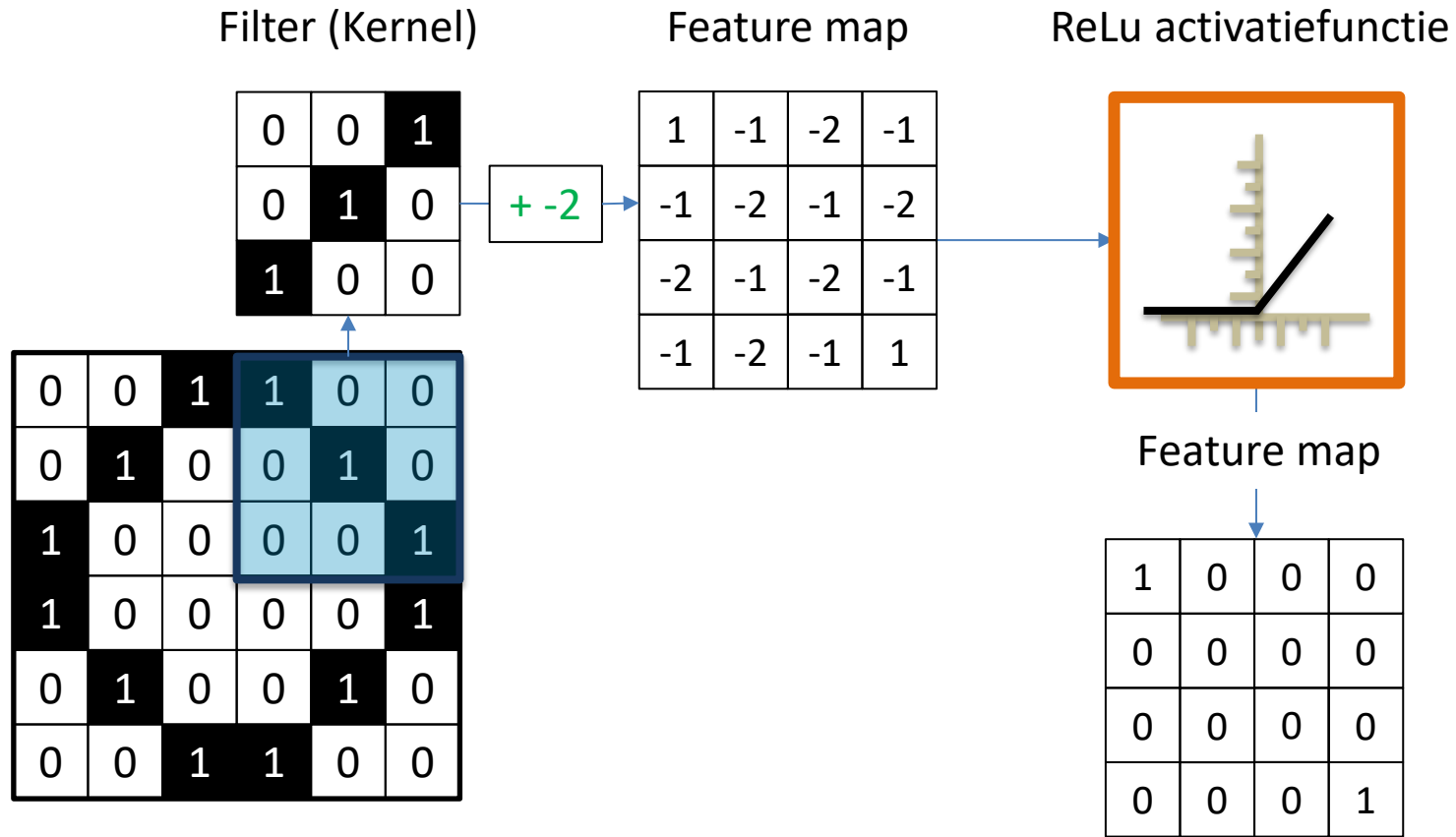
De filter is geconvuleerd met de input

CNN

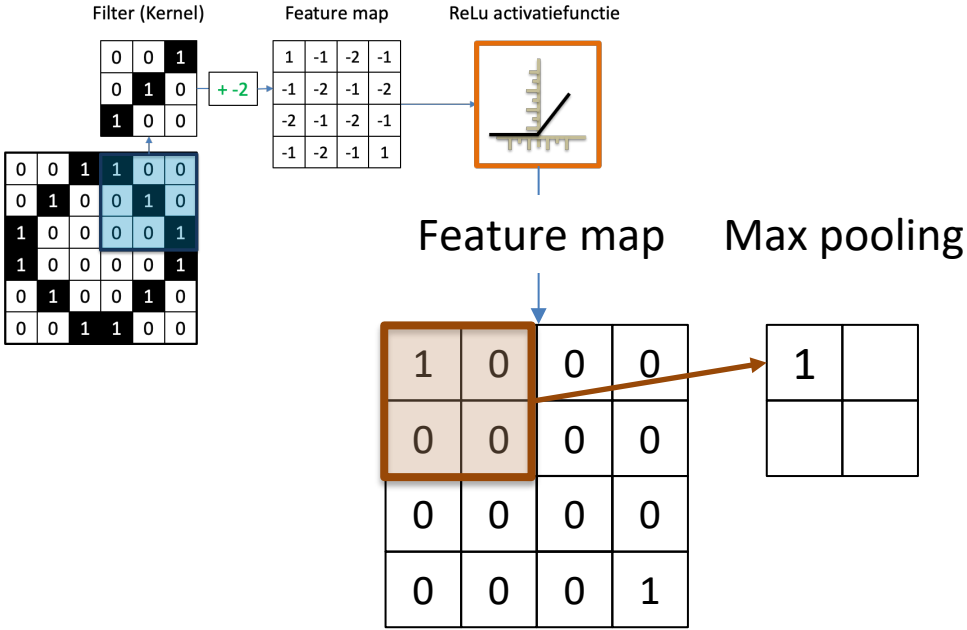


Correlatie tussen de pixels

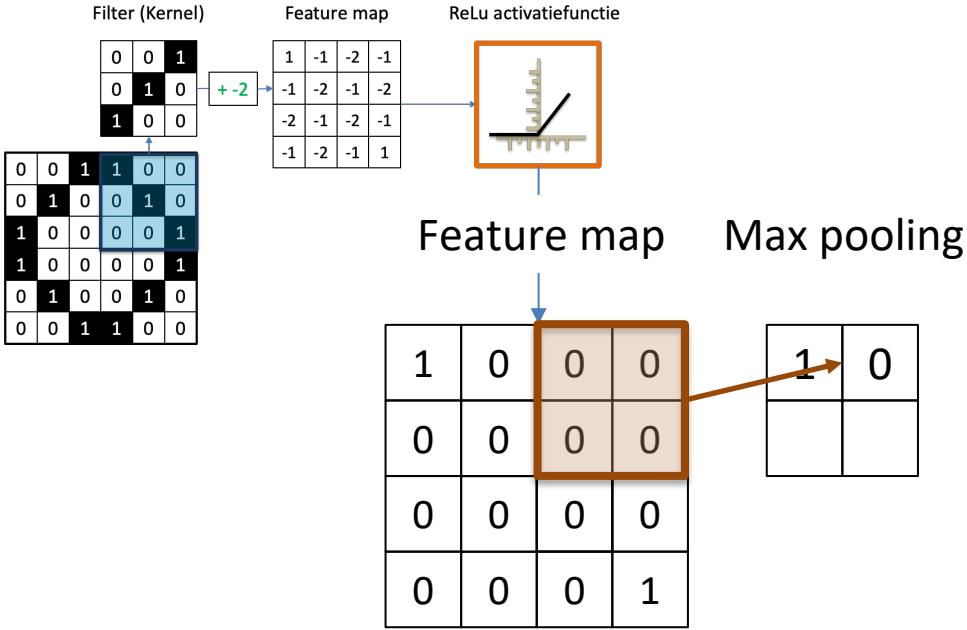
CNN



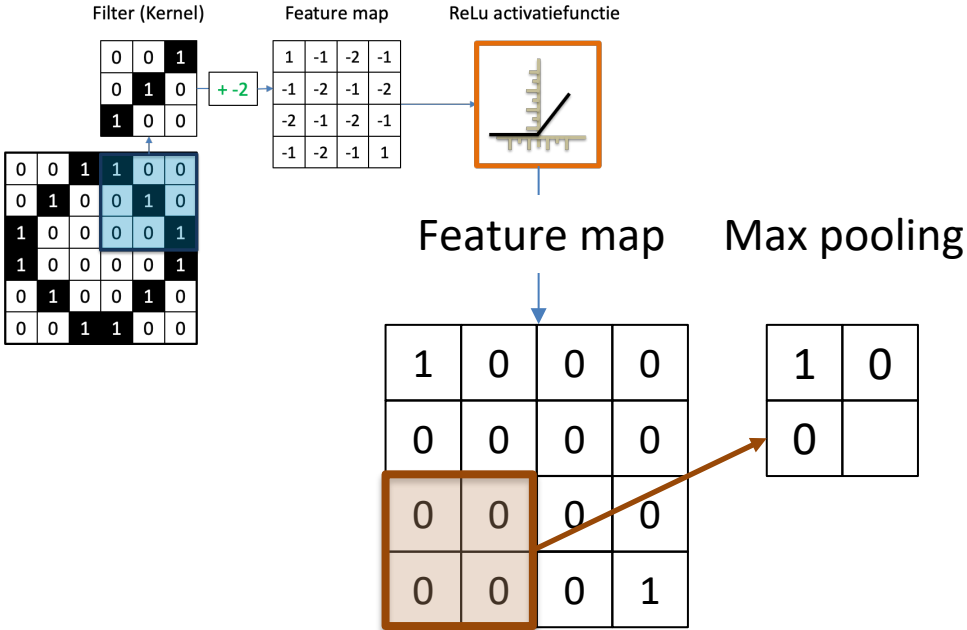
CNN



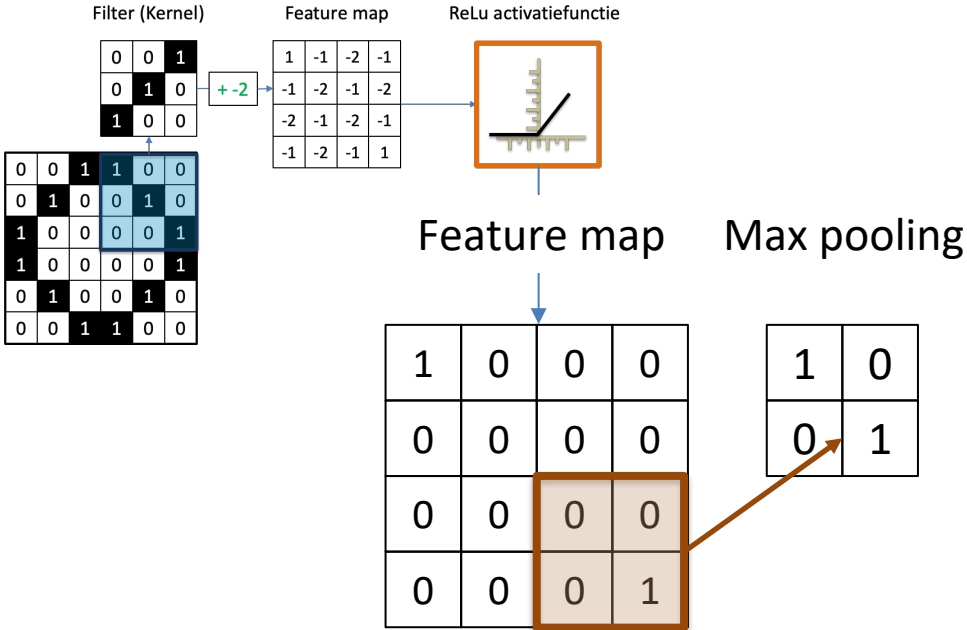
CNN



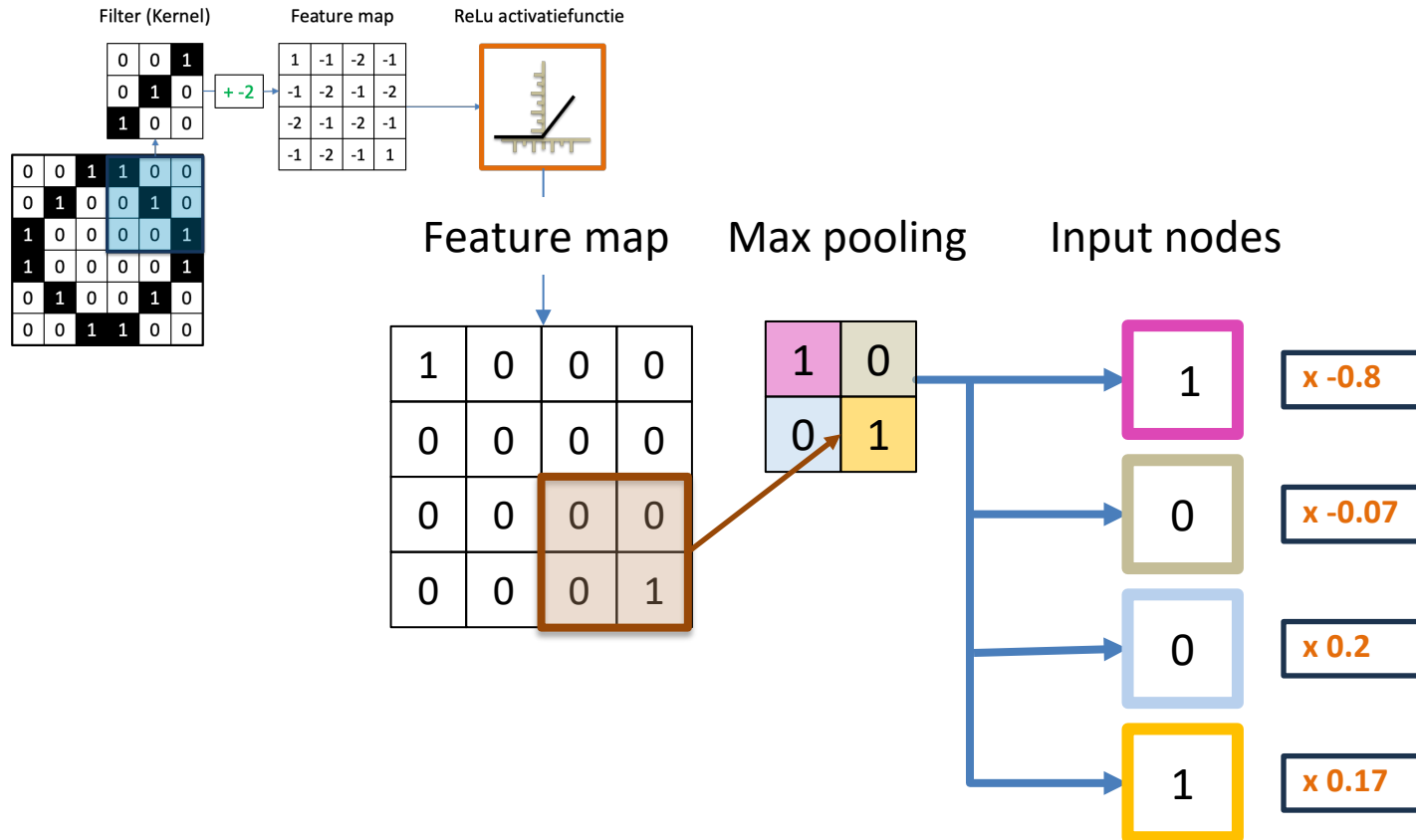
CNN



CNN

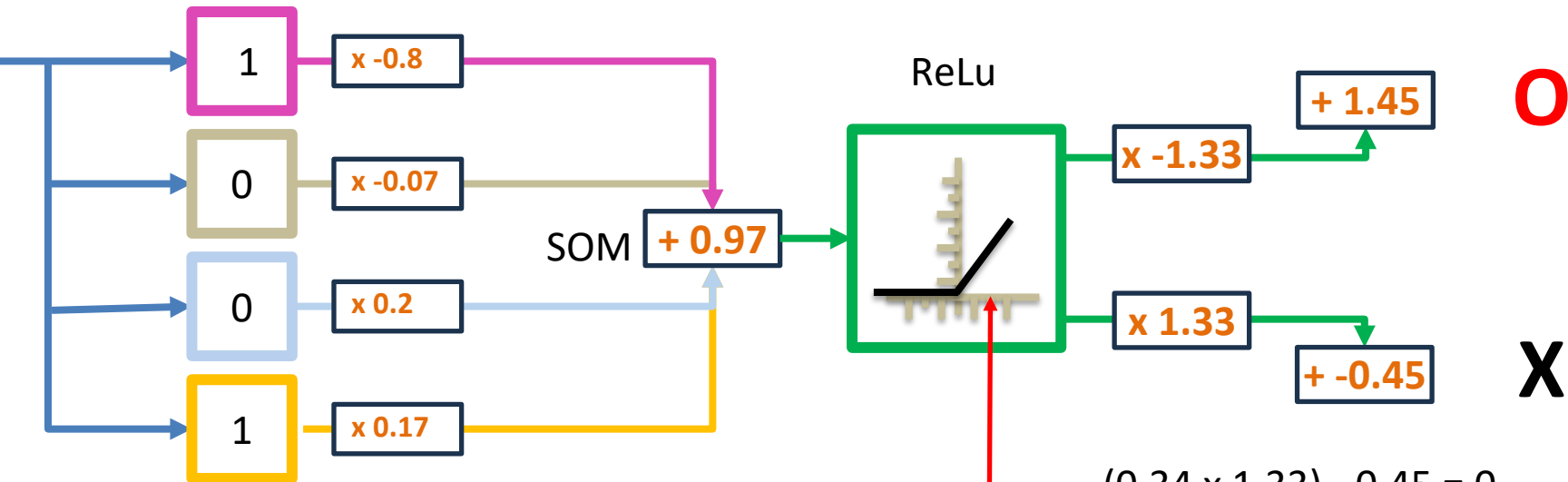


CNN



CNN

Input nodes



$$(0.34 \times -1.33) + 1.45 = 1$$

O

X

