

Context-level machine learning ter verbetering van de identificatie van lymfeklier- en botmetastasen bij prostaatkankerpatiënten met [¹⁸F]PSMA-1007-PET

R.J. Poelarends, MSc¹; J.A. van Dalen, PhD²; J.D. van Dijk, MD, PhD³; H. Stevens³; B.N. Vendel, MD³

University of Twente, Enschede¹; Department of Medical Physics² and Nuclear Medicine³, Isala, Zwolle

Abstract

Interpretation of [¹⁸F]PSMA-1007 PET can be challenging in clinical practice due to aspecific uptake in lymph nodes and bones. In this study, we aimed to improve the identification of lymph node and bone metastases in prostate cancer patients using [¹⁸F]PSMA-1007 PET through a next-level machine learning approach that incorporates context-level information. The cohort consisted of 110 prostate cancer patients who underwent [¹⁸F]PSMA-1007 PET/CT imaging for suspected metastases. Machine learning models were trained and evaluated using hotspot-specific and context-level information. Results demonstrated the superiority of the models incorporating context-level information, showing improved diagnostic performance. Leveraging context-level information in machine learning methods holds promise for improving the identification of metastases, potentially leading to a better treatment strategy.

Achtergrond

Geavanceerde beeldvormingstechnieken, zoals prostate-specific membrane antigen (PSMA)-PET/CT, hebben de detectie van primaire tumoren en metastasen bij

prostaatkanker verbeterd (1). De interpretatie van [¹⁸F]PSMA-1007 PET/CT-scans kan echter uitdagend zijn vanwege de aspecifieke opname in lymfeklieren en botstructuren (2). Het inzetten van artificiële intelligentie (AI), specifiek machine learning (ML), wordt beschouwd als een mogelijke oplossing voor het verbeteren van de interpretatie van PSMA-scans. In de afgelopen jaren zijn diverse ML-modellen ontwikkeld voor [⁶⁸Ga]Ga-PSMA-PET, [¹⁸F]DCPyl-PET en [¹⁸F]PSMA-1007-PET met classificatie van laesies, detectie van de primaire tumor en uitzaaiingen, en segmentatie als doel (3). Een veelvoorkomend type ML-model is het classificatiemodel, waarbij wiskundige principes worden toegepast om een gebied met verhoogde opname te classificeren als maligne of niet-maligne. In de rest van dit artikel wordt een gebied met verhoogde PSMA-opname aangeduid als een 'hotspot'. Om een model te trainen in het classificeren van hotspots wordt typisch gebruik gemaakt van kenmerken als SUVmax, de anatomische locatie en het type weefsel waarin de hotspot zich bevindt. Deze kenmerken worden vervolgens gekoppeld aan een label (maligne of niet-maligne), idealiter verkregen uit pathologische analyse van biopten, of anders uit beschikbare (follow-up) beeldvorming, klinische data en/of interpretaties vanuit multidisciplinaire overleggen. Ten slotte wordt het model getraind met deze eigenschappen en labels,

waarbij met wiskundige principes gepoogd wordt een logica te ontdekken in de data. Mensen zijn op dit moment meestal beter in staat om objecten te herkennen en te classificeren dan computermodellen, voornamelijk vanwege het bewuste of onbewuste gebruik van contextinformatie. Het is lastig om bijvoorbeeld een hond te identificeren op een sterk ingezoomde foto van de vacht, terwijl dit steeds eenvoudiger wordt naarmate er wordt uitgezoomd en er meer context beschikbaar is, zoals geïllustreerd in figuur 1.

Bestaande modellen maken voornamelijk gebruik van eigenschappen met betrekking tot de hotspot, met een sterk ingezoomd blikveld, waarbij de zeer informatieve contextinformatie niet gebruikt wordt (4,5). Het doel van deze studie was het ontwikkelen van een geavanceerde machine learning-methode die gebruik maakt van contextuele kenmerken in een PET scan om lymfeklier- en botmetastasen bij prostaatkankerpatiënten beter te identificeren met behulp van [¹⁸F]PSMA-1007.

