

Oponentní posudek disertační práce

Ústav: Středoevropský technologický institut VUT
Student (ka): **Ing. Lenka Štrbková**

Akademický rok: **2017/2018**

Doktorský studijní program: **Pokročilé materiály a nanovědy**

Studijní odbor: **Pokročilé nanotechnologie a mikrotechnologie**

Vedoucí disertační práce: **Prof. RNDr. Radim Chmelík, Ph.D.**

Oponent disertační práce: **Prof. RNDr. Michal Kozubek, Ph.D.**

Název pojednání práce: **Biofyzikální interpretace kvantitativního fázového zobrazení**

Aktuálnost tématu disertační práce:

Studentka se v disertační práci zabývá tématem automatizované klasifikace buněk v obrazech z optické mikroskopie a srovnává přístup bez využití a s využitím kvantitativní informace získané technikou CCHM (koherencí řízená holografická mikroskopie). Zpracovává obrazovou informaci jak ze statických dat, tak z časosběrných kvantitativních fázových obrazů. Vzhledem ke stále rostoucímu objemu obrazových dat buněk, je jistě automatizace zpracování těchto dat včetně klasifikace potřebná a užitečná. Navíc zužitkování kvantitativní informace je doposud málo probádané. Považuji proto téma disertační práce za aktuální.

Splnění stanovených cílů:

Hlavním cílem práce bylo prozkoumat a vyhodnotit různé možnosti automatizované klasifikace buněk, a to bez využití i s využitím kvantitativní informace, statické anebo i dynamické. Jako výsledek měly být popsány vhodné postupy. Jednotlivé dílčí cíle jsou popsány v kapitole 3. Bylo zejména zapotřebí tedy nejprve buňky segmentovat (v dynamickém případě i napojit výsledky segmentací jednotlivých snímků), spočítat z jednotlivých buněk různé charakteristiky (příznaky), vybrat z nich vhodnou podmnožinu a nakonec vyzkoušet různé postupy klasifikace (v této práci na bázi učení s učitelem, tedy na bázi příkladů) a srovnat jejich úspěšnost pomocí standardních metrik (accuracy, precision, recall, F-score). Lze konstatovat, že všechny tyto kroky byly provedeny, jsou v práci popsány a je učiněn závěr ohledně přínosu kvantitativní informace v klasifikačním úkolu. Stanovené cíle práce byly proto splněny.



Postup řešení problému a výsledky disertace:

Práce začíná úvodem, kde je jasně vysvětlena problematika. K této části mám jen drobný komentář, že obecně neplatí, že všechna fluorescenční barviva ovlivňují chování a morfologii buňky, například fluorescenční proteiny (GFP apod.) se naopak používají na pozorování buněk v přirozeném stavu. Následují cíle práce a shrnutí, co je ve které kapitole.

Dále je začleněna teoretická část, kterou považuji za nejslabší místo práce – je zde popsána technika CCHM v kapitole 5 (cca 3 strany) a úvod do strojového učení v sekci 6.1 (1 strana). Jelikož celá disertace je o aplikaci strojového učení, čekal bych podstatně větší než jednostránkový přehled v dané oblasti.

Další sekce 6.2 je již věnována praktické části, a to metodologii klasifikace pro potřeby daného druhu obrazových dat. Navazují kapitoly 7 a 8 zaměřené na využití metodologie ze sekce 6.2 u statických, respektive dynamických buněk. Provedená měření úspěšnosti klasifikace pro různé případy jsou dokumentována v tabulkách a grafech. Celkově je metodologie v hrubých rysech jasná, avšak některé detaily jasné nejsou a mnohé postupy nejsou přesně matematicky popsány (viz otázky níže).

Práce končí shrnutím dosažených poznatků a potvrzením, že kvantitativní informace v mikroskopii přináší lepší možnosti pro klasifikaci buněk. Nechybí ani náměty na další práci.

Význam pro praxi nebo rozvoj vědního oboru:

Jelikož se podařilo prokázat užitečnost kvantitativní mikroskopie pro klasifikaci buněk, lze konstatovat, že práce přispěla k rozvoji daného vědního oboru a vytváří základ pro další bádání směřující k eventuálnímu budoucímu uplatnění v praxi.

Formální úprava disertační práce a její jazyková úroveň:

Práce má 75 stran a je psaná anglicky, jazyková úroveň je velmi dobrá, formulace jsou jasné a výklad je srozumitelný. Text je doplněn řadou obrázků, tabulek a grafů, které vhodně doplňují výklad. Typografická úroveň je standardní, je škoda, že kromě čísel stránek nejsou použity žádné informace v záhlaví a zápatí. Členění práce je logické, ale názvy některých částí jsou poněkud matoucí, na první pohled to vypadá, že sekce 6.2, 7 i 8 se zabývají tím stejným, přitom sekce 6.2 se nezabývá klasifikací samotnou (jak vyplývá z názvu), ale pouze metodologií klasifikace. U sekce 7 není z názvu jasné, že se zabývá pouze statickými obrazy, navíc v kapitole 4 (přehled, co je ve které části) je mylně uvedeno, že se sekce 7 zabývá živými buňkami.

Zda dizertační práce splňuje podmínky uvedené v § 47 odst. 4 zákona:

(4) Studium se řádně ukončuje státní doktorskou zkouškou a obhajobou disertační práce, kterými se prokazuje schopnost a připravenost k samostatné činnosti v oblasti výzkumu nebo vývoje nebo k samostatné teoretické a tvůrčí umělecké činnosti. Disertační práce musí obsahovat původní a uveřejněné výsledky nebo výsledky přijaté k uveřejnění.¹⁾

Disertační práce splňuje uvedené podmínky.