$$(0.34 \times 1.33) - 0.45 = 0$$

$$(1 \times -0.8) + (0 \times -0.07) + (0 \times 0.2) + (1 \times 0.17) + 0.97 = 0.34$$

$$f(x) = \max(0, x)$$

$$f(0.34) = \max(0, 0.34) = 0.34$$

Overzicht

Inleiding

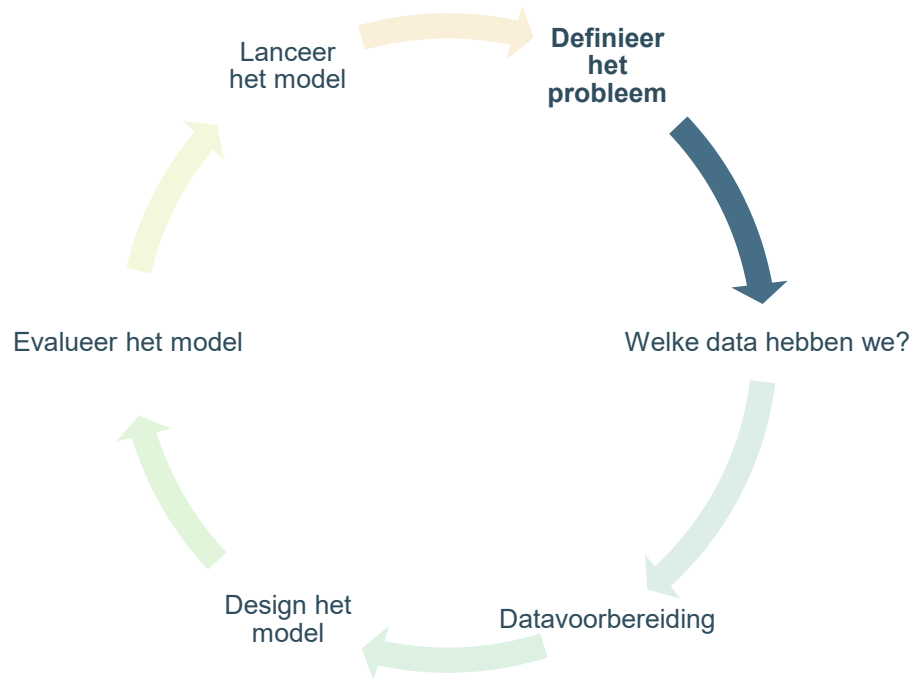
- Toenemende interesse in AI
- Welke AI modellen bestaan er?
- **Datamining in machine learning**

Onderzoeksvragen

1. Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?
2. Kan AI de diagnostiek van MDS verbeteren?

