VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2018 Markéta Navrátilová

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
DEPARTMENT OF BIOMEDICAL ENGINEERING

PŘÍZNAKY PRO ANALÝZU A KLASIFIKACI BUNĚK VE SNÍMCÍCH Z HOLOGRAFICKÉHO MIKROSKOPU
FEATURES FOR THE ANALYSIS AND CLASSIFICATION OF CELLS IN HOLOGRAPHIC MICROSCOPE IMAGES

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — Markéta Navrátilová
AUTHOR

VEDOUCÍ PRÁCE — Ing. Tomáš Vičar
SUPERVISOR

BRNO 2018

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Bakalářská práce

bakalářský studijní obor **Biomedicínská technika a bioinformatika**
Ústav biomedicínského inženýrství

Studentka: Markéta Navrátilová ***ID:*** 186677
Ročník: 3 ***Akademický rok:*** 2017/18

NÁZEV TÉMATU:

Příznaky pro analýzu a klasifikaci buněk ve snímcích z holografického mikroskopu

POKYNY PRO VYPRACOVÁNÍ:

1) Prostudujte princip snímání živých buněk na holografickém mikroskopu a prostřednictvím odborné literatury se seznamte s postupem automatické analýzy mikroskopických snímků buněk. 2) Proveďte literární rešerši příznaků používaných při analýze buněk a pro jejich klasifikaci. 3) Ve zvoleném programovacím prostředí realizujte extrakci jednoduchých příznaků na dodané nasegmentované buňky. 4) Otestujte extrahované příznaky při rozpoznání mrtvých buněk s využitím zvoleného klasifikátoru. 5) Rozšiřte databázi použitých příznaků o pokročilejší příznaky. 6) Pomocí metod pro shlukování a redukci dimenzionality analyzujte výskyt různých buněčných typů a fází. 7) Proveďte diskusi dosažených výsledků a zhodnoťte účinnost a využitelnost jednotlivých příznaků.

DOPORUČENÁ LITERATURA:

[1] SONKA, Milan. Image processing, analysis, and machine vision. 3rd ed. Toronto: Thompson Learning, c2008. ISBN 978-0495082521.

[2] Two dimensional shape and texture quantification. BANKMAN, Isaac N. Handbook of Medical Image Processing and Analysis. 2nd ed. San Diego: Academic Press, 2009, s. 261-277. ISBN 978-0-12-373904-9.

Termín zadání: 5.2.2018 ***Termín odevzdání:*** 25.5.2018

Vedoucí práce: Ing. Tomáš Vičar
Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

UPOZORNĚNÍ:
Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato bakalářská práce se zabývá problematikou příznaků používaných při analýze snímků buněk pořízených holografickým mikroskopem a pro jejich klasifikaci. Popisuje jednotlivé příznaky, stejně tak jako nástroje pro klasifikaci. Příznaky jsou extrahovány na dodaných nasegmentovaných buňkách v programovém prostředí Matlab. Na základě těchto příznaků jsou buňky klasifikovány s využitím metody podpůrných vektorů. Pomocí metod pro shlukování a redukci dimenzionality je analyzován výskyt různých buněčných typů. Zároveň je zhodnocena účinnost a využitelnost jednotlivých příznaků.

KLÍČOVÁ SLOVA

Holografická mikroskopie, příznak, klasifikace, metoda podpůrných vektorů

ABSTRACT

This thesis deals with features used for analysis and classification of cell images captured by holographic microscope. Distinctive features are described together with tools for their classification. Features are extracted on provided segmented cells with use of Matlab programming environment. Based on extracted features the cells are classified by SVM classificator. With use of clustering methods and dimensionality reduction different cell types are analyzed. Reliabity of each feature is tested.

KEYWORDS

Holographic microscopy, feature, classification, support vectores

NAVRÁTILOVÁ, Markéta. *Příznaky pro analýzu a klasifikaci buněk ve snímcích z holografického mikroskopu*. Brno, 2018, 56 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce: Ing. Tomáš Vičar

Vysázeno pomocí balíčku `thesis` verze 2.63; `http://latex.feec.vutbr.cz`

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Příznaky pro analýzu a klasifikaci buněk ve snímcích z holografického mikroskopu“ jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor(ka) uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil(a) autorská práva třetích osob, zejména jsem nezasáhl(a) nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom(a) následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora(-ky)

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu bakalářské práce panu Ing. Tomáši Vičarovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno ....

podpis autora(-ky)

OBSAH

Úvod **10**

1 Teoretická část **11**
1.1 Holografická mikroskopie 11
1.1.1 Princip holografie 11
1.1.2 Digitální holografická mikroskopie 11
1.1.3 Nekoherentní holografická mikroskopie 13
1.2 Zpracování snímků buněk 16
1.2.1 Segmentace 16
1.2.2 Tracking 16
1.2.3 Extrakce příznaků 17
1.2.4 Klasifikace 17
1.3 Příznaky pro analýzu a klasifikaci buněk 18
1.3.1 Radiometrické příznaky 18
1.3.2 Fotometrické příznaky 23
1.3.3 Dynamické příznaky 25
1.4 Klasifikátory 26
1.4.1 SVM klasifikátor 26
1.5 Analýza dat 31
1.5.1 Fisherův diskriminant 31
1.5.2 Mann-Whitneyův U-test 31
1.5.3 Analýza hlavních komponent 32

2 Praktická část **33**
2.1 Extrakce příznaků 33
2.2 Analýza buněčných fází 37
2.3 Klasifikace mrtvých buněk 44

3 Závěr **46**

Literatura **47**

Seznam symbolů, veličin a zkratek **49**

A Grafická vizualizace dat **50**

B Výsledky klasifikace **53**

C Seznam skriptů a funkcí **55**

SEZNAM OBRÁZKŮ

1.1 Princip holografického zobrazení 11
1.2 Schéma in-line DHM [2] 12
1.3 Schéma off-axis DHM [2] 13
1.4 Schéma CCHM [4] 14
1.5 Schéma MHM [3] 15
1.6 Schéma zpracování snímků buněk 16
1.7 Snímek buněk 16
1.8 Segmentovaná maska buněk 16
1.9 Tři objekty, jejichž kompaktnost je (a) 4π, (b) 15,4 a (c) 27,6 [6] . . . 19
1.10 Kódové symboly pro směry hranice pro 4-okolí a pro 8-okolí 21
1.11 Ukázka řetězového kódu 22
1.12 Nadrovina a hraniční pásmo pro lineární SVM 27
1.13 Lineárně neseparabilní data 28
1.14 Lineární oddělení původně nelineárně separabilních dat pomocí přidané dimenze [14] 29
2.1 Snímek živých buněk (vlevo) a mrtvých buněk (vpravo) 33
2.2 Boxplot pro příznaky radiální vzdálenost (vlevo) a euklidovská vzdálenost (vpravo) 34
2.3 Mitotické buňky 37
2.4 Vliv příznaků na hlavní komponenty 38
2.5 Testovací data 1 v PCA prostoru 39
2.6 Buňky č. 21 (vlevo) a č. 24 (vpravo) klasifikované jako mitotické . . . 39
2.7 Testovací sada 1 40
2.8 Testovací data 2 v PCA prostoru 40
2.9 Buňky č. 8 (vlevo nahoře) a č. 10 (vpravo nahoře) a č. 30 (dole uprostřed) klasifikované jako mitotické 41
2.10 Testovací sada 2 41
2.11 Testovací data 3 v PCA prostoru 42
2.12 Buňky č. 1 a (vlevo nahoře) a č. 10 (vpravo nahoře) klasifikované jako mitotické 42
2.13 Buňky č. 11 (vlevo nahoře) a č. 12 (vpravo nahoře) a č. 13 (dole uprostřed) klasifikované jako mitotické 43
2.14 Testovací sada 3 43

SEZNAM TABULEK

2.1 Analýza příznaků - Fisherův diskriminant 35
2.2 Analýza příznaků - Mann-Whitneyův U-test 36
2.3 Analýza příznaků - kvantily 36
2.4 Úspěšnost klasifikace buněk na trénovacích datech 44
2.5 Úspěšnost klasifikace buněk na testovacích datech 45
B.1 Úspěšnost klasifikace buněk na testovacích datech 1 a 2 za použití různých modelů klasifikátoru 53
B.2 Úspěšnost klasifikace buněk na testovacích datech 3 a 4 za použití různých modelů klasifikátoru 54

ÚVOD

V současnosti nastává rozvoj biomedicínských aplikací, s čímž souvisí i nové technologie a metody. Jednou z nich je i digitální holografická mikroskopie, ke které značně přispělo Vysoké učení technické ve spolupráci se společností TESCAN vývojem multimodálního holografického mikroskopu. Tento mikroskop umožňuje pozorovat živé buňky s využitím nekoherentního osvětlení a bez použití kontrastních látek, které by mohly ovlivňovat chování a vlastnosti buněk. S využitím tohoto mikroskopu probíhají dlouhé experimenty na živých buňkách a výsledkem je až několik stovek obrazů buněk.

Právě v těchto obrazech chceme buňky klasifikovat. Aby mohlo dojít k automatickému zpracování, musíme vhodně zvolit příznaky pro klasifikaci, což je hlavním cílem této práce.

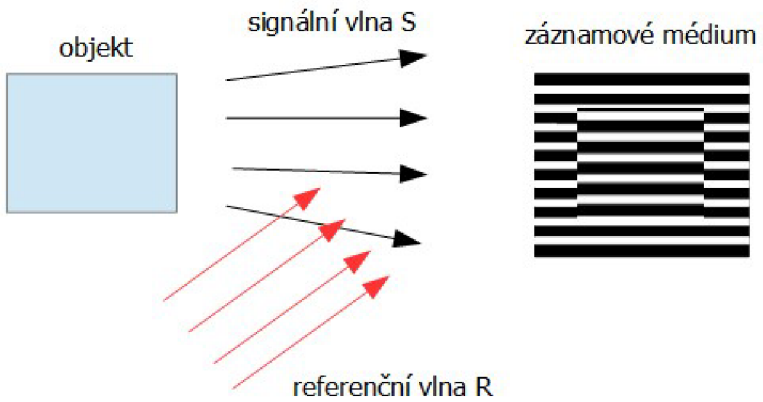
Teoretická část práce je rozdělená na pět hlavních kapitol, věnující se holografické mikroskopii, zpracování buněk, samotným příznakům a v neposlední řadě klasifikátorům. V poslední kapitole jsou stručně zmíněny postupy pro analýzu příznaků. Praktická část obsahuje extrakci jednotlivých příznaků na dodaných nasegmentovaných snímcích buněk a následnou klasifikaci buněk na základě příznakového vektoru, který vznikl z extrahovaných příznaků.

1 TEORETICKÁ ČÁST

1.1 Holografická mikroskopie

1.1.1 Princip holografie

Běžné zobrazovací metody, jako například fotografie, zaznamenávají pouze intenzitu světla. Jedinečnost holografie spočívá v zaznamenávání i fáze světelné vlny. Jelikož nám známé záznamové materiály nejsou schopné zachytit fázi světelné vlny jako takovou, musí být převedena na intenzitu. Toho se docílí pomocí referenční vlny, která interaguje se signálovou vlnovou za vzniku interferenčních proužků, jejichž intenzita je zaznamenávána, jak je vidět na obrázku 1.1. Tato intenzita již v sobě obsahuje informaci o fázi světelné vlny. Takto vzniklý obraz nezobrazuje informaci přímo, proto musí být následně zpracován pomocí numerických metod. [1]



Obr. 1.1: Princip holografického zobrazení

Holografické techniky rozdělujeme na optickou holografii a na digitální holografii. Při optické holografii se hologram zobrazí na fotografický materiál, který vyvoláme osvícením referenční vlnou. Veškeré mikroskopické techniky pracují na principu digitální holografické mikroskopie (DHM), proto se následující části zaměřují právě na ni.

1.1.2 Digitální holografická mikroskopie

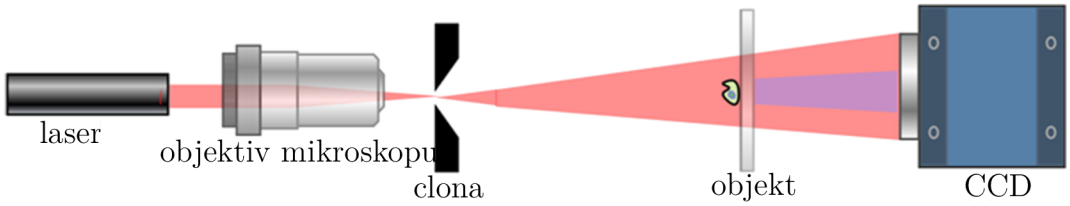
Mezi zobrazovací metody používané v biomedicínských aplikacích se řadí i rozvíjející se DHM. Různé techniky DHM, z nichž nejpoužívanější jsou in-line DHM nebo off-axis DHM, se osvědčily jako účinné nástroje pro profilování a sledování objektů s malou velikostí v 3-D objemech. Obecně se jedná o dvoustupňový proces, kdy se

nejprve zaznamená hologram, který se následně rekonstruuje pomocí numerických metod, aby se získal obraz. [2]

In-line DHM

In-line DHM systémy jsou charakterizovány nulovým úhlem mezi objektem a referenčním paprskem. Jediný světelný paprsek zaostřený do clony s otvorem o průměru řádu vlnové délky osvětluje objekt. Signální paprsek objektu je částí dopadajícího světla, které je rozptýleno objektem a nedotčený zbytek je považovaný za referenční paprsek, jak je vidět na obrázku 1.2.

