

ŽILINSKÁ UNIVERZITA
V ŽILINE

Fakulta riadenia a informatiky

Spracovanie a vyhodnocovanie grafických biomedicínskych údajov

Diplomová práca

Bc. LINDA BLAHOVÁ

Študijný program: Biomedicínska informatika

Študijný odbor: Informatika

Školiace pracovisko: Žilinská univerzita v Žiline,

Vedúci diplomovej práce: doc. Ing. Jozef Kostolný, PhD.

Žilina 2023

ZADANIE TÉMY DIPLOMOVEJ PRÁCE.

Študijný program: Biomedicínska informatika

Meno a priezvisko

Linda Blahová

Osobné číslo

560651

Názov práce v slovenskom aj anglickom jazyku

Spracovanie a vyhodnocovanie grafických biomedicínskych údajov

Processing and evaluating graphical biomedical data

Zadanie úlohy, ciele, pokyny pre vypracovanie

(Ak je málo miesta, použite opačnú stranu)

Cieľ diplomovej práce:

Cieľom práce je spracovanie prehľadu, implementácia a vyhodnotenie spoľahlivosti algoritmov na analýzu grafických údajov biomedicínskeho charakteru. Pri riešení je potrebné vybrať typ údajov, ktoré sa budú vo vybraných algoritmoch analyzovať. Nasledne sa zvolia vhodné algoritmy na implementáciu alebo knižnice, ktoré je možné použiť. Výsledkom je vyhodnotenie spoľahlivosti a vhodnosti použitia na identifikáciu a analýzu vybraného typu údajov.

Obsah:


Obsah práce:

- analýza dostupných algoritmov, knižníc pre jednotlivé typy údajov
- výber vhodných algoritmov a techník pre zvolený typ údajov
- implementácia a vyhodnotenie spoľahlivosti analýza
- overenie a validácia vybraného algoritmu

Meno a pracovisko vedúceho DP: doc. Ing. Jozef Kostolný, PhD., KI, ŽU

Meno a pracovisko tútora DP:

31 OKT. 2022


vedúci katedry
(dátum a podpis)

Čestné vyhlásenie

Vyhlasujem, že som zadanú diplomovú prácu vypracovala samostatne, pod odborným vedením vedúceho práce a používala som len literatúru uvedenú v práci.

Žilina 1. apríla 2023

Bc. Linda Blahová

Pod'akovanie

Rada by som vyjadrila úprimnú vďaku doc. Ing. Jozefovi Kostolnému, PhD. za veľkú trpezlivosť, strávený čas, ochotný prístup a cennú pomoc pri vedení tejto diplomovej práce. Vaše skúsenosti a konštruktívne nápady mi pomohli úspešne sa vyrovať s náročnosťou práce a Vaša spolupráca bola pre mňa prínosom, a priviedla ma k vytvoreniu kvalitnej a dôkladnej práce. Taktiež by som chcela za neustálu podporu, nekonečnú trpezlivosť a láskavú pomoc nie len pri vypracovaní práce, ale aj počas celého štúdia poďakovať mojej rodine a priateľom.

Abstrakt

BLAHOVÁ, Linda: Spracovanie a vyhodnocovanie grafických biomedicínskych údajov. [Diplomová práca] / Linda Blahová. – Žilinská univerzita v Žiline. Fakulta riadenia a informatiky; Katedra informatiky. – Školiteľ: doc. Ing. Jozef Kostolný, PhD.. Žilina: FRI UNIZA, 2023.

Práca sa zaoberá spracovaním obrazových biomedicínskych údajov, konkrétne mikroskopických údajov zachytávajúcich krvnú vzorku, kde je potrebné identifikovať a klasifikovať leukocyty a ich počet. Manuálne spracovanie takýchto údajov je časovo náročné a môže byť subjektívne a náchylné na ľudské chyby. Cieľom práce je preto analýza súčasných techník spracovania a dostupných nástrojov, ktoré môžu tento proces zjednodušiť a automatizovať. Práca obsahuje prehľad základných typov biomedicínskych údajov, analýzu dostupných údajov a techník na spracovanie obrazových údajov, návrh riešenia použitia dostupných techník a celkovej automatizácie procesu, implementáciu riešenia a vyhodnotenie jeho prínosov v ďalšom výskume. Výsledkom práce je riešenie, ktoré zjednoduší a zrýchli spracovanie mikroskopických údajov a zabezpečí presnejšie a spoľahlivejšie výsledky identifikácie a klasifikácie leukocytov. Toto riešenie môže nájsť uplatnenie v diagnostike rôznych ochorení, ako sú rakovina krvi, infekcie, autoimunitné ochorenia alebo alergie.

Kľúčové slová:

Leukocyty, mikroskopické obrazové údaje, segmentácia, strojové učenie, klasifikácia, ImageJ, CellProfiler, Java, Python

Abstract

The thesis deals with the processing of biomedical image data, specifically microscopic data capturing a blood sample, where leukocytes and their count need to be identified and classified. Manual processing of such data is time consuming and can be subjective and prone to human error. Therefore, the aim of this work is to analyse current processing techniques and available tools that can simplify and automate this process. The thesis includes an overview of the basic types of biomedical data, an analysis of the available data and techniques for image data processing, a proposal for a solution using the available techniques and overall automation of the process, implementation of the solution, and an evaluation of its benefits in further research. The work results in a solution that simplifies and speeds up the processing of microscopic data and provides more accurate and reliable results for leukocyte identification and classification. This solution may find application in the diagnosis of various diseases such as blood cancers, infections, autoimmune diseases, or allergies.

Keywords: Leukocytes, microscopic image data, segmentation, machine learning, classification, ImageJ, CellProfiler, Java, Python

Obsah

Úvod.....	14
1 KLASIFIKÁCIA BIOMEDICÍNSKYCH ÚDAJOV.....	15
1.1 Všeobecný prehľad biomedicínskych údajov.....	15
1.1.1 Genomické údaje.....	15
1.1.2 Proteomické údaje.....	16
1.1.3 Metabolomické údaje.....	16
1.1.4 Obrazové údaje.....	16
1.1.5 Klinické údaje.....	17
1.1.6 Údaje o správaní.....	17
1.1.7 Údaje o životnom prostredí.....	17
1.1.8 Fyziologické údaje alebo údaje o biosignáloch.....	18
1.1.9 Textové údaje.....	18
1.2 Klasifikácia obrazových biomedicínskych údajov.....	20
1.2.1 Magnetická rezonancia (MRI).....	20
1.2.2 Počítačová tomografia (CT).....	21
1.2.3 Pozitrónová emisná tomografia (PET).....	23
1.2.4 Ultrasonografia (USG).....	24
1.2.5 Rádiografia.....	26
1.2.6 Fotomikroskopia buniek.....	27
1.2.7 Záver.....	29
2 ANALÝZA PROBLEMATIKY.....	30
2.1 Analýza údajov.....	31
2.1.1 Analýza údajov zo softvérového aspektu.....	31
2.1.2 Analýza údajov z aspektu mikroskopickej anatómie (histológie).....	32
2.1.3 Spôsoby identifikácie buniek.....	33
2.2 Prehľad techník spracovania a analýzy obrazových údajov.....	34
2.2.1 Manuálna optická analýza.....	35
2.2.2 Softvérové techniky analýzy snímok.....	35
2.2.2.1 Prahovanie (Thresholding).....	35
2.2.2.2 Detekcia hrán (Edge detection).....	36
2.2.2.3 Watershed.....	36
2.2.2.4 Grafové rezy (Graph cuts).....	37
2.2.2.5 Extrakcia kontúr (Contour extraction).....	38
2.2.2.6 Logistická regresia (Logistic Regression).....	38
2.2.2.7 Stroje s podpornými vektormi (Support vector machines).....	39

2.2.2.8	Neurónové siete (Neural networks).....	40
2.3	Vyhodnotenie analýzy	40
3	NÁVRH SYSTÉMU PRE ANALÝZU LEUKOCYTOV	42
3.1	Návrh všeobecného procesu spracovania údajov	42
3.2	Návrh výberu a pedspracovania údajov.....	43
3.3	Návrh procesu spracovania a analýzy údajov.....	44
3.4	Výber prístupov spracovania údajov	45
3.4.1	Obrazové spracovanie.....	48
3.4.2	Spracovanie pomocou strojového učenia.....	48
4	IMPLEMENTÁCIA NAVRHNUTÉHO SYSTÉMU.....	50
4.1	Výber a pedspracovanie údajov.....	50
4.2	Konfigurácie metód spracovania.....	52
4.2.1	Spracovanie pomocou nástroja ImageJ	52
4.2.2	Spracovanie pomocou nástroja CellProfiler	55
4.2.2.1	Predpríprava vstupných údajov do nástroja CellProfiler	55
4.2.2.2	Použitie nástroja CellProfiler	56
4.3	Implementácia skriptov pre použitie metód spracovania	57
4.3.1	Skript pre automatické obrazové spracovanie s použitím nástroja ImageJ.....	58
4.3.1.1	Výsledky dosiahnuté obrazovým spracovaním	61
4.3.2	Skript pre automatické spracovanie strojovým učením s použitím nástroja CellProfiler	62
4.3.2.1	Výsledky spracovaním strojovým učením	66
5	OVERENIE SYSTÉMU PRE ANALÝZU LEUKOCYTOV	69
5.1	Spôsob vyhodnocovania implementovaných metód	69
5.2	Vyhodnotenie aplikovaných metód	69
5.2.1	Obrazové spracovanie snímok.....	69
5.2.1.1	Prvý experiment – Použitie techniky na jednom albume množiny údajov.....	69
5.2.1.2	Druhý experiment – Použitie techniky na celej množine údajov	72
5.2.1.3	Vyhodnotenie experimentov.....	74
5.2.2	Spracovanie snímok strojovým učením.....	75
5.2.2.1	Prvý experiment – Použitie techniky na množine údajov Raabin-WBC	75
5.2.2.2	Druhý experiment – Použitie techniky na množine údajov BCCD.....	76
5.2.2.3	Vyhodnotenie experimentov.....	77

5.3 Prínos vyhodnocovania v ďalšom výskume	78
Záver	80

Zoznam obrázkov

Obrázok 1 - Sekvencia MRI snímok mozgu [28].....	20
Obrázok 2 - Sekvencia CT snímok mozgu [32].....	21
Obrázok 3 - Sekvencia PET snímok mozgu [36]	23
Obrázok 4 - Prenatálne USG snímky nervového systému plodu [41].....	24
Obrázok 5 - Rádiografická snímka krčnej chrbtice [47].....	26
Obrázok 6 - Tkanivo jazvy po 10 násobnom zväčšení [53].....	27
Obrázok 7 - Typy bielych krviniek [58]	32
Obrázok 8 - Diagram všeobecného procesu spracovania údajov.....	42
Obrázok 9 - Diagram označení buniek v množine údajov Raabin-WBC [102].....	43
Obrázok 10 - Diagram návrhu procesu spracovania a analýzy údajov	44
Obrázok 11 - Diagram prípadov použitia navrhovaného systému	47
Obrázok 12 - Vzájomná interakcia systémových komponentov s používateľmi	50
Obrázok 13 - Ukážka pôvodnej snímky (vľavo) a predspracovanej snímky (vpravo) ...	51
Obrázok 14 - Sekvencia operácií použitá pre predspracovanie snímok	52
Obrázok 15 - Ukážka makra použitého pre nástroj ImageJ	53
Obrázok 16 - Ukážka postupného spracovania snímky v nástroji ImageJ.....	54
Obrázok 17 - Ukážka JSON Anotácie a príslušnej snímky	55
Obrázok 18 - Ukážka použitého toku príkazov pre extrakciu vlastností zo snímok	56
Obrázok 19 - Ukážka výsledkov jednotlivých krokov spracovania snímky v nástroji CellProfiler.....	57
Obrázok 20 - Diagram aktivít použitia systému pre obrazovú analýzu	58
Obrázok 21 - Ukážka identifikovaných oblastí nástrojom ImageJ.....	61
Obrázok 22 - Ukážka výstupného CSV súboru z programu CellProfiler.....	62
Obrázok 23 - Ukážka výstupného grafu a CSV súboru pre vyhodnotenie spoľahlivosti identifikácie	62
Obrázok 24 - Diagram aktivít použitia systému pre analýzu strojovým učením	63
Obrázok 25 - Ukážka orezaných obrázkov leukocytov z pôvodných snímok	66
Obrázok 26 - Ukážka výstupného CSV súboru z programu CellProfiler.....	67
Obrázok 27 – Ukážka výsledkov vyhodnotenia presnosti modelu	67
Obrázok 28 - Ukážka CSV súboru označenej a predikovanej kategórie leukocytu v testovacej množine pre 3. replikáciu	68
Obrázok 29 - Porovnanie označených (vľavo) a identifikovaných buniek (vpravo).....	69
Obrázok 30 - Experiment 1 - Počet leukocytov v kategóriách	70
Obrázok 31 - Experiment 1 - Percentuálne vyjadrenie počtu leukocytov v kategóriách71	

Obrázok 32 - Experiment 1 - Percentuálne vyjadrenie pomeru Identifikovaných ku všetkým označeným leukocytom	71
Obrázok 33 - Experiment 2 - Počet leukocytov v kategóriách	72
Obrázok 34 - Experiment 2 - Percentuálne vyjadrenie počtu leukocytov v kategóriách	73
Obrázok 35 - Experiment 2 - Percentuálne vyjadrenie pomeru Identifikovaných ku všetkým označeným leukocytom	73
Obrázok 36 - Ukážka variability snímok celej množiny údajov	74
Obrázok 37 - Ukážka identifikovaných oblastí	75
Obrázok 38 - Experiment 1 - Presnosť modelu so vstupnými údajmi Raabin-WBC	76
Obrázok 39 - Experiment 2 - Presnosť modelu so vstupnými údajmi BCCD	77

Zoznam tabuliek

Tabuľka 1 - Porovnanie výsledkov experimentov obrazového spracovania údajov.....	74
Tabuľka 2 - Porovnanie výsledkov experimentov spracovania údajov strojovým učením	78

Zoznam skratiek

Skratka	Anglický význam	Slovenský význam
CT	computed tomography	počítačová tomografia
DIP	digital image processing	digitálne spracovanie obrazu
EEG	electroencephalogram	elektroencefalogram
EHR	electronic health record	elektronická zdravotná dokumentácia
EKG	electrocardiogram	elektrokardiogram
EMG	electromyogram	elektromyogram
EMR	electronic medical record	elektronické lekárske záznamy
EOG	electrooculogram	elektrookulogram
FDG	Fluorodeoxyglucose	Fluorodeoxyglukóza
Gd	Gadolinium	Gadolínium
HDR	high dynamic range imaging	vysoký dynamický rozsah
I	Iodine	Jód
JSON	JavaScript Object Notation	zápis objektov v jazyku JavaScript
MEG	magnetoencephalogram	magnetoencefalogram
MLE	maximum likelihood estimation	metóda maximálnej vierohodnosti
MRI	magnetic resonance	magnetická rezonancia
MS	mass spectrometry	hmotnostná spektrometria
NMR	nuclear magnetic resonance	nukleárna magnetická rezonancia
PAT	photoacoustic tomography	fotoakustická tomografia
PET	positron emission tomography	pozitrónová emisná tomografia
PSM	peptid-spectrum match scoring function	skórovacia funkcia peptidovej zhody
RTG	radiography	rádiografia
TIFF	Tagged Image File Format	Tagged formát obrazového súboru

Zoznam symbolov

Symbol	Jednotka	Význam symbolu
μm	mikrometer	jednotka dĺžky
MHz	Megahertz	jednotka frekvencie
Px	Pixel (picture element)	jednotka digitálnej rastrovej grafiky
T	Tesla	jednotka magnetickej indukcie

ÚVOD

Spracovanie údajov je v súčasnosti dôležitou tematikou v mnohých vedeckých oblastiach ako napr. v medicíne, biomedicíne, biotechnológiách a ďalších. V spomenutých oblastiach sú často používané obrazové údaje, ktoré poskytujú možnosť vizualizovať anatomické a patologické procesy v tele pacienta. V medicíne a biomedicíne sa používajú rôzne zobrazovacie technológie, pričom každá poskytuje iné informácie o funkcii a stavbe tkanív v tele pacienta. Spracovanie obrazových údajov zahŕňa rôzne úrovne, ako napr. segmentácia (rozpoznávanie oblastí záujmu), či klasifikácia (rozhodovanie o zaradení objektu do konkrétnej triedy).

Spracovanie obrazových údajov je zložitý a zdĺhavý proces, čo značí, že manuálne spracovanie je časovo náročné a vyžaduje množstvo kvalifikovaných špecialistov. Okrem toho, manuálne spracovanie môže byť subjektívne a závislé na predošlých skúsenostiach, vedomostiach a kvalifikácii špecialistu. Takéto spracovanie tiež môže byť náchylné na ľudské chyby a nedostatky. Dôsledkom môže byť nesprávna segmentácia alebo identifikácia, čo môže spôsobiť nekonzistentné výsledky alebo stanovenie nesprávneho postupu liečby.

Pre účely práce boli zvolené mikroskopické údaje zachytávajúce krvnú vzorku, kde je potrebné identifikovať a klasifikovať leukocyty a ich počet. Význam vo výbere a automatizácii spracovania mikroskopických údajov vidíme práve v tom, že mikroskopických údajov je zvyčajne veľké množstvo. Ďalším dôvodom voľby je množstvo ochorení, pri ktorých vie identifikácia leukocytov podporiť diagnostiku. Príkladmi sú rakovina krvi, infekcie, autoimunitné ochorenia alebo alergie.

Preto vnímame ako cieľ práce analýzu súčasných techník spracovania a dostupných nástrojov vo forme algoritmov, knižníc, či softvérov, ktoré môžu tento proces zjednodušiť. Následne je potrebné vyhodnotiť možné oblasti na zlepšenie, implementovať a overiť riešenie, ktoré podporí automatizáciu spracovania vybraného typu biomedicínskych obrazových údajov.

V prvej kapitole je spracovaný prehľad základných typov biomedicínskych údajov, so zameraním na spracovanie obrazových medicínskych údajov. Druhá kapitola obsahuje analýzu dostupných údajov a techník a na spracovanie obrazových údajov. Tretia kapitola sa zaoberá návrhom riešenia použitia dostupných techník a celkovej automatizácie procesu. Štvrtá kapitola popisuje implementované riešenie. Piata kapitola obsahuje vyhodnotenie, overenie a prínos riešenia v ďalšom výskume.

1 KLASIFIKÁCIA BIOMEDICÍNSKÝCH ÚDAJOV

S pribúdajúcim výskumom v oblasti medicíny a digitalizáciou výsledných údajov sa v dnešnej dobe množstvo biomedicínskych údajov neustále zväčšuje, a preto je kľúčové rozdeliť ich do kategórií a správne ich klasifikovať. Je to nevyhnutné najmä pre efektívne spracovanie, analýzu, či uchovávanie biomedicínskych údajov. Biomedicínske údaje sú získavané z rôznych zdrojov, čoho následkom je skutočnosť, že údaje vyskytujúce sa v biomedicíne sú často komplexné, majú rozdielny formát a obsahujú rôzne druhy informácií, z ktorých je potrebné vybrať iba kľúčové údaje. Preto sa klasifikácia biomedicínskych údajov stala kľúčovou časťou moderného zdravotníctva. Využíva sa na optimalizáciu procesov zdravotníckeho systému, diagnostiku ochorení, výskum nových liečebných metód a v konečnom dôsledku najmä na zlepšenie zdravotnej starostlivosti. Vyhodnocovanie biomedicínskych údajov sa realizuje pomocou rôznych metód, napríklad pomocou strojového učenia či prediktívnej analýzy.

Cieľom tejto kapitoly je poskytnutie teoretického úvodu v oblasti biomedicínskych údajov. V prvej časti sú spracované všeobecné typy biomedicínskych údajov. Obsahom druhej časti kapitoly je zameranie na teóriu viazanú k obrazovým biomedicínskym údajom.

1.1 Všeobecný prehľad biomedicínskych údajov

Táto podkapitola zahŕňa oboznámenie sa so základnými najpoužívanejšími typmi biomedicínskych údajov, poskytnutie prehľadu o údajoch konkrétneho typu, spôsobe ich získavania a spracovania. Spomenuté sú aj dôvody a význam spracovania daného typu spolu s konkrétnymi príkladmi údajov.

1.1.1 Genomické údaje

Genomickým údajom je charakteristický reťazec písmen, ktoré popisujú presné poradie chemických báz prítomných v molekule nukleovej kyseliny. Tieto údaje je možné z organizmu extrahovať rôznymi spôsobmi ako napríklad odberom krvi alebo moču s využitím viacrozmernej spektroskopie nukleárnej magnetickej rezonancie (NMR). Sekvencia DNA obsahuje všetky funkčné informácie o organizme, a preto sa analýza a dekodovanie genomických údajov používajú predovšetkým na odhalenie štruktúry, zmien a rozdielov v DNA sekvencii, ktoré sú následkom mutácií génu a spôsobujú rôzne typy ochorení akými sú napríklad rakovina, cukrovka alebo srdcové choroby. Pretože genomické údaje zložitého organizmu, napr. človeka sú pomerne rozsiahle, spracovanie genomických údajov môže byť hardvérovo náročné. Príkladom takýchto typov údajov sú

údaje zo sekvenovania DNA, údaje zo sekvenovania RNA, či epigenomické údaje. [1]–[4]

1.1.2 Proteomické údaje

Pod proteomickými údajmi rozumieme maticu kvantitatívnych hodnôt prítomných proteínov, peptidov a hodnoty skórovacej funkcie peptidovej zhody (PSM). Proteomické údaje sú získavané napríklad odobratím vzorky nádorového tkaniva, moču alebo séra za pomoci spektroskopie NMR alebo röntgenovej kryštalografie. Skúmané údaje sú porovnané so vzorkou pochádzajúcou zo zdravého tkaniva. Keďže proteíny sa podieľajú na väčšine molekulárnych fyziologických procesov, následkom nesprávnej regulácie expresie proteínov sú patologické stavy ako napríklad neurodegeneratívne ochorenia. Proteomickou analýzou je možné analyzovať proteínové zmesi v čase a zostaviť overiteľné hypotézy. Medzi proteomické údaje je možné zaradiť údaje z hmotnostnej spektrometrie, údaje o interakcii proteínu s proteínom, a tiež údaje o expresii proteínov. [5], [6]

1.1.3 Metabolomické údaje

Metabolomické údaje majú podobný formát ako proteomické údaje, ktoré obsahujú kompletný súbor metabolitov. Údaje sú získavané z biologických vzoriek bunkových kultúr, krvi alebo tkanív pomocou NMR alebo hmotnostnou spektrometriou (MS). Podobne ako pri analýze proteomických údajov, vzorka metabolomických údajov je porovnaná so vzorkou zdravého jedinca. Tieto údaje môžu byť použité na profilovanie reakcií pacienta na liečbu, napríklad pri liečbe psoriázy a určiť, či by liečba bola pre pacienta prospešná, alebo by mu naopak mohla uškodiť. Metabolomické profilovanie má tiež využitie pri odhalení tichých mutácií, pri ktorých nenastávajú tiché mutácie. Príkladmi metabolomických údajov môžu byť údaje o metabolickom profilovaní alebo údaje o kvantifikácii metabolitov. [7], [8]

1.1.4 Obrazové údaje

Obrazové medicínske údaje majú grafický charakter, primárne vo forme digitálneho rasterizovaného obrazu. Údaje sa získavajú použitím prístrojov, ktoré vysielajú ionizujúce, neionizujúce alebo rádiové žiarenie, ktoré sa pri prechode rôznymi štruktúrami absorbuje alebo odrazí. Detekciou týchto javov je vytvorený obrazový výstup. Vďaka jednoduchej interpretácii daného typu údajov môžu byť anomálie posudzované lekárom, rovnako ako softvérom. Obrazové údaje majú medicínske využitie hlavne v oblasti klinickej diagnostiky, slúžia na vyšetovanie vnútornej štruktúry orgánov, tkanív či kostí a vytváraní súborov fyziologických údajov. Medzi obrazové údaje

zaraďujeme širokú škálu údajov. Medzi najznámejšie patria predovšetkým údaje zo zobrazovania magnetickou rezonanciou (MRI), údaje z počítačovej tomografie (CT), údaje z pozitronovej emisnej tomografie (PET), rádiografické údaje (RTG), údaje z fotoakustickej tomografie (PAT) a údaje z magnetoencefalografie (MEG). [9]

1.1.5 Klinické údaje

Klinické údaje zvyčajne lokalizujú, popisujú a vysvetľujú informácie o pacientovi, a ich hlavnou úlohou je zjednodušenie prístupu k zdrojom informácie. Údaje zahŕňajú napríklad diagnózu, laboratórne testy, údaje z fyzického monitorovania, demografické údaje, predpísané lieky a pod. Daný typ údajov sa získava priebežne v rámci starostlivosti o pacienta, prípadne pri klinických testovaniach. Klinické údaje sa dajú spracovať vo forme množiny štruktúrovaných metadát. Takáto forma spracovania následne ponúka jednoduchú analýzu či prepojenie údajov pacienta. Veľkou výhodou je tiež možnosť zhromažďovania podobných údajov, ktoré vedie k podpore a rozvoju poskytovania zdravotnej starostlivosti alebo pre účely výučby a výskumu lekárskech fakúlt. Do množiny klinických údajov zaraďujeme napr. údaje z elektronických lekárskech záznamov (EMR), údaje z elektronickej zdravotnej dokumentácie (EHR), údaje z lekárskech faktúr, údaje o nárokoch na zdravotné poistenie, či údaje o zdravotníckych pomôckach. [10]–[13]

1.1.6 Údaje o správaní

Údaje o správaní jedinca majú podobne štruktúrovanú formu ako klinické údaje. Údaje zahŕňajú štruktúrované informácie o psychologickom stave, správaní, presvedčeníach a rozhodnutiach pacienta. Získavanie údajov prebieha formou rozhovorov, pozorovania alebo monitorovaním pacienta pomocou nositeľných zariadení. Pre spracovanie sa používajú kvalitatívne alebo kvantitatívne metódy, v závislosti od požadovaného výsledku a zaradenia na základe spoločných znakov, s cieľom maximalizovať získané množstvo informácie. Spracovanie behaviorálnych údajov má veľký význam primárne pre diagnostiku a liečbu pacientov trpiacich duševnými chorobami ako napr. post-natálna depresia, úzkostná porucha, schizofrénia alebo post-traumatická stresová porucha. Sledovanými údajmi o správaní sú napríklad údaje o fyzickej aktivite, údaje o spánku, údaje o stravovaní a údaje o sociálnej interakcii. [14]–[17]

1.1.7 Údaje o životnom prostredí

Údaje o životnom prostredí zahŕňajú prehľad relevantných informácií, zameraných primárne na vplyv konkrétneho externého faktora na zdravie jedinca alebo populácie. Sledovanými ovplyvnenými znakmi môžu byť napríklad rozdiely v hustote kostí,

množstvo telesného tuku prípadne dentálne analýzy. Kontinuálne sledovanie môže prebiehať pomocou nositeľných zariadení, prípadne individuálnych meraní. Štúdie jednotlivých externých faktorov sú následne spájané pre vytvorenie komplexnej analýzy životného prostredia na zdravie organizmu. Výskum vystavenia rôznym typom životného prostredia je významný hlavne v pediatrii, pretože vystavenie toxínom alebo znečistenie ovzdušia má vplyv na vývin a zdravie detí, či plodu v prenatálnej fáze. Ďalším cieľom spracovania daného typu údajov je prediktívne modelovanie pre účely liečby pacienta, napríklad pri liečbe astmy. Analyzovanými údajmi životného prostredia sú údaje o kvalite ovzdušia, údaje o kvalite vody, údaje o podnebí, údaje o expozícii toxínom a znečisťujúcim látkam, atď.. [18]–[20]

1.1.8 Fyziologické údaje alebo údaje o biosignáloch

Údaje o biosignáloch sú reprezentované číselnými hodnotami meniacimi sa v čase, vytvárajúc krivku zobrazenú v grafe. V grafickom zobrazení sú zvyčajne analyzované znaky ako amplitúda, frekvencia a pod. Fyziologické údaje zachytávajú bio-elektrickú aktivitu svalov, mozgu a srdca prostredníctvom elektród rozmiestnených na tele pacienta. Kvôli rozsiahlemu množstvu nameraných údajov, môže byť výstup ďalej spracovaný algoritmami pre extrakciu významných hodnôt alebo inú transformáciu. Príkladom môže byť waveletová transformácia, pomocou ktorej môžu byť údaje analyzované v časovej aj frekvenčnej doméne. Spracovanie biosignálov je dôležité pre diagnostiku arytmií myokardu, epileptických záchvatov, aktivity a poruchy centrálného nervového systému a tiež celkovú interpretáciu zdravotného stavu jedinca. Príkladmi údajov o biosignáloch sú údaje z elektrokardiogramu (EKG), údaje z encefalogramu (EEG), údaje z elektromyogramu (EMG), údaje z elektrookulogramu (EOG) alebo údaje z magnetoencefalogramu (MEG).[21], [22]

1.1.9 Textové údaje

Textové údaje tvoria rozsiahlu časť medicínskych údajov. Sú to obsiahle neštruktúrované údaje s odbornou terminológiou, ktoré poskytujú detailný odborný popis sledovaných aspektov. Daný typ údajov je získavaný rôznymi spôsobmi v závislosti od typu a kategórie údajov. Textové údaje o pacientoch sú nadobudnuté vyšetovaním, pozorovaním a rozhovorom s pacientom, ako napríklad anamnézy, reakcie na liečbu, či lekárom vyvodené závery. Iným typom textových údajov je vedecká literatúra, popisujúca nové poznatky, princípy a metódy vyplývajúce z uskutočnených výskumov. Údaje a dokumenty sú kategorizované podľa lekárskej špecializácie. Pre klasifikáciu a zjednotenie rôznych textových údajov sa používa lematizácia – metóda pre normalizáciu textu, ktorá nahrádza a mapuje slová s podobným významom. Následne

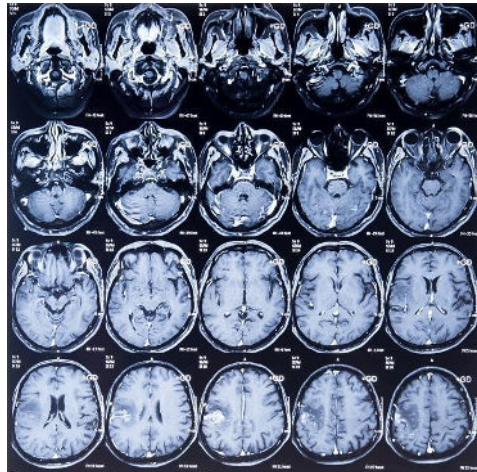
sú údaje kategorizované pomocou metód hlbokého učenia, napríklad pomocou konvolučných neurónových sietí. Spracovanie tohto typu údajov je významné primárne pre skvalitnenie zdravotnej starostlivosti, vzdelávanie a odhalenie súvislostí, spoločných znakov a tvorbu predikcie pri medicínskej diagnostike. Medzi textové údaje v oblasti medicíny zaraďujeme lekárske záznamy, vedeckú literatúru, údaje z prieskumov a rozhovorov s pacientom. Taktiež tu môžu byť zaradené rozšírené údaje príp. vygenerovaná dokumentácia z EHR. [23]–[26]

1.2 Klasifikácia obrazových biomedicínskych údajov

Primárnym zameraním práce je spracovanie obrazových biomedicínskych údajov, preto je nasledujúca časť venovaná hlbšej analýze klasifikácie a spracovania obrazových biomedicínskych údajov.