Materialen en methoden

Patiënt cohort

We hebben retrospectief 110 patiënten geïncludeerd die tussen 1 augustus 2019 en 1 februari 2021 een [¹⁸F]PSMA-1007 PET/CT hebben ondergaan, waarin PSMA-



Figuur 1. De essentie van contextinformatie bij de identificatie van honden. Een foto van ingezoomde vacht (links) kan door een ML-model geïnterpreteerd worden als een gebraden kippenpoot (midden). Het vergroten van het blikveld, ergo het toevoegen van contextinformatie, maakt de identificatie van een hond (rechts) eenvoudiger.

expressie in botstructuren of lymfeklieren zichtbaar was. Patiënten met eerdere behandelingen voor (gemetastaseerde) prostaatkanker, eerdere [¹⁸F]PSMA-1007 PET/CT-beeldvorming, of onvolledige gegevens hebben we geëxcludeerd. Hotspots werden waargenomen door visuele evaluatie in botstructuren en lymfeklieren. Deze werden door twee ervaren nucleair geneeskundigen gelabeld als maligne of niet-maligne, aan de hand van alle beschikbare beeldvorming en klinische gegevens, met de interpretatie tijdens een multidisciplinair overleg als referentie. Bij twijfel over de maligniteit van een hotspot werd de classificatie bepaald in onderling overleg.

Met behulp van de Sectra IDS7 24.2 NM fusion module zijn hotspot- en contextvariabelen verzameld vanuit iedere PET scan. De verzamelde hotspotvariabelen waren SUVmax, de anatomische locatie (thorax, wervelkolom, pelvis, para-aortaal, iliaca communis, onderste extremiteiten) en het type weefsel (bot, lymfeklier) waarin de hotspot zich bevindt. De contextvariabelen waren SUVmax van de primaire tumor, het aantal geïdentificeerde hotspots in botstructuren en lymfeklieren van de betreffende scan en een referentie SUVmax. De referentie SUVmax werd bepaald op basis van opname

in de contralaterale lymfeklieren of botstructuren. Uitzondering hierop waren hotspots in de wervelkolom waarbij de referentie werd genomen van een nabijgelegen wervel zonder hotspot. Vanuit de verzamelde variabelen werden vervolgens kenmerken (features) gecreëerd en geselecteerd met behulp van feature-engineering, waarna deze features werden verdeeld in twee datasets.

Feature-engineering

De eerste dataset bestaat uit hotspot-specifieke features, zoals ze standaard in beschikbare modellen gebruikt worden (basisbenadering). De tweede dataset bevat zowel hotspot-specifieke features als context-level features

(contextuele benadering). Binnen deze context-level groep zijn twee features toegevoegd in aanvulling op de bestaande variabelen, namelijk de maximale SUVmax van alle hotspots op de scan en de gemiddelde SUVmax van alle hotspots in de regio. Hierbij was regio gedefinieerd als de anatomische locatie waarin de hotspot zich bevindt. Wanneer een hotspot bijvoorbeeld in de ribben werd geconstateerd, werd er gekeken naar de gemiddelde SUVmax van alle hotspots in de ribben. Deze extra features zijn gekozen omdat ze een globale samenvatting geven van alle hotspots. De features per groep zijn samengevat in tabel 1.

Tabel 1. Features behorend bij de hotspot-specifieke groep en context-level groep.