Conclusie & To Do

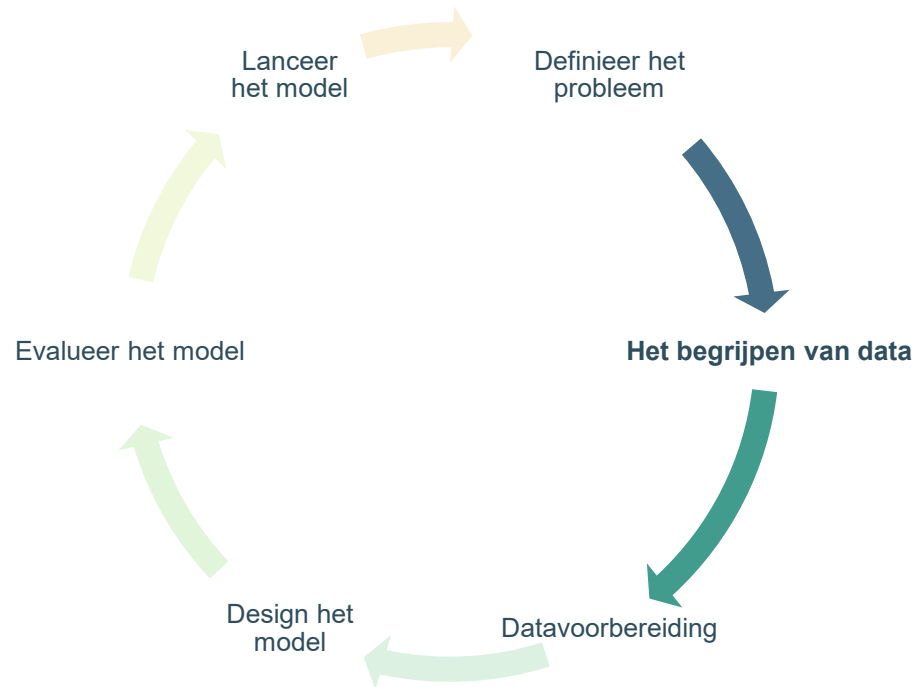
Data voor machine learning



- Hypothese
- Learning type (supervised vs unsupervised)
- Databeschikbaarheid
- Personeel

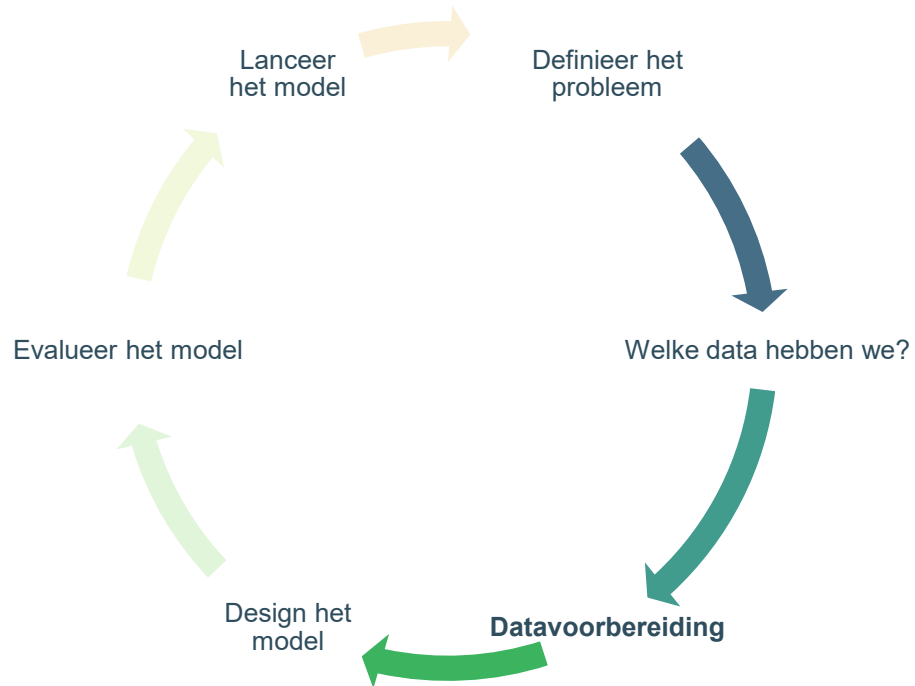
Cross Industry Standard Process for Data Mining
(CRISP-DM)

Data voor machine learning

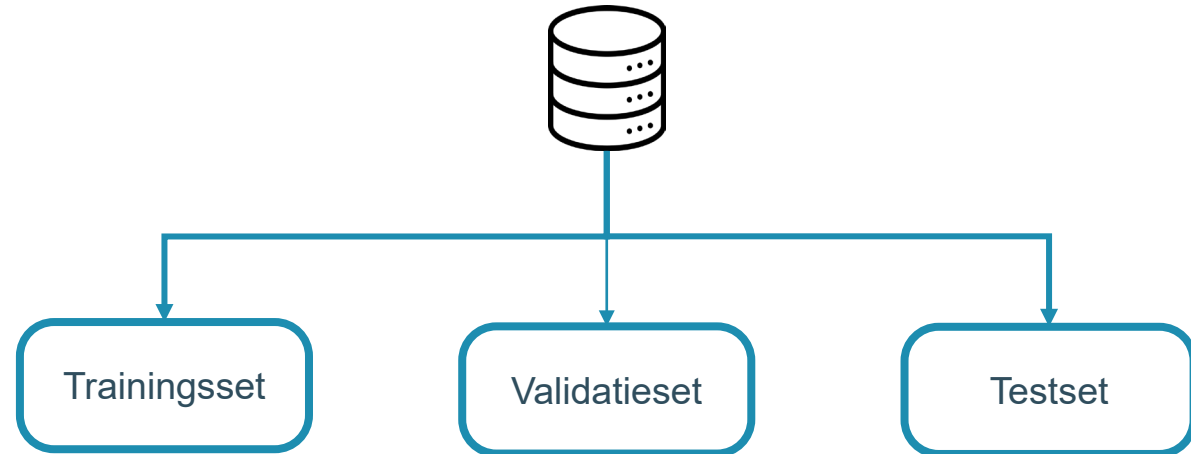


- Verzamelen van data
- Visualiseer
- Overweeg bias

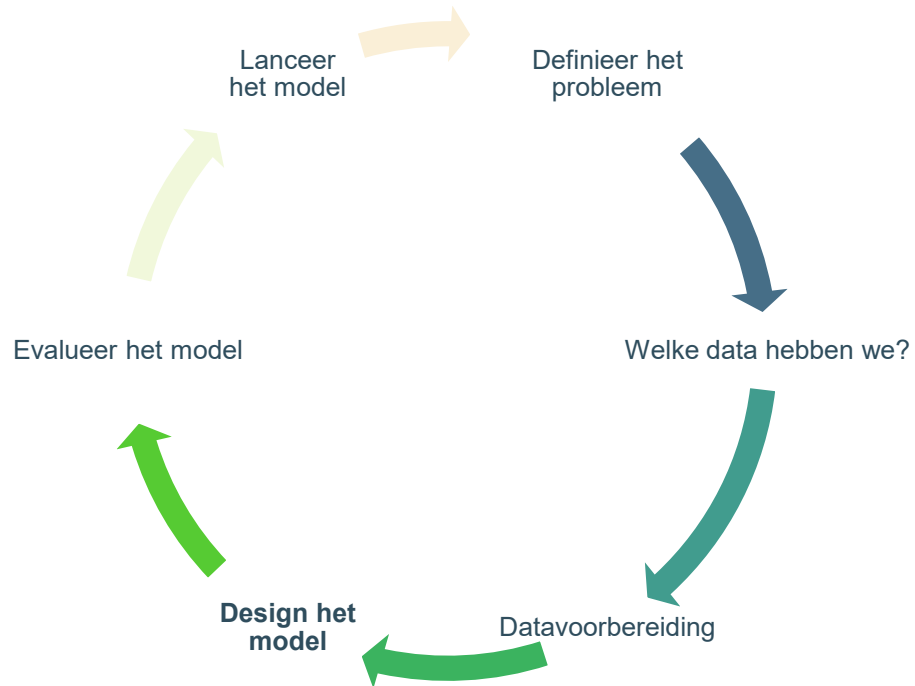
Data voor machine learning



- Kost veel tijd
- Wat doen we met ontbrekende data?
- Op welke eigenschappen gaan we selecteren?
- Data partitioneren