¹⁾ § 10 zákona č. 35/1965 Sb., o dílech literárních, vědeckých a uměleckých (autorský zákon).



Připomínky a dotazy:

- 1) Jak jsem uvedl výše, teoretická část ke strojovému učení má jednu stranu, proč nebyla začleněna podrobnější rešerše?
- 2) Jak správně uvádíte na začátku sekce 6.2, pro strojové učení s učitelem je zapotřebí dostatečného množství obrazových příkladů buněk. Co je to dostatečné množství? Zkoumala jste závislost úspěšnosti klasifikace na počtu příkladů pro trénování? V kapitole 7 máte cca 400-500 příkladů na třídu, zatímco v kapitole 8 pouze cca 50 buněk na třídu, tedy řádově méně. Jak je to tedy s potřebným počtem?
- 3) U segmentace uvádíte, že používáte algoritmus záplavy (watershed) se značkami, chybí ale popis, jak se vypočítají značky. Můžete na obhajobě vysvětlit?
- 4) V několika místech práce (např. str. 24 včetně obrázku 6, ale i na straně 43 včetně následného obrázku 17) uvádíte, že data rozdělíte na trénovací a testovací a že trénovací jsou pak anotovány expertem biologem. Proč pouze trénovací? Pakliže nemáte anotovaná testovací data, jak provádíte validaci (která mimochodem na obrázcích 6 a 17 zcela chybí)? A na co testovací data pak používáte? Podle obrázků 6 a 17 na ně pustíte klasifikaci, ale jak vyhodnotíte správnost výsledku? Na straně 34 pak píšete o k-násobné křížové validaci, kde testovací množina se postupně mění a nemůže být bez anotace. Jak to tedy je? A kolik buněk bylo tedy anotováno pro každou třídu u statického a kolik u dynamického případu?
- 5) V sekci 6.2.2 se u 4 morfometrických příznaků používá suma pixelů („sum of the pixels“), co to je? Lze sčítat intenzity pixelů, velikosti gradientů pixelů, ale ne pixely samotné. Nanejvýš je možné pixely spočítat a určit počet, ale ne sčítat.
- 6) Proč byly zvažovány pouze morfometrické a intenzitní příznaky (zde nad velikostí fáze)? Proč nebyly vůbec zvažovány texturní příznaky (např. Haralickovy), které by se u různých druhů buněk mohly lišit?
- 7) Proč byla zvolena zrovna 5-násobná křížová validace? Jak bylo číslo 5 určeno?
- 8) U tabulky 1 na straně 38 tvrdíte, že p -hodnota byla u t -testu řádově nižší u QPI příznaků než morfometrických. V tabulce je ale téměř všude stejná p -hodnota u obou druhů příznaků, a to $p < 0.001$. Můžete toto vysvětlit?
- 9) Proč se u živých buněk snímalo 48h, ale použilo se pouze 2,5h? Jak se těch 2,5h z 48h vybralo?
- 10) V sekci 8.5 není jasný postup pro sledování (tracking) buněk. Přiřazování nejbližší buňky v dalším obrázku je velmi nespolehlivá metoda. Navíc se může stát, že stejná buňka v novém obrázku bude přiřazena více buňkám ve starém (protože bude k více buňkám nejbližší), jak tento případ řešíte? Navíc zde zcela chybí ošetření dělících se buněk, co se tedy stane při dělení?
- 11) V několika místech práce se píše, že příznaky jsou normovány do intervalu 0 až 1, což je logické. Nicméně v grafu na obrázku 21 nabývají příznaky hodnot cca -2.2 až +2.2, jak je to možné?
- 12) U Fourierovy transformace (str. 49-50) je velmi zavádějící a nepřesný popis. Je uveden vzoreček spíše pro Fourierovu řadu (bez komplexních čísel) a nikoliv Fourierovu transformaci a je uvedeno, že výsledné koeficienty aproximují původní signál, zatímco u skutečné Fourierovy transformace jde o

přesné vyjádření, nikoliv aproximaci. Sumace probíhá přes množinu B , která má být podmnožinou frekvencí v jakési bázi, jaké? Můžete vysvětlit, co tedy vlastně počítáte?

- 13) Stejně tak u vlnkové transformace (str. 50-51) se mluví pouze o aproximaci, nikoliv přesném rozkladu a ve vzorečku se pracuje pouze s vlnkovou funkcí, ale chybí měřítková funkce. Dokonce se mluví o jakési množině transformací B . Mohla byste upřesnit, co a jak se tedy vlastně počítá? A pakliže se jako příznaky berou největší koeficienty, tak kolik se jich bere? A větších než jaký práh?
- 14) U dalšího příznaku, trendu, se údajně počítají koeficienty přímky proložené přes data. Přímku lze ale definovat pomocí různých dvojic koeficientů, jaké se používají zde? Oba dva?
- 15) Další příznak entropie je zcela nedefinován, můžete na obhajobě uvést definici?

Celkové zhodnocení disertační práce:

Celkově lze konstatovat, že cíle práce byly splněny a praktická část práce tvoří dobrý základ pro další rozvoj v dané oblasti. Slabinami práce jsou teoretická část a nepřesné nebo chybějící matematické definice u některých postupů.

I přes uvedené nedostatky se domnívám, že studentka prokázala tvůrčí schopnosti v dané oblasti výzkumu a práce splňuje požadavky standardně kladené na disertační práce v daném oboru.

Disertační práci Ing. Lenky Štrbkové doporučuji k obhajobě pro udělení akademického titulu "doktor" (Ph.D.).

V Brně dne 26. 1. 2018


.....
Prof. RNDr. Michal Kozubek, Ph.D.