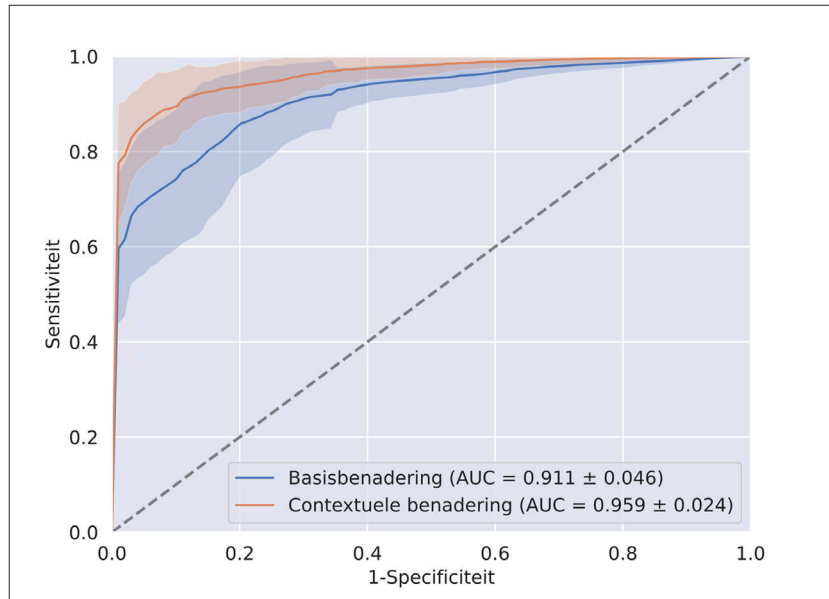
hotspot-specifiek	context-level
hotspot SUVmax weefsel type: bot lymfeklier locatie: thorax wervelkolom pelvis onderste extremiteiten para-aortaal iliaca communis	SUVmax primaire tumor referentie SUVmax aantal hotspots in lymfeklieren aantal hotspots in botstructuren maximale SUVmax van alle hotspots gemiddelde SUVmax in regio

Model training en evaluatie

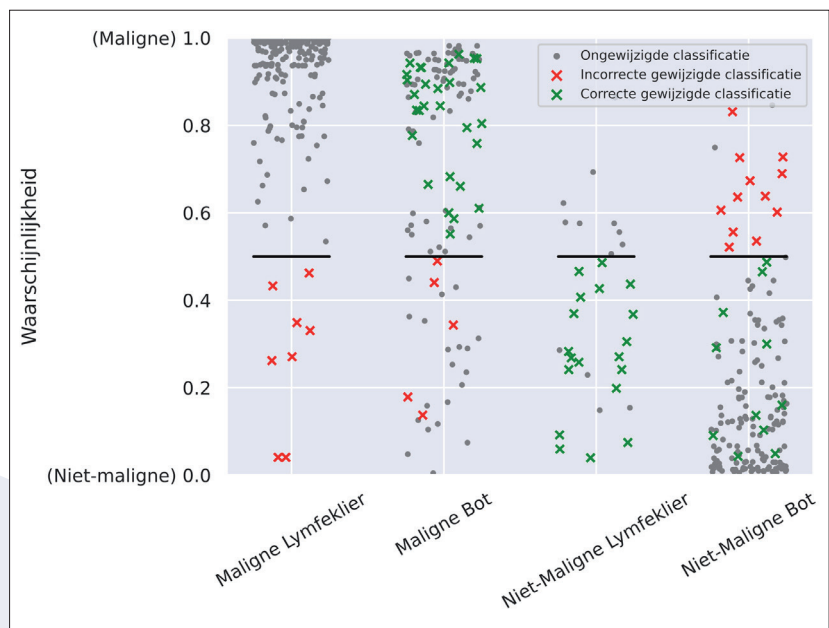
Drie veelgebruikte ML-modellen (Random Forest (6), Decision Tree (7) en K-Nearest Neighbors (8)) werden getraind en geëvalueerd op beide datasets. Er is gekozen voor deze ML-modellen omdat de keuzes die ze maken eenvoudig geïnterpreteerd en uitgelegd kunnen worden. De modellen werden getraind, getest en gevalideerd via 10x5-fold cross-validatie. Hyperparameters werden geoptimaliseerd met behulp van een grid search. De prestaties van de modellen werden geëvalueerd met behulp van een ROC-analyse, specifiek de area under the curve (AUC), specificiteit, sensitiviteit en nauwkeurigheid. Daarnaast is er gekeken naar de voorspelde waarschijnlijkheid van maligniteit per hotspot van beide modellen en de relevantie van alle features op de uiteindelijke classificaties van hotspots in lymfeklieren en botstructuren.

