

**Single-cell image analysis and machine
learning methods using the cellular
microenvironment**

Tímea Tóth

Summary of the PhD Thesis

Supervisor: Péter Horváth, PhD

Doctoral School of Biology,
Faculty of Science and Informatics, University of Szeged
Department of Biochemistry, Biological Research
Centre, Szeged



Szeged

2023



Introduction

All living entities adapt to their environment. This phenomenon can be observed as morphological differences both at the macro- and microscales. Microscopy became a crucial tool for cell biology because it allowed researchers to go beyond the limits of the human eye and recognise structural differences between individual cells.

Machine learning earned a crucial role in cell biology and in medical image analysis. One of the reasons behind that is developments in microscopy and computational cell biology led to a massive increase in data volume, frequently as large as millions of images per study. Researchers nowadays prefer these high-content screenings, however analysing the produced datasets manually is either time-consuming or downright impossible. Another reason is that there is a chance that machines could see useful properties in images that are invisible for the human eye. It is in everyone's interest to get precise medical diagnosis and reliable treatment

suggestions and to achieve this, it is important to improve the accuracy of machine learning algorithms.

Aims

Our objective was to demonstrate how considering the cellular microenvironment during machine learning training impacts classification accuracy and determine the optimal distance from the cell of interest where the classifier performs with the highest efficiency. To accomplish this, we followed three main paths: 1) use information from the cell and its environment as input for classical machine learning algorithms, 2) use a fisheye-like transformation on original biological images and use these datasets as inputs for deep learning algorithms and 3) investigate the crucial steps required in 3D, such as 3D image acquisition, 3D cell annotation, 3D cell segmentation, and use a 3D co-culture dataset in which we distinguished phenotypes.

The main findings of this thesis are as follows.

- 1) Taking into account the environmental features of cells can enhance the accuracy of machine learning-based phenotypic classification.
- 2) Combining the cellular microenvironment with fisheye transformation can improve deep learning-based single-cell phenotyping.
- 3) Incorporating neighbourhood features can increase classification accuracy in 3D cell cultures.

Results

Firstly, we analysed images of MCF-7 cell culture and urinary bladder cancer tissue sections, with classical supervised machine learning methods. We defined two approaches to represent the microenvironment: the K-nearest neighbours (KNN) and the N-distance methods. We extracted regular and neighbourhood features, and we have seen that while the local features of cells can be similar across different phenotypic classes, the neighbourhood features are able to identify them. The results of our analysis demonstrated that the incorporation

of neighbourhood features significantly enhances the accuracy of machine learning-based classification, regardless of whether the data originates from cell cultures or tissue sections. The optimal size of the neighbourhood varied according to the specific dataset and classifier used. Notably, our findings emphasised the substantial impact of neighbourhood features on tissue section phenotyping, with performance improvement close to 20%.

Secondly, we introduced a novel approach of representing images to deep-learning-based image classification networks. The fundamental concept revolves around the following notion: the original image includes both the object of interest, as well as its surrounding microenvironment within a pre-defined range. Subsequently, a fisheye-like spatial sampling technique is applied to transform the images. The fisheye sampling collects more pixels from the close proximity of the object of interest, and the resolution decreases for larger proximity. Our findings demonstrated that the proposed transformation highly outperforms conventional machine-learning techniques and deep-learning-based classifiers.

This performance was evaluated across various datasets including cell cultures, scans of cancerous tissues, and real-life images.

In the third part of the thesis, we discussed the enhancement of microscopy image quality through optical clearing and assessed the efficacy of various clearing protocols. Then, we explored the difficulties and remedies associated with annotating and segmenting cells in three dimension. Lastly, we shared our achieved progress in investigating cells and their microenvironment within 3D co-cultures consisted of carcinoma and fibroblast cells. Similar to the first part, we used KNN and N-distance measurements to determine which cells we consider neighbours, and then performed the classifications with supervised machine learning algorithms. Our results showed, that including the neighbourhood features in the machine learning process increased phenotypic classification accuracy. Furthermore, we could distinguish the two cell types with 87.5% accuracy, solely by considering the shape attributes of the cells and their neighbours.

Overall, the research demonstrates that integrating the cellular microenvironment into machine learning-driven analysis enhances single-cell phenotypic classification. This highlights the importance of considering macrostructures within cellular systems, ultimately advancing the understanding and potential applications of image analysis in various scientific contexts.

