

'Federated learning heeft enorme impact op medisch-wetenschappelijk onderzoek'

16 februari 2023



Onderzoek naar het sneller en eerder vinden van hersentumoren is niet nieuw. Ook AI wordt al jaren toegepast om op dit gebied betere resultaten te boeken. Het Erasmus MC kent bijvoorbeeld de biomedical imaging group Rotterdam (BIGR), die werkt aan geautomatiseerde en AI-toepassingen op het gebied van beelddiagnostiek. Maar federated learning, waarbij volgens FAIR-principes¹ toepassingen naar de data gaan in plaats van andersom, was nog onontgonnen terrein.

Reden voor het Erasmus MC (Rotterdam) om in 2020 aan te haken bij een grootschalig onderzoek naar in welke mate federated learning kan bijdragen aan AI-gebaseerde medisch wetenschappelijke onderzoeken (zie kader rechts). Neuroradioloog en hoogleraar in de neuroradiologie Marion Smits kende de hoofdonderzoeker van de Penn-studie, Spyridon Bakas, onder meer van een gezamenlijk consortium dat op diverse manieren probeert om hersentumoren in kaart te brengen (voor het onderzoeksdeel van Smits' functies ligt de focus op hersentumoren).

Eerste kennismaking

“Toen Bakas twee jaar geleden het federated learning onderzoek naar deze tumoren aan het opzetten was, heeft hij ook ons erbij gevraagd”, vertelt Smits. “Voor onze onderzoeksgroep was dit de eerste kennismaking met federated learning. Ook voor het Erasmus MC was het nog een vrij nieuwe vorm van AI-toepassing.”

Inmiddels zie je meer onderzoeken waarin federated learning wordt meegenomen, maar toen het Erasmus MC ermee begon, was het nog nieuw, vertelt biomedisch onderzoeker Sebastian van der Voort, actief in de onderzoeksgroep van Smits. Zijn focus bij zijn promotieonderzoek

lag op de toepassing van AI-technologie om informatie over mogelijke hersentumoren te verkrijgen die je normaal via biopten zou krijgen.

“Er zaten onder meer technische haken en ogen aan om het onderzoek op te zetten. Normaal gaat onze data naar een algoritme, plaats daarvan kwam het algoritme nu naar onze data. De software draaide dus binnen de Erasmus MC infrastructuur, onze ICT- en security-mensen hadden daar nog wel het een en ander over te zeggen. Nu is alles goed geregeld, er is een protocol opgesteld voor federated learning, zodat dit op technisch gebied geen problemen meer oplevert.”

Eenvoudiger toegang tot data

Toegang tot grote datasets is vaak een belangrijk probleem voor medisch onderzoekers, vooral wegens privacy-regels. Op zichzelf levert federated learning volgens Van der Voort niet toegang tot meer data op. Maar omdat iedereen de data in eigen beheer kan houden, is men eerder geneigd om toegang tot hun data te geven. “Ze hoeven die data niet meer uit handen te geven, wat toch altijd privacygevoelig is, maar kunnen een AI-toepassing onder hun controle op de eigen data zetten. Zo wordt toegang tot data veel eenvoudiger.”

“Vaak zijn onderzoeksgegevens ook beperkt toegankelijk omdat het oudere data betreft, zeker bij zeldzaam voorkomende tumoren. Tegenwoordig is het vrij normaal om mensen ook toestemming te vragen bij het gebruik van data voor medisch-wetenschappelijk onderzoek, maar dat is niet altijd zo geweest”, voegt Smits toe. “Niet verkregen toestemming om data te mogen delen is minder een probleem wanneer die gegevens de eigen servers niet verlaten. Federated learning omzeilt zo het probleem dat privacy-regulering soms betekent.”

Ook de diversiteit in datasets bleek groter te zijn bij federated learning, aldus Van der Voort. “Wat je voor ons soort onderzoek nodig hebt, is een goede mix in data. Stel dat je van 50.000 mensen data hebt, maar allemaal van onze scanners, dan zal een daarop ontwikkelde AI-toepassing in een ander ziekenhuis minder goed werken. Met onze studie krijg je bijdragen van zo’n grote hoeveelheid instituten, dat de databronnen heel divers zijn. Een daarop ontwikkelde AI-toepassing is veel robuuster, veel beter breed inzetbaar.”

Bijzondere uitkomsten

Wie het onderzoeksartikel leest over de federated learning-studie, realiseert zich misschien niet meteen hoe bijzonder de uitkomsten zijn, weet Smits. “Het klinkt logisch, dat we een robuust, generaliserende AI-toepassing hebben ontwikkeld. Maar het is wel bijzonder, want het lukt vaak zat niet. Er hebben 71 centra deelgenomen, wat zowel heel veel als hele diverse data heeft opgeleverd. Zij hebben allemaal geleerd hoe je een federated learning-omgeving kunt opzetten. Dit soort datasets zijn echt de toekomst, niet langer de zelf verzamelde datasets met het zelf ontwikkelde algoritme dat dan de wereld in wordt gestuurd. Federated learning zet een nieuwe standaard.”

