

OREZTA ALA MINBIZIA?

HIRU MELANOMATIK BATEAN ERRATU ETA AZAL ILUNEKO PAZIENTEAK ALBORATZEN DITUEN ALGORITMOA

Osakidetza lanean ari da Asisa aseguru-etxe pribatuaren filial baten Quantus Skin sistema osasun-zentroetan ezartzeko, 1,6 milioi euroko inbertsioa eginda. Adituek kritikatu egin dute adimen artifizialeko sistema hori, emaitza "eskasak" eta "arriskutsuak" dituelako. Algoritmoa paziente zuriekin bakarrik dago trebatuta.

 CIVIO - CC-BY-SA

Denbora urrea da. Bereziki melanomaz ari bagara, azaleko minbizirik arriskutsuena. Tumore hori lehenbailehen diagnostikatzearabakigarria da bizitzak salbatzeko, ia beste edozein minbizitan baino gehiago. Espainiako Estatuan, kalkulatzen da 2025ean 9.400 melanoma kasu inguru izango direla. Tumore oso oldarkorra da, azkar barreia daiteke eta hilabete gutxi batzuetan metastasia eragin dezake. Hori gertatzen denean, pronostikoa txarra izaten da, eta, beraz, detekzioan egindako edozein akats zorigaiztokoa izan daiteke.

Hain zuzen ere, premia horrek bultzatu du EAEn adimen artifizialaren al-

deko apustua egitera. Osakidetza lanean ari da bere osasun-zentroek eta ospitale publikoek Quantus Skin erabil dezaten, azaleko minbiziaren arriskua –melanoma barne– diagnostikatzeko diseinatutako algoritmoa. Teorian, prozesua arintzen du: lehen mailako arretatik, familiako medikuek ospitaleko dermatologia zerbitzura bidali ahal izango dituzte lesio susmagarrien irudiak, baita gaiztoak izateko probabilitatea ere, algoritmoak automatikoki kalkulatuta. Eusko Jaurlaritzaren asmoa da Quantus Skin-ek, gaur egun probetan dagoena, lehenengo zein paziente artatu behar diren erabakitzen laguntzea.

Hala ere, datuek errealitate kezkarria erakusten dute. Transmural Bio-

tehek, Quantus Skin merkaturatzen duen enpresak, hasierako azterketa bat egin zuen eta emaitza baikorrak lortu zituen, baina muga esanguratsuak zituen: erabat online egin zen eta ez zen inongo aldizkari akademikotan argitaratu, hau da, ez zuen zientzian eskatzen den ohiko kalitate-kontrola gainditu.

Geroago, Madrilgo Ramón y Cajal ospitaleko dermatologoek eta Complutense Unibertsitateko irakasleek bigarren azterketa bat egin zuten, eta argitaratu egin zuten Quantus Skin benetako eraginkortasun klinikoa ebaluatzeko. Lan horrek, Transmural Biotech enpresaren finantzaketa eta laguntza teknikoak izan zituenak, emaitza txarragoak eman zituen: algorit-



GETTY IMAGES

moak hiru melanomatik bat alde batera uzten du. Sentikortasuna %69koa da, eta horrek esan nahi du hilgarria izan daitekeen minbizi horren benetako kasuen %31 ez dituela atzematen.

Transmural Biotecheko zuzendari nagusi David Fernández Rodríguezek emaitza txarrak lotu zituen irudiak hartzean izandako akatsekin, Quantus Skin-ek emandako jarraibideak ez zirelako bete. EAEko probetan ere ikusi duten zerbait da: "Lehen arretako medikuak ez daude ondo trebatuta irudiak, argazkiak, egiteko", dio, eta horrek "medikuak prestatzeko beharra" adierazten du. Hala ere, bigarren azterlanean dermatologoek parte hartu zuten, hain zuzen ere lesio susmagarriei argazkiak ateratzen espezializatuta daudenak. Fernández Rodríguezen arabera, "irudiak ondo

moztu ondoren" fidagarritasuna hobetu egin zen, "ez zirelako aginduak zehatz-mehatz betetzen ari".