1.2.1 Magnetická rezonancia (MRI)

Magnetická rezonancia je zobrazovacia technológia, ktorá neinvazívnym spôsobom zostavuje podrobné trojrozmerné anatomické a fyziologické snímky s využitím magnetického poľa a rádiových vln. Tento typ zobrazovania sa používa primárne na zobrazovanie kĺbov, členkov, zápästia, mozgu, srdca a ciev. Obrázok 1 je ukážkou snímok magnetickej rezonancie. [9], [27]



Obrázok 1 - Sekvencia MRI snímok mozgu [28]

Prístroj pre MRI má tvar veľkej trubice, s výsuvným stolom pre pacienta v ľahu. Počas vyšetrenia je pacient umiestnený v skenovacej ploche zariadenia. Prístroj v danej oblasti vytvára silné magnetické pole, ktoré núti protóny atómov vodíka v tele pacienta k zarovnaniu s poľom. Skener MRI má zvyčajne intenzitu v rozsahu 0.5 až 1.5 T (tesla). Prístroj následne vyšle rádiovú vlnu, ktorá spôsobí vychýlenie protónov. Rezonancia jadier vodíka je spôsobená frekvenciou rádiových vln. Vypnutie prístroja spôsobí návrat magnetického vektora do stavu pokoja, čo spôsobí vyslanie rádiových vln, ktoré sú použité na vytvorenie MRI snímok. V blízkosti monitorovanej časti tela sú umiestnené cievky, ktoré slúžia ako prijímače na detekciu signálu. Výsledný signál sa zobrazí v sade prierezových snímok v sagitálnej, transverzálnej alebo koronálnej rovine. Pre zvýraznenie tkanív môže byť vyslané väčšie množstvo rádiových impulzov. Rozdielne zobrazenie rozdielnych tkanív na snímku je dosiahnuté rýchlosťou excitovaných atómov pri návrate do rovnovážneho stavu. Meranie času pre relaxáciu protónov je vykonané dvomi spôsobmi - návratom magnetického vektora do pokojového

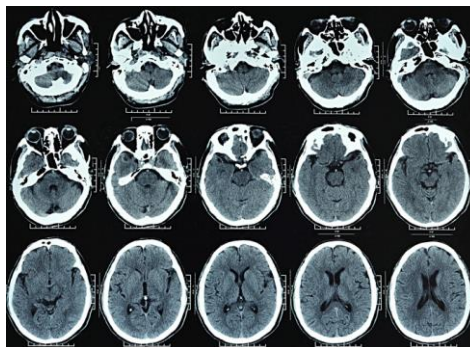
stavu (T1 relaxácia) a návratom axiálneho spinu atómu do pokojového stavu (T2 relaxácia). Rôzne tkanivá majú rôzne relaxačné časy a zmenou rýchlosti vysielania impulzov je možné dosiahnuť zdôraznenie konkrétneho typu tkaniva na snímke. Pre zvýraznenie tkanív alebo abnormalít, je možné pacientovi podať kontrastnú látku, ako napr. Gadolínium (Gd). [27], [29], [30]

Najväčšou výhodou MRI voči ostatným zdrojom obrazových biomedicínskych údajov je, že pri vytváraní snímok využíva žiarenie, ktoré sa nachádza v rádiovfrekvenčnom rozsahu v s ktorým sa bežne stretávame, a po prechode žiarenia tkanivo neostáva poškodené, z čoho vyplýva, že pri vyšetrení pomocou MRI technológie doteraz nie sú známe žiadne biologické riziká. Prednosťou magnetickej rezonancie je tiež vysoká kvalita a detailnosť vyprodukovaných snímok. [27]

Negatív magnetickej rezonancie je viacero. Vyšetrenie je väčšinou finančne nákladné, a v určitých prípadoch nemusí ponúknuť dostatok detailov pre potreby diagnostiky. V súčasnosti napríklad nedokáže rozlíšiť malígne a benígne nádorové ochorenia. Problém môžu predstavovať aj kovové implantáty v tele pacienta ako napríklad kovové svorky, chlopne alebo kardiostimulátory, ktoré môžu počas pôsobenia magnetického poľa zmeniť polohu. Pri kĺbových protézach z kovu môže byť obraz v okolí protézy skreslený. Inými negatívami môže byť poškodenie obličiek pacienta a alergia, čo znemožňuje prijatie kontrastnej látky. Ďalšou nevýhodou je hlučnosť samotného prístroja, dlhšie časové trvanie vyšetrenia, rovnako ako nedostatočný priestor pre pacienta, ktorý môže byť problémový pre pacientov trpiacich klaustrofóbiou. [27]

1.2.2 Počítačová tomografia (CT)

Ďalším významným typom pre vytváranie obrazových biomedicínskych dát je počítačová tomografia. Počítačová tomografia, podobne ako magnetická rezonancia, neinvazívnym spôsobom zostavuje prierezové snímky vnútorných štruktúr tela pacienta, na ktorých je možné odhaliť poruchy prietoku krvi a taktiež poškodenia kostí a orgánov. Obrázok 2 zobrazuje ukážku snímok počítačovej tomografie. [9], [31]



Obrázok 2 - Sekvencia CT snímok mozgu [32]

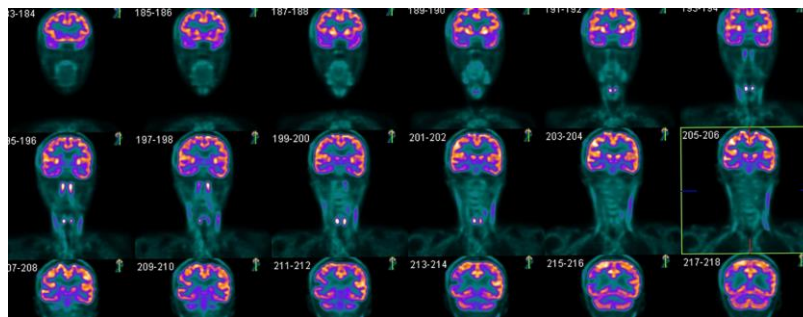
Zariadenie počítačovej tomografie môže byť navonok podobné prístroju magnetickej rezonancie. Skladá sa z pohyblivého stola na ktorom je umiestnený ležiaci pacient, a rotujúceho skenera s kruhovým otvorom, ktorým stôl postupne prechádza a vytvárajú sa prierezové snímky. Rozdiel medzi po sebe nasledujúcimi dvojrozmernými snímkami je definovaný lekárom, zvyčajne v rozmedzí 1 až 10 milimetrov. Stôl s pacientom sa pri špirálovitom vyšetrení pohybuje neustále, zatiaľ čo prístroj otáča röntgenovú trubicu obsahujúcu žiariče a detektory. Trvanie jednej rotácie skenera je jedna sekunda. Zdroj vysiela vejárovitý zväzok lúčov röntgenového žiarenia ktoré ožiaria skenovanú časť tela a na opačnej strane trubice sa nachádzajú detektory, ktoré zaznamenávajú lúče, ktoré prešli telom pacienta. V závislosti od typu tkaniva sa lúče môžu absorbovať (úplne alebo čiastočne) alebo odraziť, vďaka čomu je možné na snímke jednotlivé typy tkanív odlíšiť. Snímky sú vytvárané postupne v kraniálnom smere. Jedna fáza vyšetrenia pozostáva z 10 až 50 rotácií skenera. V závislosti od typu vyšetrenia sa celé zobrazovanie môže skladať z niekoľkých fáz. Z jednej rotácie sa zhromaždí niekoľko snímok z rôznych uhlov, ktoré sú počítačom spojené pre vytvorenie prierezu tkanív z každej rotácie. Snímka je vytvorená podľa rádiosenzitivity v porovnaní so známou hodnotou hustoty tkaniva. Prierezové snímky je následne možné zobrazovať vo všetkých anatomických rovinách, v závislosti od potrieb diagnostiky. Podobne ako pri magnetickej rezonancii, aj pri vyšetrení pomocou počítačovej tomografie existuje možnosť podania kontrastnej látky - jódu (I) pacientovi injekciou pre lepšiu vizualizáciu, napríklad pri zobrazovaní cievnych štruktúr. [9], [31], [33]

Prednosťou počítačovej tomografie je možnosť zobrazovania v reálnom čase, ktoré má využitie napríklad v neurochirurgii, pre rýchlu detekciu poškodení pri poraneniach hlavy, alebo ako intraoperačné vyšetrenie počas a bezprostredne po zákroku. Podobne ako pri magnetickej rezonancii, výhodou je vysoké rozlíšenie a množstvo zachytených detailov. Pozitívom sú tiež široké možnosti aplikácie skenov počítačovej tomografie pri diagnostike a určení liečebného procesu. [31]

Za nevýhodu je považované veľké množstvo radiácie (50 až 1000-krát väčšia dávka ako pri rádiologickom vyšetrení), ktorej je pacient počas vyšetrenia vystavený, hoci prínos vyšetrenia je v urgentnej medicíne častokrát väčší, ako negatívne účinky žiarenia. Kovové prvky ako implantáty či chirurgické svorky môžu na snímkach vytvárať odrazy a zakrývať štruktúry tela, čo môže vyústiť k nesprávnemu diagnostickému záveru. Slabou stránkou snímok pochádzajúcich z počítačovej tomografie je tiež slabé zobrazenie väzov, šliach, miechy alebo medzistavcových platničiek. [31]

1.2.3 Pozitrónová emisná tomografia (PET)

Pozitrónová emisná tomografia je ďalším typom neinvazívnej zobrazovacej technológie v medicíne, ktorá sa používa predovšetkým na vizualizáciu zmien metabolických procesov, akými sú napríklad prietok krvi či absorpcia a chemické zloženie látok. Táto metóda je založená na rádiotraktorech (rádioaktívnych látkach), pomocou ktorých je možné dané procesy prehľadne vizualizovať. Pozitrónová emisná tomografia sa využíva napríklad na diagnostiku neurónovej plasticity, mozgovej mŕtvice, demencie alebo diagnostiku pohybových porúch. Na obrázku 3 sú prezentované snímky pozitronovej emisnej tomografie. [9], [34], [35]



Obrázok 3 - Sekvencia PET snímok mozgu [36]

Prístroj pozitronovej emisnej tomografie sa podobá prístroju počítačovej tomografie. Rovnako sa skladá z pohyblivého stola, na ktorom je umiestnený ležiaci pacient, a skenera, ktorý rotuje okolo pohyblivého stola. Približne 30 až 60 minút pred vyšetrením je pacientovi podaná intravenózna látka - napr. Fluorodeoxyglukóza (FDG, $C_6H_{11}^{18}FO_5$). Po vstrebaní látky ho obehovej sústavy pacienta prebehne skenovanie pacienta. Pri procese skenovania rádioizotopové jadro vyžaruje pozitron, ktorý po zrazení s elektrónom nachádzajúcim sa v tkanive premení hmotnosť na energiu, ktorá sa prejaví vo forme dvoch fotónov. Detektorom prístroja pozitronovej emisnej tomografie je kamera, spolu s 18 až 24 prstencami zloženými z veľkého množstva scintilačných kryštálov, ktoré sú rozmiestnené okolo monitorovaného spektra. Úlohou scintilačných kryštálov je absorbovať fotóny a premieňať ionizujúce žiarenie na inú formu, napr. viditeľné svetlo alebo ultrafialové žiarenie, detekovateľné kamerou. Pri skenovaní sa využíva skutočnosť, že zdravé bunky pohltia menšie množstvo rádiotraktora, ako choré bunky. Oblasti s chorými bunkami sa tak pri vyšetrení zachytia ako "horúce miesta", ktoré skener dokáže zaznamenať a následne vytvoriť obrázky zachytávajúce postihnuté tkanivo. Vyšetrenie pomocou pozitronovej emisnej tomografie zvyčajne trvá 10 až 40 minút. Výsledkom vyšetrenia sú skeny transaxiálnych rezov, ktoré pokrývajú približne 14 až 15 centimetrový pás tela pacienta, z ktorých je v reálnom čase vytvorený 3D obraz. Vo väčšine vyšetrení pomocou pozitronovej emisnej tomografie je u pacienta snímaná

primárne oblasť medzi lebečnou spodinou a proximálnymi stehnami, pretože práve v tejto časti sú najčastejšie objavené metastatické ochorenia. Podobne, ako pri už spomenutých typoch zobrazovania, aj pri PET zobrazovaní je okrem priestorového modelu sledovať snímky v troch základných typoch anatomickej roviny. Vyšetrenie sa zvyčajne vykonáva spolu s MRI alebo CT vyšetrením, pre získanie podrobnejších informácií pre účely diagnostiky.[34], [35], [37]–[39]

Výhodou pozitronovej emisnej tomografie je zameranie sa na molekulárnu aktivitu v tele, pri ktorom je možné odhaliť aj skoré štádiá ochorenia. Tento typ zobrazovania tiež dokáže určiť presné štádium rakoviny, a tiež na základe metabolických procesov (napr. rozklad glukózy) rozlíšiť benígne a malígne nádory. V súčasnosti sa realizuje rozvoj PET/CT technológie, čo je integrácia pozitronovej emisnej tomografie a počítačovej tomografie do jedného spoločného zariadenia. Iným pozitívom je použitie vo veľkom množstve medicínskych odborov ako napríklad onkológia, neurológia, kardiológia či pri auto imúnnych ochoreniach.[34], [37]

Dôvodom, prečo sa PET vyšetrenie vo väčšine prípadov vykonáva spolu s CT alebo MRI vyšetrením, je podstatne nižšie rozlíšenie vyprodukovaného obrazu, než pri spomenutých technológiách, čo v konečnom dôsledku zvyšuje celkové náklady vyšetrenia. Taktiež, nedostatok detailov môže byť príčinou nedokonalnej lokalizácie nálezov. Inou nevýhodou je použitie rádioaktívneho materiálu, nie len z pohľadu pacienta, ale aj z pohľadu práce s takýmto typom chemickej zlúčeniny, pretože musí byť k dispozícii laboratórium, ktoré sa musí nachádzať v blízkosti skenovacieho zariadenia. Nevýhodou môže byť aj nemožnosť použitia kontrastnej látky z hľadiska komplikácií zdravotného stavu pacienta. [34], [39]

1.2.4 Ultrasonografia (USG)

Ultrasonografia je typom zobrazovacieho vyšetrenia v reálnom čase, založenom na ultrazvuku, ktoré sa používa na vizualizáciu šliach, kĺbov, svalov a vnútorných orgánov, primárne pre vyšetrenia brušnej dutiny, v kardiológii, gynekológii, urológii, či cerebrovaskulárne vyšetrenia. Obrázok 4 zobrazuje ukážku ultrasonogramu. [9], [40]



Obrázok 4 - Prenatálne USG snímky nervového systému plodu [41]

Zariadenie pre vykonávanie ultrasonografického vyšetrenia sa skladá zo snímača, ktorým je ručná sonda, generátora pre vysielanie impulzov, jednotky pre zaostrenie, kompenzačných zosilňovačov, meniča, digitálnych procesorov a zobrazovacej jednotky. Takýto diagnostický skener typicky pracuje v rozsahu frekvencie 2 až 18 MHz (Megahertzov). Pre vedenie ultrazvuku medzi snímačom a telom pacienta sa pri vyšetrení používa gél na vodnej báze. Generátor prostredníctvom sondy vysielá do tela krátke ultrazvukové impulzy, v ktorých sú obsiahnuté frekvencie z celého spomenutého rozsahu a trvajú menej ako jednu mikrosekundu. Keďže rôzne tkanivá ľudského tela sú heterogénne, priechod vln tkanivom spôsobí absorpciu časti energie a dochádza odrazu, rozptylu či lomu vln, v závislosti od akustickej impedancie charakteristickej pre daný typ tkaniva. Nižšia frekvencia je viac prenikavá, preto sa používa napríklad na vyšetrenie orgánov brucha. Vyššia frekvencia zas spôsobuje väčšiu absorpciu a lom svetla, a tak poskytuje väčšie množstvo zobrazených detailov. Aby sa pri zobrazovaní plochy zobrazovali rovnomerne, musia sa absorbované vlny kompenzovať elektronicky. Následne sú detektorom zachytávané odrazené lúče. Detektor môže použiť rôzne druhy sond, napríklad lineárnu, sektorovú, zaostrujúcu s kruhovým meničom, konvexnú alebo testovaciu s hojdačím zrkadlom. Výber sondy závisí od veľkosti monitorovanej oblasti. Je tiež možné použiť štyri základné režimy ultrazvuku – režim A, režim B, režim M a Dopplerovský režim. Režim A sa používa na snímanie jednej línie cez tela, ako funkcie hĺbky. Režim B používa lineárnu sústavu a skenuje rovinu tela v dvoch dimenziách. Režim M sa používa pri zobrazovaní pohybu, je to vlastne sekvencia skenov typu B. Dopplerovský režim využíva pri vizualizácii menič, pre doplnenie alebo zníženie frekvencie o Dopplerov posun, ktorý vzniká pri zmene vzdialeností generátora a prijímača. Ultrazvukové vyšetrenie je možné rozdeliť na laterálne, hĺbkové a axiálne. Pri vyšetrení sa zvyčajne kontrastné látky nepoužívajú, hoci existuje aj ultrazvukové vyšetrenie s kontrastom (CEUS), kedy je injekčne podaný kontrastný roztok. Vyšetrenie je neinvazívne, zvyčajne trvajúce približne 30 minút. [9], [40], [42]–[44]

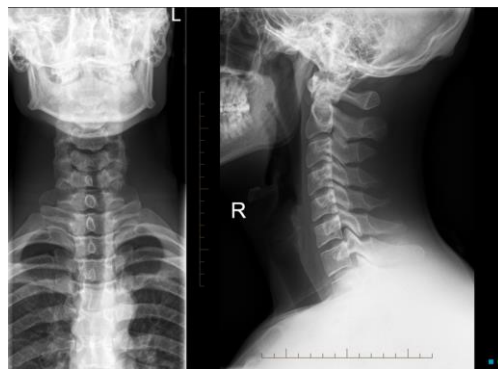
Výhodou ultrazvukovej technológie je, že nemá žiadne škodlivé vedľajšie účinky, pretože nepoužíva radiáciu ako väčšina spomenutých typov zobrazovania. Táto skutočnosť prispieva aj k tomu, že ultrasonografické vyšetrenie môže byť absolvované aj počas tehotenstva. Pozitívom je aj široké použitie nie len pri diagnostike ale aj ako pomoc pri operačnom vyšetrení v reálnom čase, najmä pri novorodencoch. Inou výhodou sú nízke náklady USG vyšetrenia. [42]–[44]

Medzi nevýhody ultrasonografu patrí pomerne nízke rozlíšenie výstupného obrazu a potreba jemných zručností a skúseností ultrasonografického pracovníka pri manipulácii so sondou. Ďalšou nevýhodou je anizotropia, fyzikálna vlastnosť nervov

a šliach, ktorá spôsobuje odrazy na ultrasonograme. Taktiež kosti blokujú ultrasonografické vlny, čo spôsobuje problémy napríklad pri vyšetrení chrbtice.[40], [44]

1.2.5 Rádiografia

Rádiografia je neinvazívnym typom zobrazovania, ktorý pre vytvorenie obrazových výstupov používa röntgenové žiarenie, ktoré má vysokú úroveň energie a preniká vnútornými štruktúrami tela. Ide o najjednoduchšiu formu vytvorenia lekárskeho obrazových snímok. Používa sa na diagnostiku a liečbu štrukturálneho poškodenia, anomálie alebo zistenie prítomnosti cudzích predmetov v tele, ale aj pri dentálnom vyšetrení a mamografii. Obrázok 5 prezentuje príklad rádiografických snímok. [45], [46]



Obrázok 5 - Rádiografická snímka krčnej chrbtice [47]

Zariadenie pozostáva z niekoľkých hlavných častí – generátor röntgenového žiarenia, systému pre detekciu žiarenia a lôžka pre pacienta. Pre potrebu vyšetrenia je pacient uložený na lôžko v potrebnej polohe a hlavica generátora je nasmerovaná na požadovanú oblasť pre skenovanie. Pacient musí počas vyšetrenia byť v nehybnom stave, pretože pohyb spôsobí rozmazanie obrazovej snímky. Počas procesu skenovania sa v generátore žiarenia sa ohrieva katóda až dôjde k vytvoreniu vysokého elektrického potenciálu medzi katódou a anódou, čo spôsobí tok elektrónov. Pri dopade elektrónov na anódu sa ich energia a rýchlosť zvyšuje a so zvyšujúcou energiou sa zvyšuje aj prenikanie žiarenia. Elektróny v anóde sú sústredované do jedného bodu, čo zabezpečuje zaostrovanie obrazu. Röntgenové žiarenie je formou elektromagnetického žiarenia s krátkou vlnovou dĺžkou, čo mu umožňuje prenikáť cez štruktúry, ktoré odrážajú viditeľné svetlo. Pre vytvorenie kontrastu medzi rôznymi typmi tkanív sa používa nastavenie kilovoltového (kV) elektrického potenciálu, pričom pri nižšom kV potenciáli dochádza k vyššej absorpcii žiarenia a naopak, použitie vysokého kV potenciálu spôsobí príliš veľký prienik žiarenia. Pre ochranu nemonitorovaného zvyšku tela pacienta pred vysielaným röntgenovým žiarením sa používa tienenie. Dĺžka skenovania sa líši od požadovaného počtu oblastí pre skenovanie. Zvyčajne trvá do 15 minút, avšak po analyzovaní výstupov môže rádiografický pracovník rozhodnúť o potrebe vykonania

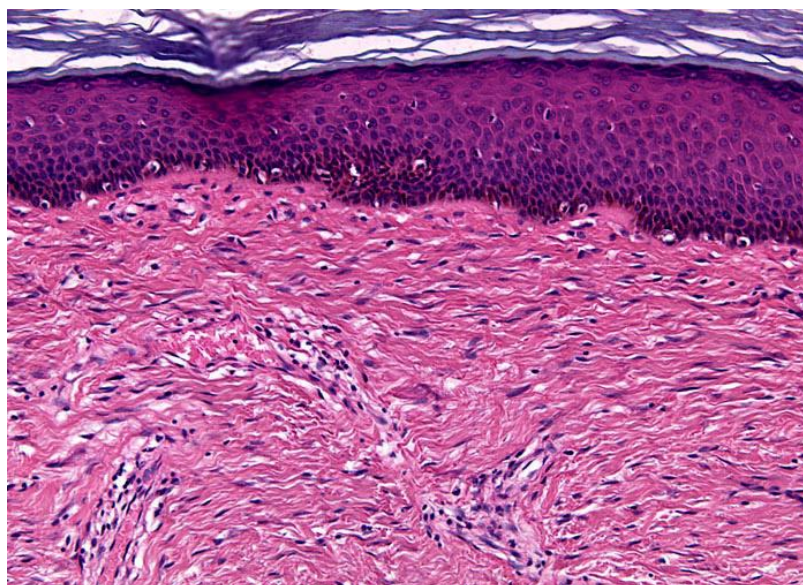
dodatočných snímok, napríklad z iných uhlov. Následne pracovník vyhodnotí vytvorené snímky a stanoví závery pre lekára ktorý rádiografické vyšetrenie odporučil. [46], [48], [49]

Výhodou použitia rádiografie je neinvazívnosť a dostupnosť tejto technológie. Pozitívom je aj rýchle vytvorenie obrazových údajov, ktoré je využité primárne pri urgentnej diagnostike. Ďalším pozitívom je, že pri správnych nastaveniach sú rádiogramy v dobrej kvalite a môžu dobre zobrazovať aj mäkké tkanivá a dokáže diagnostikovať nádory bez potreby operačného vyšetrenia. Výhodou rádiografie voči iným spomenutým zobrazovacím technológiám sú aj nižšie náklady. [46], [50]

Nevýhodou rádiografickej technológie je vystavenie pacienta radiácii, preto je potrebné vytvoriť obrazové snímky v čo najmenšom čase a požadovať čo najnižší počet snímok. Vzhľadom na spomenuté vystavenie žiareniu by rádiografické vyšetrenie nemalo prebiehať v tehotenstve. Inou nevýhodou je, že z rádiografických údajov nie je možné vyčítať priestorové informácie. Vysoká hustota kostí pri zobrazovaní spôsobuje zakrytie štruktúr v rovine paralelnej s röntgenovým lúčom. [46], [51]

1.2.6 Fotomikroskopia buniek

Fotomikroskopia buniek je v súčasnosti populárna a rozvíjajúca sa zobrazovacia technológia, pre digitálne zobrazovanie a podrobnú analýzu buniek a subcelulárnych štruktúr a procesov. Primárnym zameraním danej oblasti je automatizácia klasifikácie a segmentácie s využitím informačných technológií. Fotomikroskopia buniek je aplikovaná najmä v patológii a toxikológii na diagnostiku alebo vývoj liekov. Na obrázku 6 je zachytená mikroskopická vzorka tkaniva pacienta [52]



Obrázok 6 - Tkanivo jazvy po 10 násobnom zväčšení [53]

Systémy pre digitálne spracovanie obrazu (DIP) čoraz častejšie pre plnú automatizáciu procesov používajú motorizované mikroskopy, ktoré tiež ponúkajú grafické používateľské rozhranie pre zobrazovanie výstupov. Kľúčovými vlastnosťami mikroskopického zariadenia je možnosť výberu vhodného osvetlenia, paralelizované zobrazovanie z viacerých polôh v rámci jedného experimentu a úprava zorného poľa pre možnosť výberu počtu zobrazovaných buniek. Dôležitými aspektami zariadenia sú aj LED svetelné zdroje, možnosti optických filtrov, široká škála objektívov a vhodná kamera pre zachytenie obrazových údajov. Pre potreby mikroskopického vyšetrenia je z organizmu odobratá vzorka konkrétneho typu, podľa potreby vyšetrenia. Základnými vzorkami sú napríklad krv a krvné frakcie, tkanivo (odobraté počas chirurgického zákroku alebo pitvy), vzorka moču, vlasov, slín a podobne. Procedúra získavania vzorky sa teda líši v závislosti od typu vzorky. Dôležitou časťou vyšetrenia je stabilizácia a správne uchovanie vzorky, aby nedošlo k poškodeniu a neželaným biologickým zmenám pred následným spracovaním. Z odobratej vzorky je s použitím tekutej kultúry pripravený preparát. V prípade dlhodobého pozorovania je dôležité pripraviť preparát tak, aby nedošlo k vysušeniu. Pozorovanie preparátu zvyčajne trvá od niekoľkých minút až do 10 dní. Pri pozorovaní sa môžu použiť excitačné filtre pre detekciu fluorescenčného svetla, vyžarovaného vzorkou. Pri zachytávaní obrazu je dôležitým aspektom frekvencia snímania, ktorá závisí od dynamiky procesu. Pre zachytenie vhodných snímok je dôležitá správna expozícia. Expozičný čas by mal byť optimalizovaný pre zvýšenie pomeru signálu k šumu. Pri fluorescenčnom zobrazovaní je dôležité, aby čas expozície nebol príliš krátky, pretože vlastná endogénna autofluorescencia vzorky prekoná fluorescenčný signál, ktorý nebude dostatočne viditeľný. Vytvorené snímky sú následne analyzované lekárske pracovníkom alebo softvérom pre detekciu poškodenia alebo abnormalít nachádzajúcich sa v bunkových organelách pacienta. V závislosti od potrebných postupov sa môže čas celého procesu vyšetrenia nachádzať v rozpätí pár hodín ale aj niekoľko dní. [52], [54], [55]

Výhodou fotomikroskopického zobrazovania je jeho všeobecné využitie, s použitím špeciálnych konfigurácií pre špecifický charakter pozorovania. Okrem získania obrazových snímok zachytených v konkrétnom čase, zariadenia tiež ponúkajú sledovanie v reálnom čase, alebo vyhotovenie videa, ktoré môže poskytnúť väčší prehľad o bunkových procesoch. Inou pozitívnou vlastnosťou danej technológie je zobrazovanie výstupov na farebnej škále, ktoré napríklad pri fluoromikroskopii dokáže pomocou rôznych farebných kanálov odlíšiť rôzne fluorofóry, čo zjednodušuje proces analýzy snímku. [54]

Najväčšou nevýhodou je nutnosť dôslednej prípravy preparátu pri pozorovaní, ktorá môže byť časovo náročná. Pri zobrazovaní je nutné sa vyhnúť fototoxicite, pretože niektoré typy buniek sú náchylné na prílišné osvetlenie, čo môže spôsobiť znehodnotenie alebo nepoužiteľnosť vzorky. Taktiež, kvalita snímok je závislá od konfigurácie hardvéru nastavenej odborným pracovníkom. Negatívom môže byť pre niektoré vyšetrenia aj potreba invazívneho zákroku za účelom získania bioorganickej vzorky. [54]

1.2.7 Záver

Všetky spomenuté zobrazovacie technológie majú v medicíne naozaj široké využitie a tvoria nenahraditeľnú súčasť biomedicíny. Každý typ zobrazovania má svoje výhody aj nevýhody a každý je zameraný primárne na iné účely diagnostiky a liečby, preto sú všetky typy významné. Hoci najmä nevýhody sa môžu niekedy zdať významné, vo väčšine prípadov v praxi prínos vyšetrenia pre záchranu ľudského života prevyšuje možné negatíva vyšetrenia. Pri vyšetrení, ktoré je možné realizovať viacerými technológiami, napríklad MRI alebo CT vyšetrením, alebo pri rozhodnutí o počte snímok, by malo byť rozhodnutie zamerané na čo najvyšší prínos vyšetrenia s ohľadom na zachovanie najlepšieho možného zdravia pacienta.

Aj keď spomenuté boli iba niektoré základné spôsoby zobrazovania, je pochopiteľné, že množstvo vyprodukovaných obrazových biomedicínskych dát sa s rozvojom biomedicíny výrazne zväčšuje, a preto je aktuálnou témou spracovanie týchto údajov pomocou informačných technológií. Dôkladné spracovanie pomocou softvéru môže byť veľkým prínosom pre medicínskych pracovníkov, môže odstrániť ľudskú nedokonalosť, ale aj skrátiť časové trvanie, ktoré je v biomedicíne kľúčové.

2 ANALÝZA PROBLEMATIKY

Použitie obrazových biomedicínskych údajov prinieslo doposiaľ v oblasti zdravotníctva veľký úžitok. Na základe grafickej podoby vnútorných štruktúr tela vie už dlhý čas lekár alebo iný zdravotnícky pracovník veľmi rýchlo a presne stanoviť, prípadne vylúčiť diagnózu a navrhnúť postup liečby. V súčasnosti sa podstata verejného zdravotníctva mení a cieľom už nie je liečba ochorení, ale prevencia a rozširovanie povedomia o ochrane zdravia seba a blízkych, dôsledkom čoho je predchádzanie možným ochoreniam. Pre účely prevencie sú stanovené hypotézy príčin ochorení, ktoré je potrebné potvrdiť alebo vyvrátiť. Pre potvrdenie hypotézy s vysokou presnosťou je nevyhnutné otestovať hypotézu na niekoľkých miliónoch vzorkách, preto sa v súčasnosti vytvárajú možnosti pre hromadné zbieranie a uchovávanie údajov vo forme biobáňk či centier pre biotechnologické informácie na národnej, kontinentálnej alebo svetovej úrovni. Pri spracovaní a analýze takého rozsiahleho množstva údajov sú kritickými aspektmi čas a ľudské zdroje. Práve tu prichádzajú do popredia algoritmy, použitím ktorých je možné celý proces spracovania a analýzy výrazne optimalizovať z oboch spomenutých hľadísk.

Druhým aspektom verejného zdravotníctva, ktoré práve prechádza zmenou je obrat zo zamerania na ochorenia na zameranie na jednotlivca a rozvoj precíznej medicíny, kde je kladený dôraz práve na analýzu rôznych údajov o pacientovi pre určenie tých, ktoré môžu korelovať s diagnostikovaným ochorením. Nadobudnuté dáta sú podľa získaných znalostí nakoniec klasifikované a v konečnom dôsledku rozširujú znalosti v danej oblasti, čím opäť prispievajú k znalostiam podporujúcim prevenciu pred ochoreniami.

Pre správnu identifikáciu prevencie pred ochorením je kľúčové nájsť pôvod ochorenia, zvyčajne na úrovni buniek alebo vnútrobunkových štruktúr a ich zmien. Fotomikroskopické zobrazovanie buniek pacienta je preto v dnešnej dobe obzvlášť v popredí nie len z medicínskej, ale aj z technologickej perspektívy. Inovácie v informačných technológiách z toho dôvodu sú a budú pri správnom použití veľkým prínosom pre rozvoj verejného zdravotníctva. Nielen pre získavanie, uchovanie a správu dát v biomedicínskych informačných systémoch, ale aj pre vytváranie spoľahlivých analýz a záverov z nadobudnutých dát. Príkladom takejto analýzy môže byť automatizovaná identifikácia karcinogénnych buniek vo fotomikroskopickej snímke tkaniva alebo krvi pacienta. Iným príkladom a zároveň zameraním tejto práce je sústredenie sa na spracovanie, analýzu a identifikáciu leukocytov vo fotomikroskopických snímkach.

2.1 Analýza údajov

Vybraným typom údajov pre spracovanie boli fotomikroskopické snímky krvnej vzorky. Pre potreby spracovania a identifikácie leukocytov a ich druhov je potrebné vybraný typ údajov zanalyzovať zo softvérového hľadiska, i z hľadiska mikroskopickej anatómie.

2.1.1 Analýza údajov zo softvérového aspektu

Pre jednotné spracovanie údajov softvérom a možnosť následného porovnania výsledkov a určenie spoľahlivosti je potrebné určiť štandardné vlastnosti vzorky, ktoré budú musieť byť splnené pred samotným spracovaním.

Prvým aspektom je obrazový formát vzorky. Celá vybraná množina údajov musí mať jednotný formát. Mikroskopické snímky sú vo väčšine prípadov štandardného formátu ako napr. JPG, JPEG, TIFF alebo PNG. K snímkam sú v niektorých prípadoch pridružené súbory obsahujúce dodatočné informácie vo formáte XML alebo JSON.

Druhým aspektom je obsah snímky. V závislosti od použitého hardvéru pre zachytenie snímky sa na snímke môže objaviť aj časť okuláru, ktorá je pre nás nežiadúcim prvkom. Preto je možné orezanie snímky tak, aby obsahovala iba približenú časť preparátu. Orezanie je možné vykonať v rôznych štádiách. Existujú mikroskopické zariadenia s priamou podporou orezania, prípadne je možné počítačové orezanie za pomoci softvéru s daným špecifickým účelom, alebo možnosť orezania je taktiež štandardom u nástrojov pre spracovanie a analýzu obrazových snímok.