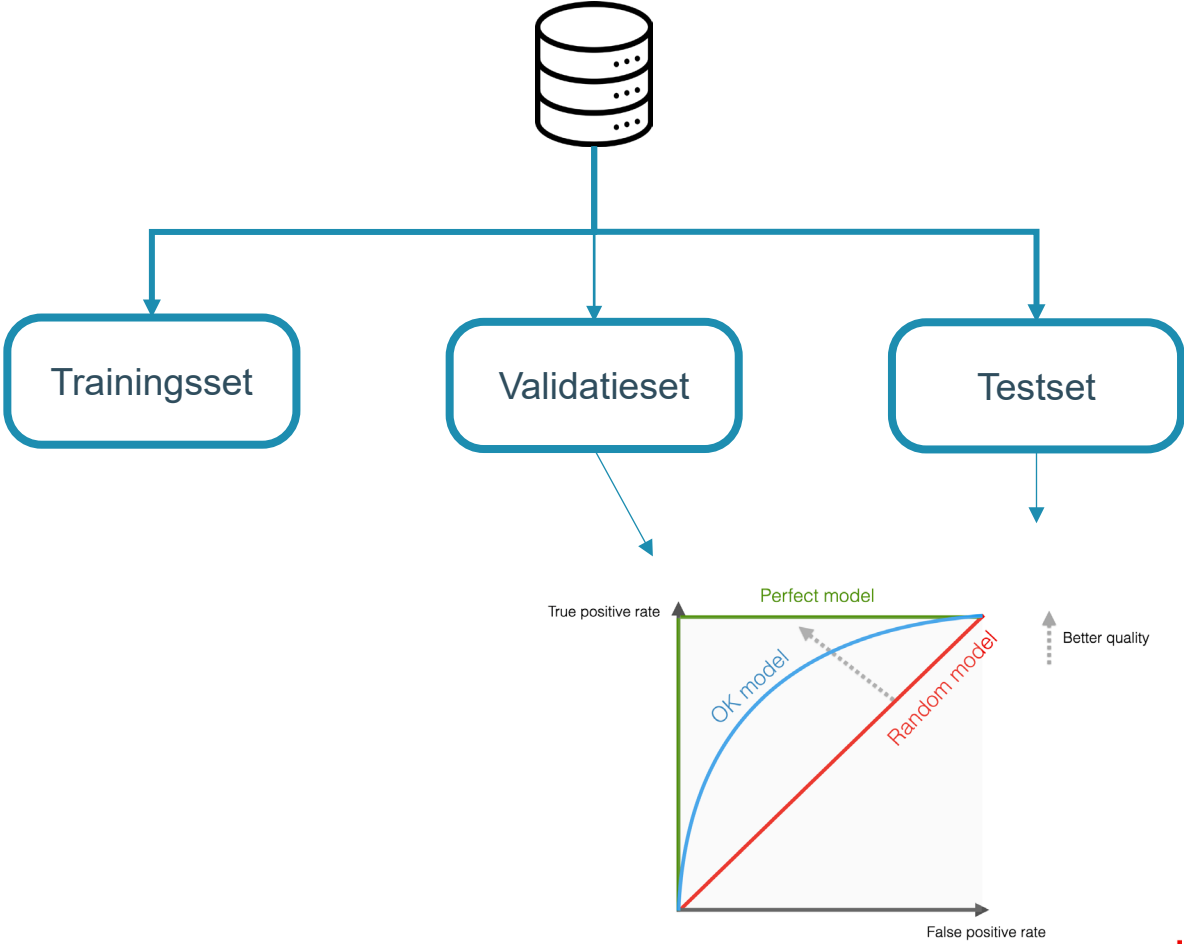
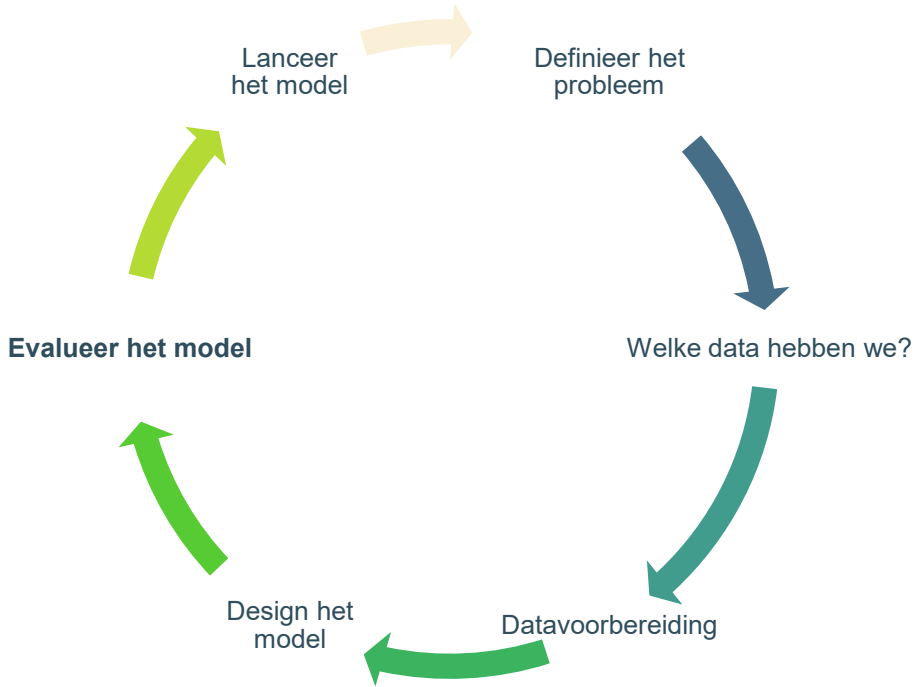


Data voor machine learning

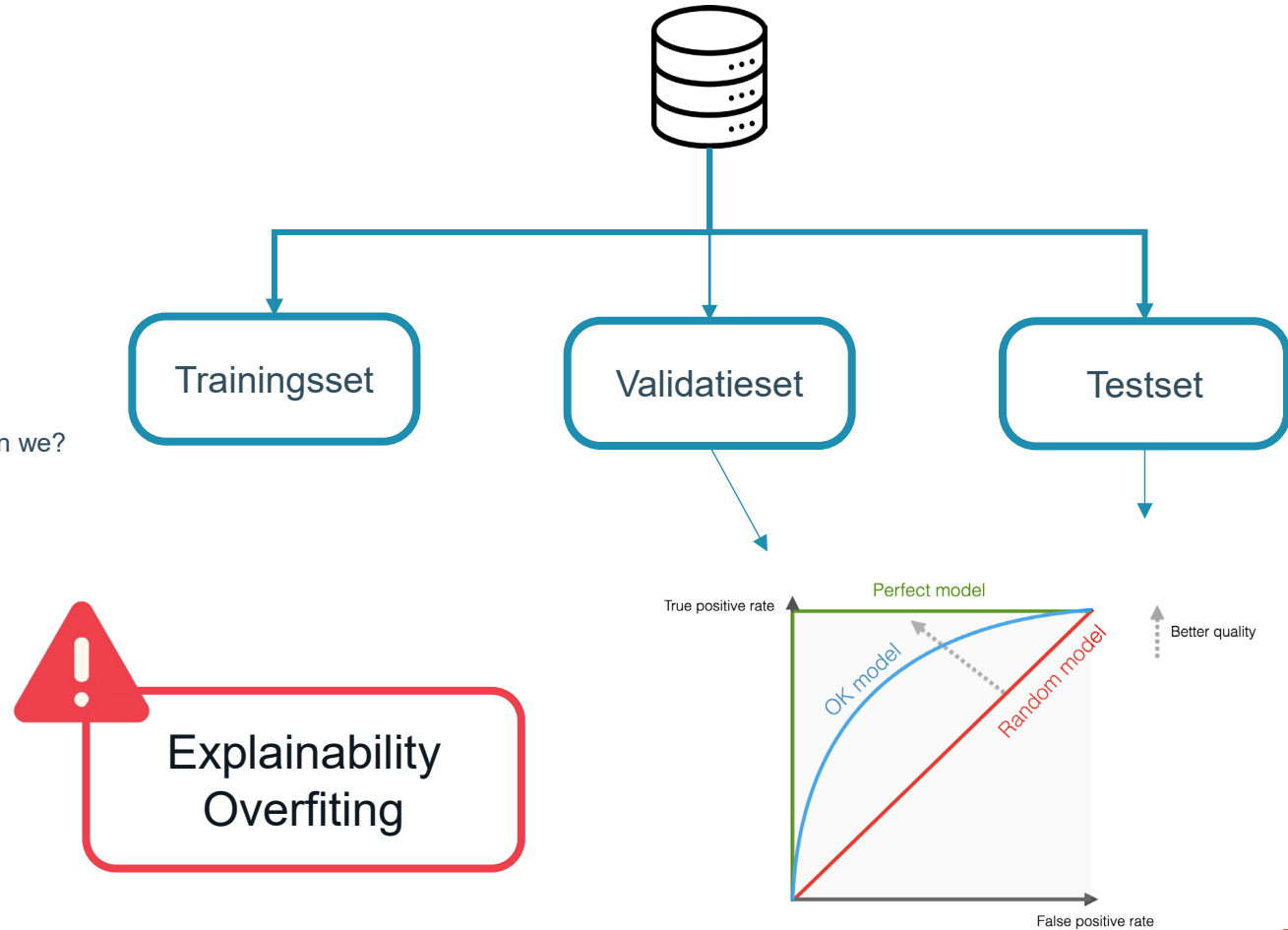
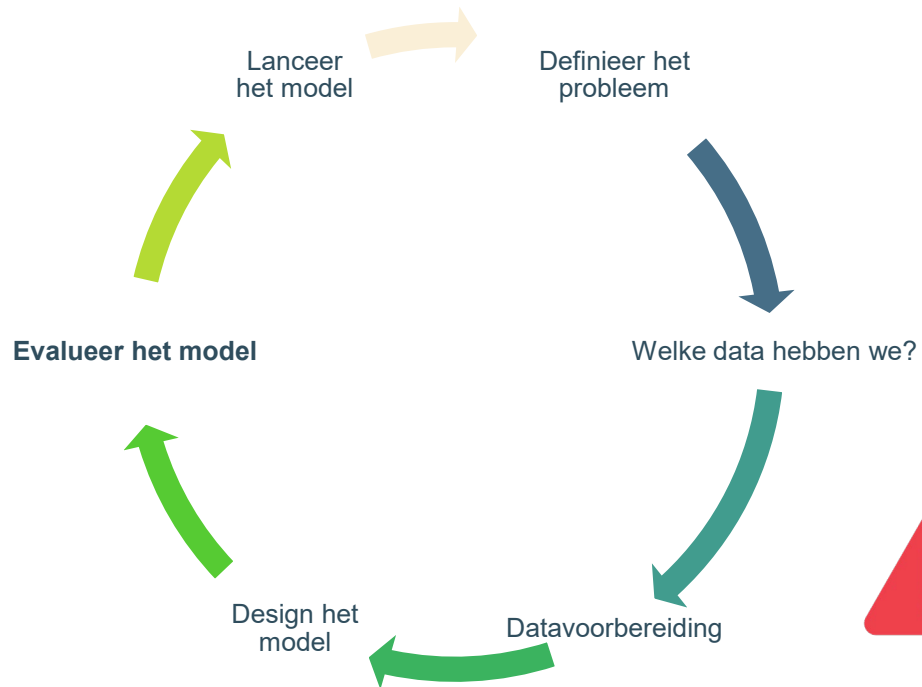