ITURRI INDEPENDENTEEK KRITIKATUTAKO SISTEMA

"Azaleko minbiziarentzat", %70eko sentzibilitatea izatea "oso txarra da". Bartzelonako Clinic Ospitaleko Larruazaleko Minbiziaren Unitateko zuzendaria da Josep Malvehy Guilera medikua. Datu "oso eskasa" dela dio: "Norbaiti tresna hori ematen badiozu argazki bat egin eta melanoma bat izan daitekeen ala ez atzemateko, eta hirutik batean erratzen bada, ez da egokia lehen arretako ingurune batean azaleko minbizi bahetzeko, gehiago eskatu behar diozu". Mallorcako Son Llàtzer Ospitaleko dermatologo eta Dermapixelen egile Rosa

Taberner Ferrer-en ustez, "negatibo faltsuak %31 izateak arriskutsua dirudiela ezin uka. Baheketa gisa kaskarra da oso".

Hala ere, Transmural Biotecheko zuzendari nagusia arazoa minimizatzen saiatu da. Bere produktuari mesede egiten dioten datuak soilik oinarri hartuta, Quantus Skin-en sentzibilitate baxuaz hitz egitea saihestu du. Bere eraginkortasun klinikoa aztertu zuen lan beraren arabera, sistemak bi aldiz huts egiten du, gainera: bere espezifikotasunak positibo faltsuen %19,8 eragiten du, hau da, kalteberak diren bost orbanetatik bat melanomarekin nahasten du. Horrek esan nahiko luke Quantus Skin-ek gomendatuko lukeela artatutako pertsonen ia %20 beharrik gabe espezialistengana bideratzea.



M. PÉREZ - EFE

1,6 milioi euroko kontratu publikoa Asisarentzat

2022an, Osakidetza ia 1,6 milioi euroko kontratua esleitu zion Transmural Biotech enpresari, "irudi medikoetan trebatutako adimen artifizialeko algoritmoak" martxan jartzeko. "Gutxienez" %85eko sentsibiltatea eta espezifikotasuna exijitu zuten. Bartzelonako Unibertsitateko eta Clinic Ospitaleko spin-off enpresa bat da, Asisa aseguru-etxe pribatuaren jabetzakoa. Osakidetzaaren arabera, baldintza-agirietan hainbat algoritmo jaso zituzten arren, azkenean bi bakarrik aukeratu ziren, "asistentzia-inpaktu" eta "osasun-errendimendu" handiagoak zituztelako. Horietako bat Quantus Skin zen; Civiok jakin ahal izan duenez, erabakia alde bakarrez hartu zen, dagozkion espezialistekin kontsultatu gabe. Otsailean, Osakidetza ziurtatu zuen

Quantus Skin-ek "baliozkotze-faseak" gaingitu eta "integrazio-fasean" zegoela. Bere eraginkortasun klinikoari buruz Civiok egindako galderei erantzunez, orain Osakidetza adierazi du probetan jarraitzen duela eta "emaitzen arabera" hartuko duela erabakia. Hala ere, Osakidetza arduradunek ez dute erantzun ba ote zekiten Quantus Skin-en eraginkortasun klinikoari buruz argitaratutako datuak (%69,1eko sentikortasuna eta %80,2ko espezifikotasuna) kontratuak eskatzen duen % 85aren azpitik zeudela. EAEko esleipenaz gain, Transmural Biotech-ek beste kontratu publiko bat baino ez du Katalunian, askoz zenbateko txikiagokoa (25.000 euro), erradiologiako adimen artifizialeko algoritmoak ziurtatzeko.

Azterlanean, egileek -Madrilgo Ramón y Cajal Ospitaleko dermatologoak eta Madrilgo Complutense Unibertsitateko irakasleak- defendatzen dute hobe dela Quantus Skin-ek espezifikotasun handia izatea (positibo faltsu gutxi), sentsibiltate txikiaren kontura bada ere (negatibo faltsu gehiago), ez baita aplikatuko ziurtasunez diagnostikatzeko, baheketa gisa baizik. Hau da, lehen mailako arretatik kasuak iragazten laguntzeko. Hipotesi

horren arabera, kontsulta espezializatuak saturatzea saihestu daiteke horrela, eta itxaron-zerrendak eta gastuak murriztu.

