

# Medische Beeldbewerking met behulp van Artificial Intelligence

**U heeft vast wel eens om medische redenen een röntgenscan, MRI, CT of echo ondergaan. Beeldvorming speelt namelijk een belangrijke rol in de medische wereld, niet alleen in de kliniek, maar ook voor onderzoekers. In de kliniek helpt beeldvorming bij het stellen van een diagnose en het geven van een prognose, maar ook bij het plannen en begeleiden van een ingreep zoals een operatie of bestraling. Artificial intelligence kan helpen om automatisch metingen te verrichten in de grote hoeveelheid medische beelden die gemaakt wordt.**

In medisch onderzoek wordt beeldvorming gebruikt voor het bestuderen van de uiterlijke kenmerken van een ziekte, het ziekteverloop en tevens voor het bepalen van de effecten van een behandeling. Hierbij is het vergelijken van beelden van groot belang, zowel tussen patiënten als van dezelfde patiënt op verschillende tijdstippen. Bij voorkeur wordt dit vergelijken gedaan op een kwantitatieve manier: met het gebruik van zogenaamde *kwantitatieve beeldbiomarkers*. Dit zijn waarden die iets kunnen vertellen over een ziekte, zoals afmetingen of vormen van weefsels of structuren. Zo is bijvoorbeeld het volume en de vorm van bepaalde hersengebieden voorspellend voor de ontwikkeling van verschillende vormen van dementie, is littekenweefsel in de hersenen voorspellend voor MS, is aderverkalking kenmerkend voor verschillende hart- en vaatziekten en kraakbeenvolume en -vorm voor artrose. Door het gebruik van deze kwantitatieve beeldbiomarkers kunnen we hopelijk beter begrijpen wat de verschillen zijn tussen patiënten en gezonde personen en hoe ziektes zich ontwikkelen. Ook kan, door het vergelijken van waarden van een patiënt met een database met waarden van gezonde personen, worden bepaald hoe we ziektes beter kunnen herkennen.

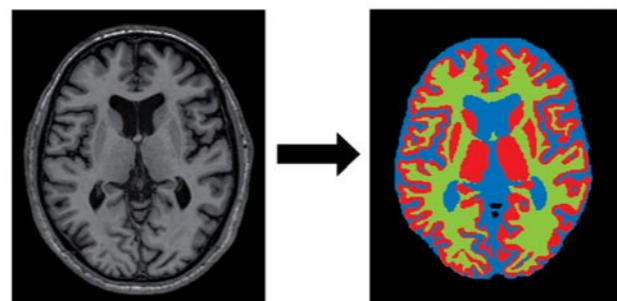
MRI- en CT-scans worden het meest gebruikt voor het extraheren van kwantitatieve beeldbiomarkers, omdat ze driedimensionale beelden geven. Om deze beeldbiomarkers te kunnen meten in MRI- en CT-beelden, moeten deze beelden gesegmenteerd worden in de weefsels

Dr. ir. A. van Opbroek heeft een achtergrond in de toegepaste wiskunde (BSc) en biomedische technologie (MSc). Tijdens haar PhD en PostDoc in de medische beeldbewerking bij het Erasmus MC heeft ze veel kennis en ervaring opgedaan met machine learning en artificial intelligence en de toepassing hiervan in de medische praktijk. De afgelopen 2,5 jaar heeft ze gewerkt bij TNO als onderzoeker AI en deep learning.



of structuren waarin men geïnteresseerd is. We nemen als voorbeeld de simpele beeldbiomarkers 'volume van de grijze stof', 'volume van de witte stof' en 'volume van het hersenvocht' in de hersenen. Grijze stof is het weefsel in de hersenen dat informatie opslaat, witte stof transporteert signalen door de hersenen en hersenvocht voorziet de hersenen van voedingsstoffen. Het volume van de witte en grijze stof neemt af naarmate we ouder worden, maar bij mensen met ontwikkelende dementie meer dan bij gezonde personen. Het volume van deze twee stoffen zijn hiermee voorspellend voor dementie. Hersenvocht zit overal waar geen witte of grijze stof zit en neemt dus juist toe.