- Kies het juiste model
 - Decision trees
 - Random forests
 - Gradient boosting
 - Neurale netwerken (bv Convolutionele neurale netwerken)
 - Support vector machines
 - Logistische regressie

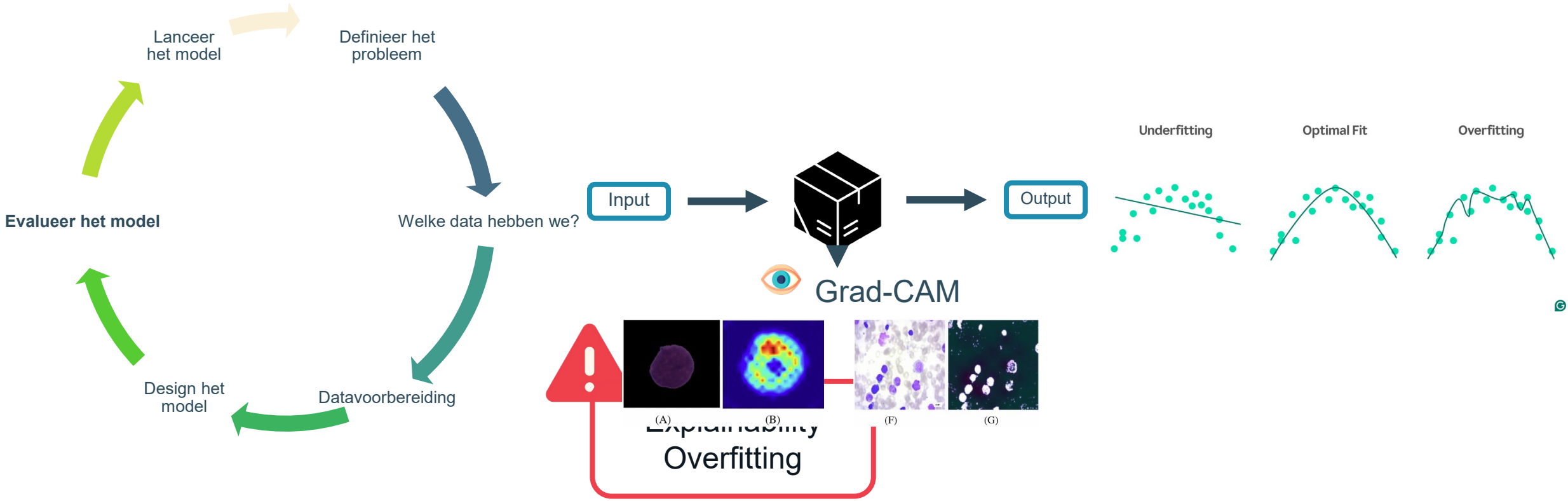
Data voor machine learning



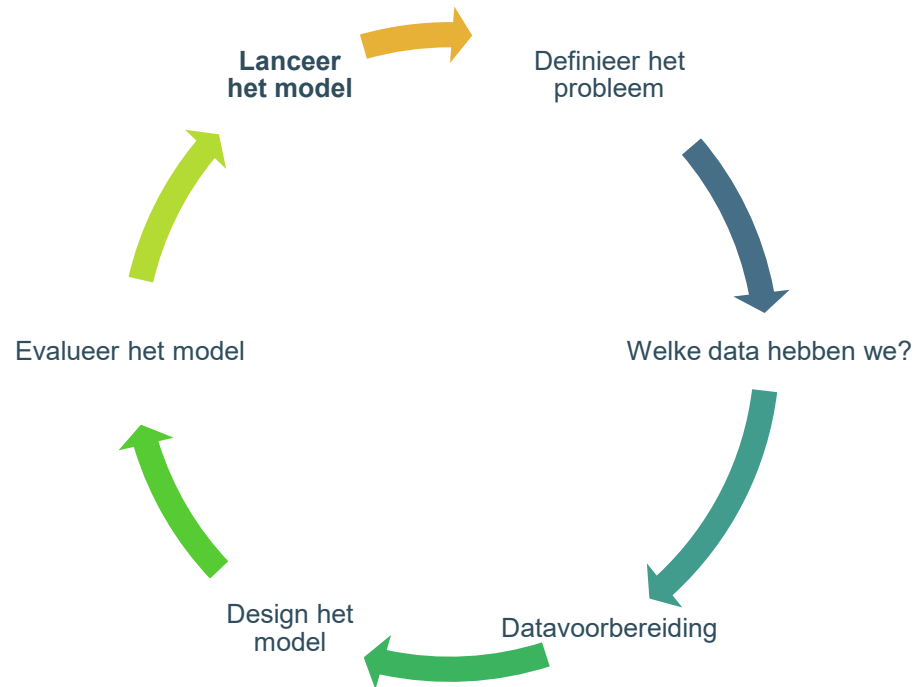
Data voor machine learning



Data voor machine learning



Data voor machine learning



Overzicht

Inleiding

- Toenemende interesse in AI
- Welke AI modellen bestaan er?
- Datamining in machine learning

Onderzoeksvragen

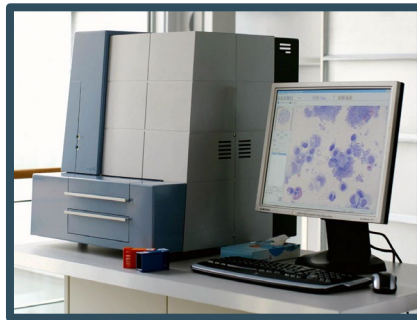
1. Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?
2. Kan AI de diagnostiek van MDS verbeteren?

Conclusie & To Do

Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?