Civiok kontsultatutako adituek zailtasun jarri dute algoritmoaren atzeko estrategia. Minbizia diagnostikatzeko estandar idealik ez dagoen arren -neurri batean tumore bakoitzaren agresibitatearen arabera- delako, Quantus Skin-ek lortutakoa onargarria izatetik

urrun dago. "Azkar hazteko eta pazientea hiltzeko arrisku handia duten lesioetan, melanomaren diagnostikoa egitean erratuz gero, horretan ez naiz oso tolerantia izango. Gutxienez %92, %93 eta % 94ko sentsibiltateak eskatu behar ditut", dio Bartzelonako Clinic Ospitaleko Azaleko Minbiziaren Unitateko zuzendari Josep Malvehy Guilerak.

"Bahetze gisa erabili nahi badute, sistemak sentikortasun super handia izan

beharko luke, espezifikotasun txikixeago baten kontura", azaldu du Taberner Ferrerrek. Beste era batera esanda, hobe da horrelako algoritmo bat behar baino zuhurragoa izatea: hobe da batzuetan pertsona osasuntsuengan alarma faltsua sortu eta huts egitea, benetako minbizi kasu bat alde batera uztea baino.

LARRUAZAL ILUNA, DIAGNOSTIKO EEZAGUNA

Quantus Skin-en arazoak bere sentsibilitate txikitik haratago doaz. Azterlanak melanomaren diagnostikoaren eragin-kortasun klinikoa ebaluatu zuen, ez ordea azaleko minbizi ohikoagoak baina hain erasokorrak ez direnak, hala nola kartzinoma basozelularra eta zelula ezkatadunen kartzinoma. Horietan ere Quantus Skin erabili daiteke. Egileek ez zuten aztertu nola eragiten dion larruazalaren koloreak algoritmoaren funtzionamenduari, nahiz eta aitortzen duten hori dela beren ikerketaren muga nagusietakoa.

Quantus Skin sare neuronalatan oinarrituta dago eta azaleko minbiziak detektatzen ikasi du ia bakarrik pertsona zuriengan. Algoritmoa, lehenik eta behin, International Skin Imaging Collaboration (ISIC) erakundearen 56.000 irudi baino gehixeagorekin elikatu zen. Mediku-argazkien gordailu publikoa da, batez ere Mendebaldeko ospitaleetan bilduak, eta gehienak larruazal argiak dituzten pazienteenak dira. Ondoren, Madrilgo Ramón y Cajal ospitaleko 513 pazienteren irudiak erabiliz probatu zen Quantus Skin. Guztiak zuriak ziren.

Quantus Skin elikatzeko erabili zen datu-multzoak "gizon eta emakume kaukasiarren" irudiak ditu, Transmural Biotecheko zuzendari nagusiak baieztatu duenez. "Ez dut nahi gutxiengo etnikoen gaian eta horretan guztian sartu, tresna Osakidetzak erabiltzen baitu. Eskura jartzen dudana tresna bat da, dituen mugekin", dio Fernández Rodríguezek. Azal ilunagoetan trebakuntzarik ez badauka ere, Eusko Jaurlaritzak adierazi du ez dela beharrezkoa "inola-ko neurririk ezartzea" berdintasuna eta diskriminaziorik eza bermatzeko, EAEko algoritmoen eta adimen artifizialeko sistemen katalogoan dagoen Quantus Skin-eko fitxan diotenaren arabera. Hala ere, sare neuronalak ia eskusiboki pertsona zurien irudiekin entrenatu direnez, litekeena da azal ilunagoetan akats gehiago izatea, hala nola ijito-et-

niako gizabanakoetan eta Latinoamerikako edo Afrikako jatorria dutenengan.

"Oso erraza da algoritmoak huts eginaraztea", azaldu dio Civiori AEBetako Texasko Unibertsitateko dermatologia irakasle Adewole Adamsonnek. Izan ere, 2018an ohartarazi zuen adimen artifizialak diskriminazioa eragin zezakeela, modu inklusiboan eta askotarikoan garatzen ez bazen. Quantus Skin baino haratago doan arazoa da.

