

Strojno učenje

- Deset milijonov evrov vreden projekt, koordiniran iz Slovenije, bo v prihodnjih petih letih zaposloval 50 podoktorskih raziskovalcev z vsega sveta
- Med raziskovanimi področji tudi napovedovanje višine gladine morja, personalizirana medicina in raziskovanje komunikacije delfinov

Tekst: **Katja Željan**



Fotografije: Shutterstock

Na Agenciji RS za okolje se bodo v okviru projekta SMASH ukvarjali z napovedovanjem in razumevanjem ekstremnih dogodkov v atmosferi, na kopnem in v morju.

Univerzi v Novi Gorici je skupaj s partnerji – Univerzo v Ljubljani, Inštitutom Jožef Stefan, Agencijo RS za okolje in Inštitutom informacijskih znanosti Maribor – uspelo pridobiti koordinacijo t. i. COFUND projekta v vrednosti 10 milijonov evrov, ki ga sofinancira program za raziskave in razvoj Obzorje Evropa. V njem bo kar 25 pridruženih partnerjev, med drugim univerze v Berkeleyju, Amsterdamu, Washingtonu in Londonu ter Inštitut CERN, s projektom SMASH pa nameravajo v naslednjih petih letih pritegniti 50 vrhunskih podoktorskih raziskovalcev in raziskovalcev z vsega sveta, vključno s slovenskimi raziskovalci in raziskovalkami v tujini, ki bodo z uporabo metod strojnega učenja, umetne inteligence in superračunalnika Vega raziskovali na različnih področjih naravoslovja in humanistike. Z delom naj bi začeli v letošnjem juliju.

Kot pravi dr. **Gabrijela Zaharijaš**, glavna vodja in koordinatorica projekta SMASH na Univerzi v Novi Gorici, je novogoriška univerza glede na svojo velikost že doslej pridobila veliko domačih in tujih raziskovalnih projektov, je pa SMASH zagotovo največji med njimi in hkrati največji med tistimi, ki jih je Evropska komisija doslej odobrila Sloveniji. Rektor Univerze v Novi Gorici dr. **Boštjan Golob** pojasnjuje, da je SMASH v principu namenjen osnovnim raziskavam v znanosti, razširili pa ga bodo tudi na bolj aplikativna področja. V ospredju bo pet znanstvenih področij, in sicer splošno strojno učenje za znanstveno uporabo, aplikacije strojnega učenja pri odprtih vprašanjih v astrofiziki, strojno učenje pri raziskavah podnebnih sprememb, računalništvo kot pomoč v raziskavah človeške in živalske komunikacije ter aplikacija orodij umetne inteligence na področju personalizirane medicine in ved o življenju.

Prikaz računalniško simuliranega trka gruč protonov v detektorju ATLAS na nadgrajenem HL-LHC, ko se bo sočasno lahko zgodilo tudi 200 trkov (reakcij), ki so na prikazu obarvani modro. Iz vsake reakcije nato letijo na novo nastali delci v detektor in naloga računskih znanstvenih metod je razločiti in razumeti vseh 200 reakcij. Očitno je, da so nam lahko metode strojnega učenja v veliko pomoč ...