V súvislosti s orezaním snímky je pre spoľahlivú detekciu nutnosť zadefinovania správneho očakávaného rozmeru bunky, keďže väčšina softvérov očakáva zadanie veľkosti bunky v pixeloch (px).

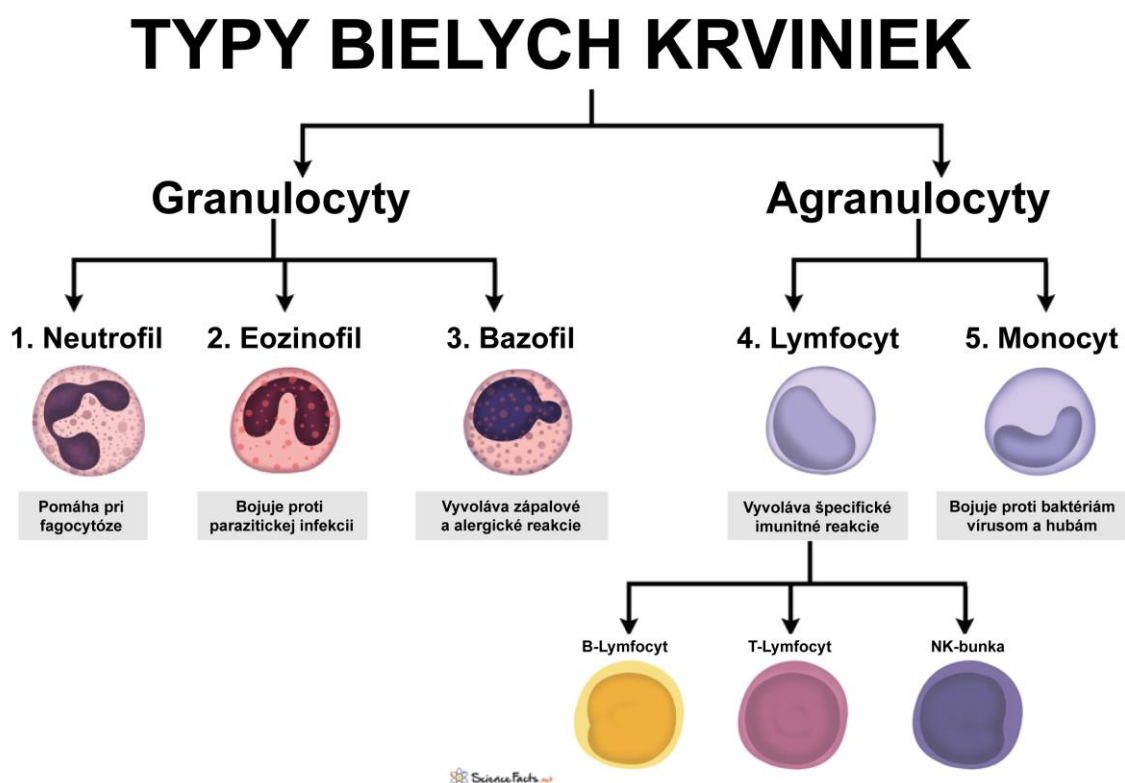
Pred vykonaním analýzy počtu buniek sa na snímkach môže vyskytnúť aj problematika autofluorescencie buniek. Autofluorescencia je vlastnosť buniek, ktorá môže spôsobiť nesprávnu identifikáciu bunky softvérom a skresliť tak vyprodukované výsledky, čo v konečnom dôsledku znižuje spoľahlivosť softvéru. Preto je pri analýze snímky aj potrebné určiť či je na snímke autofluorescencia prítomná, a určiť potrebné predpracovanie snímky za účelom odstránenia alebo zmiernenia autofluorescencie buniek. [54]

Pre niektoré metódy spracovania, ako napríklad prahovanie, môže byť žiadúca aj úprava farebného spektra snímok. Snímky by v ideálnom prípade mali byť zostavené tak, aby sa rovnaké typy buniek zobrazovali v rovnakom zafarbení a bolo tak možné jednoducho určiť o aký typ bunky sa jedná. V prípade že množina údajov je zostavená zo snímok z rôznych zariadení alebo rôznych podmienok, je pravdepodobné, že bude potrebné

určenie jednotnej farebnej škály a následne zvolená technika pre predspracovanie sady snímok tak, aby všetky snímky zodpovedali určenému štandardu. Technikou by mohla byť napríklad korekcia alebo normalizácia farieb snímky.

2.1.2 Analýza údajov z aspektu mikroskopickej anatómie (histológie)

Na vytvorenie preparátu krvnej vzorky je použitý tenký náter odobratej krvi na podložnom sklíčku zafarbený špeciálne pre potreby mikroskopického vyšetrenia konkrétneho typu krviniek. Na zachytenej snímke preparátu môže byť pozorované väčšinové množstvo erytrocytov a zriedkavejší výskyt trombocytov a leukocytov, ktoré sú našim zameraním. Leukocyty sú pravými bunkami – obsahujú jadro a ostatné bunkové organely, ktoré je možné na snímkach pozorovať, analyzovať a klasifikovať tak typ leukocytu. Základnými typmi leukocytov sú granulocyty a agranulocyty, podľa obsahu granúl v cytoplazme. Jadro granulocytov je členité a v cytoplazme sa nachádzajú špecifické granuly. Delenie granulocytov sa ďalej vetví na neutrofilné granulocyty, eozinofilné granulocyty a bazofilné granulocyty. Agranulocyty sa delia na lymfocyty a monocyty. Typy bielych krviniek sú zobrazené na obrázku 7. [56], [57]



Obrázok 7 - Typy bielych krviniek [58]

Neutrofilné granulocyty tvoria približne 60% leukocytov. Je možné ich identifikovať vďaka veľkému počtu granúl fialovej farby v cytoplazme. Neutrofilny majú veľkosť 12-14 mikrometrov (μm). Obsahujú jedno jadro, ktoré môže mať dva až päť segmentov. V krvi

sa nachádza najväčší počet neutrofilov s trojsegmentovaným jadrom. V jadre sa nachádza kondenzovaný chromatín. [56], [57]

Eozinofilné granulocyty tvoria len 1 až 3% leukocytov. Vo vyvinutom štádiu obsahujú dvojlaločné jadro a veľké množstvo charakteristických červených až červenofialových granúl. Granule sú acidofilné a obsahujú toxické proteíny. Priemerná veľkosť eozinofilov je 12 – 17 μm . [56], [57]

Bazofilné granulocyty tvoria približne iba 0,5% všetkých leukocytov obsiahnutých v organizme. Jadro bazofilov je zvyčajne dvojsegmentové a v cytoplazme obsahujú veľké granuly ktorých sfarbenie je tmavomodré až fialovo-čierne. Bazofily majú štandardný priemer 14 – 16 μm . Granule bazofilov obsahujú histamín, heparín, serotonín, či leukotriény. [56], [57]

Lymfocyty tvoria približne 24 – 40% z celkového počtu leukocytov. Veľkosťou sa podobajú skôr erytrocytom – väčšina lymfocytov má priemer 6 – 9 μm . Len približne 10% lymfocytov má priemer 10 – 14 μm . Jadro lymfocytov je malé a obsahuje tmavý chromatín. Menšie lymfocyty majú iba malé množstvo cytoplazmy sfarbenej do bledomodra až fialova. [56], [57]

Monocyty tvoria len 3 - 8% leukocytov organizmu. Monocyty sú charakteristické rozsiahlou plazmatickou časťou a veľkým obličkovitým alebo laločnatým jadrom. Jadro monocytov je umiestnené excentricky. S priemerom až 20 μm sa veľkosťou zaraďujú medzi najväčší typ leukocytov. Z veľkej časti sa skladajú z cytoplazmy ktorá obsahuje veľmi jemné ružové až fialové granule. [56], [57]

2.1.3 Spôsobu identifikácie buniek

Obrazové snímky sú po spracovaní analyzované z obsahovej stránky. Pre potreby práce je cieľom primárne odlíšenie a identifikácia leukocytov od erytrocytov a trombocytov. Analýza leukocytov môže mať viacero rôznych úrovní, v závislosti od obsahu snímky.

Pokiaľ snímka zachytáva väčšinu plochy preparátu, zachytených leukocytov bude pravdepodobne niekoľko. V takomto prípade je predmetom skúmania počet leukocytov, spolu s veľkosťou plôch jednotlivých leukocytov, alebo veľkosťou plochy ako celku pre daný typ bunky. Táto identifikácia je potrebná napríklad pre diagnostiku leukocytyózy alebo leukopénie. U zdravého dospelého jedinca je bez ohľadu na pohlavie počet leukocytov $4 - 10 \cdot 10^9$ v krvi s objemom 1 liter. Leukocytyóza značí nadmerné množstvo leukocytov v periférnej krvi v dôsledku nadmernej tvorby leukocytov v kostnej dreni (dreňová leukocytyóza) alebo v dôsledku vzniku malígnych nádorov, bakteriálnych infekcií, či intoxikácií (distribučná leukocytyóza) Leukopénia, naopak, značí pokles

množstva leukocytov. Leukopénia sa môže prejaviť pri hladovaní alebo pobyte v chladnom prostredí. V prípade hlbšej analýzy je možné identifikovať samotné typy leukocytov, kde sa následne môže sledovať pomer granulocytov ku agranulocytom alebo sledovanie nadbytku, prípadne nedostatku leukocytov konkrétneho druhu pre diagnostiku ochorení ako napr. eozinofília, lymfocytóza či neutropénia. [57]

Iným prípadom sú obrazové údaje, ktoré zachytávajú iba jednu bunku požadovaného typu. Takéto snímky vznikajú zvyčajne ako druhá úroveň spracovania, po prvotnej, už spomenutej identifikácii buniek určeného typu orezaním snímky tak, aby obsahovala iba jednu bunku daného druhu. Vo vytvorených snímkach sa následne aplikuje jednobunková analýza pre účely diagnostiky, ktorá pozostáva primárne z kvalitatívneho hodnotenia konkrétnej bunky. Kvalitatívne hodnotenie môže zahŕňať napríklad analýzu granúl a iných vnútrobunkových organel, či iné sledovanie vnútorných abnormalít, ktoré spôsobujú poruchy mechanizmu pre usmrcovanie baktérií alebo poruchy funkcie lyzozómov. Dôsledkom spomenutých abnormalít môže byť napríklad chronická granulomatózna choroba (neschopnosť usmrtenia niektorých mikroorganizmov fagocytmi) alebo Chediakov-Higashiho syndróm (porucha vyprázdňovania lyzozómov), ktoré majú za následok oslabenie imunitného systému organizmu a vo väčšine prípadov spôsobujú úmrtie jedinca v skorom veku. [57], [59]

Pri oboch spomenutých úrovniach spracovania je možné pre lepšiu identifikáciu použiť kontrastné látky. Aplikujú sa dva typy kontrastných látok na báze fluorescenčných prvkov, ktorým je zvyčajne fosfor (P). Prvý typ látok sa aplikuje výlučne na patologické vzorky, pretože je pre organizmus toxický. Druhý typ kontrastných látok je možné aplikovať na živé bunky, pretože nemajú škodlivé vlastnosti a fluorescencia je dočasná. Aktuálne existuje viacero komerčných produktov s rôznymi vlastnosťami ktoré sú špecifické pre konkrétne aplikácie. Pri identifikácii buniek je potrebné použiť mikroskop vybavený špeciálnym laserom alebo filtrami, ktoré pri rôznych vlnových dĺžkach osvetlenia zachytávajú rôzne typy fluorescencie. Spracovanie fluorescenčných snímok sa od bežnej techniky spracovania líši tým, že je potrebné snímky spracovávať po jednotlivých farebných kanáloch. [60], [61]

2.2 Prehľad techník spracovania a analýzy obrazových údajov

Spracovanie obrazu s cieľom získania kvantitatívnych a kvalitatívnych hodnôt pozorovaných buniek môže mať niekoľko fáz. V závislosti od požadovaného výstupu je možné urobiť iba časť postupnosti fáz, ktorá zabezpečí dosiahnutie požadovaného výsledku.

Za prvú fázu môžeme považovať segmentáciu obrazu pre identifikáciu jednotlivých buniek. Druhá fáza nadväzuje na obraz spracovaný prvou fázou a zahŕňa kvantifikáciu vlastností buniek. Poslednou, treťou fázou je možnosť sledovať a porovnávať identifikované vlastností buniek navzájom a vyvodenie záverov. Pre realizáciu spomenutých fáz sa štandardne používajú nižšie popísané metódy a platformy, v minulosti hlavne manuálneho charakteru. V súčasnosti sú rozvíjané najmä softvérové techniky analýzy a to primárne pre ich výhody akými sú zníženie trvania spracovania a analýzy z časového hľadiska, ale aj odstránenie chybovosti ľudského faktora. [54]

2.2.1 Manuálna optická analýza

Proces manuálnej optickej analýzy zvyčajne vykonáva laboratórny pracovník s potrebným vyškolením pre nadobudnutie znalostí pre rozpoznávanie typov a vlastností pozorovaných buniek a štruktúr. Prvotne sa analýza realizovala priamym pozorovaním preparátu v mikroskope a ručným písomným značením počtu a vlastností buniek. Výhodou takejto metódy je napríklad to, že okrem mikroskopu nie je potrebný dodatočný hardvér pre spracovanie, či pamäťové médium pre ukladanie snímok. Daná metóda má však podstatne viac nevýhod, napríklad nemožnosť replikácie a neskoršej opätovnej analýzy konkrétneho záberu, je časovo náročná, vysoko závislá na ľudskom faktore a môže spôsobiť únavu a poškodenie zraku pracovníka.

Neskôr sa s využitím informačných technológií manuálna optická analýza začala vykonávať zachytením fotografických snímok pozorovaného preparátu a následnou manuálnou analýzou týchto snímok. Pri takejto analýze laboratórny pracovník postupne prechádzal zachytené bunky a za pomoci dostupných grafických nástrojov počítača vyznačoval a spočítaval zachytené bunky a ich podstatné znaky. Hoci sa takýmto spôsobom vyriešil problém replikácie, spomenutá metóda je stále časovo náročná a závislá od schopností a chybovosti človeka.

2.2.2 Softvérové techniky analýzy snímok

V tejto časti sú spomenuté niektoré aktuálne používané a rozvíjané techniky pre softvérovú automatizáciu analýzy biomedicínskych obrazových údajov, čo prispieva k urýchleniu a zjednodušeniu procesu medicínskej diagnostiky.

2.2.2.1 Prahovanie (Thresholding)

Prahovanie je najčastejšie používanou technikou segmentácie obrazových biomedicínskych údajov. Používa sa odlíšenie objektov záujmu, napríklad leukocytov, od pozadia a iných nezaujímavých štruktúr prítomných na obrazovej snímke. Podstatou

metódy je na základe úrovne kontrastu určiť prahovú hodnotu pre farebný alebo 8-bitový obraz tak, aby pri binarizácii obrazu došlo k farebnému odlíšeniu podstatných pozorovaných častí snímky, zobrazených napríklad bielou farbou od nepodstatných, zobrazených čiernou farbou, alebo naopak. Algoritmy pre určenie prahovania môžu byť založené napríklad na informácii o tvare histogramu odtieňov sivej prítomných v obraze, na základe lokálnych charakteristík, či na zhlukovaní. Niektoré softvéry a knižnice pre spracovanie obrazových dát ponúkajú okrem metódy prahovania s vlastnými parametrami aj možnosť využitia metódy pre automatické prahovanie. Príkladmi často používaných metód prahovania sú napríklad Otsu, IsoData, Huang alebo Shanbhag. [62]–[64]

Príkladom nástroja pre identifikáciu typov buniek a analýzu bunkových štruktúr a ich procesov pomocou techniky prahovania, aj v spojení s použitím fluorescenčných látok je open-source nástroj Cell-ID predstavený v [65]. Príkladom použitia je spracovanie a identifikácia zmien bunkových štruktúr kvasiniek a buniek ľudských embryonálnych obličiek (HEK 293) prezentované v článku [66].

2.2.2.2 Detekcia hrán (Edge detection)

Detekcia hrán je tiež technikou používanou pri obrazovej segmentácii. Môže byť použitá samostatne pre identifikáciu hrán a odlíšenie rôznych objektov alebo štruktúr zobrazených na snímke, alebo môže byť použitá napríklad spoločne s metódou prahovania. Detekcia hrán pre oddelenie rôznych objektov používa zmeny intenzity na identifikáciu kontúr objektu, napríklad identifikáciu tvaru buniek. Proces predprípravy snímky pre detekciu obrysov môže zahŕňať zaostrenie a vylepšenie či lokáciu objektu na snímke. Algoritmom pre detekciu hrán je prvotne zo snímky odstránený šum. Následne sú aplikované spomenuté vylepšenia obrazu, ktorý je nakoniec prevedený na binárny obraz. Štandardne používanými algoritmami pre detekciu hrán sú napríklad Canny, Sobel, Laplacian alebo Prewitt. [67]

CellProfiler je štandardne používaný softvér pre odlíšenie jadra a tela buniek v obrazových biomedicínskych snímkach, ktorý na rozlíšenie štruktúr používa práve algoritmy pre detekciu hrán. Príkladom použitia je identifikácia zmien tvaru či lokácie bunkového jadra pred a po aplikácii liečebnej procedúry. [68]

2.2.2.3 Watershed

Technika s názvom watershed sa taktiež používa pri segmentácii obrazu pri identifikácii počtu objektov prítomných na snímke, pre oddelenie prekrývajúcich sa objektov, akými môžu byť napríklad prekrývajúce sa bunky. Watershed sa zvyčajne používa po

aplikovaní transformácie obrazu na binárny. Použitie techniky watershed môže teda spresniť detegovaný počet objektov vytvorením hrán medzi viacerými prekrývajúcimi sa objektami, u ktorých sa binarizáciou spôsobilo zjednotenie do jedného objektu. Algoritmus techniky watershed najprv každému pixelu priradí intenzitu pixelu na základe jeho vzdialenosti od pozadia obrazu. Následne sú identifikovaním lokálnych miním z intenzít označené stredy objektov, ktoré v topografickej reprezentácii predstavujú údolia postupne zaplňané vodou. V bodoch, kde pri zaplňaní dôjde k stretnutiu dvoch nádrží sú identifikované hranice medzi dvomi prekrývajúcimi sa objektami. V softvérovom spracovaní môže byť vodná hladina reprezentovaná maticou, pomocou ktorej sú matematickými operáciami vyhodnotené súradnice hraníc. [69]

Príkladom softvéru implementujúceho techniku watershed je open-source projekt PhyloCell vytvorený v jazyku MATLAB. PhyloCell okrem segmentácie a analýzy obrazovej snímky umožňuje aj kvantitatívnu analýzu časozbernej sekvencie snímok. Príkladom použitia spomenutého softvéru s metódou watershed je segmentácia mitochondriálnych štruktúr počas procesu delenia kvasinkových buniek. [70]

2.2.2.4 Grafové rezy (Graph cuts)

Grafové rezy sú ďalšou technikou pre segmentáciu obrazových údajov. Pomocou týchto rezov je možné v obrazovej snímke identifikovať hranice regiónov ktoré zhlukujú pixely s podobnými vlastnosťami do oblastí, napríklad oddelenie popredia snímky od pozadia. Podstatou algoritmu pre vytvorenie grafových rezov je vytvorenie ohodnoteného grafu kde uzly reprezentujú pixely, a ohodnotenie hrán reprezentuje odlišnosť pixelov na základe príznakov ako napríklad farba alebo intenzita. Následne je vybraný východzí a koncový uzol, ktoré reprezentujú popredie a pozadie snímky. V ďalšom kroku je každému uzlu priradená značka podľa minimálnej vzdialenosti k počiatočnému alebo koncovému uzlu s prihliadnutím na minimalizovanie rozdielov medzi susednými uzlami. Značky uzlov sa postupne aktualizujú podľa susedných uzlov až kým nedôjde k ustáleniu značiek. Značky pixelov reprezentujú či pixel, resp. oblasť pixelov patrí do popredia alebo pozadia snímky, na základe čoho môže byť následne obraz zjednodušený alebo binarizovaný. Pre nájdenie minimálneho rezu alebo maximálneho toku sa používajú rôzne algoritmy ako napríklad Edmondsov-Karpov algoritmus alebo Goldbergov algoritmus (Push Relabel Algorithm). [71], [72]

CellX je open-source softvérový nástroj pre segmentáciu a analýzu buniek na mikroskopických obrazových snímkach, tiež dostupný ako package pre MATLAB. Použitím grafického rozhrania CellX je možná jednoduchá parametrizovaná

segmentácia buniek s využitím grafových rezov, napríklad pre dekódovanie informácií o plazmatickej membráne kvasinkových buniek s odolnosťou voči šumu. [73]

2.2.2.5 Extrakcia kontúr (Contour extraction)

Extrakcia kontúr je metódou segmentácie obrazových snímok pre detekciu uzavretých kontúr a tvaru objektov prítomných na snímke, napríklad určenie kontúr subcelulárnych štruktúr za pomoci viacerých hodnôt prahovania. Predspracovanie obrazu pre použitie spomenutej metódy môže vyžadovať použitie prahovania alebo detekcie hrán. Extrakcia kontúr môže byť realizovaná lokálne, napríklad heuristickým prístupom alebo globálne, napríklad pomocou zhlukovania alebo s použitím aktívnych obrysových (deformovateľných) modelov, tzv. „hadov“. Algoritmus s použitím hada s gradientným vektorovým tokom (GVF) na začiatku vytvára pole síl GVF pre riadenie hada, ktoré ho priťahuje k hraniciam objektu, vytvorí sa mapa hrán za pomoci určenia intenzity pixelov a určí sa počiatočná poloha hada. Následne sa pomocou matematických rovníc iteratívnym prístupom postupne vypočítajú hranice objektu. S vyšším počtom iterácií sa presnosť kontúr zvyšuje, preto je potrebné určiť dostatočný počet iterácií. Často používanými algoritmami pre detekciu kontúr v obrazových snímkach sú už spomenuté techniky aktívnych kontúr, Freemanov (reťazový) kód alebo Douglasov-Peuckerov algoritmus. [74]–[76]

Segmentačný systém PombeX dokáže na obrazových snímkach s gradientom zaostrenia a variujúcim kontrastom segmentovať bunky aplikovať korekcie pre zlepšenie detekcie bunkovej membrány a jadra. Kontúry štruktúr sú následne získané pomocou GVF hada a validované pomocou metód strojového učenia. Príkladom použitia je analýza obrazovej sekvencie správania a zmien v proteínoch pivných kvasiniek „in vivo“. [77]

2.2.2.6 Logistická regresia (Logistic Regression)

Logistická regresia je metóda strojového učenia so supervízorom používaná pre klasifikáciu dát. Cieľom algoritmu je naučiť sa funkciu, ktorá dokáže priradiť vstupným dátam správnu triedu na základe ich vlastností. Napríklad v medicíne sa môže používať na klasifikáciu pacientov do skupiny s ochorením na základe ich zdravotného stavu a symptómov, či predikciu priebehu epidémií. Logistická regresia môže byť binárna, multinomická alebo ordinálna. Zvyčajne sa používa pre klasifikáciu diskrétnych hodnôt, ale môže byť použitá aj pre riešenie problémov regresnej analýzy a predikciu spojitej náhodnej premennej. Binárna logistická regresia využíva logistickú funkciu na predikciu výstupu, ktorá transformuje výsledok lineárnej funkcie do rozsahu $[0, 1]$. Táto funkcia sa využíva na výpočet pravdepodobnosti, ktorá symbolizuje určitosť, s akou vstupné dáta

patria do danej triedy. Údaje sú teda klasifikované do triedy, pre ktorú bola vypočítaná najvyššia pravdepodobnosť zhody. V prípade logistickej regresie môžu byť parametre funkcie učené napr. pomocou metódy maximálnej vierohodnosti (MLE), ktorá maximalizuje pravdepodobnosť, že model správne klasifikuje vstupné dáta. V prípade, že predikovaná trieda môže nadobúdať viac ako 2 hodnoty, používa sa multinomická logistická regresia. [78]–[80]

Weka je open-source softvér pre segmentáciu, klasifikáciu a analýzu obrazových dát, ktorý umožňuje použitie metód strojového učenia, napr. logistickej regresie, na riešenie problémov v mikroskopii. Príkladom použitia nástroja Weka s využitím logistickej regresie je klasifikácia nádorových buniek v obrazových snímkach tkaniva, na základe ich morfológických vlastností. Weka umožňuje aj použitie ďalších techník strojového učenia, ako je stroj s podpornými vektormi (SVM) alebo náhodné lesy (RF). [81], [82]

2.2.2.7 Stroje s podpornými vektormi (Support vector machines)

Stroje s podpornými vektormi je technika strojového učenia so supervízorom, ktorú je možné použiť pre segmentáciu, klasifikáciu alebo regresnú analýzu obrazových dát, napríklad pre oddelenie konkrétneho typu bunky od komplexného pozadia, či klasifikáciu fázy bunkového cyklu. Zámerom algoritmu spomenutej techniky je v N-rozmernom priestore určiť hyperrovinu, ktorá oddelí a klasifikuje údaje, v našom prípade pixely na obrazovej snímke, na základe maximálnej vzdialenosti. Podporné vektory sa získavajú spojením najbližších dvojíc bodov na opačných stranách hyperplochy s cieľom maximalizovať šírku medzi podpornými vektormi, čo zabezpečí istú klasifikáciu na množine pre natrénovanie algoritmu, a dobrú zhodu klasifikácie testovacej množiny s trénovacou množinou. Hyperplocha je daná rovnicou zahŕňajúcou váhový vektor kolmý na plochu, ktorý je možné vypočítať ako lineárnu kombináciu podporných vektorov. Pre klasifikáciu testovacích údajov je použitá matematická funkcia, ktorá určuje, do ktorej triedy testované údaje patria. Algoritmus môže byť lineárny alebo nelineárny s využitím napríklad polynómických funkcií alebo radiálnych bázových funkcií. [83]–[87]

DISCO (Data Informed Segmentation of Cell Objects) je systém pre segmentáciu buniek v mikrofluidických zariadeniach. Príkladom je použitie fyzikálnych a morfológických obmedzení a časových informácií pohybu a rastu buniek pre segmentáciu buniek. Pomocou dvoch klasifikátorov s podpornými vektormi je na snímkach klasifikované popredie a pozadie, či extrakcia príznakov z popredia na určenie jadra a okrajov bunky. [88]

2.2.2.8 Neurónové siete (Neural networks)

Neurónové siete sú technikou hlbokého učenia, používanou podobne ako stroje s podpornými vektormi, pre segmentáciu, klasifikáciu, či regresnú analýzu obrazových údajov, napríklad pre klasifikáciu rôznych typov buniek, alebo tkanivových štruktúr, ale aj na vylepšenie obrazu pre zvýšenie presnosti diagnostiky. Konvolučné neurónové siete a viacvrstvový perceptrón sú príkladmi dopredných neurónových sietí s učiteľom, používanými najmä pre segmentáciu obrazu. Autoenkóder a samoorganizujúce sa mapy sú typom neurónových sietí bez učiteľa, ktoré sa používajú napríklad na extrakciu príznakov alebo odstránenie šumu. Neurónové siete sa skladajú z niekoľkých vrstiev zložených z jednotlivých neurónov. Prvou vrstvou je vstupná vrstva, ktorá prijíma vstupné údaje. Pod prvou vrstvou sa nachádza niekoľko ďalších skrytých vrstiev, ktoré s údajmi vykonávajú rôzne matematické operácie. Úlohou výstupnej vrstvy je vykonanie rozhodnutia, napríklad klasifikácie. Konkrétny postup algoritmu vytvorenia a použitia neurónovej siete je závislý na architektúre a špecifickej úlohe. Všeobecným postupom pri vytváraní neurónovej siete je vytvorenie aproximačného modelu, konfigurácia súboru údajov a nastavenie architektúry. Následne je potrebné tréningovanie neurónovej siete a zlepšovanie výkonnosti a presnosti. Po dosiahnutí dostatočnej presnosti je možné neurónovú sieť použiť pre testovanie na novej množine údajov. [89]–[93]

Cell-ACDC je príkladom modulárneho softvéru vytvoreného v jazyku Python pre segmentáciu a vykonanie poloautomatizovaného procesu anotácie fáz bunkového cyklu neoznačených buniek. Príkladom použitia softvéru Cell-ACDC s použitím dvoch zakomponovaných neurónových sietí YeaZ a CellPose je analýza snímok pučiacich kvasiniek pomocou sledovania fluorescenčného proteínu mKate2 a následný výpočet objemu buniek a zistenie vzájomnej korelácie v objemoch buniek. [94]

2.3 Vyhodnotenie analýzy

Spracovanie mikroskopických údajov krvnej vzorky vyžaduje relevantné znalosti o fyziológii skúmanej vzorky a schopnosť použiť ich v spracovaní. Z pohľadu softvérového spracovania je nevyhnutné aby snímky boli v jednotnom formáte a tiež je potrebné určiť, či je potrebné ich predspracovanie, napr. normalizáciou farieb.

Spracovanie formou manuálnej optickej analýzy je neefektívne, preto volíme ako vhodnejší spôsob použitie softvérových techník. Ako už bolo uvedené, existuje množstvo knižníc a softvérov ktoré už obsahujú bežne používané algoritmy pre spracovanie obrazových údajov, preto nie je nutná ich implementácia.

Softvérové techniky spracovania obrazových údajov je možné rozdeliť do troch kategórií. Prvou kategóriou sú algoritmy pre segmentáciu údajov. Príkladom takýchto techník je prahovanie, detekcia hrán alebo extrakcia kontúr. Druhou kategóriou sú metódy strojového učenia. Medzi metódy strojového učenia zaraďujeme napríklad spomenutú logistickú regresiu a stroje s podpornými vektormi. Treťou kategóriou sú techniky hlbokého učenia. V hlbokom učení sú zaradené rôzne typy neurónových sietí.

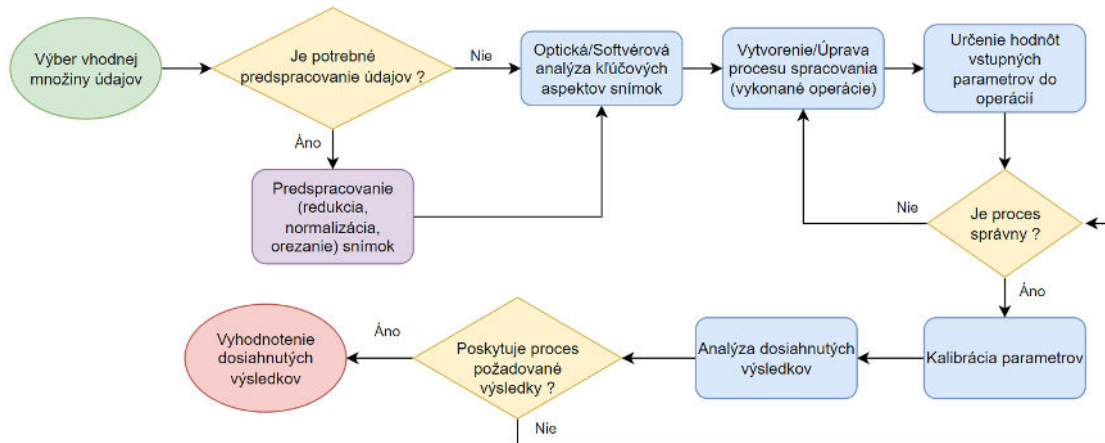
Popísané kategórie je možné klasifikovať ako úrovne, pričom prvá kategória je najjednoduchšia úroveň spracovania. Výber kategórie, ktorou budú snímky spracovávané závisí od požadovaného výstupu. Pomocou prvej úrovne môžu byť v údajoch identifikované oblasti záujmu. Druhá úroveň sa vie na základe poskytnutej už klasifikovanej množiny oblastí záujmu naučiť identifikovať a zaradiť do tried nové údaje. Tretia úroveň sa vie naučiť klasifikovať neoznačené bunky bez akéhokoľvek tréningu.

Všetky spomenuté techniky a úrovne môžu byť aplikované pre spracovanie mikroskopických snímok vzorky krvi. V práci sa zameriame na použitie vybraných techník a softvérov z prvej a druhej úrovne. Výber techník bol vyhodnotený analýzou často používaných metód a ich výsledkov v existujúcich štúdiách ([95]–[97] pre prvú úroveň, [98], [99] pre druhú úroveň). Pre prvú úroveň bude použitý nástroj ImageJ a technika prahovania. Druhá úroveň bude obsahovať použitie softvéru CellProfiler a logistickej regresie. Aplikáciou týchto dvoch úrovní je možné demonštrovať ich rozdiely, možnosti a limitácie.

3 NÁVRH SYSTÉMU PRE ANALÝZU LEUKOCYTOV

3.1 Návrh všeobecného procesu spracovania údajov

Všeobecný proces spracovania mikroskopických údajov pozostáva z niekoľkých za sebou nasledujúcich krokov. Postup procesu je zobrazený na obrázku 8.