**"Azkar hazteko eta
pazientea hiltzeko
arritsu handia duten
lesioetan, melanomaren
diagnostikoa egitean
erratz gero, horretan ez
naiz oso tolerantzia izango.
Gutxienez %92, %93 eta
% 94ko sentsibilitateak
eskatu behar ditut".**

Josep Malvehy Guilera,
Bartzelonako Clinic Ospitaleko
Azaleko Minbiziaren Unitateko
zuzendaria

Haren iragarpenak baieztatu egin dira. Dermatologian, algoritmoak batez ere paziente zurien irudiekin elikatzen direnean, "azal ilunetan fidagarritasun diagnostikoa" murrizten da, Taberner Ferrerrek dioenez. Suediako First Derm enpresaren Skin Image Search algoritmoak, batez ere larru zurien argazkiekin trebatuak, %70etik %17ra murriztu zuen bere zehaztasuna azal iluna zuten pertsonekin erabili zenean. Ikerketa berriagoek baieztatu dute horrelako algoritmoek okerrago funtzionatzeko dutela pertsona beltzetan, eta hori ez da arazo

teknikoengatik gertatzen, entrenamendu-datuak dibertsitate faltagatik baizik.

Melanoma pertsona zurietan askoz ere ohikoagoa den minbizi bada ere, kasu horietan azal iluna duten pertsonak biziraupen global txikiagoa dute. Avery Smith ingeniari estatubatuarra ondo ezagutzen ditu zenbaki horiek. Bere bikotekideari, Latoya Smithi, melanoma bat atzeman zioten ezkondu eta urte eta erdira. "Asko harritu ninduten etniaren araberrako biziraupen-tasek. Latoya, afroamerikarra izanik, behealdean zegoen. Nik ez nekien hori, bete-betean jo ninduen arte. Beldurgarria izan zen", kontatu dio Civiori. Diagnostikoa egin ostean, 2011 amaieran, Latoya hil egin zen.

Onduz geroztik, dermatologia inklusiboagoa lortzeko eta algoritmoek desberdintasunak ez handitzeko lanean ari da Avery Smith. Bereiziki talde kalteberetan algoritmoak izan dezaket "in-paktuaz" ohartarazteko, Smithen ez du adimen artifiziala "tresna" gisa aipatzen, guraize soil bat izango bailitzan: "Marketerako terminoa da, jendeak ulertzeko modu bat. Baina adimen artifiziala askoz gehiago da".

Anabel K. Arias legelaria ere, Kontsumitzaile eta Erabiltzaileen Federazioko (CECU) bozeramailea, ondorio horiei buruz mintzo da: "Diagnostiko goiztiarra egiteko adimen artifiziala erabiltzean, aitzat hartu behar da litekeena dela biztanleriaren zati bat azpiordezkatuta egotea. Kasu horretan, diagnostikoa okerra izan daiteke eta eragina izan dezake pertsonaren osasunean; kaltea ere sor dezake".

ALGORITMOAREN AURREAN IKUSEZIN

"Pertsonok konfiantza handia dugu adimen artifizialaren gainean, benetakoak ez diren objektibotasun-ezaugarri batzuk ematen dizkiogu", dio Deustuko Unibertsitateko psikologia esperimentaleko katedradun Helena Matute Greñok. AA orok jasotzen duen informazioa erabiltzen du erabakiak hartzeko. Datu horiek onak ez badira edo osatu gabe badaude, baliteke huts egitea. Sistematikoki erratzen denean, algoritmoak akatsak egiten ditu, eta horiei 'alborapen' esaten diegu. Eta akats horiek pertsona-talde jakin bati gehiago eragiten badiote -jatorriagatik, azalaren koloreagatik, generoagatik edo adinagatik- alborapen diskriminatzaileez ari gara



EH11 KOLORE EKIMENAK 2016 BILBON EGINDAKO KALEJIRA BAT. TXENG MENG

Quantus Skinek kontuan hartzen ez duen aniztasuna

2025. urtearen hasieran, EAEn atzeritar jatorriko 316.942 pertsona zeuden, Ikuspegi Immigrazioaren Euskal Behatokiaren datuen arabera. 60.000 baino gehiago Magrebetik eta Saharaz hegoaldeko Afrikatik etorritakoak dira, eta

164.000 pertsona inguru, berriz, Latinoamerikatik, azal-tonuetan aldaera asko duen zonaldea. Horiei gehitu beharko litzaizkieke azal zuria ez duten pertsonak, hemen jaio eta bizi direnak.