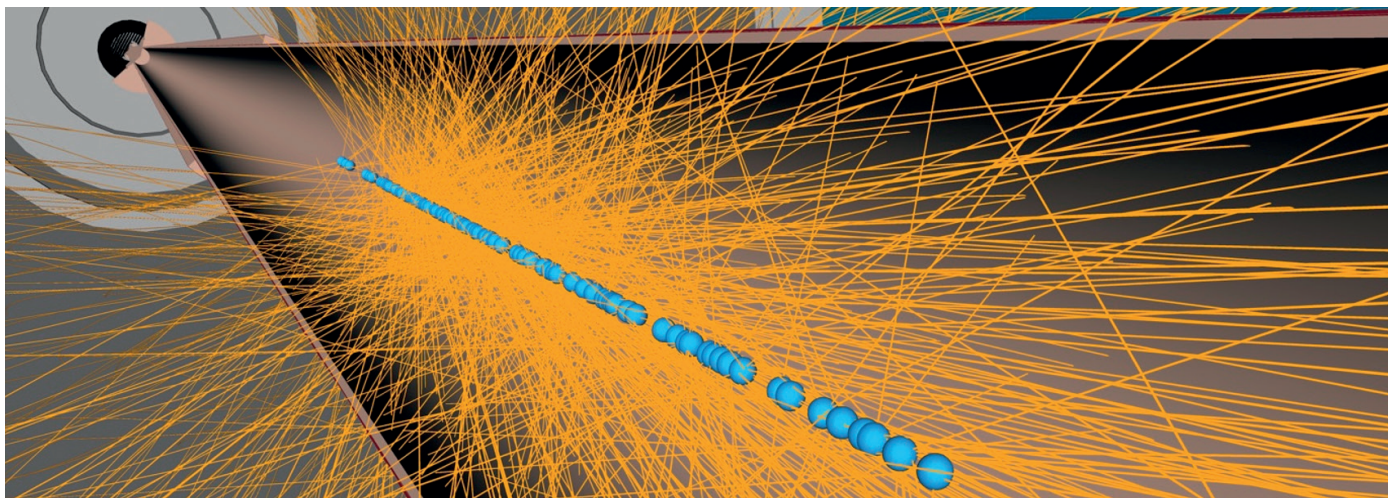
Izjemna priložnost za slovensko znanost

Hkrati izpostavlja, da bosta Univerza v Novi Gorici in država Slovenija tudi po zaključku projekta veliko pridobili. »Po eni strani je to zagotovo 50 odličnih podoktorskih raziskovalcev v najproduktivnejši dobi, s čimer se bodo močno okrepile aktivnosti na znanstvenih področjih, ki uporabljajo metode umetne inteligence. Seveda bi radi v projekt pritegnili tudi podjetja, ki že uporabljajo metode strojnega učenja in so pripravljena z nami sodelovati.«

Dr. **Danijel Skočaj**, predstojnik Laboratorija za umetne zaznavne sisteme na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, poudarja, da je strojno učenje v zadnjem desetletju doživelo pravo renesanso. »Strojno učenje danes uporabljamo na različnih področjih, od fizike do medicine in meteorologije, ne nazadnje tudi v industriji. Na vseh omenjenih področjih imamo opravka z veliko količino podatkov, človek pa iz te množice težko izlušči določene vzorce ali pa nas ti vzorci omejujejo. Zato ne gre drugače, kot da postavimo avtomatske algoritme, ki znajo delati s temi podatki. Tolikšna količina podatkov, kot jo sproduciramo v zadnjih letih, je namreč neprimerljiva s prejšnjimi obdobji oziroma s celotno zgodovino človeštva,« izpostavlja.

Pomoč najzmogljivejšega računalnika v Sloveniji

Projekt SMASH je usmerjen v raziskave, kjer se za obdelavo velike količine podatkov uporablja globoko in strojno učenje kot temelj za umetno inteligenco. Kot poudarja dr. **Dejan Valh**, vodja Sektorja za superračunalništvo Inštituta informacijskih znanosti Maribor, zaradi specifičnih zahtev običajni računalniški sistemi niso dovolj računske zmogljivi in je z njimi mogoče izvesti le krajše izračune z omejeno količino podatkov, z manj procesorskimi jedri in manj pomnilnika. Nimajo namreč dovolj kapacitet za shranjevanje vele podatkov, nad katerimi se izvaja modeliranje, kompleksne računske algoritme, analize in preučevanje podatkov z namenom pridobivanja hitrih in verodostojnih informacij, to pa omogoča nova odkritja. »Superračunalniki premostijo te pomanjkljivosti z večjim številom zelo zmogljivih računalnikov, imenovanih računski vozlišča, ki imajo veliko računskih jeder in ogromno pomnilnika, med seboj pa so tesno povezani v izredno hitro komunikacijsko omrežje z majhnimi





Vzpostavlanje kliničnega odločitvenega sistema, ki bi že zelo zgodaj pomagal napovedati, kakšno je dejansko stanje nedonošenčkov in katera oblika zdravljenja bi bila zanje najprimernejša.

zakasnitvami. Na tak način omogočajo tekmo s časom,« pojasnjuje. Sistemi za shranjevanje podatkov na superračunalnikih po Valhovi navedbah omogočajo velike hitrosti branja, zapisovanja in premikanja teh podatkov do procesorjev ali grafičnih pospeševalnikov, ki so še posebej primerni za strojno učenje. »HPC Vega, najzmogljivejši superračunalnik v Sloveniji, ima na voljo preko 1000 takih vozlišč, skupaj premore preko 120 tisoč procesorskih jeder in 240 izredno zmogljivih grafičnih pospeševalnikov. Omogoča shranjevanje 19.000 terabajtov podatkov, kar bi lahko enačili s kapaciteto diskov v enakem številu osebnih računalnikov. Postdoktorski raziskovalci v projektu SMASH bodo imeli možnost izkoristiti potencial superračunalnika HPC Vega, ki že drugo leto polno deluje,« izpostavlja.