Obrázok 8 - Diagram všeobecného procesu spracovania údajov

Proces spracovania obrazových snímok začína výberom vhodnej množiny údajov. Množina by mala byť jednotná ako celok, teda snímky by mali byť rovnakého formátu, nasnímané v rovnakých svetelných podmienkach s použitím jedného mikroskopu s jednotným priblížením a zachytené kamerou s rovnakými nastaveniami. Výhodou by bol výber množiny údajov, ktorá má už expertmi identifikované alebo klasifikované leukocyty, pre možnosť overenia spoľahlivosti vybraných implementovaných techník. Návrh procesu výberu vhodnej množiny údajov bude bližšie popísaný v kapitole 3.2.

Pri vytváraní množiny údajov môže dôjsť k pochybeniu laboratórneho pracovníka alebo použitej hardvérovej techniky. V tom prípade je ďalším krokom v procese spracovania je pedspracovanie údajov. V množine dát sa môžu nachádzať snímky poškodených alebo znehodnotených preparátov, ktoré je potrebné odstrániť. Ďalej sa môžu objaviť rozdiely vo farebnej škále snímok, napríklad z dôvodu nepresnosti pri príprave preparátu, alebo v zmenách pri automatickom nastavení kamery. Taktiež môže byť v závislosti od použitej techniky spracovania potrebné množinu údajov orezať, aby sa zo snímky odstránila zachytená plocha tubusu. V prípade že sú všetky snímky v poriadku, tento krok môže byť preskočený.

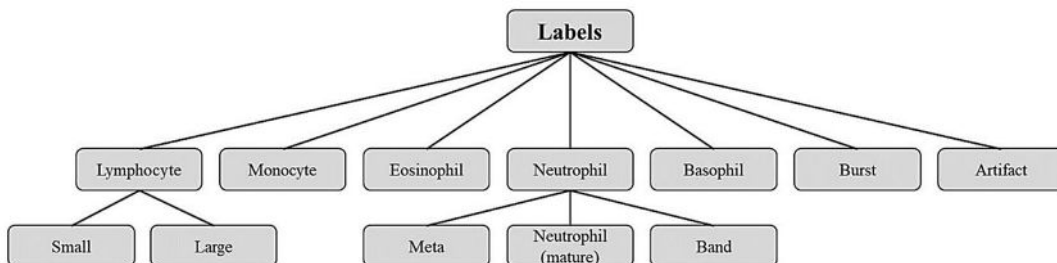
Ďalším krokom je optická alebo softvérová analýza kľúčových aspektov snímky, na základe ktorých je možné identifikovať hľadané objekty na snímke. Kľúčovými aspektmi môže byť napríklad priemer alebo rozloha bunky, tvar bunky, farba, intenzita a podobne.

Následne je možné pomocou vykonanej analýzy stanoviť proces spracovania mikroskopických snímok a určiť hodnoty ich parametrov, potrebné pre identifikáciu leukocytov na snímkach.

Po spracovaní je potrebné overiť správnosť implementovaného procesu a v prípade jeho správnosti vykonať kalibráciu vstupných parametrov. Nakoniec je potrebné manuálne skontrolovať dosiahnuté výsledky a ich uspokojivosť. V prípade že výsledky nie sú dostačujúce, je nutné sa vrátiť k prehodnoteniu aplikovaného procesu spracovania snímok, prípadne ku kalibrácii parametrov procesu.

3.2 Návrh výberu a predspracovania údajov

Po zvažovaní rôznych možností množín údajov (napr. LISC [100], BCCD [101], iné voľne dostupné údaje), bola za najvhodnejšiu zvolená množina Raabin-WBC [102]. Dôvodmi výberu boli najmä veľký počet snímok, rôzny počet leukocytov na snímkach (pre účely identifikácie počtu leukocytov na snímke), a anotácie pre snímky vo formáte JSON, ktoré obsahujú polohu a typ buniek identifikovaných špecialistami v danej oblasti. Anotácie klasifikujú typy leukocytov zobrazené na obrázku 9. Anotácie môžu byť použité pre overenie správnosti obrazového spracovania, ale aj tréning a overenie spoľahlivosti techniky strojového učenia.



Obrázok 9 - Diagram označení buniek v množine údajov Raabin-WBC [102]

Vybraná množina údajov obsahuje vzorky zo 73 filmov periférnej krvi. Krvné preparáty boli zhotovené použitím roztoku GIEMSA. Obrazové snímky boli zhotovené so 100 násobným zväčšením mikroskopmi Olympus Cx18 a Zeiss. Snímky boli zachytené telefónmi Samsung Galaxy S5 a LG G3 a sú vo formáte JPG s rozlíšením 3120x4160. Celkovo je v množine obsiahnutých 3609 lymfocytov, 795 monocytov, 10 862 neutrofilov, 1066 eozinofilov a 301 bazofilov.[102]

Hoci je množina veľmi rozsiahla (približne 40 000 snímok), obsahuje snímky z rôznych mikroskopov zachytené rôznymi kamerami. Snímky sú v rámci mikroskopov ďalej rozdelené do rôznych albumov. Väčšina albumov okrem obyčajných snímok obsahuje aj snímky s použitím vysokého dynamického kontrastu (HDR). Preto bude potrebné vybrať

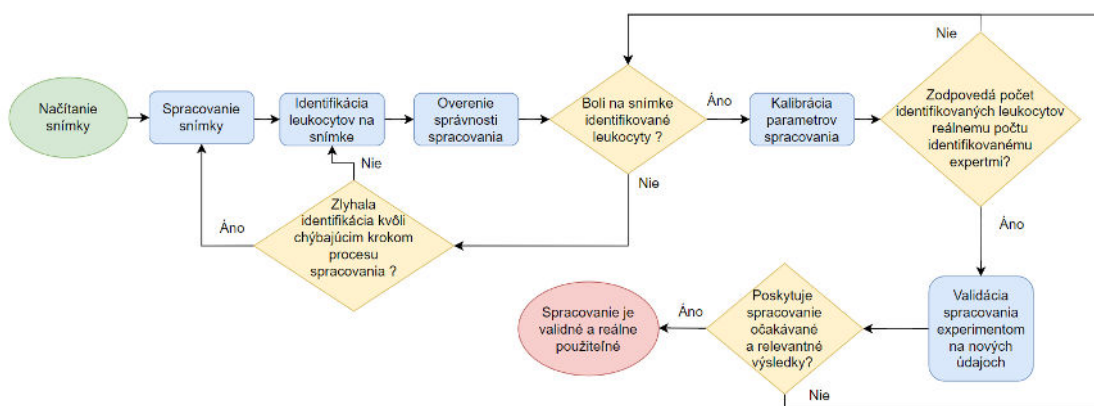
jednotnú podmnožinu údajov. V albumoch sa nachádzajú aj poškodené snímky, ktoré bude potrebné vynechať.

Navrhovaným použitím je výber množiny s použitím druhého mikroskopu (Zeiss mikroskop). Táto množina obsahuje albumy Album 4, Album 5, Album 6, cml-baso1.3, sent, sent2 a sentremained1. Snímky v albume cml-baso1,3 sú od ostatných albumov výrazne odlišné, preto navrhujeme spojenie a použitie všetkých albumov okrem spomenutého albumu. Albumy obsahujú väčšie množstvo snímok bez použitia HDR technológie, preto navrhujeme použiť iba snímky ktoré nie sú označené ako HDR. Na záver bude potrebné všetky vybrané snímky manuálne opticky skontrolovať a odstrániť poškodené snímky.

Po definovaní množiny údajov pre potreby práce, môže byť potrebné predspracovanie údajov. Predspracovanie údajov zahŕňa použitie softvéru pre prácu s rastrovou grafikou, akým je napríklad Adobe Photoshop alebo GIMP. Predspracovanie môže byť použité za účelom zjednotenia a zdôraznenia grafických aspektov obrazu, napr. kontrastu, vyváženia bielej a podobne. Dôvodom použitia predspracovania je najmä kombinácia viacerých albumov s miernymi odlišnosťami grafických vlastností.

3.3 Návrh procesu spracovania a analýzy údajov

Všeobecný návrh postupu spracovania a vyhodnotenia obrazových údajov je popísaný na obrázku 10.



Obrázok 10 - Diagram návrhu procesu spracovania a analýzy údajov

Prvými dvomi krokmi je načítanie a spracovanie snímky vybranou technikou s použitím dostupných nástrojov alebo knižníc. Spracovanie začína načítaním snímky a postupné kroky spracovania a vykonané operácie sa líšia v závislosti od možností, ktoré vybraná technika ponúka, či od cieľa, ktorý jej použitím chceme dosiahnuť. Návrh postupnosti

operácií spracovania bude preto priblížený samostatne pre vybrané techniky v kapitolách 3.4.1 a 3.4.2.

Ďalším krokom postupu je identifikácia leukocytov na spracovanej snímke. Identifikácia môže byť dosiahnutá rôznymi spôsobmi, napríklad automatizovanými optickými technikami, alebo extrakciou vlastností a príznakov. Výstup identifikácie závisí od stanovených požiadaviek, môže sa napríklad jednať o vypočítanú polohu a plochu bunky, spočítanie počtu buniek na snímke, či určenie konkrétneho typu bunky (napr. lymfocyt, neutrofil) prípadne percentuálnych pomerov daných typov.

Výsledky identifikácie musia následne prejsť verifikáciou. Je potrebné overenie, či nami zvolená implementácia je schopná na obrazových snímkach identifikovať stanovený cieľ. Verifikácia môže byť realizovaná napríklad spustením navrhovaného procesu použitia techniky s predpokladaným výstupom a následnou kontrolou, či výstup zodpovedá očakávaniam. Verifikácia zahŕňa vytvorenie štatistík pre porovnanie výsledkov softvérovej identifikácie s výsledkami daných laboratórnymi pracovníkmi.

Po verifikácii modelu nasleduje kalibrácia parametrov techniky. Pri kalibrácii je potrebné hodnoty parametrov odhadnúť na základe predošlých pokusov, resp. iných dostupných údajov. Kalibrácia zabezpečuje, aby výstupy techniky zodpovedali realite, teoretickým predpovediam a poskytovali tak čo najpresnejšie výsledky.

Ďalším krokom je validácia implementovanej techniky. Pri validácii sa kalibrovaná technika aplikuje na nové údaje. Takýmto použitím sa určí, či sú výstupy techniky relevantné aj na iných obrazových snímkach, než na množine, ktorá sa použila pre vytvorenie a kalibrovanie.

Po validovaní je možné pre získanie presnejších výsledkov vykonať väčšie množstvo simulácií, ktoré zabezpečia relevantnejšie informácie o reálnych získaných výstupoch. Tiež môže byť zanalyzovaný vplyv jednotlivých prvkov na techniku ako celok, a môžu byť identifikované možné rozšírenia.

3.4 Výber prístupov spracovania údajov

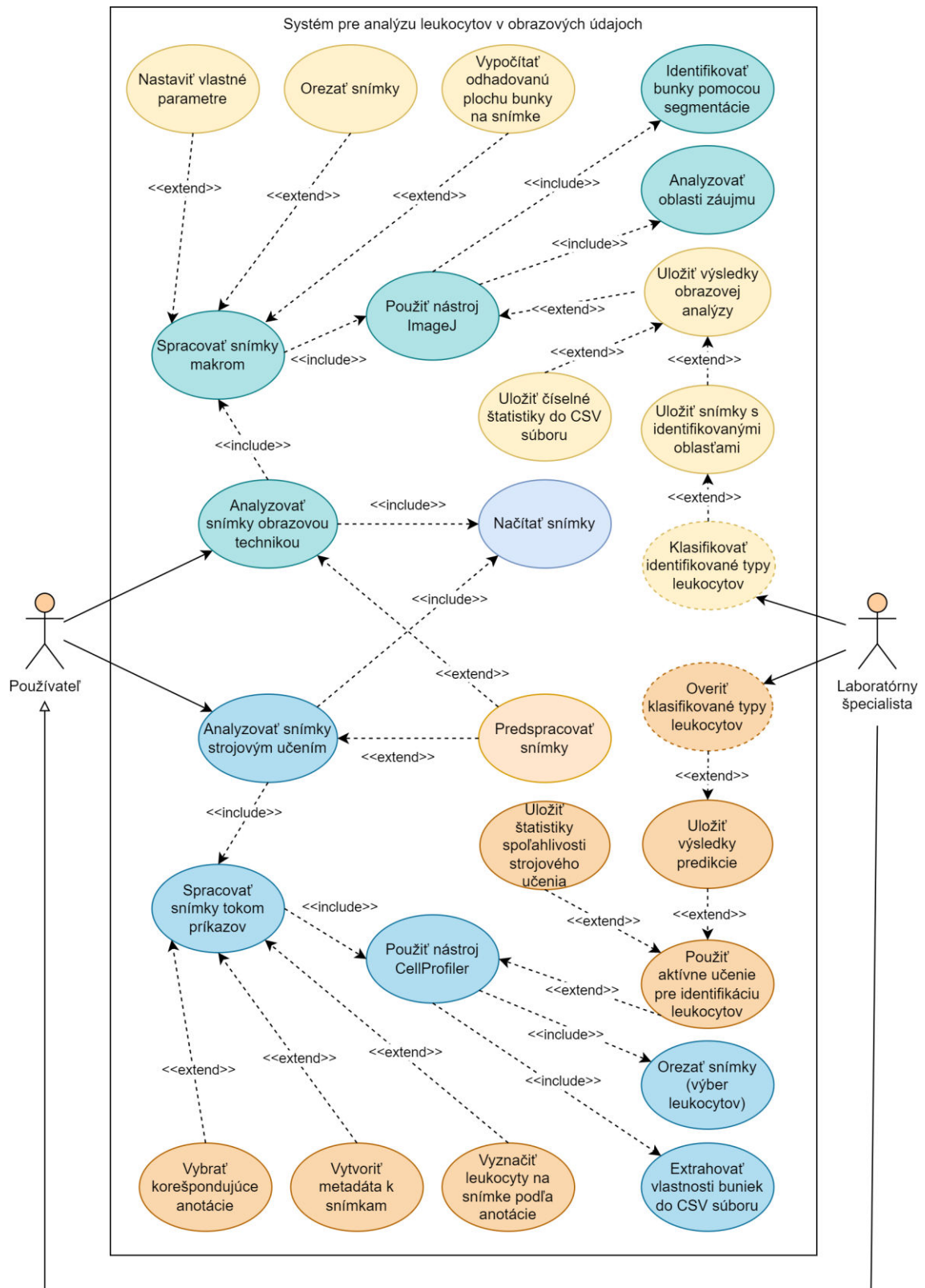
Pre spracovanie mikroskopických obrazových údajov boli vybrané dva prístupy. Prvým prístupom je obrazové spracovanie na základe grafických vlastností obrazu. Druhým prístupom je spracovanie pomocou extrakcie vlastností snímok do matematickej podoby a následné použitie pre model strojového učenia pre spracovanie a klasifikáciu získaných údajov.

Spomenuté prístupy boli vybrané pre zdôraznenie ich kontrastu, na základe odlišných možností, ktoré individuálne ponúkajú, ale aj odlišných nárokov ktoré vyžadujú. Kým pre

použitie obrazového spracovania je postačujúca aj menšia množina vstupných údajov, poskytuje menej možností spracovania a výstupov. Prístup pomocou strojového učenia má zas vyššie nároky na veľkosť množiny, avšak na druhej strane je dostupné veľké množstvo knižníc a štúdií pre spracovanie a podrobnejšiu klasifikáciu údajov, ktorá môže presahovať limity obrazového spracovania.

Okrem individuálneho použitia uvedených techník je možné aj vytvorenie kombinácie – použitie obrazového spracovania a následná aplikácia algoritmov strojového učenia na výstupných dátach z obrazového spracovania. Takéto prepojenie však pravdepodobne nebude plne automatizované a bude vyžadovaná asistencia odborníka.

Obrázok 11 popisuje prípady použitia navrhovaného systému. Aktivity v diagrame ohraničené prerušovanou čiarou sú návrhom pre použite a možnosť rozšírenia systému v budúcnosti.



Obrázok 11 - Diagram prípadov použitia navrhovaného systému

3.4.1 Obrazové spracovanie

Obrazové spracovanie mikroskopických snímok ponúka veľké množstvo dostupných open-source softvérov alebo knižníc, už spomenutých v kapitole 2.2.2. akými sú napríklad ImageJ alebo knižnice open-cv, či jfree. Pre potreby práce bolo zvolené použitie softvéru ImageJ, ktorý obsahuje množstvo implementovaných už spomenutých algoritmických techník pre segmentáciu obrazu. Rozšírenie a automatizácia použitia nástroja ImageJ môže byť implementovaná v jazyku Java, napríklad s použitím doplnkových knižníc napr. jfree.

Použitie softvéru ImageJ je možné automatizovať pomocou vstavanej možnosti makra, ktoré predstavuje kód, primárne v programovacom jazyku Java, kde je možné definovať sekvenciu krokov spoločne s hodnotami, ktoré má softvér vykonať. Takýmto spôsobom je možné stanoviť jednotnú postupnosť krokov segmentácie a analýzy obrazu, ktorá môže byť aplikovaná na snímku, alebo množinu snímok bez potreby manuálneho vykonávania daných operácií. Príkladom operácií je odčítanie pozadia (subtract background), určenie prahovej hodnoty pre vytvorenie binárneho obrazu (threshold), či analýza odlišných častí snímky (analyze particles). Pri volaní makra pomocou dodatočného Java kódu je možné makro upravovať parametricky, a následne spracovávať výstupy, čo pridáva dodatočné možnosti a poskytuje plnú automatizáciu celého procesu spracovania a vyhodnocovania.

Pre vytvorenie a kalibráciu parametrov makra môže byť použitá menšia náhodne vybraná vzorka z celej množiny údajov. Pre overenie správnosti zvolených parametrov bude následne makro otestované na celej množine. Uložené výstupné matematické aj obrazové údaje po spracovaní snímky makrom môžu byť následne načítané a spracované v spomenutej implementácii. Na základe týchto výstupov je možné vyhodnotiť spoľahlivosť identifikácie. Ako metrika pre vyhodnotenie spoľahlivosti môže byť použité napríklad skóre presnosti (accuracy score), určené ako pomer správne identifikovaných leukocytov ku všetkým identifikovaným leukocytom, alebo pomer správne identifikovaných leukocytov k všetkým leukocytom identifikovaným expertom.

3.4.2 Spracovanie pomocou strojového učenia

Pre techniku spracovania obrazových mikroskopických údajov pomocou strojového učenia môžu byť použité rôzne programovacie jazyky, avšak v súčasnosti sú často používanými napríklad Python alebo MATLAB. Model strojového učenia vyžaduje extrakciu vlastností z obrazových údajov, preto sa navrhuje použiť už dostupný softvér s názvom CellProfiler. Predspracovanie údajov, orezanie snímok a vytvorenie metadát potrebných ako vstup pre použitie nástroja CellProfiler, rovnako ako spracovanie

výstupných dát z tohto nástroja, a ich následná aplikácia v modeli pre aktívne strojové učenie a vyhodnotenie výstupov bude implementovaná v programovacom jazyku Python.

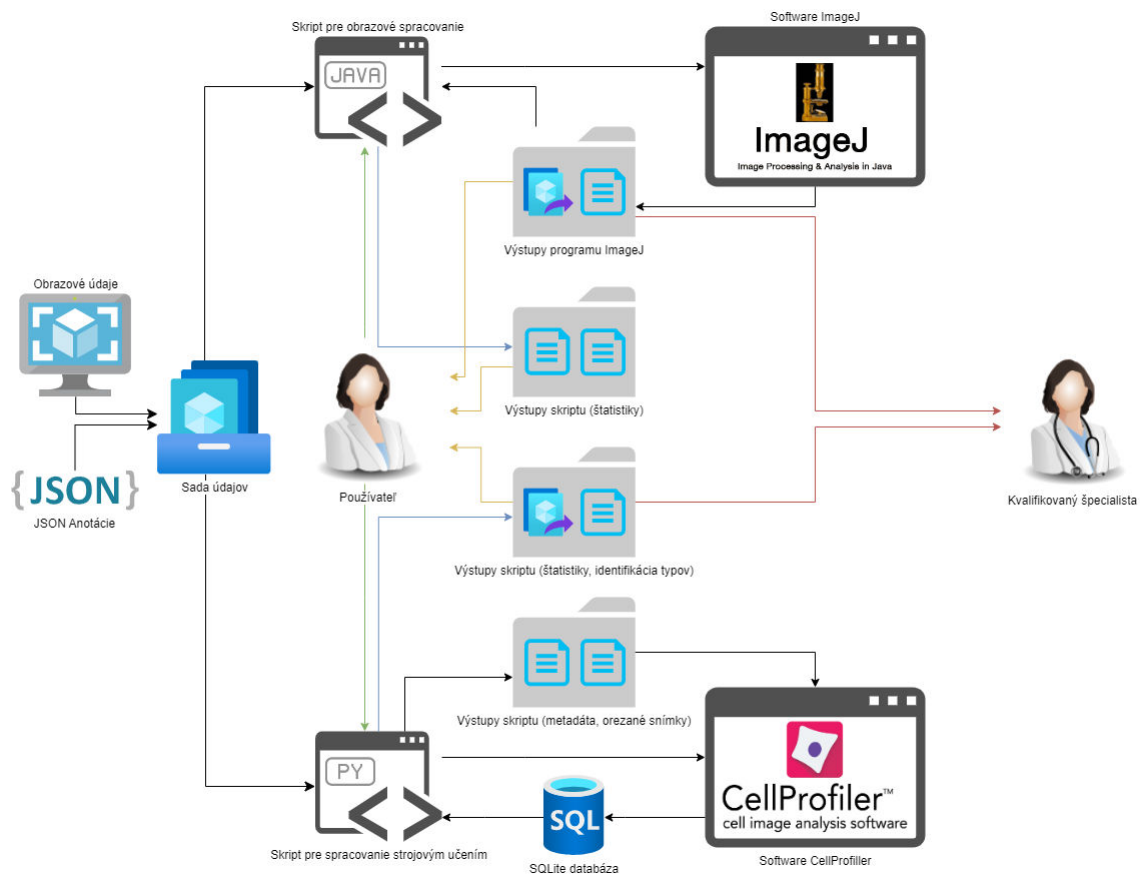
Extrakcia vlastností z obrazových snímok v programe CellProfiler používa tok spracovania príkazov (pipeline). Vstupom do toku príkazov sú obrazové snímky pre spracovanie, a metadáta obsahujúce napríklad určenie výstupnej triedy pre tréning modelu strojového učenia, v prípade že sú dostupné, alebo sa dajú extrahovať z názvov snímok. Do toku je podľa potreby možné vybrať rôznu kombináciu predpripravených algoritmov a operácií pre spracovanie snímok a extrakciu vlastností do matematickej podoby. Príkladom môže byť napríklad identifikácia primárnych objektov na snímke, výpočet rôznych vlastností ich intenzity, textúry, granularity a pod. Predlohou pre zostavenie kľúčovej postupnosti toku bude postup predstavený v článku Nimroda Bermana, v sekcii Extrakcia vlastností buniek pomocou nástroja CellProfiler [103].

Výstupné údaje zo spomenutého nástroja budú ďalej použité pre tréning a testovanie modelu aktívneho strojového učenia.

Aktívne strojové učenie je vylepšenou technikou bežného strojového učenia, v ktorej sa kladie dôraz na identifikáciu minimálneho počtu údajov, potrebných pre natréningovanie modelu. Aktívne strojové učenie je možné v jazyku Python aplikovať na rôzne techniky klasifikácie, ako napríklad logistická regresia, náhodné lesy, či stroje s podpornými vektormi. Spoľahlivosť kalibrovaného aktívneho strojového učenia je možné určiť rôznymi bežne používanými metrikami. Príkladmi daných metrick sú matica chybovosti (confusion matrix), presnosť klasifikácie (accuracy score), skóre F1 (F1 score) alebo plocha pod krivkou (area under curve).

4 IMPLEMENTÁCIA NAVRHNUTÉHO SYSTÉMU

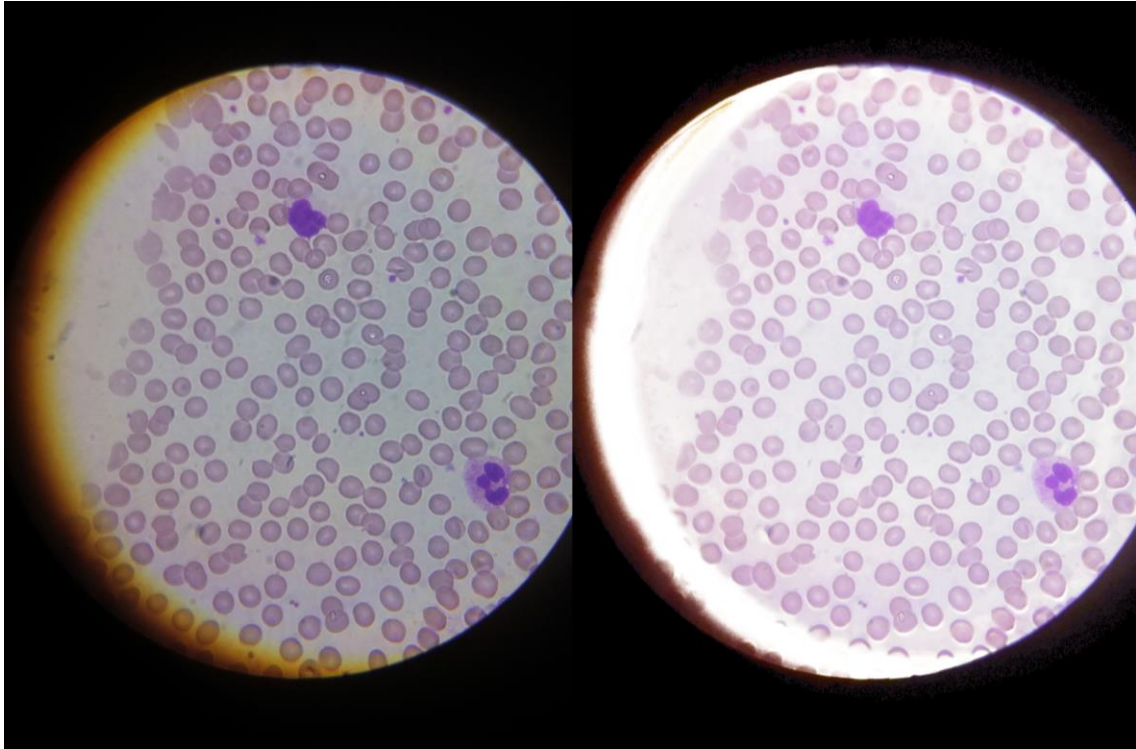
Táto kapitola popisuje detailný postup a popis prvkov, ktoré boli implementované pre vytvorenie systému pre spracovanie, identifikáciu a analýzu leukocytov na biomedicínskych obrazových snímkach. Obrázok 12 graficky popisuje vzájomnú interakciu používateľa a špecialistu s jednotlivými vstupmi, komponentmi a výstupmi systému.



Obrázok 12 - Vzájomná interakcia systémových komponentov s používateľmi

4.1 Výber a predspracovanie údajov

Na celkovú množinu údajov Raabin-WBC bol aplikovaný navrhovaný postup selekcie a predspracovania údajov. Vybraná množina údajov obsahuje 2114 obrazových snímkov s rozlíšením 3120x4160 px. Na novú množinu bolo pri predspracovaní aplikovaných niekoľko akcií. Obrázok 13 zobrazuje pôvodnú snímku a snímku po predspracovaní.



Obrázok 13 - Ukážka pôvodnej snímky (vľavo) a predspracovanej snímky (vpravo)

Prvou akciou bolo potlačenie šumu. Pri potlačení šumu boli nastavenia zamerané na odstránenie farebného šumu a odstránenie artefaktov JPEG, so zachovaním detailov a miernym zaostrením. Druhou operáciou bola identifikácia a nastavenie úrovni snímky pre automatické vylepšenie jasů a kontrastu snímky. V treťom kroku bol dodatočne navýšený jas snímky, pre zosvetlenie farieb. Následne bolo na snímke aplikované vyváženie farieb, primárne pre úroveň stredných tónov s voľbou zachovania svetlosti. Väčšina snímok obsahovala chromatickú aberáciu objektívu mikroskopu v červeno-žltých odtieňoch, ktorá bola pre zlepšenie identifikácie buniek odstránená použitím akcie pre nahradenie farby. Nežiadúce farby boli vo výslednej snímke nahradené bielou farbou. Poslednou operáciou predspracovania bolo uloženie upravenej snímky prepísaním pôvodnej snímky. Presné kroky a hodnoty parametrov sú uvedené na Obrázku 14.

Z postupu vykonaných akcií bolo vytvorené makro, ktoré bolo možné spustiť skúšobnej verzii programu Adobe Photoshop 2022. Program v plnej verzii je možné použiť zakúpením licencie na 1 mesiac. Dostupnou náhradou open-source programu bez potreby zakúpenia licencie je napríklad softvér GIMP, ktorý však ponúka menšie množstvo možností spracovania. Dôvodom použitia softvéru Adobe Photoshop 2022 bolo najmä prívetivé používateľské rozhranie a jednoduchšie vytvorenie makra, než v programe GIMP, čo urýchlilo celý proces predspracovania údajov.

Historie	Akce	Poznámky
✓	<input type="checkbox"/>	▼ wbc_preprocessing
✓	<input type="checkbox"/>	▼ Potlačit šum Barevný šum: 59% Zostřeni: 35% S volbou Odstranit artefakt JPEG s: Parametry kanálu seznam Parametry kanálu Kanál: složený kanál Síla: 9 Zachovat detaily: 52 Parametry kanálu Kanál: červená kanál Síla: 0 Parametry kanálu Kanál: zelená kanál Síla: 0 Parametry kanálu Kanál: modrá kanál Síla: 0 Název nastavení: "Výchozí"
✓	<input type="checkbox"/>	▼ Úrovní Druh přednastavení: Vlastní Nastavení: nastavení úrovní seznam nastavení úrovní Kanál: složený kanál S volbou Automaticky vylepšit jas a kontrast
✓	<input type="checkbox"/>	▼ Jas a kontrast Jas: 23 Kontrast: 0 Bez volby Použít starší
✓	<input type="checkbox"/>	▼ Vyvážení barev Stín - úrovně: 0, 0, 0 Úrovně středních tónů: 13, -14, 6 Úrovně světel: 0, 0, 0 S volbou zachovat světlost
✓	<input type="checkbox"/>	▼ Nahradit barvu Neurčitost: 68 Minimum: barva Lab Světlost: 49,42 Pl: 4,12 B: 12,42 Maximální: barva Lab Světlost: 86,61 Pl: 27,27 B: 54,56 Odstín: -47 S-sytost: -67 L-světlost: 100
✓	<input type="checkbox"/>	Uložit

Obrázok 14 - Sekvencia operácií použitá pre predspracovanie snímok

4.2 Konfigurácie metód spracovania

Konfigurácia metód spracovania popisuje detailné postupy a parametre zvolené pre spracovanie údajov pomocou spomenutých dostupných softvérov ImageJ a CellProfiler.