2001

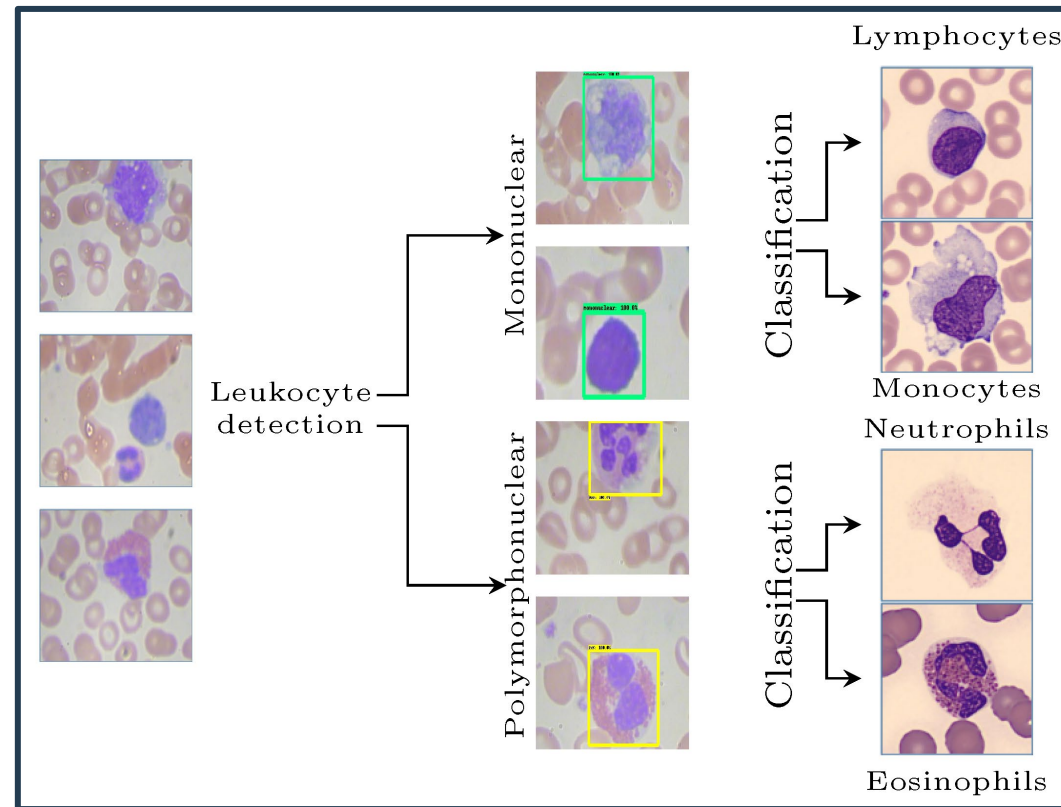


→ 2025

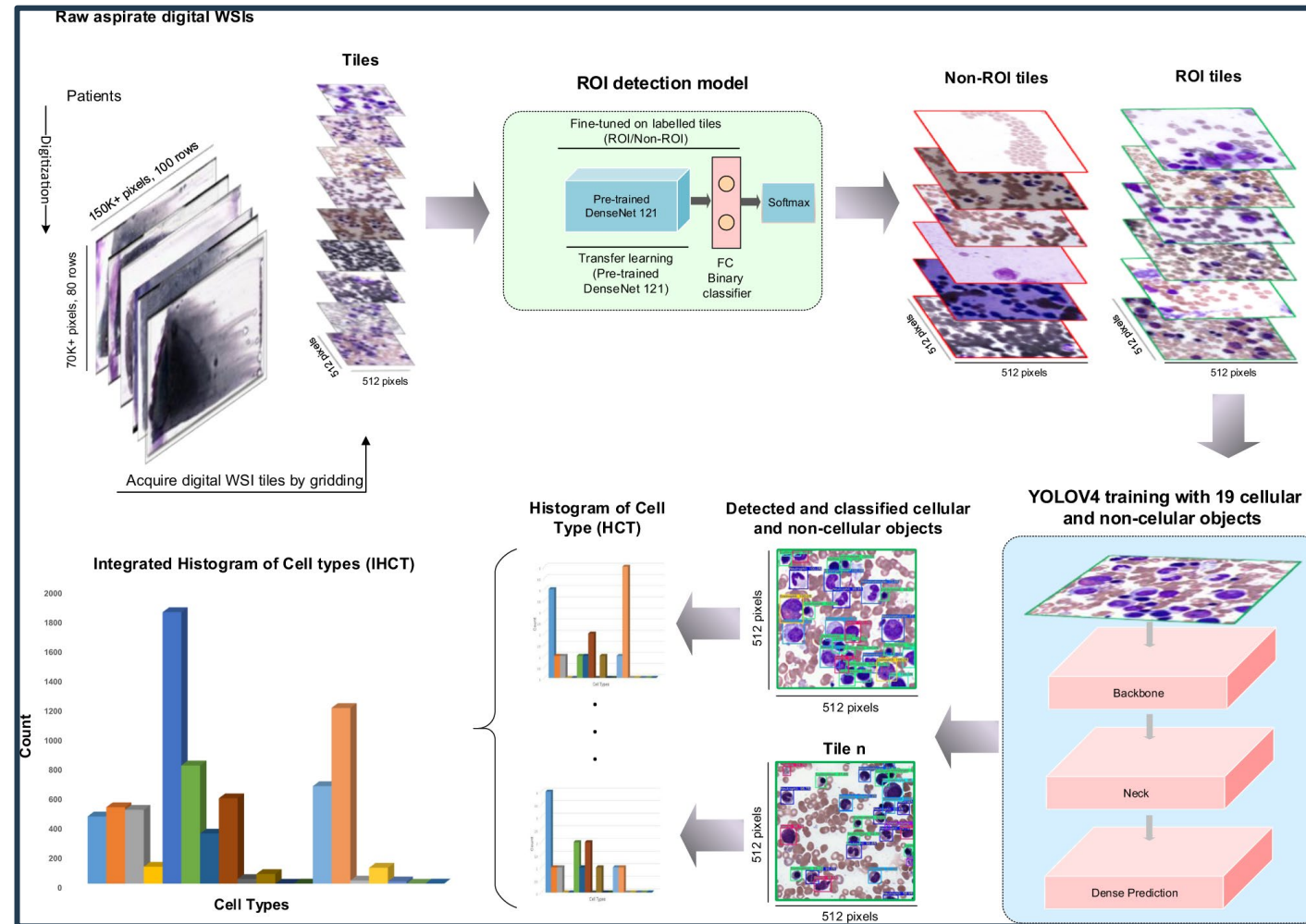
AI in CellaVision

- Machine learning als sinds de jaren 90
 - Eigen bibliotheken voor classificatie gebaseerd op traditionele ML modellen/ image processing = "feature engineering"
- Moderne systemen integreren meer deep learning methodes
 - Volledige afbeeldingen kunnen als data fungeren en het systeem leert zichzelf patronen aan om cellen te classeren
- Grootste uitdaging
 - Data kwaliteit en kwantiteit
- Toekomst
 - Volledige classificatie zonder human in the loop
 - Integratie met andere toestellen (celtellers)

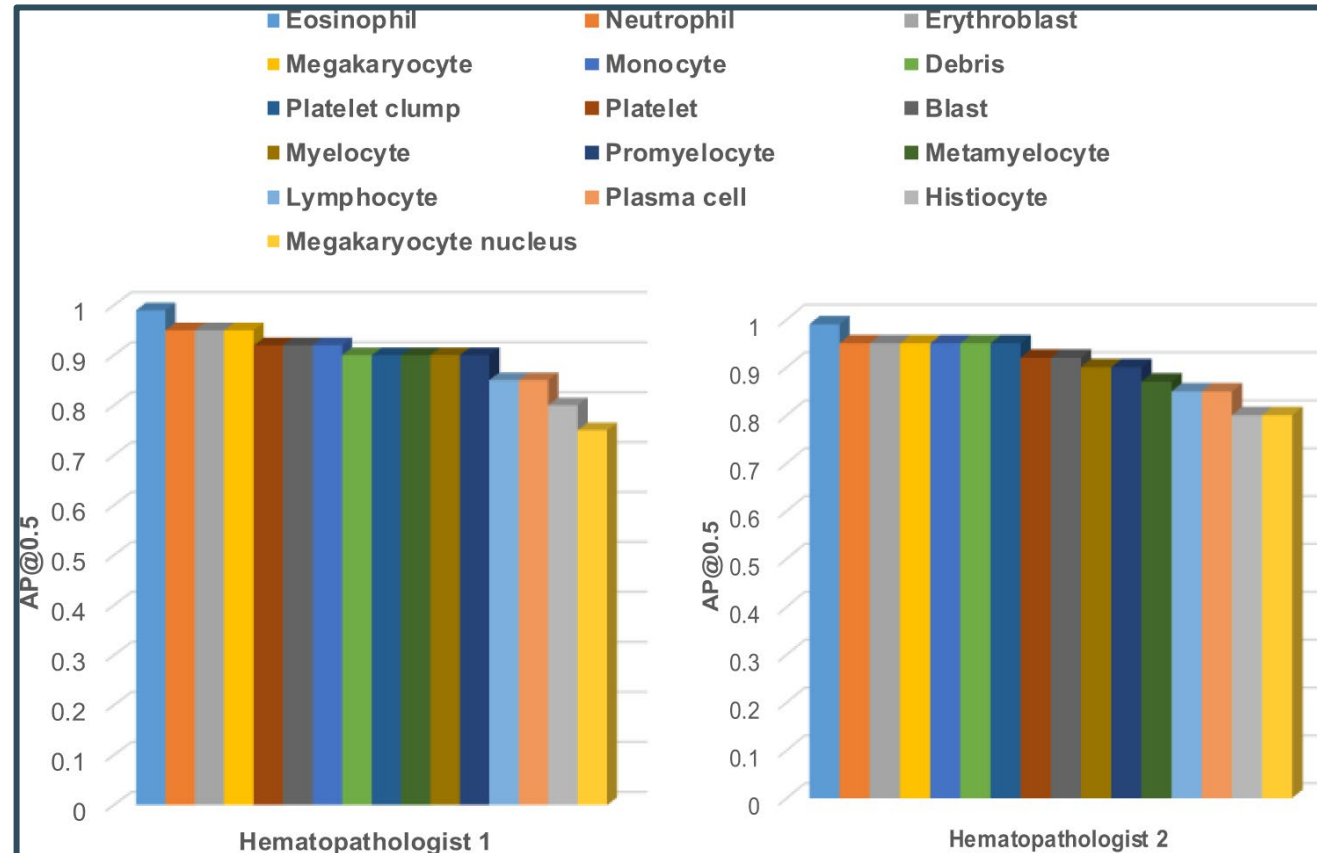
Deep learning in periferer bloed



Deep learning in beenmergaspiraat



Deep learning in beenmergaspiraart



Overzicht

Inleiding

- Toenemende interesse in AI
- Welke AI modellen bestaan er?
- Datamining in machine learning

Onderzoeksvragen

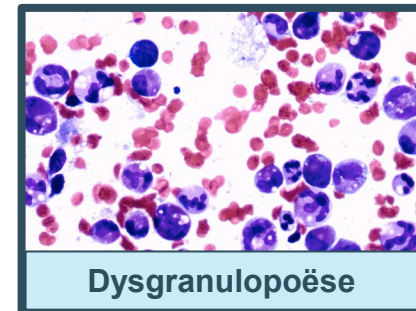
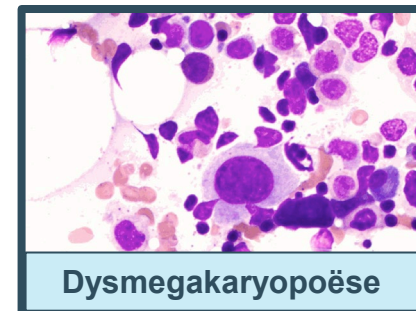
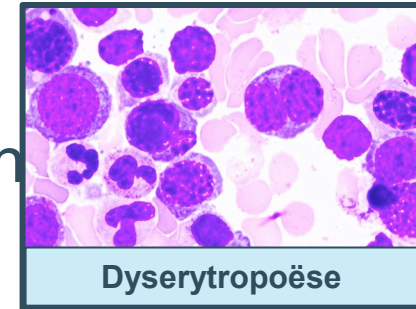
1. Hoe kan AI een rol spelen binnen het cytomorfologisch onderzoek?
2. Kan AI de diagnostiek van MDS verbeteren?

Conclusie & To Do

MDS

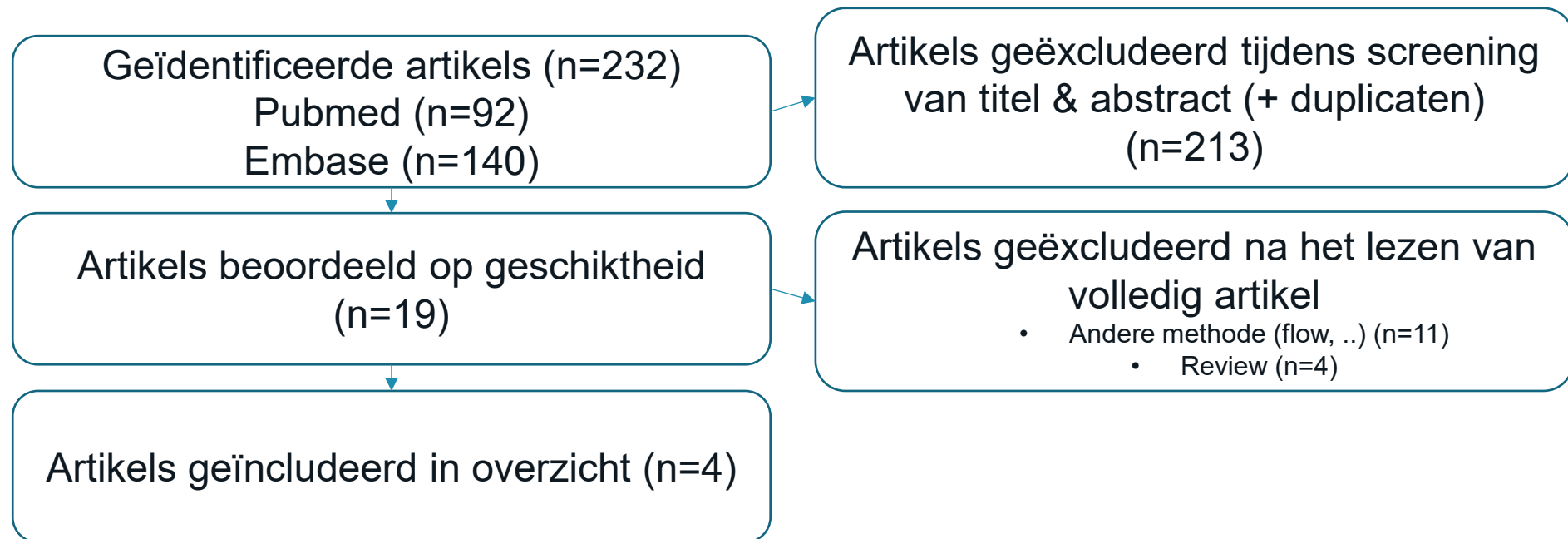
Definitie

- Klonale aandoeningen van hematopoëtische stamcellen
- Ineffectieve hematopoëse → cytopenieën
- Risico op progressie naar AML (~30%)



Is er een voordeel in het gebruik van AI-gebaseerde image analyse bij het cytomorfologisch onderzoek van MDS?