V případě koherentního světelného zdroje není clona nezbytná. Pole objektu je automaticky zarovnáno s referenčním paprskem a interference objektu a referenčního paprsku vede ke vzniku interferogramu, který je zaznamenán CCD kamerou a poté přenesen do počítače, kde se numericky rekonstruuje. [2]



Obr. 1.2: Schéma in-line DHM [2]

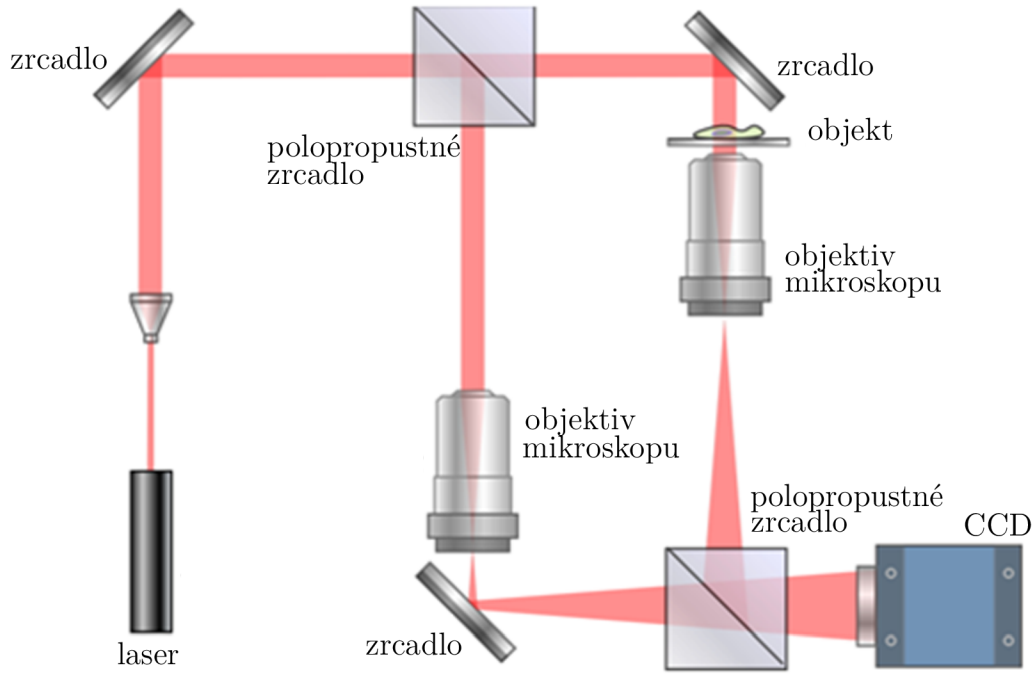
Tyto systémy však vyžadují zachycení více než jednoho interferogramu k rekonstrukci signální vlny, což může být omezujícím faktorem při zobrazování rychle se měnících jevů. [3]

Off-axis DHM

Off-axis DHM systémy, na rozdíl od in-line DHM, jsou charakterizovány nenulovým úhlem mezi objektem a referenčním paprskem, jinými slovy, světelný paprsek je rozdělen na paprsek signální a paprsek referenční. Tyto systémy neumožňují použití nekoherentního zdroje světla. Pro rekonstrukci signální vlny objektu je však zapotřebí pouze jeden zachycený interferogram, který činí tyto systémy vhodné pro zobrazování rychle se měnících jevů a přináší vysokou stabilitu měření fáze. [3]

Na obrázku níže 1.3 je zobrazeno schéma off-axis DHM systému. Jedno rameno obsahuje objekt (vzorek) a objektiv, který promítne zvětšený obraz objektu na CCD kameru. Druhým ramenem prochází referenční paprsek. Referenční a signálové vlny jsou posunuty o takový úhel, aby se zabránilo překrývání referenčního a signálového

obrazu. Zachycený holografický snímek se následně zpracuje pomocí numerických metod. [2]



Obr. 1.3: Schéma off-axis DHM [2]

1.1.3 Nekoherentní holografická mikroskopie

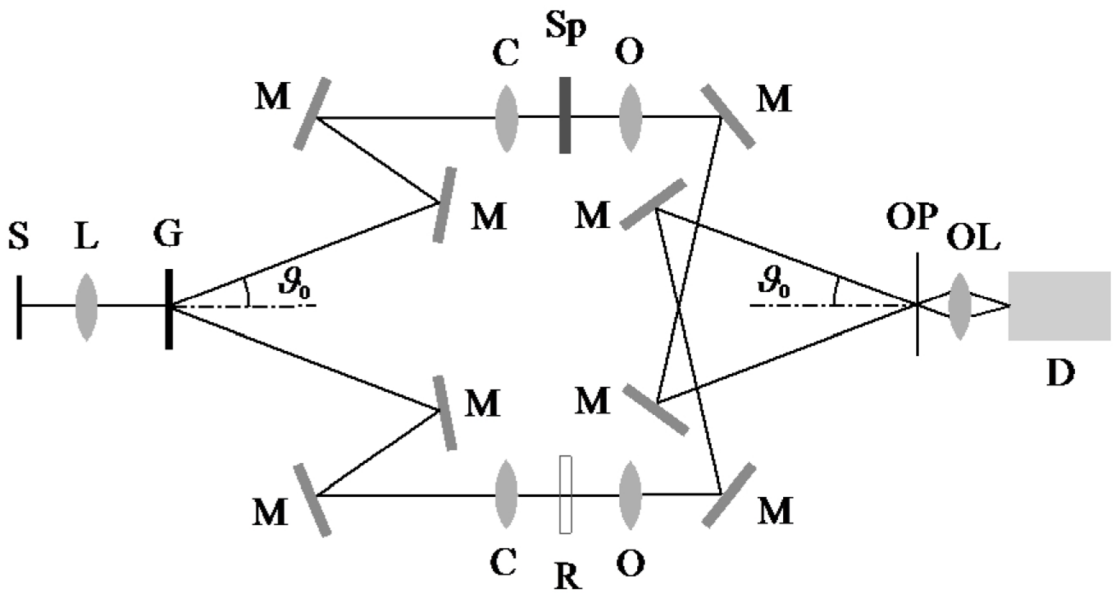
Tento typ holografické mikroskopie využívá off-axis zapojení spolu s achromatickým interferometrem, což umožňuje použití i nekoherentního zdroje světla. Achromatický interferometr použitý v tomto zapojení zaznamenává jak fázi, tak amplitudu výstupu pomocí koherentního referenčního svazku a vykazuje tak vlastnost prostorové invariance. Výsledkem je hradlování koherence a silné potlačení šumu, díky čemuž se dosahuje vysoké přesnosti měření a kvality obrazu. [3] [4]

Koherencí řízený holografický mikroskop

Schéma koherencí řízeného holografického mikroskopu (CCHM - Coherence-Controlled Holographic Microscope) je vyobrazeno na obrázku níže (Obr. 1.4).

Difrakční mřížka G je zobrazena ve výstupní rovině OP, přičemž 1. a -1. difrakční řád se šíří přes oddělená ramena, signální a referenční. Tato ramena jsou totožná a zrcadlově symetrická. Obě ramena se skládají z identických objektivů O a kondenzorů C s konečnou tubusovou délkou. Rovina difrakční mřížky G a výstupní rovina

OP jsou v obou ramenech opticky sdružené s rovinou objektu. Konečný interferenční obraz je vytvořen ve výstupní rovině OP, kde se oba difrakční řády překříží. [4]



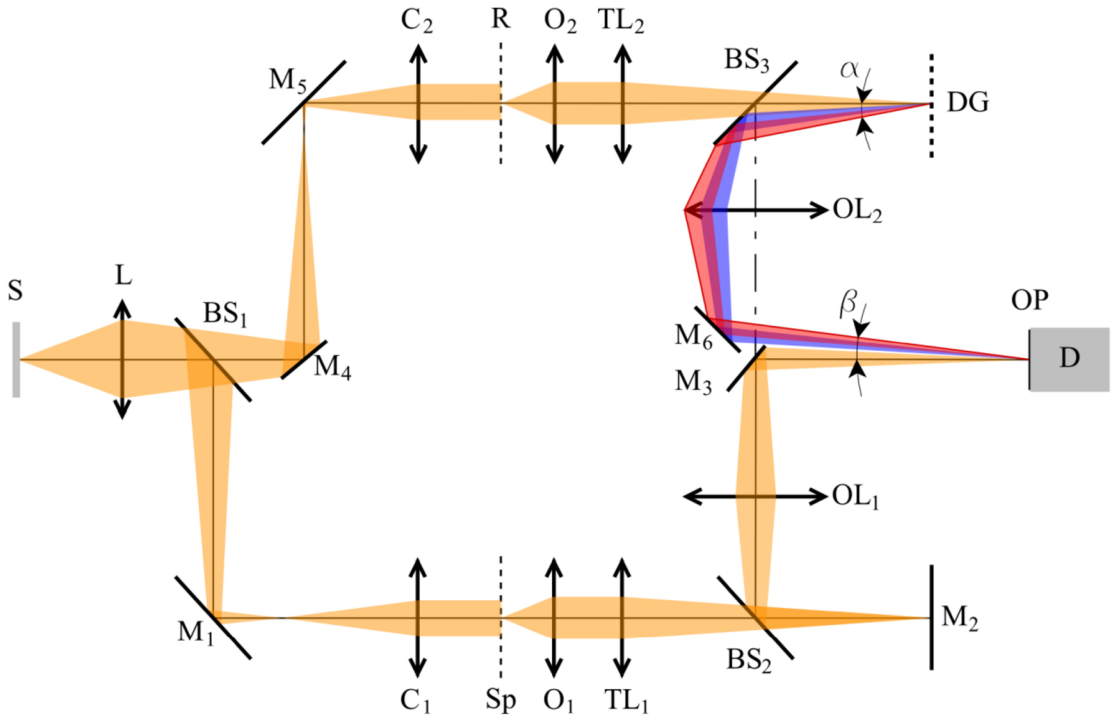
Obr. 1.4: Schéma CCHM [4]

S. . . zdroj světla, L. . . kolektor, M. . . zrcadla, G. . . difrakční mřížka, C. . . kondenzor, R. . . referenční komůrka, Sp. . . pozorovaný objekt, O. . . objektiv, OP. . . výstupní rovina, OL. . . výstupní čočka, D. . . detektor.

Multimodální holografický mikroskop

Tento mikroskop vznikl ve spolupráci VUT se společností TESCAN a byl použit k zaznamenání dat zkoumaných v této práci. Jeho konfigurace vychází z původního CCHM a je vidět na obrázku 1.5. Nové uspořádání využívá Machova-Zenderova interferometru. Podobně jako u CCHM se světlo šíří přes oddělené optické dráhy - signální a referenční rameno. Obě ramena jsou tvořena totožnými prvky, a to kondenzory C, objektivy O a tubusovými čočkami TL. Difrakční mřížka DG je umístěna v referenčním ramenu a zobrazena na výstupní rovině OP. Rovina difrakční mřížky DG a objektové roviny Sp, R jsou opticky sdružené s výstupní rovinou OP objektivem a výstupními čočkami OL. Konečný obraz není tvořen přímo referenčním paprskem ve výstupní rovině OP, protože pro zobrazování je použit pouze 1. difrakční řád, zatímco jiné difrakční řády jsou odfiltrovány v ohniskové rovině výstupních čoček. Avšak jakmile se signální a referenční paprsek překříží na výstupní rovině, vzniknou interferenční proužky odpovídající vrypům interferenční mřížky o

konstantní frekvenci pro všechny vlnové délky a je tak dosaženo achromatičnosti interferometru.



Obr. 1.5: Schéma MHM [3]

S. . . zdroj světla, L. . . kolektor, M. . . zrcadlo, BS. . . dělič svazku, C. . . kondenzor, Sp. . . pozorovaný vzorek, R. . . referenční komůrka, O. . . objektiv, TL. . . tubusová čočka OP. . . výstupní rovina, D. . . detektor, DG. . . difrakční mřížka, OL. . . výstupní čočka.

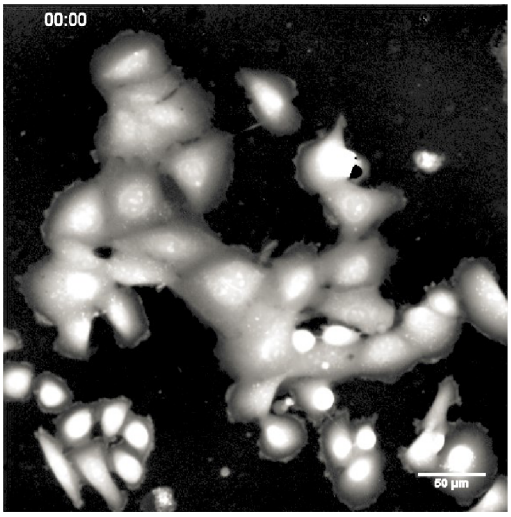
1.2 Zpracování snímků buněk

Obecné schéma zpracování snímků buněk se řídí následujícím schématem:

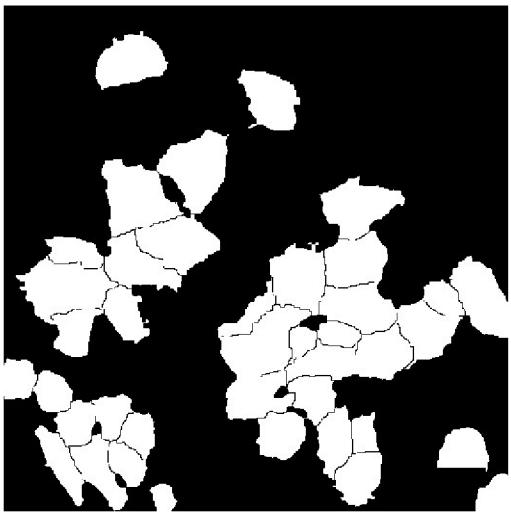
Obr. 1.6: Schéma zpracování snímků buněk

1.2.1 Segmentace

Segmentace je jedním z nějdůležitějších kroků vedoucí k analýze obrazu. Během segmentace dojde k rozpoznávání relevantních objektů (např. buněk) a jejich oddělení od pozadí v každém snímku. Úskalím segmentace je nejednoznačnost dat (podobnost objektu s pozadím) a šum. [5] [8]Na následujícím obrázku je zobrazen snímek buněk (Obr. 1.7) a jeho nasegmentovaná maska (Obr. 1.8).



Obr. 1.7: Snímek buněk



Obr. 1.8: Segmentovaná maska buněk

Mezi metody segmentace se řadí segmentace prahováním, segmentace na základě detekce hran a segmentace narůstáním oblastí. [5]

1.2.2 Tracking

Trackingem je označováno nalezení sobě odpovídajících objektů ve snímcích a následné sdružování segmentovaných objektů z jednoho snímku do dalšího. Nejjednodušším přístupem je propojení každého segmentovaného objektu ve snímku s nej-

bližším objektem v následujícím snímku. „Nejbližší“ může být myšleno jednak na vzdálenost, ale i na nejbližší podobnost intenzity, objemu, popř. dalších vlastností. Toto řešení funguje dobře v případě, kdy objekty jsou dostatečně rozdílné alespoň v jednom z příznaků. [8]

1.2.3 Extrakce příznaků

Příznaky představují vlastnosti objektů, které jsou pro objekt specifické a tudíž na jejichž základě můžeme objekt klasifikovat. Jeden příznak zpravidla ke správné klasifikaci nestačí, proto se vytváří tzv. příznakový vektor. Příznaky se dělí na radiometrické, které popisují tvarové vlastnosti objektu, a fotometrické, které popisují jasové vlastnosti objektu. [5]

1.2.4 Klasifikace

Rozpoznávání objektů je založeno na přiřazování objektů do určitých tříd, kdy počet tříd je obvykle předem znám. Tento proces má na starost tzv. klasifikátor. Klasifikátor se rozhoduje pro přidělení objektů k příslušným třídám na základě naučených vzorů, jedná se tedy o metodu strojového učení s učitelem. [5]

1.3 Příznaky pro analýzu a klasifikaci buněk

Pro správnou klasifikaci buněk se musí stanovit jejich specifické vlastnosti, typické pro daný typ. Tyto vlastnosti se označují jako příznaky. Příznaky musí být spolehlivé a dostatečně rozlišující. Hlavní dva druhy příznaků jsou příznaky radiometrické a fotometrické. V případě časových záznamů lze využít i příznaky dynamické.

1.3.1 Radiometrické příznaky

Radiometrické příznaky se rozdělují na příznaky založené na regionech a na příznaky založené na hranicích. První jmenované popisují metrické vlastnosti objektů vypočítané z plošného rozložení pixelů. Druhé jmenované popisují geometrické vlastnosti objektu, které jsou odvozené na základě transformování hranice objektu do jednorozměrného signálu a jeho analyzování. Mezi radiometrické příznaky založené na regionech se řadí velikost, obvod, kompaktnost, excentricita, podlouhlost, konvexnost, lokální momenty a další. Možností je nespočet, záleží, jaký příznak bude vhodně popisovat náš objekt. Mezi radiometrické příznaky založené na hranicích patří radiální vzdálenost, řetězový kód a Fourierovy příznaky.