4.2.1 Spracovanie pomocou nástroja ImageJ

Spracovanie pomocou nástroja ImageJ sa vykonáva spustením makra s definovanou postupnosťou krokov. Jednotlivé kroky makra sú uvedené na obrázku 15.

```

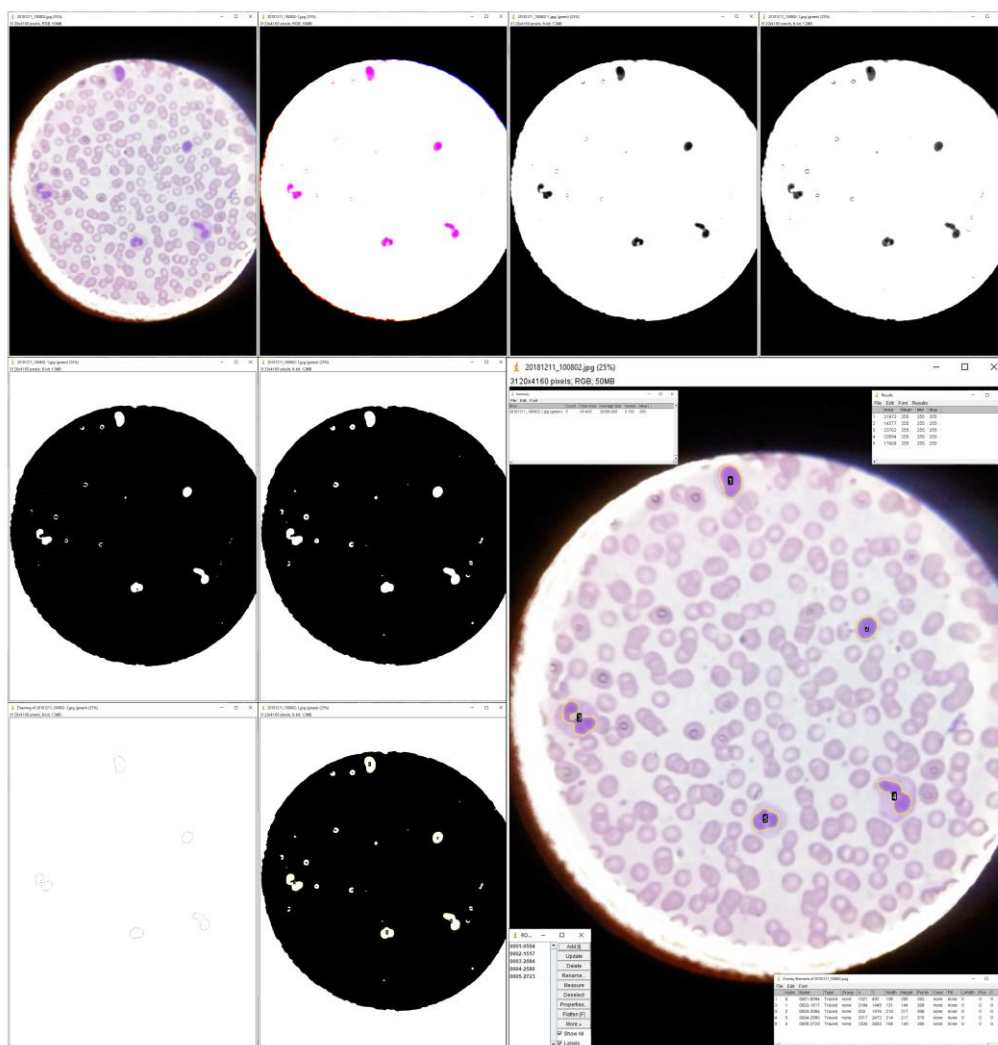
1  imgnm = "fileNameParam";
2  title = imgnm + ".jpg";
3  title1 = imgnm + "-1.jpg";
4  open("sampleDirParam" + title );
5  run("Duplicate...", " ");
6  getStatistics(mean,std);
7  if ((mean > isNotDarkParam && std > isLightSTDPParam) || std >= isDarkSTDPParam) {
8      setMinAndMax(lightMinParam, lightMaxParam);
9  } else if ((mean < isNotLightParam && std < isNeutralSTDPParam)
10     || (mean > isNeutral && std < isDarkSTDPParam)) {
11     setMinAndMax(neutralMinParam, neutralMaxParam);
12 } else {setMinAndMax(darkMinParam, darkMaxParam);}
13 run("Split Channels");
14 selectWindow(""+ title1+" (green)");
15 run("Enhance Contrast...", "saturated=enhContrastParam normalize equalize");
16 run("Auto Threshold", "method=Minimum ");
17 for (i=0; i<dilateCountParam; i++) {run("Dilate");}
18 run("Analyze Particles...", "size=cellSizeMinParam-cellSizeMaxParam circularity=0.10-1.00 \
19     show=Outlines display exclude clear summarize overlay add");
20 selectWindow(title);
21 roiManager("Show All");
22 run("Flatten");
23 saveAs("Png", "outputDirParam/stats/img/"+imgnm+".png");
24 saveAs("Results", "outputDirParam/stats/results/Results_"+imgnm+".csv");
25 roiManager("List");
26 saveAs("Results", "outputDirParam/stats/"+imgnm + ".csv");
27 run("Quit");

```

Obrázok 15 - Ukážka makra použitého pre nástroj ImageJ

Prvým krokom makra je načítanie snímky. Vybrané okno sa zdublikuje, aby bolo možné na novom okne vykonávať transformácie a konečné výstupy zobrazíť na pôvodnom formáte snímky. Ďalej je z histogramu snímky získaný priemer a štandardná odchýlka. Pomocou týchto hodnôt je na duplikáte parametricky nastavený interval svetlosti a kontrastu (riadky 7-12) tak, aby zdôraznil iba časti obrazu ktoré sú farebne podobné odtieňu leukocytov na mikroskopických snímkach. Po nastavení hodnôt intervalu je aplikované rozdelenie farebných kanálov. Ďalej sa pracuje iba so zeleným kanálom, pretože jadrá leukocytov sú v ňom najvýraznejšie. Ďalej je na snímku aplikované zdôraznenie kontrastu s normalizáciou a ekvalizáciou histogramu. Následne je použitá metóda pre segmentáciu obrazu – prahovanie, ktoré oddelí zvýraznené objekty od pozadia snímky. Hodnota prahu je vypočítaná automaticky a zvolenou metódou pre výpočet prahu bola metóda minimálnej chyby („minimal error“). Aplikácia metódy „Dilate“ spôsobí rozšírenie identifikovaných hraníc. Metóda je použitá z dôvodu rôzneho granulovitého tvaru jadier leukocytov. Pri veľmi tenkých spojoch jadier často dochádza k identifikácii lalokov jadier ako osobitných objektov. Operácia „Dilate“ môže v tomto prípade pomôcť rozdelené segmenty opäť spojiť. Po spracovaní snímky spomenutými metódami je zavolaná operácia automatickej analýzy častíc „Analyze particles“. Táto analýza vyžaduje rôzne vstupné parametre ako napríklad požadovaná plocha a kruhová krivka bunky. Vstupné parametre pomohli pri identifikácii odstrániť nesprávne identifikované ako napríklad krvné doštičky alebo oblasti tieňov buniek. Interval plochy bol vopred určený výpočtom s použitím vzorcov pre výpočet obsahu kruhu, prepočtu plochy z jednotky μm do jednotky px a vzorcom pre výpočet mikroskopického zväčšenia. Známymi premennými boli priemer leukocytov (6 - 20 μm), veľkosť pixelu na snímke

(1,12 μm) a použitie 100 násobného zväčšenia. Výpočet bol overený manuálnou selekciou plôch na snímke a použitím funkcie „Measure“ v programe ImageJ. Hodnoty parametra kruhovej krivky boli tiež odhadnuté na základe teoretických vedomostí o tvare jadier leukocytov. Použité možnosti výstupov analýzy sú zobrazenie hraníc, spočítanie či sumarizácia identifikovaných regiónov. Funkcionalita „ROI Manager“ (region of interest) poskytuje detailnejšie informácie o výsledkoch analýzy, napr. súradnice identifikovaných regiónov v dvojrozmernom priestore. Pred použitím daného rozšírenia je nutné vybrať okno s pôvodnou snímku zavolaním metódy selectWindow. Použitím možnosti „Show all“ sa na pôvodnej snímke zobrazia hranice identifikovaných oblastí a metóda „Flatten“ zabezpečí spojenie vrstiev snímky a regiónov do jednej vrstvy. Po zjednotení vrstiev je snímka uložená vo formáte PNG. Číselné výsledky a ich sumarizácia sú uložené do CSV súborov určených pre ďalšie spracovanie. Posledným krokom makra je ukončenie programu. Spracovanie snímky makrom je zobrazené na obrázku 16.



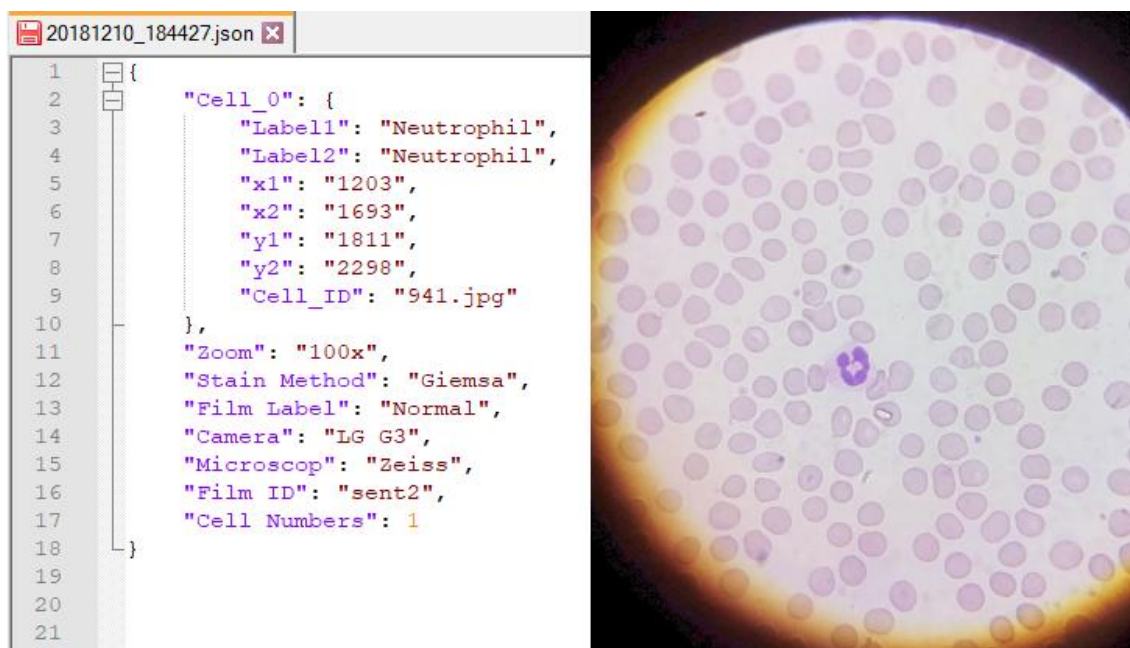
Obrázok 16 - Ukážka postupného spracovania snímky v nástroji ImageJ

4.2.2 Spracovanie pomocou nástroja CellProfiler

Spracovanie snímok pomocou nástroja CellProfiler je vlastne procesom transformácie obrazových dát na textové pomocou identifikácie a výberu kľúčových vlastností snímky. Pre čo najpresnejšiu identifikáciu a následnú predikciu typov leukocytov je potrebné extrahovať vlastnosti nie z celej mikroskopickkej snímky, ale zo snímky, ktorá obsahuje iba oblasť leukocytu. Preto použite tejto techniky potrebné dodatočné softvérové predspracovanie údajov ktoré vyžaduje znalosti o polohe a type bunky. Znalosť typu bunky bude potrebná neskôr, pre naučenie a overenie správnosti modelu strojového učenia, avšak pri použití strojového učenia je vhodné (ale nie nutné) zahrnúť ju už medzi vstupné dáta vkladané do nástroja CellProfiler.

4.2.2.1 Predpríprava vstupných údajov do nástroja CellProfiler

V procese predprípravy boli použité spomenuté anotácie, ktoré boli dostupné v rámci množiny údajov Raabin-WBC. Anotácie sa viažu ku každej snímke osobitne, sú formáte JSON a obsahujú niekoľko údajov, napr. číslo, označenie a súradnice identifikovanej bunky, počet identifikovaných buniek, či informácie o príprave preparátu a použitom hardvéri. Príklad anotácie je uvedený na obrázku 17.



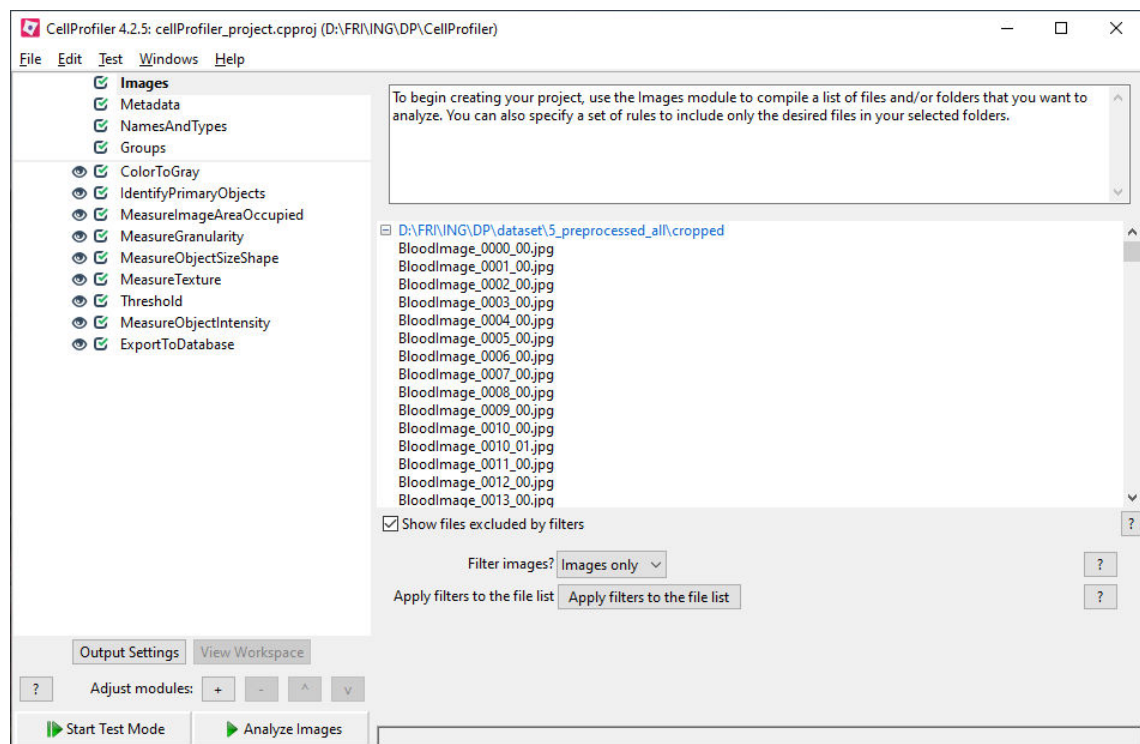
Obrázok 17 - Ukážka JSON Anotácie a príslušnej snímky

Keďže z celkovej množiny údajov bola pre spracovanie pre potreby práce vybraná iba reprezentatívna vzorka, bolo rovnako potrebné vybrať iba korešpondujúce anotácie. Tento výber bol realizovaný skriptom, ktorý bude bližšie popísaný v kapitole 4.3.2. Ďalej bolo potrebné pripraviť výrezy snímok obsahujúcich leukocyty ako primárne objekty, podľa informácií o polohe načítaných z anotácií. Vyrezaným snímkam sa určilo nové

pomenovanie v tvare BloodImage_ID1_ID2.jpg, kde ID1 predstavuje poradové číslo snímky v priečinku a ID2 predstavuje poradové číslo bunky na danej snímke. Pri orezaní snímok boli zároveň vytvorené metadáta obsahujúce názov súboru, identifikačné čísla snímky a bunky na snímke a typ bunky. Detailnejší popis skriptu s orezaním snímok a tvorbou metadát sa nachádza v kapitole 4.3.2.

4.2.2.2 Použitie nástroja CellProfiler

Spracovanie pomocou nástroja CellProfiler vyžaduje vytvorenie toku príkazov. Na obrázku 18 je uvedená postupnosť použitého toku.

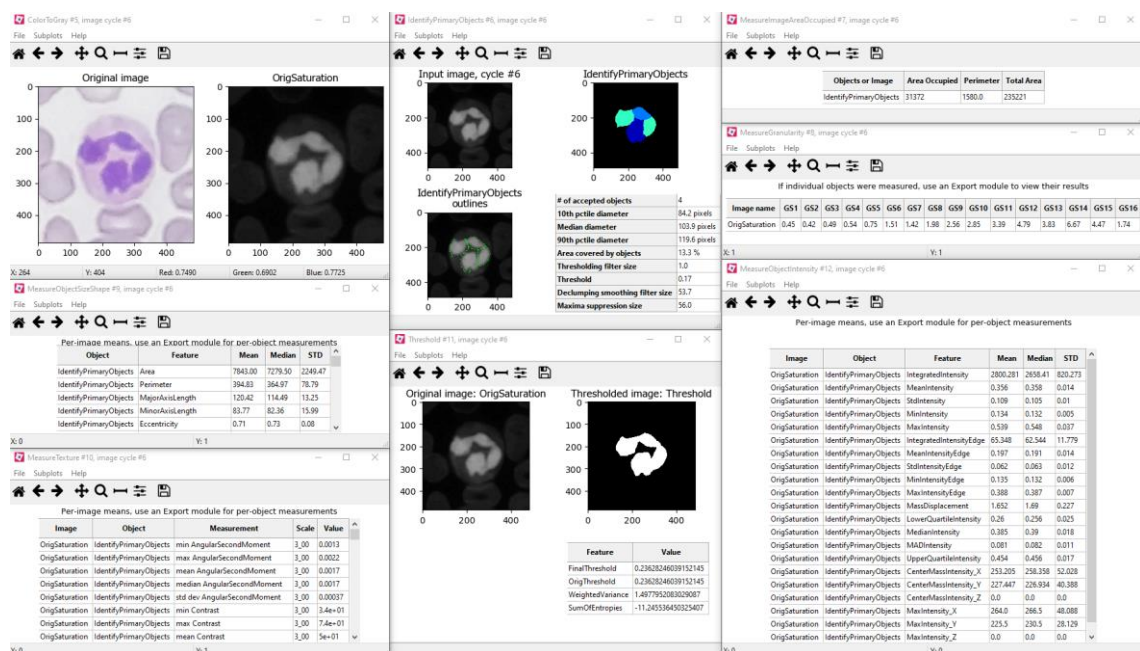


Obrázok 18 - Ukážka použitého toku príkazov pre extrakciu vlastností zo snímok

Do nástroja CellProfiler bolo najprv vložených 2916 orezaných snímok leukocytov, spoločne so spomenutými metadátami. Metadáta a snímky bolo potrebné vzájomne mapovať cez pomocou regulárneho výrazu (RegEx) z názvu snímok. Potrebné bolo aj vytvorenie názvu množiny v sekcii „NamesAndTypes“, pod ktorým sa množina identifikovala v ďalších operáciách.

Následne bolo potrebné zostaviť tok príkazov pre spracovanie množiny údajov. Prvou operáciou je prevedenie farebnej snímky do odtieňov šedej pomocou vstavanej funkcie „ColorToGray“. Na upravenej snímke prebehol algoritmus identifikácie primárnych objektov – identifikácia jadra leukocytu. Aplikácia metódy „MeasureImageAreaOccupied“ zanalyzuje snímku a vráti číselné informácie o rozsahu plochy primárneho objektu na snímke. Granularita bunky prítomnej na snímke je ohodnotená pomocou operácie

„MeasureGranularity“. Na identifikovaný primárny objekt je tiež aplikovaná metóda „MeasureObjectSizeShape“, ktorá vyhodnotí jeho veľkosť a tvar. Ďalšou aplikovanou operáciou s názvom „MeasureTexture“ je číselné vyhodnotenie textúry bunky. Metóda „Threshold“ vyhodnotí hranicu prahovania potrebného pre segmentáciu primárneho objektu. Poslednou operáciou pre extrakciu vlastností je metóda „MeasureObjectIntensity“, ktorá analyzuje a meria intenzitu pixelov primárneho objektu. Po vykonaní všetkých operácií sa výstupné údaje uložia do tabuľky v SQLite databáze, vytvorenej automaticky nástrojom CellProfiler. Údaje kvôli veľkému počtu snímok a vykonaných operácií nástroj CellProfiler nedokázal exportovať priamo do súboru CSV, čo bolo dôvodom použitia databázy. Výstupné údaje okrem vstupných údajov a údajov vrátených spomenutými operáciami obsahujú napríklad aj údaje o čase ktorý zabralo vykonanie danej operácie pre danú snímku a pod. Výstupné údaje boli v poslednom kroku z databázy získané vykonaním príkazu „Select“, ktorý vráti iba požadované stĺpce. Vybrané údaje boli exportované do súboru CSV určeného na ďalšie spracovanie pre účely strojového učenia. Obrázok 19 popisuje spracovanie snímky nástrojom CellProfiler.



Obrázok 19 - Ukážka výsledkov krokov spracovania snímky v nástroji CellProfiler

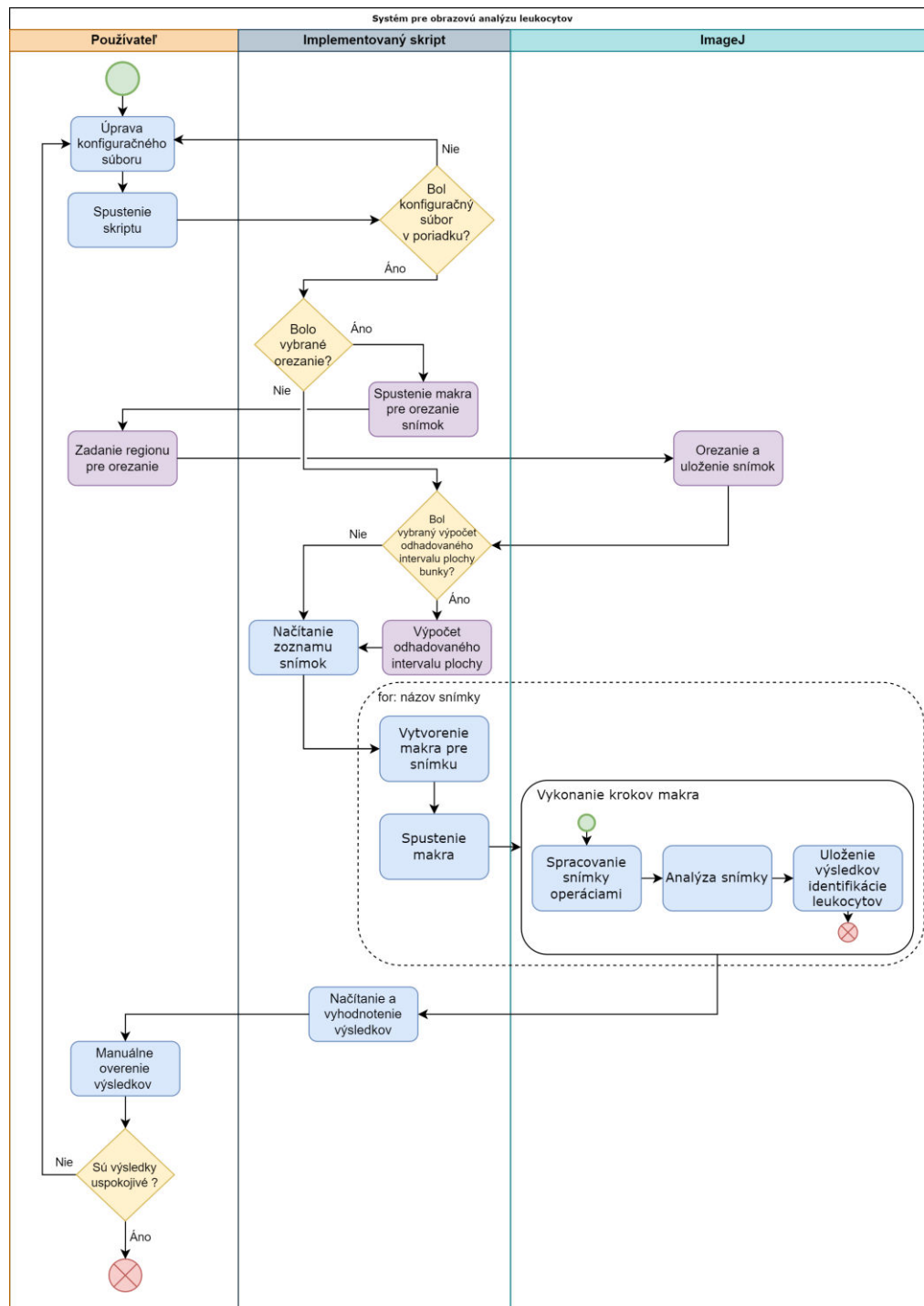
4.3 Implementácia skriptov pre použitie metód spracovania

Proces použitia softvérov ImageJ a CellProfiler vyžaduje okrem ich samotného použitia ďalšie kroky, či už v rámci prípravy vstupných údajov alebo spracovania a analýz výstupných údajov. Väčšinu dodatočných operácií, rutinných prác a tiež použitie techník

je možné automatizovať pomocou kódu. To bolo dôvodom implementácie skriptov, ktoré zabezpečujú zjednodušenie a urýchlenie celého procesu spracovania údajov.

4.3.1 Skript pre automatické obrazové spracovanie s použitím nástroja ImageJ

Skript vykonávajúci obrazové spracovania snímok s použitím softvéru ImageJ bol implementovaný v jazyku Java. Obrázok 20 popisuje diagram aktivít obrazovej analýzy.



Obrázok 20 - Diagram aktivít použitia systému pre obrazovú analýzu

Vstupným argumentom skriptu je konfiguračný súbor, ktorý obsahuje povinné vstupné a výstupné cesty v súborovom systéme, ako aj nepovinné parametre pre manuálne nastavenie parametrov makra. Detailnejší popis konfiguračného súboru sa nachádza v Git repozitári uvedenom v prílohe A.

Na začiatku skriptu bola implementovaná kontrola, či boli zadané všetky potrebné vstupy. Potrebnými vstupmi je cesta ku konfiguračnému súboru, ktorý musí obsahovať cestu k súboru ImageJ.exe, cestu k adresáru obsahujúcemu vstupné dáta a cestu k súboru, kde je potrebné uložiť výstupné dáta. Pokiaľ boli spomenuté hodnoty parametrov v poriadku, skript zo súboru načíta všetky zadané parametre a ich hodnoty a uloží ich do údajovej štruktúry HashMap.

Následne sú pomocou načítaných hodnôt inicializované potrebné premenné, napr. systémový príkaz pre spustenie softvéru ImageJ.

Pokiaľ bol v konfiguračnom súbore zadaná hodnota „y“ pre argument s názvom „crop“, ďalším krokom skriptu je spustenie aplikácie ImageJ s makrom, vytvoreným pre orezanie snímok. Makro pre orezanie obsahuje iba niekoľko krokov. Prvým krokom makra je načítanie snímok. Následne je používateľ vyzvaný pre označenie regiónu, ktorý požaduje zo snímky vyrezať, a operácie „Crop“ a „Image Sequence...“ zabezpečia orezanie a uloženie snímok vo formáte JPG do adresára „CroppedSamples“. Posledným krokom je ukončenie programu ImageJ. Použitie makra pre orezanie snímok nie je povinné a slúži len pre rozšírenie procesu automatizácie o ďalšie možnosti, ktoré môžu byť v určitých prípadoch potrebné. Podmienkou pre správne dávkové orezanie snímok je aby všetky snímky boli zachytené z rovnakého uhla. Pre potreby práce makro nebolo použité práve z dôvodu, že spomenutej podmienke použitá množina snímok nevyhovovala. Z tohto dôvodu makro nebolo popísané v kapitole 4.2.1, avšak bolo vytvorené v rámci implementácie, a je možné ho v prípade potreby použiť.

Ďalším krokom skriptu je načítanie adresára s množinou údajov na spracovanie, pomocou nájdenia cesty k adresáru v mape argumentov. Z adresára sú do štruktúry ArrayList načítané názvy všetkých snímok.

Pred samotným spracovaním snímok softvérom ImageJ musí byť vytvorené už vyššie popísané makro. Pre automatizáciu použitia makra bola vytvorená predloha – „MacroImageProcessingTemplate“, ktorá obsahuje všetky operácie makra popísané v kapitole 4.2.1. Predloha sa odlišuje iba tým, že konkrétne číselné hodnoty a cesty k priečinkom sú nahradené názvami parametrov. Predloha makra je v skripte načítaná do premennej typu StringBuilder.

Podstatou skriptu je postupná iterácia listom načítaných názvov súborov. V každej iterácii sa z premennej s obsahom predlohy makra urobí kópia. V tejto kópii sú následne všetky názvy parametrov nahradené ich skutočnými zadanými hodnotami, napr. cesta, názov spracovávanej snímky, hodnoty plochy bunky a pod. Nepovinné parametre majú v programe preddefinované hodnoty, ktoré boli vybrané na základe analýzy a experimentov. Tieto hodnoty sa použijú v prípade, že používateľ nedefinoval vlastné hodnoty daných parametrov v konfiguračnom súbore. Po vykonaní úprav sa obsah makra zapíše a uloží do súboru „MacroImageProcessing“.

Ďalším krokom je použitie príkazu operačného systému pre spustenie aplikácie ImageJ s vytvoreným makrom. Táto akcia je realizovaná pomocou triedy „ProcessBuilder“, dostupnej v balíčku `java.lang` a spôsobí vytvorenie a beh nového procesu. Aby sa predišlo zahltenu systému, hlavné vlákno skriptu čaká na dokončenie spusteného procesu. Takto sú postupne spracované všetky snímky s už spomenutým ukladaním výstupov. Realizácia spracovania snímok pomocou sekvencie v tomto prípade, po vyhodnotení požiadaviek a dôsledkov, nebola použitá. Hoci výhodou sekvenčného spracovania bolo kratšie trvanie procesu, spracovanie vyžadovalo vysoké pamäťové nároky na použitý hardvér. Hlavnou nevýhodou však bol formát, v akom boli podané výstupné výsledky. Pri sekvenčnom spracovaní snímok softvér ImageJ nedokázal po identifikácii objektov priradiť nájdené regióny k pôvodným snímkam. To znamená, že pokiaľ sa na snímke nachádzali 3 leukocyty, snímka bola uložená 3-krát, pričom na každej bol označený jeden z identifikovaných regiónov. Našou požiadavkou však bola možnosť spočítavania buniek na snímke, preto bolo takéto spracovanie nevyhovujúce a museli sme sa uspokojiť s postupným spracovaním snímok, aj keď daný prístup predlžuje celkový čas spracovania.

Po spracovaní všetkých snímok sú postupne načítané obsahy CSV súborov ktoré vznikli ako výstup z použitia aplikácie ImageJ, spoločne s pôvodnými JSON anotáciami snímok. Keďže obrazové spracovanie často u granulocytov rozdelí jadro na niekoľko častí a počíta s nimi ako s rôznymi oblasťami, pre správne určenie spoľahlivosti je potrebné identifikovať, ktoré oblasti sú v blízkosti (nachádzajú sa v jednej bunke). Pre identifikáciu bol vytvorený algoritmus, ktorý z identifikovaných oblastí vytvoril obdĺžniky a porovnal ich centrá. Pokiaľ je euklidovská vzdialenosť centier menšia ako maximálny priemer leukocytu (20 μm), regióny sú označené ako jedna bunka. Následne je nutné určiť, či boli označené bunky technikou ImageJ identifikované správne. Dané porovnanie je realizované algoritmom, ktorý porovnáva polohy označených buniek na snímke s polohou určenou v anotáciách. Pokiaľ bola označená bunka už nájdená v anotáciách, je označená a započítaná ako správne určená, a ďalej je z porovnania vynechaná. Po

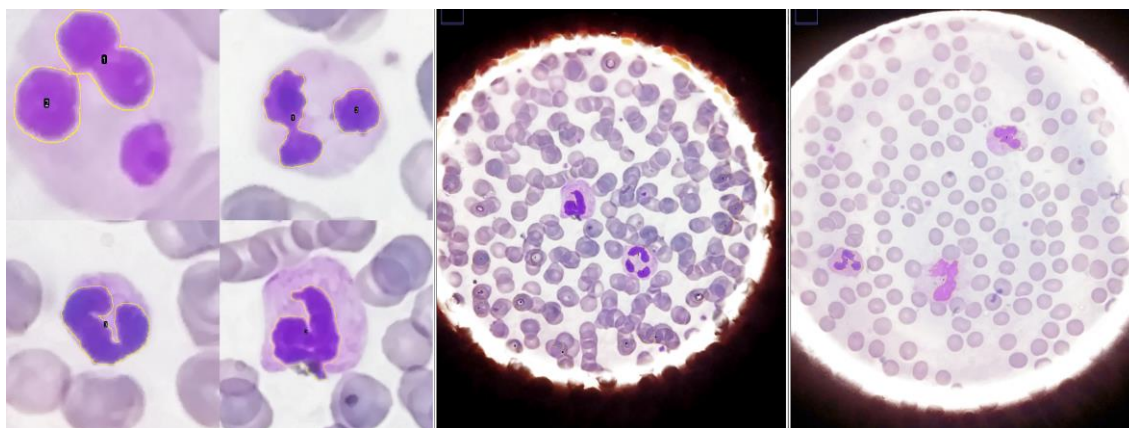
prejdení všetkých správne označených buniek je vypočítaný počet buniek, ktoré boli identifikované technikou, ale nenachádzajú sa v anotáciách (falošné pozitíva) a naopak, počet buniek ktoré neboli identifikované, ale v anotáciách sa nachádzajú (falošné negatíva).

Na záver sú vypočítané štatistiky vyhodnocujúce úspešnosť použitej techniky. Prvou štatistikou je percentuálne vyhodnotenie, koľko snímok bolo vyhodnotených ako skutočných pozitív spolu s falošnými pozitívami (boli nájdené všetky bunky, ktoré boli v anotáciách), falošných pozitív a falošných negatív. Percentuálna spoľahlivosť správnej identifikácie skutočných pozitív je uložená v stĺpcovom grafe. Druhým typom štatistiky sú skladané stĺpcové grafy a CSV súbory, obsahujúce počty všetkých troch spomenutých kategórií pre každú snímku.

4.3.1.1 Výsledky dosiahnuté obrazovým spracovaním

Obrazové spracovanie snímok s použitím techniky ImageJ poskytuje možnosť identifikácie leukocytov na mikroskopických snímkach na základe grafických vlastností snímky.

Prvým z výsledkov je obrazová snímka (obrázok 21), na ktorej sú žltou farbou ohraničené regióny, ktoré boli identifikované ako jadrá leukocytov. Regióny sú v rámci jednej snímky očíslované, pre vyhodnotenie počtu identifikovaných oblastí.



