

## 4. Félévi beszámoló

Ribli Dezső ([dkrib@caesar.elte.hu](mailto:dkrib@caesar.elte.hu))

Témavezető: Csabai István

A dolgozat címe: *Mesterséges neuronhálózatok modelljeinek elemzése illetve alkalmazása adatintenzív tudományos problémákban*

2018. 05. 30.

### **Bevezetés:**

Bonyolult komplex jelenségek esetén gyakran a kizárólag mérési adatokra támaszkodó, magas dimenziós statisztikai módszerek képesek a legjobban előrejelezni kísérleti, mérési eredményeket. Ezt a hozzáállást nevezik gépi tanulásnak, aminek egyik legfontosabb eszközei a mesterséges neuronhálók. A tudományos vizsgálatok során bonyolult rendszerekről gyűjtött mérési adatok robbanszerű növekedése miatt, a gépi tanulás további térnyerése várható az elkövetkező években évtizedekben.

A mesterséges neuronhálók alkalmazásában 2012 óta hatalmas előrelépések történtek. Képfelismerési feladatokban a korábbi módszerekhez képest tízedére csökkentették a hibát, a legjelentősebb teszten, és így már olyan eredményeket értek el, amik bőven összemérhetőek az emberi pontossággal, azaz egy ember is csak nagy erőfeszítés árán lenne képes hasonló pontosságot elérni, mint a neurális háló alapú módszerek.

Az áttörést az indította el, hogy megjelentek a párhuzamos számítások elvégzésére képes grafikus chippek és ezeknek az általános célú programozását lehetővé tevő szoftverek.

Az elmúlt 6 évben felismerve a módszerben lehetséges alkalmazási potenciált, a mesterséges neuronháló kutatások központjává az egyetemi műhelyek helyett leginkább az óriás technológiai cégek kutatóközpontjai váltak, pl.: Facebook, Google, Microsoft, Baidu, Nvidia. A doktori kutatásaim során a mesterséges neuronháló modellek alkalmazási lehetőségeit vizsgálom adatintenzív tudományos kérdésekben, továbbá a modelleket igyekszem olyan szempontból elemezni, ami a tudományos kérdések és megoldások további megértését segíti.

### **Az előző három félévben elért kutatási eredmények összegzése:**

Az első három félév során sikeresen vettem részt egy nagy nemzetközi tudományos adatelemzéssel foglalkozó versenyen a The Digital Mammography DREAM Challenge-en. 1200 regisztrált versenyző közül a kompetitív szakaszban 2. helyen végeztem, a verseny második szakaszában pedig az én módszerem továbbfejlesztett változata lett kimagaslóan a leghatékonyabb.

A munka során sikerrel ismertem fel, hogy a feladat más megközelítéssel oldható meg, mint amivel mások próbálkoznak. Várhatóan az elkövetkező években az én megközelítéssel azonos vagy hasonló elven működő megoldások jelennek meg a mindennapi orvosi gyakorlatban, elsősorban az Egyesült Államokban, de remélhetőleg hazánkban is. Az eredményeimet összefoglaltam egy a Scientific Reportsban idén megjelent cikkben.

További elősorban rákkutatással kapcsolatos projektben vettem részt kisebb-nagyobb számítógépes, tudományos adatfeldolgozási feladattal, melyekből számos cikk született, amiknek társszerzője vagyok.

## **Az aktuális félévben elvégzett kutatások ismertetése:**

### *Mammográfia:*

Ebben a félévben megjelent a mammográfiás kutatásaimat leíró cikk a Scientific Reports folyóiratban. Rengeteget dolgoztunk kórházi partnerek keresésével, hogy további adatok segítségével tovább tudjuk fejleszteni, vizsgálni a megoldásunkat.

