

## 2. félévi beszámoló

**Báskay János** (baskayj@student.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Pollner Péter

A dolgozat címe: Hálózat-statisztikák adattudományi alkalmazásai

### **Bevezetés**

A 21. század információs forradalma megteremtette a lehetőséget, hogy a természeti és társadalmi jelenségek tanulmányozása során korábban elképzelhetetlennek tűnő adatmennyiséget gyűjthessünk össze. Hamar világossá vált, hogy ezeket manuálisan feldolgozni, majd előrejelzéseket alkotni lehetetlen feladat, erre a problémára nyújtott megoldást a különböző gépi tanulási módszerek megjelenése. Ezek segítségével az adatok közötti korrelációk könnyedén összegyűrhatók egy robosztus modellé, mellyel klasszifikációs vagy regressziós problémák oldhatók meg.

Ezzel párhuzamosan gyors fejlődésen ment keresztül a Hálózattudomány is, melynek keretein belül rendelkezésre áll számos eszköz a komplex rendszerek megértésére és modellezésére.

PhD tanulmányaim során szeretnék a Hálózattudomány eszközeinek felhasználásával és gépi tanulási módszerek interdiszciplináris alkalmazásán keresztül átfogó ismereteket elsajátítani az Adattudomány területén.

### **Aktuális félévben elvégzett kutatások ismertetése**

#### ***Túlélés jóslás gépi tanulással (Gépi tanulás kis adatszetten)***

Ebben a félévben első sorban a korábbi szemeszterben folytatott munkában tártam fel a lehetséges hibákat, így a korábbi beszámolóban prezentált módszertan eltéréseit sorolom fel.

A validációhoz 10-Fold-os keresztvalidációra tértem át, ugyanis ez számít „industry standard”-nek, ezáltal a modell teljesítményét könnyebb összehasonlítani mások munkájával. Változtattam továbbá a feature szelekción is, tekintve, hogy a validációnál jelentősen csökkent a számításigény a bootstrap-módszerhez képest, így megfelelő volt a greedy feature selection-t használni.

A következő szemeszterben szeretném majd megvizsgálni, hogy javul-e a túlélési idő becslése, hogyha a túlélési idő helyett annak logaritmusát prediktálnám. (Ugyanis exponenciális-szerű eloszlást követ, melynek következtében viszonylag kevés a hosszú túlélésű minta, ezzel jelentősen növelve a predikciós modell hibáját.) Emellett fontos még a stabilitásvizsgálat is, hogy a jövőben beérkező újabb betegekre is használható legyen a létrehozott model/módszertan. Amennyiben ezek sikerrel járnak a kapott módszertan publikációra kész lesz.

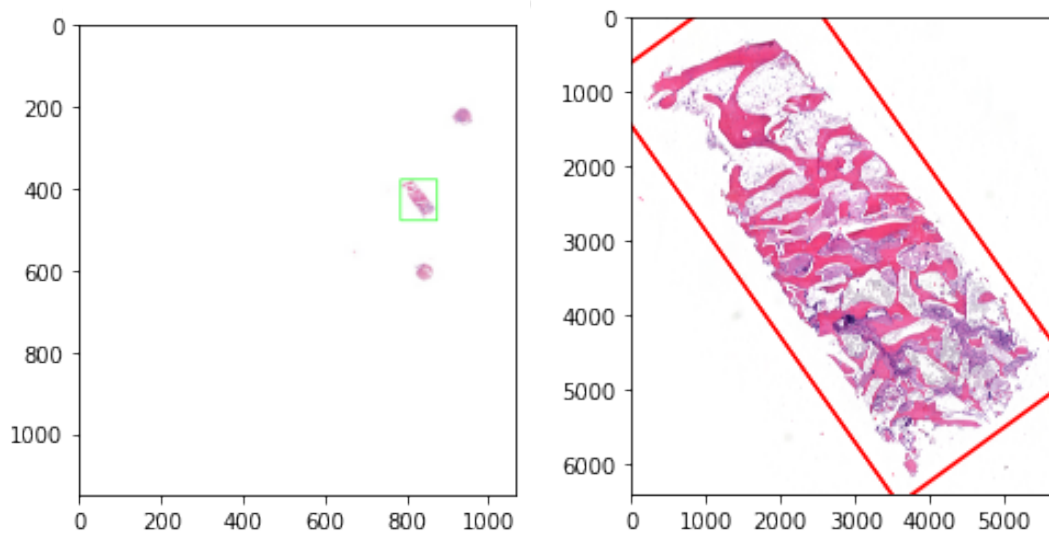
#### ***Képszegmentáció fogászok által készített patológias metszeteken***