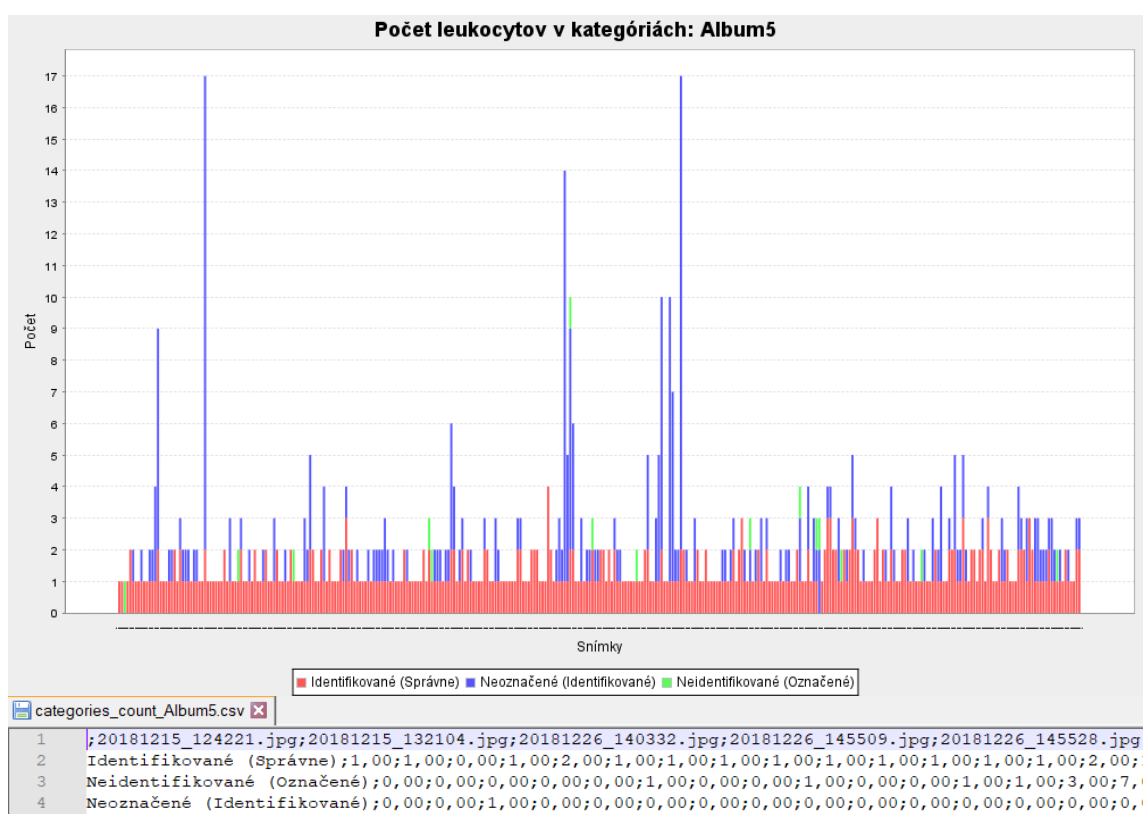
Obrázok 21 - Ukážka identifikovaných oblastí nástrojom ImageJ

Druhým výsledkom je CSV súbor Results_NÁZOV.csv (obrázok 22), kde „NÁZOV“ reprezentuje názov obrazovej snímky. Tento súbor obsahuje číselné vyhodnotenie analýzy snímky. Pre každú identifikovanú bunku obsahuje jej identifikačné číslo, plochu, súradnice bunky v dvojrozmernom priestore a tiež výšku a šírku bunky. Všetky spomenuté údaje sú uvedené v pixeloch.

1	, Index, Name, Type, Group, X, Y, Width, Height, Points, Color, Fill, IWidth, Pos, C, Z, T
2	1, 0, 0001-2064, Traced, none, 1421, 1973, 161, 183, 298, none, none, 0, 0, 0, 0, 0
3	2, 1, 0002-2490, Traced, none, 2204, 2362, 97, 256, 292, none, none, 0, 0, 0, 0, 0
4	3, 2, 0003-3260, Traced, none, 982, 3160, 159, 201, 330, none, none, 0, 0, 0, 0, 0

Obrázok 22 - Ukážka výstupného CSV súboru z programu CellProfiler

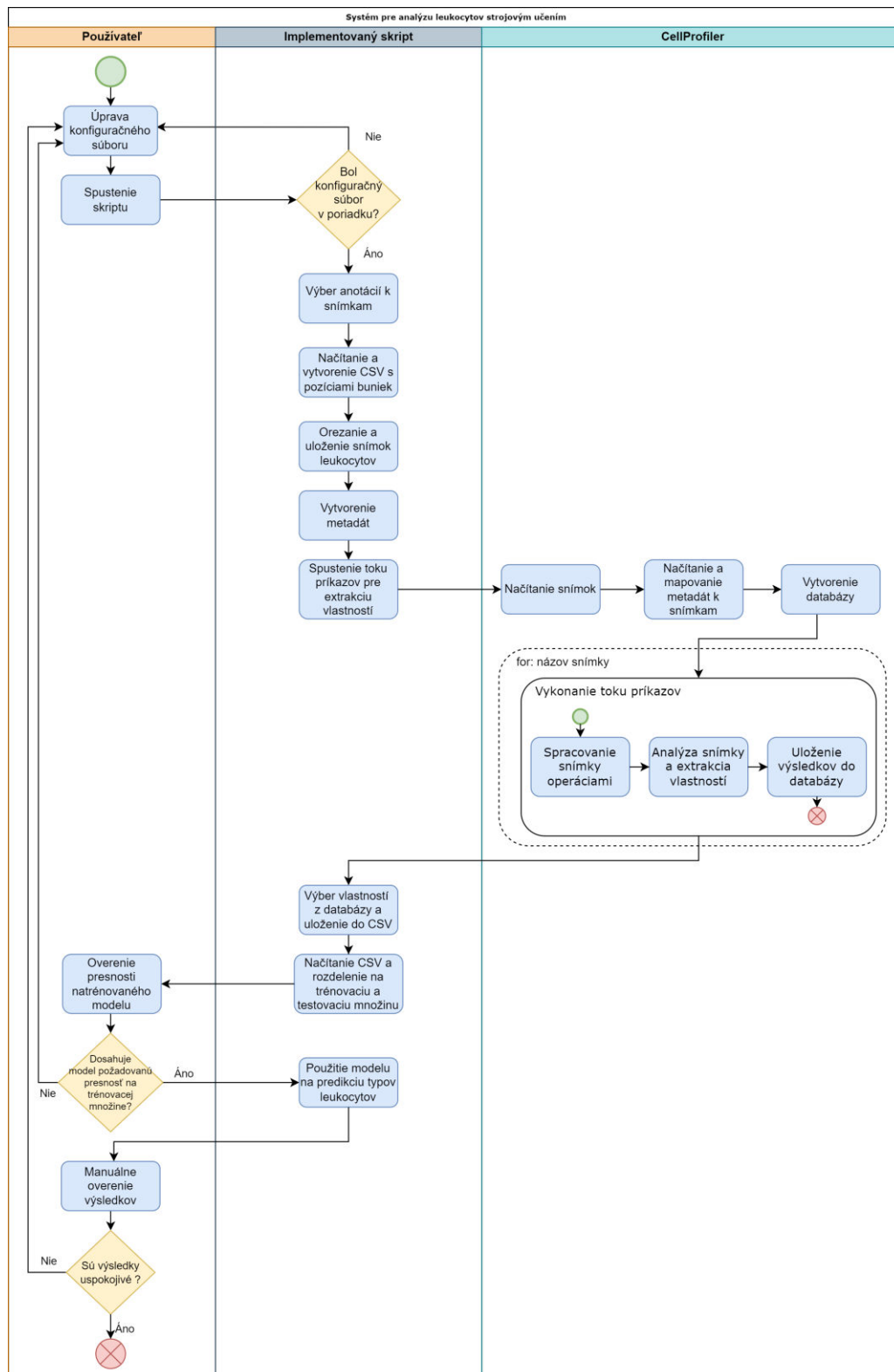
Tretím výsledkom je vyhodnotenie spoľahlivosti identifikácie (obrázok 23). Spoľahlivosť zahŕňa percentuálne vyhodnotenie správne identifikovaných oblastí leukocytov pre všetky snímky aj pre každú snímku osobitne. Pre každú snímku je ďalej vyhodnotený počet správne identifikovaných oblastí, počet identifikovaných oblastí, ktoré sa nenachádzajú v anotáciách a počet oblastí, ktoré neboli identifikované, napriek tomu, že v anotáciách sa nachádzajú. Tieto výsledky sú ukladané CSV súbore a grafoch.



Obrázok 23 - Ukážka výstupného grafu a CSV súboru pre vyhodnotenie spoľahlivosti identifikácie

4.3.2 Skript pre automatické spracovanie strojovým učením s použitím nástroja CellProfiler

Skript pre spracovanie údajov strojovým učením s použitím softvéru CellProfiler bol implementovaný v jazyku Python. Obrázok 24 obsahuje diagram aktivít pri spracovaní snímkov strojovým učením.



Obrázok 24 - Diagram aktivít použitia systému pre analýzu strojovým učením

Vstupným parametrom skriptu je cesta ku konfiguračnému súboru. Tento súbor je tiež popísaný v Git repozitári uvedenom v prílohe A.

Prvým krokom skriptu je použitie funkcie „select_annotations“. Cieľom tejto funkcie je výber anotácií k vybranej množine snímkov. Táto funkcia najprv zabezpečuje načítanie

všetkých názvov snímok zo zadaného adresára do listu, pomocou knižnice os. Rovnakým spôsobom sú načítané JSON anotácie, ktoré boli dostupné k pôvodnej množine údajov. Ako už bolo spomenuté, množina mohla byť zredukovaná, napr. odstránením poškodených snímok. Aby sme používateľa odbremenili od manuálneho odstraňovania nepotrebných anotácií, ďalším krokom metódy je vytvorenie nového priečinka s názvom „new_annotations“ pomocou knižnice shutil. Následne je prejdený celý zoznam názvov priečinkov a pokiaľ sa daný názov nachádza v zozname anotácií, súbor s daným názvom sa skopíruje do adresára „new_annotations“.

Druhou použitou funkciou je „export“. Cieľom tejto funkcie je načítanie anotácií, výber informácií o polohe bunky pre každú bunku v každej snímke, a ich exportovanie do spoločného CSV súboru. Na začiatku sú pomocou funkcie „glob“ z knižnice glob načítané všetky názvy JSON anotácií prítomné v adresári „new_annotations“ a je vytvorený prázdny list, ktorý bude slúžiť na priebežné ukladanie načítaných informácií. Pre zachovanie prehľadnej adresárovej štruktúry je najskôr vytvorený priečinok pre uloženie CSV súboru. Ďalším krokom je iterácia cez list názvov anotácií. V každej iterácii je pomocou knižnice json otvorený a pomocou funkcie „load“ načítaný obsah aktuálne prechádzaného súboru. Použitie knižnice json uľahčuje prácu s obsahom súboru, pretože poskytuje možnosť pristupovať k obsahu súboru pomocou operátora „[]“. V súbore sa nachádza informácia o počte expertmi identifikovaných leukocytov. Táto informácia je načítaná a funkcia pokračuje vnoreným cyklom pre načítanie informácií pre všetky identifikované leukocyty v anotácii. V rámci spomenutého cyklu sa postupne pre jednotlivé súbory zvyšuje premenná „image_id“, ktorá určuje poradové číslo spracovaného súboru. Okrem toho je z názvu súboru odvodený názov snímky, ku ktorej anotácia patrí. Vo vnorenom cykle sú pre konkrétnu bunku načítané x a y súradnice pixelov, ktoré štvorcovcom ohraničujú danú bunku. V rámci vnoreného cyklu sa podobne ako pre súbory, v tomto prípade pre jednotlivé leukocyty, zvyšuje premenná „cell_id“, ktorá určuje poradové číslo bunky na danej snímke. Načítaný je aj typ leukocytu (označený značkou „Label2“), ktorý bol identifikovaný druhým expertom. Všetky spomenuté extrahované údaje sú nakoniec vložené do údajovej štruktúry list. Tento list bude tvoriť jeden riadok výsledného CSV súboru. Riadok je pomocou funkcie „append“ pridaný do listu vytvoreného na začiatku tejto metódy. Po prejdení oboch cyklov je z listu obsahujúceho všetky načítané informácie pomocou knižnice pandas vytvorený objekt typu „DataFrame“ so stĺpcami v rovnakom poradí: filename(názov obrazovej snímky), image_id (poradové číslo snímky), cell_id (poradové číslo bunky na snímke), wbc_type (typ leukocytu), xmin (minimálna súradnica na osi x), xmax (maximálna súradnica na osi x), ymin (minimálna súradnica na osi y), ymax (maximálna súradnica na osi y).

Posledným krokom je uloženie „DataFrame“ tabuľky do CSV súboru, pomocou metódy „to_csv“, s názvom súboru „metadata.csv“ bez možnosti indexovania riadkov (parameter je nastavený na hodnotu False).

Treťou funkciou, ktorá je použitá v procese spracovania údajov je metóda „crop_and_create_metadata“. Cieľom tejto funkcie je načítanie obrazových snímok a orezanie a uloženie oblastí leukocytov na základe CSV súboru vytvoreného predošlou funkciou a doplnenie CSV súboru o názov novej snímky obsahujúcej samostatný leukocyt. Táto časť používa doplnkovú funkciu crop_cell s parametrom „row“, ktorý reprezentuje riadok z CSV súboru. Funkcia crop_cell pomocou knižnice cv2 načíta zo zadaného adresára snímku s názvom, ktorý sa nachádza v stĺpci „filename“. Ďalej sú načítané súradnice osí x a y, na základe ktorých je označená časť snímky. Pre označenú časť je vytvorené nové pomenovanie súboru, vo formáte BloodImage_ID1_ID2.jpg, kde ID1 predstavuje načítanú hodnotu stĺpca image_id vo formáte štyroch číslíc a ID2 je hodnota stĺpca cell_id vo formáte dvoch číslíc. Funkcia crop_cell sa nakoniec pokúsi o uloženie snímky do zadaného adresára pre výstupné dáta. V prípade úspešného uloženia metóda vráti názov uloženej snímky, inak vráti slovo „error“. Funkcia „crop_and_create_metadata“ v prvom kroku vytvorí nový adresár pre zapisovanie orezaných snímok s názvom „cropped“, a povolí práva pre zapisovanie do adresára pre všetkých používateľov. Ďalším postupom je opätovné načítanie spomenutého CSV súboru z predošlej metódy, z ktorého je vytvorený objekt „DataFrame“. Dataframe pre všetky riadky aplikuje funkciu crop_cell a výsledok metódy uloží do nového stĺpca s názvom „cell_filename“ v tabuľke „DataFrame“. Tabuľka je nakoniec opäť rovnakým spôsobom ako v predošlej metóde uložená, prepísaním spomenutého CSV súboru. Takto je k snímkam vytvorený súbor s metadátami.

Ďalšou použitou funkciou je „run_cellprofiler“. Táto funkcia pomocou konzoly spustí súbor s preddefinovaným tokom príkazov v programe CellProfiler. Súbor tiež obsahuje potrebné konfigurácie, či cesty k adresáru s orezanými snímkami a súbor s metadátami.

Po dokončení procesu predošlej funkcie je použitá funkcia „select_metadata“, ktorej úlohou je otvoriť spojenie s databázou pomocou knižnice sqlite3, ktorá obsahuje výstupné dáta z predošlej funkcie. Vykonaním príkazu zo súboru select.sql sú načítané hodnoty z vybraných stĺpcov. Po načítaní údajov sú postupne zapísané do CSV súboru.

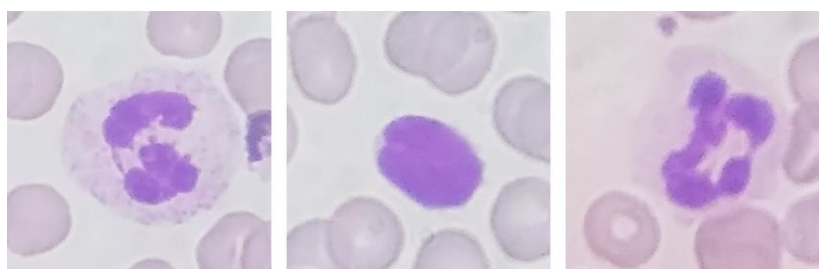
Uložené údaje z databázy sú ďalej spracované funkciou „active_learning“. Na začiatku sú údaje načítané a sú z nich odstránené stĺpce, ktoré nie sú potrebné pre spracovanie alebo riadky, ktoré nie je možné binárne klasifikovať ako granulocyty alebo agranulocyty. Ostatné riadky sú podľa príslušného typu leukocyty zaradené do skupiny granulocytov

alebo agranulocytov, nahradením hodnoty v stĺpci „Image_Metadata_wbc_type“. Hodnoty tejto triedy sú transformované do celočíselného intervalu v rozsahu <0;1> a identifikované ako trieda, ktorá bude predikovaná. Následne je množina rozdelená na dve množiny. Prvá množina (65%) slúži pre trénovanie modelu a druhá (35%) pre vytvorenie predikcií. Množina pre trénovanie je ďalej rozdelená na množiny pre trénovanie a overenie trénovanie modelu v pomere 60% a 40%. Použitým klasifikátorom je logistická regresia, implementovaná v knižnici sklearn. V logistickej regresii boli použité atribúty pre výber algoritmu klasifikácie (saga), váhy pre balancovanie pravdepodobností (podľa veľkosti množín údajov), maximálny počet iterácií (1000) a stratégia dopytovania podľa neistoty („uncertainty sampling“). Logistická regresia je zaobalená použitím objektu „ActiveLearner“ dostupným v knižnici modAL. Ďalším krokom bolo určenie základnej veľkosti trénovacej množiny (2 snímky), ktorá sa v každej ďalšej iterácii zväčšuje. Postupné zväčšovanie trénovacej množiny dáva možnosť určiť jej minimálnu veľkosť, pri ktorej model dosahuje požadovanú presnosť. V každom kroku zväčšovania (epochy) sa pomocou prvej množiny model natrénuje a vyhodnotí spoľahlivosť pomocou metód „fit_transform“, „predict_proba“ a „balanced_accuracy_score“. Po ukončení daného počtu epoch sa model použije pre predikciu na druhej množine s uložením výsledkov do CSV súboru. Celý proces strojového učenia je replikovaný 100-krát a sú ukladané a vypočítané priemerné výsledky štatistík vo forme grafov.

4.3.2.1 Výsledky spracovaním strojovým učením

Spracovanie snímok použitím techniky strojového učenia a nástroja CellProfiler poskytuje možnosť identifikácie typov leukocytov na mikroskopických snímkach obsahujúcich jednotlivé leukocyty na základe extrahovaných matematických vlastností snímky.

Prvým výsledkom spracovania sú orezané snímky leukocytov (obrázok 25) podľa pôvodných JSON anotácií. Anotácie boli súčasťou množiny údajov a na základe buniek identifikovaných expertmi sú z pôvodných snímok vyrezané a uložené označené časti, vhodné pre ďalšie spracovanie.



Obrázok 25 - Ukážka orezaných obrázkov leukocytov z pôvodných snímok

Druhým výsledkom je databáza a výstupný CSV súbor (obrázok 26) z nástroja CellProfiler, ktoré obsahujú extrahované vlastnosti snímok v číselnej forme. Databáza je typu SQLite, z ktorej je exportovaný CSV súbor s vybranými stĺpcami tabuľky.

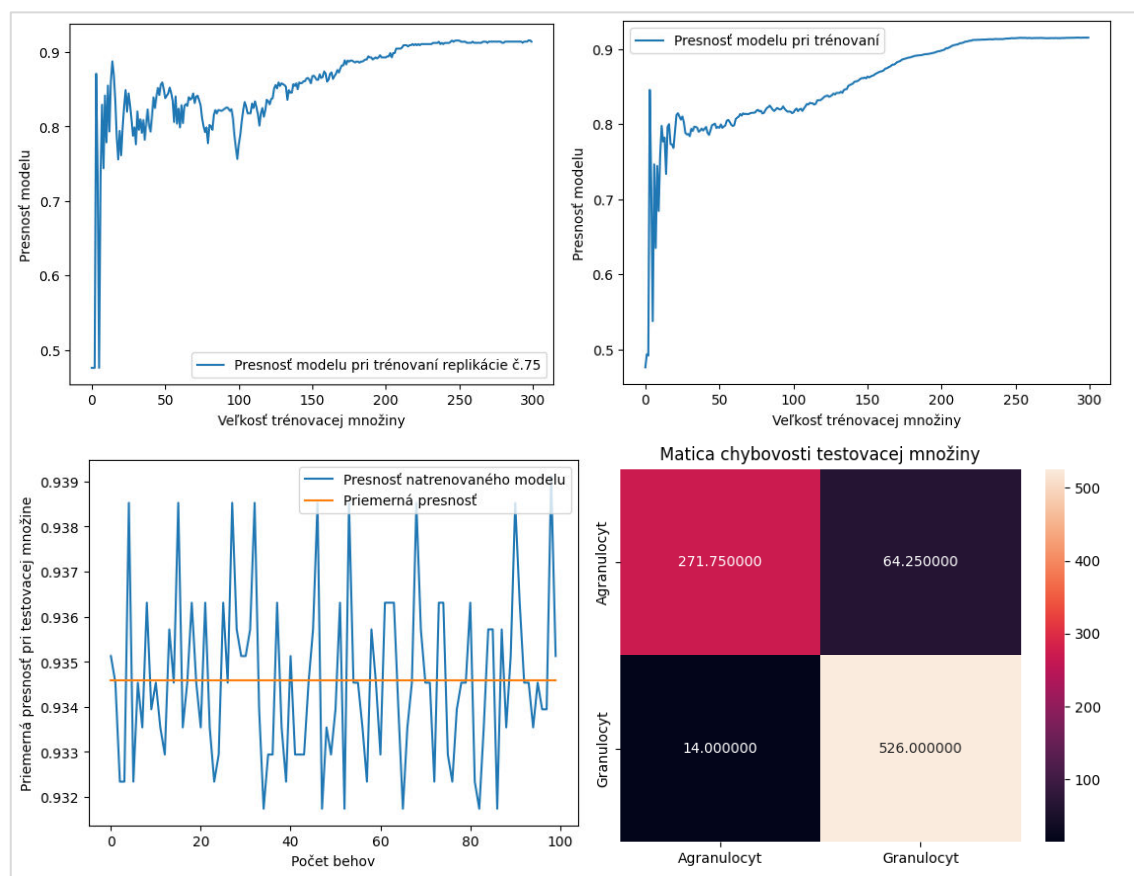
```

1 Image_Metadata_cell_filename,Image_Metadata_image_id,Image_Metadata_cell_id,Image_Metadata_wbc_type,Image_AreaOccupied_AreaOccupied_IdentifyPrimaryObjects
2 BloodImage_0000_00.jpg,0,0,large_lymph,20779.0,550.0,236682.0,1,2,23159749611735,1,60289947757244,0,535973302056277,1,38221423169693,1,32908105678845,1,55
3 BloodImage_0001_00.jpg,1,0,small_lymph,25373.0,645.0,248460.0,1,2,39769131599794,2,15694751686632,0,856656126999701,0,91345289844753,2,6768941469627,0,824
4 BloodImage_0002_00.jpg,2,0,Neutrophil,30255.0,1252.0,236682.0,3,4,11958880242552,5,03287611864479,2,97102823410941,3,8953477131646,3,79281846994205,4,4747
5 BloodImage_0003_00.jpg,3,0,Neutrophil,38449.0,1304.0,237654.0,3,7,3404549549089,5,29918180000716,1,95198516544025,5,69308901729959,4,03395657679976,4,163
6 BloodImage_0004_00.jpg,4,0,Unknn,58724.0,1730.0,236682.0,2,0,770978993501204,1,0183426820548,3,26949429493146,2,60377762944238,4,35262841053003,5,7422825
7 BloodImage_0005_00.jpg,5,0,Neutrophil,31372.0,1580.0,235221.0,4,2,84524872301726,3,38736656990509,4,7850020109445,3,83376034964968,6,67147692294344,4,4710
8 BloodImage_0006_00.jpg,6,0,Neutrophil,42267.0,1825.0,240084.0,3,8,57736147077588,6,3281732522373,5,88079692667603,4,14719436747104,2,06831184581673,0,092
9 BloodImage_0007_00.jpg,7,0,Neutrophil,32402.0,1477.0,237168.0,3,2,59721031050796,5,37614332647943,6,56326722530266,8,83938487624742,2,7451933585993,1,004
10 BloodImage_0008_00.jpg,8,0,Neutrophil,47620.0,2051.0,242064.0,3,7,19007619370626,6,41683259221906,3,2226333527857,0,96436643158549,0,00936018223410916,2,
11 BloodImage_0009_00.jpg,9,0,Neutrophil,30493.0,1581.0,238630.0,4,9,80671726441136,8,56622737859467,7,62492903436921,3,86641002582321,0,910839967213792,1,20
12 BloodImage_0010_00.jpg,10,0,Artifacts,23162.0,990.0,236680.0,2,2,6683819163927,8,34820636442882,11,0505646768118,12,2449042671962,4,0302997275382,3,8421424
13 BloodImage_0010_01.jpg,10,1,Neutrophil,28093.0,1194.0,243522.0,2,1,00503744726309,1,414273217256,1,69446548116075,2,8703951186441,12,3988605559114,10,1013
14 BloodImage_0011_00.jpg,11,0,small_lymph,25231.0,603.0,237169.0,1,2,1981425274665,1,52772749319716,1,57798493448136,0,720131842848962,0,986342264571761,0,3
15 BloodImage_0012_00.jpg,12,0,Neutrophil,36905.0,1979.0,235710.0,4,5,51421820711833,9,94179769070399,3,84078994930553,1,66930626687438,1,01602135496631,1,74
16 BloodImage_0013_00.jpg,13,0,Neutrophil,23907.0,1194.0,238108.0,3,8,20185633021705,6,34940168102487,5,78642409846698,4,42661673491884,2,40063219874435,1,04
17 BloodImage_0013_01.jpg,13,1,Neutrophil,20603.0,956.0,237165.0,2,1,73177504071897,5,30826103035717,10,713531043609,6,31543819985086,6,15645442924396,2,7076

```

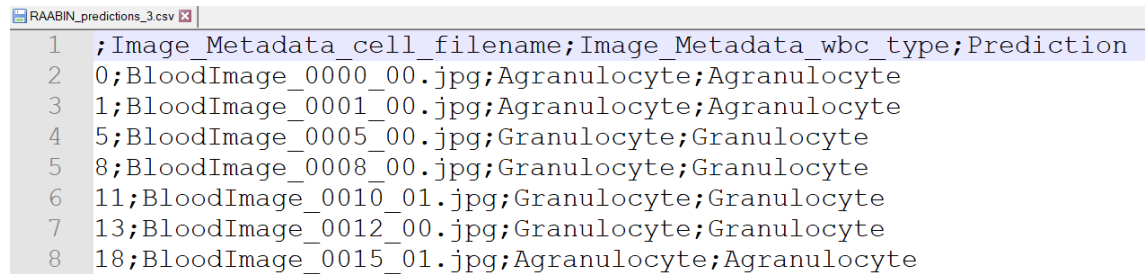
Obrázok 26 - Ukážka výstupného CSV súboru z programu CellProfiler

Tretím výsledkom sú uložené grafy (obrázok 27) a CSV súbory vyjadrujúce spoľahlivosť algoritmu strojového učenia. Pre každú replikáciu sú ukladané čiarové grafy percentuálnej spoľahlivosti predikcie (metrikou vyvážené skóre presnosti) pre jednotlivé veľkosti trénovacej množiny. Pre celkový počet replikácií je okrem grafu ktorý zobrazuje priemer všetkých grafov percentuálnej spoľahlivosti ukladaný aj graf zobrazujúci jednotlivé percentuálne spoľahlivosti a priemernú percentuálnu spoľahlivosť modelu na trénovacej množine. Ďalej je pre replikácie ukladaná priemerná matica chybovosti pre trénovaciu aj testovaciu množinu.



Obrázok 27 – Ukážka výsledkov vyhodnotenia presnosti modelu

Štvrtým výsledkom sú CSV súbory (obrázok 28), s údajmi druhej (testovacej) množiny spojené s predikciou typu leukocyty. Tieto údaje je možné použiť pre ďalšie vytváranie analýz a grafické porovnanie predikovaných klasifikovaných typov leukocytov s označenými.



```
1 ;Image_Metadata_cell_filename;Image_Metadata_wbc_type;Prediction
2 0;BloodImage_0000_00.jpg;Agranulocyte;Agranulocyte
3 1;BloodImage_0001_00.jpg;Agranulocyte;Agranulocyte
4 5;BloodImage_0005_00.jpg;Granulocyte;Granulocyte
5 8;BloodImage_0008_00.jpg;Granulocyte;Granulocyte
6 11;BloodImage_0010_01.jpg;Granulocyte;Granulocyte
7 13;BloodImage_0012_00.jpg;Granulocyte;Granulocyte
8 18;BloodImage_0015_01.jpg;Agranulocyte;Agranulocyte
```

Obrázok 28 - Ukážka CSV súboru označenej a predikovanej kategórie leukocytu v testovacej množine pre 3. replikáciu

5 OVERENIE SYSTÉMU PRE ANALÝZU LEUKOCYTOV

5.1 Spôsob vyhodnocovania implementovaných metód

Vyhodnotenie implementovaných metód spočíva v definícii ich limitácií a spoľahlivosti.

V procese verifikácie boli overené zvolené kroky a celkové výsledky použitých techník. Verifikácia zahŕňala spustenie techník s prípadmi použitia, kedy bol jasný očakávaný výsledok a vytvorenie grafov štatistík. Pre obrazové spracovanie to bolo porovnanie výsledkov softvérovej identifikácie s výsledkami danými laboratórnymi pracovníkmi. Strojové spracovanie bolo vyhodnocované metrikou vyváženého skóre presnosti („balanced accuracy score“).

Validácia implementovaných metód bola realizovaná experimentmi popísanými v tejto kapitole. Pre každú techniku boli vybrané 2 experimenty, na ktorých prebehlo overenie relevantnosti implementácie na rôznych vstupných údajoch.

5.2 Vyhodnotenie aplikovaných metód

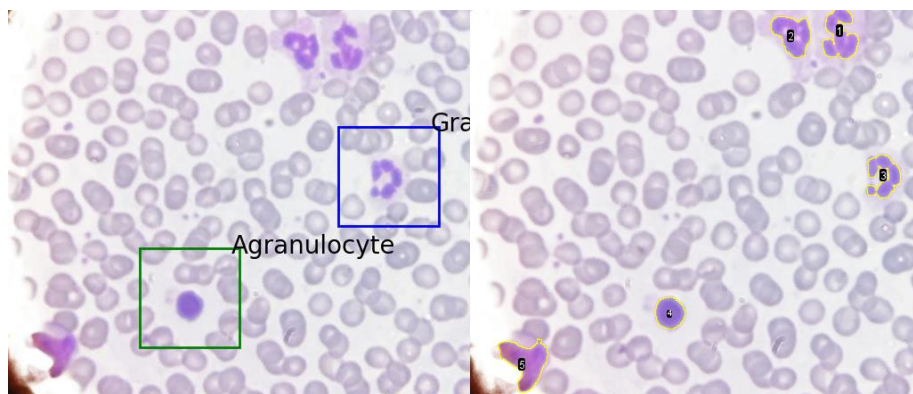
Experimenty sa realizovali na množine údajov Raabin-WBC a BCCD, ktoré boli bližšie popísané v kapitole 3.2. Množina údajov BCCD bola aplikovaná kvôli menšiemu rozsahu údajov len ako doplnkový experiment.

5.2.1 Obrazové spracovanie snímok

Cieľom experimentov bolo overiť spoľahlivosť identifikácie obrazového spracovania údajov na priemernej a veľkej variabilnej množine údajov.