Resultaten

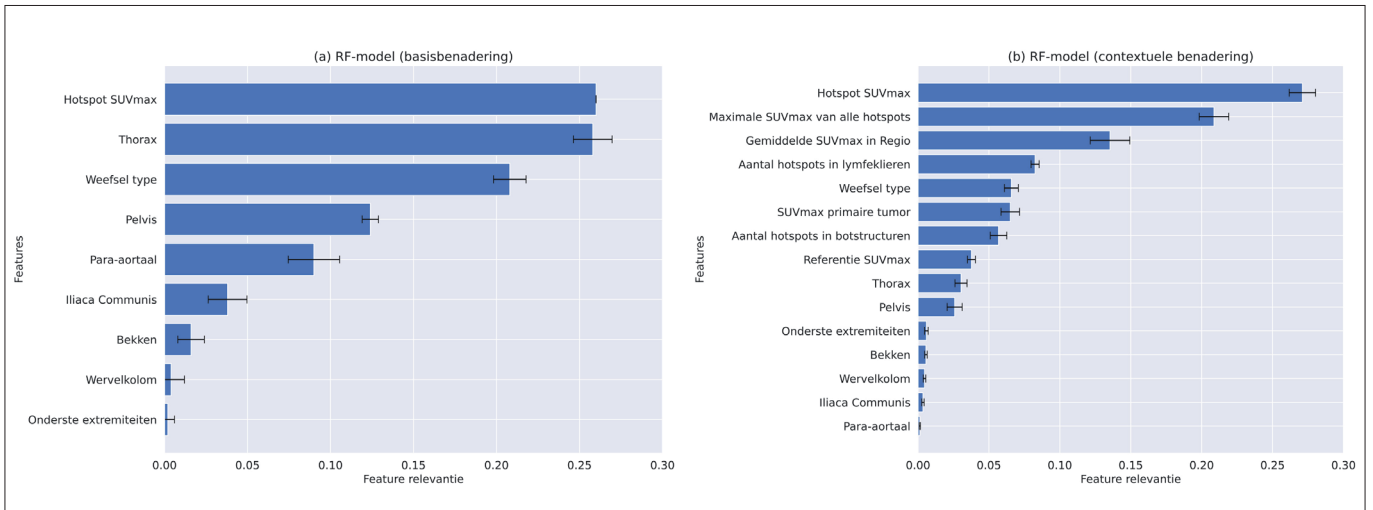
De gemiddelde leeftijd van de geïnccludeerde patiënten was 72 (range 38-84), met een mediane iPSA-waarde van 14 ng/mL (range 2-527). Het klinische TNM-stadium was overwegend cT4 (n=41), gevolgd door cT1 (n=31) en cT2 (n=31). De meerderheid had geen lymfekliermetastasen (N0, n=73 vs N1, n=37) en weinig hadden metastasen op afstand (M0, n=85 vs M1, n=25). International Society of Urological Pathology (ISUP)-scores varieerden, met de meeste patiënten in ISUP 5 (n=51). Hotspots werden waargenomen in 290 botstructuren en 273 lymfeklieren. In totaal waren 119 (41%) botstructuren en 243 (89%) lymfeklieren als maligne gelabeld. Het Random Forest-model (RF-model) presteerde beter dan de Decision Tree- en K-Nearest Neighbors-modellen op minstens twee van de evaluatiestatistieken ($p < 0,001$), voor zowel de basis- als de contextuele benadering. De ROC-curves van beide