Velikost

Nejjednodušším příznakem každého objektu je jeho velikost. Velikost udává počet pixelů, z nichž je objekt tvořen.

Obvod

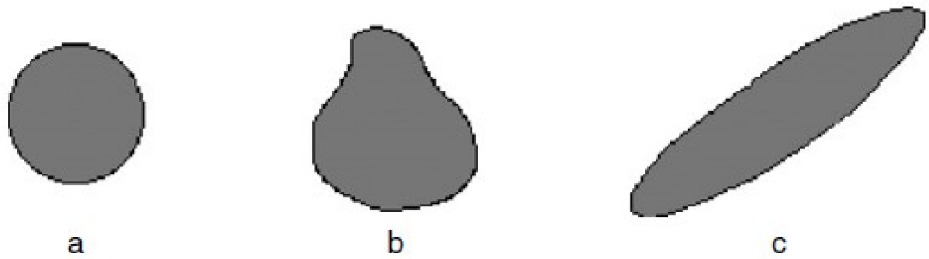
Obvod označuje počet pixelů, které tvoří hranici objektu.

Kompaktnost

Kompaktnost C udává míru podobnosti objektu vůči kruhu. Vypočítá se jako poměr druhé mocniny obvodu P k ploše objektu A.

$$C = \frac{P^2}{A} \tag{1.1}$$

Objekty perfektně kruhového tvaru mají nejnižší hodnotu C (4π). Se zvyšující se složitostí tvaru roste kompaktnost. Tato vlastnost poskytuje dobrou korelaci mezi kompaktností a vnímáním nepravidelnosti tvaru. Neslouží ale jako přesný ukazatel složitosti tvaru. Výhodou této charakteristiky je výpočetní jednoduchost, stejně tak jako nezávislost vůči měřítku, natočení či posunutí. [6]



Obr. 1.9: Tři objekty, jejichž kompaktnost je (a) 4π, (b) 15,4 a (c) 27,6 [6]

Excentricita

Excentricita je definována jako poměr hlavní osy A a na ni kolmé vedlejší osy B. Nabývá hodnot od 0 do 1, kdy 0 znamená, že objektem je kruh a 1 znamená, že se jedná o úsečku. [5] [7]

Podlouhlost

Podlouhlost se vypočítá jako poměr stran obdélníka opsaného objektu. Opsaný obdélník musí mít co nejmenší obsah, čehož se docílí postupným natáčením. [5]

Konvexnost

Konvexnost udává míru podobnosti objektu ke své konvexní schránce. Vypočítá se jako poměr plochy objektu k ploše konvexního obalu. Nabývá hodnot od 0 do 1, přičemž hodnota 1 platí pro konvexní (vypouklé) objekty. [5]

Lokální momenty

Lokální momenty udávají informaci o rozložení a tvaru souboru bodů. 2D obraz je reprezentován třemi hodnotami (i, j, f), kde (i, j) značí prostorové souřadnice a $f(i, j)$ intenzitu pixelu. [6]

Pro lokální moment $p + q$ řádu obrazu o MxN pixelech platí:

$$m_{pq} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i^p j^q f(i, j) \qquad p, q = 0, 1, 2, 3, ... \tag{1.2}$$

Lokální momenty v závislosti na jejich řádu poskytují velké množství informací. Například u binárních obrazů lokální moment nultého řádu udává velikost, protože platí:

$$m_{00} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i^0 j^0 f(i, j) \tag{1.3}$$

$$m_{00} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i,j) \tag{1.4}$$

Jelikož binární obraz jsou pouze nuly a jedničky, nultý moment spočítá počet všech jedniček (nuly se v součtu neprojeví) a tím se získá velikost.

Aby se dosáhlo nezávislosti na rotaci, používá se centrovaný moment:

$$\mu_{pq} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (i-\overline{i})^p (j-\overline{j})^q f(i,j) \qquad p,q = 0,1,2,3,... \tag{1.5}$$

kde $\overline{i}, \overline{j}$ jsou souřadnice centroidu, které se získají dle:

$$\overline{i} = \frac{m_1 0}{m_0 0} \qquad \overline{j} = \frac{m_0 1}{m_0 0} \tag{1.6}$$

jelikož pro výpočet $m_1 0$ platí rovnice 1.7 a 1.8, které sčítají všechny i souřadnice všech pixelů s hodnotou 1.

$$m_{10} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i^1 j^0 f(i,j) \tag{1.7}$$

$$m_{10} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i \cdot f(i,j) \tag{1.8}$$

Stejně tak pro výpočet $m_0 1$ platí rovnice 1.9 a 1.10, které sčítají všechny j souřadnice všech pixelů s hodnotou 1.

$$m_{01} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i^0 j^1 f(i,j) \tag{1.9}$$

$$m_{01} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} j \cdot f(i,j) \tag{1.10}$$

Po získání součtu i a j souřadnic, se obě hodnoty vydělí celkovou plochou, kterou udává nultý moment $m_0 0$. Tak se získá jejich průměrná hodnota, která udává polohu centroidu $\overline{i}, \overline{j}$.

Radiální vzdálenost

Radiální vzdálenost $d(n)$ udává vzdálenost každého pixelu hranice $(x(n), y(n))$ od centrálního bodu. Za centrální bod je považován centroid (x_c, y_c) a výsledná sekvence hodnot radiální vzdálenosti se získá vysledováním všech pixelů hranice N. [6]

$$d(n) = \sqrt{(x(n)-x_c)^2 + (y(n)-y_c)^2} \qquad n = 0,1,...,N-1 \tag{1.11}$$

Aby byla zajištěna invariance vůči měřítku, výsledná hodnota radiální vzdálenosti $d(n)$ se normalizuje:

$$r(n) = \frac{d(n)}{max[d(n)]} \tag{1.12}$$

Normalizovaná hodnota $r(n)$ může být použita pro výpočet obecného momentu m_p nebo centrálního momentu μ_p, kde p značí řád momentu:

$$m_p = \frac{1}{N} \sum_{n=0}^{N-1} [r(n)]^p \tag{1.13}$$

$$\mu_p = \frac{1}{N} \sum_{n=0}^{N-1} [r(n) - m_1]^p \tag{1.14}$$

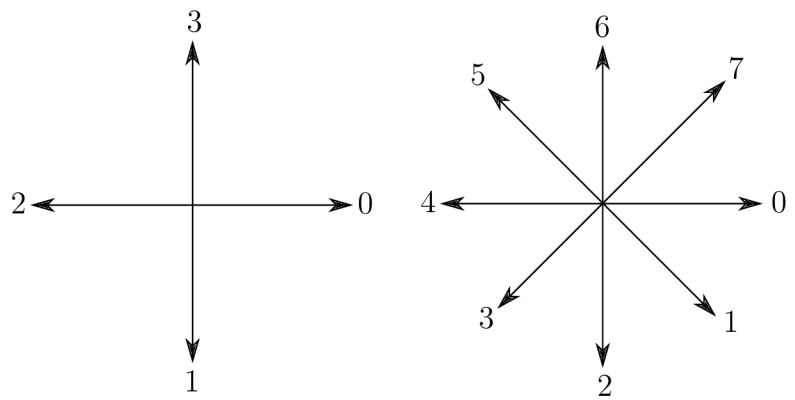
Normalizované momenty invariantní vůči měřítku, natočení či posunutí se získají jednoduchou úpravou:

$$\overline{m_p} = \frac{m_p}{\mu_2^{p/2}} \tag{1.15}$$

$$\overline{\mu_p} = \frac{\mu_p}{\mu_2^{p/2}} \qquad p \neq 2 \tag{1.16}$$

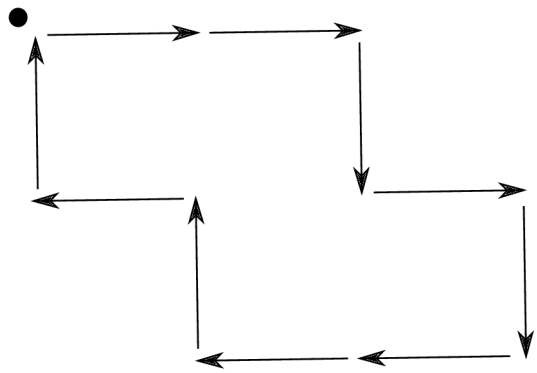
Řetězový kód

Jednou z metod určení hranic objektu je řetězový kód, někdy označovaný po svém autorovi jako Freemanův kód. Tvar objektu je v tomto případě definován relativní pozicí po sobě jdoucích bodů jeho hranice. Každému bodu hranice je přiděleno označení, které má význam určitého směru pohybu vzhledem k pozici z předchozího bodu. V případě 4-okolí se standardně používá označení číslicemi nula až tři, v případě 8-okolí označení číslicemi nula až sedm. Jiná označení jsou možná, nesmí se ale zapomenout vždy uvést pravidla tvorby kódu. [5] [6]



Obr. 1.10: Kódové symboly pro směry hranice pro 4-okolí a pro 8-okolí

V případě označení znázorněného výše na Obr. 1.10 by řetězový kód obrazce na Obr. 1.11 vypadal takto: 0, 0, 1, 0, 1, 2, 2, 3, 2, 3.



Obr. 1.11: Ukázka řetězového kódu

Jelikož při popisu objektů o velkých rozměrech a vysokém rozlišení by řetězový kód byl příliš velký, obvykle se připravuje hranice pro získání řetězového kódu zmenšením rozlišení pomocí nové x-y mřížky. Výchozím bodem pro začátek řetězového kódu může být kterýkoli z těchto bodů. Každá hranice tak může být reprezentována tolika řetězovými kódy, kolik má objekt výchozích bodů. Aby se zabránilo této nejednoznačnosti, volí se za začátek bod, který vyprodukuje řetězový kód s nejmenší číselnou hodnotou.

Řetězový kód invariantní vůči natočení lze získat pomocí rozdílu směru mezi dvěma po sobě následujícími číslicemi. Tak se dostane číslo udávající počet kroků mezi jednotlivými body. Za kladný krok se označuje směr proti směru hodinových ručiček. Například v řetězovém kódu na Obr. 1.11 průchod z prvního bodu do druhého neznamená změnu směru, tudíž první prvek nového řetězce je 0. Průchod z druhého bodu do třetího je posunut o jeden krok po směru hodinových ručiček, tedy - 1. Z třetího bodu do čtvrtého se směr změní o jeden krok proti směru hodinových ručiček, tedy 1. [6] Tímto způsobem se získá jednoznačný řetězový kód invariantní vůči rotaci (pro objekt na Obr. 1.11 by tento kód byl: 0, -1, 1, -1, -1, 0, -1, 1, -1, -1).

Fourierovy příznaky

Každý pixel hranice objektu může být reprezentován komplexním číslem. Z těchto čísel je vytvořena sekvence komplexních čísel, nazývaná radiální funkce $c(n)$. Radiální funkce je vždy počítána vzhledem k nějakému bodu, nejčastěji centroidu (x_c, y_c): [6]

$$c(n) = x(n) - x_c + j(y(n) - y_c) \qquad n = 0, 1, ..., N-1 \tag{1.17}$$

kde n je index pixelu a N počet všech pixelů hranice.

Aplikací Fourierovy transformace na radiální funkci $c(n)$ se získají spektrální koeficienty $d(u)$:

$$d(u) = \frac{1}{N} \sum_{n=0}^{N-1} c(n) e^{-} \frac{j2\pi nu}{N} \qquad u = 0, 1, ..., N-1 \tag{1.18}$$

Získané koeficienty obsahující veškeré informace o tvaru hranice a jsou nezávislé na translaci, což je dáno přesnou segmentací. Dále jsou nezávislé na rotaci - otočení se projeví cyklickým posunem prvků radiální funkce, což se projeví změnou fáze $d(u)$, ale amplituda zůstane beze změny. [6]

1.3.2 Fotometrické příznaky

Fotometrické příznaky odráží optické vlastnosti objektu. Kromě základních jasových ukazatelů popsaných v následující části, se mezi ně řadí histogram a jeho deskriptory a statistické momenty.

Základní fotometrické příznaky

Mezi základní fotometrické příznaky objektu Ω patří průměrná jasová úroveň objektu, maximální a minimální jasová úroveň a diference extrémů jasové funkce.

$$průměrná\ jasová\ úroveň = \frac{1}{N} \sum_{(x,y)\in\Omega} f(x,y) \tag{1.19}$$

$$maximální\ jasová\ úroveň = max(f(x,y)) \qquad (x,y) \in \Omega \tag{1.20}$$

$$minimální\ jasová\ úroveň = min(f(x,y)) \qquad (x,y) \in \Omega \tag{1.21}$$

$$diference\ extrémů = B_{max} - B_{min} \tag{1.22}$$

Histogram a jeho deskriptory

Histogram $h(z)$ udává četnost jednotlivých jasových úrovní z v obraze. Je to tedy vektor, který obsahuje počet pixelů v obraze na každé šedé úrovni. Jako příznak je vhodnější použít jeho deskriptory, mezi které patří průměrná hodnota histogramu, kontrast, energie a entropie. [6]

$$H_{mean} = \frac{1}{N} \sum_{z=1}^{N} q \cdot h(z) \tag{1.23}$$

$$H_{contrast} = \frac{1}{N} \sum_{z=1}^{N} (z \cdot h(z) - H_{mean})^2 \tag{1.24}$$

$$H_{energy} = \frac{1}{N} \sum_{z=1}^{N} h(z)^2 \tag{1.25}$$

$$H_{entropy} = \frac{1}{N} \sum_{z=1}^{N} h(z) \cdot log_2 h(z) \tag{1.26}$$

Další možností, jak využít histogram k získání příznaků, je získat pozici, na které se nachází maximální či minimální jasová úroveň. Díky velkému rozsahu využití je histogram významným ukazatelem při analýze a klasifikaci obrazů.

Statistické momenty

Statistické momenty jsou založené na histogramech intenzity, kde histogram $h(k)$ udává četnost zastoupení jasových úrovní K v obrazu v rozsahu hodnot 0 až $K-1$. Obrazy, jejichž textura je hladká, mají blízké jasové úrovně všech pixelů a naopak, obrazy, jejichž textura je hrubá, mají jasové úrovně pixelů rozdílné. Statistické momenty jsou odvozeny od histogramu a proto neobsahují prostorové informace, což je jejich nevýhodou. [6]

Průměrná hodnota histogramu μ udává průměrnou jasovou úroveň objektu a lze ji označit jako první statistický moment:

$$\mu = \frac{1}{K} \sum_{k=0}^{K-1} K \cdot h(k) \tag{1.27}$$

N-tý moment pro daný histogram se vypočítá dle:

$$m_n = \frac{1}{K} \sum_{k=0}^{K-1} (k - \mu)^n h(k) \tag{1.28}$$

Druhý moment m_2, rozptyl, je často používaný ukazatel, jelikož je výpočetně nenáročný a dobře koreluje s vnímáním drsnosti. Jinými slovy, poskytuje informaci o kontrastu. Třetí moment m_3, udává sklon, což vypovídá o asymetrii. Čtvrtý moment m_4 dává informace o kurtóze, tedy rovnoměrnosti. [6]

1.3.3 Dynamické příznaky

Pokud analyzované obrazy jsou časové záznamy, lze využít dynamických příznaků pro další analýzu. Zpravidla se jedná o vývoj/změnu v čase, z čehož vyplývá, že nejvýznamnějším příznakem bude rychlost.

