

2. félévi beszámoló

Borkovits Bendegúz (borbende@phys-gs.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Pollner Péter

A dolgozat címe: Képdiaosztika adattudományi alapú fejlesztése

2024.05.29.

1. Bevezetés

Folytattam az előző féléves beszámolóban bemutatott projektet. A Semmelweis Egyetemen vastagbélrákos szövetekből elkészített patológiai metszetekkel dolgoztam, azzal a különbséggel, hogy betekintést nyerhettem ezeknek az elkészítésébe és így én is készíthettem különböző metszeteket. Ezután a Budapesti Műszaki és Gazdaságtudományi Egyetemen (BME), az Alkalmazott Biotechnológia és Élelmiszertudományi Tanszék laborjában közép-infravörös tartományban spektrumokat vettem fel a metszetekről egy infravörös mikroszkóp segítségével.

2. Az aktuális félévben elvégzett kutatások ismertetése

Az előző félévben elbírálásra leadott kéziratot két javítási ciklus után megjelentették az adott folyóiratban. A részletek a Publikációk szekciójában tekinthetők meg. A kézirat témája a leírt spektrumok megkülönböztetése gépi tanulós modellekkel aszerint, hogy rákos vagy nem rákos szöveti területről lett felvéve a spektrum. Így bináris klasszifikáció volt a modellek feladata rákos (pozitív) és nem rákos (negatív) osztályok között. Meg kell említeni, hogy "rák" alatt első stádiumú (carcinoma) vastagbélrákot értünk és nem részei az adathalmaznak a leírt májmetasztázisos (harmadik stádiumú) szövetekről felvett spektrumok. A háttértől elválasztott és utána egyes tartományaikon megszürt spektrumokkal futtatott modellek validált eredményei az alsó TABLE 3 jelölésű, a cikkből átemelt táblázatban látható. Hét különböző modellt használtunk, hogy feltérképezzük, hogy melyik modell a legmegfelelőbb erre a feladatra. Összességében az infravörös spektrumokkal 70-80 %-os pontossággal megmondhatjuk, hogy az adott spektrum rákos vagy nem rákos szövetdarabról lett felvéve. A modelleket a velük kapott

eredmények alapján összehasonlítottuk a sum of ranking differences módszer segítségével és megkaptuk, hogy az egy-dimenziós U-Net mély neurális háló állt a legközelebb az ideális (0 % SRD) modellhez, ahogyan a FIGURE 5 feliratú ábrán látható.

TABLE 3 The efficiency metrics of the machine learning models when their input was the data that remained after performing K -means clustering and then we applied the blanking filters.

Methods	ACC	AUC	Sensitivity	Specificity
Random forest classifier	0.76 ± 0.09	0.84 ± 0.10	0.81 ± 0.10	0.72 ± 0.18
XGBoost Classifier	0.76 ± 0.11	0.84 ± 0.13	0.80 ± 0.11	0.72 ± 0.19
Linear Discriminant Analysis	0.75 ± 0.16	0.80 ± 0.23^a	0.74 ± 0.24	0.74 ± 0.18
Support Vector Classifier	0.72 ± 0.13	-	0.75 ± 0.16	0.67 ± 0.21
MLP with 3 Dense layers	0.81 ± 0.10	0.88 ± 0.10	0.81 ± 0.17	0.81 ± 0.19
2D convolutional neural network	0.83 ± 0.14	0.89 ± 0.13^a	0.79 ± 0.21	0.88 ± 0.14^a
1D U-Net neural network	0.81 ± 0.11	0.89 ± 0.10	0.79 ± 0.19	0.83 ± 0.14

Note: The errors were obtained by taking the standard deviation from running the models on 12 different train-test split sets, while the values themselves are averages obtained the same way.

Abbreviations: ACC, overall accuracy; AUC, area under ROC curve score; MLP, multilayer perceptron.

^a Of course, the efficiency metrics cannot achieve a score higher than 1; these occurred only because of our use of standard deviation as error bars.

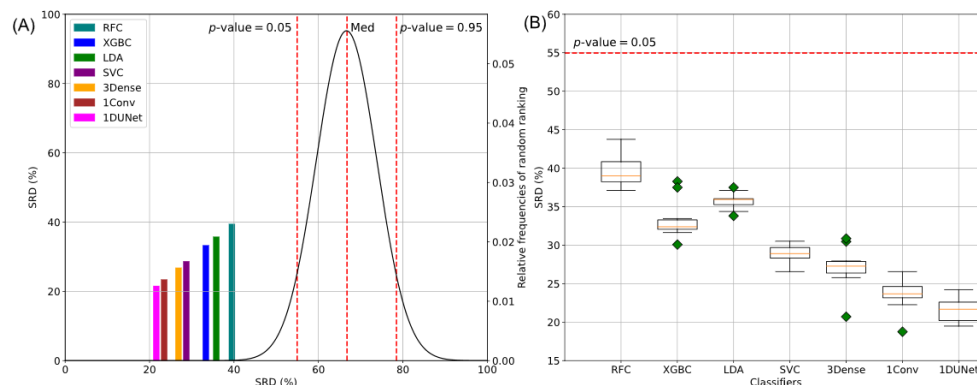


FIGURE 5 (A) Visualization of the sum of ranking differences-based ranking of the models. The black curve represents the SRDs of the random numbers generated during the comparison. Every model stands apart from the random distribution, meaning that each model behaved deterministically. (B) Ten-fold cross-validation was performed to determine the exact ranking of our models. The green markers denote outliers and the yellow lines showcase the median of each box.

A továbbiakban belekezdtem Kontsek Endre és Török Csongor segítségével egy újabb, az előzőhöz köthető projektbe, amelyben arra a kérdésre keressük a választ, hogy mi az optimális patológiai metszetvastagság közép-infravörös spektroszkópiai és tömegspektrometriai mérésekhez, analízishez. Közben bekapcsolódtam egy szintén a patológiához köthető projektbe, ami bizonyos sejtek whole slide image-eken (WSI) való gépi tanulásos detekciójából áll. Ebből a két projektből még nem született publikálásra méltó eredmény.

3. Publikációk

Március 13-án publikálták a Q3-as *Journal of Chemometrics* folyóiratnál *Classification of mid-infrared spectra from tissue microarray images of colorectal cancer* című kéziratunkat, melyben első szerző voltam.

4. Tanulmányi tevékenység az aktuális félévben

A következő kurzusokat végeztem el ebben a félévben:

- Fraktálnövekedés EA. (FIZ/3/004E)

5. Konferenciák az aktuális félévben

Nem vettem részt konferencián ebben a félévben.

6. Oktatási tevékenység az aktuális félévben

- Tudományos modellezés számítógépes laboratórium (dsscimodf20lm, labor, 7 kredit), projektvezetés, konzultáció, előadások és beszámolók értékelése és javítása (heti 4 óra)

7. Elismerések

Továbbra is részesülök a DKÖP-2023 Doktori Kiválósági Ösztöndíj Program által nyújtott támogatásban.