A tavaszi szemeszter során belekezdtem egy második nagyobb terjedelmű projectbe. A kutatás első felének célja, hogy elkészítsek egy olyan mélytanulásra épülő módszertant, mellyel a Semmelweis Egyetem fogorvosai egyszerűen ki tudnak értékelni patológias metszeteket, és az ezekből kapott eredményeket összehasonlítva a non-invazív ám rosszabb felbontású 3DCT képekkel megbizonyosodjanak róla, hogy 3DCT alapján is sikeresek-e a végrehajtott beavatkozások.

A felhasznált metszetek fog transzplantációra váró betegek álkapcsából vették. Erre azért volt szükség, ugyanis hogyha egy fogat eltávolítanak, és a pótlásra hosszú ideig nem kerül sor akkor azon a környéken az álkapocs szerkezete megváltozik, és szivacsossá válik. Sajnos

ez a szivacsos szerkezet nem alkalmas implantátumok fogadására. (Nem képes a rágás során fellépő erővel megbirkózni.) Viszont, hogyha az érintett területre beültetésre kerül egy kisebb tömör csontdarab, akkor a környező csontszövet szerkezete is megváltozik, kevésbé lesz szivacsos, ezzel alkalmassá téve a fogimplantátum fogadására. A metszeteken ezért a csontszövetet kell szegmentálni, és annak különböző tulajdonságait (felszín, térfogat, fraktáldimenzió...) vizsgálni.

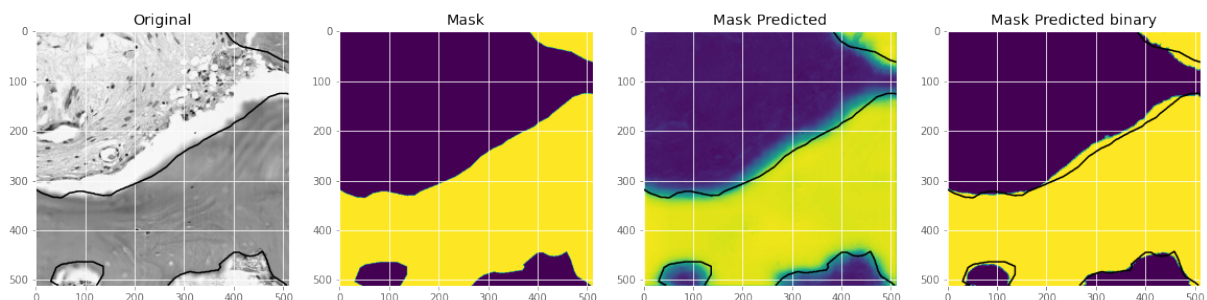
A módszertan első lépése az orvosi mikroszkóp által készített „multi-resolution” kép előkészítése a neurális háló számára. Ehhez első körben meg kell találni a felvételen azt a tartományt, ami a metszetet ábrázolja, majd ezt a részt kivágni és a lehetséges legjobb felbontásban lementeni.



Továbbá mivel a metszetek orientációja nem mindig egyezik meg, lehetőség van egy ferde téglalapot is illeszteni, melynek affin transzformációjával a metszetek majd egymásra forgathatók. (Erre a 3D rekonstrukcióhoz lesz szükség.)

A kapott metszeteket ezután kis egymást átfedő csempékre osztjuk szét, melyek felbontása már megegyezik a neurális háló bemeneti felbontásával, odafigyelve arra, hogy a tanításhoz és validációhoz használt képeknél a csempézés megegyezzen az eredeti képen és az orvosok által készített bináris maszkon is.