Rychlost

Rychlost popisuje pohyb sledovaného objektu v obraze. Lze vypočítat ze změny polohy centroidu x, y, přičemž i udává číslo snímku.

$$v = \sqrt{(x_{i+1} - x_i)^2 + (y_{y+1} - y_i)^2} \tag{1.29}$$

Euklidovská vzdálenost jasových hodnot

Dalším dynamickým příznakem je euklidovská vzdálenost d_E jasových hodnot objektu v po sobě jdoucích snímcích. Jedná se o vyjádřeni změny jasu v jednotlivých pixelech, kde i označuje číslo pixelu a N počet všech pixelů objektu.

$$d_E = \sqrt{\sum_{i=1}^{N} (f(x_{i+1}, y_{i+1}) - f(x_i, y_i))^2} \tag{1.30}$$

1.4 Klasifikátory

Klasifikátory mají za úkol přiřadit objekty na základě jejich příznaků do tříd, jejichž počet je obvykle znám předem. Mezi nejznámější klasifikátory se řadí SVM klasifikátory, umělé neuronové sítě a rozhodovací stromy či Bayesův klasifikátor. Při klasifikaci buněk se využívá i Mahalanobisův klasifikátor nebo algoritmus k-nejbližších sousedů (k-NN, z ang. *k-nearest neighbours*). Pro klasifikaci buněk v této práci byl použitý SVM klasifikátor, jelikož je přesný a dobře použitelný na biologická data, protože umí pracovat s velkým množstvím příznaků. [10]

1.4.1 SVM klasifikátor

Support vector machine (SVM) je metoda strojového učení s učitelem pro klasifikaci problémů do dvou skupin. Zakládá se na nalezení tzv. nadroviny, která rozděluje příznakový prostor tak, aby příznaky náležící různým třídám ležely v opačných poloprostorech. Vzdálenost nadroviny od nejbližšího bodu v obou poloprostorech by měla být co největší. Tyto nejbližší body se nazývají podpůrné vektory (anglicky *support vectors*), z čehož vznikl název této metody.

Lineární SVM - separabilní případ

Nejjednodušším případem jsou lineární klasifikátory pro separovatelná trénovací data (příznaky), která lze lehce rozdělit nadrovinou. Předpokládejme, že trénovací data jsou $(\vec{x_i}, y_i)$, kde $\vec{x_i}$ je vektor příznaků a y_i informace od učitele. y_i nabývá hodnot ± 1 podle toho, do jaké třídy příznaky patří.

Pro rovnici nadroviny platí:

$$\vec{w} \cdot \vec{x} + b = 0 \tag{1.31}$$

kde $\vec{w}$ je normála nadroviny a $\|\vec{w}\|$ její Euklidovská norma.

Zároveň platí:

$$\vec{x_i} \cdot \vec{w} + b \geq +1 \qquad y_i = +1 \tag{1.32}$$

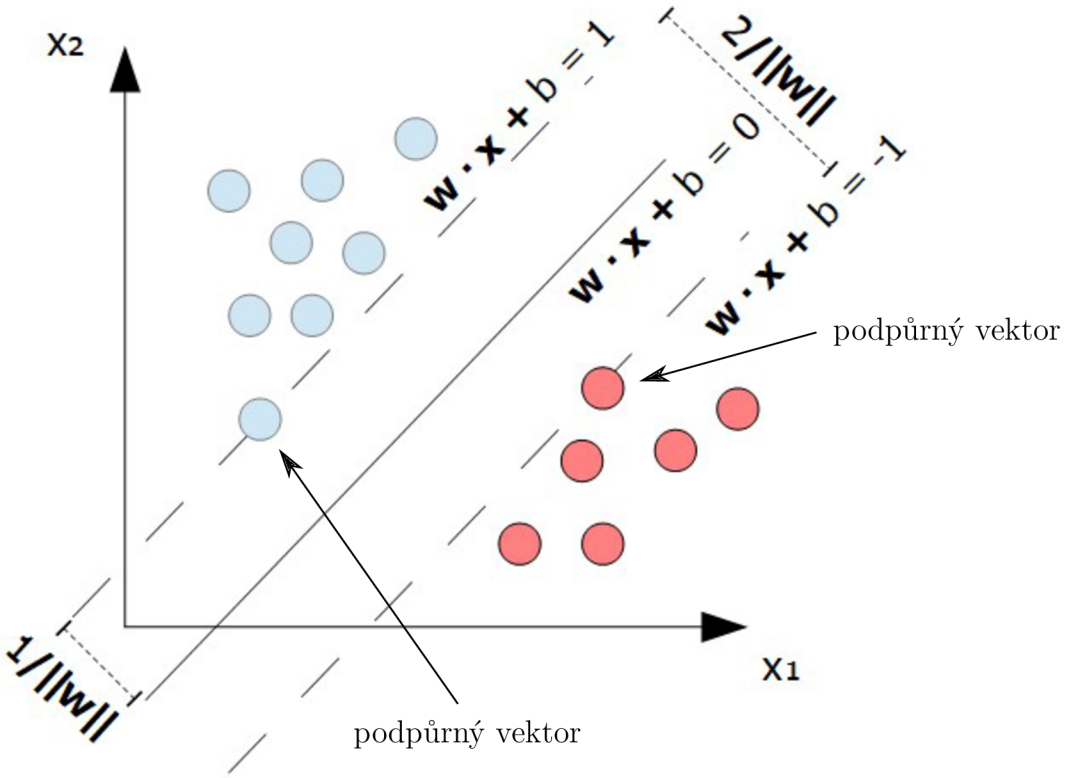
$$\vec{x_i} \cdot \vec{w} + b \leq -1 \qquad y_i = -1 \tag{1.33}$$

Zmíněné podmínky můžeme sloučit do jedné nerovnice:

$$y_i(\vec{x_i} \cdot \vec{w} + b) - 1 \geq 0 \qquad \forall i \tag{1.34}$$

Vzdálenost obou nadrovin je pak $2/\|w\|$ (Obr. 1.12), což jinými slovy označuje šířku hraničního pásma. Jelikož cílem je, aby vzdálenost obou nadrovin byla maximální, hledáme takové parametry $\vec{w}$ a b, aby se minimalizovala $\|w\|/2$, za podmínky že platí 1.34. K usnadnění výpočtů se využívá hledání minima $\|w\|^2/2$. Zároveň je

zřejmé, že nadrovina je definována daty k ní nejblíže a řešení nezávisí na ostatních datech.[9] [11]



Obr. 1.12: Nadrovina a hraniční pásmo pro lineární SVM

Z úlohy lze odvodit duální problém L_D za účelem nalezení maxima vzhledem k duálním proměnným α_i:

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\alpha_i\alpha_j y_i y_j \vec{x_i} \cdot \vec{x_j} \tag{1.35}$$

za omezujících podmínek:

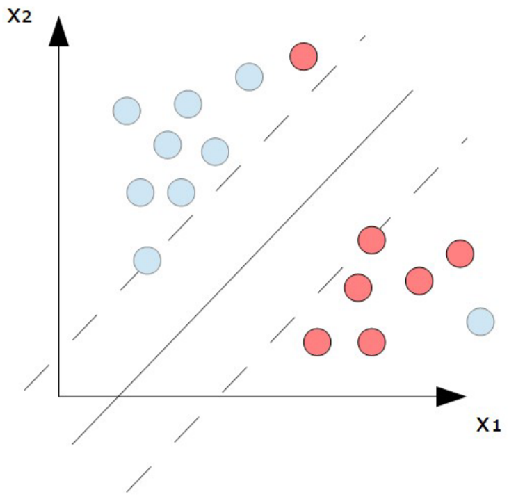
$$\alpha_i \geq 0 \qquad \forall i \tag{1.36}$$

a

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{1.37}$$

Lineární SVM - neseparabilní případ

Výše zmíněný algoritmus nejde použít, když máme neseparabilní data, která nelze jednoznačně lineárně oddělit s nulovou chybou (Obr. 1.13). Týká se to zejména dat, která jsou zašuměná a nebo se třídy částečně překrývají.



Obr. 1.13: Lineárně neseparabilní data

Abychom i přesto mohli použít lineární SVM, provede se zmírnění podmínek 1.32 a 1.33 zavedením relaxačních proměnných $\xi_i, i = 1, ..., l$, čímž pro prvotní podmínky dostaneme [9]:

$$\vec{x_i}\vec{w} + b \geq +1 - \xi_i \qquad y_i = +1 \tag{1.38}$$

$$\vec{x_i} \cdot \vec{w} + b \leq -1 + \xi_i \qquad y_i = -1 \tag{1.39}$$

$$\xi_i \geq 0 \qquad \forall i \tag{1.40}$$

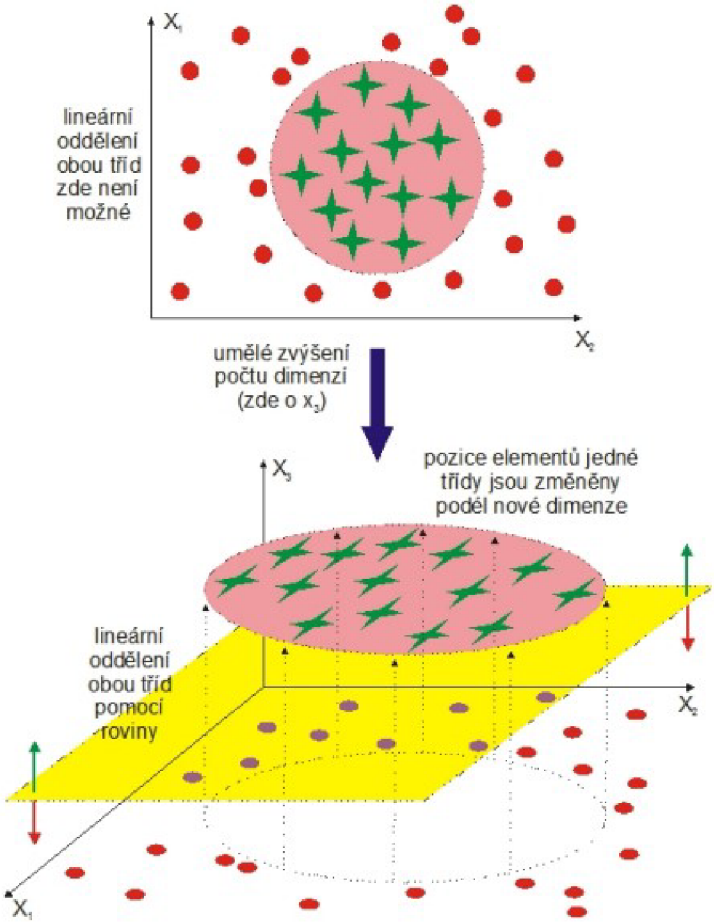
Primární problém pak vypadá tak, že se hledá minimum $\|w\|^2/2$ rozšířené o co nejmenší chybu $C \sum_{i=1}^{l} \xi_i$ (rovnice 1.41).

$$(\vec{w}, b, \xi_i) = argmin(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i) \tag{1.41}$$

C označuje konstantu zvolenou uživatelem. Čím větší je tato hodnota, tím větší je penalizace za chybu a tím pádem pak bude malý počet chybně klasifikovaných objektů, ale zároveň bude hraniční pásmo úzké. Snahou je tedy najít takové C, aby bylo ideálním kompromisem. [10]

Nelineární SVM

V případě, že máme data, která nejsou lineárně separabilní, se využívá tzv. jádrového triku (z anglického kernel trick), díky němůž se data převedou do jiné dimenze, kde už lineárně separabilní jsou (Obr. 1.14). [9]



Obr. 1.14: Lineární oddělení původně nelineárně separabilních dat pomocí přidané dimenze [14]

Využití jádrového triku je možné z toho důvodu, že při řešení optimalizačního problému se vstupní data vyskytují jen ve formě skalárního součinu $x_i \cdot x_j$ (rovnice 1.35). Tudíž není potřeba vypočítat přesnou transformaci dat do prostoru o vyšší dimenzi, pouze se skalární součiny v původním prostoru nahradí vnitřním produktem dat v prostoru o vyšší dimenzi - hodnotami jádrové funkce K.

Výraz pro nalezení maxima pak vypadá následovně:

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\alpha_i\alpha_j y_i y_j K(\vec{x_i}, \vec{x_j}) \tag{1.42}$$

Nejběžněji používanými jádrovými funkcemi jsou polynomiální funkce, přičemž d označuje stupeň polynomu:

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^d \tag{1.43}$$

a radiální bázová funkce, kde σ je volný parametr:

$$K(x_i, x_j) = exp(-\|x_i - x_j\|^2/2\sigma^2) \tag{1.44}$$

Tímto způsobem lze nalézt optimální rozdělovací nadrovinu pro jakákoli data za použití dostatečného počtu dimenzí [9]

1.5 Analýza dat

Během analýzy dat se využívá statistických metod pro popis a vyhodnocení těchto dat. V následujících podkapitolách jsou stručně popsány tři metody využité v praktické části této práce - Fisherův diskriminant, Mann-Whitneyův U-test a analýza hlavních komponent.

1.5.1 Fisherův diskriminant

Fisherova diskriminant D_F udává, jak jsou třídy od sebe oddělitelné. Obecně se vypočítá jako rozptyl mezi třídami vydělený rozptylem uvnitř jednotlivých tříd, konkrétně pak vztahem 1.45, kde $\mu_ž$ označuje střední hodnotu daného příznaku pro živé buňky, μ_m střední hodnotu pro mrtvé buňky, $\sigma_ž$ směrodatnou odchylku hodnot jednotlivých příznaků pro živé buňky a σ_m směrodatnou odchylku pro mrtvé buňky. Jelikož se počítá se středními hodnotami a směrodatnými odchylkami, předpokládá se normální rozdělení hodnot.[15]

$$D_F = \frac{|\mu_ž - \mu_m|^2}{\sigma_ž^2 + \sigma_m^2} \tag{1.45}$$

1.5.2 Mann-Whitneyův U-test

Jinou možnou metodou pro analýzu je Mann-Whitneyův U-test, jinak známý jako Wilcoxonův dvouvýběrový test (ang. *Wilcoxon Rank-Sum test*). Na rozdíl od Fisherova diskriminantu nepředpokládá normální rozložení dat a tudíž je obecněji aplikovatelný.