Figuur 1 geeft een voorbeeld van een 2-dimensionale plak van een MRI-hersenebeeld en een bijbehorende segmentatie in achtergrond, witte stof, grijze stof en hersenvocht. Het handmatig maken van een dergelijke segmentatie is echter enorm arbeidsintensief, zeker in 3D. Daarnaast is een handmatige segmentatie ook subjectief; een ander persoon, maar ook dezelfde persoon op een ander moment, zal een andere segmentatie genereren. Hierdoor kan het lastig zijn segmentaties met elkaar te vergelijken, zeker als er slechts kleine verschillen zijn tussen beelden. Om deze twee redenen wordt er de afgelopen decennia veel onderzoek gedaan naar het automatisch genereren van dergelijke segmentaties met behulp van *machine learning of artificial intelligence*.



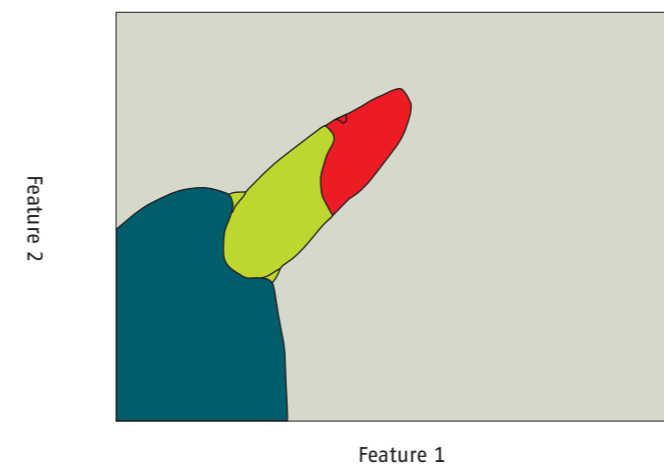
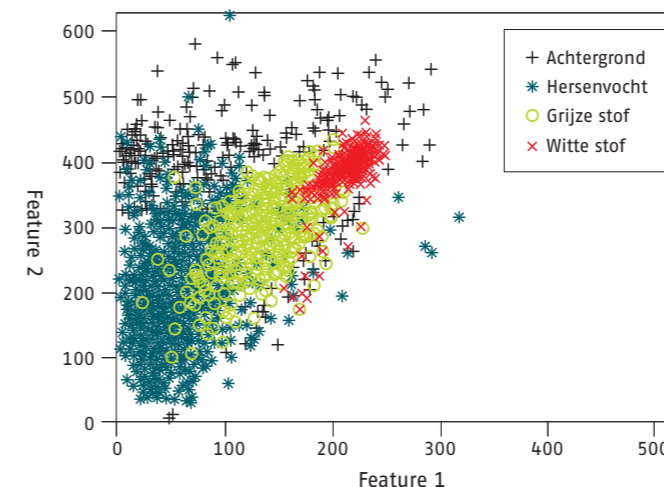
**Figuur 1:** Voorbeeld van medische beeldsegmentatie. Links: een plak van een MRI-beeld van de hersenen. Rechts: een segmentatie van het beeld in achtergrond (zwart) en de drie hersenweefsels: hersenvocht (blauw), grijze stof (rood) en witte stof (groen).

Machine learning heeft zich de afgelopen jaren bewezen als waardevolle techniek voor automatische medische beeldsegmentatie. Hierbij worden voorbeelden gebruikt om een beslissingsmodel te trainen. Het voordeel van het automatisch bepalen van beslissingen aan de hand van voorbeelden is dat een ontwikkelaar niet zelf hoeft te programmeren hoe beslissingen worden gemaakt, wat heel ingewikkeld en arbeidsintensief zou zijn.

In ons voorbeeld bestaan de genoemde voorbeelden uit MRI-scans en bijbehorende handmatig gemaakte segmentaties van de hersenen, deze worden ook wel de trainingsbeelden genoemd. Uit deze *trainingsbeelden* wordt vervolgens een groot aantal *trainingsamples* geëxtraheerd voor elk van de verschillende classes, d.w.z. de verschillende weefsels die gesegmenteerd zijn. Deze trainingsamples bestaan in het algemeen uit voxels (drie-dimensionale pixels) of stukjes bestaande uit een klein aantal naastgelegen voxels (bijvoorbeeld 3x3x3). Voor al deze trainingsamples worden vervolgens *features* gemeten: beeld eigenschappen van het sample. Voorbeelden

van zulk soort features zijn bijvoorbeeld de grijswaarde (intensiteit) in het MRI beeld, locatie in het beeld (weergegeven in bijvoorbeeld  $x,y,z$  coördinaten), maar ook buurtinformatie zoals de gemiddelde grijswaarde rondom het sample, of verschil in grijswaarde met naastgelegen samples. Door het meten van deze features in alle trainings-samples kunnen deze worden voorgesteld als punten in een *feature ruimte* (zie Figuur 2). Deze feature ruimte is vaak hoogdimensionaal, omdat veel features (tientallen, honderden, of zelfs duizenden features) worden gemeten.