Z njim je svoje raziskave izvajalo že več kot 500 znanstvenikov, raziskovalcev, razvijalcev, inženirjev, profesorjev in študentov tako iz Slovenije kot iz vse Evrope. Projekti, ki jih izvajajo, so iz enakih pa tudi iz drugih znanstvenih disciplin, kot so kemijske vede in razvoj materialov, biokemija, bioinformatika in strojništvo. »Kar nekaj projektov temelji na uporabi digitalnih dvojčkov, ki omogočajo posnemanje stvari iz narave v digitalnem okolju, mnogo pa je tudi projektov z učenjem in uporabo umetne inteligence, ki v razvitem svetu postaja vse bolj uporabna,« še dodaja Valh.

Od leve proti desni: dr. Danijel Skočaj, dr. Matjaž Ličer, dr. Gabrijele Zaharijaš, dr. Dejan Valh, dr. Vida Groznik, dr. Boštjan Golob, dr. Artur Stepanov, dr. Borut Paul Kerševan

Procesi, ki jih še nismo videli

Projekt SMASH bo povezan tudi z odprtimi vprašanji v astrofiziki, konkretnije z Velikim hadronskim trkalnikom v CERN-u. »Fizika delcev nenehno pritiska na meje našega znanja o najbolj temeljnih zakonih, ki urejajo naše vesolje. Tak napredek je v širši javnosti vzbudil veliko zanimanja, na primer odkritje Higgsovega bozona v letu 2012. Naslednji izziv za HEP je nadgradnja Velikega hadronskega trkalnika v Evropskem laboratoriju za fiziko delcev CERN v Ženevi na LHC z visoko svetilnostjo (HL-LHC¹), ki je načrtovana za leto 2029. Trkalnik bo močno povečal svojo svetilnost, število trkov delcev na časovno enoto, in eksperimenti bodo zabeležili do 10-krat več podatkov v sistemih za shranjevanje. To bo povečalo možnost, da bi lahko opazili redke procese, ki jih še nismo videli in lahko pomenijo nova odkritja ter tudi omogočijo meritve že znanih procesov z veliko natančnostjo. Vendar pa je prepoznavanje redkih pojavov v zapletenih dogodkih visokoenergijskih trkov postalo izredno težko in zahteva razvoj in uporabo najnaprednejših računskih metod z uporabo strojnega učenja ter najzmogljivejših računalniških sistemov, kot so danes superračunalniki, na primer slovenski superračunalnik Vega. Tako je cilj sodelovanja znanstvenikov na področju eksperimentalne fizike delcev v projektu SMASH pridobitev novih sodelavcev in znanj ter prodornih idej na področju uporabe metod strojnega učenja in superračunalnikov pri iskanju novih odkritij na eksperimentu ATLAS na Velikem hadronskem trkalniku,« pravi dr. **Borut Paul Kerševan** z Inštituta Jožef Stefan v Ljubljani.



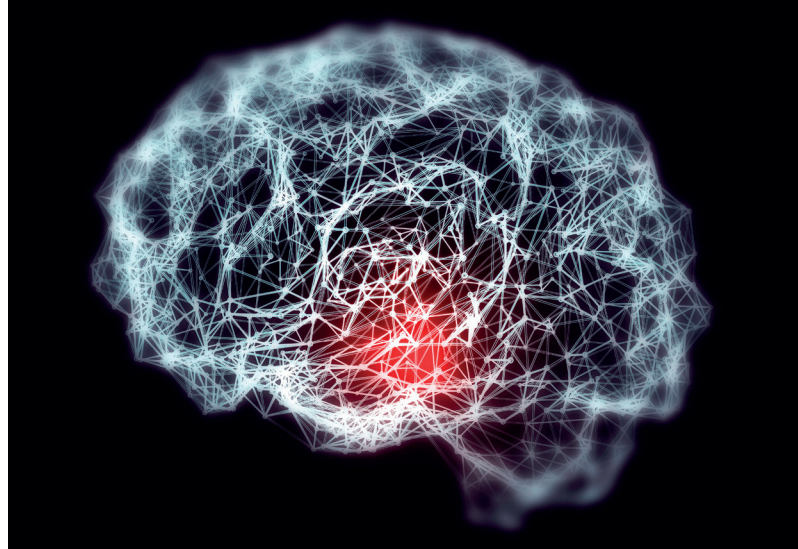
Napovedovanje ekstremnih dogodkov v atmosferi in oceanih