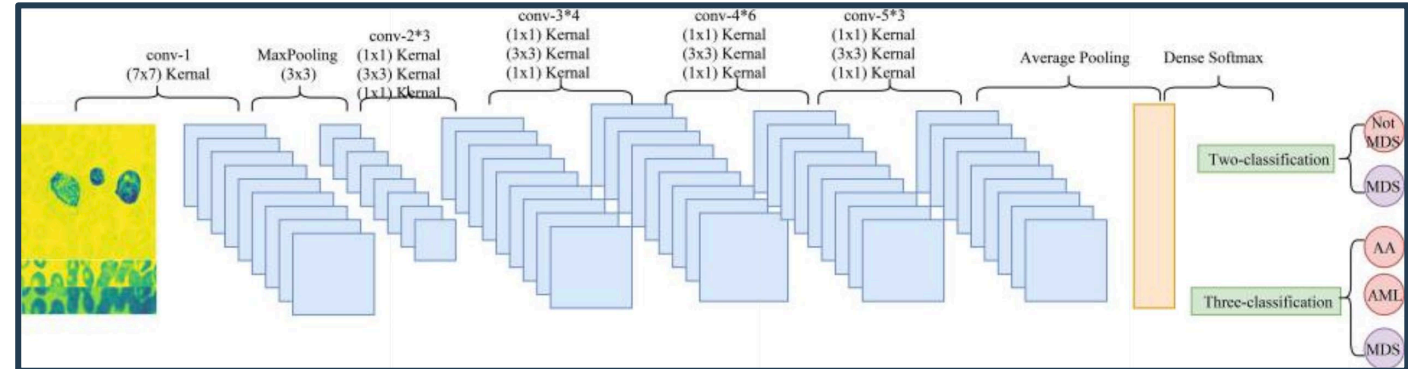
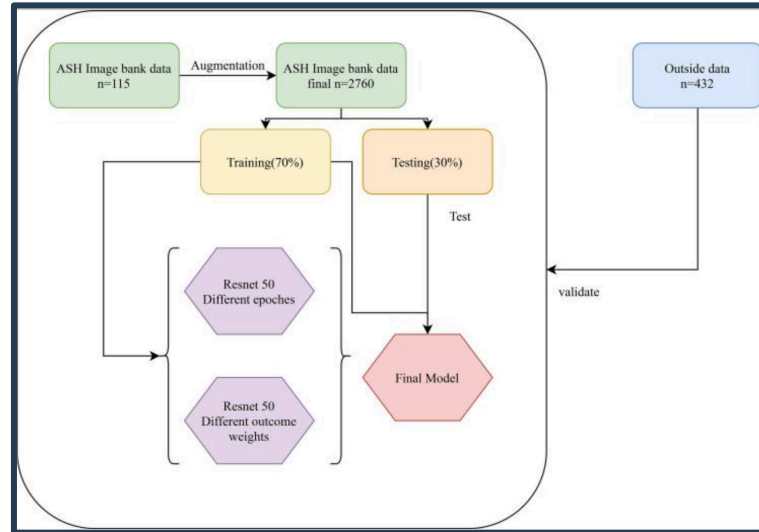
- Pubmed/Embase: 'MDS', 'Myelodysplastic Syndrome', 'AI', 'Artificial Intelligence', 'Machine learning', 'Deep learning'
- Laatste 10 jaar



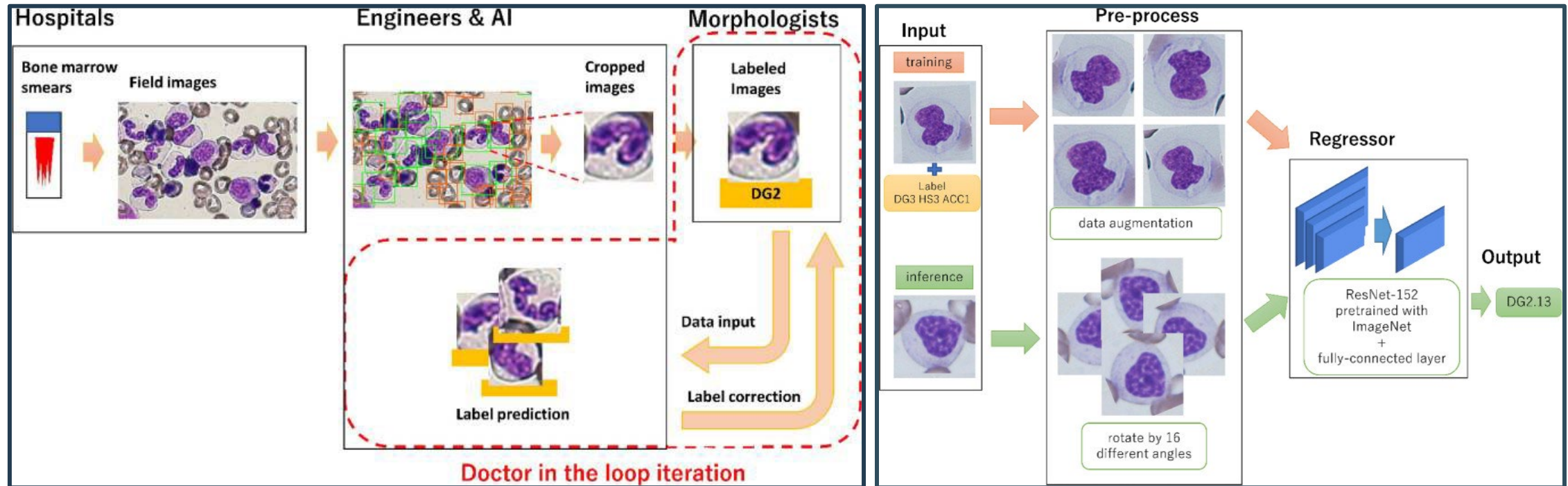
Auteur	Jaar	Dataset	Model	Outcome	Resultaten	Voordelen	Nadelen
Wang M. et al.	2022	115 BMS van ASH image databank (32 MDS, 26 AA, 57 AML)	ResNet-50 (CNN)	Onderscheid tussen MDS en geen MDS	MDS vs. non-MDS AUC 0.985 (testset) AUC 0.942 (validatieset) Accuracy 91.4% (testset), Accuracy 92.1% (validatieset) Sens 99.2% (testset) Spec 88.1% (testset) Spec 93.8% (validatieset)	Zeer goede prestatie	Vereist een clinicus
		432 BMS van klinische data voor externe validatie (214 MDS, 115 AA, 103 AML)		Onderscheid tussen MDS, AA, AML		Drie klassen (AA vs. MDS vs. AML) AUC 0.968 (testset) AUC 0.948 (validatieset) Accuracy 92.9% (testset), Accuracy 91.5% (validatieset) Sens 85.7% (testset) Sens 88.7% (validatieset)	Intern en extern gevalideerd
Mori J. et al.	2020	35 BMS met MDS 1797 labeled single-cell afbeeldingen	Faster R-CNN met ResNet-101 ResNet-152	hypogranulatie in neutrofielen (representatief voor dysplasie) voor het herkennen van MDS	Mild – Matige/ernstige dysplasie AUC 0.944 - 0.921 Accuracy 97.2% - 98.2% Sens 91.0% - 85.2% Spec 97.7% - 98.9% PPV 76.3% - 80.6% NPV 99.3% - 99.2%	Goede prestatie Efficient design Open-source code	Enkel op basis van hypogranulariteit in de neutrofielen Niet gebalanceerde datasets Misclassificatie tussen hyposegmentatie en immature granulocyten Geen externe validatie Selectiebias

Auteur	Jaar	Dataset	Model	Outcome	Resultaten	Voordelen	Nadelen
Lee N. et al.	2022	34 BMS met MDS 24 normale BMS 8065 labeled single cells in 8 categorieën	U-Net segmentatie InceptionV3 CNN classificatie	Automatische detectie en classificatie van dysplastische cellen in beenmergaspiraten van patiënten met MDS	Dysplastische granulocyten (GD) AUC 0.996 Sens 90% Spec 99.9% Erytroïde dysplasie AUC ? Sens 79% Spec 99.2% Megakaryocytaire dysplasie AUC ? Sens 89.9% Spec 94.8%	Hoge classificatie accuraatheid Acht verschillende categorieën dus een robuuster model Grad-CAM heatmap voor explainable AI Open-source code	Niet gebalanceerde data Beperkte dataset
Acevedo A. et al.	2021	20670 afbeeldingen van normale en dysplastische neutrofielen	Acht eigen CNN modellen werden getraind en beste werd geselecteerd: "DysplasiaNet"	Automatische detectie van hypogranulaire neutrofielen in het perifeer bloed.	Hypogranulaire neutrofielen AUC 0.982 Sens 95.5% Spec 94.3% PPV 94% Accuracy 94.85%	Hoge classificatie accuraatheid Minder complex model met minder rekenkracht Uitgebreide dataset Explainable AI Externe validatie	Enkel op basis van hypogranulariteit in de neutrofielen en enkel binair (geen gradering)