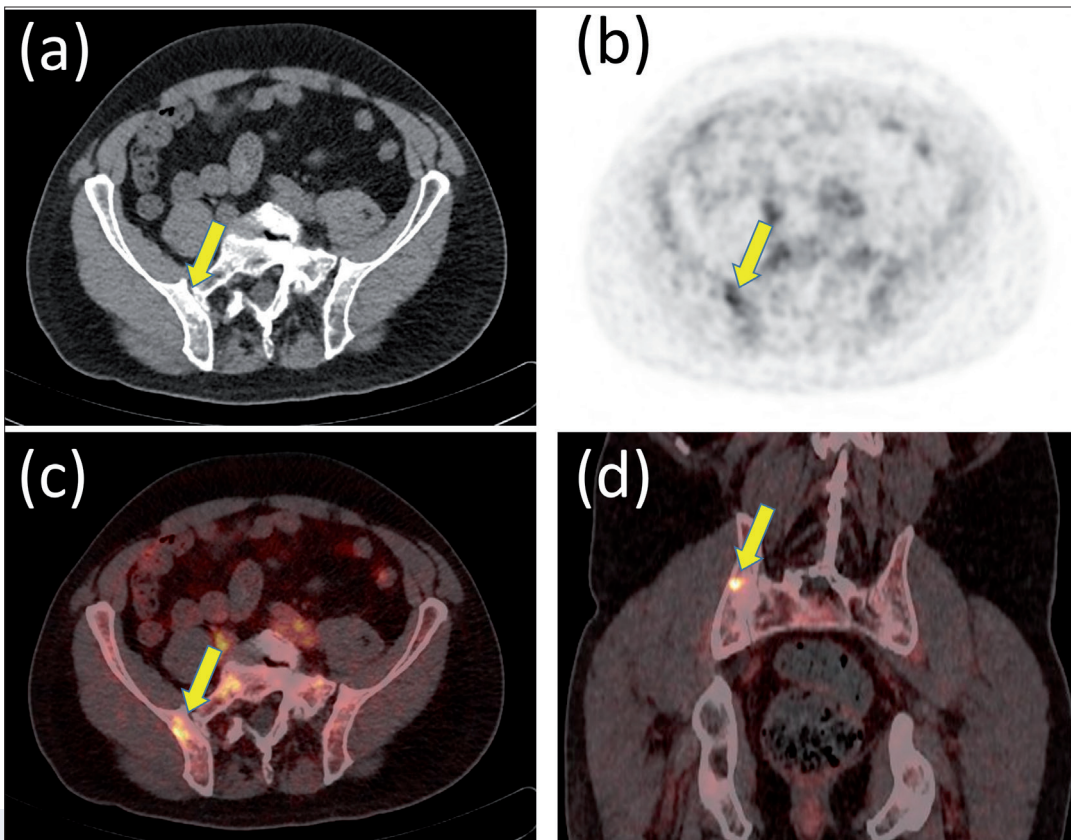
Figuur 2. De ROC-curves van de basisbenadering (blauw) en de contextuele benadering (oranje) van het RF-model. Het contextuele model leidde tot een hogere AUC dan de conventionele basisbenadering (0,96 vs 0,91, $p < 0,001$).



Figuur 3. De voorspelde waarschijnlijkheid op maligniteit per hotspot door het RF-model (contextuele benadering). Elke stip of kruis representeert een hotspot, waarbij een hoge en lage waarschijnlijkheid respectievelijk een maligne en niet-maligne classificatie aangeeft. Een kruis geeft een hotspot weer die anders is geclassificeerd dan door het basismodel ten gevolge van de context-level informatie, waarbij rood en groen respectievelijk een incorrect gewijzigde en correcte gewijzigde classificatie weergeven. De lijn ter hoogte van 0.5 representeert de drempelwaarde, waarbij een hotspot met een waarschijnlijkheid boven deze drempel als maligne wordt geclassificeerd.



Figuur 4. De relevantie van verschillende features voor de basisbenadering (a) en de contextuele benadering (b) van het RF-model. Hoe hoger de waarde, hoe sterker het model afhankelijk is van de desbetreffende feature.



Figuur 5. [¹⁸F]PSMA-1007 (184 MBq) PET/CT scan van een 62-jarige man met prostaatkanker (cT1c), PSA 8 ng/mL, Gleason score 8, ISUP 4: transversale doorsnede CT (a), PET (b), gefuseerde PET/CT (c), en coronale doorsnede gefuseerde PET/CT (d). Geen verdenking op lymfkliermetastasering of metastasering op afstand. Solitair PSMA-focus (SUVmax=7,2) waargenomen in het rechter os ileum (zie pijl), op CT indifferente. Op basis van visuele beoordeling werd een metastase niet uitgesloten. Het basis ML-model scoorde deze hotspot als maligne, het context-level ML-model als waarschijnlijk (87%) niet-maligne, overeenkomstig follow-up data van deze patiënt.