Gyakran éri az a kritika a neurális hálókat, hogy fekete doboz szerűek és nem lehet érteni, hogyan működnek, ezért elkezdtem vizsgálni az eddig fejlesztett modell rejtett részleteit, hogy jobban megértsük a módszer működését. Az itt kapott érdekes eredményeket elkezdtem összefoglalni egy cikk formájában. Az együttműködő orvosok radiológiai szempontból értelmezhetően találják a talált motívumokat, részletesebb eredményekre várok, hogy befejezhessem a munkát.

### *Kozmológiai paraméterek becslése gyenge gravitációs lencsézés térképek alapján:*

A hagyományos kép-klasszifikációs alkalmazások mellett a deep learning a fizikai modellezéseket is segítheti. Az Univerzum főbb paramétereit - mint pl. a barionikus anyag aránya vagy a fluktuációkat jellemző  $\sigma_8$  -, meghatározhatjuk a különböző paraméterekkel indított N-test szimulációkból számolt és az észlelt gravitációs lencsés összevetésével. A közelmúltban a témakörben megjelent tanulmány deep learninggel kozmológiai paraméterek becslésére vállalkozik gyenge lencsézés térképek alapján. A korábbi munka deep learninggel és gépi tanúlással kapcsolatos része, sajnos hagyott kívánni valókat maga után, ezért elkezdtük a témával foglalkozni, Pataki Bálint mesterszakos hallgatóval. Munkánk eredménye során sikerült nagyban javítani a korábbi munka eredményein, és a A neuronháló "fekete dobozába nézve", a paraméterek elemzésével sikerült egy egyszerű asztrofizikailag is értelmezhető paraméter-mentes becslő módszert is létrehozni, ami a korábbiaknál pontosabb szimulációs adatokon, és a neuronhálókkal ellentétben akár mérési adatokon is potenciálisan pontosabb lehet.

Az eredményeket jelenleg írjuk le cikk formájában.

### *Egyéb:*

Tovább segítettem Rácz István doktoranduszunk a gamma felvillanások távolságbecslésével. Az előző félévek során elvégzett adatelemző munkáim eredményeképpen két cikk is megjelent a nagyon rangos Cell folyóiratban, amikben társszerző vagyok.

## **Publikációk:**

*Ribli, D., Horváth, A., Unger, Z., Pollner, P. and Csabai, I., 2018. Detecting and classifying lesions in mammograms with Deep Learning. Scientific reports, 8(1), p.4165.*

Mammográfiás kutatásaimat leíró cikk. Ebben a cikkben mind a kérdés feltevését, mind a programozási munka oroszlánrészét én végeztem és a cikket is én írtam meg.

*Pipek, O., Ribli, D., Molnár, J., Póti, Á., Krzystanek, M., Bodor, A., Tusnady, G.E., Szallasi, Z., Csabai, I. and Szüts, D., 2017. Fast and accurate mutation detection in whole genome sequences of multiple isogenic samples with IsoMut. BMC bioinformatics, 18(1), p.73.*

Ebben a cikkben egy új mutáció kereső módszert dolgoztunk ki, a partnerek által végzett speciális kísérlet egyedi tulajdonságait kiaknázva. A munkát még mesterszakos hallgató koromban kezdtem. A módszer kidolgozását Pipek Orsolyával együtt végeztük, a cikk megírása az ő munkája, a megjelent szoftvert én írtam.

*Szikriszt, B., Póti, Á., Pipek, O., Krzystanek, M., Kanu, N., Molnár, J., Ribli, D., Szeltner, Z., Tusnady, G.E., Csabai, I. and Szallasi, Z., 2016. A comprehensive survey of the mutagenic impact of common cancer cytotoxics. Genome biology, 17(1), p.99.*

Az előző cikk ennek a publikációnak a megjelenéséhez szükséges adatelemző munka kifejlesztését írja le. A munkát még mesterszakos hallgató koromban kezdtem. A cikkben megjelenő adatok elemzésének egy részét én végeztem.