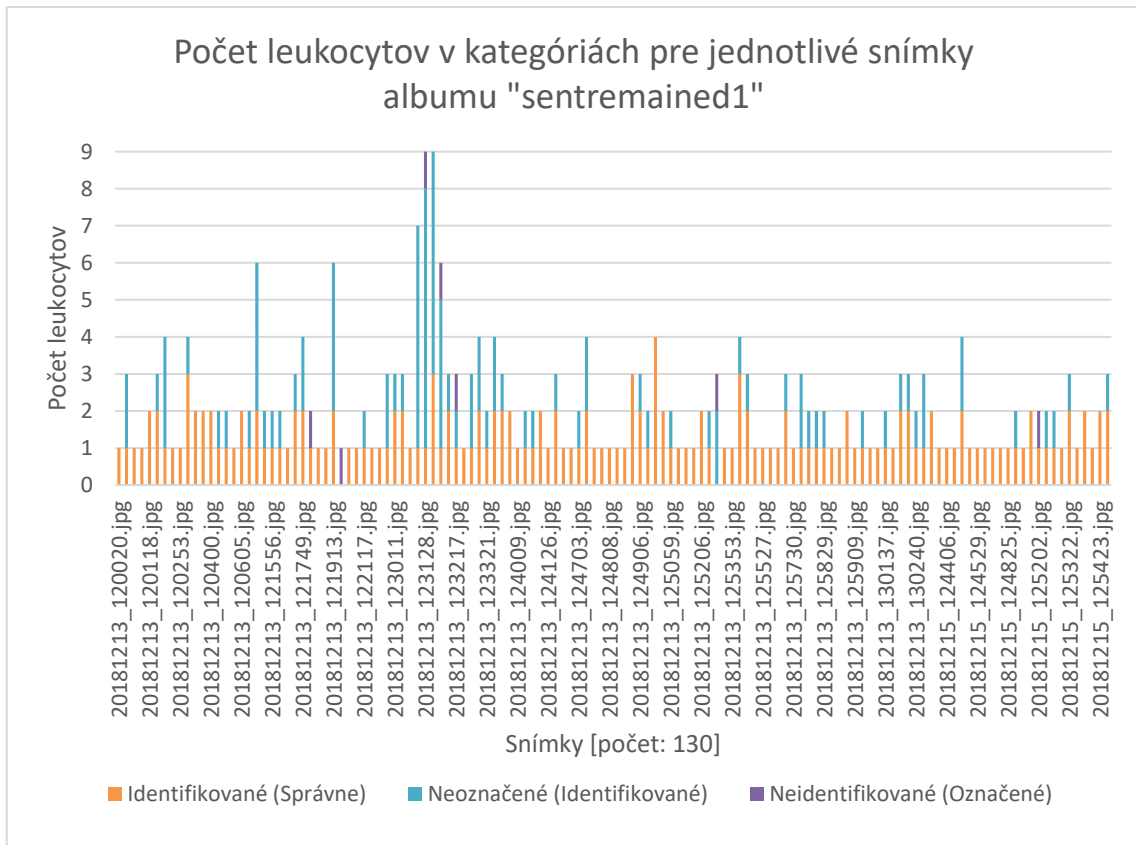
5.2.1.1 Prvý experiment – Použitie techniky na jednom albume množiny údajov

V prvom experimente bol implementovaný systém použitý pre spracovanie Albumu „sentremained1“ z množiny údajov Raabin-WBC, s obsahom 130 snímok.



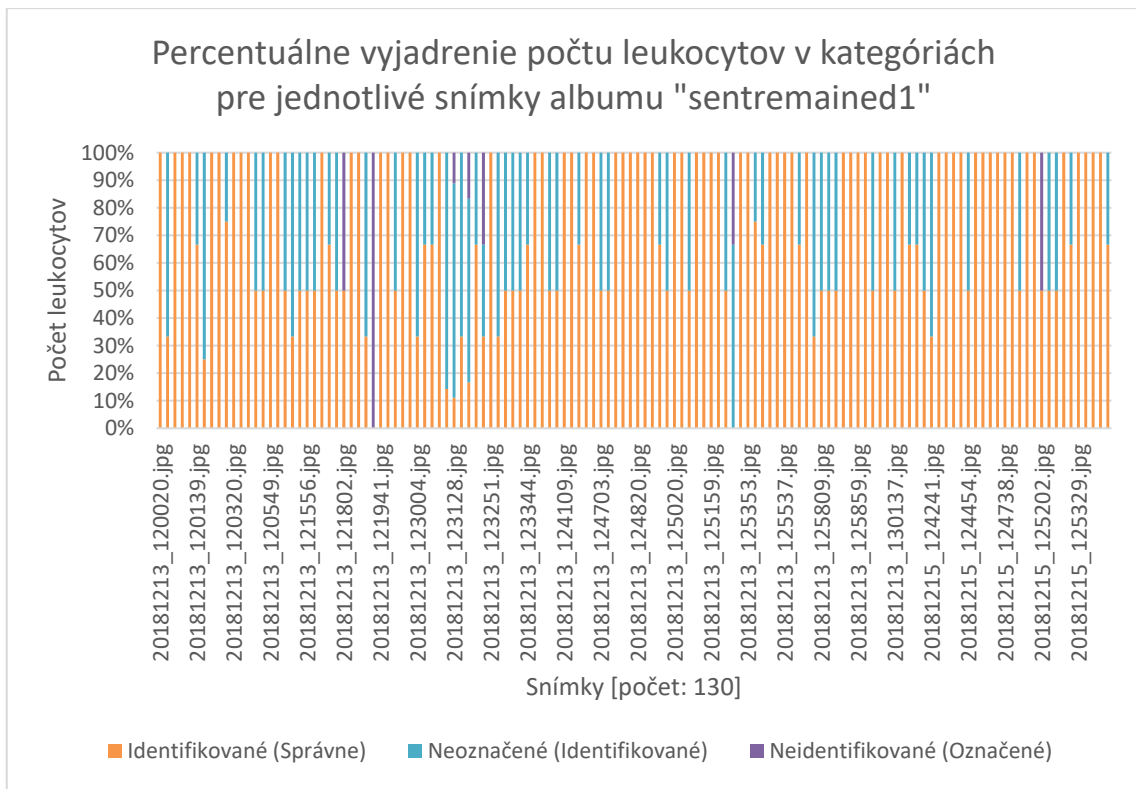
Obrázok 29 - Porovnanie označených (vľavo) a identifikovaných buniek (vpravo)

Výsledné štatistiky sú zobrazené na obrázkoch 30 – 32. Na obrázku 30 je zobrazený počet leukocytov pre jednotlivé snímky vybraného albumu. Táto štatistika zobrazuje spoľahlivosť identifikácie a chybovosť. Na grafe je možné vidieť vyšší počet neoznačených buniek, ktoré ale vo väčšine prípadov neznamenajú zvýšenú chybovosť ale neboli v anotáciách označené (obrázok 29). Pri vyššom počte (6 a viac) neoznačených buniek na snímke je možné túto situáciu považovať ako chybnú identifikáciu. Kategória neidentifikovaných buniek obsahuje chýbajúcu identifikáciu buniek na snímkach, ktoré boli zaznamenané v anotáciách.

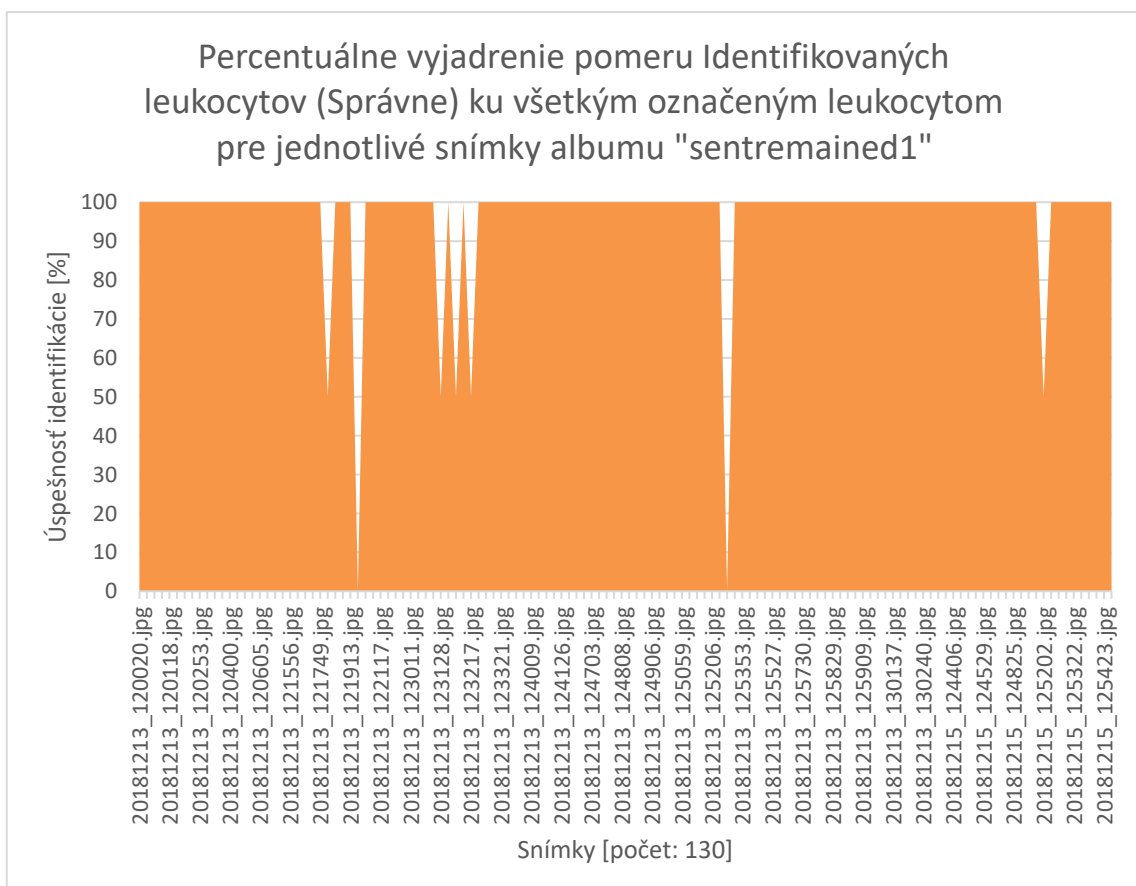


Obrázok 30 - Experiment 1 - Počet leukocytov v kategóriách

Obrázok 31 zobrazuje rovnakú štatistiku ako obrázok 30, ale prepočítanú v percentách. Táto štatistika je uvedená, pretože je pri veľkom počte neoznačených buniek (ako napr. v experimente 2) prehľadnejšia. Na obrázku 32 je zobrazená percentuálna úspešnosť identifikácie leukocytov v pomere k označeným leukocytom.



Obrázok 31 - Experiment 1 - Percentuálne vyjadrenie počtu leukocytov v kategóriách

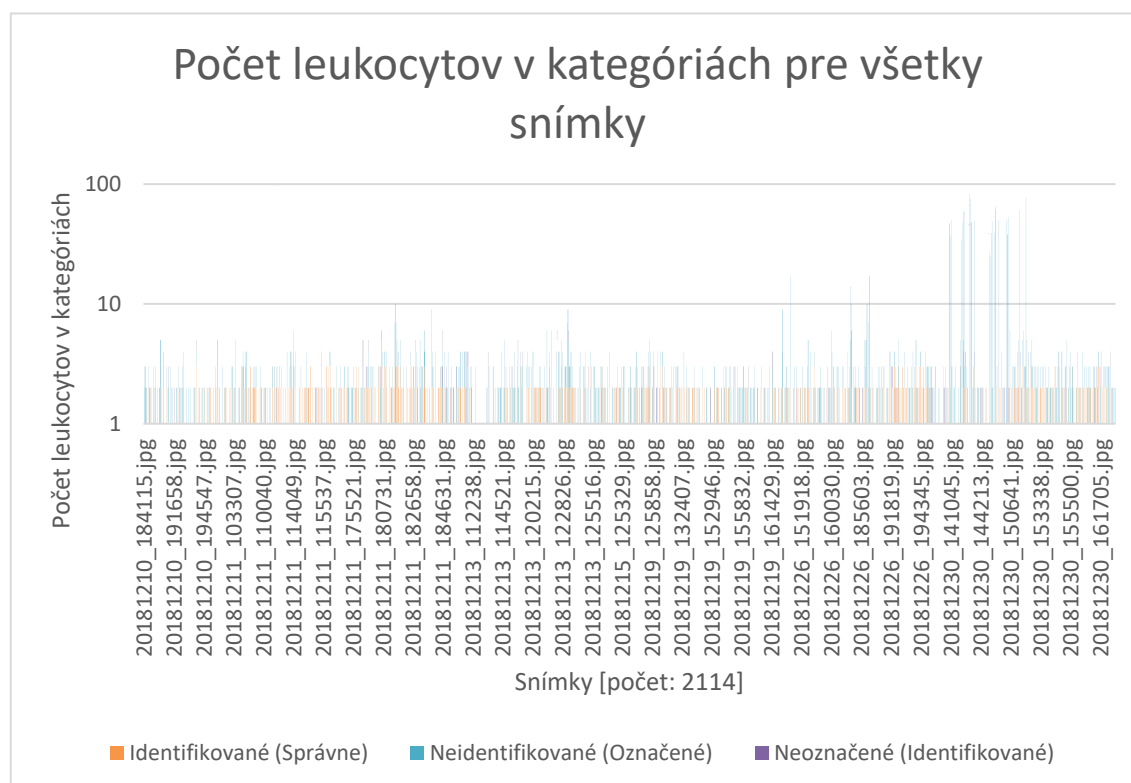


Obrázok 32 - Experiment 1 - Percentuálne vyjadrenie pomeru Identifikovaných ku všetkým označeným leukocytom

5.2.1.2 Druhý experiment – Použitie techniky na celej množine údajov

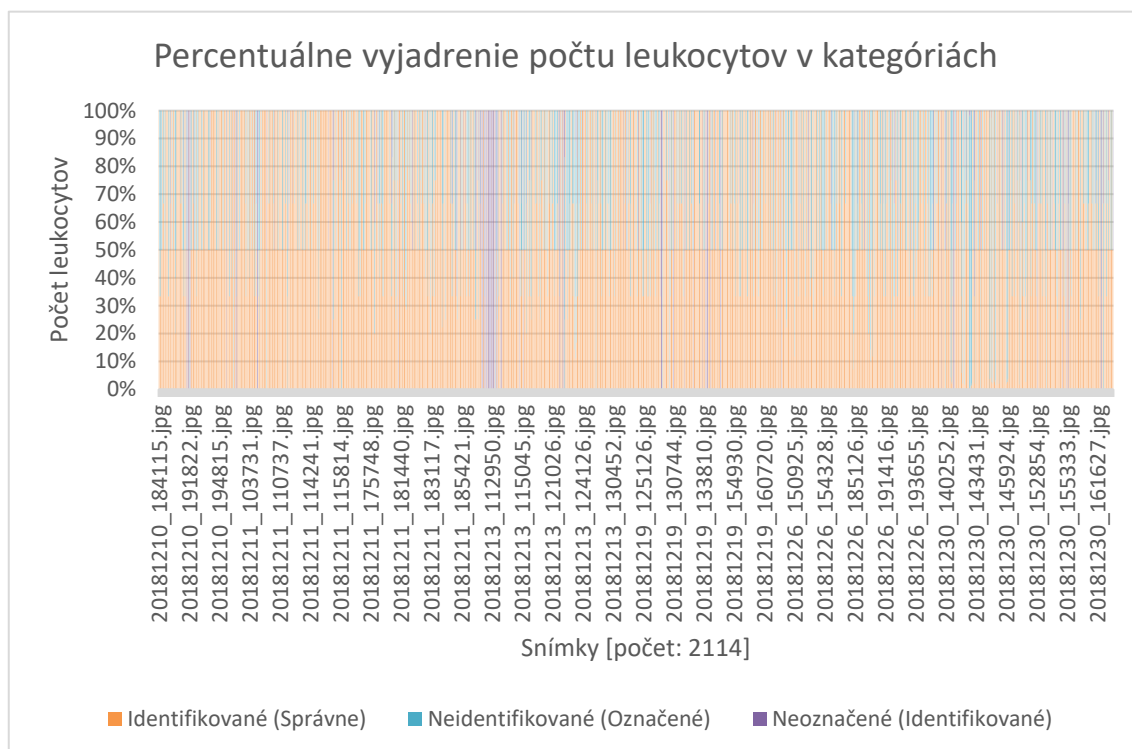
Druhý experiment bol realizovaný pre celú množinu údajov (2114 snímok), s cieľom zistiť spoľahlivosť techniky na veľkej a variabilnej referenčnej vzorke.

Na obrázku 33 je možné vidieť zvýšený počet neidentifikovaných buniek, ktoré ale chápeme ako chybu rozpoznávania na snímke. Po dodatočnej analýze snímok sa zistila rozdielna optická kvalita (zhoršená) oproti ostatným snímkam. Preto by bolo pre ďalšiu analýzu tieto snímky vyhodnocovať samostatne s rozdielnymi parametrami.

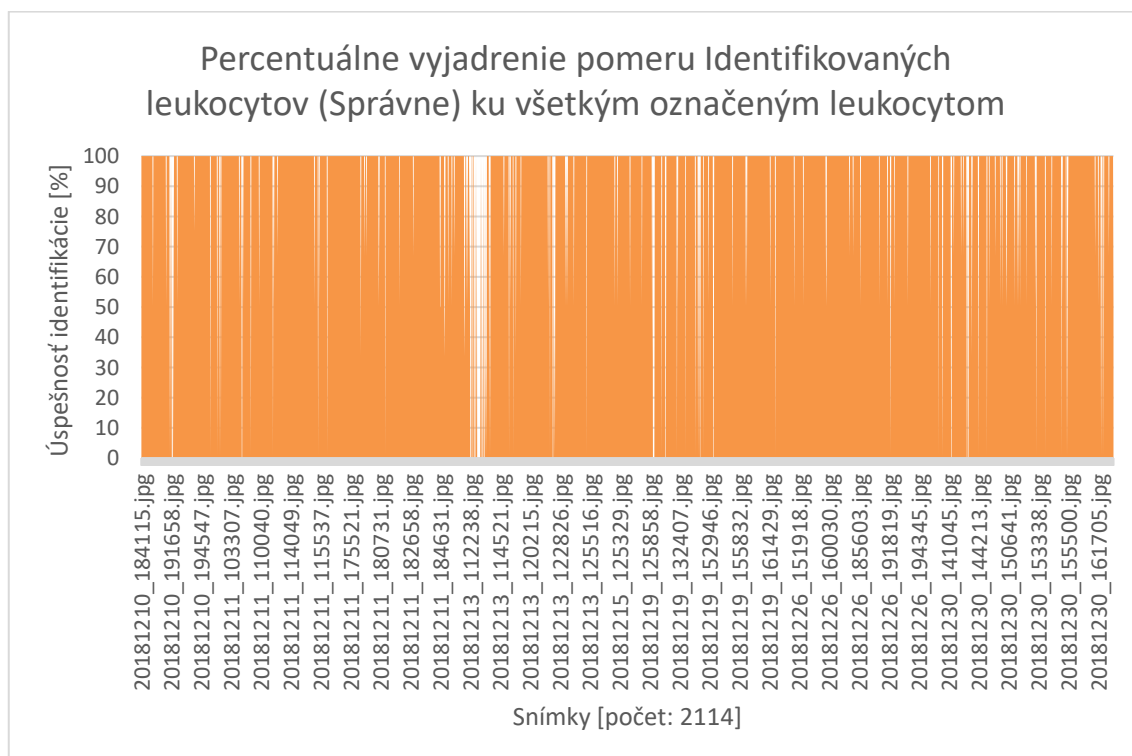


Obrázok 33 - Experiment 2 - Počet leukocytov v kategóriách

Na obrázku 34 je podiel spomenutých kategórií vyjadrený v percentách, pričom môžeme pri zohľadnení chýbajúcich anotácií stále pozorovať vysokú percentuálnu spoľahlivosť implementovanej techniky. Obrázok 35 uvádza spoľahlivosť techniky pri správnej identifikácii označených leukocytov.



Obrázok 34 - Experiment 2 - Percentuálne vyjadrenie počtu leukocytov v kategóriách



Obrázok 35 - Experiment 2 - Percentuálne vyjadrenie pomeru Identifikovaných ku všetkým označeným leukocytom

5.2.1.3 Vyhodnotenie experimentov

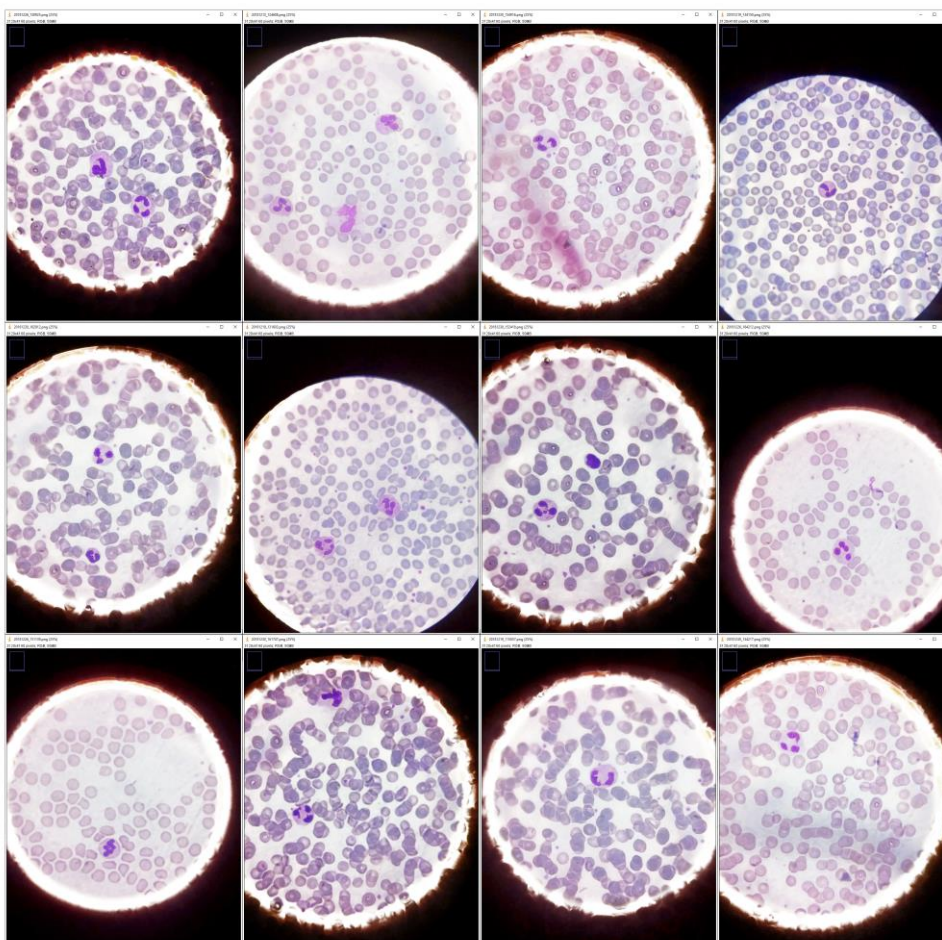
V tabuľke 1 je sumarizované porovnanie spoľahlivosti implementovanej techniky na uvedených experimentoch.

Tabuľka 1 - Porovnanie výsledkov experimentov obrazového spracovania údajov

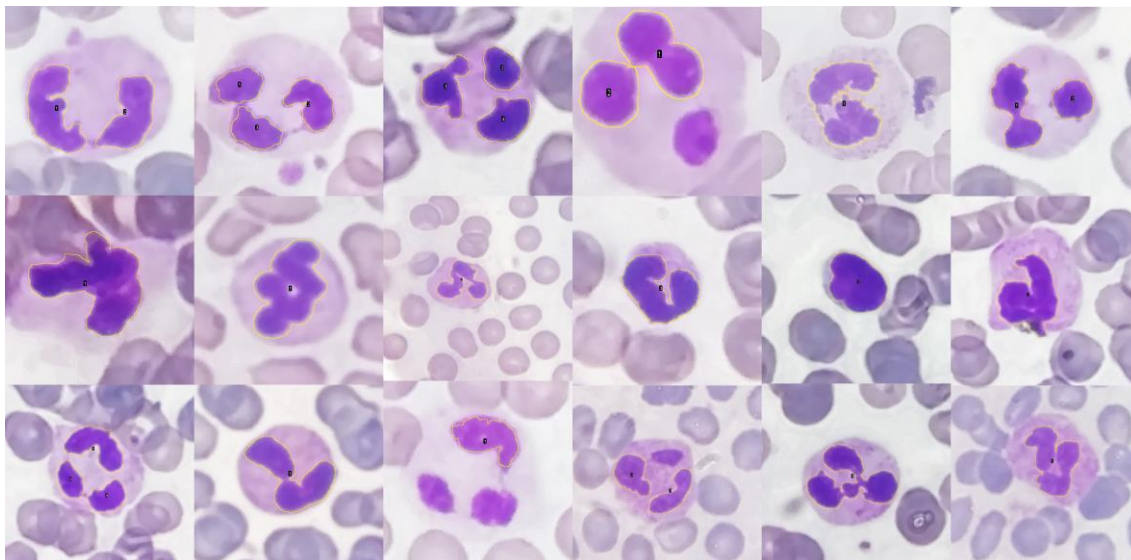
Experimenty	Počet snímok	Identifikované ¹ [%]	Identifikované ² (Správne) [%]	Neidentifikované (Označené) [%]	Neoznačené (Identifikované) [%]
Experiment 1	130	94,62	53,08	5,38	44,62
Experiment 2	2114	91,49	59,98	8,51	34,34

¹⁾ Snímky obsahujúce všetky Identifikované leukocyty z označených leukocytov (môžu obsahovať navyše neidentifikované leukocyty)
²⁾ Snímky obsahujúce iba všetky Identifikované leukocyty z označených leukocytov

Spoľahlivosť dosiahnutá v druhom experimente bola nižšia než v prvom experimente, práve z dôvodu väčšej variability vzorky, ktorú môžeme vidieť na obrázkoch 36 a 37. Použitie tejto techniky je teda výhodnejšie v prípade menšej a podobnej množiny údajov, napríklad pri vzorke z jedného pokusu. Keďže technika pracuje iba s grafickými vlastnosťami snímky, je veľmi citlivá na rozdiely a zmeny v týchto aspektoch, od čoho sa odvíja aj jej väčšia chybovosť.



Obrázok 36 - Ukážka variability snímok celej množiny údajov



Obrázok 37 - Ukážka identifikovaných oblastí

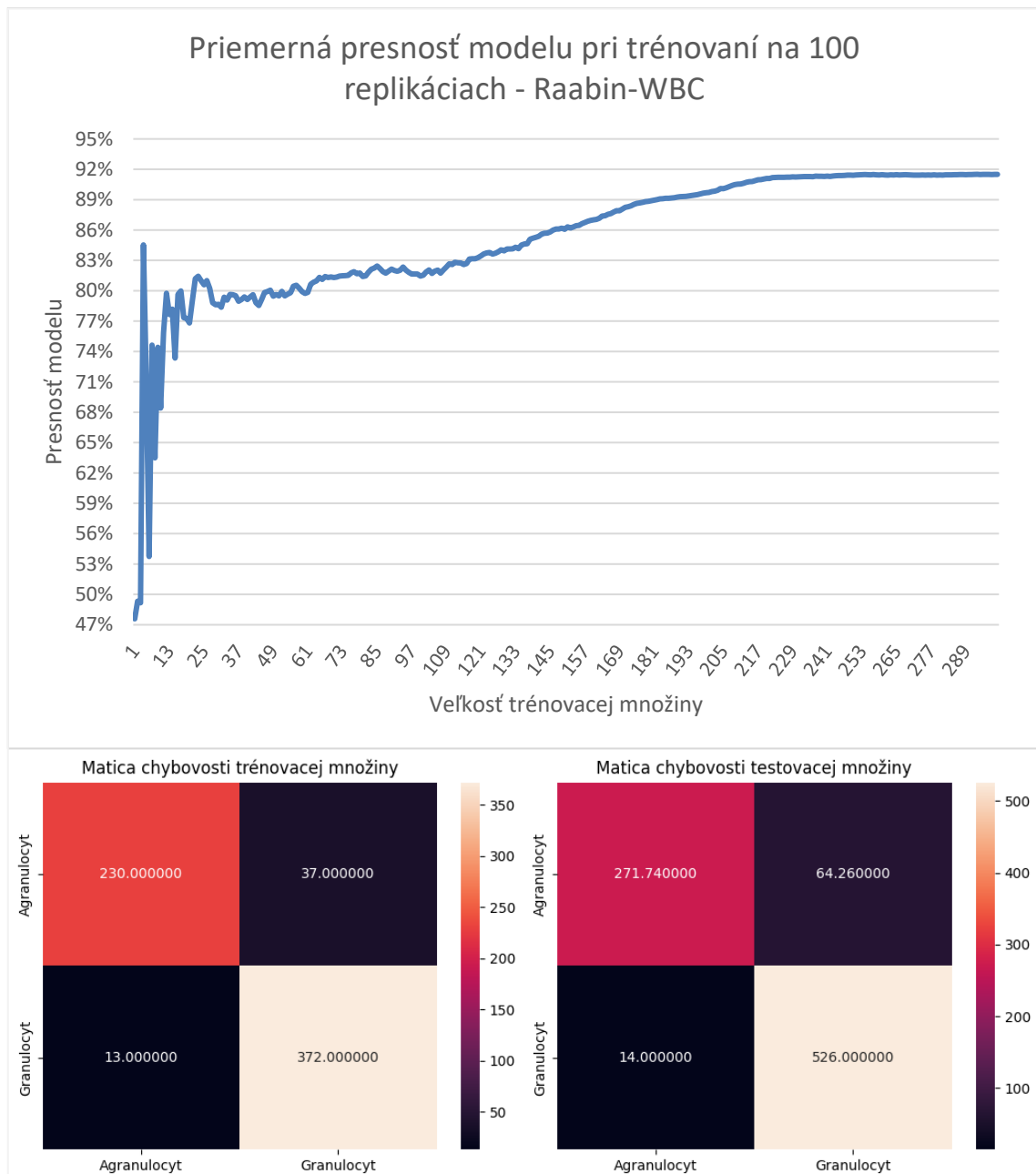
5.2.2 Spracovanie snímok strojovým učením

Cieľom experimentov bolo overiť implementáciu spracovania snímok strojovým učením na rôznych množinách údajov. Pre oba experimenty bolo realizovaných 100 replikácií, za účelom eliminácie náhodnosti výberu snímok z celkovej trénovacej množiny.

5.2.2.1 Prvý experiment – Použitie techniky na množine údajov Raabin-WBC

V prvom experimente bola technika použitá pre všetky snímky označených leukocytov v množine Raabin-WBC (2504 snímok). Množina bola rozdelená na dve množiny pre natréovanie a otestovanie modelu.

Obrázok 38 zobrazuje presnosť modelu na dvoch typoch grafov. Prvý graf uvádza priemernú presnosť modelu pri tréovaní s postupným zväčšovaním trénovacej množiny. Ustálenie spoľahlivosti klasifikácie leukocytov môžeme pozorovať pri veľkosti trénovacej množiny 220 snímok – cca 91%. Spodné grafy zobrazujú maticu chybovosti, ktorá demonštruje spoľahlivosť klasifikácie jednotlivých tried.



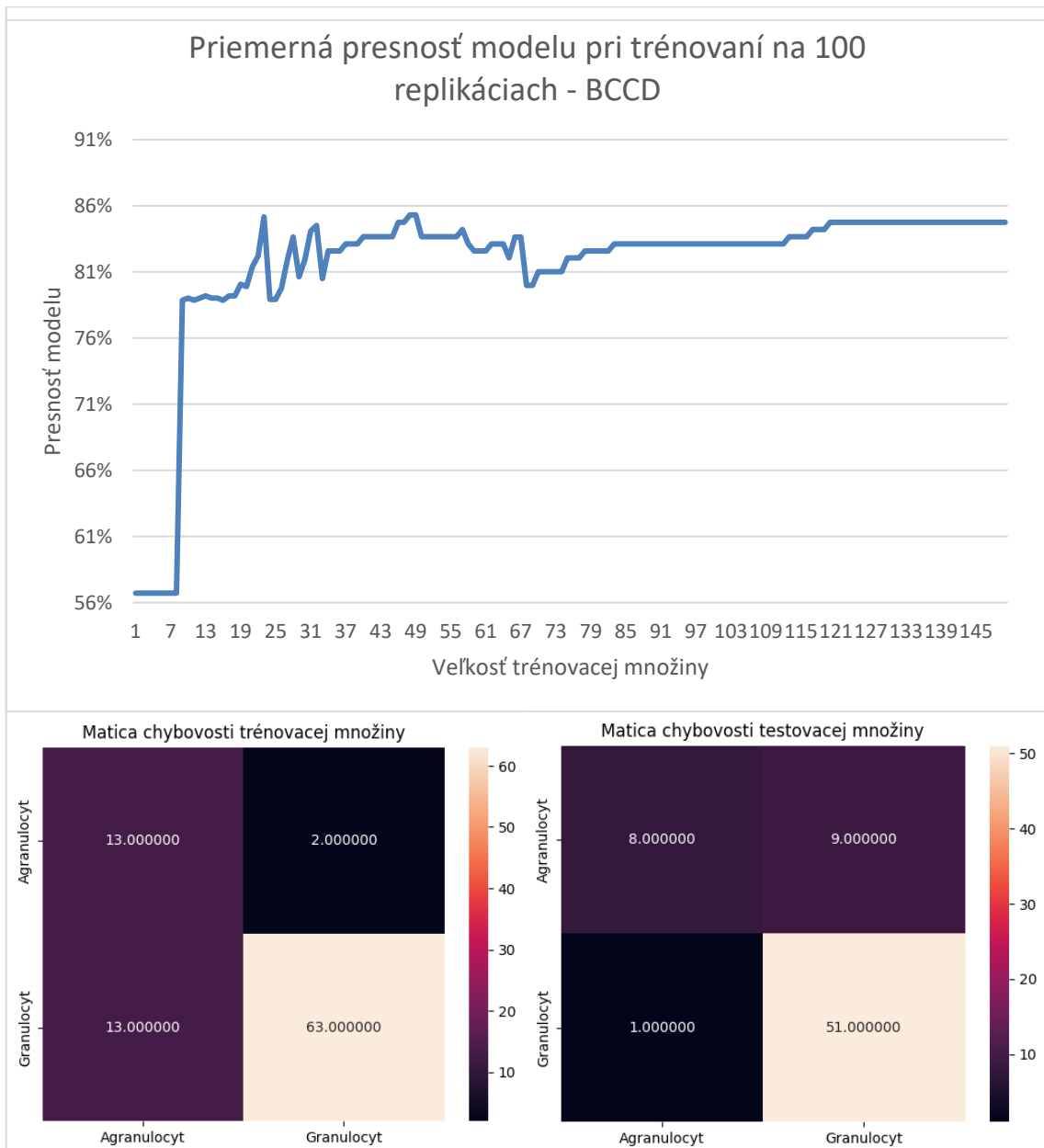
Obrázok 38 - Experiment 1 - Presnosť modelu so vstupnými údajmi Raabin-WBC

5.2.2.2 Druhý experiment – Použitie techniky na množine údajov BCCD

V druhom experimente bola technika použitá pre všetky označené snímky v množine BCCD. V tejto množine sa však nachádzalo iba 344 snímok leukocytov, preto boli niektoré parametre (napr. maximálna veľkosť použitej testovacej množiny, pomer testovacej a tréňovacej množiny a počet tried klasifikácie) adekvátne upravené k počtu snímok.