RF-modellen zijn zichtbaar in figuur 2. Het RF-model met contextuele features presteerde beter dan het RF-model met hotspot-specifieke features wat betreft de AUC (0,96 vs 0,91, $p < 0,001$), specificiteit (0,86 vs 0,79, $p < 0,001$) en sensitiviteit (0,91 vs 0,85, $p < 0,001$). Het RF-model was nauwkeuriger in het classificeren van hotspots in lymfeklieren dan in hotspots in botstructuren voor zowel de basisbenadering (78,3% vs 88,8%, $p < 0,001$) als de contextuele benadering (93,6% vs 86,5%, $p < 0,001$).

Het toevoegen van contextinformatie als aanvulling op het basismodel zorgt voor reclassificatie van 15% van de hotspots zoals weergegeven in figuur 3. Het valt op dat het context-level model de classificaties van lymfeklieren alleen wijzigt van maligne naar niet-maligne, wat resulteert in een verhoogde specificiteit en een iets verlaagde sensitiviteit in de classificatie van lymfeklieren.

De SUVmax van de hotspot leverde de meest relevante bijdrage van de verschillende features voor beide RF-modellen, zoals weergegeven in figuur 4. Het valt op dat de features die de locatie van de hotspot omschrijven bijna geen waarde toevoegen aan het model dat ook contextuele features gebruikt. De toegevoegde contextuele features, zoals de maximale SUVmax van alle hotspots en de gemiddelde SUVmax in de regio, dragen in grote mate bij aan de classificatie van de hotspots. Het prestatieverschil tussen de basisbenadering en de contextuele benadering is klinisch inzichtelijk gemaakt middels een voorbeeld, geïllustreerd in figuur 5.

Discussie

In dit onderzoek hebben we laten zien dat ML algoritmen geschikt zijn voor het classificeren van hotspots op een [¹⁸F]PSMA-1007 PET scan, waarbij het toevoegen van contextuele informatie de prestaties significant verbetert.

Het meenemen van een klein aantal contextuele features maakt het model specifiek en sensitiever.

Het best presterende model blijkt beter in het classificeren van maligne lymfeklieren dan van maligne botlaesies. Dit wordt mogelijk veroorzaakt door het feit dat onze dataset procentueel gezien een groot aantal maligne lymfeklieren bevat, waardoor het model de neiging heeft om zich vooral te specialiseren in het herkennen van deze maligniteiten. Eerdere studies naar specifieke opname in botstructuren op [¹⁸F]PSMA-1007-PET/CT vonden een sterke relatie tussen de aanwezigheid van meerdere botmetastasen en de kans op maligniteit van een specifieke hotspot, wat overeenkomt met onze bevindingen (9,10). In tegenstelling tot deze studies laat ons model echter een lage afhankelijkheid zien voor de locatiefeatures (9,10). Deze tegenstrijdige uitkomst wordt mogelijk veroorzaakt door het relatief kleine aantal locatiegroepen die we hebben gebruikt, waardoor de locaties niet specifiek genoeg zijn om informatie toe te voegen.

Het gelimiteerde aantal geïncludeerde patiënten in deze studie kan mogelijk hebben geleid tot een systematische afwijking in de evaluatiestatistieken. Daarnaast kan de subjectieve selectie van hotspots door het ontbreken van een specifiek protocol hebben geresulteerd in suboptimale resultaten. Bovendien kan de onevenwichtige verdeling van maligne en niet-maligne hotspots van invloed zijn geweest op de resultaten. De afwezigheid van een op histologie gebaseerde referentie en het ontbreken van een reproduceerbaarheidsanalyse creëren onzekerheid in de betrouwbaarheid van de labels, waardoor een volledig correcte labelling niet gegarandeerd is en een bias in de resultaten kan ontstaan. Ondanks een eventuele suboptimale verdeling van maligne en niet-maligne hotspots, een mogelijk