Hodnoty dat obou tříd jsou sloučeny a je vytvořeno jejich pořadí. Následně se hodnoty spolu s jejich pořadím rozdělí zpět do původních souborů a následně každou třídu reprezentuje jen pořadí jednotlivých hodnot a nikoli hodnoty samotné. Pořadí každé třídy se následně sečtou (T_1, T_2) a dle vztahů 1.46 a 1.47, kde m a n označují počet hodnot jednotlivých tříd, se vypočítá výsledná statistická hodnota.

$$U_1 = mn + \frac{m(m+1)}{2} - T_1 \tag{1.46}$$

$$U_2 = mn + \frac{n(n+1)}{2} - T_2 \tag{1.47}$$

Z obou hodnot se pak vybere jejich minimum U (rovnice 1.48), které je výsledkem statistického testu a porovnává se s kritickou hodnotou. Pokud je výsledná hodnota menší než kritická hodnota, nulová hypotéza shody distribučních funkcí obou tříd se zamítá. Jinými slovy, distribuční funkce jsou rozdílné a tudíž se jedná o dobře využitelný příznak.

$$U = min(U_1, U_2) \tag{1.48}$$

Jinou možností interpretace statistického testu je pomocí p-hodnoty. P-hodnota udává nejnižší možnou hladinu významnosti pro zamítnutí hypotézy.[16]

1.5.3 Analýza hlavních komponent

Analýza hlavních komponent (PCA - z ang. *Principal Component Analysis*) je metodou pro redukci dimenzionality. Tyto metody obecně odstraňují nadbytečnost komponent (příznaků) bez významné ztráty informací. PCA je založena na generování nových příznaků, které jsou lineární kombinací původních příznaků. Každý příznak z datové množiny v d rozměrném prostoru je mapován do k rozměrného podprostoru, který je menší než původní prostor. Data jsou transformovány tak, aby byl zachován co největší rozptyl hodnot. [17]

Toho se docílí tak, že se nejprve vypočítá kovarianční matice C mezi všemi příznaky. V případě, že $d = 3$, a máme tedy 3 dimenze x, y, z, kovarianční matice vypadá následovně:

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix} \tag{1.49}$$

kde jednotlivé kovariance mezi dimenzemi se vypočítají dle rovnice 1.50, kde n označuje počet příznaků a $\overline{x}, \overline{y}$ průměrné hodnoty pro daný příznak.

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)} \tag{1.50}$$

Jelikož výsledná kovarianční matice je čtvercová, lze z ní vypočítat matici vlastních vektorů W (ang. *eigenvectores*), která bude o rozměrech d x d.

Nová data Y jsou kombinací původních dat X a matice W, přičemž Y je stejně jako X o rozměru n x d, kde n označuje počet příznaků a d udává rozměr prostoru. [18]

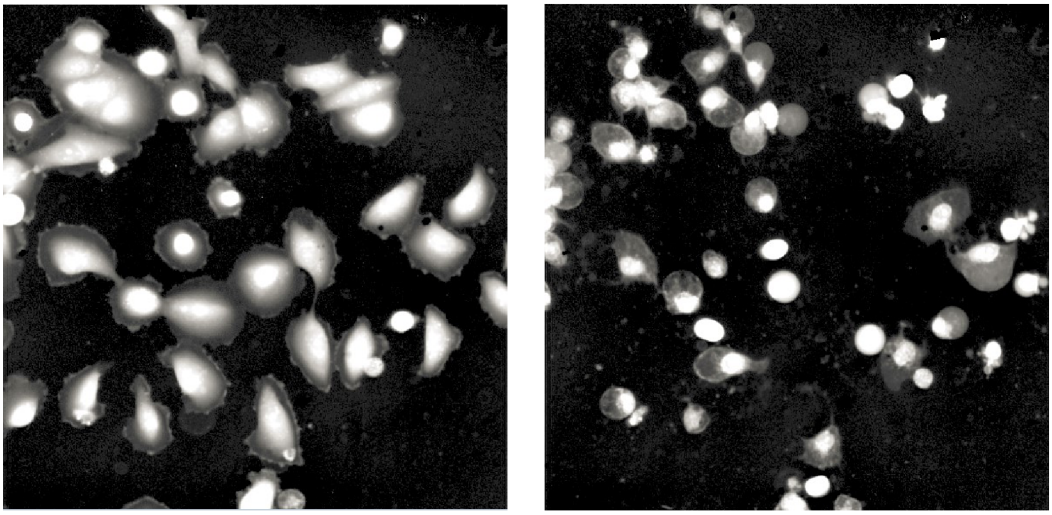
$$Y = XW \tag{1.51}$$

2 PRAKTICKÁ ČÁST

První část praktické části se zaměřuje na extrakci jednotlivých příznaků a zhodnocení jejich využitelnosti. V druhé části jsou na základě těchto příznaků buňky klasifikovány.

2.1 Extrakce příznaků

Zpracovávaná data jsou záznamy buněčného pozorování, obsahující několik set snímků buněk v čase. Tyto záznamy jsou uložené v souboru tif, do něhož je ukládá program holografického mikroskopu. Pro další práci se soubory buněk se vychází z předpokladu, že na začátku pozorování jsou všechny buňky živé, zatímco na konci pozorování jsou všechny buňky mrtvé (obrázek **??**).



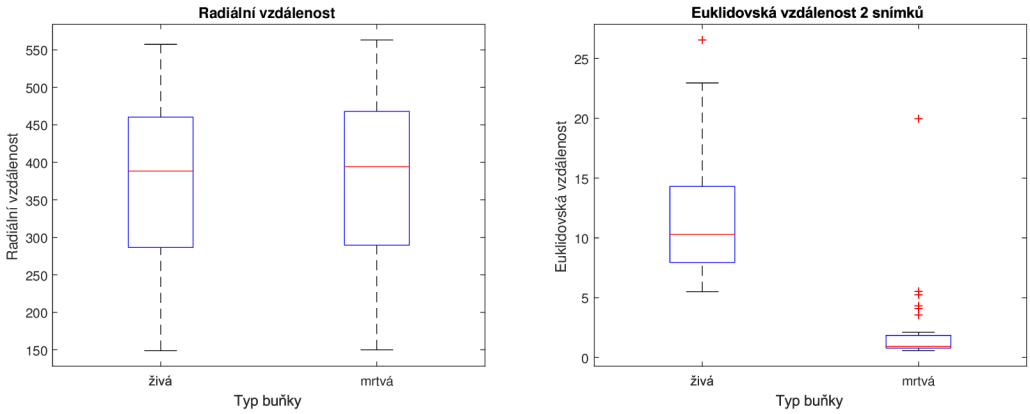
Obr. 2.1: Snímek živých buněk (vlevo) a mrtvých buněk (vpravo)

Pomocí dodaných nasegmentovaných masek se z jednotlivých buněk extrahují příznaky. Příznaky se extrahovaly vždy z prvního a posledního snímku, přičemž příznaky extrahované z prvního snímku klasifikujeme jako příznaky pro živé buňky a příznaky z posledního snímku jako příznaky pro mrtvé buňky.

Mezi extrahovanými příznaky byly: velikost, obvod, kompaktnost, excentricita, průměrná radiální vzdálenost, průměrná intenzita, hmotnost, maximální intenzita, minimální intenzita, směrodatná odchylka intenzity, rychlost pohybu, euklidovská vzdálenost hodnot buňky na 2 po sobě jdoucích snímcích, Fourierovy deskriptory, pravděpodobnost jednotlivých jasových úrovní, maximální hodnota histogramu a pozice, na které se nachází a entropie.

K posouzení využitelnosti jednotlivých příznaků se využije jejich boxplotů, ze kterých je na první pohled zřejmé, pro které příznaky jsou jejich hodnoty odlišné

pro živé a pro mrtvé buňky. Boxploty, neboli krabicové grafy, velmi dobře analyzují hodnoty, jelikož zobrazují jak jejich 1. a 3. kvartil a medián, tak i minimální hodnotu, maximální hodnotu a odlehlé hodnoty. Na obrázku 2.2 jsou zobrazeny boxploty pro 2 vybrané příznaky, radiální vzdálenost a euklidovskou vzdálenost hodnot ve 2 snímcích. Jak je z obrázků patrné, radiální vzdálenost nebude vhodným příznakem, jelikož hodnoty tohoto příznaků jsou velmi podobné pro živé i mrtvé buňky. Naopak euklidovská vzdálenost je hodně vypovídající příznak, protože jak je vidět, veškeré hodnoty se liší, ať už jejich rozsah nebo jednotlivé kvartily. Boxploty pro všechny příznaky jsou obsaženy v příloze A.



Obr. 2.2: Boxplot pro příznaky radiální vzdálenost (vlevo) a euklidovská vzdálenost (vpravo)

Z posouzení všech boxplotů lze vyvodit, že velmi vhodnými příznaky jsou velikost, obvod, hmotnost, rychlost a euklidovská vzdálenost. Naopak radiální vzdálenost, minimální jas a hodnota a pozice maxima histogramu se ukázaly jako nevhodné. Ostatní příznaky jsou na pomezí a je vhodné provést další analýzu dříve, než se rozhodne o jejich použití při klasifikaci.

Boxplotů se nedá využít v případě Fourierových deskriptorů, pravděpodobností jasových úrovní a entropie jelikož se nejedná o jednu hodnotu, ale soubor hodnot. Jejich využitelnost se zjistí experimentálně při samotné klasifikaci - zda jejich použití zvyšuje úspěšnost správné klasifikace či nikoli.

Další metodou pro analýzu využitelnosti jednotlivých příznaků je výpočet Fisherova diskriminantu D_F. Vypočítané hodnoty zobrazuje tabulka 2.1. Čím větší je hodnota Fisherova diskriminantu, tím lépe jsou jednotlivé třídy od sebe oddělitelné a příznak má tak větší váhu. Jako významné se ukázaly velikost, obvod, kompaktnost, hmotnost a euklidovská vzdálenost (v tabulce tučně).

í využitou statistickou metodou pro analýzu byl Mann-Whitneyův U-test. Výsledné p-hodnoty jsou zobrazeny v tabulce 2.2, jednotlivé kvantily v tabulce 2.3.

Tab. 2.1: Analýza příznaků - Fisherův diskriminant

Příznak	$\mu_ž$	$\sigma_ž$	μ_m	σ_m	**DF**
velikost	2857,61	1209,85	1436,93	742,65	**1,002**
obvod	231,34	62,84	147,11	46,60	**1,159**
kompaktnost	19,76	4,54	15,70	2,73	**0,587**
excentricita	0,72	0,17	0,62	0,16	0,174
radiální vzdálenost	375,70	113,26	374,87	115,83	0,000
průměrná intenzita	0,48	0,18	0,64	0,47	0,109
hmotnost	3150,91	1167,26	1949,20	995,44	**0,614**
maximální jas	1,65	0,63	2,05	0,91	0,128
minimální jas	-0,01	0,07	-0,05	0,21	0,023
směrodatná odchylka jasu	0,44	0,18	0,54	0,33	0,080
rychlost	2,53	4,65	0,39	0,49	0,209
euklidovská vzdálenost	11,66	5,02	2,29	3,76	**2,236**
max histogramu - hodnota	0,31	0,11	0,30	0,15	0,002
max histogramu - pozice	2,79	3,01	4,14	7,33	0,029

Nulová hypotéza, která se testovala na hladině významnosti $\alpha = 0.05$, je: *Distribuční funkce hodnot obou tříd (živé a mrtvé buňky) jsou shodné.* Příznaky, mající hodnoty pro které se nulová hypotéza zamítá, jsou velikost, obvod, kompaktnost, excentricita, hmotnost, rychlost a euklidovská vzdálenost. Tyto příznaky mají tedy rozdílné distribuční funkce a je vhodné je použít.

Tab. 2.2: Analýza příznaků - Mann-Whitneyův U-test

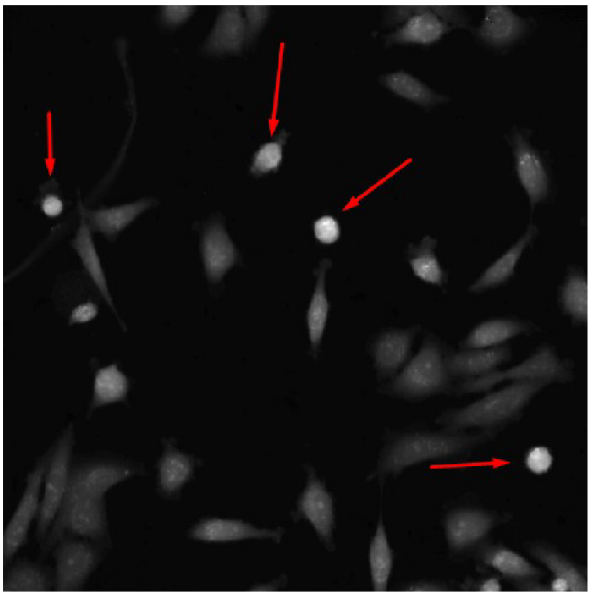
Příznak	**p-hodnota**
velikost	0,18*10-4
obvod	0,05*10-4
kompaktnost	0,18*10-3
excentricita	0,77*10-2
radiální vzdálenost	1,00
průměrná intenzita	0,82
hmotnost	0,14*10-3
maximální jas	0,08
minimální jas	0,48
směrodatná odchylka jasu	0,62
rychlost	**0,02*10-7**
euklidovská vzdálenost	**0,02*10-7**
max histogramu - hodnota	0,75
max histogramu - pozice	0,69

Tab. 2.3: Analýza příznaků - kvantily

Příznak	**Q0,5**$_{ž}$	**Q0,75**$_{ž}$	**Q0,5**$_{m}$	**Q0,75**$_{m}$
velikost	2769,00	3878,00	1305,00	1730,00
obvod	225,13	274,27	130,00	174,31
kompaktnost	18,50	21,41	15,27	17,42
excentricita	0,76	0,85	0,66	0,72
radiální vzdálenost	388,38	460,32	393,98	467,87
průměrná intenzita	0,41	0,54	0,41	0,86
hmotnost	3333,99	3771,27	1687,25	2606,56
maximální jas	1,36	2,01	1,66	2,59
minimální jas	0,01	0,02	0,02	0,03
směrodatná odchylka jasu	0,36	0,58	0,39	0,71
rychlost	1,44	1,73	0,25	0,46
euklidovská vzdálenost	10,29	14,31	0,93	1,85
max histogramu - hodnota	0,32	0,37	0,31	0,42
max histogramu - pozice	2,00	2,00	2,00	2,00

2.2 Analýza buněčných fází

Dalším z cílů bylo analyzovat buněčné fáze pomocí metod pro redukci dimenzionality, konkrétně metodou PCA. K analýze tentokrát byly použity statické snímky buněk. Předmětem zájmu byly mitotické buňky. Jsou to ty buňky, které jsou jasně ohraničené, kulaté, hodně světlé a nemají postupný ztmavující se přechod k okrajům (obrázek 2.3).

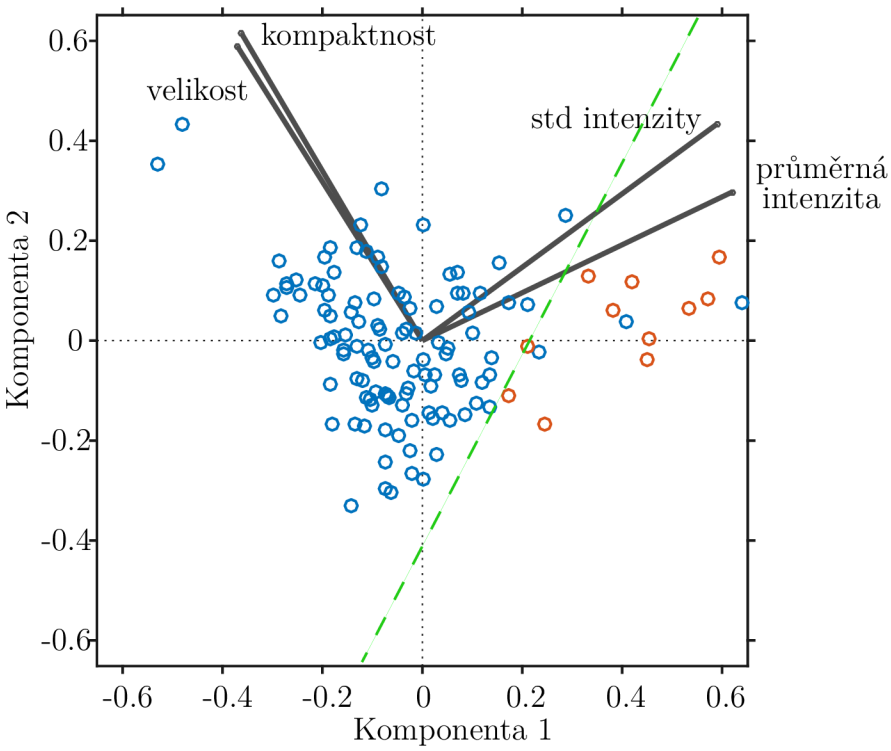


Obr. 2.3: Mitotické buňky

Prvotní analýza byla provedena na souboru buněk ze tří snímků, aby se získal větší počet vzorků. Vzhledem k charakteristice buněk byly pro PCA vybrány příznaky velikost, kompaktnost, průměrná intenzita a směrodatná odchylka intenzity. Výsledkem PCA jsou dvě proměnné, jedna udávající váhy, pro jednotlivé příznaky a příslušné komponenty, a druhá udávající výsledné skóre.