Journal of Clinical Epidemiology-n argitaratutako txosten batek erakutsi zuen medikuntzarako AARI buruzko azterlanen %12k bakarrik aztertzen zuela ea alborapenik ba ote zuen. Alborapen ohikoena arrazarena zen, eta ostean generoarena eta adinarena; eta gehienak historikoki diskriminazioak jasan zituzten taldeei eragiten zieten. Akats horiek gerta daitezke entrenamendutatuak ez badira behar bezain anitzak eta orekatuak: algoritmoek biztanleriaren zati batengandik bakarrik ikasten badute, okerrago funtzionatzen dute bestelakoak diren pertsona edo talde minoritarioetan.

Akatsak ez dira larruazalaren kolore mugatzen. Aurpegia ezagutzeko teknologia komertzialek askoz ere huts gehiago egiten dute emakume beltzak

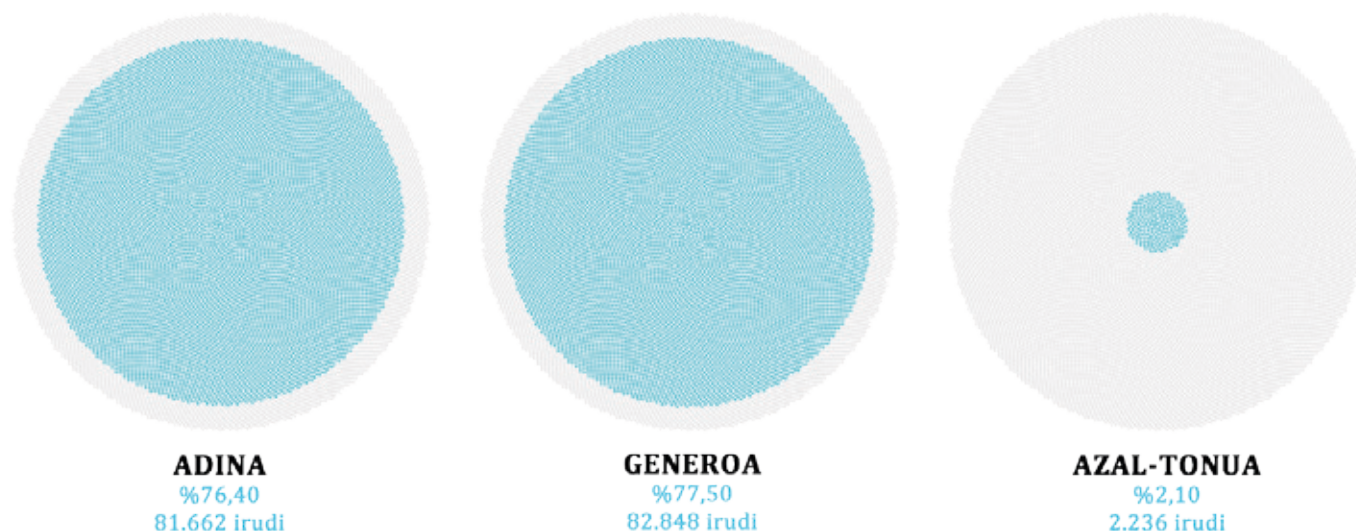
sailkatzerakoan, historikoki gizon zurien irudiekin entrenatu direlako. Antzeko zerbait gertatzen da toraxeko erradiografiak aztertzen dituzten algoritmoekin edo gaixotasun kardiobaskularrak atzematen dituztenekin. Horien errendimendu diagnostikoa okerragoa da emakumeengan, entrenamendutatuak desorekatuta badaude. Bestalde, gibleko gaixotasunak aurreikusteko gehien erabiltzen den datu-multzoetako bat erabat alboratuta dago –datuen %75 gizonenak dira–, eta, beraz, algoritmoek askoz ere huts gehiago egiten dute emakumeekin. Britania Handian, transplanteak lehenesteko algoritmoak pertsona gazteenak diskriminatzen zituen. Zein da arrazoa? Datu mugatuekin entrenatu izana. Hurrengo bost urteetarako biziraupena bakarrik hartzen bai-

tzuten kontuan, eta ez organo berri bat jasoko zuten pazienteek irabaziko zuten bizitzaren iraupena.