Összefoglalás

Ez a dolgozat a mikroszkopikus képelemzéssel kapcsolatos kutatásainkat mutatja be. Célunk az volt, hogy egysejtes szinten, különösen a sejtbiológia és a patológia összefüggésében értékes felismeréseket nyerjünk összetett képadathalmazokból. Megmutattuk, hogy a sejtek mikrokörnyezetének figyelembevétele milyen hatással van az egyes sejtek fenotípus jellemzésére, felügyelt gépi tanulás és mélytanulási technikák segítségével.

Először MCF-7 sejttenyészetről és húgyhólyagrákos szöveti metszetekről készült képeket elemeztünk klasszikus felügyelt gépi tanulási módszerekkel. Két

megközelítést használtunk a mikrokörnyezet meghatározására: a K-közelebbi szomszéd (KNN) és az N-távolság módszerét. Lokális és szomszédsági jellemzőket vontunk ki és azt láttuk, hogy míg a sejtek helyi jellemzői hasonlóak lehetnek a különböző fenotípus osztályokban, a szomszédsági jellemzők képesek megkülönböztetni őket. Eredményeink azt mutatták, hogy a szomszédsági jellemzők bevonása jelentősen növeli a gépi tanuláson alapuló osztályozás pontosságát, függetlenül attól, hogy az adatok sejt kultúrákból vagy szöveti metszetekből származnak. A szomszédság optimális mérete az adott adathalmaztól és az alkalmazott osztályozótól függően változott. Eredményeink kiemelik a szomszédsági jellemzők jelentős hatását a szöveti metszetek fenotipizálása esetében, ahol a teljesítményjavulás megközelítette a 20%-ot.

Ezután a képek reprezentálásának újszerű megközelítését mutattuk be, amelyek a mélytanuláson alapuló képosztályozó hálózatok bemeneteként funkcionálhatnak. Az elvárásaink a következők voltak a bemeneti képekkel kapcsolatban: az eredeti kép tartalmazza mind a

számunkra érdekes objektumot, mind pedig az azt körülvevő mikrokörnyezetet egy előre meghatározott tartományon belül. Ezt követően egy halszemoptika-szerű térbeli mintavételi technikát alkalmazunk a képek átalakítására. A halszemoptika-szerű mintavételezés több képpontot gyűjt össze a vizsgált objektum közvetlen közeléből, mint a távolabb elhelyezkedő képterületekről. Eredményeink azt mutatták, hogy a javasolt transzformáció nagymértékben felülmúlja a hagyományos gépi tanulási technikákat és a mélytanuláson alapuló osztályozókat. Ezt a teljesítményt különböző adathalmazokon, köztük sejttenyészeteken, rákos szöveti metszeteken és valós életből származó képeken értékeltük.

Harmadik lépésben 3D sejt kultúrákkal foglalkoztunk. Vizsgáltuk a mikroszkópos képminőség optikai tisztítással történő javítását és különböző tisztítási protokollok hatékonyságának mérését. Ezután a sejtek háromdimenziós annotálásával és szegmentálásával kapcsolatos nehézségeket és ezek megoldásait tárgyaltuk. Végül megosztottuk a sejtek és mikrokörnyezetük vizsgálatában elért eredményeinket a karcinóma- és

fibroblaszt sejtekből álló 3D ko-kultúrák esetében. Az első részhez hasonlóan KNN és N-távolság mérésekkel határoztuk meg, hogy mely sejteket tekintjük szomszédnak, majd felügyelt gépi tanulási algoritmusokkal végeztük el az osztályozásokat. Eredményeink azt mutatták, hogy a szomszédsági jellemzők bevonása a gépi tanulási folyamatba növelte a fenotípus osztályozási pontosságot. 87,5%-os pontossággal tudtuk megkülönböztetni a két sejtípust a szferoidokon belül, kizárólag a sejtek és szomszédaik alakj jellemzőinek figyelembevételével.

Összességében a kutatás azt mutatja, hogy a sejtek mikrokörnyezetének integrálása a gépi tanulás által vezérelt elemzésbe javítja az egysejtes fenotípus osztályozást. Ez rávilágít a mikrostruktúrák sejtes rendszereken belüli figyelembevételének fontosságára, ami végső soron a képelemzés megértését és lehetséges alkalmazásait segíti elő különböző tudományos kontextusokban.