Pro interpretaci se využije vykreslení biplotu. Všechny čtyři příznaky jsou reprezentovány úsečkou. Její délka a směr udávají, jak každý příznak přispívá ke dvěma hlavním komponentám (obrázek 2.4).

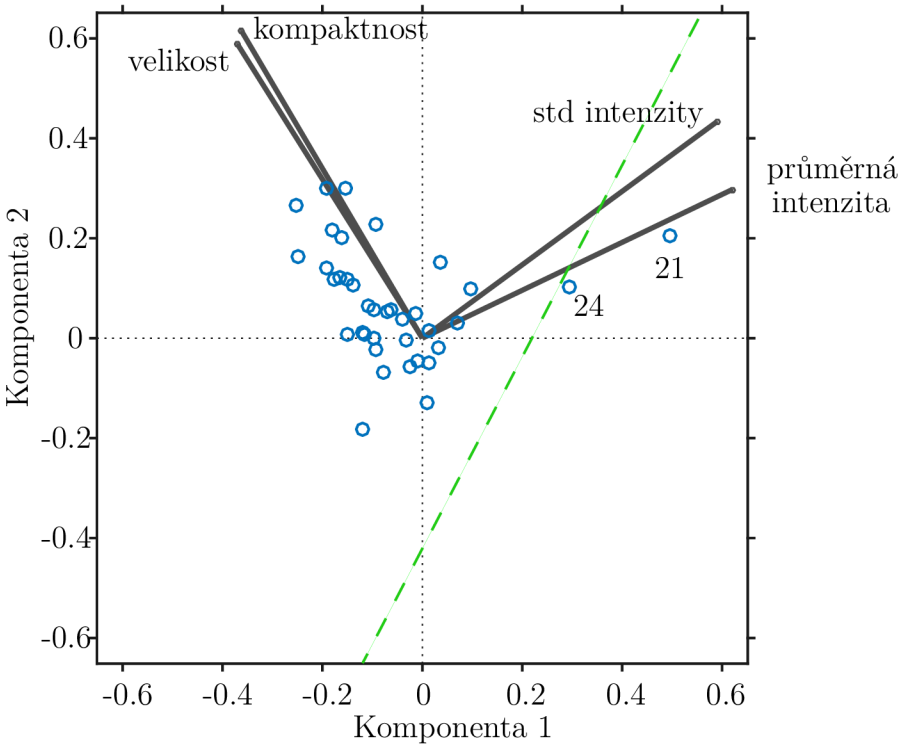
První komponenta je vykreslená na vodorovné ose a má kladné váhy (pravá strana grafu) pro příznaky průměrná intenzita a směrodatná odchylka jasu a záporné váhy (levá strana grafu) pro velikost a kompaktnost. Druhá komponenta je zobrazena na svislé ose a má kladné váhy pro všechny příznaky (horní část grafu). Barevné kroužky zobrazují jednotlivé buňky, mitotické červeně a ostatní modře. Z uvedeného lze vyvodit, že mitotické buňky jsou ty buňky s nejvyššími hodnotami první komponenty. Zelená přerušovaná čára odděluje PCA prostor, přičemž pravá část prostoru obsahuje mitotické buňky, čehož lze využít při klasifikaci.



Obr. 2.4: Vliv příznaků na hlavní komponenty

Bez předchozí znalosti, zda se jedná o mitotické buňky či nikoli, se následně v PCA prostoru vykreslí testovací sady buněk spolu s rozdělující čárou. Z dříve uvedeného vyplývá, že buňky, zobrazené v pravé části prostoru od oddělující zelené čáry, jsou buňky mitotické, což se jednoduše ověří vykreslením jednotlivých buněk.

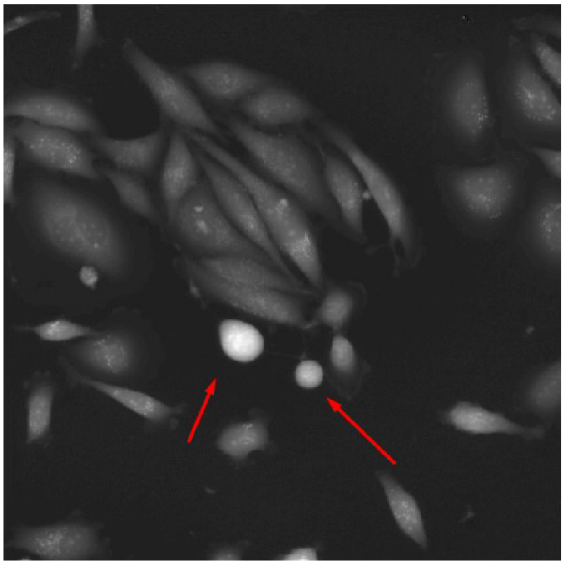
Dle obrázku 2.5, buňky číslo 21 a 24 v první sadě testovacích dat jsou mitotické. Obrázek 2.6 zobrazuje vykreslené samostatné buňky. Buňky jsou ohraničené, kulaté a velmi světlé bez ztmavujícího přechodu k okrajům, tudíž se opravdu jedná o mitotické buňky.



Obr. 2.5: Testovací data 1 v PCA prostoru

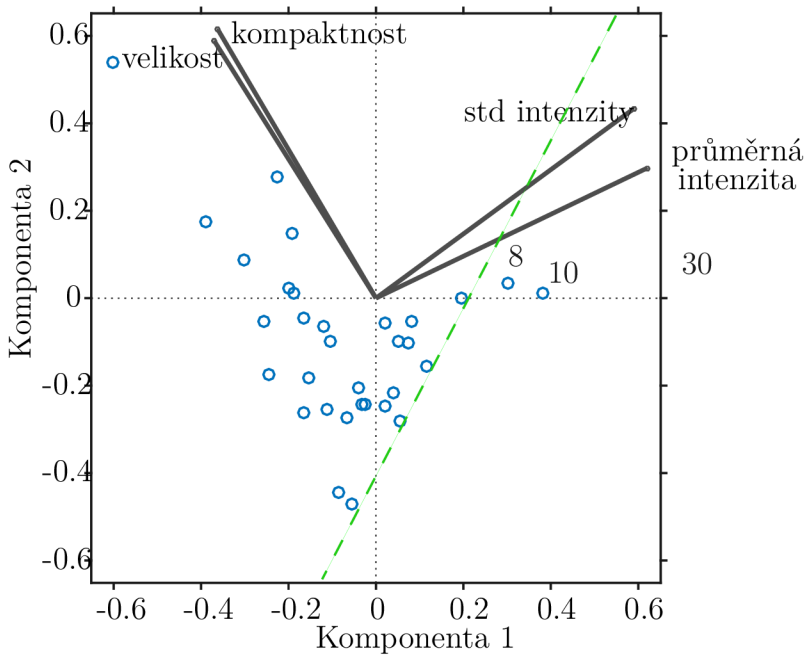
Obr. 2.6: Buňky č. 21 (vlevo) a č. 24 (vpravo) klasifikované jako mitotické

Obrázek 2.7 zobrazuje snímek se všemi buňkami. Mitotické jsou jen ty dvě, které se určili z PCA prostoru, tudíž byla analýza úspěšná.



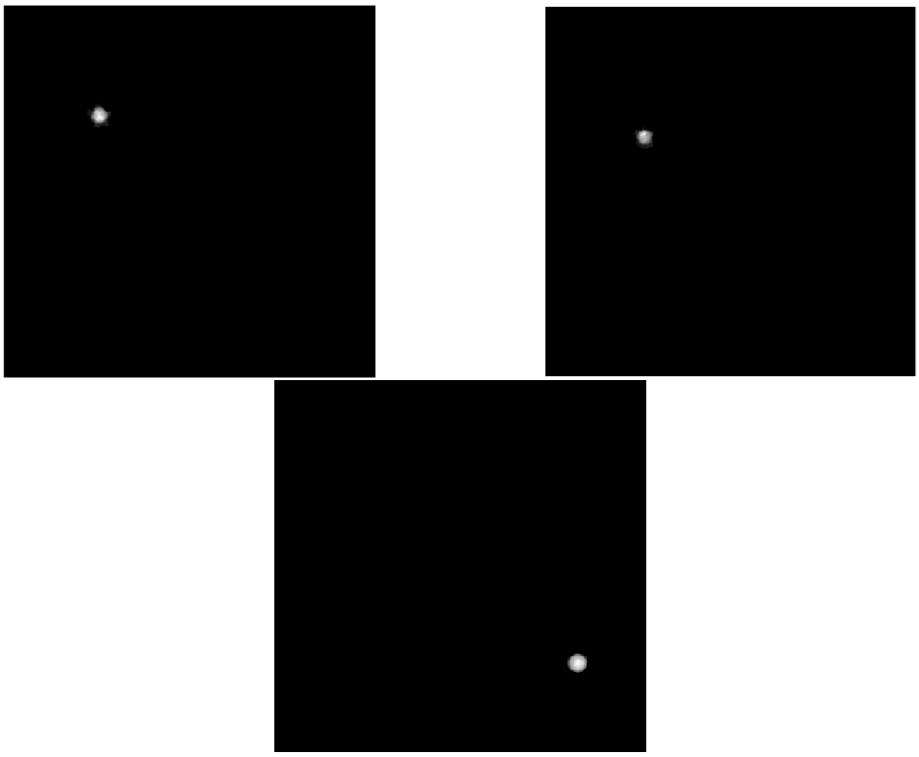
Obr. 2.7: Testovací sada 1

Stejný postup se aplikoval na testovací data 2 a 3. V PCA prostoru jsou buňky číslo 8, 10 a 30 (buňka č. 30 přesahuje měřítko grafu) zobrazeny jako mitotické.



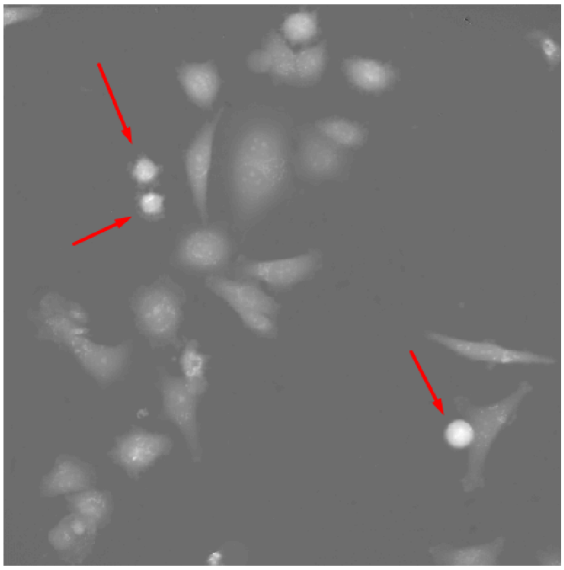
Obr. 2.8: Testovací data 2 v PCA prostoru

Jednotlivé buňky pak vypadají takto:



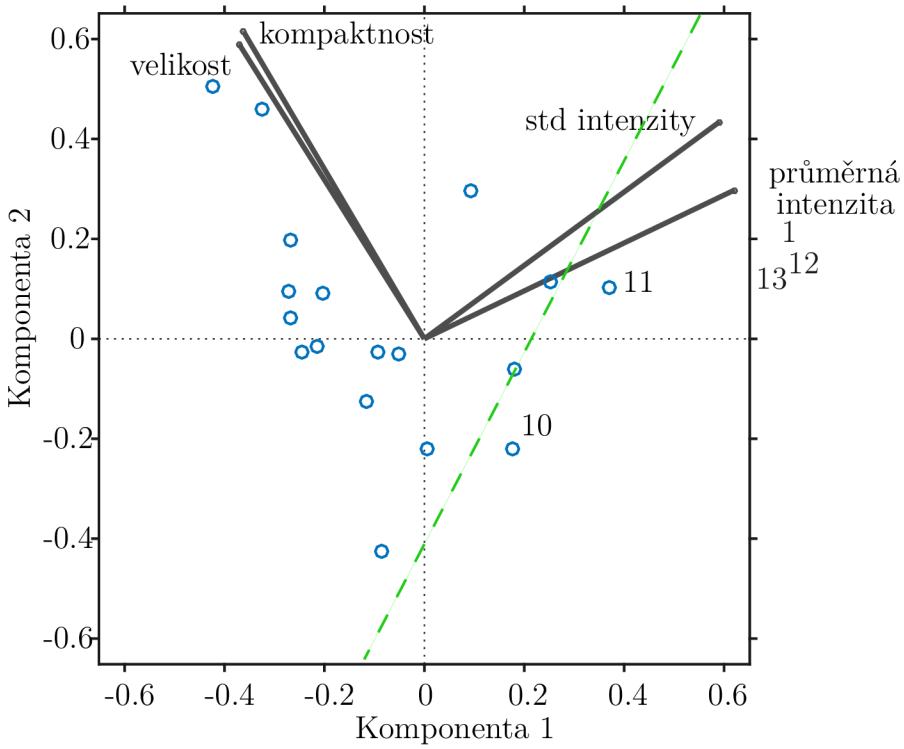
Obr. 2.9: Buňky č. 8 (vlevo nahoře) a č. 10 (vpravo nahoře) a č. 30 (dole uprostřed) klasifikované jako mitotické

Snímek všech buněk (obrázek 2.10) opět potvrzuje správné rozpoznání všech mitotických buněk od ostatních.