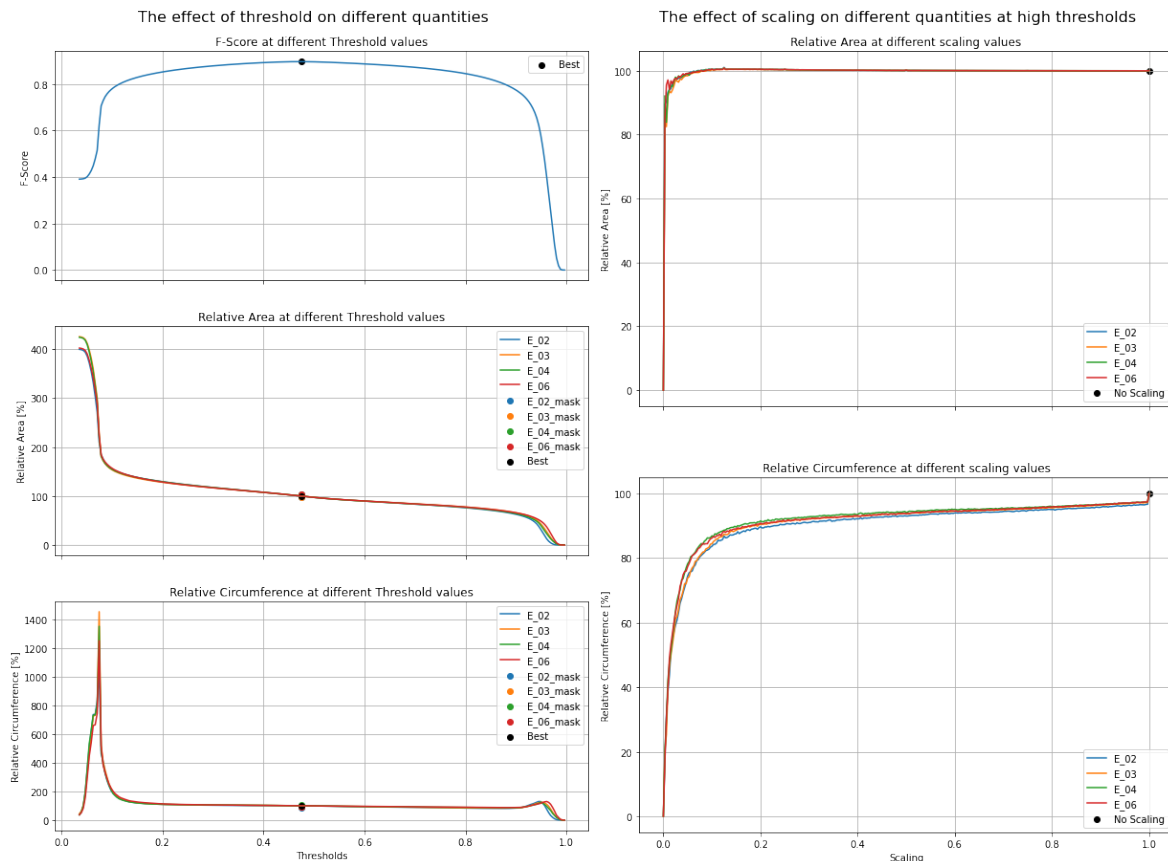
Ezek után egy batch-normalizációt tartalmazó UNET<sup>1</sup> neurális hálót tanítok be ahol a loss függvény binary cross entropy és az előre szegmentált képek 30%-t teszem félre validációra. Kb. 50 epoch után a validáció pontossága 94%. Az eredmények magukért beszélnek:



<sup>1</sup> Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". [arXiv:1505.04597](https://arxiv.org/abs/1505.04597)

A betanított hálóval ezek után ki lehet értékelni olyan mintákra is, melyekhez nem tartozik bináris maszk. A kapott csempékből rekonstruálható a teljes metszet képe, az átfedő részeknél pedig átlagolva kiküszöbölhetőek a neurális háló által létrehozott szél-közeli artifact-ok.

Mivel a neurális háló tulajdonképpen egy heatmap-et készít a beadott képekhez, a predikció pontossága növelhető, hogyha egy bilateral Filter segítségével zajszűrést hajtunk végre, majd a heatmap-en meghúzott threshold-ot optimalizáljuk. Az optimalizáció úgy tehető meg, hogy felvesszük a precision-recall görbét, majd azt a határértéket választjuk ki, melyen az F-score ( $2 * pr * rc / (pr + rc)$ ) a legnagyobb. Az így megválasztott threshold környékén a fizikai mennyiségek, mint a felület térfogat kb. állandó marad, még a felbontás változtatása esetén is.



## Tanulmányi tevékenység az aktuális félévben

- Adatexploráció és vizualizáció (FIZ/3/085)
- Kvantuminformáció-elmélet (FIZ/3/060E)

## Oktatási tevékenység az aktuális félévben

Az alapszagos Fizika hallgatók számára tartott *Modern fizika laboratóriumi gyakorlatok* című tárgy *Spektrofotometria* mérés keddi gyakorlatait tartottam meg 7 alkalommal és javítottam ki a hallgatók által beadott jegyzőkönyveket.