Na obrázku 39 je zobrazená priemerná presnosť modelu. Je možné si všimnúť zníženú spoľahlivosť oproti predošlému experimentu, čo je spôsobené malým počtom snímok vo

vzorke. Matice chybovosti potvrdzujú problém v množstve snímok potrebnom pre získanie požadovanej vyššej úspešnosti (aspoň 220 snímok).



Obrázok 39 - Experiment 2 - Presnosť modelu so vstupnými údajmi BCCD

5.2.2.3 Vyhodnotenie experimentov

V tabuľke 2 sú uvedené porovnania číselných parametrov a výsledkov jednotlivých množín údajov a presnosti modelu vyhodnocovanej metrikou vyváženého skóre

presnosti („balanced accuracy score“) s váhami, kde váhami bol počet agranulocytov a granulocytov v množine.

Tabuľka 2 - Porovnanie výsledkov experimentov spracovania údajov strojovým učením

Experimenty	Počet snímok	Celková veľkosť trérovacej / testovacej množiny pri tréovaní	Veľkosť testovacej množiny pri predikcii	Priemerná presnosť modelu pri tréovaní [%] / veľkosť použitej ¹⁾ trérovacej množiny	Priemerná presnosť modelu pri testovaní predikcie [%]
Experiment 1	2504	976 / 652	876	91,57 / 300	93,48
Experiment 2	344	184 / 91	69	84,75 / 150	83,44

¹⁾ Pri aktívnom učení sú z celkovej trérovacej množiny náhodne vyberané prvky, ktoré tvoria použitú trérovaciu množinu

Hoci sú výsledky dosiahnuté v druhom experimente horšie, je to hlavne z dôvodu nízkého počtu snímok. Pri prvom experimente bolo potrebných približne 210 snímok, pre dosiahnutie spoľahlivosti nad 90%. Pri 150 snímkach bola spoľahlivosť približne 85%. V druhom experimente bola pri rovnakom počte snímok a minimálnej kalibrácii parametrov dosiahnutá spoľahlivosť 84,75% pri tréovaní modelu a 83,44% pri predikcii, čo preukazuje relevantnosť a použiteľnosť implementovaného riešenia na rôznych množinách údajov. Zároveň bolo experimentmi dokázané, že nevyhnutnou podmienkou spoľahlivého spracovania údajov technikou strojového učenia je väčšia množina údajov.

5.3 Prínos vyhodnocovania v ďalšom výskume

Uskutočnené experimenty v dostatočnej miere potvrdili úspešnosť navrhnutých a implementovaných techník. Pri každej technike je možné nájsť silné a slabé stránky, ktoré je možné využiť pri rozhodovaní a maximalizácii úspešnosti realizovaného experimentu.

Zvolené techniky boli počas implementácie pripravené na ďalšie rozšírenie. Prvou možnosťou je využitie techniky obrazového a strojového spracovania spoločne. Výhodou spojenia týchto techník je možnosť aplikácie na novej množine snímok, ktorá ešte nemá vytvorené anotácie buniek k ich kategorizácii. Obrazové spracovanie vie identifikovať pozície leukocytov a ohraničiť ich tvar. Na základe týchto znalostí je možné odborníkom jednoduchšie vytvoriť anotácie, pretože oblasti buniek sú už identifikované, a navyše je postačujúce klasifikovať iba pomernú časť množiny snímok. Zvyšná časť môže byť predikovaná natrénovaným modelom aktívneho strojového učenia. Takéto využitie však vyžaduje dodatočnú implementáciu, ktorá nebola predmetom záverečnej práce.

Aktuálne je možné systém aplikovať aj na zjednodušenie vytvárania anotácií v nových množinách údajov.

Ďalším príkladom rozšírenia tejto práce je implementácia do webovej aplikácie formou knižnice. Toto rozšírenie je možné aplikovať aj s aktuálnym stavom implementácie systému. Využitie si to môže nájsť v informačnom systéme na uchovávanie a analýzu biomedicínskych údajov v štýle softvérovej biobanky, aktuálne riešenej v inej diplomovej práci.

Reálny prínos implementovaného systému môžu nájsť biomedicínski špecialisti, zameriavajúci sa na počítanie a analýzu typov leukocytov na mikroskopických snímkach. Použitím tohto systému je možné automatizovať a urýchliť proces analýzy, a tiež odstrániť ľudskú chybovosť. Príkladom použitia systému môže byť zhodnotenie počtu leukocytov pred a po aplikácii liečiva. Iným príkladom aplikácie je možnosť optimálnejšieho vyhodnotenia diagnostiky ochorení, ako napr. leukémie.

ZÁVER

Riešenie problematiky spracovania a vyhodnocovania údajov je v súčasnosti aktuálnou témou vo viacerých oblastiach. Nakoľko je problematika obsiahla, bolo potrebné sa najskôr zamerať na prehľad klasifikácie biomedicínskych údajov. Z tohto prehľadu vieme určiť viacero oblastí, kde je možné klasifikáciu aplikovať v medicínskych a biomedicínskych procesoch.

Jednou z týchto oblastí je analýza obrazových údajov, do ktorej môžeme zaradiť napr. magnetickú rezonanciu, počítačovú tomografiu, ultrasonografiu, rádiografiu alebo fotomikroskopiou buniek. Rozvoj tejto oblasti je prínosom v množstve liečebných procesov, ktoré skvalitňujú životnú úroveň a zdravie populácie. Spoločným znakom analýz spomenutých typov techník sú podobné algoritmické a softvérové postupy pri spracovaní a vyhodnocovaní snímok, preto je rozvoj jednej techniky prínosný a ľahko aplikovateľný aj v ďalších technikách.

Práca sa zameriava na analýzu fotomikroskopických snímok, ktoré si nájdu široké uplatnenie v odvetviach biomedicíny a biomedicínskej informatiky. Cieľom je identifikovať a klasifikovať typy a počet leukocytov na obrazových snímkach. Pestrosť typov a vlastností leukocytov je rozsiahla, preto je ich klasifikácia náročným procesom. Pre zjednodušenie je preto v práci klasifikácia generalizovaná na dve hlavné kategórie leukocytov – granulocyty a agranulocyty. V súčasnosti už existuje množstvo dostupných techník a softvérov, ktoré je možné aplikovať v procese identifikácie a klasifikácie.

Práca sa zameriava na návrh systému pre spracovanie a analýzu takýchto snímok, pre čo bol potrebný výber správnej množiny údajov. V návrhu sa definovali dve techniky spracovania – obrazové a strojové. Práca sa zaoberá popisom implementácie týchto techník samostatne, keďže každá technika pristupuje k riešeniu stanoveného problému iným spôsobom. Obrazové spracovanie je implementované pomocou JAVA skriptov s využitím nástroja ImageJ. Strojové spracovanie využíva techniku aktívneho strojového učenia, implementovaného v PYTHON skriptoch s aplikáciou softvéru CellProfiler. Vytvorený systém umožňuje okrem spracovania aj vyhodnocovať spoľahlivosť a výsledky použitia techník. Navrhnuté riešenie bolo overené na sadách experimentov. Tieto experimenty demonštrujú správnosť tohto systému, a zároveň stanovujú odporúčania pre jeho optimálne použitie.

Prínosom práce je spracovanie teórie klasifikácie biomedicínskych údajov, ich použitie a vyhodnocovanie. Ďalším prínosom je vytvorenie systému pre spracovanie a analýzu leukocytov, pre ktorý sa vyhodnotila spoľahlivosť implementovaných techník. Nemaľou výhodou je rozšírenie znalostí v jednotlivých odboroch, ktoré je možné využiť pri budúcom profesijnom raste a podrobnejšom rozvoji tejto problematiky.

Zoznam použitej literatúry

- [1] Inc. Pure Storage, “What Is Genomics Data?” <https://www.purestorage.com/knowledge/what-is-genomics-data.html> (cit 01. marec 2023).
- [2] S. Ghatak, R. B. Muthukumar, a S. K. Nachimuthu, “A Simple Method of Genomic DNA Extraction from Human Samples for PCR-RFLP Analysis”, *J Biomol Tech*, s. jbt.13-2404-001, dec. 2013, doi: 10.7171/jbt.13-2404-001.
- [3] National Human Genome Research Institute, “Genomic Data Science”, 2022. <https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science> (cit 01. marec 2023).
- [4] P. Schmieder, “Multidimensional NMR Spectroscopy”, v *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, Springer Berlin Heidelberg, s. 1204–1208. doi: 10.1007/3-540-29623-9_5020.
- [5] S. Al-Amrani, Z. Al-Jabri, A. Al-Zaabi, J. Alshekaili, a M. Al-Khabori, “Proteomics: Concepts and applications in human medicine”, *World J Biol Chem*, roč. 12, č. 5, s. 57–69, sep. 2021, doi: 10.4331/wjbc.v12.i5.57.
- [6] A. Schmidt, I. Forne, a A. Imhof, “Bioinformatic analysis of proteomics data”, *BMC Syst Biol*, roč. 8, č. S2, s. S3, mar. 2014, doi: 10.1186/1752-0509-8-S2-S3.
- [7] E. Riekeberg a R. Powers, “New frontiers in metabolomics: from measurement to insight”, *F1000Res*, roč. 6, s. 1148, júl. 2017, doi: 10.12688/f1000research.11495.1.
- [8] A. Cambiaghi, M. Ferrario, a M. Masseroli, “Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration”, *Brief Bioinform*, s. bbw031, apr. 2016, doi: 10.1093/bib/bbw031.
- [9] S. Wang *et al.*, “Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects”, *Information Fusion*, roč. 76, s. 376–421, dec. 2021, doi: 10.1016/j.inffus.2021.07.001.
- [10] Institute of Medicine (US) Roundtable on Value & Science-Driven Health Care., “Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary.”, 2010. <https://www.ncbi.nlm.nih.gov/books/NBK54290/> (cit 22. marec 2023).

-
- [11] U. of W. Health Sciences Library, "DATA RESOURCES IN THE HEALTH SCIENCES", 2023. <https://guides.lib.uw.edu/hsl/data/findclin> (cit 02. marec 2023).
- [12] Longwood Research Data Management, "Biomedical Metadata". <https://datamanagement.hms.harvard.edu/collect-analyze/documentation-metadata> (cit 02. marec 2023).
- [13] J. Barowski a A. Cheprasov, "Medical Data Types and Examples", 2022. <https://study.com/learn/lesson/medical-data-types-examples.html> (cit 20. február 2023).
- [14] R. A. Cochran, S. S. Feldman, N. V Ivankova, A. G. Hall, a W. Opoku-Agyeman, "Intention to Use Behavioral Health Data From a Health Information Exchange: Mixed Methods Study", *JMIR Ment Health*, roč. 8, č. 5, s. e26746, máj. 2021, doi: 10.2196/26746.
- [15] J. C. Yau *et al.*, "TILES-2019: A longitudinal physiologic and behavioral data set of medical residents in an intensive care unit", *Sci Data*, roč. 9, č. 1, s. 536, sep. 2022, doi: 10.1038/s41597-022-01636-4.
- [16] L. A. Palinkas, "Qualitative and Mixed Methods in Mental Health Services and Implementation Research", *Journal of Clinical Child & Adolescent Psychology*, roč. 43, č. 6, s. 851–861, nov. 2014, doi: 10.1080/15374416.2014.910791.
- [17] World Health Organization, "Mental disorders", 08. jún 2022. <https://www.who.int/news-room/fact-sheets/detail/mental-disorders> (cit 23. február 2023).
- [18] Regis College, "8 Environmental Factors that Affect Health", 22. apríl 2021. <https://online.regiscollege.edu/blog/environmental-factors-that-affect-health/> (cit 24. február 2023).
- [19] National Research Council (US) Committee on Environmental Epidemiology; National Research Council (US) Commission on Life Sciences., "Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology.", 1997. <https://www.ncbi.nlm.nih.gov/books/NBK233639/> (cit 22. február 2023).
- [20] J. Alper, Ed., *Informing Environmental Health Decisions Through Data Integration*. Washington, D.C.: National Academies Press, 2018. doi: 10.17226/25139.
- [21] "Practical Biomedical Signal Analysis Using MATLAB®", 2012.

- [22] J. Rajeswari a M. Jagannath, “Advances in biomedical signal and image processing – A systematic review”, *Inform Med Unlocked*, roč. 8, s. 13–19, 2017, doi: 10.1016/j.imu.2017.04.002.
- [23] I. Spasic a G. Nenadic, “Clinical Text Data in Machine Learning: Systematic Review”, *JMIR Med Inform*, roč. 8, č. 3, s. e17984, mar. 2020, doi: 10.2196/17984.
- [24] L. Qing, W. Linhong, a D. Xuehai, “A Novel Neural Network-Based Method for Medical Text Classification”, *Future Internet*, roč. 11, č. 12, s. 255, dec. 2019, doi: 10.3390/fi11120255.
- [25] Y. Sreekumar a P. K. Nizar Banu, “Clinical Text Classification of Medical Transcriptions Based on Different Diseases”, 2022, s. 613–623. doi: 10.1007/978-981-16-9885-9_50.
- [26] A. Devaney, “9 most common types of medical text datasets”, 31. august 2022. <https://www.centaurilabs.com/post/9-most-common-types-of-medical-text-datasets> (cit 27. február 2023).
- [27] Inc. (RSNA). Radiological Society of North America, “Body MRI”, 06. december 2022. <https://www.radiologyinfo.org/en/info/bodymr> (cit 04. marec 2023).
- [28] Lymphoma Australia, “MRI scan - Lymphoma Australia”. <https://www.lymphoma.org.au/lymphoma/tests-diagnosis-and-staging/scans-and-lymphoma/mri-scan/> (cit 09. marec 2023).
- [29] S. Wang *et al.*, “Advances in data preprocessing for bio-medical data fusion: An overview of the methods, challenges, and prospects”, *Information Fusion*, roč. 76. Elsevier B.V., s. 376–421, 01. december 2021. doi: 10.1016/j.inffus.2021.07.001.
- [30] A. Berger, “Magnetic resonance imaging”, *BMJ : British Medical Journal*, roč. 324, č. 7328, s. 35, jan. 2002, doi: 10.1136/BMJ.324.7328.35.
- [31] A. REHMAN, M. A. SHEIKH, a M. M. WARAICH, “CT Scan”, *The Professional Medical Journal*, roč. 16, č. 04, s. 579–582, jan. 2023, doi: 10.29309/tpmj/2009.16.04.2746.
- [32] Two Views, “What is a Cranial CT scan? | Two Views”. <https://www.two-views.com/ct/cranial-scan.html> (cit 17. marec 2023).
- [33] “What is Computed Tomography? | FDA”. <https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/what-computed-tomography> (cit 04. marec 2023).

- [34] “PET Scan: What It Is, Types, Purpose, Procedure & Results”. <https://my.clevelandclinic.org/health/diagnostics/10123-pet-scan> (cit 04. marec 2023).
- [35] M. Kapoor a A. Kasi, “PET Scanning”, *StatPearls*, okt. 2022, Cit: 04. marec 2023. [Online]. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK559089/>
- [36] The University of California, “Brain PET scan”. <https://www.ucsfhealth.org/medical-tests/brain-pet-scan-> (cit 18. marec 2023).
- [37] S. S. Anand, H. Singh, a A. K. Dash, “Clinical Applications of PET and PET-CT”, *Med J Armed Forces India*, roč. 65, č. 4, s. 353, 2009, doi: 10.1016/S0377-1237(09)80099-3.
- [38] A. Berger, “Positron emission tomography”, *BMJ: British Medical Journal*, roč. 326, č. 7404, s. 1449, jún. 2003, doi: 10.1136/BMJ.326.7404.1449.
- [39] L. K. Griffeth, “Use of PET/CT scanning in cancer patients: technical and practical considerations”, *Proc (Bayl Univ Med Cent)*, roč. 18, č. 4, s. 321, okt. 2005, doi: 10.1080/08998280.2005.11928089.
- [40] A. Carovac, F. Smajlovic, a D. Junuzovic, “Application of Ultrasound in Medicine”, *Acta Informatica Medica*, roč. 19, č. 3, s. 168, 2011, doi: 10.5455/AIM.2011.19.168-171.
- [41] A. Etchegaray, S. Juarez-Peñalva, F. Petracchi, a L. Igarzabal, “Prenatal genetic considerations in congenital ventriculomegaly and hydrocephalus”, *Child’s Nervous System*, roč. 36, č. 8, s. 1645–1660, aug. 2020, doi: 10.1007/s00381-020-04526-5.
- [42] “Contrast-Enhanced Ultrasound (CEUS): What It Is & Procedure”. <https://my.clevelandclinic.org/health/diagnostics/22754-contrast-enhanced-ultrasound-ceus> (cit 07. marec 2023).
- [43] G. M. Lanza, “Ultrasound Imaging: Something Old or Something New?”, *Invest Radiol*, roč. 55, č. 9, s. 573, sep. 2020, doi: 10.1097/RLI.0000000000000679.
- [44] “Ultrasound-pros and cons | European Society for Paediatric Anaesthesiology”. <https://www.euroespa.com/science-education/specialized-sections/espa-pain-committee/us-regional-anaesthesia/ultrasound-pros-and-cons/> (cit 07. marec 2023).
- [45] Food and Drug Administration, “Radiography”, <https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/radiography>, 28. september 2020.

- [46] B. O'Sullivan a S. Goergen, "Plain Radiograph/X-ray", <https://www.insideradiology.com.au/plain-radiograph-x-ray/>, 31. august 2017.
- [47] A. Dixon, "Normal cervical spine radiographs", *Radiopaedia.org*, nov. 2014, doi: 10.53347/RID-32505.
- [48] PrimedeQ India Private Limited, "How does an X-Ray machine work? What are the main components of an X-Ray unit?", <https://www.primedeq.com/blog/how-does-an-x-ray-machine-work-what-are-the-main-components-of-an-x-ray-unit/>, 20. jún 2019.
- [49] Diagnostic Imaging Systems, "Physics of Radiology", <https://vetxray.com/resource-center/continuing-education/physics-of-radiology/>.
- [50] J.-Y. Chen, Y.-J. D. Green, M. D. Sabatinos, a S. A. Forsburg, "PombeX: Robust Cell Segmentation for Fission Yeast Transillumination Images", *PLoS One*, roč. 8, č. 12, s. 81434, 2013, doi: 10.1371/journal.pone.0081434.
- [51] RF Wireless World, "Advantages of X-Ray | Disadvantages of X-Ray", <https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-X-Ray.html>.
- [52] J. v Lorenzo-Ginori, "Digital Image Processing in Cell Imaging: A Review on Research Developments Digital Image quality analysis and optimization in Medical Studies. View project Análisis de la señal ECG mediante Distribuciones Tiempo-Frecuencia View project DIGITAL IMAGE PROCESSING IN CELL IMAGING: A REVIEW ON RESEARCH DEVELOPMENTS", 2014. [Online]. Available at: <https://www.researchgate.net/publication/277246844>
- [53] "Scar Tissue at 10x Magnification | Nikon's MicroscopyU". <https://www.microscopyu.com/gallery-images/scar-tissue-at-10x-magnification-1> (cit 12. marec 2023).
- [54] A. P. Cuny, F. P. Schlottmann, J. C. Ewald, S. Pelet, a K. M. Schmoller, "Live cell microscopy: From image to insight", *Biophys Rev*, roč. 3, č. 2, s. 021302, jún. 2022, doi: 10.1063/5.0082799.
- [55] J. B. Vaught a M. K. Henderson, "Biological sample collection, processing, storage and information management". [Online]. Available at: <http://biospecimens.cancer.gov/bestpractices>
- [56] S. Paxton, M. Peckham, A. Knibbs, S. Paxton, A. Knibbs, a M. Peckham, "White Blood Cells", *The Leeds Histology Guide*. 2003.

- [57] A. Prof. MUDr. Čalkovská DrSc., K. Prof. MUDr. Javorka DrSc., M. Prof. MUDr. Javorka PhD., D. Doc. MUDr. Mokrá PhD., a I. Doc. MUDr. Tonhajzerová PhD., *Fyziológia človeka pre nelekárske študijné programy*. Martin: Vydavateľstvo Osveta, 2017.
- [58] “Types of White Blood Cells (Leukocytes) with Functions and Picture”. <https://www.sciencefacts.net/types-of-white-blood-cells.html> (cit 13. marec 2023).
- [59] M. S. Blumenreich, “The White Blood Cell and Differential Count”, *Clinical Methods: The History, Physical, and Laboratory Examinations*, 1990, Cit: 05. marec 2023. [Online]. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK261/>
- [60] Y. Wang, H. Schellenberg, V. Walhorn, K. Toensing, a D. Anselmetti, “Binding Mechanism of Fluorescent Dyes to DNA Characterized by Magnetic Tweezers”, *Mater Today Proc*, roč. 4, s. S218–S225, 2017, doi: 10.1016/j.matpr.2017.09.190.
- [61] “Flow Cytometry Fluorophores and Dyes | Proteintech Group”. <https://www.ptglab.com/support/flow-cytometry-protocol/flow-cytometry-fluorophores-and-dyes/> (cit 07. marec 2023).
- [62] M. Makem a A. Tiedeu, “An efficient algorithm for detection of white blood cell nuclei using adaptive three stage PCA-based fusion”, *Inform Med Unlocked*, roč. 20, s. 100416, jan. 2020, doi: 10.1016/J.IMU.2020.100416.
- [63] G. Hanumant a R. Manza Babasaheb Ambedkar, “Segmentation of Thermal Images Using Thresholding-Based Methods for Detection of Malignant Tumours”, 2016, doi: 10.1007/978-3-319-47952-1_11.
- [64] F. Al-Hafiz, S. Al-Megren, a H. Kurdi, “Red blood cell segmentation by thresholding and canny detector”, *Procedia Comput Sci*, roč. 141, s. 327–334, 2018, doi: 10.1016/J.PROCS.2018.10.193.
- [65] A. Gordon, A. Colman-Lerner, T. E. Chin, K. R. Benjamin, R. C. Yu, a R. Brent, “Single-cell quantification of molecules and rates using open-source microscope-based cytometry”, *Nat Methods*, roč. 4, č. 2, s. 175–181, feb. 2007, doi: 10.1038/NMETH1008.
- [66] A. Chernomoretz, A. Bush, R. Yu, A. Gordon, a A. Colman-Lerner, “Cell-ID Software for Microscope-Based Cytometry”, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, roč. CHAPTER, č. SUPPL. 84, s. Unit, 2008, doi: 10.1002/0471142727.MB1418S84.

- [67] D. Vikram Mutneja, “Methods of Image Edge Detection: A Review”, *Journal of Electrical & Electronic Systems*, roč. 04, č. 02, 2015, doi: 10.4172/2332-0796.1000150.
- [68] C. McQuin *et al.*, “CellProfiler 3.0: Next-generation image processing for biology”, *PLoS Biol*, roč. 16, č. 7, s. e2005970, júl. 2018, doi: 10.1371/JOURNAL.PBIO.2005970.
- [69] “OpenCV: Image Segmentation with Watershed Algorithm”. https://docs.opencv.org/4.x/d3/db4/tutorial_py_watershed.html (cit 07. marec 2023).
- [70] S. Fehrmann, C. Paoletti, Y. Goulev, A. Ungureanu, H. Aguilaniu, a G. Charvin, “Supplemental information Aging Yeast Cells Undergo a Sharp Entry Into Senescence Unrelated to the Loss of Mitochondrial Membrane Potential”.
- [71] Y. Boykov a G. Funka-Lea, “Graph Cuts and Efficient N-D Image Segmentation”, *Int J Comput Vis*, roč. 70, č. 2, s. 109–131, 2006, doi: 10.1007/s11263-006-7934-5.
- [72] “Image Segmentation with Graph Cuts”. <https://julie-jiang.github.io/image-segmentation/> (cit 08. marec 2023).
- [73] S. Dimopoulos, C. E. Mayer, F. Rudolf, a J. Stelling, “Accurate cell segmentation in microscopy images using membrane patterns”, *Bioinformatics*, roč. 30, č. 18, s. 2644–2651, sep. 2014, doi: 10.1093/BIOINFORMATICS/BTU302.
- [74] “Active Contours, Deformable Models, and Gradient Vector Flow”. <https://iacl.ece.jhu.edu/Projects/gvf/> (cit 07. marec 2023).
- [75] A. Cartas-Ayala, “Gradient Vector Flow Snakes”.
- [76] Sonka *et al.*, “Edge Contour Extraction”, *Pitas*, section 5.5.
- [77] J. Y. Peng, Y. J. Chen, M. D. Green, S. A. Sabatinos, S. L. Forsburg, a C. N. Hsu, “PombeX: Robust Cell Segmentation for Fission Yeast Transillumination Images”, *PLoS One*, roč. 8, č. 12, s. e81434, dec. 2013, doi: 10.1371/JOURNAL.PONE.0081434.
- [78] T. Hastie, R. Tibshirani, a J. Friedman, “The Elements of Statistical Learning”, 2009, doi: 10.1007/978-0-387-84858-7.
- [79] M. Jordan, J. Kleinberg, a B. Schölkopf, “Pattern Recognition and Machine Learning”, 2006.

- [80] S. Swaminathan, “Logistic Regression — Detailed Overview | by Saishruthi Swaminathan | Towards Data Science”, 2018. <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc> (cit 10. marec 2023).
- [81] Anuranjeeta, K. K. Shukla, A. Tiwari, a S. Sharma, “Classification of Histopathological images of Breast Cancerous and Non Cancerous Cells Based on Morphological features”, *Biomedical and Pharmacology Journal*, roč. 10, č. 1, s. 353–366, mar. 2017, doi: 10.13005/BPJ/1116.
- [82] “How to Use Machine Learning Algorithms in Weka - MachineLearningMastery.com”. <https://machinelearningmastery.com/use-machine-learning-algorithms-weka/> (cit 10. april 2023).
- [83] “Support Vector Machine — Introduction to Machine Learning Algorithms | by Rohith Gandhi | Towards Data Science”. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47> (cit 07. marec 2023).
- [84] M. Wang, X. Zhou, F. Li, J. Huckins, R. W. King, a S. T. C. Wong, “Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy”, *Bioinformatics*, roč. 24, č. 1, s. 94–101, jan. 2008, doi: 10.1093/BIOINFORMATICS/BTM530.
- [85] A. Yang, Y. Bai, H. Liu, K. Jin, T. Xue, a W. Ma, “Application of SVM and its Improved Model in Image Segmentation”, *Mobile Networks and Applications*, roč. 27, č. 3, s. 851–861, jún. 2022, doi: 10.1007/S11036-021-01817-2/FIGURES/10.
- [86] A. Ng, “CS229 Lecture notes Support Vector Machines”.
- [87] “Linear vs. Non-linear Support Vector Machines: Which is More Accurate? | by Kenneth Ross | Medium”. <https://medium.com/@krr3wf/linear-vs-non-linear-support-vector-machines-which-is-more-accurate-d2380fcd57e6> (cit 07. marec 2023).
- [88] E. Bakker, P. S. Swain, a M. M. Crane, “Morphologically constrained and data informed cell segmentation of budding yeast”, *Bioinformatics*, roč. 34, č. 1, s. 88–96, jan. 2018, doi: 10.1093/BIOINFORMATICS/BTX550.
- [89] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, a W. Enbeyle, “Deep Neural Networks for Medical Image Segmentation”, *J Healthc Eng*, roč. 2022, 2022, doi: 10.1155/2022/9580991.

- [90] “1.17. Neural network models (supervised) — scikit-learn 1.2.2 documentation”. https://scikit-learn.org/stable/modules/neural_networks_supervised.html (cit 08. marec 2023).
- [91] A. Khazri, “Self Organizing Maps. (Kohonen’s maps) | by Achraf KHAZRI | Towards Data Science”. <https://towardsdatascience.com/self-organizing-maps-1b7d2a84e065> (cit 09. marec 2023).
- [92] “Auto-Encoder: What Is It? And What Is It Used For? (Part 1) | by Will Badr | Towards Data Science”. <https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726> (cit 08. marec 2023).
- [93] “Build a neural network in 7 steps | Neural Designer”. <https://www.neuraldesigner.com/learning/user-guide/design-a-neural-network> (cit 08. apríl 2023).
- [94] F. Padovani, B. Mairhörmann, P. Falter-Braun, J. Lengefeld, a K. M. Schmoller, “Cell-ACDC: a user-friendly toolset embedding state-of-the-art neural networks for segmentation, tracking and cell cycle annotations of live-cell imaging data”, *bioRxiv*, s. 2021.09.28.462199, okt. 2021, doi: 10.1101/2021.09.28.462199.
- [95] A. Sengur, Y. Akbulut, U. Budak, a Z. Comert, “White Blood Cell Classification Based on Shape and Deep Features”, *2019 International Conference on Artificial Intelligence and Data Processing Symposium, IDAP 2019*, sep. 2019, doi: 10.1109/IDAP.2019.8875945.
- [96] S. Mayala a J. B. Haugsøen, “Threshold estimation based on local minima for nucleus and cytoplasm segmentation”, *BMC Med Imaging*, roč. 22, č. 1, s. 1–12, dec. 2022, doi: 10.1186/S12880-022-00801-W/FIGURES/11.
- [97] E. P. Mandyartha, F. T. Anggraeny, F. Muttaqin, a F. A. Akbar, “Global and Adaptive Thresholding Technique for White Blood Cell Image Segmentation”, *J Phys Conf Ser*, roč. 1569, č. 2, s. 022054, júl. 2020, doi: 10.1088/1742-6596/1569/2/022054.
- [98] H. Chen *et al.*, “Accurate classification of white blood cells by coupling pre-trained ResNet and DenseNet with SCAM mechanism”, *BMC Bioinformatics*, roč. 23, č. 1, s. 1–20, dec. 2022, doi: 10.1186/S12859-022-04824-6/FIGURES/10.
- [99] A. Elen a M. Kamil Turan, “Classifying White Blood Cells Using Machine Learning Algorithms”, *Geliştirme Dergisi International Journal of Engineering Research and Development UMAGD*, roč. 11, č. 1, s. 141, 2019, doi: 10.29137/umagd.498372.

- [100] S. H. Rezaatfighi a H. Soltanian-Zadeh, “Automatic recognition of five types of white blood cells in peripheral blood”, *Computerized Medical Imaging and Graphics*, roč. 35, č. 4, s. 333–343, jún. 2011, doi: 10.1016/J.COMPMEIMAG.2011.01.003.
- [101] M. Mohamed, B. H. Far, B. Far, a A. Guaily, “An efficient technique for white blood cells nuclei automatic segmentation”, doi: 10.1109/ICSMC.2012.6377703.
- [102] Z. M. Kouzehkanan *et al.*, “A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm”, *Sci Rep*, roč. 12, č. 1, s. 1123, dec. 2022, doi: 10.1038/S41598-021-04426-X.
- [103] N. Berman, “Active Learning for Cell Imaging. Prioritize labeling of cell imaging... | by Nimrod Berman | Towards Data Science”, 31. júl 2022. <https://towardsdatascience.com/active-learning-for-cell-imaging-99fc7dcdbc82> (cit 08. marec 2023).

Prílohy

Zoznam príloh

Príloha A	Odkaz na zdrojový kód systému.....	93
Príloha B	Pamäťové médium	94

Príloha A | Odkaz na zdrojový kód systému

Zdrojový kód implementovaného systému je dostupný v Git repozitári:
https://github.com/lindajblahova/DiplomaThesis_GraphicalBioDataProcessing

Príloha B | Pamäťové médium

Priložené pamäťové médium obsahuje prácu v elektronickej podobe.