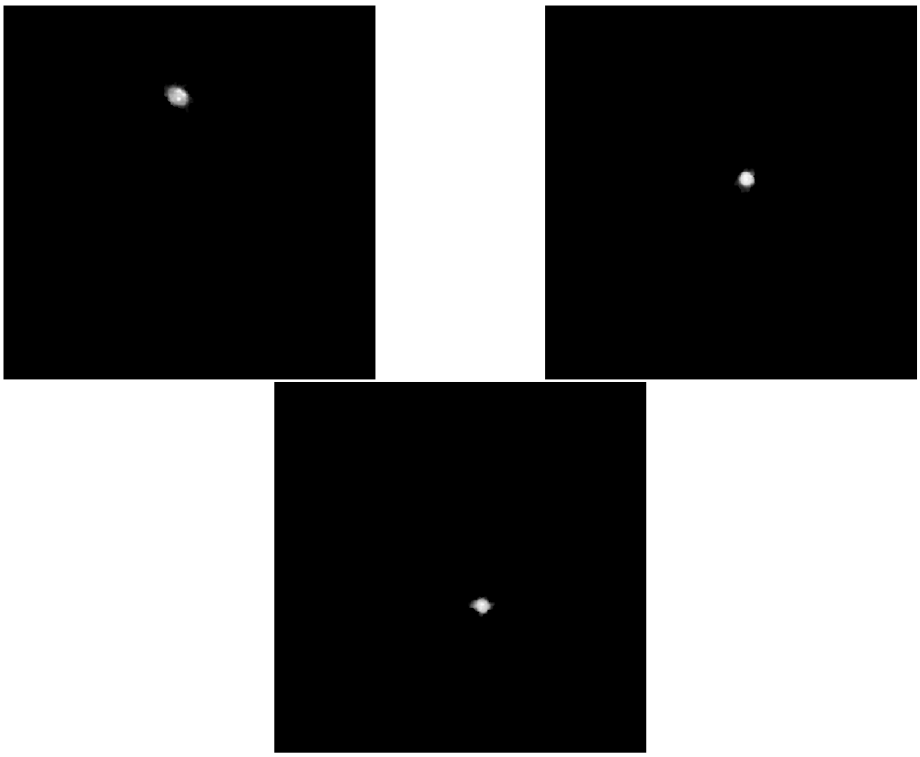
Obr. 2.10: Testovací sada 2

Na obrázku 2.11 je 3. testovací sada buněk v PCA prostoru. Jako mitotické se zařadily buňky č. 1, 10, 11, 12 a 13, které jsou pak vyobrazeny na obrázku 2.12 a 2.13.



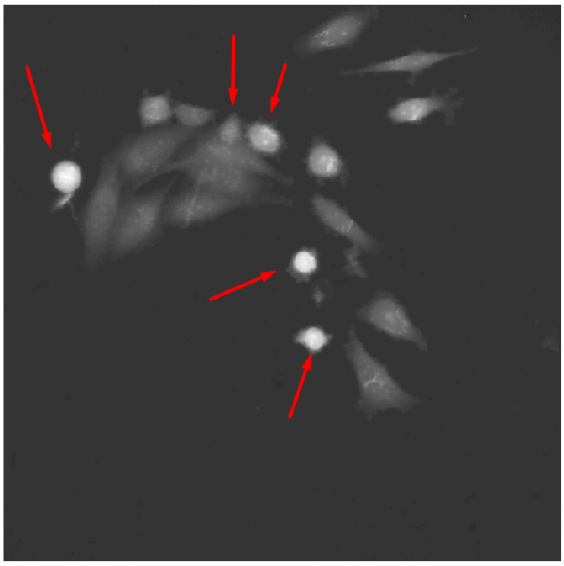
Obr. 2.11: Testovací data 3 v PCA prostoru

Obr. 2.12: Buňky č. 1 a (vlevo nahoře) a č. 10 (vpravo nahoře) klasifikované jako mitotické



Obr. 2.13: Buňky č. 11 (vlevo nahoře) a č. 12 (vpravo nahoře) a č. 13 (dole uprostřed) klasifikované jako mitotické

Ze snímku všech buněk (obrázek 2.14) lze vidět, že buňka č. 10 nebyla dobře klasifikovaná. Při zpětném pohledu do PCA prostoru je vidět, že tato buňka se nejvíce blíží hraniční čáře a zároveň je od ostatních klasifikovaných buněk více ve spodní části prostoru.



Obr. 2.14: Testovací sada 3

2.3 Klasifikace mrtvých buněk

Vybrané příznaky se použily jako trénovací data pro jednoduchý SVM klasifikátor. Tento klasifikátor se prvně otestoval na těch samých datech, které se použily jako trénovací.

V následující tabulce 2.4 je zobrazen počet správně a nesprávně klasifikovaných buněk za použití různých příznaků a různého kernelu pro klasifikátor. V případě použití 5 příznaků se vybraly ty, které měly na základě předchozích analýz největší vypovídající hodnoty: velikost, obvod, hmotnost, rychlost a euklidovskou vzdálenost. Při rozšíření na 7 příznaků se přidaly kompaktnost a excentricita, ke kterým se následně přidaly postupně pravděpodobnost jasu (pst jasu), Fourierovy deskriptory (FD) a entropie, aby se mohla porovnat účinnost klasifikace s využitím těchto příznaků, jelikož jejich aplikovatelnost nebylo možné analyzovat předchozími postupy. Buď se nevyužilo žádné jaderné funkce, takže klasifikátor byl lineární, nebo se použila radiální bázová funkce (rbf) s automatickým nastavením scale a nebo polynomiální funkce 3. či 5. stupně.

Tab. 2.4: Úspěšnost klasifikace buněk na trénovacích datech

Typ SVM	**správně**	**špatně**	**úspěšnost** [%]
všechny příznaky/žádný kernel	29	27	51.8
všechny příznaky/rbf	29	27	51.8
5 příznaků/žádný kernel	55	1	98.2
5 příznaků/rbf	56	0	100
7 příznaků/žádný kernel	54	2	96.4
7 příznaků/rbf	56	0	100
7 příznaků/polynom 3. stupně	56	0	100
7 příznaků/polynom 5. stupně	56	0	100
7 příznaků + pst jasu/žádný kernel	56	0	100
7 příznaků + pst jasu/rbf	56	0	100
7 příznaků + pst jasu/polynom 3. st.	56	0	100
7 příznaků + pst jasu/polynom 5. st.	30	26	53.6
7 příznaků + FD/žádný kernel	56	0	100
7 příznaků + FD/rbf	56	0	100
7 příznaků + FD/polynom 3. st.	56	0	100
7 příznaků + FD/polynom 5. st.	28	28	50
7 příznaků + entropie/žádný kernel	29	27	51.8
7 příznaků + entropie/rbf	29	27	51.8

Za použití všech příznaků klasifikátor nefunguje správně, ani při klasifikaci stejných dat, na kterých byl naučený. Z toho vyplývá, že nemá význam tento model využívat na testovacích datech. Stejný případ nastává za použití 7 vybraných příznaků a entropie a při použití polynomiální funkce 5. stupně u 7 vybraných příznaků a pravděpodobnosti jasu a u 7 vybraných příznaků s Fourierovými deksriptory. Jelikož je úspěšnost klasifikace za použití entropie znatelně nižší, než za použití stejných 7 příznaků bez entropie, lze entropii vyhodnotit jako nevyužitelný příznak.

Ostatní modely klasifikátorů byly použity pro testování na dalších datech. Nejlepší výsledky klasifikace se ukázaly při použití modelu se 7 vybranými příznaky spolu s pravděpodobnostmi jasových úrovní za použití lineárního klasifikátoru. Výsledky klasifikace zobrazuje následující tabulka 2.5.

Tab. 2.5: Úspěšnost klasifikace buněk na testovacích datech

Data	**správně**	**špatně**	**úspěšnost** [%]
data_test1	64	16	80
data_test2	71	13	84.5
data_test3	60	14	81.1
data_test4	74	30	71.2

Celková úspěšnost je 79.2%. Na základě provedených klasifikací lze vyvodit, že příznak pravděpodobnost jasu má pozitivní vliv na klasifikaci, zatímco Fourierovy deskriptory nikoli, jelikož úspěšnost klasifikace snižovaly. Výsledky klasifikace pro ostatní modely klasifikátorů jsou obsaženy v příloze B.

3 ZÁVĚR

Bakalářská práce byla zaměřena na příznaky pro klasifikaci buněčných snímků pořízených holografickým mikroskopem.

V první části byl stručně popsán princip zachycení snímků a jejich postup zpracování. Podrobněji jsou vysvětlené jednotlivé příznaky, jednak radiometrické - popisující metrické vlastnosti buněk, jednak fotometrické - pracující s jasovými hodnotami segmentovaných buněk. Zmíněny jsou dynamické příznaky, které byly využity v praktické části. Základní princip fungování využitého SVM klasifikátoru je popsán a na konci teoretické části jsou shrnuty postupy pro analýzu příznaků.

Druhá, praktická část se zaměřuje na samotnou extrakci příznaků v programovém prostředí Matlab. Příznaky byly extrahovány na dodaných nasegmentovaných buňkách. Následovalo zhodnocení využitelnosti jednotlivých příznaků. Využitelnost se hodnotila na základě boxplotů jednotlivých příznaků a dle hodnot Fisherova diskriminantu a p-hodnoty při neparametrickém testu. Na základě této analýzy se z vybraných příznaků se vytvořil příznakový vektor, který se následně využil pro samotnou klasifikaci buněk, a to jednak mitotických buněk pomocí metod pro redukci dimenzionality, a jednak mrtvých buněk s využitím SVM klasifikátoru.

Mitotické buňky byly úspěšně odděleny v PCA prostoru, jehož hlavní komponenty byly vypočítány z příznaků dobře popisujících tyto buňky - velikost, obvod, průměrná intenzita a směrodatná odchylka intenzity. Jako nejúspěšnější řešení klasifikačního problému mrtvých buněk se ukázala lineární klasifikace při použití příznaků velikost, obvod, kompaktnost, excentricita, hmotnost, rychlost, euklidovská vzdálenost hodnot 2 snímků a pravděpodobnost jasu. Vzhledem k charakteristice dat lze metody považovat za poměrně přesné, nicméně jejich spolehlivost je velmi závislá na přesnosti segmentace.

LITERATURA

[1] HARIHAN, P.: *Basics of holography.* New York, NY: Cambridge University Press, 2002, xii, 161 p. ISBN 05-210-0200-1.

[2] YU, Xiao, Jisoo HONG, Changgeng LIU a Myung K. KIM. Review of digital holographic microscopy for three-dimensional profiling and tracking. *Optical Engineering.* 2014-11-01, vol. 53, issue 11, s. 112306-. DOI: 10.1117/1.OE.53.11.112306. Dostupné z URL: <http://opticalengineering.spiedigitallibrary.org/article.aspx?doi=10.1117/1.OE.53.11.112306>

[3] SLABÝ, Tomáš, Pavel KOLMAN, Zbyněk DOSTÁL, Martin ANTOŠ, Martin LOŠŤÁK a Radim CHMELÍK. Off-axis setup taking full advantage of incoherent illumination in coherence-controlled holographic microscope. *Optics Express.* 2013, vol. 21, issue 12, s. 14747-. DOI: 10.1364/OE.21.014747. Dostupné z URL: <https://www.osapublishing.org/oe/abstract.cfm?uri=oe-21-12-14747>

[4] KOLMAN, Pavel, Radim CHMELÍK. Coherence-controlled holographic microscope. *Optics Express.* 2010, vol. 18, issue 21, s. 21990-. DOI: 10.1364/OE.18.021990. Dostupné z URL: <https://www.osapublishing.org/oe/abstract.cfm?uri=oe-18-21-21990>

[5] SONKA, M.: *Image processing, analysis, and machine vision.* 3rd ed. Toronto: Thompson Learning, c2008. ISBN 978-0495082521.

[6] BANKMAN, I. H.: *Handbook of Medical Image Processing and Analysis.* 2nd ed. San Diego: Academic Press, 2009, s. 261-277. ISBN 978-0-12-373904-9.

[7] WANG, P. et al. Feature-based analysis of cell nuclei structure for classification of histopathological images. *Digital Signal Processing.* 2018, vol. 78, s. 152-162. DOI: <https://doi.org/10.1016/j.dsp.2018.03.008>

[8] MEIJERING, E. et al. Methods for Cell and Particle Tracking. Imaging and Spectroscopic Analysis of Living Cells. 2012, vol. 504, s. 183-200.

[9] BURGES, Christoper J. C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery.* 1998, 2(2):121-167. Dostupné z URL: <http://cmp.felk.cvut.cz/cmp/courses/recognition/Labs/svm2/SVM_tutorial-Burges98.pdf>

[10] CORTES C., VAPNIK V. Support-Vector networks. *Machine Learning.* 1995, 20: 273-297.

[11] SCHÖLKOPS B., SMOLA A.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. Cambridge, MA: MIT Press, 2002, s. 648 . ISBN 0-262-19475-9.

[12] LIU Song, Piyushkumar A. MUNDA, Jagath C. RAJAPAKSE. Features for Cells and Nuclei Classification. *33rd Annual International Conference of the IEEE EMBS*. Boston, Massachusetts USA, 2011.

[13] LORENZO-GINORI J.V. et al. Cervical Cell Classification Using Features Related to Morphometry and Texture of Nuclei. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. CIARP 2013. Lecture Notes in Computer Science, vol 8259. Berlin: Springer, s. 222-229.

[14] Support vector machines (SVM), IS MUNI. Dostupné z URL: <`https://is.muni.cz/el/1433/podzim2006/PA034/09_SVM.pdf`>

[15] HANYÁŠOVÁ, L. *Metody texturní analýzy v medicínských obrazech*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. s. 33-34. Dostupné z URL: <`https://dspace.vutbr.cz/bitstream/handle/11012/4016/diplomka.pdf?sequence=1`>

[16] KUBANOVÁ, J.: *Statistické metody pro ekonomickou a technickou praxi*. Bratislava: Statis, 2008 (3). s. 247. ISBN 978-80-85659-47-4.

[17] ILIN Alexander, TAPANI Raiko: Practical Approaches to Principal Component Analysis in the Presence of Missing Values. Journal of Machine Learning Research. 11, 2010. s. 1957-2000.