In deze feature ruimte wordt vervolgens een beslissingsmodel, een *classifier*, geoptimaliseerd die zo goed mogelijk de samples van de verschillende classes kan onderscheiden. Er bestaan heel veel verschillende soorten classifiers die de feature ruimte op verschillende manieren opsplitsen (bijvoorbeeld per feature of alle features tegelijkertijd), met convexe of niet-convexe doelfuncties en computationeel intensief of juist heel snel. Het is belangrijk voor ontwikkelaars om een goede afweging te maken in de classifier die het meest geschikt is. Een van de belangrijkste eigenschappen hierbij is dat zogenaamde *over-training* moet worden voorkomen, wat inhoudt dat de classifier weliswaar de trainingsdata heel goed beschrijft, maar niet generaliseert naar nieuwe data. Dit kan bijvoorbeeld worden voorkomen door een classifier te kiezen die niet teveel opschuift als een klein deel van de data wordt weggenomen of door het wegnemen van niet-informatieve features. Veelgebruikte classifiers het afgelopen decennium zijn de *support vector machine* en *random forest*, maar de afgelopen jaren wordt het veld vooral gedomineerd door *deep neural networks*, vanwege hun superieure prestatie en hun vermogen om zelf de beste features te kunnen bepalen.



**Figuur 2:** 2D voorbeeld van een feature ruimte en een classifier. Feature 1: MRI intensiteit (grijswaarde) van de voxel, Feature 2: gemiddelde intensiteit van de naburige voxels. Boven: samples van 4 verschillende classes in de feature ruimte. Onder: een classifier die beschrijft waar in de feature ruimte de verschillende classes zich bevinden.

Wanneer de classifier geleerd is, kan het gebruikt worden om een nieuw beeld, het testbeeld te segmenteren. Hiertoe wordt het *testbeeld* eerst opgesplitst in samples, de *testsamples*. Voor deze testsamples worden dezelfde features gemeten als voor de trainingsamples om ze in de feature ruimte te kunnen plaatsen. Tenslotte wordt de getrainde classifier toegepast om te bepalen tot welke klasse elk testsample behoort. Dit geeft vervolgens een classificatie voor elk sample in het beeld: de segmentatie. De kwaliteit van de gegenereerde automatische segmentatie kan vervolgens worden bepaald als voor het testbeeld ook een handmatige segmentatie beschikbaar is, bijvoorbeeld door het meten van het percentage correct gesegmenteerde voxels. Als de kwaliteit van de methode goed genoeg is kan deze vervolgens worden toegepast op nieuwe beelden om zo snel en accuraat een segmentatie te maken.

Dit soort machinelearningmethodes werken in het algemeen ontzettend goed en worden in rap tempo ontwikkeld voor alle denkbare ziektes. Voor veel toepassingen grenst de kwaliteit van de gegenereerde segmentaties aan die van mensen, of weet deze zelfs te overtreffen, in een fractie van de tijd die mensen hiervoor nodig hebben. Op deze manier worden door onderzoekers waardevolle nieuwe inzichten verkregen over het verloop en de diagnose van ziektes en kan gemakkelijker de invloed van behandelingen kwantitatief worden gemeten.

De afgelopen jaren worden dergelijke beeldanalysemethodes ook meer en meer toegepast in de kliniek. Ze ondersteunen artsen bijvoorbeeld met het stellen van een diagnose door automatische metingen te vergelijken met referentiewaarden van gezonde personen. Normaal gesproken krijgen artsen gedurende hun loopbaan een getraind oog dat kan inschatten wanneer er iets mis is (bijvoorbeeld ernstig gekrompen grijze stof), maar door automatische beeldanalyse kan dit kwantitatief worden gemeten en vergeleken. Op deze manier kan kunstmatige intelligentie helpen om diagnoses niet alleen sneller maar ook betrouwbaarder te stellen. ■