*Zámborszky, J., Szikriszt, B., Gervai, J.Z., Pipek, O., Póti, Á., Krzystanek, M., Ribli, D., Szalai-Gindl, J.M., Csabai, I., Szallasi, Z. and Swanton, C., 2017. Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. Oncogene, 36(6), p.746.*

Az a cikk az előzőhöz hasonló munkafolyamat részeként szültett, csak nem gyógyszerek hanem gének kiesésének mutagenikus hatását írja le. A munkát még mesterszakos hallgató koromban kezdtem. A cikkben megjelenő adatok elemzésének egy részét én végeztem.

*Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J.I., Watkins, T.B., Nicol, D. and Stares, M., 2018. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell, 173(3), pp.595-610.*

Ehhez a cikkhez a mutációs spektrumok elemzését végeztük Pipek Orsolyával.

*David Y. Takeda, Sandor Spisák, Ji-Heui Seo, Connor Bell, Edward O'Connor, Keegan Korthauer, Dezso Ribli, Istvan Csabai, Norbert Solymosi, Zoltan Szállási, Paloma Cejas, Xintao Qiu, Henry Long, Viktória Tisza, Pier Vitale Nuzzo, Mersedeh Rohanizadegan, Mark M. Pomerantz, William C. Hahn, Matthew L. Freedman., 2018. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. Cell, Accepted.*

A cikkhez egy gén kópia szám változást mérő adatbázist elemeztem, ami a cikk kiindulási alapjához volt szükséges, a munkám a fig.1A, fig S1A ábrákon látható.

Előkészületben lévő publikációk:

- A félév során a mammográfiát elemző deep learning módszer részleteinek megértését tárgyaló munkán dolgoztam sokat. A cikk vázlatát megírtam, ha az orvos kollégák eredményei megérkeznek, idén nyáron szeretném befejezni és beküldeni a cikket.
- A félév második felében a deep learning alapú kozmogiai paraméter becsléssel foglalkoztam sokat, Pataki Bálint mesterszakos hallgató segítségével. A munk már előrehaladott állapotban van, idén nyáron szeretném befejezni és beküldeni a cikket.
- Zsuppán R., Ribli, D., Horváth, A., Unger, Z., Pollner, P., & Csabai, I, Bánsághi Z., Bérczi V.: A számítógépes mélytanulás és a mammográfia hazai lehetőségei, Orvosi Hetilap, beküldésre előkészítve (2018)
- A mammográfiai a versenyben az élen végzőkkel közösen írt cikk előrehaladott állapotban van, több érdekes további eredmény született, ebben a megállapodás szerint 3. szerzőként fogok szerepelni, remélhetőleg ebben az évben beküldésre kerül.
- Az előző beszámolómban említett transzkripció faktor kötőhely predikcióval foglalkozó cikk, amiben konzorciumi tagként leszünk megemlítve, még mindig készül.

### **Tanulmányi tevékenységek az aktuális félévben:**

A félévben két órára jártam: “Gráfok a bioinformatikában” és “Környezeti áramlások fizikája”.

### **Konferenciák a képzés alatt:**

A The Digital Mammography DREAM Challenge eredményeknek köszönhetően meghívott előadóként adtam elő a *RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges, New York, USA (2017)* nemzetközi konferencián.

Előadtam egy szemináriumon a Kútvölgyi Kórház Onkológiai Klinikáján, Deep learning a mammográfiában címmel.

### **Oktatási tevékenység**

Gyakorlat vezetőként részt vettem a *Fizika numerikus módszerei* című gyakorlat oktatásában. Előadásokat tartottam a, Csabai István professzor által az idén először meghirdetett, Deep learning és gépi tanulás a tudományokban című speciális előadás keretében. Továbbá segédkeztem a tantárgy tematikájának elkészítésében, a házi feladatok összeállításában és kiértékelésében. A tárgy nagy sikert aratott, több mint 50 diák vette fel, és 30-nál több diák végezte el a rendszeres feladatokat és végezte el a tárgyat.