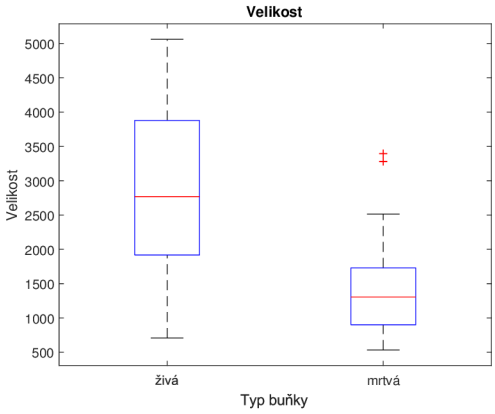
[18] SMITH Lindsay I.: A Tutorial on Principal Component Analysis. 2002. Dostupné z URL: <`http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf`>

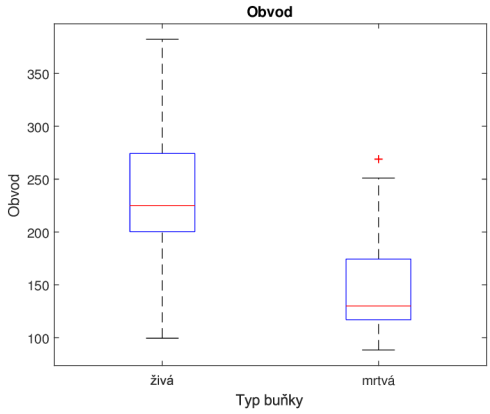
SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

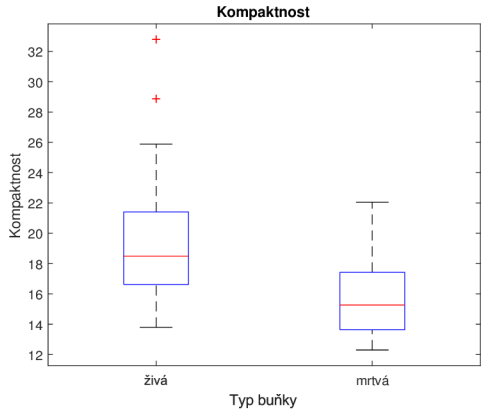
DHM	digitální holografická mikroskopie
CCHM	koherencí řízený holografický mikroskop(z anglického *Coherence-Controlled Holographic Microscope*
k-NN	algoritmus k nejbližších sousedů (z anglického *k-nearest neighbours*
SVs	podpůrné vektory (z anglického *Support Vectores*
SVM	support vector machines
rbf	radiální bázová funkce
PCA	analýza hlavních komponent (z anglického *Principal Component Analysis*
FD	Fourierovy deskriptory
pst jasu	pravděpodobnost jasu
C	kompaktnost
A	obsah plochy objektu
P	obvod objektu
m_{pq}	lokální moment p+q řádu
μ_{pq}	centrovaný moment p+q řádu
Ω	objekt
$h(z)$	histogram
L_P	primární Langrangeova funkce
L_D	duální Langrangeova funkce
α_i	Langrangeovy multiplikátory

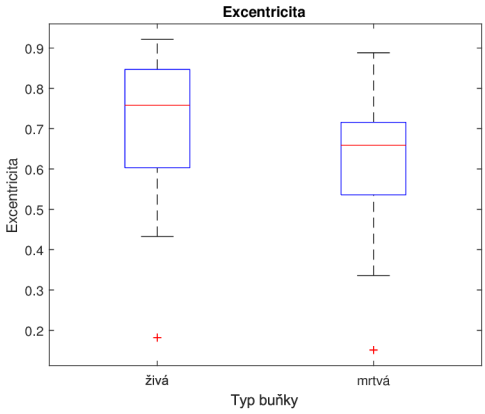
A GRAFICKÁ VIZUALIZACE DAT

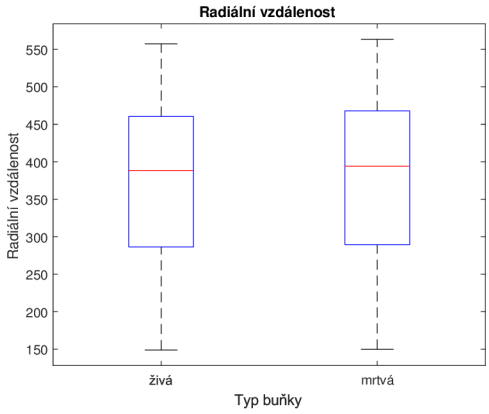
Boxploty s hodnotami pro jednotlivé příznaky

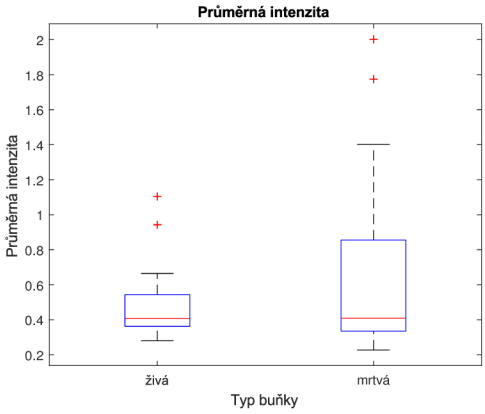


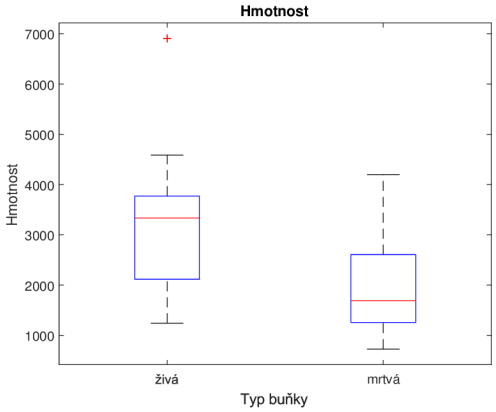


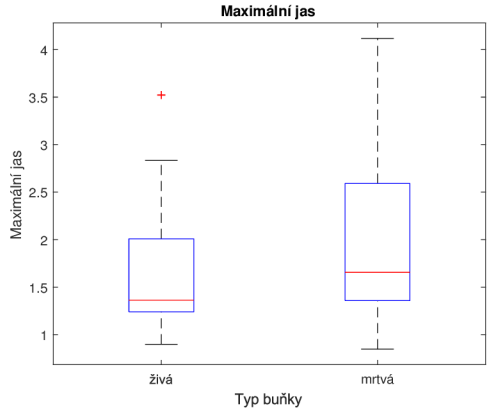


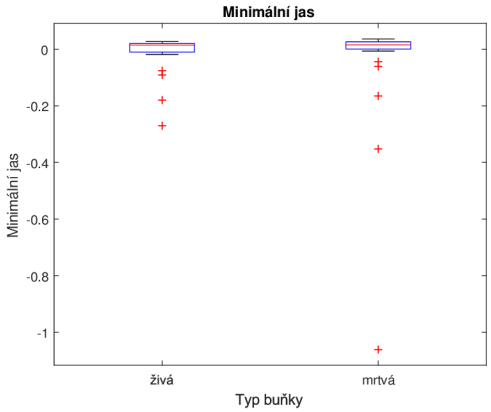


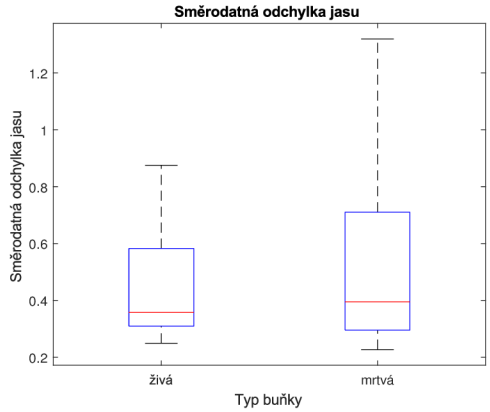


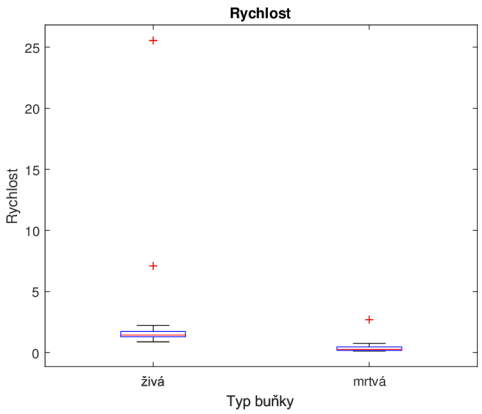


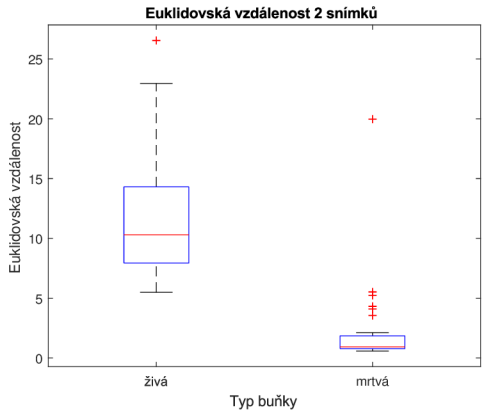

Hmotnost
Hmotnost
7000
6000
5000
4000
3000
2000
1000
živá
mrtvá
Typ buňky

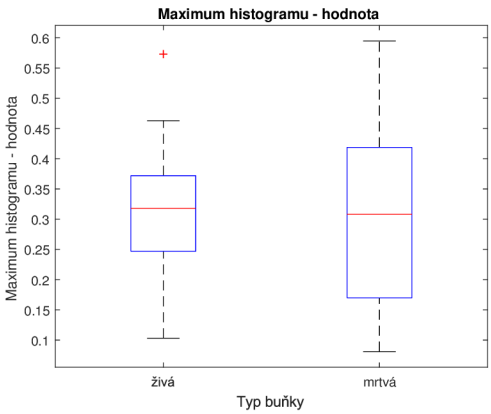

Maximální jas
Maximální jas
4
3.5
3
2.5
2
1.5
1
živá
mrtvá
Typ buňky

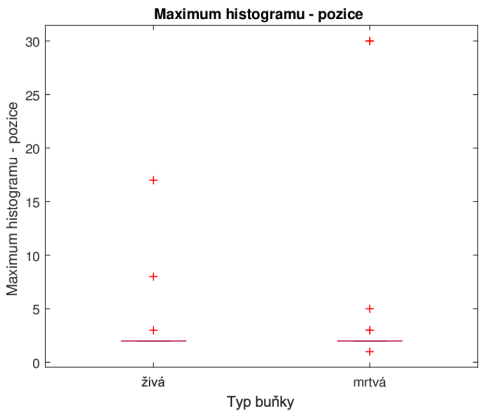

Minimální jas
Minimální jas
0
-0.2
-0.4
-0.6
-0.8
-1
živá
mrtvá
Typ buňky


Směrodatná odchylka jasu
Směrodatná odchylka jasu
1.2
1
0.8
0.6
0.4
0.2
živá
mrtvá
Typ buňky


Rychlost
Rychlost
25
20
15
10
5
0
živá
mrtvá
Typ buňky


Euklidovská vzdálenost 2 snímků
Euklidovská vzdálenost
25
20
15
10
5
0
živá
mrtvá
Typ buňky


Maximum histogramu - hodnota
Maximum histogramu - hodnota
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
živá
mrtvá
Typ buňky


Maximum histogramu - pozice
Maximum histogramu - pozice
30
25
20
15
10
5
0
živá
mrtvá
Typ buňky

B VÝSLEDKY KLASIFIKACE

Tab. B.1: Úspěšnost klasifikace buněk na testovacích datech 1 a 2 za použití různých modelů klasifikátoru

data_test1	**správně**	**špatně**	**úspěšnost** [%]
5 příznaků/žádný kernel	50	30	62.5
5 příznaků/rbf	54	26	67.5
7 příznaků/žádný kernel	53	27	66.3
7 příznaků/rbf	53	27	66.3
7 příznaků + pst jasu/žádný kernel	64	16	80
7 příznaků + pst jasu/rbf	42	38	52.5
7 příznaků + FD/žádný kernel	56	24	70
7 příznaků + FD/rbf	42	38	52.5
7 příznaků/polynom 3. stupně	55	25	68.8
7 příznaků/polynom 5. stupně	57	23	71.3
7 příznaků + pst jasu/polynom 3. st.	61	19	76.3
7 příznaků + FD/polynom 3. st.	52	28	65
data_test2	**správně**	**špatně**	**úspěšnost** [%]
5 příznaků/žádný kernel	53	31	63.1
5 příznaků/rbf	63	21	75
7 příznaků/žádný kernel	53	31	63.1
7 příznaků/rbf	60	24	71.4
7 příznaků + pst jasu/žádný kernel	71	13	84.5
7 příznaků + pst jasu/rbf	48	36	57.1
7 příznaků + FD/žádný kernel	61	23	72.6
7 příznaků + FD/rbf	43	41	51.2
7 příznaků/polynom 3. stupně	67	17	79.8
7 příznaků/polynom 5. stupně	66	18	78.6
7 příznaků + pst jasu/polynom 3. st.	63	21	75
7 příznaků + FD/polynom 3. st.	55	29	65.5

Tab. B.2: Úspěšnost klasifikace buněk na testovacích datech 3 a 4 za použití různých modelů klasifikátoru

data_test3	**správně**	**špatně**	**úspěšnost** [%]
5 příznaků/žádný kernel	45	29	60.8
5 příznaků/rbf	59	15	79.7
7 příznaků/žádný kernel	50	24	67.6
7 příznaků/rbf	52	22	70.3
7 příznaků + pst jasu/žádný kernel	60	14	81.1
7 příznaků + pst jasu/rbf	38	36	51.4
7 příznaků + FD/žádný kernel	51	23	68.9
7 příznaků + FD/rbf	39	35	52.7
7 příznaků/polynom 3. stupně	62	12	83.8
7 příznaků/polynom 5. stupně	61	13	82.4
7 příznaků + pst jasu/polynom 3. st.	58	16	78.4
7 příznaků + FD/polynom 3. st.	45	29	60.8
data_test4	**správně**	**špatně**	**úspěšnost** [%]
5 příznaků/žádný kernel	59	45	56.7
5 příznaků/rbf	65	39	62.5
7 příznaků/žádný kernel	59	45	56.7
7 příznaků/rbf	52	52	50
7 příznaků + pst jasu/žádný kernel	74	30	71.2
7 příznaků + pst jasu/rbf	54	50	51.9
7 příznaků + FD/žádný kernel	76	28	73.1
7 příznaků + FD/rbf	52	52	50
7 příznaků/polynom 3. stupně	61	43	58.7
7 příznaků/polynom 5. stupně	61	43	58.7
7 příznaků + pst jasu/polynom 3. st.	75	29	72.1
7 příznaků + FD/polynom 3. st.	59	45	56.7

C SEZNAM SKRIPTŮ A FUNKCÍ

bakalarska_prace - hlavní skript obsahující extrakci příznaků a klasifikaci buněk
PCA - skript pro analýzu buněčných typů pomocí PCA
biplotG - funkce pro vykreslení komponent PCA
imshow4 - funkce pro zobrazení 2D, 3D nebo 4D snímků
velikost - funkce pro výpočet velikosti
obvod - funkce pro výpočet obvodu
kompaktnost - funkce pro výpočet kompaktnosti
excentricita - funkce pro výpočet excentricity
rad_vzdalenost - funkce pro výpočet radiální vzdáleností
intenzita - funkce pro výpočet průměrné intenzity a hmotnosti
jas - funkce pro výpočet maximální intenzity, minimální intenzity a směrodatné odchylky intenzity
rychlost - funkce pro výpočet rychlosti
euklidovska_vzdalenost_hodnot - funkce pro výpočet euklidovské vzdálenosti hodnot ve 2 po sobě jdoucích snímcích
centerobject - funkce pro vycentrování objektu doprostřed snímku (nezbytné pro výpočet Fourierových deskriptorů)
gfd - funkce pro výpočet Fourierových deskriptorů
pravdepodobnost_jasu - funkce pro výpočet pravděpodobnosti jednotlivých jasových úrovní
deskriptory_histogramu - funkce pro výpočet deskriptorů histogramu (maximální hodnota a její pozice, entropie)

D OBSAH PŘILOŽENÉHO CD

marketa_navratilova_BP.pdf - elektronická verze bakalářské práce

skripty_a_funkce - složka obsahující skripty a funkce pro extrakci příznaků a klasifikaci

data_SVM - složka obsahující ukázková data pro extrakci příznaků a klasifikaci mrtvých buněk

data_zaznam - ukázkový záznam snímků buněk zobrazující prvních 5 a posledních 5 snímků z původního záznamu

data_zaznam_maska.mat - nasegmentovaná maska pro ukázkový záznam buněk

SVMModel_pst_jasu_linear.mat - vybraný SVM klasifikační model

data_PCA - složka obsahující ukázková data pro klasifikaci mitotických buněk pomocí PCA

data_stat.tif - ukázkový snímek buněk

data_stat_maska.png - nasegmentovaná maska pro ukázkový snímek buněk

loadings.mat - váhy pro výpočet nového skóre z původních příznaků

prumer_x.mat - průměr příznaku x pro normalizaci

std_x.mat - směrodatná odchylka příznaku x pro normalizaci