Na Agenciji RS za okolje se bodo v okviru projekta SMASH ukvarjali z napovedovanjem in razumevanjem ekstremnih dogodkov v atmosferi in oceanih. »Poskušali bomo razviti nove sisteme strojnega učenja, ki nam bodo pomagali napovedovati ekstremne pojave, kot so toče, suše, ekstremne dogodke v morju, morske in atmosferske vročinske valove ter hidrološke, hudourniške in obalne poplave,« pojasnjuje dr. **Matjaž Ličer**, ki se ukvarja s fizikalno oceanografijo na Agenciji RS za okolje in Morski biološki postaji Nacionalnega inštituta za biologijo. Kot ugotavlja, se nova komunikacija strojnega učenja z drugimi strokovnimi področji v zadnjem času hitro razvija, zato si na Agenciji RS za okolje želijo, da bi inkorporirali fizikalne zakone v pametne inteligentne sisteme ter da bi se slednji učili ne samo iz podatkov, temveč tudi iz fizikalnih zakonov. Obalne poplave v Jadranu so sicer doslej napovedovali z matematično-fizikalnimi simulacijami celotnega Jadranskega bazena.

»Ti izračuni so numerično intenzivni, obenem pa niso imuni za težave. Ena od osrednjih težav je dejstvo, da je polna gladina morja v Severnem Jadranu nelinearno odvisna od časovne razlike med vremenskim in plimnim vplivom, ki sta med seboj neodvisna. Majhne napake v napovedovanju hitrosti vetra nad morjem lahko zato vodijo v velike napake v napovedani gladini morja v Kopru,« poudarja dr. Ličer. Težave lahko zmanjšajo, če pri simulacijah stanja morja namesto enega možnega razvoja vremena nad morjem upoštevajo več deset razvojev vremena, ki se med seboj rahlo razlikujejo.

»Seveda to pomeni nekaj desetkrat dražji izračun kot v primeru ene same simulacije. Zato smo s skupino prof. dr. Mateja Kristana s Fakultete za računalništvo in informatiko Univerze v Ljubljani ter s kolegoma Lojzeto Žustom in Markom Rusom razvili HIDRO, novo metodo napovedovanja obalnih poplav z globokimi mrežami. Ta je po statističnih metrikah uspešnejša pri napovedovanju obalnih poplav od matematično-fizikalnih modelov oceana, obenem pa je v prognostičnem smislu milijonkrat hitrejša, saj ji ni treba reševati zapletenih diferencialnih enačb v ukrivljeni geometriji

Prepoznavanje redkih pojavov v zapletenih dogodkih visokoenergijskih trkov zahteva razvoj in uporabo najnaprednejših računalniških metod z uporabo strojnega učenja ter najzmogljivejših računalniških sistemov, kot so danes superračunalniki.



Na področju personalizirane medicine se trenutno osredotočajo predvsem na zgodnjo detekcijo in monitoring nevrogenativnih bolezni, kot sta Parkinsonova bolezen in demenca.

celotnega Jadranskega bazena, temveč napoveduje gladino le za Koper,« pojasnjuje.

Globoki model napovedovanja obalnih poplav v Agenciji RS za okolje vsakodnevno že uporabljajo za prognozo gladine slovenskega morja, a omenjeni raziskovalci bi radi v prihodnje ta model še izboljšali in ga prenesli tudi na druga svetovna morja. »En takšen poskus je že v teku,« priznava dr. Matjaž Ličer. Bi se torej lahko od slovenskih znanstvenikov v prihodnje učile tudi druge države? »Lahko si predstavljate, da so obalne poplave težava za vse obmorske države, še zlasti pa imajo te težave regionalna morja, ki so zaprta. Baltik in Sredozemski bazen sta že takšna primera. Seveda bi radi preverili, kako se sistemi, ki smo jih razvili za Jadransko morje, učinkoviti tudi za napovedovanje obalnih poplav znotraj teh bazenov,« odgovarja sogovornik. Zanimanje za takšne meritve po njegovih navedbah obstaja, zato v tem trenutku že sodelujejo s kolegi z Baltika.

