

3. félévi beszámoló  
**Pataki Bálint Ármin** ([patbaa@gmail.com](mailto:patbaa@gmail.com))  
Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program  
Témavezető: Csabai István  
A dolgozat címe: Gépi tanulás a tudományokban  
2020.01.15.

### **Bevezetés:**

A doktori munkám kutatási témája nem változott, továbbra is gépi tanulási módszerek különböző tudományos területeken való alkalmazhatóságát kutatom.

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

Nyáron elindult egy izgalmas együttműködés a SOTE II. sz. Patológiai Intézete és a Komplex Rendszerek Fizikája Tanszék között. A kollaboráció különböző patológiai metszetek elemzése gépi tanuló algoritmusokkal, végső soron pedig orvosi diagnosztikai döntéstámogató rendszer (CAD, computed aided diagnosis) fejlesztése.

Vastagbélrák hazánkban, és a világon is a vezető rákos megbetegedések és halálesetek között található. A vastagbélrák szűrések egy előzetes teszt után szövettannal folytatódnak, melynek során a bélből leválasztott polipokat patológusok mikroszkóp alatt megvizsgálják, és az így kapott diagnózis alapján dől el, milyen kezelés szükséges. Ezeket a tárgylemezeket azonban patológiai szkennerek segítségével digitalizálhatjuk is. A projekt során konvolúciós neurális hálózatokat tanítunk szöveti képeken, hogy különböző szöveti elváltozásokat pontosan felismerjen. Nehézséget jelent, hogy egy-egy ilyen beszkenelt kép hatalmas, JPEG tömörítéssel is 5-10 GB lehet. A projekt során patológus kollégák jelölik be a képeken az elváltozásokat, amit igyekszünk megtanítani a modelleknek.

A projekt jól halad, a kezdeti eredmények biztatóak.

A félév során témavezetőmmel, Csabai Istvánnal részt vettünk egy Dream Challenges által szervezett gépi tanulási versenyen, ahol génexpressziós adatokból kellett minél pontosabban meghatározni, hogy egy adott nő a terhesség hanyadik hetében jár. A versenyen első helyezést értünk el. Mint helyezettet, meghívtak a projekt New York-ban rendezett konferenciájára, ahol prezentáltam a módszerünket. A verseny szervezői az eredményeket publikálni szeretnék tudományos folyóiratban, mely cikken a versenyen dobogósok társszerzőként lesznek feltüntetve. A publikáció várható időpontja ismeretlen.

<https://www.elte.hu/content/elte-s-aranyerem-new-yorkbol.t.19979>

### **Publikációk:**

A félévben két publikáció lett elfogadva, amelyeken társszerző vagyok:

- *Dezső Ribli, Bálint Ármin Pataki, José Manuel Zorrilla Matilla, Daniel Hsu, Zoltán Haiman, István Csabai, Weak lensing cosmology with convolutional neural networks on noisy data, Monthly Notices of the Royal Astronomical Society, Volume 490, Issue 2, December 2019, Pages 1843–1860, <https://doi.org/10.1093/mnras/stz2610>*

- *Amid C, Pakseresht N, Silvester N, Jayathilaka S, Lund O, Dynovski LD, Pataki B, Visontai D, Xavier BB, Alako BTF, Belka A, Cisneros JLB, Cotten M, Haringhuizen GB, Harrison PW, Höper D, Holt S,*

Hundahl C, Hussein A, Kaas RS, Liu X, Leinonen R, Malhotra-Kumar S, Nieuwenhuijse DF, Rahman N, Dos S Ribeiro C, Skiby JE, Schmitz D, Stéger J, Szalai-Gindl JM, Thomsen MCF, Cacciò SM, Csabai I, Kroneman A, Koopmans M, Aarestrup F, Cochrane G. *The COMPARE Data Hubs. Database (Oxford)*. 2019 Jan 1;2019:baz136. doi: 10.1093/database/baz136. PMID: 31868882; PMCID: PMC6927095.