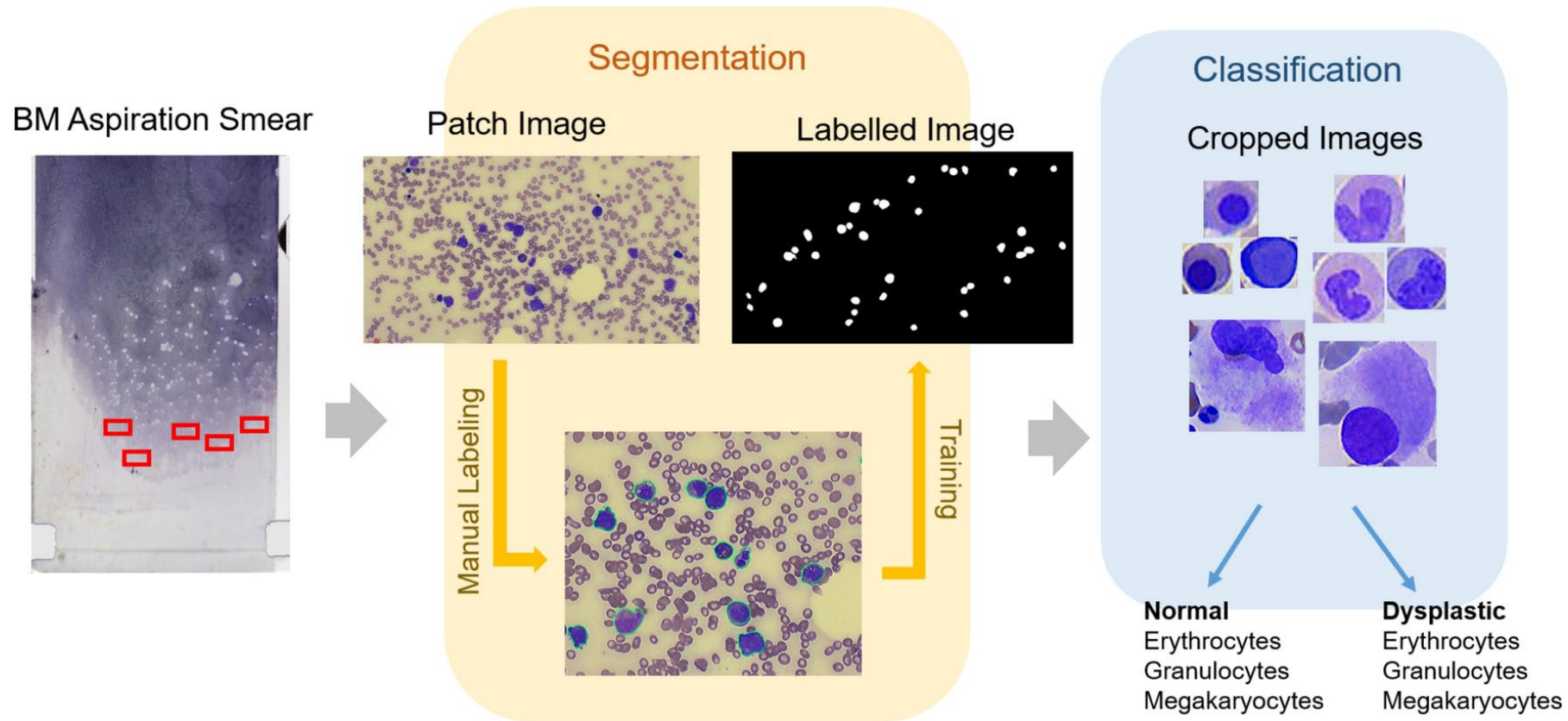
A Deep Learning Model for the Automatic Recognition of Aplastic Anemia, Myelodysplastic Syndromes, and Acute Myeloid Leukemia Based on Bone Marrow Smear



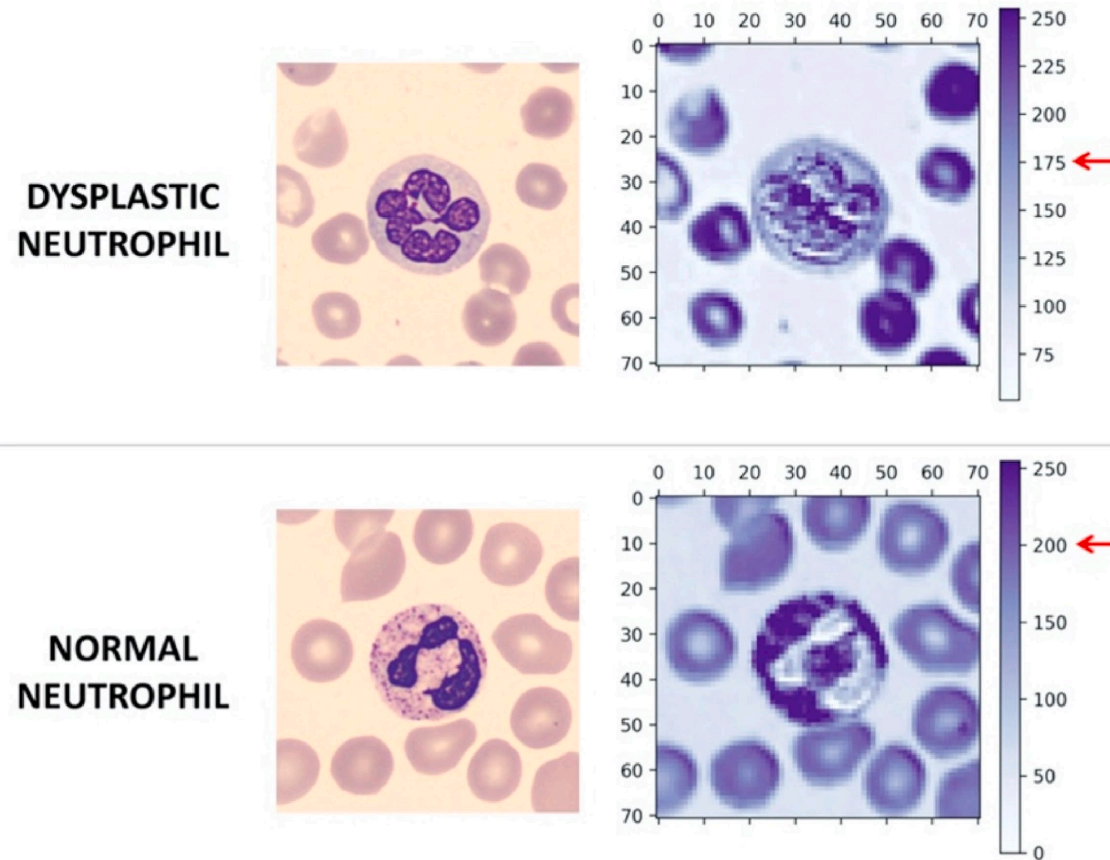
Assessment of dysplasia in bone marrow smear with convolutional neural network



Deep learning application of the discrimination of bone marrow aspiration cells in patients with myelodysplastic syndromes



A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes



Conclusie

- In het algemeen beperkte datasets.
- Grotere studies zijn nodig voor implementatie.
- Kan standaardisatie tussen ziekenhuizen verhogen.
- Deep learning kan een aanvullende rol innemen bij het diagnosticeren van MDS maar kan een expert niet vervangen.
- Nood aan explainable AI
- Regularisatie

To Do

Referenties

- Ketan Paranjape, Michiel Schinkel, Richard D Hammer, Bo Schouten, R S Nannan Panday, Paul W G Elbers, Mark H H Kramer, Prabath Nanayakkara, The Value of Artificial Intelligence in Laboratory Medicine: Current Opinions and Barriers to Implementation, *American Journal of Clinical Pathology*, Volume 155, Issue 6, June 2021, Pages 823–831, <https://doi.org/10.1093/ajcp/aqaa170>
- Jahangiri, M., Rahim, F., Saki, N., & Saki Malehi, A. (2021). Application of Bayesian Decision Tree in Hematology Research: Differential Diagnosis of β -Thalassemia Trait from Iron Deficiency Anemia. *Computational and mathematical methods in medicine*, 2021, 6401105. <https://doi.org/10.1155/2021/6401105>
- Muhammad, D., Salman, M., Keles, A., & Bendeche, M. (2025). ALL diagnosis: can efficiency and transparency coexist? An explainable deep learning approach. *Scientific reports*, 15(1), 12812. <https://doi.org/10.1038/s41598-025-97297-5>
- Cheuque, C., Querales, M., León, R., Salas, R., & Torres, R. (2022). An Efficient Multi-Level Convolutional Neural Network Approach for White Blood Cells Classification. *Diagnostics (Basel, Switzerland)*, 12(2), 248. <https://doi.org/10.3390/diagnostics12020248>
- Tayebi, R. M., Mu, Y., Dehkharghanian, T., Ross, C., Sur, M., Foley, R., Tizhoosh, H. R., & Campbell, C. J. V. (2022). Automated bone marrow cytology using deep learning to generate a histogram of cell types. *Communications medicine*, 2, 45. <https://doi.org/10.1038/s43856-022-00107-6>
- Wang, M., Dong, C., Gao, Y., Li, J., Han, M., & Wang, L. (2022). A Deep Learning Model for the Automatic Recognition of Aplastic Anemia, Myelodysplastic Syndromes, and Acute Myeloid Leukemia Based on Bone Marrow Smear. *Frontiers in oncology*, 12, 844978. <https://doi.org/10.3389/fonc.2022.844978>
- Mori, J., Kaji, S., Kawai, H., Kida, S., Tsubokura, M., Fukatsu, M., Harada, K., Noji, H., Ikezoe, T., Maeda, T., & Matsuda, A. (2020). Assessment of dysplasia in bone marrow smear with convolutional neural network. *Scientific reports*, 10(1), 14734. <https://doi.org/10.1038/s41598-020-71752-x>
- Acevedo, A., Merino, A., Boldú, L., Molina, Á., Alférez, S., & Rodellar, J. (2021). A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes. *Computers in biology and medicine*, 134, 104479. <https://doi.org/10.1016/j.compbiomed.2021.104479>