Pomoč nedonošenčkom

V dobi strojnega učenja in umetne inteligence naj bi, kot kaže, napredovala tudi t. i. personalizirana medicina. Dr. **Vida Groznik**, članica Laboratorija za umetno inteligenco na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, pravi, da se na področju personalizirane medicine trenutno osredotočajo predvsem na zgodnjo detekcijo in monitoring nevrogenativnih bolezni, kot sta Parkinsonova bolezen in demenca. Svoje znanje uporabljajo tudi na drugih področjih, denimo v farmaciji in onkologiji. V onkologiji se tako ukvarjajo z glioblastomom, tumorjem na možganih, ki ima zelo slabo prognozo. »Ugotoviti poskušamo, kateri so tisti faktorji v tumorju, na katere moramo targetirati zdravljenje, da bi bila prognoza čim boljša in da bi bilo življenje ljudi, ki dobijo tak tumor, daljše in bolj kakovostno,« izpostavlja.

Še eno področje, ki je v personalizirani medicini izjemno zanimivo, pa je po besedah dr. Groznikove spremljanje kognitivnega upada in motoričnih spretnosti pri nedonošenčkih, torej pri otrocih, ki se rodijo pred 37. tednom nosečnosti. »Pri njih je povečana možnost, da bodo imeli težave pri nadaljnjem razvoju. Zato poskušamo novejšo tehnologijo povezati s standardnimi testi, ki jih izvajajo pediatri,





Ena od možnosti uporabe strojnega učenja je na področju odkrivanja vzorcev komunikacije pri živalih, ki imajo visoko stopnjo inteligenčnih in kognitivnih sposobnostih, na primer pri delfinih.

in na tak način gradimo klinični odločitveni sistem, ki bi že zelo zgodaj pomagal napovedati, kakšno je dejansko stanje teh otrok in katera oblika zdravljenja bi bila zanje najprimernejša. Tako bi tudi pediatri lahko zelo dobro vedeli, v katero smer zdravljenja naj gredo, da ne bi zamudili tistih najbolj ključnih momentov v razvoju nedonošenčkov,« še dodaja.

Kot ugotavlja dr. Vida Groznik, je apliciranje strojnega učenja na te domene izjemno zanimivo, zato se njihova raziskovalna skupina ukvarja tudi z nadgradnjo metodologij strojnega učenja in z razločljivo umetno inteligenco. »Seveda si želimo, da bi ljudje sprejeli stvari, ki jih razvijamo, in jih tudi uporabili. Uporabili pa jih bodo takrat, ko bodo tem sistemom popolnoma zaupali. To velja tako za splošno populacijo kot za zdravnike. Zdravnik mora vedeti, zakaj je računalnik rekel, da ima določena oseba težave in da ima mogoče demenco ali Parkinsonovo bolezen,« poudarja sogovornica, ki hkrati priznava, da je varovanje osebnih podatkov pri uporabi metod strojnega učenja izjemno pomembno.

Ob naravoslovju bo projekt SMASH odprl tudi vrata raziskavam v humanistiki. Primer je jezikoslovje. Kot pojasnjuje dr. **Artur Stepanov** iz Centra za kognitivne znanosti jezika Univerze v Novi Gorici, je ena od možnosti uporabe strojnega učenja na področju odkrivanja vzorcev komunikacije pri živalih, ki imajo visoko stopnjo inteligenčnih in kognitivnih sposobnostih, na primer pri delfinih.

»Delfini so izredno dobro organizirani in pri svojih vsakodnevnih opravilih delujejo usklajeno, dogovorijo se, na primer, glede lova ali varovanja mladičev. Pri raziskovanju njihove komunikacije doslej še nismo prepoznali jezika kot sistema strukturnih vzorcev, morda zato, ker smo pričakovali, da bomo našli glasove, besede, stavke. V nasprotju z jezikoslovjem strojno učenje ni omejeno s predstavami, kakšen bi jezik moral biti, zato lahko prepozna vzorce, ki niso človeški jezik, so mu pa po svoji funkcionalni strukturi sorodni. V projektu SMASH bomo posneli še veliko več oglašanj delfinov in nadaljevali raziskovanje delfinščine s sodobno tehnologijo. Upamo, da bo to pomembno prispevalo k razširitvi našega znanja o komunikaciji delfinov, o komunikaciji v živalskem svetu in tudi o naravi jezika,« napoveduje dr. Stepanov.

V ta namen so se na Univerzi v Novi Gorici že povezali s kolegi v Bolgariji, natančneje z Univerzo Varna, skupaj pa naj bi raziskovali akustične signale, ki jih oddajajo delfini, da bi bolje razumeli njihov način komuniciranja. Največji izziv, na katerega bi radi znanstveniki na ta način odgovorili, je, ali je sistem komuniciranja pri delfinih približno tako močan kot pri ljudeh, ko gre za sposobnost izražanja. ■

Strojno učenje danes uporabljamo na različnih področjih, od fizike do medicine in meteorologije, ne nazadnje tudi v industriji.