niet volledig correcte labelling en daardoor bias in de resultaten, zien we dat de prestaties van het ontwikkelde ML-model verbeteren bij het toevoegen van contextuele informatie. Voor toekomstige verbeteringen raden we aan om radiomische features toe te voegen vanuit de CT-scan, waardoor ook de structurele informatie over het botweefsel kan worden gebruikt. Moazemi et al. voerden een studie uit naar hotspot-classificatiemodellen met behulp van [⁶⁸Ga]Ga-PSMA-PET/CT, waarbij ze ontdekten dat ML-modellen die PET- en CT-radiomische features gebruiken beter presteren dan modellen die features gebruiken van slechts een van de modaliteiten (4). Het gebruik van CT-radiomische features kan leiden tot een verbeterd model voor het classificeren van hotspots in botstructuren. Daarnaast is het van belang om meer specifieke anatomische locatiefeatures toe te voegen om het model verder te verbeteren. Een uitgebreidere patiëntengroep en een objectievere selectie van hotspots zijn wenselijk voor een verbeterde analyse. Ten slotte is het gebruik van referentielabels op basis van histologische data essentieel voor een nauwkeurigere analyse.

Conclusie

We hebben succesvol een ML-model ontwikkeld om lymfeklier- en botmetastasen bij prostaatkankerpatiënten te identificeren met behulp van [¹⁸F]PSMA-1007 PET. Het ontwikkelde model kan door nucleair geneeskundigen en nucleair radiologen worden gebruikt als ondersteuning bij de diagnostiek en verslaglegging. De resultaten van ons onderzoek benadrukken het potentieel van het benutten van contextuele informatie in ML-methoden om de identificatie van metastasen te verbeteren, wat kan leiden tot betere diagnostiek en

daarmee uiteindelijk een betere behandelstrategie voor de patiënt. ♦

Referenties

1. Schlemmer HP, Krause BJ, Schütz V, et al. Imaging of Prostate Cancer. *Dtsch Arztebl Int.* 2021;118:713-9
2. Hoberück S, Löck S, Borkowetz A, et al. Intraindividual comparison of [68Ga]-Ga-PSMA-11 and [18F]-F-PSMA-1007 in prostate cancer patients: a retrospective single-center analysis. *EJNMMI Res.* 2021;11:109
3. Lindgren Belal S, Frantz S, et al. Applications of Artificial Intelligence in PSMA PET/CT for Prostate Cancer Imaging. *Semin Nucl Med.* 2024;54:141-9
4. Moazemi S, Khurshid Z, Erle A, et al. Machine Learning Facilitates Hotspot Classification in PSMA-PET/CT with Nuclear Medicine Specialist Accuracy. *Diagnostics.* 2020;10:622
5. Erle A, Moazemi S, Lütje S, et al. Evaluating a Machine Learning Tool for the Classification of Pathological Uptake in Whole-Body PSMA-PET-CT Scans. *Tomography.* 2021;7:301-12
6. Breiman L. Random forests. *Mach Learn.* 2001;45:5-32
7. Quinlan JR. Induction of decision trees. *Mach Learn.* 1986;1:81-106
8. Cover TM, Hart PE. Nearest Neighbor Pattern Classification. *IEEE Trans Inf Theory.* 1967;13:21-7
9. Phelps TE, Harmon SA, Mena E, et al. Predicting Outcomes of Indeterminate Bone Lesions on 18F-DCFPyL PSMA PET/CT Scans in the Setting of High-Risk Primary or Recurrent Prostate Cancer. *J Nucl Med.* 2023;64:395-401
10. Arnfield EG, Thomas PA, Roberts MJ, et al. Clinical insignificance of [18F]PSMA-1007 avid non-specific bone lesions: a retrospective evaluation. *Eur J Nucl Med Mol Imaging.* 2021;48:4495-507