Publications

MTMT ID: 10064460

The graduation procedure is based on the following publications:

1. **Toth, T.**, Balassa, T., Bara, N., Kovacs, F., Kriston, A., Molnar, C., Haracska, L., Sukosd, F., Horvath, P. (2018) Environmental properties of cells improve machine learning-based phenotype recognition accuracy. Scientific Reports 8, 10085

IF: 4.011

2. **Toth, T.**, Bauer, D., Sukosd, F., Horvath, P. (2022) Fisheye transformation enhances deep-learning-based single-cell phenotyping by including cellular microenvironment. Cell Reports Methods, 2(12), 100339.

IF: 3.8

Other publications:

3. Diosdi, A., Hirling, D., Kovacs, M., **Toth, T.**, Harmati, M., Koos, K., Buzas, K., Piccinini, F., Horvath, P. (2021) A quantitative metric for the comparative evaluation of optical clearing protocols for 3D multicellular spheroids. Computational and Structural Biotechnology Journal 19, 1233-1243

IF: 6.155

4. Diosdi, A., Hirling, D., Kovacs, M., **Toth, T.**, Harmati, M., Koos, K., Buzas, K., Piccinini, F., Horvath, P. (2021) Cell lines and clearing approaches: a single-cell level 3D light-sheet fluorescence microscopy dataset of multicellular spheroids. Data in Brief 36, 107090

IF: not applicable

5. Tasnadi, E., **Toth, T.**, Kovacs, M., Diosdi, A., Pampaloni, F., Molnar, J., Piccinini, F., Horvath, P. (2020) 3D-Cell-Annotator: an open-source active surface tool for single-cell segmentation in 3D microscopy images. Bioinformatics, Volume 36, Issue 9, May 2020, Pages 2948–2949

IF: 6.937

6. Piccinini, F., Balassa, T., Carbonaro, A., Diosdi, A., **Toth, T.**, Moshkov, N., Tasnadi, E., Horvath, P. (2020) Software tools for 3D nuclei segmentation and quantitative analysis in multicellular aggregates. Computational and Structural Biotechnology Journal 18, 1287-1300

IF: 7.271

7. Hollandi, R., Szkalicity, A., **Toth, T.**, Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., Kovacs, M., Migh, E., Goodman, A., Balassa, T., Koos, K., Wang, W., Caicedo, J.C., Bara, N., Kovacs, F., Paavolainen, L., Danka, T., Kriston, A., Carpenter, A.E., Smith, K., Horvath, P. (2020) nucleAIzer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. Cell Systems 10 (5), 453-458. e6

IF: 10.304

Cumulative IF: 38.478

Nyilatkozat

Alulírott **Dr. Horváth Péter** a jelölt (Tóth Tímea) témavezetőjeként kijelentem, hogy a „Single-cell image analysis and machine learning methods using the cellular microenvironment” című disszertáció a jelölt saját munkája, melyet a témavezetésem mellett önállóan készített el. Kijelentem, hogy a disszertáció megfelel az SZTE TTIK Biológia Doktori Iskola követelményeinek.

Felelős szerzőként kijelentem, hogy Tóth Tímea jelentősen hozzájárult az alább felsorolt, a doktori eljárás alapjául szolgáló tudományos publikációk eredményéhez

- **Toth, T.**, Balassa, T., Bara, N., Kovacs, F., Kriston, A., Molnar, C., Haracska, L., Sukosd, F., Horvath, P. (2018) Environmental properties of cells improve machine learning-based phenotype recognition accuracy. *Scientific Reports* 8, 10085

- **Toth, T.,** Bauer, D., Sukosd, F., Horvath, P. (2022) Fisheye transformation enhances deep-learning-based single-cell phenotyping by including cellular microenvironment. Cell Reports Methods, 2(12), 100339.

Igazolom továbbá, hogy az ebben a dolgozatban bemutatott eredményeket más PhD dolgozat nem mutatta be és a jövőben sem használják fel tudományos fokozat megszerzéséhez.

Szeged, 2023. szeptember 12.

.....

Dr. Horváth Péter

Szegedi Biológiai Kutatóközpont