Két kézirat pedig publikálva lett a BioRxiv-on, folyóiratba való publikálásuk pedig folyamatban van. Ezek közül az egyik a korábbi beszámolómban taglalt antibiotikum rezisztenciás kutatás, melyiken első szerző vagyok:

- Matamoros, S., Hendriksen, R., Pataki, B., Pakseresht, N., Rossello, M., Silvester, N., ... & Schultsz, C. (2019). *Accelerating surveillance and research of antimicrobial resistance-an online repository for sharing of antimicrobial susceptibility data associated with whole genome sequences*. *BioRxiv*, 532267.

- Pataki, B. Á., Matamoros, S., van der Putten, B. C., Remondini, D., Giampieri, E., Aytan-Aktug, D., ... & Schultsz, C. (2019). *Understanding and predicting ciprofloxacin minimum inhibitory concentration in Escherichia coli with machine learning*. *bioRxiv*, 806760.

Így összességében már három publikáció jelent meg a doktorim kezdete óta (és kettő pedig elfogadásra vár):

- Ribli, Dezső, Bálint Ármin Pataki, and István Csabai. "An improved cosmological parameter inference scheme motivated by deep learning." *Nature Astronomy* 3.1 (2019): 93.

- Dezső Ribli, Bálint Ármin Pataki, José Manuel Zorrilla Matilla, Daniel Hsu, Zoltán Haiman, István Csabai, *Weak lensing cosmology with convolutional neural networks on noisy data*, *Monthly Notices of the Royal Astronomical Society*, Volume 490, Issue 2, December 2019, Pages 1843–1860, <https://doi.org/10.1093/mnras/stz2610>

- Amid C, Pakseresht N, Silvester N, Jayathilaka S, Lund O, Dynovski LD, Pataki B Á, Visontai D, Xavier BB, Alako BTF, Belka A, Cisneros JLB, Cotten M, Haringhuizen GB, Harrison PW, Höper D, Holt S, Hundahl C, Hussein A, Kaas RS, Liu X, Leinonen R, Malhotra-Kumar S, Nieuwenhuijse DF, Rahman N, Dos S Ribeiro C, Skiby JE, Schmitz D, Stéger J, Szalai-Gindl JM, Thomsen MCF, Cacciò SM, Csabai I, Kroneman A, Koopmans M, Aarestrup F, Cochrane G. *The COMPARE Data Hubs. Database (Oxford)*. 2019 Jan 1;2019:baz136. doi: 10.1093/database/baz136. PMID: 31868882; PMCID: PMC6927095.

### **Tanulmányi tevékenység az aktuális félévben:**

A korábbi félévekben több tárgyat végeztem, így már csak két tárgyam van hátra, ezek közül az egyiket végeztem el a mostani félévben.

- Preklinikai modellek a daganatkutatásban FIZ/3/082

A tárgyra még nem kaptam meg az érdemjegyet jelen beszámoló írásakor.

### **Konferenciák az aktuális félévben:**

2019.11.04-én előadtam a RECOMB/ISCB Conference on Regulatory & Systems Genomics konferencián New York-ban a terhességi idő becsléséről. A konferencián való részvétel szervezők általi finanszírozása volt a verseny nyeresége.

2019.11.25-26-án részt vettem a Wigner által szervezett HEPTech AIME19 AI & ML konferencián, itt csak hallgatósággként vettem részt.

2019.12.09-én előadtam a SOTE-n szervezett 'szakmai találkozón' a fentebb említett patológiai projekt eredményeiről.

<http://semmelweis.hu/hirek/2019/12/10/strategiai-szakmai-talalkozo-a-mesterseges-intelligenciarol/>

**Oktatási tevékenység az aktuális félévben:**

A félév során két tantárgyat tartottam, mindkettőt Ribli Dezső PhD hallgatóval heti váltásban:

- Data mining and machine learning dsminingf17vm (EA + gyak) 4 kredit
- Számítógépes alapismeretek szamalapf19la (gyak) 3 kredit

A Data mining-ot and machine learning kurzust angolul tartottuk, és PhD-s hallgatók is jártak rá.