Het raamwerk van federated learning zoals de onderzoeksgroep dat heeft opgezet, is publiekelijk beschikbaar. Elke onderzoeksgroep kan dat gaan gebruiken voor hun eigen federated learning toepassingen. “Wij weten dat het voor een brede groep instellingen werkt, dus het raamwerk heeft zichzelf bewezen”, benadrukt Van der Voort. “Datzelfde geldt voor de robuustheid van de AI-toepassing. We hebben een aantal deelnemers niet meegenomen in het trainen van het algoritme. Na de ontwikkeling ervan hebben we het algoritme bij hen toegepast en daar werkte het even goed als bij de rest.”

Goede eerste stap

De technologie en het raamwerk zijn dus solide, maar Van der Voort verwacht dat het voorlopig een lastige drempel voor nieuwe federated learning-onderzoeken zal blijven om potentiële deelnemers er van te overtuigen dat het een veel veiliger en privacy-bestendiger manier is om hun data in te zetten. “Er zullen nog nieuwe geslaagde federated learning onderzoeken nodig zijn om die horde weg te nemen. Maar we hebben een goede eerste stap gezet.”

Smits is ervan overtuigd dat die drempel tijdelijk zal zijn, zodra onderzoekers en instellingen beseffen hoeveel hordes er weg genomen worden om data te delen. “Je weet dat data in eigen beheer blijft en alleen gebruikt worden voor het specifieke onderzoek. Je weet dat jij er de credits voor krijgt voor wat betreft jouw bijdrage. En omdat er geen enkele plek is waar alle data bij elkaar staan, kan niemand anders er iets mee gaan doen, of kan er door een hack veel data op straat komen te liggen. Dit opent denk ik juist de deuren voor meer grotere onderzoeksprojecten.”

Ook voor niet-onderzoeksinstellingen wordt deelname veel eenvoudiger, voorziet Smits. Nu is dat vaak lastig. De data moeten goed geanonimiseerd zijn, de toestemming moet goed geregeld zijn, de data delen moet veilig gebeuren. Dus vaak blijven datasets beperkt tot die van academische ziekenhuizen. “Straks kan iedereen meedoen, ook uit landen waarin de zorg anders geregeld is, en kunnen normaal ondervertegenwoordigde sociaal-economische klassen beter meegenomen worden. Dat helpt ook enorm in het voorkomen van bias in je datasets of je AI-toepassing.”

Inzetten in praktijk

Concreet heeft het onderzoek een algoritme opgeleverd dat in theorie ingezet kan worden om onderzoek naar hersentumoren sterk te verbeteren. Smits wil het dan ook het liefst zo snel mogelijk in de klinische praktijk in zetten.

“Zo’n toepassing kan de grootte van tumoren beter en sneller bepalen dan dat ik dat kan. Veel van wat wij ontwikkelen, is echter vooral wetenschappelijk interessant, maar nog niet geschikt voor toepassing op gebruikersniveau. Daarin kunnen we met commerciële partijen sparren. Zij hebben die kennis wel en kunnen een toepassing helpen doorontwikkelen naar de praktijk. Dus ik denk dat samenwerkingsverbanden zoals tussen Intel en Penn Medicine goede tandems vormen. Het is evident dat je daarbij wel belangenverstremming moet voorkomen.”

Over de studie

Om de behandeling van ziekten te verbeteren, moeten onderzoekers toegang krijgen tot grote hoeveelheden medische data – in de meeste gevallen datasets die de drempel overschrijden die één instelling kan produceren. De toegankelijkheid van data is echter al lang een probleem in de gezondheidszorg vanwege wetten inzake databescherming, zoals in Nederland de AVG. Hierdoor was het bijna onmogelijk om medisch onderzoek en gegevensuitwisseling op grote schaal te realiseren zonder de gezondheidsinformatie van patiënten in gevaar te brengen.

In 2020 kondigden Intel en Penn Medicine een overeenkomst aan om samen te werken en federated learning te gebruiken om tumordetectie te verbeteren en de behandelingsresultaten te verbeteren van een zeldzame vorm van kanker – glioblastoma (GBM). Het onderzoek werd gefinancierd door het Informatics Technology for Cancer Research programma van het National Cancer Institute van de National Institutes of Health. In de federated learning studie werd een wereldwijde dataset uit 71 instellingen op zes continenten onderzocht.

Intel's federated learning-hardware en software voldoen volgens deze IT-aanbieder aan de eisen van dataprivacy en beschermen data-integriteit, privacy en veiligheid door middel van vertrouwelijk computergebruik.

Lees [hier](#) het artikel over de studie in Nature Communications.

Referentie

1. Artikel over toepassing van FAIR-principes in de zorg (2018): [Link](#)