Gauza bera gertatzen da adinarekin lotutako alborapenekin. Horiek bereziki problematikoak dira pediatrian.

ALBORAPENAK SAIHESTEA: OZTOPO-LASTERKETA

Alborapenak saihesteko irtenbidea egon badago: "Entrenamendu multzoak ahalik eta zabalena behar du izan", azaldu du Erradiologia Medikoaren Espainiako Elkarteko bozeramaile eta Bartzelonako Bellvitge Unibertsitate Ospitaleko erradiologo Antonio López Ruedak. Baina hau ezin da beti egiaztatu. Orain arte, mediku-irudiak erabiltzen dituzten Espainiako Estatuan ezarritako adimen artifizialeko sistema gehienek ez dituzte entrenamen-



ITURRIA: DAVID WEN, SAAD M KHAN, ANTONIO JI XU, HUSSEIN IBRAHIM, LUKE SMITH, JOSÉ CABALLERO ET AL. (2022). CHARACTERISTICS OF PUBLICLY AVAILABLE SKIN CANCER IMAGE DATASETS: A SYSTEMATIC REVIEW. THE LANCET DIGITAL HEALTH 4(1): E64-E74.

Datu-base mediko gehienek ez dute larruazalaren kolorea erregistratzen

Hogeita bat datu-base publikotan jasotako 106.950 irudi dermatologikoei egindako azterketa batek erakutsi zuenez, pazienteen larruazalaren tonua ozta-ozta jaso zuen %2,1ek.

du-datuak argitaratzen. Eta datu-multzoak eskuragarriak direnean, beste arazo bat da metadaturik ez jasotzea –hala nola jatorria, generoa, adina edo larruazal mota–, zeintzuek lagunduko luketen inklusibo eta orekatuak diren egiaztatzen.

"Teoriak dio nire biztanleriaren %90ak larru zuria badu, azal mota horrekin entrenatu behar dudala, prebalentzia ere garrantzitsua baita erabakiak hartzerakoan; nire inguruan entrenatu behar dut", azpimarratu du José Juan Pereyra Rodríguez Sevillako Virgen del Rocío Unibertsitate Ospitaleko doktoreak. Kanpoan garatutako sistemen kasuan, egokiena da ospitaleek algoritmoen errendimendua ebaluatzea beren paziente-taldeetan. Lopez-Rueda ere horren aldekoa da, edozein adimen artifizial martxan jarri aurretik "tokiko datuekin berriz entrenatzearen" alde egiten baitu.

"Alborapenak dituzten algoritmoek teknologia iraultzaile horren onura potentzialak kentzen dizkiete pazienteei"

Adewole Adamson,
dermatologiako irakaslea
Texasko Unibertsitatean

"Oso garestia da enpresarentzat eta ospitalearentzat, baina hori da benetan funtzionatuko lukeena", dio erradiologoak.

Biztanleriaren ezaugarriak posta-kodearen arabera ere aldatu egiten dira "Clinic Ospitalean [Bartzelonako erdigunean] software bat garatzen badut eta Bellvitgen [kanpoaldean] ezartzen badut, ez dit balio; eta alderantziz ere ez", dio Lopez-Ruedak. Alborapen algoritmikoen ondorioak benetan kaltegarriak izan daitezke: diagnostiko desegokiak egiteagatik pazienteei kalte egitea. "Alborapenak dituzten algoritmoek teknologia iraultzaile horren onura potentzialak kentzen dizkiete pazienteei", dio Adamson, eta gaineratu du: "Arazoaren sustraia ez dago algoritmoetan, horien diseinuari eta garapenari eskainitako hausnarketan eta zaintzan baizik". ●